

Unveiling Behavioral Types of Nonverbal Immediacy: A Hybrid Approach Using EFA and LCA

Chengwei Cao

New York University, New York, United States

Cc8933@nyu.edu

Abstract. This study, based on a large-scale survey of over 130,000 respondents, explored the structural and behavioral diversity captured by the Nonverbal Immediacy Scale (NIS). We used exploratory factor analysis (EFA) to reveal the underlying dimensions of nonverbal behavior and combined it with latent class analysis (LCA) to group individuals. Overall, the analyses revealed a multifactor structure and identified three behavioral profiles that differentiate communication tendencies. The latent classes differed in demographic characteristics, particularly gender and education level. Combining dimensional and categorical analysis, this study provides new insights into inter-individual differences in nonverbal immediacy within a large population.

Keywords: NIS, EFA, LCA, Analysis.

1. Introduction

Nonverbal communication, such as eye contact and facial expressions, plays an important role in everyday interpersonal interactions by conveying emotions and attitudes. This concept was first introduced by Mehrabian (1967) through his “liking leads to immediacy behavior” theory, and has since been extended and empirically validated by numerous subsequent studies. [1,2] Among related constructs, Nonverbal Immediacy refers to the extent to which individuals express approachability, engagement, and participation through physical behaviors. It has been widely recognized as a key psychological indicator of communication quality and interpersonal closeness. [3]

To systematically assess this construct, Richmond, McCroskey, and Johnson (2003) developed the Nonverbal Immediacy Scale (NIS), which has been broadly adopted in studies on interpersonal communication, teacher-student interaction, and cross-cultural expression of nonverbal cues.[4,5]

While previous studies have examined the factor structure or behavioral profiles of nonverbal immediacy separately, few have combined the two. This study uses both EFA and LCA to explore its internal structure and individual differences.[6-8] This study first conducts Exploratory Factor Analysis (EFA) to examine the underlying multidimensional structure of the NIS and clarify the latent dimensions represented by the items. Building on this, Latent Class Analysis (LCA) is applied to identify subgroups of individuals who exhibit systematically different patterns of nonverbal behavior, as reflected in their response profiles.

By integrating dimensional and typological modeling approaches, this study aims to deepen our understanding of the latent structure underlying nonverbal immediacy and provide a theoretical foundation for future individualized communication strategies and psychological interventions.

2. Factor Structure and Latent Class Modeling

2.1. Factor Analysis Model

This study first employs Exploratory Factor Analysis (EFA) to extract the latent dimensions underlying the items in the Nonverbal Immediacy Scale (NIS), aiming to uncover the construct structure of nonverbal immediacy.

Factor analysis assumes that each observed variable X_i can be represented as a linear combination of a set of common factors $F_1, F_2, \dots, F_m$, along with a unique error term ϵ_i . [9] The general model is expressed as:

$$X_i = \lambda_{\{i1\}F_1} + \lambda_{\{i2\}F_2} + \dots + \lambda_{\{im\}F_m} + \varepsilon_i \quad (1)$$

where λ_{ij} denotes the factor loading of variable X_i on the j -th common factor, and ε_i represents the unique (specific) error associated with X_i . [9]

In terms of parameter estimation, this study adopts the Principal Axis Factoring method for factor extraction and applies Varimax orthogonal rotation to enhance interpretability, allowing for the identification of relatively independent and meaningful factor dimensions.

2.2. Latent Class Analysis Model

To identify the latent heterogeneity in individuals' nonverbal immediacy behaviors, this study further applies **Latent Class Analysis (LCA)** to model responses across the 26 items of the NIS scale.

LCA is a probabilistic, discrete latent variable model that assumes the distribution of observed variables is generated by a mixture of several unobserved latent classes [10]. The general model is expressed

$$P(Y_i) = \sum_{c=1}^C \pi_c \prod_{j=1}^J P(Y_{ij} | Z_i = c) \quad (2)$$

$Y_1, \dots, Y_J$ denote the J observed variables (i.e., the questionnaire items); C is the latent class variable with K possible categories; π_c represents the prior probability of class c ; $P(Y_j | Z_i = c)$ denotes the conditional probability of response to item j given membership in class c .

3. Model Building

3.1. Data Processing

To ensure data quality, we first confirmed that the dataset contained no missing values across the main scale items (Q1–Q26) and vocabulary recognition items (VCL1–VCL16). In this dataset, a value of zero does not indicate a valid response but instead reflects an unattempted or skipped question; however, no such cases were observed.

As an additional quality control measure, the questionnaire included deliberately fabricated vocabulary items (e.g., VCL6, VCL9, VCL12). Participants who claimed to recognize two or more of these fake words were excluded, as this may indicate inattentive or dishonest responding. Furthermore, respondents whose reported age fell outside the range of 10 to 70 years were also removed. After these screening steps, the final cleaned dataset included 130,169 valid observations.

Furthermore, as the Nonverbal Immediacy Scale (NIS) includes several negatively worded items, for example, “I avoid eye contact with others”. Direct summation of item scores without adjustment could result in misleading interpretations. To address this, reverse scoring was applied to 13 items based on the original scale design, ensuring a consistent direction of measurement. The total nonverbal immediacy score (total_score) was then calculated by summing the responses to all 26 items. Higher scores indicate a greater self-reported tendency toward nonverbal immediacy in social interactions.

3.2. Exploratory Factor Analysis (EFA)

All negatively worded items were reverse-coded prior to analysis to maintain consistent scale polarity. The responses were then standardized, and factor extraction was performed using the principal axis factoring method.

The number of factors was determined using multiple criteria, including the Kaiser rule (eigenvalues > 1), visual inspection of the scree plot, and parallel analysis. To improve interpretability, a varimax rotation was applied to the factor loadings. The resulting factor scores—representing

distinct dimensions of nonverbal immediacy such as expressiveness or behavioral engagement—were retained for use in subsequent modeling.

3.3. Latent Class Analysis

Latent class analysis (LCA) was used to uncover unobserved subpopulations characterized by distinct response patterns on the 26-item NIS. The model was estimated using raw item scores as input (after reverse-coding the necessary items), and models with one to five latent classes were compared. Model fit was evaluated using standard information criteria—Bayesian Information Criterion (BIC), Akaike Information Criterion (AIC) assess classification quality.

4. Descriptive Statistics and Latent Class Interpretation

4.1. Computing Total Score

First, the total score of each participant was computed based on the 26-item NIS scale. As shown in Figure 1, the score ranged from 25 to 130, with the majority concentrated around 90–95. The distribution is left-skewed, indicating that most participants reported relatively high levels of nonverbal immediacy.

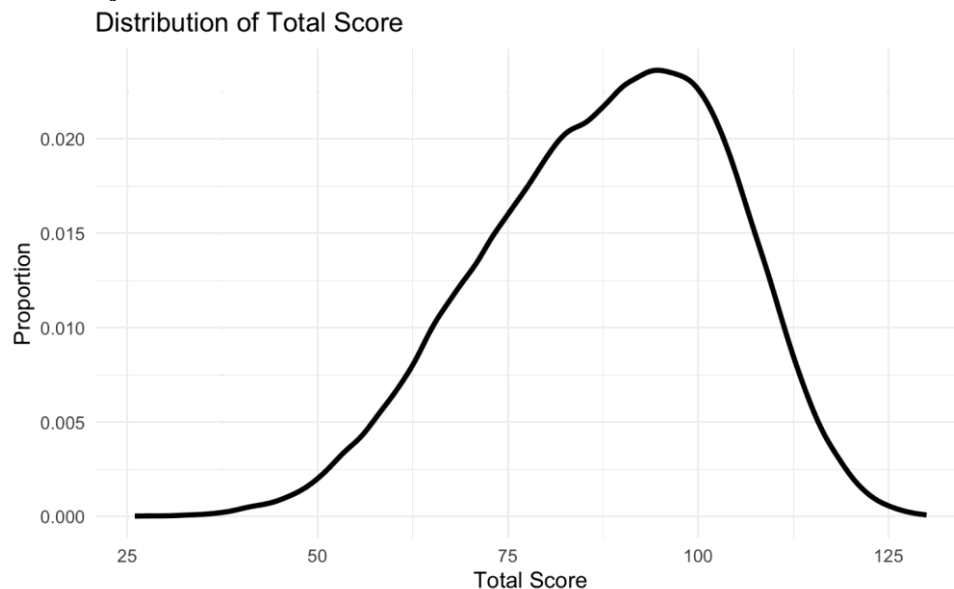


Fig. 1 Distribution of Total Score

Next, this paper explored gender differences in total scores. While both male and female participants showed similar distributions (Figure 2), female participants exhibited a slightly right-skewed distribution with a higher peak, suggesting generally higher scores. The age distribution (Figure 3) is also examined, which showed that most participants were between 16 and 24 years old, with a median age of 20. The strong right skew and concentration around young adulthood indicate that the sample is predominantly composed of adolescents and young adults.

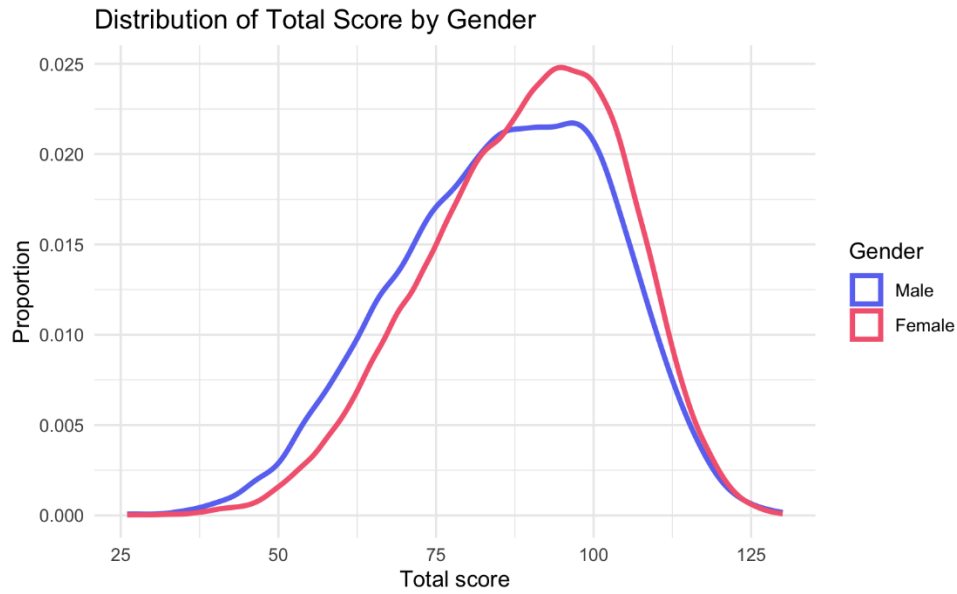


Fig. 2 Distribution of Total Score By Gender

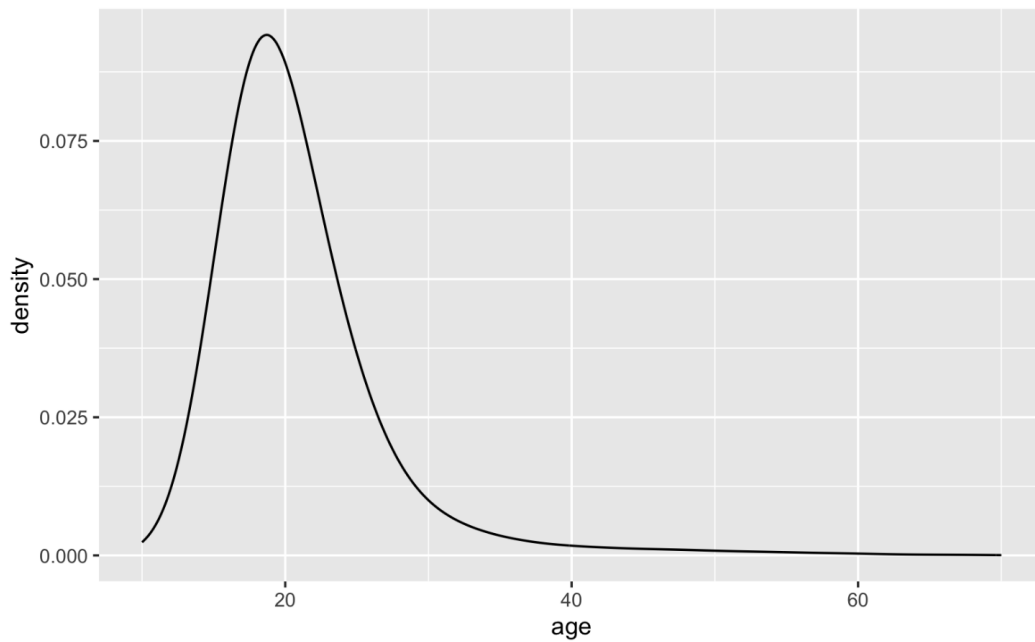


Fig. 3 Distribution of Age

To examine whether the NIS scale captures multiple dimensions of nonverbal immediacy, we conducted an exploratory factor analysis. The method was chosen to uncover potential latent dimensions that explain the covariance structure among the 26 items. Responses were standardized prior to analysis. Compared to PCA or IRT, factor analysis is more suitable for identifying underlying psychological constructs.

The Fig.4 revealed a sharp drop after the first factor and a visible elbow after the second factor, suggesting that a two-factor solution is appropriate.

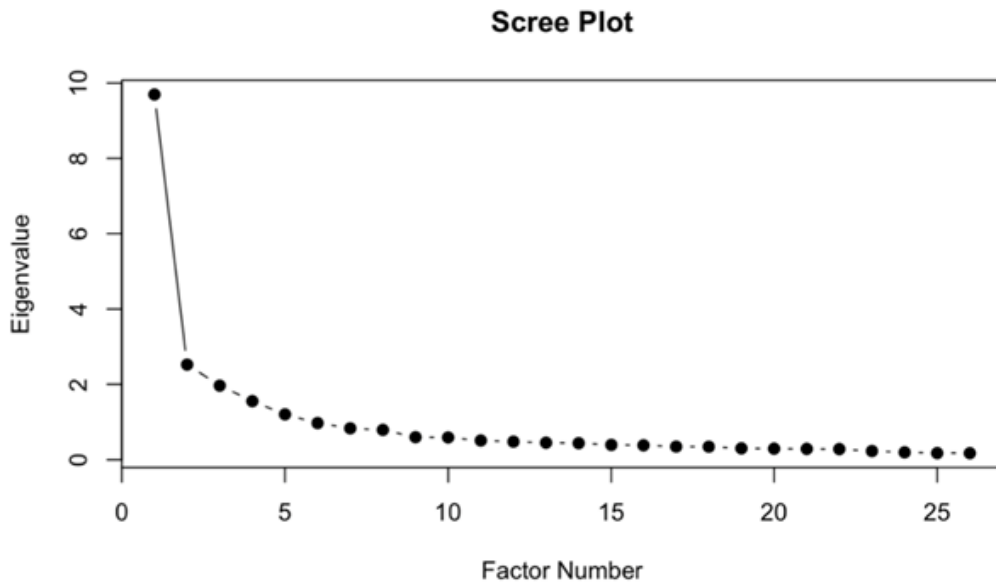


Fig. 4 Scree Plot

4.2. Factor Analysis Process

Next figure 5 shows about loadings on factor 1 to figure out the feature of factor 1. Factor 1 appears to reflect nonverbal immediacy related to eye contact and visual engagement during communication. it can be seen that Q8, Q17, Q22 and Q4 have high loading on factor1. Q8 is 'I avoid eye contact while talking to people'. Q17 is 'I look directly at people while talking to them.' Q22 is 'I maintain eye contact with people when I talk to them.' and Q4 is 'I look over or away from others while talking to them.' All 4 questions are about eye contact. This suggests that Factor 1 captures a latent trait describing how much individuals engage in or avoid direct visual connection with others during interpersonal interaction.

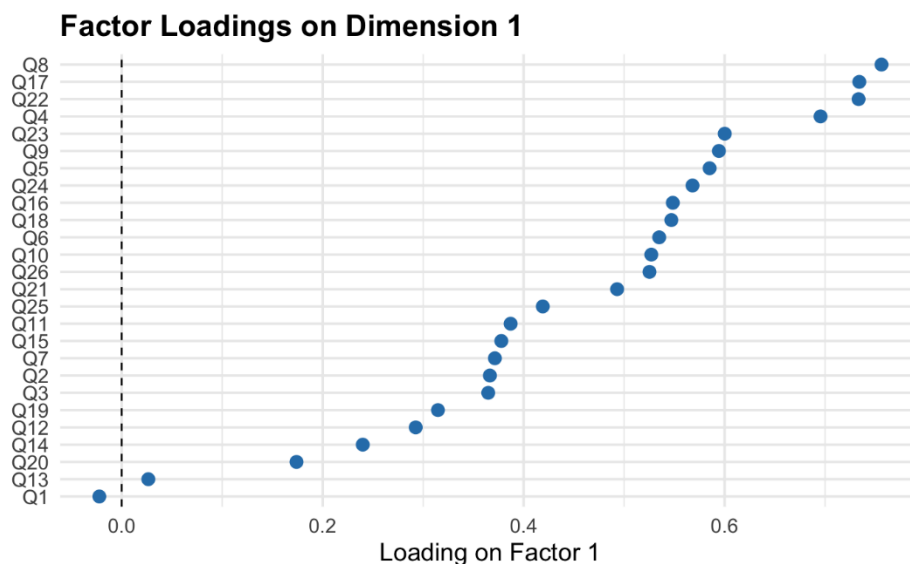


Fig. 5 Factor Loadings on Dimension 1

For Factor on dimension 2(Figure 6). , it can be seen that Q13, Q14,Q1 and Q12 has the highest loadings. Q1 is 'I use my hands and arms to gesture while talking to people. ' Q12 is 'I use a variety of vocal expressions when I talk to people. 'Q13 is 'I gesture when I talk to people. ' Q14 is 'I am animated when I talk to people. ' These indicate that this dimension reflects how much an individual uses expressive behaviors such as gestures, tone variation, and body movement to enhance interpersonal interaction.

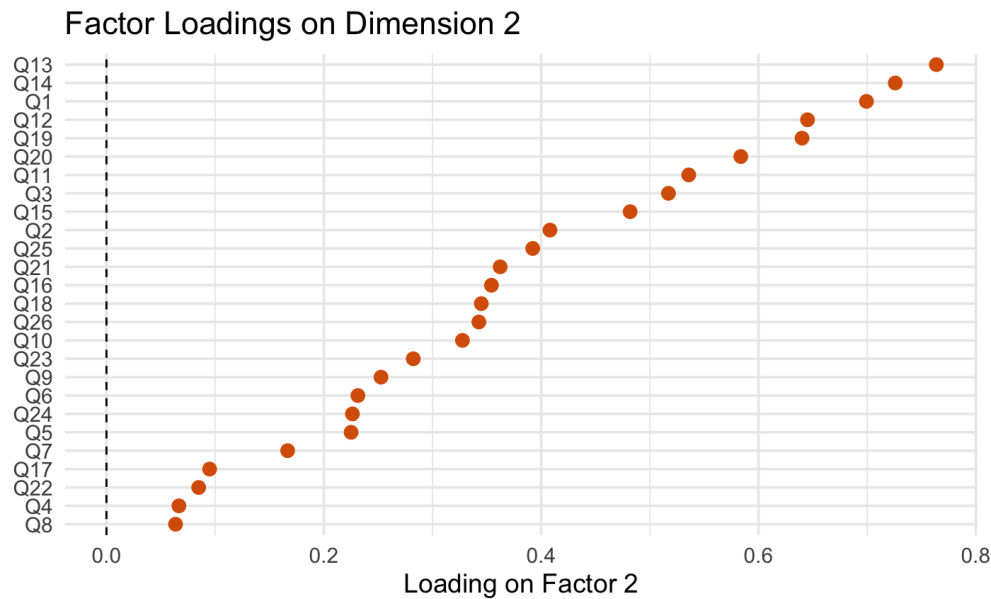


Fig. 6 Factor Loadings on Dimension 2

4.3. LCA Interpretation

To explore potential heterogeneity in participants' nonverbal communication patterns, latent class analysis (LCA) is conducted using item responses. Next step is estimating and comparing models with 1 to 5 classes based on BIC, AIC, and entropy. Although BIC consistently decreased, the decrease from 3 to 4 classes was relatively small compared to the decrease from 2 to 3 classes. Considering statistical criteria and interpretability, a three-class solution is chosen.

The response patterns of the three classes revealed distinct behavioral tendencies. Individuals in Class 1 tended to score highly on avoidance-related items while showing lower levels of immediacy-related behaviors. This suggests that participants in this group are more likely to avoid close physical proximity and eye contact during social interactions, reflecting a more withdrawn or reserved communication style.

In contrast, Class 2 and Class 3 both exhibited higher overall levels of nonverbal immediacy, but differed in how this immediacy was expressed. Participants in Class 2 provided more moderate and cautious responses, particularly on items involving social distance and gaze behavior, indicating a more measured and perhaps socially regulated communication style. Meanwhile, Class 3 participants demonstrated a tendency to choose more extreme response options—for instance, “never” or “always”—across various items. This pattern suggests a more consistent and confident nonverbal style, with stronger behavioral tendencies across different communication scenarios.

When comparing total scores across latent classes, we can observe a clear pattern. Participants from class 1 tend to have the lowest scores, class 2 have moderate, and class 3 with the highest scores. This finding aligns with our previous interpretation of latent classes (Figure 7).

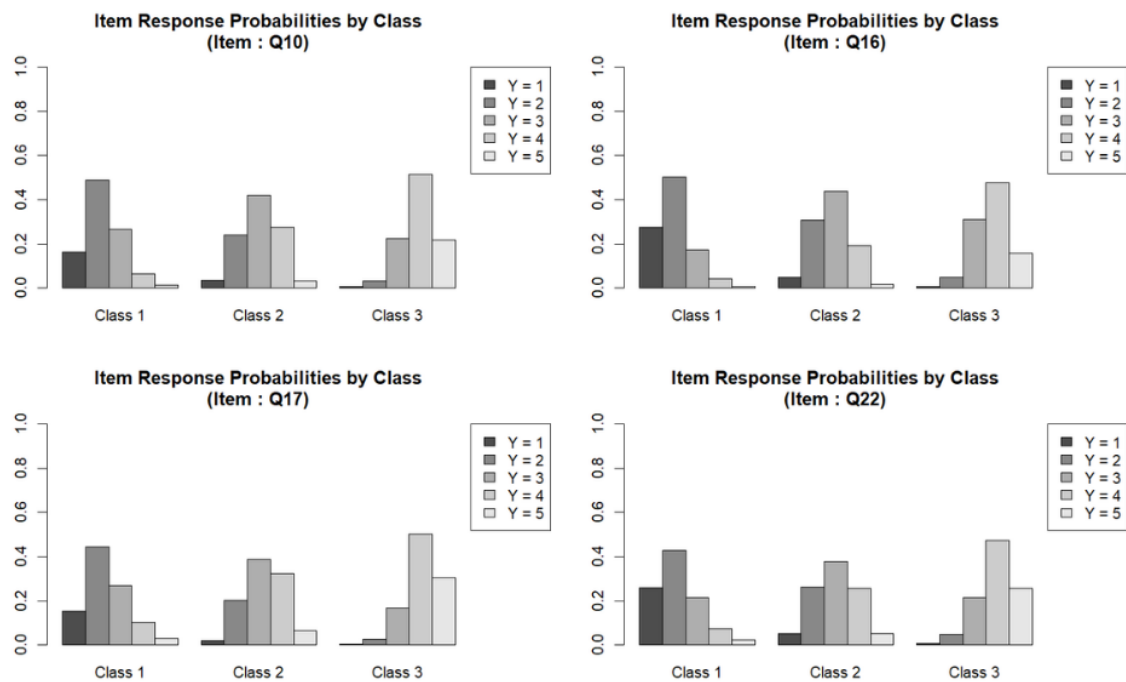


Fig. 7 Item Response Probabilities by Class

After interpreting the response pattern of different classes, this paper explore how participants in different latent classes varied across demographic variables. Among all demographic factors, only gender and education showed clear differences among classes. In particular, class 1 included a slightly higher proportion of male participants, while class 2 and 3 have relatively more female participants. For education level, participants from class 1 have relatively lower education level compared with other two classes (Figure 8,9).

However, this pattern should be interpreted carefully. As the descriptive statistics part suggests, the survey contains a disproportionately high number of females, and the majority of respondents were between 16 and 24. This kind of data distribution may influence our observed pattern.

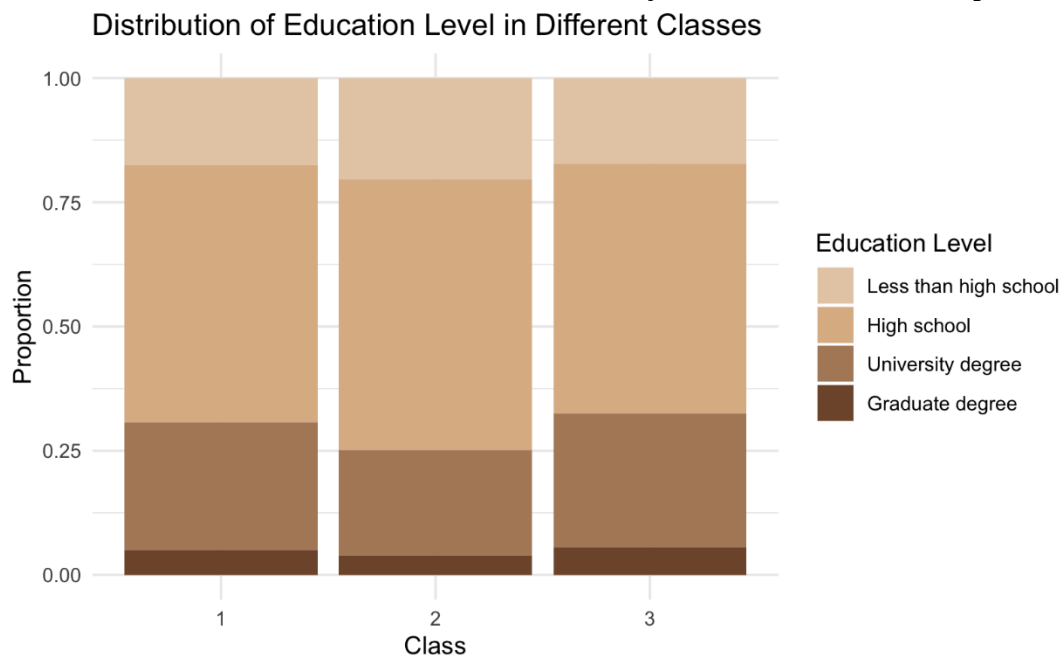


Fig. 8 Distribution of Education by class

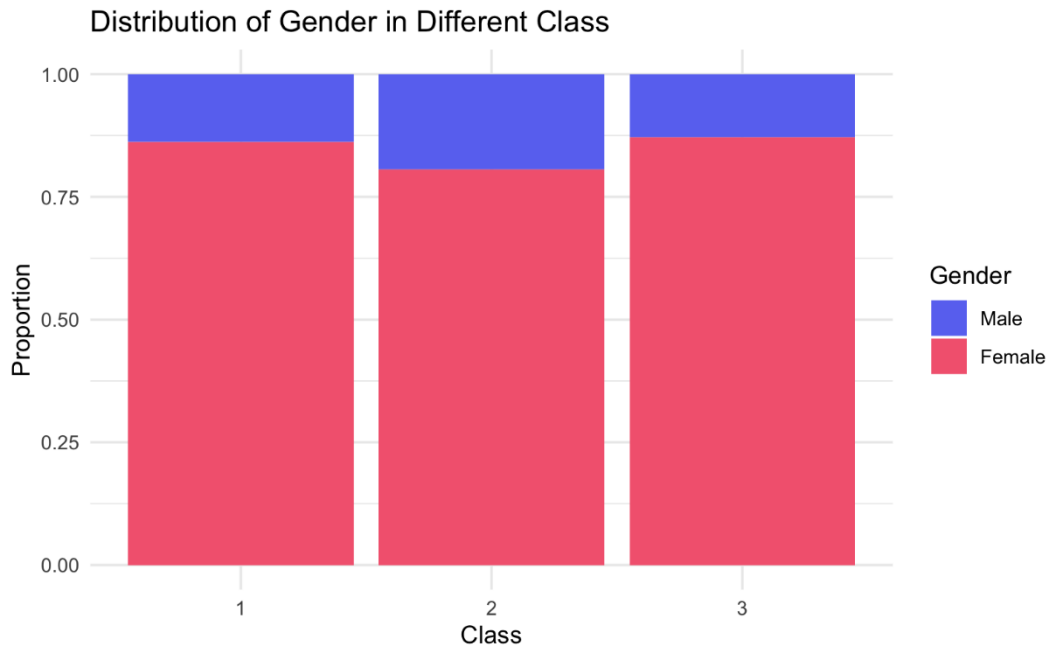


Fig. 9 Distribution of Gender by class

5. Conclusion

Based on questionnaire data from the Nonverbal Immediacy Scale, this paper constructs a statistical modeling framework to explore individual differences in nonverbal behavior. First, the raw data was cleaned and scored, reverse-coded, and a total score was calculated to measure individuals' overall level of nonverbal immediacy. Subsequently, principal component analysis (PCA) and factor analysis (FA) were used to explore the underlying structure between the items and extract key dimensions.

Furthermore, latent class analysis (LCA) was used to cluster the sample to identify latent groups with distinct response patterns. By comparing metrics such as BIC and entropy across models with different numbers of classes, a three-class model was ultimately selected as the optimal solution. The results showed significant differences in nonverbal behavior across classes: Class 1 tended to exhibit avoidance behaviors, while Classes 2 and 3 exhibited varying degrees of nonverbal engagement. Further demographic analysis revealed significant differences between classes in terms of gender and education level. Class 1 had a slightly higher proportion of males and relatively lower educational attainment, while Classes 2 and 3 were predominantly female and highly educated.

Although this study revealed the underlying structure and demographic differences underlying questionnaire response patterns, it still has certain limitations. For example, the sample had a significant gender imbalance, which may have affected the representativeness of the results. Furthermore, the analysis was based solely on self-reported data and was not validated with behavioral observations or other external variables. Future research could further incorporate multi-source data and attempt to link latent class results with specific behavioral manifestations or psychological variables to enhance the model's explanatory power and practical value.

References

- [1] A. Mehrabian, "Attitudes inferred from neutral verbal communications," *Journal of Consulting Psychology*, vol. 31, no. 4, pp. 414–417, 1967.
- [2] P. Puertas-Molero, F. Zurita-Ortega, G. González-Valero, and J. L. Ortega-Martín, "Design and validation of the Non-Verbal Immediacy Scale (NVIS) for the evaluation of non-verbal language in university professors," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1159, 2022.

- [3] V. P. Richmond, J. C. McCroskey, and A. D. Johnson, "Development of the nonverbal immediacy scale (NIS): Measures of self- and other-perceived nonverbal immediacy," *Communication Quarterly*, vol. 51, no. 4, pp. 504–517, 2003.
- [4] J. C. McCroskey, A. Sallinen, J. M. Fayer, V. P. Richmond, and R. A. Barraclough, "Nonverbal immediacy and cognitive learning: A cross-cultural investigation," *Communication Education*, vol. 45, no. 3, pp. 200–211, 1996.
- [5] P. Puertas-Molero, F. Zurita-Ortega, G. González-Valero, and J. L. Ortega-Martín, "Design and validation of the Non-Verbal Immediacy Scale (NVIS) for the evaluation of non-verbal language in university professors," *International Journal of Environmental Research and Public Health*, vol. 19, no. 3, p. 1159, 2022.
- [6] S. Cao, L. Liu, Q. Zhu, Z. Zhu, J. Zhou, P. Wei, and M. Wu, "Association between dietary patterns and plasma lipid biomarker and female breast cancer risk: Comparison of latent class analysis (LCA) and factor analysis (FA)," *Frontiers in Nutrition*, vol. 8, 2021. [Online].
- [7] W. Zhang, B. Wang, D. Zhao, C. Zhao and M. Xu, "A neural multi-scale feature fusion model with hybrid attention for corn pest recognition," *Journal of Bionic Engineering*, vol. 21, no. 3, pp. 621–633, May 2024,
- [8] D. Sotres-Alvarez, A. H. Herring, and A. M. Siega-Riz, "Latent Class Analysis Is Useful to Classify Pregnant Women into Dietary Patterns," *The Journal of Nutrition*, vol. 140, no. 12, pp. 2253–2259, 2010,
- [9] L. L. Thurstone, "Multiple factor analysis," *Psychological Review*, vol. 38, no. 5, pp. 406–427, 1931.
- [10] C. E. Collins and M. M. Lanza, "Latent class analysis as a model-based clustering method for categorical data," *Nurs. Res.*, vol. 65, no. 6, pp. 456–465, 2016