ISSN: 1309 – 6575
Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi
Journal of Measurement and Evaluation in Education and Psychology
Research Article; 2024; 15(1); 35-49

Comparison of Cluster Analysis and Latent Class Analysis for the Detection of Fake Responses on Personality Tests*

İbrahim ŞAHİN ** Seher YALÇIN ***

Abstract

Personality tests reveal the trait being measured, allowing test takers to present themselves differently than they really are. Research suggests that such deception in personality tests can have a negative impact on criterion-related validity. This study compared the effectiveness of cluster analysis and latent class analysis in detecting faking behavior in personality tests. A post-test control group design was used with 543 11th-grade students from eight different high schools in Sanliurfa province during the academic year 2021–2022. Participants in the experimental group were asked to respond in a specific way in order to score higher on the test, believing that their placement in the university depended on the result of the personality test indicating that they had a "positive" profile. Conversely, the control group was asked to present themselves truthfully and give honest answers. In this study, the initial focus was to assess the validity and reliability of the personality test scores. A comparison was then made between the scores of the participants in the experimental and control groups for each sub-dimension of the personality test to determine if there was a significant difference. The findings showed that there was a significant difference in the mean scores between the two groups, with the experimental group having a higher mean score. In addition, the results of cluster analysis and latent class analysis showed that latent class analysis outperformed cluster analysis in detecting fake respondents with a lower error rate.

Keywords: Fake responding, cluster analysis, latent class analysis

Introduction

The decision-making process involves gathering relevant information, comparing it with certain criteria, and reaching a conclusion. Consequently, decision-making can be considered an evaluative process (Turgut & Baykul, 2019). In various stages of education, some decisions need to be made. These decisions may be related to school management, teaching methods, curriculum, selection, placement, classification of individuals, or students' career goals (Thorndike & Thorndike-Christ, 2013). Evaluating data obtained through measurement processes plays a crucial role in determining the effectiveness of educational programs and methods, identifying students' learning deficiencies and achievements, and guiding them toward areas where they can be successful, considering their interests and abilities (Baykul, 2015).

The literature indicates that personality tests are frequently used in research, self-exploration, and clinical decision-making processes. Research purposes for using personality tests include measuring the effectiveness of treatment methods or interventions, helping individuals gain self-awareness under the guidance of a counselor, and making treatment decisions in clinical settings (Thorndike & Thorndike-Christ, 2013). In the field of measurement and evaluation, there has been a focus on examining the applicability of certain assessment tools used in student guidance services as adapted tests in computer-based environments. For example, a self-assessment inventory was employed in one study (Aybek &

*This study was presented at the VIII. International Congress on Measurement and Evaluation in Education and Psychology on September 21-23, 2022. Additionally, this study has been derived from a Master's Thesis conducted under the supervision of Assoc. Prof. Dr. Seher YALÇIN and prepared by İbrahim ŞAHİN.

** Teacher., Republic of Türkiye Ministry of National Education, Şanlıurfa-Türkiye, ibrahimsahinpdr@gmail.com, ORCID ID: 0000-0003-4972-3552
*** Assoc. Prof., Ankara University, Faculty of Education, Ankara- Türkiye, yalcins@ankara.edu.tr, ORCID ID: 0000-0003-0177-6727

To cite this article:
Şahin, İ., Yalçın, S. (2024). Comparison of cluster analysis and latent class analysis for the detection of fake responses on personality tests, *15*(1), 35-49. https://doi.org/10.21031/epod.1327395

Received: 14.07.2023
Accepted: 4.03.2024

Çıkrıkçı, 2018), while another study used the Skills Confidence Vocational Interest Inventory (Şimşek & Tavşancıl, 2022) to help students recognize their abilities, interests, and values.

Career counseling is one field where personality tests are widely used. Personality traits play a crucial role in various processes, including career choice, career planning, and job satisfaction. When individuals align their career choices with their personality traits, it has a positive impact on their productivity and job satisfaction (Pişkin, 2020). Holland (1973) viewed career choice as a reflection of personality and argued that just as individuals possess personality traits, different occupations also require specific personality traits. Holland categorized these personality traits into six types: realistic, investigative, artistic, social, enterprising, and conventional. According to Holland, individuals tend to gravitate toward professions that allow them to utilize their abilities, attitudes, and values that are consistent with their personality traits. Working in jobs that are compatible with one's personality traits can lead to occupational satisfaction. Törnroos et al. (2019) examined the relationship between personality traits and occupational satisfaction, and their findings are consistent with Holland's perspective. The research indicated that individuals in the same occupation share similar personality traits, and occupational satisfaction increases when there is a match between the average personality traits associated with an occupation and the individual's own personality traits. Moreover, certain personality traits have a greater impact on individuals' occupational choices. The influence of personality traits on job satisfaction and work efficiency emphasizes the importance of accurately measuring personality in career planning and occupational selection. This can be achieved by developing measurement tools that provide valid and reliable results while minimizing the occurrence of fake responses.

Self-report personality inventories have both advantages and disadvantages. On the one hand, individuals themselves are considered to be the best source of accurate information about their own personalities. On the other hand, there are weaknesses associated with this method, such as individuals' lack of sufficient self-knowledge or unwillingness to share certain information about themselves with others. These limitations have led to the need for alternative methods in personality measurement (Cohen & Swerdlik, 2009). In addition, cognitive factors such as inattention, rapid responding, etc., and response styles (such as always tending to give an intermediate response) are also associated with fake responding and inconsistent responding on the scales (Demetriou et al., 2015; Wetzel et al., 2016).

In maximum performance tests, individuals may attempt to give fake responses. In these tests, individuals only have the opportunity to present themselves as less successful than they are. However, in typical response tests, they can present themselves as either better or worse than they truly are (Mehrens & Lehmann, 1991). For instance, when taking an intelligence test, individuals are not expected to perform at a level higher than their current ability, excluding the guessing effect. However, this differs for typical behavioral tests such as personality tests, where individuals with extroverted personalities may intentionally present themselves as introverts. The use of personality tests to measure the suitability of applicants for a job has increased steadily. Meta-analytic studies conducted since the early 1990s have shown that personality tests have an unprecedented level of validity and predictability in personnel selection (Rothstein & Goffin, 2006).

Self-report personality tests operate under the assumption that test-takers will give honest responses, but this assumption may not always be possible. Many respondents may be unwilling to disclose the truth about themselves, even if they are aware of it (Kubinger, 2002). In the context of personnel selection, the use of personality tests is based on two fundamental assumptions. The first assumption is that the instrument effectively measures the intended trait. For instance, if an individual scores high on items measuring honesty in a personality test, it is assumed that he or she is honest in real life. The second assumption is that test scores can predict individuals' future performances. While there is evidence supporting these assumptions, there are valid reasons that remain skeptical about their real-world realization (Adair, 2014). Personality tests often give away the trait they are intended to measure, which allows test takers to present themselves as different from who they are. Research indicates that most job applicants tend to exaggerate their positive traits to increase their chances of being selected, and this deliberate distortion undermines criterion-related validity (Huber, 2017). A meta-analysis by Viswesvaran and Ones (1999) revealed that the personality scores of job applicants were 0.48–0.65 standard deviations higher than those of current employees.

Respondents may also tend to display certain socially acceptable characteristics, even if they do not genuinely possess them. They may prefer to provide responses that ensure social approval rather than reflecting their true views or personality traits (Mehrens & Lehmann, 1991). For example, an individual might respond to items related to freedom of expression in a personality inventory in a way that portrays them as supportive of freedom of expression, even if they are actually intolerant of differing opinions. Some respondents avoid extreme responses and instead opt for moderate responses, making it challenging to gather accurate information from such individuals (Kline, 1999). On the other hand, in personality research, some specifically developed scales were needed to examine the effects of social desirability (Erzen et al., 2021). Additionally, methods such as item response theory models can be used statistically to evaluate the agreement between observed responses and model-predicted responses. These analyzes evaluate how well the model fits the real data and whether this fit is meaningful (Embretson & Reise, 2000). Moreover, statistical methods such as latent profile analysis are also used in the literature to detect classes that react carelessly at the extreme (Maniaci & Rogge, 2014).

The literature highlights the correlation between personality traits and occupational satisfaction (Kang & Malvaso, 2023). When individuals work in jobs that are not aligned with their personality traits, it can have a negative impact on their professional satisfaction and subsequently reduce their productivity. Therefore, the act of faking responses in personality tests should be regarded as more than just an attempt to deceive; it can result in a waste of time and resources. Considering that the process of guiding individuals toward suitable professions begins in secondary education, placing students in university programs that align with their personality traits can lead to a more successful career journey. Consequently, high school students were chosen as the target group for this study. When examining studies conducted in the literature (Huber, 2017; Widhiarso & Himam, 2015; Yankov, 2019), it is seen that the results obtained from the normal process and the directed and encouraged fake responders are compared and that fake responders often have similar response patterns. This study was necessary because of the negative impact of fake responding behavior on the validity and reliability of the scores obtained from the measurement tool, the fact that personality tests are used in important decisions such as hiring individuals, and the importance of detecting intentional errors involved in the measurement process.

In this context, the research aims to address the following sub-objectives:

1. Is there a significant difference in the mean scores of students in experimental and control groups for each sub-dimension of the personality test?
2. To what extent can cluster analysis (CA) detect fake and honest responders in the administered personality test?
3. To what extent can latent class analysis (LCA) detect fake and honest responders in the administered personality test?

Method

A post-test control group design was used in this study. The participants consisted of 11th-grade students from eight different high schools in Sanliurfa during the 2021–2022 academic year. The participants were divided into two groups: the experimental group and the control group. Both groups were administered the Quick Big Five Personality Test (HBBKT). The experimental group was instructed to answer the inventory in a specific way that would present them as the most suitable candidates for admission to a university department, considering that their scores on the inventory would be evaluated for university admission. The control group, on the other hand, was informed that the results of the inventory would be used only for the purposes of a study and were asked to answer the inventory honestly, reflecting their true selves. Assuming that the experimental group gave fake responses and the control group gave honest responses, we examined how accurately the statistical analyses used could classify the respondents. The control group was informed that the results obtained from the inventory would only be used for research purposes and were asked to answer the inventory honestly, reflecting their true selves. No explanation was given to the control group about the experimental design of the study.

Participants
The total initial sample consisted of 705 students, with 363 students in the experimental group and 342 students in the control group. After eliminating data with missing values and extreme outliers during data cleaning, the final sample consisted of 266 students in the experimental group, 277 students in the control group, and 543 students in total. As stated by Dibao-Dina et al. (2014), statistical power is maximum in a sample of equal size. Therefore, the participants in the experimental and control groups were close to each other. Descriptive statistics including the distribution of the study participants by gender are shown in Table 1.

Table 1.
Descriptive Statistics of the Participants

		Experimental Group	Control Group	Total
Gender	Female	130	140	270
	Male	136	137	273
	Total	266	277	543

As seen in Table 1, the distribution of students by gender is relatively similar in both the experimental and control groups. The mean age of the experimental group was 16.1 years with a standard deviation (SD) of 0.36, while the mean age of the control group was slightly higher at 16.38 years with an SD of 0.60. Overall, when considering both groups together, the mean age of the entire study group was 16.24 years with an SD of 0.52.

Data Collection Tool
The Quick Big Five Personality Test (HBBKT) was used in this study. The test, developed by Vermulst and Gerris (2005) and adapted to Turkish by Morsünbül (2014), is based on the Five Factor Theory of Personality. It measures five personality traits: Extraversion, Agreeableness, Emotional Stability, Conscientiousness, and Openness to Experience. The test consists of 30 items, with six items measuring each personality factor. Each item was scored on a seven-point scale. The criterion-related validity of the adapted test was established by examining its relationship with self-concept salience, depression, anxiety, and life satisfaction. The internal consistency reliability of the sub-dimensions of the adapted test ranged from 0.71 to .81, and the test-retest correlation coefficients ranged from 0.80 to .87. In their study, Kutlu and Pamuk (2017) used the adapted test in a Turkish sample of 285 students, reporting Cronbach's alpha values ranging from .69 to .81. Rassart et al (2013) applied the test in a Belgian sample consisting of 366 participants aged between 15 and 20 years, reporting Cronbach's alpha values ranging between .75 and .90. Van der Linden et al. (2010) applied the test to a Dutch sample (mean age 14 years and 10 months) and reported acceptable model-data fit with the following statistics: χ^2=29.24, df=11, NNFI=.92, CFI=.98, RSMEA=.06. They also reported Cronbach's alpha values between .66 and .83.

Data Collection and Analysis
The necessary permissions from the Ministry of National Education and the ethics committee were obtained for the study. The data were collected in November 2021, and the test-retest application was carried out in April 2022. Before implementation, the purpose of the study was explained to school administrators, students, and parents, and informed consent was obtained. We began to prepare the data for analysis by examining missing data. In dealing with nonrandom missing data, it is recommended to delete missing data (Büyüköztürk et al., 2020). For this reason, in this study, data belonging to 60 participants were removed from the dataset as they contained nonrandom missing data.
The second step was to look for outliers. In order to detect univariate outliers, the raw scores in the dataset were converted into standard z-scores. In large samples (n>100), the z-range is accepted as "-4, +4" (Büyüköztürk et al., 2020). In this context, the data of 10 participants with univariate outliers were excluded from the study. In detecting multivariate outliers, the Mahalanobis D^2 statistic is used (Hair et al., 2014). Data from 40 participants with multivariate outliers were deleted. These procedures left data from 266 participants in the experimental group and 277 participants in the control group.

In the third step, the distributions of the data belonging to the experimental and control groups were examined. To do this, the experimental and control groups were considered separately, and the mean, mode, median, standard deviation, kurtosis, and skewness values of each item were checked. Table 2 shows the item statistics for the experimental and control groups.

Table 2.
Item Statistics for Experimental and Control Groups

	Experimental Group					Control Group				
Item	Mode	Median	Mean	Skewness	Kurtosis	Mode	Median	Mean	Skewness	Kurtosis
I 1	4	4	4.25	-0.06	-0.58	6	6	5.84	-0.71	-0.15
I 2	7	6	5.69	-0.48	-0.56	4	4	4.18	0.01	-0.89
I 3	4	4	4.41	-0.21	-0.42	4	4	4.15	0.02	-0.97
I 4	7	5	5.10	-0.55	-0.41	3	4	3.82	0.36	-0.50
I 5	5	5	4.86	-0.53	-0.35	6	6	5.56	-0.52	-0.66
I 6	7	6	6.07	-0.69	-0.34	7	6	5.63	-0.80	0.10
I 7	7	6	5.43	-0.68	-0.29	3	3	3.56	0.16	-0.72
I 8	6	5	5.23	-0.63	-0.26	6	5	5.03	-0.78	.031
I 9	5	5	4.98	-0.62	-0.12	4	4	4.01	0.03	-1.01
I 10	5	5	4.94	-0.72	-0.11	7	6	6.08	-0.73	-0.33
I 11	6	5	5.11	-0.73	0.00	3	4	3.87	0.24	-0.81
I 12	6	6	5.23	-0.77	0.00	6	5	4.82	-0.51	-0.35
I 13	5	5	5.08	-0.73	0.09	4	4	4.04	0.03	-0.99
I 14	5	5	4.95	-0.60	0.16	5	5	5.34	-0.24	-0.31
I 15	6	5	5.17	-0.82	0.16	6	6	5.71	-0.73	0.08
I 16	7	6	6.27	-0.94	0.36	1	3	2.75	0.67	-0.51
I 17	7	6	5.59	-0.96	0.48	4	5	4.64	-0.48	-0.43
I 18	6	6	6.07	-0.93	0.63	5	5	4.83	-0.46	-0.46
I 19	6	6	5.43	-1.06	0.67	7	6	5.52	-0.55	-0.43
I 20	7	6	6.07	-1.00	0.70	7	5	4.82	-0.57	-0.73
I 21	6	6	5.91	-1.06	1.16	3	4	4.04	0.09	-0.96
I 22	7	6	5.77	-1.26	1.31	6	6	5.66	-0.78	0.24
I 23	7	6	6.06	-1.26	1.41	4	4	4.39	-0.24	-0.81
I 24	7	6	5.85	-1.23	1.49	7	6	5.58	-0.73	-0.72
I 25	6	6	5.66	-1.21	1.56	5	5	4.79	-0.48	-0.56
I 26	7	6	6.05	-1.33	1.62	7	5	4.48	-0.14	-1.02
I 27	7	6	6.24	-1.19	1.62	5	5	4.67	-0.56	-0.17
I 28	6	6	5.63	-1.29	1.88	7	6	6.09	-0.94	0.37
I 29	7	6	6.11	-1.53	2.83	1	3	3.23	0.43	-0.66
I 30	7	7	6.25	-2.06	5.77	7	6	5.60	-0.49	-0.64

Analyzing Table 2 separately for the experimental and control groups, it can be seen that the mode, median, and mean values for most items are either equal or closely similar between the two groups. Additionally, the skewness and kurtosis values of all items in the control group fall within the range of ±1, except for the 19th and 28th items in the experimental group, which have skewness and kurtosis values within the range of ±2. Considering the instructions given to the participants in the experimental group, it was expected that this group would have higher scores than the control group. Therefore, these findings are consistent with the objectives of the study. Items 29 and 30 have kurtosis values of 2.83 and 5.77, respectively. It shows that experimental group members gave extreme responses to these items. It may be interpreted as meaning that the students in the experimental group thought that the most important characteristics they should have to be accepted into the university program were the characteristics represented by these items. As a result, the mode, median, and mean values in this group approach 7, indicating a departure from the normal distribution. This suggests that the students followed the given instructions appropriately. In contrast, the data from the control group show a distribution closer to the normal distribution compared with the experimental group, supporting the assumption that the students in the control group gave honest answers in accordance with the instructions given. Considering the data as a whole, it was concluded that the normality assumption was met, allowing the data to be analyzed without any intervention. In addition, the assumption of multivariate normality was examined with Bartlett's Test of Sphericity, and it was concluded that the test result was significant; that is, this assumption was met.

Before the LCA and the CA were carried out, the validity and reliability of the measurement tool were assessed. Table 3 shows the Cronbach's Alpha reliability coefficients for the experimental and control groups.

Table 3.
Cronbach's Alpha and McDonald’s Omega Reliability Coefficients for Experimental and Control Groups

	Cronbach's Alpha		McDonald’s Omega	
	Experimental Group	Control Group	Experimental Group	Control Group
Agreeableness	.71	.62	.73	.68
Extraversion	.71	.81	.72	.82
Conscientiousness	.79	.80	.80	.81
Emotional Stability	.74	.67	.75	.69
Openness to Experience	.65	.63	.67	.67
Entire Test	.86	.80	.86	.81

When interpreting the calculated Cronbach's Alpha value to assess internal consistency, R. B. Kline (2005) suggests that values of 0.70 and above are considered 'acceptable', .80 and above are considered 'very good’, and .90 and above are considered 'excellent'. Additionally, Hair et al. (2014) mentioned that values of 0.60 and above may be acceptable if there is evidence of good construct validity. Nunnally & Bernstein (1994) suggested that McDonald’s omega coefficient can be interpreted like Cronbach’s alpha, and values above .70 can be considered acceptable. Upon reviewing Table 3, it can be seen that the omega and alpha coefficients of each sub-dimension are close to each other, and all sub-dimensions have reliability coefficients within the acceptable range. Besides, since the reliability coefficients of the control group scores were lower on some subscales, a test-retest method was employed to reinforce the reliability assessment. The first phase of the test-retest was conducted on April 6, 2022, followed by the second phase on April 19, 2022, at Sanliurfa Social Sciences High School, with 39 students participating. The results of the test-retest study are given in Table 4.

Table 4.
Reliability Coefficients for Test-Retest Application

Sub-Dimension	r
Agreeableness	.63
Extraversion	.83
Conscientiousness	.80
Emotional Stability	.74
Openness to Experience	.76

When analyzing Table 4, it can be concluded that there is a strong correlation between the first and second administrations in the sub-dimensions, with the exception of the Agreeableness sub-dimension, where a moderate relationship is observed. Confirmatory factor analysis (CFA) was conducted to verify the original factor structure and assess the measurement tool's construct validity. The software used for CFA was LISREL 8.7, utilizing Maximum Likelihood (ML) Estimation as the estimation method. Prior to the analysis, the dataset was prepared by removing missing data and outliers. The data were then divided into experimental and control groups, and CFA was performed separately for each group. The goodness of fit of the CFA model was assessed based on the $\chi2$/sd, CFI, RMSEA, and SRMR values. The results of the goodness of fit statistics obtained from the analysis are presented in Table 5.

Table 5.
Confirmatory Factor Analysis Statistics

	χ^2	df	χ^2/df	CFI	RMSEA	SRMR
Experimental Group	800.31	395	2.02*	.93	.062	0.06
Control Group	967.19	395	2.45*	.86	.072	0.07

*p<.001

The first statistic used to assess the model-data fit is the chi-square test. If the chi-square test is not significant, it suggests a good model-data fit. However, this test tends to become significant as the sample size increases (Hair et al., 2014). Therefore, the ratio of the chi-square value to the degrees of freedom, denoted as $\chi 2/df$, can be used as an indicator of model-data fit. A $\chi 2/df$ ratio of 3 or lower indicates a good fit, while a value between 3 and 5 indicates an adequate fit (Sümer, 2000). Examining Table 5, it can be seen that the chi-square tests for both experimental and control groups are significant, but their respective $\chi 2/df$ values are less than 3. This finding indicates a good model-data fit. Another measure used to assess the fit is the Comparative Fit Index (CFI), which ranges from 0 to 1. A CFI value close to 1 indicates a good fit. CFI values of 0.90 or higher are considered acceptable for model-data fit (Westland, 2019). The CFI coefficient of the experimental group exceeded the acceptable level, whereas the CFI value of the control group was close to the acceptable level. Furthermore, a root means square error of approximation (RMSEA) value of 0.05 or lower indicates a good model-data fit (Schumacker & Lomax, 2004). Browne and Cudeck (1993, as cited in Keith, 2015) suggested that RMSEA values of 0.08 or lower are acceptable, whereas values of 0.10 or higher indicate poor model-data fit. In this study, the RMSEA values for both experimental and control groups are within an acceptable range. The SRMR value is interpreted in the same way as RMSEA; therefore, according to SRMR, it can be stated that the model-data fit of both groups is at an acceptable level.

For the first sub-objective of the study, an independent samples t-test was conducted to examine whether there was a significant difference between the participants' scores in the experimental and control groups for each sub-dimension of the personality test. The means of both groups were compared, and the significance of the mean differences was assessed. Additionally, the eta-square effect size was calculated for the significant findings.

For the second and third sub-objectives of the study, the effectiveness of CA and LCA in identifying fake respondents was assessed. Clusters and latent classes obtained from each analysis were named based on the available data, and then the accuracy rates of the analyses were calculated. The correct classification rate is determined by dividing the number of subjects classified as true negative and true positive by the total number of subjects, multiplied by 100 (Hair et al., 2014). In this study, classification accuracies were calculated by dividing the total number of correctly classified participants by the total number of participants.

CA and LCA

CA is a method used to categorize objects based on predetermined criteria, with the goal of identifying the highest similarity within objects and the greatest differentiation between categories. These objects can be respondents to a test, products, or other items under investigation (Hair et al., 2014). In this study, the clustering analysis used the two-step method, which was determined to be suitable for the dataset using SPSS software. The two-step method is designed for large datasets with a predetermined number of clusters and combines hierarchical and nonhierarchical CA techniques (Everitt et al., 2011).

LCA, on the other hand, is a statistical approach that aims to classify individuals into homogeneous subgroups based on their observable response patterns to a series of measurement tools (Geiser, 2013). These latent classes represent unobservable subgroups, where individuals within each subgroup share certain characteristics but differ significantly from individuals in other subgroups (Vermunt & Magidson, 2005). Traditional LCA is similar to CA in that it seeks to identify homogeneous subgroups within a heterogeneous population, often referred to as latent class CA (Vermunt & Magidson, 2002). The data were analyzed using SPSS and Latent Gold software packages.

Results

Results of the Comparison of Scores Achieved by Participants in the Experimental and Control Groups on the Sub-Dimensions of the Measurement Instrument

An independent samples t-test was carried out to assess whether there was a significant difference in the scores obtained by the participants in the experimental and control groups on the sub-dimensions of the measurement tool. The findings of the independent samples t-test are given in Table 6.

Table 6.
Independent Samples T-Test Findings

	Experimental Group (N=266)		Control Group (N=277)		t	Effect Size (η^2)
	Mean	Sd	Mean	Sd		
Agreeableness	36.04	3.97	33.92	4.44	5.85**	.060
Extraversion	29.12	5.93	25.24	7.69	6.58**	.074
Conscientiousness	33.39	5.81	28.10	7.05	9.51**	.143
Emotional Stability	31.47	5.94	23.16	6.27	15.82**	.316
Openness to Experience	35.44	4.17	32.32	4.72	8.15**	.109

**p<.001

When examining Table 6, it is clear that the independent samples t-test conducted for each sub-dimension shows statistically significant results. There was a significant difference in favor of the experimental group across all sub-dimensions. In other words, participants in the experimental group scored higher than the control group on all sub-dimensions. Upon analyzing the effect size values, it can be inferred that the differences in mean scores resulting from group membership are moderate in the sub-dimensions of Agreeableness, Extraversion, and Openness to Experience, whereas they are high in the sub-dimensions of Conscientiousness and Emotional Stability. This indicates that participants in the experimental group portrayed themselves as individuals with more positive traits, aligning with the study objectives. In other words, it shows that the students fulfilled what they were told in the experimental design and that the experimental procedure was effective.

Findings Regarding CA

Within the scope of the second sub-objective of the study, the participants were divided into two groups using cluster analysis. Participants' responses to the test items were used as input for grouping. Since the study consisted of experimental and control groups, the analysis was limited to two groups.

The clusters formed after the analysis were initially labeled as K1 and K2. Separate examinations were made for each sub-dimension, and the groups were named. In this study, the actual group membership of each individual in the clusters is known by the researchers. Therefore, these groups can be named by considering which of the experimental or control groups the majority of individuals in the clusters formed by the analysis are from. It can be said that the new group, consisting mostly of individuals from the experimental group, represents the experimental group, and the other group represents the control group. However, in real life, it remains unclear to which of the groups (fake or honest respondents) the participants belong. Thus, we tried to identify which of the clusters formed by the analysis represents the experimental group and which represents the control group by using information other than the actual group memberships of the individuals. This was done by initially analyzing the size of the clusters. The number of participants in the clusters formed by the analysis and the number of participants in the actual experimental and control groups are summarized in Table 7.

Table 7.
Cluster Sizes Generated by CA

		CA Results		
		K1	K2	Total
Agreeableness	Experimental Group	48	218	266
	Control Group	86	191	277
	Total	134	409	543
Extraversion	Experimental Group	105	161	266
	Control Group	186	91	277
	Total	291	252	543
Conscientiousness	Experimental Group	187	79	266
	Control Group	96	181	277
	Total	283	260	543
Emotional Stability	Experimental Group	191	75	266
	Control Group	50	227	277
	Total	241	302	543
Openness to Experience	Experimental Group	189	77	266
	Control Group	124	153	277
	Total	313	230	543

As shown in Table 7, the sub-dimension of "Conscientiousness" demonstrates the highest similarity between the sizes of the clusters formed during the analysis and the actual group sizes, whereas the sub-dimension of "Agreeableness" exhibits the greatest differentiation. It can be inferred that the Agreeableness sub-dimension has the lowest classification accuracy, even without cluster labeling. Two possible scenarios can arise from this observation. Assuming K1 as the experimental group and K2 as the control group for the Agreeableness sub-dimension, the analysis indicates a higher type two error rate, and vice versa, a higher type one error rate.
Upon examining the dataset, the clusters generated by the analysis for each sub-dimension and the matching rates between the actual experimental and control groups were analyzed. Consequently, it was determined that K1 corresponds to the experimental group and K2 corresponds to the control group for the "Agreeableness" and "Extraversion" sub-dimensions, while the opposite was true for the remaining sub-dimensions. After naming the clusters, the goodness of fit was assessed for each sub-dimension using the chi-square test, and the accurate classification rate was calculated. The reconstructed distribution table, along with the classification accuracy rate and chi-square test findings for each sub-dimension, are given in Table 8.

Table 8.
Classification Accuracy Table regarding CA

		CA Results			Classification Accuracy (%)	Chi-sq Test	
		Exp.	Control	Total		χ^2	df
Agreeableness	Exp.	218	48	266	55.99	12.34**	1
	Control	191	86	277			
	Total	409	134	543			
Extraversion	Exp.	161	105	266	63.90	41.78**	1
	Control	91	186	277			
	Total	252	291	543			
Conscientiousness	Exp.	187	79	266	67.77	69.08**	1
	Control	96	181	277			
	Total	283	260	543			
Emotional Stability	Exp.	191	75	266	76.97	158.84**	1
	Control	50	227	277			
	Total	241	302	543			
Openness to Experience	Exp.	189	77	266	62.98	38.40**	1
	Control	124	153	277			
	Total	313	230	543			

**p<.001

Table 8 shows that the Emotional Stability sub-dimension achieved the highest accurate classification rate in the cluster analysis, with a rate of 76.9%. On the other hand, the Agreeableness sub-dimension had the lowest accurate rate of classification at 55.9%.
The variation in classification accuracy across sub-dimensions can be attributed to several factors. This disparity may stem from the underlying mathematical principles of the analysis itself and potential inconsistencies in participants' adherence to the provided instructions. Even if some participants provided appropriate responses, they might have been assigned to an incorrect cluster. For instance, an individual who genuinely possessed more positive traits and was instructed to respond honestly could have been misclassified as a fake respondent.

Findings Regarding LCA

Within the scope of the third sub-objective of the study, LCA was used to categorize the participants into fake and honest respondent groups based on their responses to the test. A similar approach was adopted as in CA. Initially, the classes generated by the analysis were labeled as S1 and S2. Subsequently, the data were analyzed to determine which class represented the experimental group and which represented the control group. The accuracy of this determination was then confirmed by comparison with the existing dataset. The sizes of the classes formed by the LCA for each sub-dimension are presented in Table 9, which provides a comparison with the existing experimental and control groups.

Table 9.
Class Sizes Generated by LCA

		LCA Results		
		S1	S2	Total
Agreeableness	Exp.	266	0	266
	Control	63	214	277
	Total	329	214	543
Extraversion	Exp.	193	73	266
	Control	120	157	277
	Total	313	230	543
Conscientiousness	Exp.	86	180	266
	Control	184	93	277
	Total	270	273	543
Emotional Stability	Exp.	209	57	266
	Control	73	204	277
	Total	282	261	543
Openness to Experience	Exp.	104	162	266
	Control	192	85	277
	Total	296	247	543

Upon analyzing Table 9, it becomes clear that the results obtained from LCA closely align with the actual group sizes in the Consciousness and Emotional Stability sub-dimensions. Specifically, in the Conscientiousness sub-dimension, all participants from the experimental group were assigned to the S1 class. This observation without explicitly labeling the latent classes may indicate a high level of accurate classification or possibly suggest the opposite scenario. To gain further insights, the dataset was examined, latent classes were labeled, and their correspondence with the experimental and control groups was comparatively tabulated. Classification accuracy rates were calculated for each sub-dimension, and a chi-square test was conducted. These findings are presented in Table 10.

Table 10.
Classification Accuracy Table regarding LCA

		LCA Results			Classification	Chi-sq Test	
		Exp.	Control	Total	Accuracy (%)	χ^2	df
Agreeableness	Exp.	266	0	266	88.40	339.17**	1
	Control	63	214	277			
	Total	329	214	543			
Extraversion	Exp.	193	73	266	64.45	47.50**	1
	Control	120	157	277			
	Total	313	230	543			
Conscientiousness	Exp.	180	86	266	67.03	63.09**	1
	Control	93	184	277			
	Total	273	270	543			
Emotional Stability	Exp.	209	57	266	76.05	148.22**	1
	Control	73	204	277			
	Total	282	261	543			
Openness to Experience	Exp.	162	104	266	65.19	49.96**	1
	Control	85	192	277			
	Total	247	296	543			

**$p<.001$

Table 10 shows that LCA achieved the highest classification accuracy in the Agreeableness sub-dimension. It accurately classified 88.40% of the participants within this sub-dimension. Furthermore, in the actual application, all participants from the experimental group were correctly classified into the experimental group. The relatively lower rates of the correct classification in other sub-dimensions may be due to inconsistent response patterns among students or the characteristics of the measurement tool employed. Particularly in the Openness to Experience sub-dimension, the presence of inconsistent responses from students in both the experimental and control latent classes may have led to decreased classification accuracy.

Comparison of CA and LCA

The classification accuracy rates of LCA and CA, as applied for the purposes of this study, are comparatively presented in Table 11.

Table 11.

Classification Accuracy Rates of and LCA

	Classification Accuracy Rate	
	CA (%)	LCA (%)
Agreeableness	55.99	88.40
Extraversion	63.90	64.45
Conscientiousness	67.77	67.03
Emotional Stability	76.97	76.05
Openness to Experience	62.98	65.19

Upon reviewing Table 11, it is evident that LCA achieves a higher accuracy rate for classification in the Agreeableness sub-dimension. In addition, it achieves a nearly equal correct classification rate in the Conscientiousness and Emotional Stability sub-dimensions. CA has the highest accurate classification rate of 76.97% in the Emotional Stability sub-dimension, while its lowest accuracy rate was recorded in the Agreeableness sub-dimension at 55.99%. On the other hand, LCA achieves its highest level of accurate classification rate in the Agreeableness sub-dimension with a rate of 88.4%, while its lowest level is in the Extraversion sub-dimension with a rate of 64.45%. For each analysis, false positive and false negative rates were calculated for each sub-dimension. These rates are shown in Table 12.

Table 12.

False Positive and False Negative Rates for CA and LCA

	CA		LCA	
	False Positive (%)	False Negative (%)	False Positive (%)	False Negative (%)
Agreeableness	68.95	18.05	22.47	0
Extraversion	32.85	39.47	43.32	27.44
Conscientiousness	34.66	29.70	33.57	32.33
Emotional Stability	18.05	28.20	26.35	21.43
Openness to Experience	44.77	28.95	30.69	39.10

Upon examining the false positive and false negative rates of the analyses, it is evident that both analyses exhibit a higher tendency towards false positive. However, in the Extraversion and Emotional Stability sub-dimensions, CA exhibits a higher false negative classification rate. A comparable pattern can be seen in LCA. Here, the false positive classification rate is higher than the false negative classification rate, apart from in the Openness to Experience sub-dimension. These findings suggest that both analyses are more likely to misclassify honest respondents as fake respondents rather than including fake respondents in the honest respondent category.

Discussion and Conclusion

The study initially analyzed the reliability and validity levels of scores derived from the personality test taken by two groups: the experimental group consisting of fake respondents and the control group consisting of honest respondents. Both groups' internal consistency levels were deemed acceptable. Additionally, a test-retest application conducted on the control group revealed moderate stability in the Agreeableness sub-dimension and high stability in the remaining sub-dimensions. The confirmatory factor analysis conducted to assess construct validity yielded goodness-of-fit values that were close to or above the acceptable thresholds. Thus, the construct validity of the scores obtained from the personality test was supported for the study group.

Significant differences were found in the mean scores of the participants between the experimental group and the control group in the personality test, favoring the experimental group. Upon analyzing the effect size values, it was determined that the level of differences in mean scores resulting from group membership was moderate in the Agreeableness, Extraversion, and Openness to Experience sub-dimensions and high in the Conscientiousness and Emotional Stability sub-dimensions.

When assessing the capacity of CA and LCA to detect fake respondents in the personality test, it was found that LCA exhibited higher classification accuracy in the Agreeableness, Extraversion, and Openness to Experience sub-dimensions, while achieving an equivalent level of accuracy in the Conscientiousness and Emotional Stability sub-dimensions. Consequently, the findings of this study suggest that LCA performed better than CA in detecting fake respondents in personality tests. The divergence in results between the two analyses may be attributed to the mathematical foundations underlying the analyses or the response patterns of the students. This study aligns with the study conducted by Widhiarso and Himam (2015), which examined the detection of fake respondents by CA and LCA. Both studies indicated that CA had a higher frequency of type one errors, while LCA demonstrated higher classification accuracy. Widhiarso and Himam reported classification accuracy rates of 51% to 65% for CA and 55% to 67% for LCA, whereas the current study achieved classification accuracy ranging from 56% to 77% for CA and 65% to 88% for LCA. Thus, the two studies are consistent in terms of which analysis type had higher type one errors and higher classification accuracy. The disparity in classification accuracy levels may be attributed to variations in the study group or the measurement tool used.

Compared to Widhiarso and Himam’s (2015) research with a similar objective, this study exhibited higher classification accuracy values in CA. While the prior study achieved its highest classification accuracy in the Openness to Experience sub-dimension, the current study attained the highest accuracy in the Emotional Stability sub-dimension. Both studies consistently indicate that relying solely on CA for the detection of fake respondents is insufficient.

As with CA, LCA yielded higher classification accuracy values compared to the study conducted by Widhiarso and Himam (2015). The prior study reported classification accuracy ranging from 55% to 68%, while the current study achieved values between 65% and 88%. This finding aligns with the outcomes of a study conducted by Magidson and Vermunt (2002) on simulation data with known group memberships, demonstrating that LCA exhibited higher classification accuracy. Given the higher classification accuracy of LCA in the present study, it can be inferred that the findings of both studies are consistent with each other.

Both CA and LCA tend to produce more type one errors than type two errors. In other words, they are more likely to misclassify honest respondents as fake responders. This aspect should be considered during the evaluation process. Additionally, both analyses tend to label individuals with higher mean scores as fake respondents. It is important to keep in mind that individuals with genuinely positive characteristics may be mistakenly labeled as fake respondents by these analyses. Tabachnick and Fidell (2013) argued that the outcomes resulting from type one and type two errors may vary depending on the research objective. In this study, LCA and CA are not considered as methods that can detect fake respondents with complete accuracy but as one of methods that can be used to detect these respondents. Assigning someone who is actually a fake respondent to the honest category by the analysis may lead to this individual not being checked for fake responding. On the other hand, a higher rate of type one error would result in further assessments of individuals who are actually honest respondents, leading to a waste of time and effort. Consequently, it is preferable to have a lower rate of type two errors in this study. Practitioners should consider both situations when making decisions. Furthermore, while it is commonly assumed that individuals' responses to paper-and-pencil measurements are honest and precise, this cannot be conclusively proven by solely relying on such methods. As a solution, it is recommended that researchers employ biometric devices to compare and verify the results of paper and pencil measurements.

In this study, the participants who were instructed to give fake responses were told to think that their admission to a university department would be based on their test scores without specifying which department it was. In future studies, providing a clearer description of the fake personality structure for the group asked to give fake responses may be beneficial. Moreover, this study only examined LCA and CA among the methods used to detect fake responding behavior. Future studies could explore other analyses and include individuals from different age groups beyond the limited group that participated to this study voluntarily.

Declarations

Conflict of Interest: The authors of the article declare that they have no conflict of interest with any person or organization that may be a party to this study.

Ethical Approval: It is declared that scientific and ethical principles have been followed while carrying out and writing this study and that all the sources used have been properly cited.

References

Adair, C., K. (2014). *Interventions for addressing faking on personality assessments for employee selection: A meta-analysis.* [Doctoral Dissertation, DePaul University], College of Science and Health Theses and Dissertations. https://via.library.depaul.edu/csh_etd/93

Aybek, E. C. & Çıkrıkçı, R. N. (2018). Kendini değerlendirme envanteri'nin bilgisayar ortamında bireye uyarlanmış test olarak uygulanabilirliği [Applicability of the Self Assessment Inventory as a Computerized Adaptive Test]. *Turkish Psychological Counseling and Guidance Journal*, *8* (50), 117-141.

Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klâsik test teorisi ve uygulaması.* (3. Baskı). Pegem Akademi.

Büyüköztürk, Ş. Şekercioğlu, G. & Çokluk, Ö. (2020). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları.* (6. Baskı). Pegem Akademi.

Cohen, R. J., & Swerdlik, M. (2009). *Psychological testing and assessment* (7th ed.). McGraw-Hill.

Demetriou, C., Uzun Özer, B. & Essa, C. A. (2015). *Self-Report Questionnaires*. The Encyclopedia of Clinical Psychology, First Edition. Edited by Robin L. Cautin and Scott O. Lilienfeld. John Wiley & Sons, Inc. https://doi.org/10.1002/9781118625392.wbecp507

Dibao-Dina, C., Caille, A., Sautenet, B., Chazelle, E., & Giraudeau, B. (2014). Rationale for unequal randomization in clinical trials is rarely reported: a systematic review. *Journal of Clinical Epidemiology*, *67*(10), 1070–1075. https://doi.org/10.1016/j.jclinepi.2014.05.015

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists.* Lawrence Erlbaum Associates Publishers.

Erzen, E., Yurtçu, M., Ulu Kalın, Ö., & Koçoğlu, E. (2021). Development Of Social Desirability Scale: Validity and Reliability Study. *Electronic Journal of Social Sciences*, *20* (78); 879-891. https://doi.org/10.17755/esosder.774947

Everitt, B., Landau, S., Leese, M., & Stahl, D. (2011). *Cluster analysis* (5th ed.). Wiley.

Geiser, C. (2013). *Data analysis with Mplus*. Guilford.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2014). *Multivariate data analysis* (7th ed.). Pearson Education Limited.

Holland, J. L. (1973). *Making vocational choices: A theory of careers*. Prentice-Hall.

Huber, C. R. (2017). *Faking and the validity of personality tests: using new faking-resistant measures to study some old questions.* [Doctoral Dissertation, Minnesota University]. https://hdl.handle.net/11299/185605.

Kang, W., & Malvaso, A. (2023). Associations between Personality Traits and Areas of Job Satisfaction: Pay, Work Itself, Security, and Hours Worked. *Behavioral Sciences*, *13*(6), 445. http://dx.doi.org/10.3390/bs13060445

Keith, T. Z. (2015). *Multiple regression and beyond: An introduction to multiple regression and structural equation modeling* (2nd ed.). Routledge.

Kline, P. (1999). *The handbook of psychological testing* (2nd ed.). Routledge.

Kline, R. B. (2005). *Principles and practice of structural equation modeling*. (2nd ed.). Guilford.

Kubinger, K. D. (2002). On faking personality inventories. *Psychologische Beiträge, 44* (1), 10-16. https://www.proquest.com/docview/212173478

Kutlu, M., & Pamuk, M. (2017). Üniversite öğrencilerinde cep telefonunun problemli kullanımının kişilik bağlamında incelenmesi [Investigation of university students' problematic usage of mobile phone in the context of personality]. *Journal of Human Sciences*, *14*(2), 1263–1272. https://www.j-humansciences.com/ojs/index.php/IJHS/article/view/4073

Magidson, J., & Vermunt, J. (2002). Latent class models for clustering: A comparison with K-means. *Canadian Journal of Marketing Research*, *20*(1), 36-43.

Maniaci, M. R., & Rogge, R. D. (2014). Caring about carelessness: Participant inattention and its effects on research. *Journal of Research in Personality*, *48*, 61–83.

https://doi.org/10.1016/j.jrp.2013.09.008

Mehrens, W. A., & Lehmann, I. J. (1991). *Measurement and evaluation in education and psychology* (4th ed.). Wadsworth.

Morsünbül, Ü. (2014). Hızlı büyük beşli kişilik testi Türkçe versiyonu geçerlik ve güvenirlik çalışması [The validity and reliability study of the Turkish version of Quick Big Five Personality Test]. *Düşünen Adam The Journal of Psychiatry and Neurological Sciences*, *27*(4), 316-322. https://doi.org/10.5350/dajpn2014270405

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Pişkin, M. (2020). Kariyer gelişim sürecini etkileyen faktörler. B. Yeşilyaprak (Ed.), *Mesleki Rehberlik ve Kariyer Danışmanlığı Kuramdan Uygulamaya* (s. 45-90) içinde. Pegem Akademi.

Rassart, J., Luyckx, K., Goossens, E., Apers, S., Klimstra, T. A., & Moons, P. (2013). *Personality traits, quality of life and perceived health in adolescents with congenital heart disease. Psychology & Health, 28(3), 319 -335.* https://doi.org/10.1080/08870446.2012.729836

Rothstein, M. G., & Goffin, R. D. (2006). The use of personality measures in personnel selection: What does current research support? *Human Resource Management Review*, *16*(2), 155-180. https://doi.org/10.1016/j.hrmr.2006.03.004

Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. (2nd ed.). Lawrence Erlbaum Associates.

Sümer, N. (2000). Yapısal eşitlik modelleri: Temel kavramlar ve örnek uygulamalar. *Türk Psikoloji Yazıları*. *3*(6), 49-74.

Şimşek, A. S., & Tavşancıl, E. (2022). Validity and reliability of Turkish version of skills confidence inventory. *Turkish Psychological Counseling and Guidance Journal*, *12*(64), 89-107. https://doi.org/10.17066/tpdrd.1096008

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics.* (6th ed.) Pearson.

Thorndike, R. M., & Thorndike-Christ, T. M. (2013). *Measurement and evaluation in psychology and education* (8th ed.). Pearson.

Törnroos, M., Jokela, M., & Hakulinen, C. (2019). The relationship between personality and job satisfaction across occupations. *Personality and Individual Differences*, *145*, 82-88. https://doi.org/10.1016/j.paid.2019.03.027

Turgut, M. F., & Baykul, Y. (2019). *Eğitimde ölçme ve değerlendirme* (8. Baskı). Pegem Akademi.

van der Linden, D., Scholte, R. H. J., Cillessen, A. H. N., Nijenhuis, J. t., & Segers, E. (2010). Classroom ratings of likeability and popularity are related to the Big Five and the general factor of personality. *Journal of Research in Personality*, *44*(5), 669–672. https://doi.org/10.1016/j.jrp.2010.08.007

Vermulst, A. A., & Gerris, J. R. (2005). *QBF: Quick big five persoonlijkheidstest handleiding (Quick big five personality test manual).* Leeuwarden: LDC.

Vermunt, J. K., & Magidson, J. (2002). *Latent class cluster analyses*. In J. A. Hagenaars & A. L. McCutcheon (Eds.), Applied latent class analysis (pp. 89-107). Cambridge University Press.

Vermunt, J. K., & Magidson, J. (2005). *Latent GOLD 4.0 user's guide.* Statistical Innovations Inc.

Viswesvaran, C., & Ones, D. S. (1999). Meta-analyses of fakability estimates: Implications for personality measurement. *Educational and Psychological Measurement, 59*(2), 197-210. https://doi.org/10.1177/00131649921969802

Westland, J. C. (2019). *Structural equation models: From paths to networks*. (2nd ed.). Springer.

Wetzel, E., Böhnke, J. R., & Brown, A. (2016). Response biases. In F. T. L. Leong, D. Bartram, F. M. Cheung, K. F. Geisinger, & D. Iliescu (Eds.), *The ITC international handbook of testing and assessment* (pp. 349–363). Oxford University Press. https://doi.org/10.1093/med:psych/9780199356942.003.0024

Widhiarso, W., & Himam, F. (2015). Employee recruitment: Identifying response distortion on the personality measure. *Electronic Journal of Business Ethics and Organization Studies, 20*(1), 14-21. http://urn.fi/URN:NBN:fi:jyu-201505061735

Yankov, G. P. (2019). Faking on personality tests: *The relationship between intelligence and personality*. [Doctoral dissertation, Bowling Green State University]. https://scholarworks.bgsu.edu/psychology_diss/206