



University of Hormozgan

# Iranian Journal of Educational Research

Print ISSN: 1735 - 563X    Online ISSN: 2980 - 874X

Homepage: <http://ijer.hormozgan.ac.ir>



Educational and Behavioral  
Research Center

## The Impact of Demographic Characteristics on Response Styles in Personality Assessment: Examining the Role of Gender, Age, Education, and Their Interactions Using Multidimensional Item Response Theory Models

Hassan Karimi<sup>1</sup> , Asghar Minaei<sup>2</sup> , Ali Moghadamzade<sup>3</sup> 

1. Department of Psychology, Sav.C., Islamic Azad University, Saveh, Iran

2. Associate Professor, Department of Educational Measurement, Faculty of Psychology and Education, Allameh Tabatabaii, Tehran, Iran, [asgharminaei@gmail.com](mailto:asgharminaei@gmail.com)

3. Associate Professor, Department of Educational Measurement, Faculty of Psychology and Education, University of Tehran, Tehran, Iran

### Article Info

### ABSTRACT

**Article type:**

Research Article

**Article history:**

Received 2 Feb. 2025

Received in revised form 3

Apr. 2025

Accepted 15 Apr. 2025

Published online 01 Jun. 2025

**Keywords:**

Response styles,

Multidimensional question-answer theory,

Generalized multidimensional partial validity model

**Objective:** The present study was conducted with the aim of investigating the effect of demographic characteristics (gender, age and education level) on answering styles (extreme, moderate and agree) in personality questionnaires and using advanced models of multidimensional question-answer theory.

**Methods:** This descriptive-correlation study was conducted on the data of 17,994 participants who completed the Neo personality questionnaire. Three different models of multidimensional question-answer theory were used for data analysis, and finally, the generalized multidimensional partial validity model was selected as the best model. One-way and two-way analysis of variance was used to test the hypotheses.

**Results:** The findings showed that men have a significantly higher extreme and agreeable response style than women. With increasing age, the tendency to extreme response style increases and the tendency to moderate and agreeable styles decreases. Also, higher education levels were associated with more extreme response style and less agreeable response style. Furthermore, a significant interaction between gender and level of education was observed in predicting extreme response style.

**Conclusions:** The results of this research emphasize the importance of considering demographic characteristics in the process of designing, implementing and interpreting personality questionnaires. These findings can lead to increasing measurement accuracy and reducing biases caused by response styles.

**Cite this article:** Karimi, H., Minaei, A. & Moghadamzade, A. (2025). The Impact of demographic characteristics on response styles in personality assessment: examining the role of gender, age, education, and their interactions using multidimensional item response theory models. *Iranian Journal of Educational Research*, 4 (2), 1-16.

DOI: <https://doi.org/10.22034/4.2.1>



© The Author(s).

DOI: <https://doi.org/10.22034/4.2.1>

Publisher: University of Hormozgan.

## Introduction

In recent decades, the study of response styles in self-report tests has become one of the important topics of psychometrics. Response styles, as systematic patterns in responding independent of the content of questions, can threaten measurement accuracy and become a source of unwanted variance (Pellninger, 2017). Among the most important of these styles, extreme response style, agreeable response style, and moderate response style can be mentioned.

Demographic characteristics such as gender, age and education level play an important role in the formation of these styles. Researches have shown that these factors can affect the tendency of people to choose middle or extreme options and also the level of agreement with the questionnaire sentences (Wijters, Givens and Shilwartz, 2010). However, the results of existing studies are not uniform and the interactive role of these variables has been less investigated.

This is important because demographic differences in response styles may lead to measurement error; In such a way that people with the same level of a personality trait will receive different scores simply because of the different answering style. Typically, these response styles are considered sources of common method bias that should be controlled for (Charles Wegnambs, 2022).

The development of multidimensional question-answer theory models has made it possible to more accurately separate response styles from personality traits (Bolt and Johnson, 2009). Despite this, few studies have investigated the role of gender, age and education and their interaction in shaping response styles. In order to fill this gap, the present study analyzes the effect of demographic characteristics on the three main response styles using multidimensional models of question-answer theory.

## Material and Methods

The current research is a fundamental-applied study in terms of its purpose and descriptive-correlation in terms of data collection method. In this study, the relationships between the variables have been examined without manipulating them. Research data were collected from the responses of 17,994 participants to the short form (60 questions) Neo Personality Questionnaire. Participants answered the questionnaire voluntarily through an online platform. In order to ensure the quality of the data, duplicate and incomplete answers were removed from the final analysis.

## Research tool

The main tool of the research was the Neo Personality Questionnaire (NEO-FFI), which measures the five main dimensions of personality in the form of 60 items. Each dimension consists of 12 items, and responses are set on a five-point Likert scale from "strongly disagree" to "strongly agree" ( Paul Costa & McCrae, 1999 ).

In this research, three main response styles were investigated:

- Extreme response style: frequency of using options at the ends of the spectrum ("strongly disagree" and "strongly agree").
- Middle response style: the frequency of using the option in the middle of the spectrum ("I have no opinion").
- Agree response style: the frequency of using the agree options ("agree" and "strongly agree"), regardless of the content of the items.

## Analysis method

To measure the response styles more accurately, the E design matrix with multiple dimensions was used, in which 60 items, 5 response classes, 8 dimensions (5 personality dimensions and 3 response styles) and gamma parameters were considered. The loadings of agreeable response style were considered constant across items. To ensure the convergence of the model, the variance of all dimensions was considered equal to 1 and 2500 integration points were used in the estimation of the model. Also, to control the effect of reverse items, the answers were decoded in 27 reverse items of the questionnaire. This approach allowed simultaneous assessment of response styles and personality dimensions.

Inspired by the framework of models proposed by Henninger and Meisser (2020), three different models of question-answer theory were used:

1. Model I: generalized multidimensional partial validity with fixed scoring weights and estimated detection parameters.
2. The second model: the generalized stochastic threshold model.
3. The third model: the random threshold model

These three models have different complexity respectively and provide the possibility of comprehensive examination of response styles. The first model allows response styles and personality dimensions to be modeled simultaneously; A second model with random thresholds

models individual differences in response scale use; And the third model, as the simplest model, examines personality dimensions independently of response styles.

### **The focus of analysis in this article**

In this article, unlike the other authors' article that dealt with the relationship between response styles and personality dimensions (Kirimi et al., 1404), the focus is on demographic characteristics. In this way, four main questions were examined:

1. The difference in response styles between women and men.
2. The difference in response styles in different age groups.
3. The difference in response styles at different levels of education.
4. Interaction between gender and level of education in predicting response styles.

Fit indices such as deviance, AIC, BIC and CAIC were used to select the appropriate model. All analyzes were performed using the TAM software package in the R environment (Kiefer et al., 2017).

Each of the models used in this research examines response styles from a different perspective. As Henninger and Meisser (2019, 2020) have shown, the first model with fixed scoring weights and estimated detection parameters allows for a more detailed examination of the impact of response styles on different questions; A second model with random thresholds allows us to model individual differences in response scale use; And the third model, as a simpler version of the random threshold model, provides a tool for independent estimation of response styles and their comparison in different demographic groups.

The simultaneous use of these three models in the current research, which is considered a comprehensive approach, enables a deeper understanding of how response styles appear in different groups (based on gender, age, education and the interaction of gender and education).

**Table 1.** Fit models and specification of estimated parameters

Parameters	The first model	The second model	The third model
Questions threshold parameter	240	240	240
The slope parameter of the questions	77	120	0
Variance/correlation parameter	28	10	19
Diagnostic parameters	random	constant	Fixed and equal to 1
Personality dimensions	concurrent modeling	concurrent modeling	Independent modeling
Guess parameters	•	•	0
Regression parameters	•	•	0
Delta parameter	•	•	0
Model complexity	High	medium	low
Duration of analysis	4.4 days	5.1 days	2.9 hours
Date of analysis	2024-04-27	2024-04-04	2024-03-29

### **The first model: generalized multidimensional partial validity with fixed scoring weights and estimated diagnostic parameters<sup>1</sup>**

In this model, the design matrix E is used and the discrimination parameters of the items are considered randomly. This model allows us to simultaneously model response styles and personality dimensions and examine the influence of each on individuals' responses to items. As seen in (Table 1) the number of parameters, this model has 240 threshold parameters, 77 slope parameters and 28 variance/correlation parameters. This model is implemented using Quasi Monte Carlo integration method with 2500 integration points and 119 iterations.

### **The second model: the generalized stochastic threshold model<sup>2</sup>**

In this model, the design matrix E is also used, but the items' detection parameters are fixed. Also, the variance of personality dimensions is fixed to 1 and the correlations between different dimensions are assumed to be zero. This model has a simpler structure than the first model and is suitable for investigating the effect of response styles on people's responses to items. This model has 240 threshold parameters, 120 slope parameters and 10 variance/correlation parameters. The model is implemented using Quasi Monte Carlo integration method with 2500 integration points

<sup>1</sup> Generalized Multidimensional PCM with fixed scoring weights and estimated discrimination parameters (Falk & Cai, 2016)

<sup>2</sup>Generalized Random Threshold Model (adapted from Wang & Wu, 2011)

and 4000 iterations. Like the first model, this model is also implemented using Monte Carlo quasi-integration method with 2500 integration points and 4000 iterations.

### **The third model: random threshold model**

In this model, the design matrix B is used and the parameters of item detection are considered constant and equal to 1. Also, the variance of personality dimensions is fixed to 1 and the correlations between different dimensions are assumed to be zero. This model is the simplest IRT model used in this research and is suitable for examining personality dimensions independently of response styles. This model has 240 threshold parameters and 19 variance/correlation parameters and its fitting took only 2.9 hours.

These three models were compared with fit indices (Deviance, AIC, BIC, and CAIC) and the model that showed the best fit was used to test the research questions (difference in response styles in gender, age groups, education levels, and gender  $\times$  education interaction). All analyzes were performed using the TAM package in the R software environment (Kiefer et al., 2017).

Each of the models used in this research examines response styles from a different perspective. As Henninger and Meisser (2019, 2020) have shown, the first model with fixed scoring weights and estimated detection parameters allows for a more detailed examination of the impact of response styles on different questions; A second model with random thresholds allows us to model individual differences in response scale use; And the third model, as a simpler version of the random threshold model, provides a tool for independent estimation of response styles and their comparison in different demographic groups.

## **Results**

In order to answer the research questions, first three different models of multidimensional question-answer theory were fitted to select the most suitable model for data analysis. The comparison of these models showed that the generalized multidimensional partial validity model with fixed scoring weights and estimated detection parameters (the first model) has the best fit with the data. This model allows more accurate estimation of response styles (extreme, moderate and agree) by considering the random detection parameters for the items and hence it was chosen for the analysis of demographic differences.

**Table 2.** Fit indices of three models

Indicator	Description	The first model	The second model	The third model
Deviance	The value of the model deviation, which indicates the fit of the model.	2,774,822	2,813,269	2,931,279
Log Likelihood	The logarithm of the model's likelihood to fit the data.	-1,387,411	-1,406,635	-1,465,639
AIC	Akaike information criterion for model fit.	2,775,512	2,814,009	2,931,797
AIC3	A corrected version of AIC with the addition of parameter penalty.	2,775,857	2,814,379	2,932,056
BIC	Bayesian information criterion for model fit.	2,778,202	2,816,894	2,933,816
aBIC	Adjusted version of BIC with adjusted penalty.	2,777,106	2,815,718	2,932,993
CAIC	AIC proportional to the number of samples.	2,778,547	2,817,264	2,934,075
AICc	AIC corrected for sample size.	2,775,526	2,814,025	2,931,804
GHP	GHP log penalty for the number of parameters.	1.28539	1.30322	1.35777

Table 2 shows the fit indices of three models. As can be seen, the first model has lower values than the other two models in most indicators (Deviance, AIC, BIC, CAIC, etc.). For example, the value of Deviance in the first model (2,774,822) is lower than the second model (2,813,269) and the third model (2,931,279). Also, the AIC value in the first model (2,775,512) is lower than the second model (2,814,009) and the third model (2,931,797).

Based on this, the first model was chosen as the most appropriate model, and in the following, the findings of the research are reported based on this model. At this stage, the difference in response styles (extreme, moderate and agree) based on gender, age, level of education and the interaction between gender and education were investigated.

To answer the research questions, after selecting the best model (generalized multidimensional partial validity model), estimates related to three response styles (extreme, moderate, and agree) were extracted. Then, differences between demographic groups (gender, age, education level and gender  $\times$  education interaction) were analyzed using one-way analysis of variance test.

**Table 3.** ANOVA test results for gender differences in response styles

PR significance level (>F)	F value	mean square (MEANSQ)	sum of squares (SUMSQ)	degrees of freedom (DF)	Response style
< 2e-16*** <sup>3</sup>	115.2	94.51	95	1	ERS
0.0801	3.063	0.030270	0.03	1	MRS
1.19e-05***	19.19	12.86	13	1	ARS

As can be seen in Table 3, the results of the one-way analysis of variance test showed that the gender differences in extreme and agreeable response styles are significant, while no significant difference between women and men was observed in the middle response style. Therefore, it can be concluded that men have a greater tendency to use extreme options (totally disagree/totally agree) and favorable options (agree/totally agree) than women, but there is no significant difference between the two sexes in using the middle option.

To investigate the difference in response styles in different age groups, one-way analysis of variance (ANOVA) was performed for the three dimensions of extreme response style, moderate response style, and agreeable response style. The test results were significant for all three dimensions and showed that the response patterns are different in age groups (Table 4).

**Table 4.** ANOVA test results of response styles for age group

Response style	Source of changes	degree of freedom	sum of squares (SUMSQ)	mean square (meansq)	value (F)	Significant level PR(>F)
ERS	age group	3	286	95.27	117.6	<2e-16***
	The remainders	17990	14570	0.81		
MRS	age group	3	0.56	0.18533	18.8	3.58e-12***
	Remainders	17990	177.30	0.00986		
ARS	age group	3	71	23.808	35.69	<2e-16***
	Remainders	17990	12001	0.667		

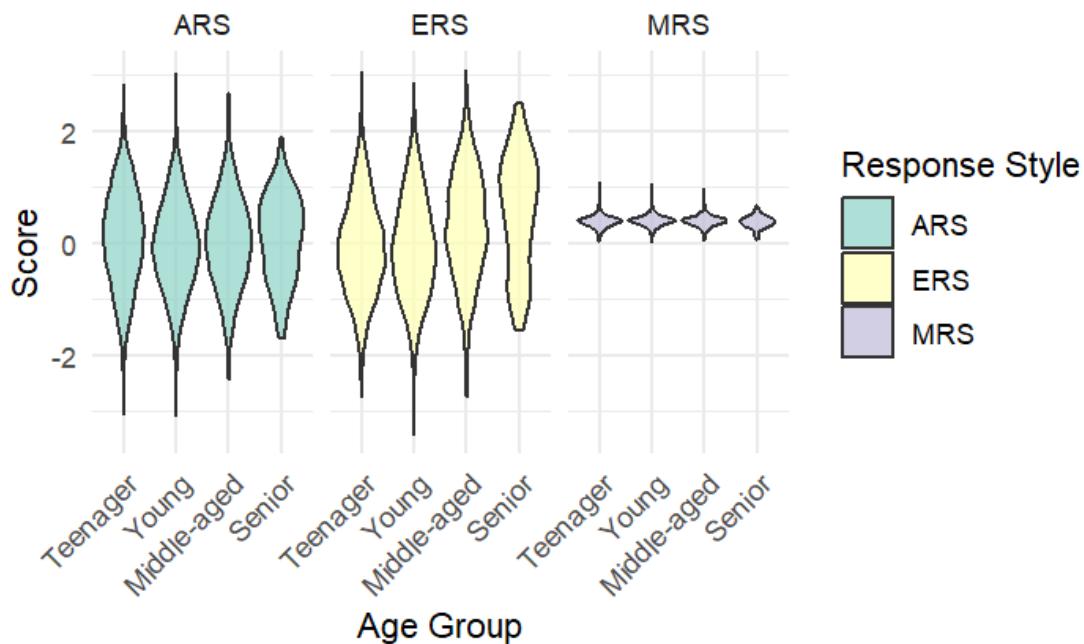
As seen in Table 4, the ANOVA test results showed that there is a significant difference between age groups in all three response styles. For the extreme response style, the value of  $F = 117.6$  and  $p < 0.001$ , for the medium response style, the value of  $F = 18.8$  and  $p < 0.001$ , and for the agreeable response style, the value of  $F = 35.69$  and  $p < 0.001$  was obtained. These results show that the differences between age groups in the use of extreme, moderate and agreeable options are significant.

<sup>3</sup> Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Also, further investigations using Tukey's post hoc test determined that in the extreme response style, the scores of middle-aged and elderly people were significantly higher than teenagers and young people, which shows that the tendency to choose extreme options increases with age.

In the agreeable response style, a significant difference was observed between some age groups. In particular, middle-aged and elderly people had lower scores than teenagers, but significant differences between other groups were less.

In the middle response style, young and middle-aged people had significantly lower scores than teenagers, but no significant differences were observed between the elderly and other groups. These results show that with increasing age, the tendency to choose middle options decreases.



**Figure 1.** Response styles in age groups

As seen in Figure 1, the trend of changes of all three response styles in different age groups is visible. This graph shows well the increasing pattern of the extreme response style and the decreasing pattern of the moderate and agreeable response style with increasing age, which was obtained in the statistical results.

One-way analysis of variance was used to investigate the difference in response styles at different levels of education. The studied educational levels included "sub-diploma", "diploma", "post-diploma", "bachelor's degree", "master's degree" and "doctorate". The results of this analysis are presented in Table 5.

**Table 6.** ANOVA test results of response styles for education level

Response style	Source of changes	degree of freedom	sum of squares (SUMSQ)	mean square (meansq)	value (F)	Significant level PR(>F)
ERS	Education level	5	288	57.64	71.18	<0.001
	The remainders	17988	14567	0.81		
MRS	Education level	5	0.38	0.07506	7.607	<0.001
	Remainders	17988	177.48	0.00987		
ARS	Education level	5	29	5.744	8.578	<0.001
	Remainders	17988	12044	0.670		

The results of the analysis of variance showed that the level of education had a statistically significant effect on all three response styles: extreme, moderate and agreeable. For the extreme response style, the F value obtained (71.18) with a significance level ( $p < 0.001$ ) indicates a strong and significant effect. In the case of the middle response style, although the F value (7.607) is smaller, it was still significant at the ( $p < 0.001$ ) level. And for the agreeable response style, the results showed an F value equal to (8.578) and a significance level ( $p < 0.001$ ), which indicates a significant effect of education on this response style.

Also, supplementary investigations using Tukey's post hoc test showed that the pattern of education's effect on each response style is different. In the extreme response style, a linear and significant increasing trend was observed, so that the scores of people with university degrees (especially master's and doctorate) were significantly higher than the scores of people with under-diploma education.

In contrast, for the middle response style, significant differences were observed mainly between lower education groups (sub-diploma and diploma), and at higher education levels, the differences were not statistically significant. Finally, for the affirmative response style, a general downward trend was seen, meaning that people with high education (bachelor's and higher) scored

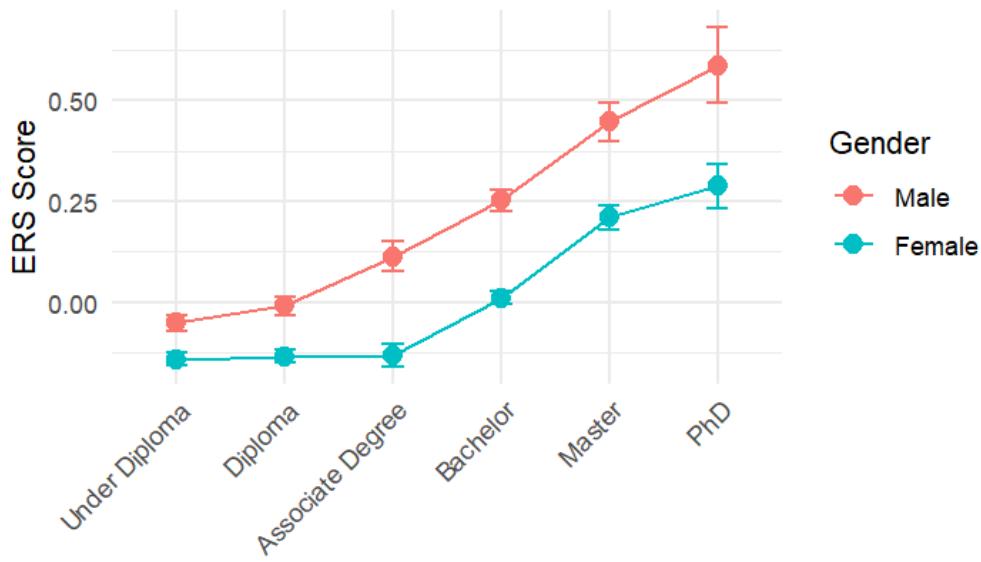
significantly lower than the under-diploma group, which indicates their less tendency to respond affirmatively.

In order to investigate the interactive effect of gender and education level on response styles, two-way analysis of variance was used. In this analysis, gender (male and female) and education level (six groups) were considered as independent variables and scores of three response styles were considered as dependent variables. The main goal was to investigate whether the effect of education level on response styles is the same for men and women. The results of this analysis for each response style are presented separately in Tables 7 to 9.

**Table 7.** ANOVA results for the interaction between gender and education in extreme response style

source of variance	degrees of freedom (DF)	sum of squares (SUM SQ)	mean square (MEAN SQ)	F value	P
gender	1	95	94.51	117.673	< 2e-16 ***
Education level	5	300	59.94	74.629	< 2e-16 ***
The interaction of gender and education	5	19	3.82	4.756	0.000241 ***
the remainder	17982	14442	0.80		

The results of two-way analysis of variance for extreme response style (Table 7) showed that in the case of extreme response style, both gender and education level had a significant effect on response styles, and the interaction between these two variables was also significant. This indicates that the effect of gender on response styles is different according to education level.



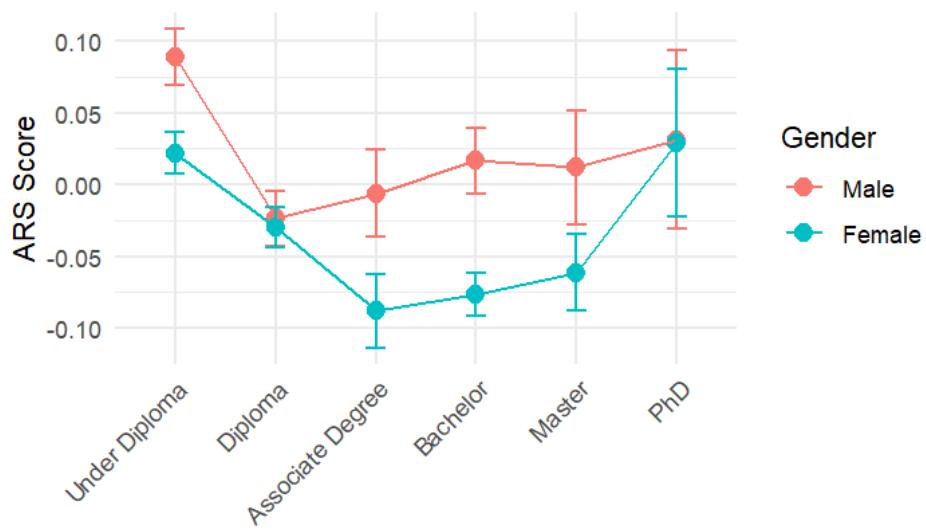
**Figure 2.** Interaction between gender and education in extreme response style

Figure 2 shows this interaction graphically. As can be seen in this graph, in general, men have higher extreme response style scores than women at all educational levels. However, the shape of the slope of the curves is different for the two sexes. It seems that with the increase in the level of education, the gap between the scores of men and women in this response style increases, which indicates a significant interaction between these two variables.

**Table 8.** ANOVA results for the interaction between gender and education on median response style

source of variance	degrees of freedom (DF)	sum of squares (SUM SQ)	mean square (MEAN SQ)	F value	p
gender	1	13	12.862	19.232	1.16e-05 ***
Education level	5	28	5.665	8.470	5.11e-08 ***
The interaction of gender and education	5	5	0.960	1.435	0.208
the remainder	17982	12027	0.669		

In the case of agreeable response style, the results are similar to the middle response style and the interaction between gender and education level is not significant, but the main effects of gender and education are significant. These results show that the effects of gender and education level on response styles may be different in different dimensions and in some cases the interaction between these two variables is effective.



**Figure 3.** Interaction between gender and education in agreeable response style

This finding shows that although women and men, as well as different educational groups, differ from each other in the score of agreeable response style, the pattern of influence of education is the same for both sexes. Figure 3 also shows this issue clearly, so that the trend of changes in the yes answer style scores at different educational levels for men and women is almost the same and parallel.

In summary, the findings showed that the interaction between gender and education was significant only for extreme response style. This means that in the interpretation of the extreme response style score, one should simultaneously pay attention to the gender and the level of education of the individual, because the effect of education on this response style is different for women and men. On the contrary, for two response styles, middle and agree, the effects of gender and education work independently and can be considered separately.

## Discussion

The present study investigated the effect of demographic characteristics on response styles in personality questionnaires by applying advanced models of multidimensional question-answer theory. The findings of this study, which is based on the data analysis of 17,994 participants, showed that gender, age and education level are significantly related to response styles.

The present study showed that men show more tendency towards extreme response style, which is in line with the findings of Kim and Bolt (2020). On the contrary, women obtained higher scores in medium and agreeable response styles, which is consistent with the results of van Wernberg (2013).

In terms of the effect of age, the findings of the current research showed that the tendency to respond excessively increases with age, which is in line with the studies of Greenleaf (1992) and Ramsted et al. (2010). Also, teenagers showed the greatest tendency to respond in the affirmative. Regarding the effect of education, higher levels of education were associated with more extreme responding, which is consistent with Winkler et al.'s (1982) findings. Also, higher education was associated with less agreeable responses, which is consistent with the results of Johnson et al. (2005) and Harzing (2006). However, it should be noted that Marin et al.'s (1992) study had conflicting results in this regard.

Identifying a significant interaction between gender and education in the extreme response style is one of the unique findings of this research, which is consistent with the results of Kim and Bolt (2020) and shows that the effect of education on this response style is different for men and women. This finding is consistent with the research results of Wetzel et al. (2016), who showed that the agreeable response style is more influenced by the main effects of demographic variables than their interactive effects.

### **Importance of findings and practical applications**

These findings emphasize the importance of considering demographic characteristics in the design, implementation and interpretation of personality questionnaires. Paying attention to these factors can lead to increasing the accuracy of measuring personality traits and reducing biases caused by response styles.

### **Suggestions for future studies**

It is suggested that more advanced models be used in future studies to better understand the underlying mechanisms of the effect of these variables on response styles. Also, examining these relationships in different cultural contexts and with different measurement tools can lead to more generalization of the findings.

### **Data availability statement**

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

### **Ethics statement**

The studies involving human participants were reviewed and approved by the ethics committee of Islamic Azad University. The patients/participants provided their written informed consent to participate in this study.

### **Author contributions**

All authors contributed to the study conception and design, material preparation, data collection, and analysis. All authors contributed to the article and approved the submitted version.

### **Funding**

The authors did (not) receive support from any organization for the submitted work.

### **Conflict of interest**

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## **References**

Bolt, D. M., & Johnson, T. R. (2009). Addressing score bias and differential item functioning due to individual differences in response styles. *Applied Psychological Measurement*, 33(4), 335-352.

Bolt, D. M., & Newton, J. R. (2011). Multiscale measurement of extreme response style. *Educational and Psychological Measurement*, 71, 814-833.

Costa, P. T., & McCrae, R. R. (1999). The Five-Factor Theory of Personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of Personality: Theory and Research* (pp. 139-153). New York: Guilford.

Falk, C. F., & Cai, L. (2016). A flexible full-information approach to the modeling of response styles. *Psychological Methods*, 21(3), 328-347.

Greenleaf, E. A. (1992). Measuring extreme response style. *Public Opinion Quarterly*, 56(3), 328-351.

Henninger, M., & Meiser, T. (2020). Different approaches to modeling response styles in divide-by-total item response theory models (part 2): Applications and novel extensions. *Psychological Methods*, 25(5), 577-595.

Johnson, T., Kulesa, P., Cho, Y. I., & Shavitt, S. (2005). The relation between culture and response styles: Evidence from 19 countries. *Journal of Cross-Cultural Psychology*, 36(2), 264-277.

Karimi, H., Minaei, A. & Moghadamzadeh, A. (2025). The relationship between five personality traits and response styles: A study of gender differences using multidimensional IRT models. *Journal of developmental and educational psychology*, Accepted.

Kiefer, T., Robitzsch, A., & Wu, M. (2017). TAM: Test analysis modules (Version 2.8-21) [Computer software].

Kim, N., & Bolt, D. M. (2021). A mixture IR Tree model for extreme response style: Accounting for response process uncertainty. *Educational and Psychological Measurement*, 81(4), 742-764.

Marin, G., Gamba, R. J., & Marin, B. V. (1992). Extreme response style and acquiescence among Hispanics: The role of acculturation and education. *Journal of Cross-Cultural Psychology*, 23(4), 498-509.

Plieninger, H. (2017). Mountain or molehill? A simulation study on the impact of response styles. *Educational and Psychological Measurement*, 77(1), 32-53.

Plieninger, H. (2021). Developing and applying IR-Tree models: Guidelines, caveats, and an extension to multiple groups. *Organizational Research Methods*, 24(3), 654-670.

Rammstedt, B., Goldberg, L. R., & Borg, I. (2010). The measurement equivalence of Big-Five factors' markers for persons with different levels of education. *Journal of Research in Personality*, 44(1), 53-61.

Scharl, A., & Gnambs, T. (2022). The impact of different methods to correct for response styles on the external validity of self-reports. *European Journal of Psychological Assessment*.

Wang, W. C., & Wu, S. L. (2011). The random-effect generalized rating scale model. *Journal of Educational Measurement*, 48(4), 441-456.

Watson, D. (1992). Correcting for acquiescent response bias in the absence of a balanced scale: An application to class consciousness. *Sociological Methods & Research*, 21(1), 52-88.

Wetzel, E., Lüdtke, O., Zettler, I., & Böhnke, J. R. (2016). The stability of extreme response style and acquiescence over 8 years. *Assessment*, 23(3), 279-291.

Winkler, J. D., Kanouse, D. E., & Ware, J. E. (1982). Controlling for acquiescence response set in scale development. *Journal of Applied Psychology*, 67(5), 555-561.