www.editada.org

International Journal of Combinatorial Optimization Problems and Informatics, 18(1), 2027, 372-384. ISSN: 2007-1558.
https://doi.org/10.61467/2007.1558.2027.v18i1.1511

Detection of Dietary Self-Regulation Patterns Using Agglomerative Hierarchical Clustering (AHC) for Ultra-Processed Foods Among Schoolchildren

Raquel Estrella Barrón [1], Airam Anerim Guillen Suárez [2], Martha Alicia Lara Heredia [3], Freddy De la Cruz Ruiz [4], José Antonio Córdova Hernández [5], Reinerio Zapata Salazar [6]

[1,2,3] Facultad de Ciencias de la Salud. Universidad Autónoma del Carmen, Ciudad del Carmen, Campeche, México

[4,5,6] División Académica Multidisciplinaria de Comalcalco. Universidad Juárez Autónoma de Tabasco, Comalcalco, Tabasco, México

Email address(es): restrella@pampano.unacar.mx[1], aaguillen@pampano.unacar.mx[2], alara@pampano.unacar.mx[3], antonio.cordova@ujat.mx[5]

✉ Corresponding author: reinerio.zapata@ujat.mx[6], freddy.delacruz@ujat.mx[4]

Abstract. This research analyzed food self-control patterns regarding ultra-processed foods (UPF) in 236 schoolchildren (aged 8–12) using a cross-sectional design. A pictorial instrument assessed health, taste, and preference for 50 foods. Calculating a Self-Control Index and applying AI (Agglomerative Hierarchical Clustering with Canberra distance and full linkage) alongside Multidimensional Scaling, results showed significantly lower self-control for UPF versus healthy foods (Wilcoxon test). Clustering identified three profiles: high UPF attraction (n=189), balanced/healthy preference (n=45), and dissociation between knowledge and behavior (n=2). Findings highlight children's vulnerability to UPF, the need for differentiated interventions, and demonstrate AI-integrated indices as valuable for early risk profiling and precision nutrition.

Keywords: Food Self-Control, Ultra-processed Foods, Nonparametric statistics, Hierarchical Grouping Algorithm.

Article information
Received: Jun 26, 2026
Accepted: Aug 11, 2026

1 Introduction

Ultra-processed food (UPF) consumption has emerged as a major public health concern due to its increasing incorporation into dietary patterns and its association with poorer diet quality and a range of adverse health outcomes (Lane et al., 2024). Its presence has also been documented in the Mexican population, including young adults (Estrella Barrón & Telumbre Terrero, 2024). In this context, understanding how perceptions and preferences toward these foods are shaped during childhood is important for identifying processes that may influence subsequent food choices.

Children's food preferences result from the interaction of sensory characteristics, individual experiences, and environmental influences (Beckerman et al., 2017). Among the dimensions involved in this process, taste constitutes a major determinant of food evaluation and choice. However, hedonic appeal and preference for consumption do not necessarily represent the same phenomenon. Studies conducted in pediatric populations have demonstrated distinctions between components such as liking and wanting (Liem & Zandstra, 2009), as well as differential contributions of taste, health, and desire to eat during food choice processes (Pearce et al., 2020).

Perceived healthiness represents another important dimension. Children and adolescents are capable of differentiating foods according to their perceived healthfulness (Bucher et al., 2016), while taste and health-related evaluations may exert different levels of influence on the formation of food preferences and dietary decisions (van Meer et al., 2019; Serrano-Gonzalez et al., 2021). Furthermore, Rigo et al. (2023) reported that sensory, nutritional, and health-related cues jointly contribute to food choices among children aged 5 to 12 years. Collectively, these findings suggest that perceived healthiness, taste, and consumption preference are related but distinct dimensions, and that their joint assessment may provide a more comprehensive characterization of children's food evaluations.

Nevertheless, the relationships among these dimensions may exhibit substantial heterogeneity across schoolchildren, which may not be fully captured by individual descriptive measures. In the present study, the Food Preference Regulation Index (FPRI) was used to quantify the discrepancy between taste ratings and the stated preference for consuming a food, weighted by its perceived healthiness.

However, summary indicators may conceal heterogeneity in individual response patterns. From a multivariate perspective, Agglomerative Hierarchical Clustering (AHC) enables the exploration of similarities among observations and the grouping of participants according to shared patterns across multiple dimensions without requiring the prior specification of categories. Complementarily, Multidimensional Scaling (MDS) provides a representation of dissimilarity relationships in a lower-dimensional space, facilitating the visualization of underlying structures and gradients within the data. The integration of these techniques therefore offers an exploratory strategy for characterizing heterogeneity in children's food evaluations beyond comparisons based solely on summary measures.

Within this framework, the aim of the present study was to analyze food preference regulation toward ultra-processed and healthy foods among schoolchildren using the Food Preference Regulation Index (FPRI), and to explore multidimensional patterns of perceived healthiness, taste, and consumption preference through Agglomerative Hierarchical Clustering (AHC) and Multidimensional Scaling (MDS). This approach seeks to provide a quantitative and multivariate characterization of children's food evaluations and to examine the heterogeneity in how schoolchildren integrate hedonic and health-related attributes when evaluating different foods.

2 Methodology

2.1. Study design and population

An observational, cross-sectional study was conducted with a sample of 236 schoolchildren between 8 and 12 years of age, enrolled in a primary school in Ciudad del Carmen, Campeche, Mexico. Data collection took place between August and December 2023. The sample was selected using stratified random sampling to ensure adequate representation of the school population. All procedures complied with ethical principles for research involving human subjects, and informed consent was obtained from parents or guardians prior to participation.

2.2. Instrument and data collection

Food consumption and perception were assessed using a computerized pictographic instrument, previously validated in similar studies (Ha et al., 2019). The questionnaire included 50 foods represented with real images (300 × 300 pixels, 72 dpi), of which:

- 25 corresponded to ultra-processed foods (e.g., sweets, fried foods, processed meats).
- 25 were healthy foods (e.g., fruits, beans, vegetables).

Each participant independently rated three attributes of each food item: perceived healthiness (the extent to which the food was considered healthy), taste (the extent to which the food was perceived as pleasant or palatable), and consumption preference (the extent to which the participant liked consuming the food). Ratings were obtained using a five-point visual scale based on colored facial expressions.

2.3. Data preparation and cleaning

The data were originally collected in long format, with three records per student corresponding to the dimensions of perceived healthfulness, taste, and consumption preference for each of the 50 foods evaluated. Based on this initial structure, three procedures were conducted sequentially: variable standardization, missing data treatment, and data format transformation.

Variable standardization. To facilitate analytical traceability and clearly distinguish between the two food groups assessed by the instrument, systematic prefixes were assigned to variable names: UPF_ for the 25 ultra-processed foods and HF_ for the 25 healthy foods. Additionally, categorical variables were numerically recoded to enable their inclusion in subsequent statistical analyses.

Missing data treatment. Of the 35,400 expected observations, 84 missing values (0.24%) were identified, distributed across 31 of the 50 variables (62%), with an average of 2.7 missing observations per affected variable (approximately 0.38%). The limited extent of missingness and its uniform distribution, without concentration in specific items, participants, or dimensions, were consistent with a Missing Completely At Random (MCAR) mechanism. As formalized by Rubin (1976), MCAR refers to a condition in which the probability of a value being missing is independent of both the observed and unobserved data; that is, missingness occurs purely by chance. Under this assumption, simple univariate imputation methods do not introduce systematic bias into parameter estimates, although they may lead to a slight underestimation of variance (Little & Rubin, 2019).

Given that the proportion of missing values was well below the 5% threshold, beyond which the bias associated with simple imputation techniques becomes more consequential (Schafer, 1999), a univariate central tendency imputation approach was adopted. The choice of the specific measure of central tendency was guided by the distributional characteristics of each variable, as assessed using the Coefficient of Variation (CV):

The Equation 1:

$$CV = \frac{s}{\bar{x}} * 100 \qquad 1$$

where:

- CV = Coefficient of Variation, expressed as a percentage.
- s = sample standard deviation of the variable.
- $\bar{x}$ = arithmetic mean of the variable.

A decision rule was established whereby, when CV ≤ 30%, the distribution was considered homogeneous and the arithmetic mean was deemed representative of the distribution's central tendency; therefore, mean imputation was applied. Conversely, when CV > 30%, the distribution was considered heterogeneous or skewed, a condition under which the mean loses representativeness because of its sensitivity to extreme values. In such cases, median imputation was employed, given its robustness to skewed distributions (Manterola & Otzen, 2014; von Hippel, 2013). This criterion operationalizes the classical recommendation of using the mean for symmetric distributions and the median for skewed distributions, a widely adopted data preprocessing approach that preserves both sample size and the distributional integrity of individual variables (García-Laencina et al., 2010).

Data format transformation. Following the completion of variable standardization and missing-data imputation within the original data structure, the dataset was reorganized from long to wide format, such that each student was represented by a single record. The final analytical matrix consisted of 236 students × 50 variables per dimension. All procedures were performed using IBM SPSS Statistics Version 31. Data quality was verified through duplicate detection, random comparisons between original and transformed records, and assessments of the consistency of descriptive statistics before and after the transformation process.

2.4. Derived variables

Based on the ratings of perceived healthiness, taste, and consumption preference assigned to each food item, the Food Preference Regulation Index (FPRI) was developed as a composite indicator of the discrepancy between a food's hedonic appeal and the stated preference for consuming it, weighted by perceived healthiness.

For each food item, the index was calculated using the following formula:

- FPRIfood = Perceived Healthiness × (Taste − Consumption Preference)

where perceived healthiness refers to the extent to which the child considered the food to be healthy, taste reflects the perceived palatability of the food, and consumption preference indicates the degree to which the child liked consuming that food.

In this formulation, the difference between taste and consumption preference determines the direction of the index, while perceived healthiness acts as a weighting factor that amplifies or attenuates this difference. Positive FPRI values indicate that the perceived taste of a food exceeded the participant's preference for consuming it. A value of zero indicates agreement between both ratings, whereas negative values indicate that consumption preference exceeded the perceived taste rating.

Higher positive values reflect a larger gap between the perceived palatability of a food and the actual desire to consume it, adjusted for its perceived healthiness. This discrepancy was interpreted as a greater degree of regulation of food preference in relation to the hedonic appeal of taste.

Using individual food-level scores, three summary indicators were computed: a global FPRI, based on all 50 food items evaluated; FPRI-UPF, calculated using only the 25 ultra-processed foods; and FPRI-HF, calculated using the 25 healthy foods. These indicators were derived exclusively from the ratings obtained through the assessment instrument. Therefore, they do not directly measure actual food consumption behavior, nor do they represent a general psychological construct of self-control.

2.5. Statistical analysis

Descriptive analyses (mean, median, mode, standard deviation, variance, range, and percentile intervals) were conducted to characterize the Food Preference Regulation Indices (FPRIs).

Given the non-normal distribution of the data, nonparametric tests were applied:

- Kruskal-Wallis test, to evaluate differences in food preference regulation across school groups.
- Wilcoxon signed-rank test, to compare the FPRI for healthy foods (FPRI-HF) and the FPRI for ultra-processed foods (FPRI-UPF).

2.6. Analysis Using an Agglomerative Hierarchical Clustering Algorithm

The Agglomerative Hierarchical Clustering (AHC) algorithm was employed to examine and analyze the relationship between unhealthy food consumption and taste preferences among children exhibiting lower levels of self-control. This AHC algorithm generates clusters based on the chosen linkage criterion and utilizes a distance matrix Dn of dimensions (n × n), which is computed through a specified distance metric.

The methodology followed in this study is described in (Hernández-Gómez et al., 2023; Giordani et al., 2020).

The construction of AHC involves the following sequential steps:

1. Initially, each object under investigation is assigned to an individual cluster.
2. Subsequently, the dissimilarity matrix is computed between every pair of objects within the dataset being analyzed.
3. Linkage criteria are then established to determine how clusters will be merged during the analysis.
4. Finally, steps 2 and 3 are iteratively repeated until all objects are consolidated into a single comprehensive cluster.

AHC is a bottom-up clustering technique where clusters are constructed from subclusters, which can themselves contain additional subclusters. This approach produces a multi-level hierarchical structure of clusters, visually represented through a tree-like diagram called a dendrogram. In this research, the Canberra distance metric was employed as the dissimilarity measure between paired data points within the numerical dataset. The Canberra distance is a metric used to compute the distance (d) between vectors x and y in an n-dimensional real vector space (Lance & Williams 1966; Lance & Williams 1967).

This technique is frequently employed to assess similarity in numerical datasets and to create symmetric clusters. The method involves calculating the sum of fractional components, where each fraction represents the absolute difference between corresponding coordinates of the object pairs. Data points featuring both numerator and denominator values of zero are omitted from the calculation, as they are considered mathematically undefined. The mathematical formulation for the Canberra distance is provided in Equation 2.

$$d(x,y) = \sum \frac{|x_i - y_i|}{|x_i| + |y_i|} \quad 2$$

Where d(x, y) represents the distance between vectors x and y. The complete linkage method was selected for constructing the AHC. This approach determines the similarity between two clusters by calculating the maximum distance between their constituent objects, which establishes the criterion for cluster merging (Pacheco, 2015).

The Equation 3 (Giordani et al., 2020):

$$\max_{i_{C1}, i_{C2}} d\,(i_{C1}, i_{C2}),\ i_{C1} \in C_1,\ i_{C2} \in C_2 \qquad 4$$

Where iC1 and iC2 denote any two observations considered as clusters. The term d(iC1, iC2) represents the maximum distance between these two clusters. The notations iC1 ∈ C1 and iC2 ∈ C2 indicate that a specific observation belongs to its corresponding cluster, either C1 or C2. The complete linkage method typically generates more compact clusters, which proves advantageous for data analysis.

Multidimensional Scaling (MDS) (Buja et al., 2008) was also applied to the dissimilarity matrix to project the data into a smaller space and facilitate the visualization of clustering patterns.

2.7. Software

Data processing and cleaning were performed in SPSS v31 and LibreOffice Calc, while clustering and MDS analyses were run in R (v4.5.1, 2025). Clustering analyses were conducted using the R programming language, specifically R version 4.5.1 (2025-06-13)– "Great Square Root" Copyright (C) 2025 The R Foundation for Statistical Computing Platform: x86_64-pc-linux-gnu (Debian 12 bookworm).

2.8. Ethical considerations

Based on the approval issued by the Ethics Committee (Official Letter FCS-SE/125, June 29, 2023), the project from which this article derives was submitted to ethical evaluation prior to its execution, complying with the principles established for research on human subjects, in accordance with current national regulations and the criteria of the Declaration of Helsinki. Written informed consent was obtained from parents or legal guardians prior to the inclusion of schoolchildren in the study, ensuring confidentiality, anonymity, and the exclusive use of the information for academic and scientific purposes.

3 Results and Discussion

3.1. Descriptive Statistics

Table 1 presents the descriptive statistics for the Food Preference Regulation Index (FPRI) for the complete set of foods evaluated (overall FPRI) and according to food type: ultra-processed foods (FPRI-UPF) and healthy foods (FPRI-HF).

Positive mean values were observed for all three measures. The FPRI-UPF showed the highest mean score (1.13; SD = 1.27), followed by the overall FPRI (0.96; SD = 1.12) and the FPRI-HF (0.79; SD = 1.40). Median values followed a similar pattern, with scores of 0.96 for ultra-processed foods, 0.74 for the overall index, and 0.56 for healthy foods.

According to the operationalization of the index, these positive values indicate that, on average, taste ratings exceeded the stated preference for consuming the foods. The higher score observed for the FPRI-UPF reflects a greater weighted discrepancy between the hedonic appeal of taste and the preference for consuming ultra-processed foods compared with healthy foods. The Variability was greatest for the FPRI-HF (SD = 1.40), followed by the FPRI-UPF (SD = 1.27) and the overall FPRI (SD = 1.12).

Table 1. Descriptive statistics for the Food Preference Regulation Index (FPRI), by food type

Statistics		Overall FPRI	FPRI-UPF	FPRI-HF
Mean		.96	1.13	.79
Median		.74	.96	.56
Mode		-.06	.00	.00
SD		1.12	1.27	1.40
Variance		1.27	1.61	1.95
Rank		6.48	9.80	9.28
Min		-1.82	-3.88	-3.28
Max		4.66	5.92	6.00
Percentiles	25	,1150	,2000	-,0900
	50	,7400	,9600	,5600
	75	1,6550	1,7300	1,6400

Overall FPRI: Overall Food Preference Regulation Index; FPRI-UPF: Food Preference Regulation Index for Ultra-Processed Foods; FPRI-HF: Food Preference Regulation Index for Healthy Foods.sss

3.2. Wilcoxon test results

The Kruskal-Wallis test revealed statistically significant differences in the overall Food Preference Regulation Index (FPRI) across the nine school groups evaluated ($\chi^2 = 35.2$, df = 8, $p < 0.001$). The effect size associated with these differences was $\varepsilon^2 = 0.148$.

Post hoc pairwise comparisons using the Dwass-Steel-Critchlow-Fligner test identified seven statistically significant pairwise differences (Table 2). These differences primarily involved the 4th Grade B and 5th Grade B groups.

Table 2. Post Hoc Comparisons of the Overall Food Preference Regulation Index (FPRI) Across School Groups Showing Statistically Significant Differences.

Comparison	p-value
4to A vs 4to B	0.047
4to B vs 4to C	0.038
4to B vs 5to A	0.005
4to B vs 6to B	0.009
4to C vs 5to B	0.026
5to A vs 5to B	0.020
5to B vs 6to B	0.027

The differences observed across school groups indicate that food preference regulation was not homogeneous within the study population. This variability is consistent with existing evidence showing that food preferences during childhood are shaped by multiple influences and experiences and may vary considerably among children. Beckerman et al. (2017) describe the development of children's food preferences as a dynamic process in which biological predispositions interact with environmental experiences, while Smith et al. (2017) demonstrate that individual differences in food preferences are strongly associated with environmental factors.

Nevertheless, the differences identified among school groups do not allow conclusions to be drawn regarding the factors underlying the observed variability in the FPRI. Although previous research has linked family, social, and food environment characteristics to the development of food preferences, these variables were not assessed in the present study. Therefore, school group membership should be interpreted solely as a grouping factor within which differences in FPRI were observed, rather than as a causal exposure or determinant of food preference regulation.

The concentration of significant pairwise comparisons in the 4th Grade B and 5th Grade B groups further suggests the presence of distinct patterns within the sample. These findings should be interpreted with caution, as post hoc analyses identify which groups differ from one another but do not, by themselves, explain the mechanisms underlying such differences. A more comprehensive characterization of these patterns would require future studies to incorporate family and school environmental variables, food exposure measures, and other factors associated with the development of food preferences.

3.3. Differences in FPRI According to Food Type

The Wilcoxon signed-rank test was used to compare the Food Preference Regulation Index for healthy foods (FPRI-HF) and ultra-processed foods (FPRI-UPF). The distribution of ranks showed that 136 schoolchildren had FPRI-HF values lower than their FPRI-UPF values, 91 had FPRI-HF values higher than their FPRI-UPF values, and 10 showed identical values in both categories (Table 3).

The difference between the indices was statistically significant ($Z = -3.739$, $p < 0.001$), indicating differences in the distribution of the FPRI according to food type. Consistent with the descriptive findings, the FPRI was higher for ultra-processed foods (mean = 1.13) than for healthy foods (mean = 0.79).

According to the operationalization of the index, this finding indicates that ultra-processed foods exhibited, on average, a greater discrepancy between taste ratings and the stated preference for consuming them, weighted by perceived healthiness. This pattern is consistent with a greater relative regulation of preference in the face of the hedonic appeal of ultra-processed foods.

Table 3. Rank Distribution in the Wilcoxon Test.

Ranges				
		N	**Mean Rank**	**Sum of Ranks**
FPRI-HF - FPRI-UPF	Negative Ranks	135[a]	122,49	16781,00
	Positive Ranks	91[b]	102,47	9325,00
	Ties	10[c]		
	Total	236		
a. FPRI-HF < FPRI-UPF				
b. FPRI-HF > FPRI-UPF				
c. FPRI-HF = FPRI-UPF				

The observed difference between FPRI-UPF and FPRI-HF should be interpreted in light of the fact that taste, consumption preference, and perceived healthiness are related but distinct constructs. Studies that have examined these attributes separately have shown that children integrate both hedonic and health-related information when forming food preferences and making food choices, and that the relative influence of each attribute varies according to the food item and individual child characteristics (Serrano-Gonzalez et al., 2021; Rigo et al., 2023). Consequently, a food receiving a high taste rating does not necessarily imply an equally high preference for consuming it.

This distinction among components of food-related responses has also been reported in experimental studies. Pearce et al. (2020) found that health, taste, and desire-to-eat ratings contributed differently to food choice processes in children, while Liem and Zandstra (2009) demonstrated that liking and wanting may follow distinct patterns in response to food exposure. These findings support the rationale for analyzing hedonic appeal and stated consumption preference as separate dimensions rather than assuming that they represent the same underlying phenomenon.

The higher FPRI-UPF score observed in the present study may therefore be interpreted as reflecting a greater separation between the perceived taste appeal of ultra-processed foods and the stated preference for consuming them, while also taking into account the level of healthiness attributed to these products by the schoolchildren. This finding is relevant because previous research has shown that taste is a major determinant of children's food decisions (Bruce et al., 2016; Serrano-Gonzalez et al., 2021), and those considerations of healthfulness can modify those decisions (van Meer et al., 2019). Taken together, these results suggest that children's food evaluations are influenced not only by sensory appeal but also by the interaction between hedonic and evaluative dimensions.

Nevertheless, this interpretation differs from a direct assessment of dietary self-control. Ha et al. (2019) examined the relationship among perceived healthiness, taste, and preference in children using an independent measure of self-control and reported associations between this construct and evaluations of both healthy and unhealthy foods. Similarly, Serrano-Gonzalez et al. (2021) assessed self-control success through food choices made in situations involving conflicts between taste and healthfulness. In the present study, neither food choice behavior nor an independent measure of self-control was evaluated. Therefore, the FPRI should be understood as an indicator of the relative regulation of food preference, derived from the discrepancy between the assessed dimensions, rather than as a direct measure of inhibitory capacity, actual food consumption, or psychological self-control.

3.4. Pattern shown in the dendrogram created with Agglomerative Hierarchical Clustering (AHC) Algorithm

The dendrogram generated through agglomerative hierarchical clustering, utilizing the Canberra dissimilarity metric and complete linkage method, illustrates a hierarchical structure in the food perception patterns among the 236 students.

The analysis identified three groups with distinct patterns in the evaluation of healthy and ultra-processed foods. Responses showing greater similarity across the dimensions of taste, consumption preference, and perceived healthiness were clustered together at earlier stages, forming shorter branches, whereas greater differences in response patterns were reflected by longer branches.

A predominant cluster (Group 1, red; n = 189) was identified, characterized by a stronger attraction toward ultra-processed foods (UPFs). A second cluster (Group 2, blue; n = 45) exhibited a more balanced preference pattern or a greater orientation toward healthy foods. A third, smaller cluster (Group 3, green; n = 2) consisted of atypical profiles displaying discordant patterns among perceived healthiness, taste, and consumption preference. This latter group showed the greatest inconsistency across the three evaluated dimensions.

Overall, the dendrogram revealed substantial heterogeneity in food perception and preference patterns among the schoolchildren included in the study (Figure 1).

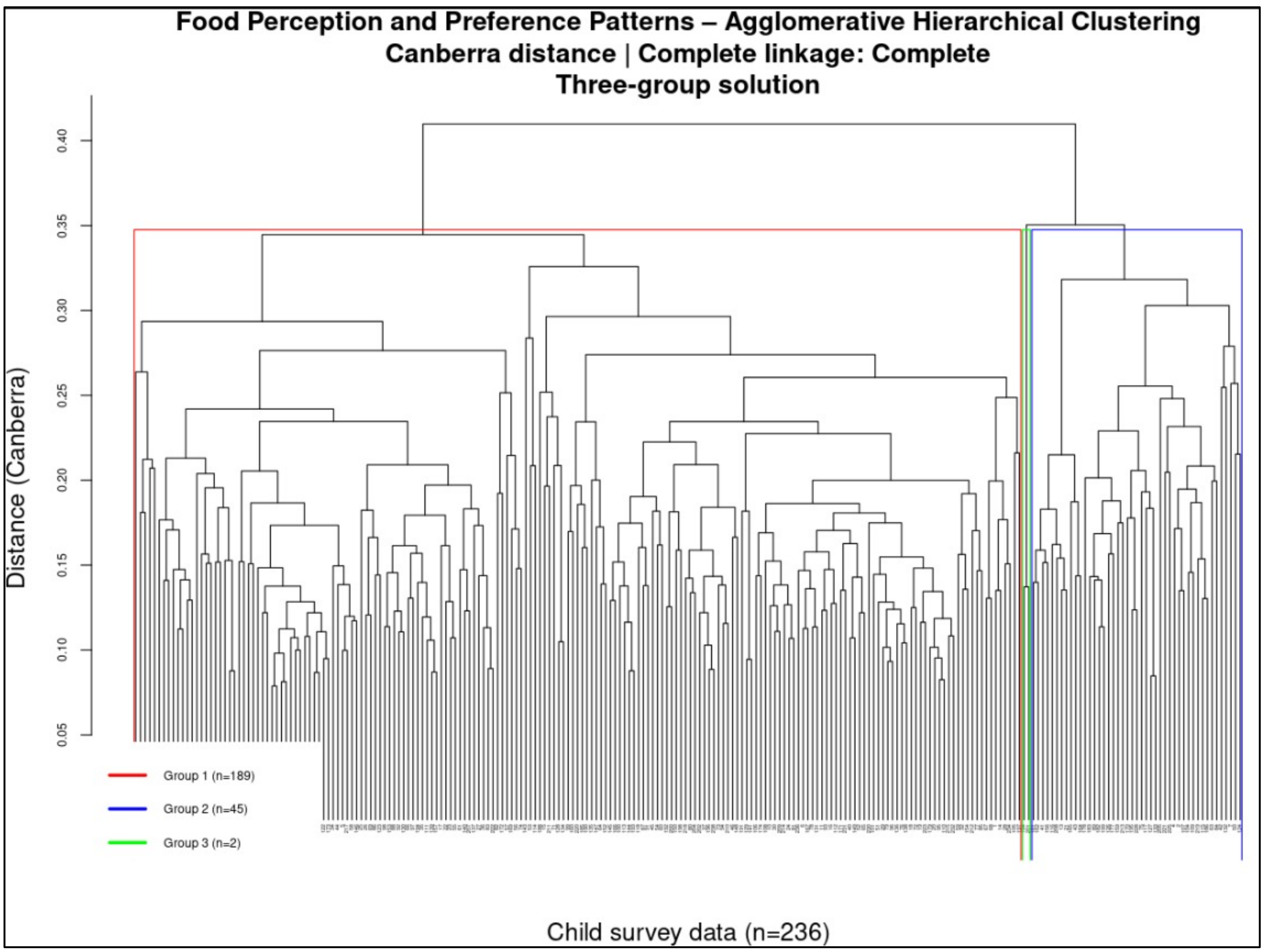

Fig. 1 *Clustering pattern revealed with agglomerative hierarchical clustering algorithm.*

The identification of distinct patterns is consistent with the multidimensional nature of children's food preferences. Evidence indicates that food preferences develop through the interaction of sensory characteristics of foods with individual and environmental experiences. Consequently, distinct patterns of food evaluation may emerge even within relatively homogeneous pediatric populations (Beckerman et al., 2017). In particular, the predominance of a group characterized by a stronger attraction toward ultra-processed foods (UPFs) is consistent with studies highlighting the central role of taste in food evaluation and food choice during childhood (Bruce et al., 2016; Serrano-Gonzalez et al., 2021).

The presence of a second group exhibiting a more balanced pattern across food categories further supports the existence of heterogeneity in the integration of the evaluated attributes. Experimental studies have shown that taste and perceived healthiness may contribute differentially to food evaluation and choice, and that health considerations can modify the relative influence of hedonic attributes during food-related decision-making (Pearce et al., 2020; van Meer et al., 2019). In this context, the pattern observed in Group 2 may reflect a different relationship among taste, perceived healthiness, and consumption preference, although the available data do not allow this configuration to be attributed to any specific self-regulatory mechanism.

Group 3 consisted of only two schoolchildren and displayed a pattern that was clearly differentiated from those of the main clusters. Although their responses suggest greater discordance among perceived healthiness, taste, and consumption preference, the very small number of cases prevents this cluster from being considered a stable or generalizable profile. Accordingly, its interpretation should be regarded as strictly exploratory. Nevertheless, the presence of these cases highlights the existence of individual patterns that are not represented by the predominant configurations and that could be examined in studies with larger sample sizes.

Overall, the Agglomerative Hierarchical Clustering (AHC) analysis identified heterogeneity in schoolchildren's patterns of food perception and preference, thereby complementing the differences previously observed in the Food Preference Regulation Index (FPRI). However, these clusters should be understood as exploratory patterns derived from the assessed dimensions rather than as diagnostic categories or psychological profiles of self-regulation. Their interpretation is limited to the relationships observed among perceived healthiness, taste, and consumption preference within the study sample.

3.5. Multidimensional Scaling (MDS)

Multidimensional Scaling (MDS), based on the Canberra dissimilarity matrix, was used to represent similarities and differences in schoolchildren's response patterns within a two-dimensional space. The projection showed a spatial differentiation among the three groups identified through the hierarchical clustering analysis, although with varying degrees of overlap between them (Figure 2). The two-dimensional representation showed that Group 1 (red; n = 189) contained the majority of the participants, whereas Group 2 (blue; n = 45) exhibited a distribution that was partially overlapping with that of Group 1. Group 3 (green; n = 2) was positioned separately from the two main groups in the multidimensional space.

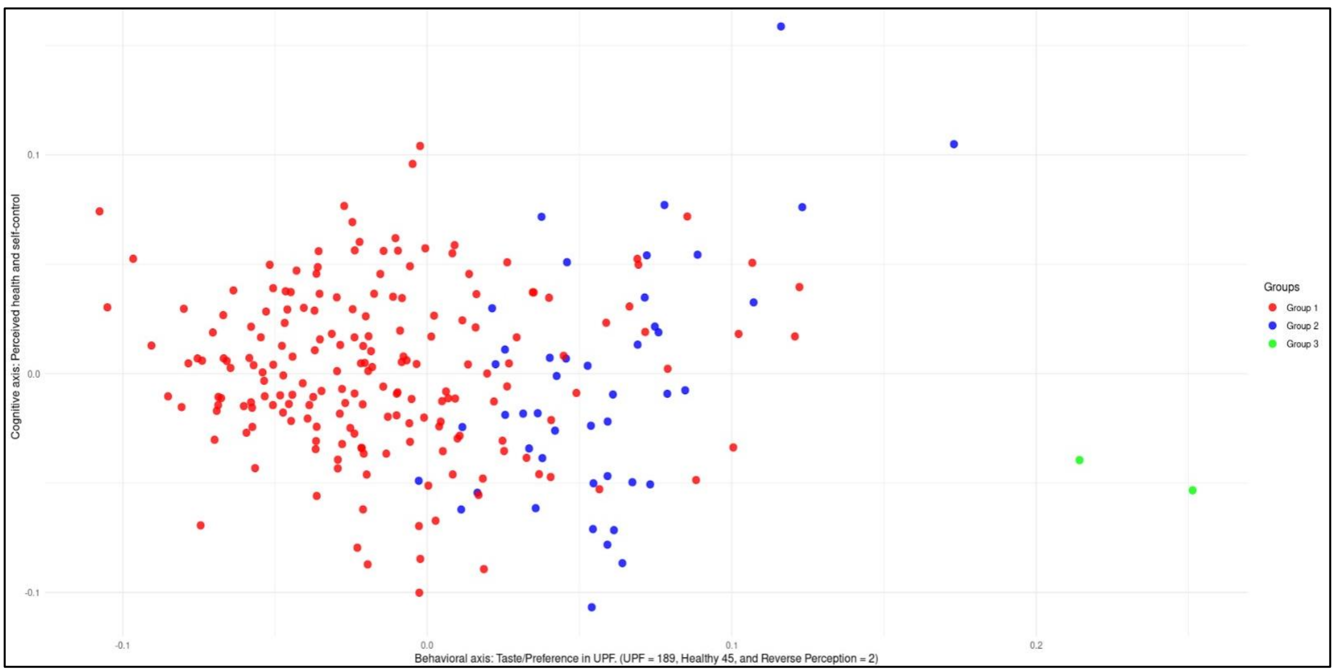

Fig. 2 *Patterns revealed by Multidimensional Scaling (MDS).*

The MDS representation complemented the results of the hierarchical clustering analysis by showing that the observed heterogeneity did not correspond to a fully distinct separation among schoolchildren. The partial overlap between Groups 1 and 2 suggests the existence of gradual patterns of similarity across the evaluated dimensions rather than mutually exclusive categories of food-related responses. This interpretation is consistent with evidence describing children's food preferences as the result of the variable integration of attributes such as taste and perceived healthiness during food evaluation and choice processes (Serrano-Gonzalez et al., 2021; Pearce et al., 2020). The spatial separation of Group 3 should be interpreted with particular caution because it comprised only two participants and therefore does not provide sufficient evidence to support the existence of a stable profile within the study population.

Overall, the findings indicate that children's food evaluations integrate perceived healthiness, taste, and consumption preference, and that these three dimensions do not always operate in the same manner. This distinction has important implications for the design of food and nutrition education strategies, as providing information about the nutritional quality of foods may be insufficient if hedonic attributes involved in the formation of preferences and food choices are overlooked. Experimental evidence has shown that taste plays a major role in children's food-related decisions and that directing attention toward healthfulness can influence the decision-making process (Bruce et al., 2016; van Meer et al., 2019).

From this perspective, interventions targeting school-aged children could address two complementary objectives: promoting the recognition of food healthfulness and fostering experiences that align food preferences with healthier dietary choices. However, the findings of the present study do not allow conclusions to be drawn regarding the effectiveness of specific intervention strategies, nor do they permit the observed patterns to be attributed to particular characteristics of the school or family environment. Accordingly, these implications should be regarded as directions for future interventions and research rather than definitive conclusions.

Methodologically, this study combines an indicator derived from ratings of perceived healthiness, taste, and consumption preference with multivariate techniques designed to explore heterogeneity in food-related response patterns. The Food Preference Regulation Index (FPRI) quantifies the weighted discrepancy between the hedonic appeal of taste and the stated preference for consuming foods, whereas Agglomerative Hierarchical Clustering (AHC) and Multidimensional Scaling (MDS) provide complementary perspectives on the similarities and differences among schoolchildren's response patterns.

Several limitations of this approach should be acknowledged. The FPRI is an operational measure specifically developed for this study and does not constitute a direct or previously validated measure of dietary self-control. Likewise, the identified clusters represent exploratory patterns within the analyzed sample rather than diagnostic categories or behavioral profiles that can be generalized to broader populations. This consideration is particularly important for Group 3, which consisted of only two participants.

4 Conclusions

The study revealed differences in food preference regulation among schoolchildren according to food type. The FPRI was significantly higher for ultra-processed foods than for healthy foods, indicating a greater weighted discrepancy between the hedonic appeal of taste and the stated preference for consuming UPFs. This finding suggests that children's food evaluations integrate the dimensions of taste, consumption preference, and perceived healthiness, which, although related, do not necessarily operate in an equivalent manner.

Furthermore, the differences observed across school groups, together with the patterns identified through Agglomerative Hierarchical Clustering (AHC) and visualized using Multidimensional Scaling (MDS), demonstrated heterogeneity in the way schoolchildren evaluated foods. These techniques complemented the FPRI by providing a multivariate perspective on the relationships among the assessed dimensions. However, the resulting clusters should be interpreted as exploratory patterns within the study sample rather than as psychological profiles or diagnostic categories.

Taken together, the FPRI and multivariate analyses provide an approach for examining the relationship between hedonic and evaluative components of children's food preferences. These findings may contribute to the development of future research aimed at further investigating the factors associated with the formation of preferences for ultra-processed and healthy foods, incorporating measures of actual food consumption as well as individual and contextual variables to assess the utility and reproducibility of the identified patterns.

Acknowledgments

The authors express their sincere gratitude to the administration, teaching staff, and administrative personnel of the elementary school “Josefa Ortiz de Dominguez” in Ciudad del Carmen, Campeche, for the support provided for this study. Their invaluable collaboration and willingness to assist were fundamental to administering the research instruments to the students. Likewise, deep gratitude is extended to all the students who participated voluntarily and enthusiastically in this project.

Funding

This research was conducted without the support of any funding institution.

Conflict Of Interest

“The authors state that there are no conflicts of interest in writing the manuscript.”

Data Availability

The data set generated during and/or analyzed in this study is available upon reasonable request to the corresponding author.

References

Beckerman, J. P., Alike, Q., Lovin, E., Tamez, M., & Mattei, J. (2017). The development and public health implications of food preferences in children. *Frontiers in Nutrition, 4*, Article 66. https://doi.org/10.3389/fnut.2017.00066

Buja, A., Swayne, D. F., Littman, M. L., Dean, N., Hofmann, H., & Chen, L. (2008). Data visualization with multidimensional scaling. *Journal of Computational and Graphical Statistics, 17*(2), 444–472. https://doi.org/10.1198/106186008X318440

Estrella Barrón, R., & Telumbre Terrero, J. Y. (2024). Caracterización del consumo de alimentos ultraprocesados en alumnos universitarios del área de Ciencias de la Salud. *Ciencia Latina Revista Científica Multidisciplinar, 8*(5), 1609–1622. https://doi.org/10.37811/cl_rcm.v8i5.13633

García-Laencina, P. J., Sancho-Gómez, J. L., & Figueiras-Vidal, A. R. (2010). Pattern classification with missing data: A review. *Neural Computing and Applications, 19*(2), 263–282. https://doi.org/10.1007/s00521-009-0295-6

Giordani, P., Ferraro, M. B., & Martella, F. (2020). Hierarchical clustering. In *An introduction to clustering with R* (pp. 9–73). Springer. https://doi.org/10.1007/978-981-13-0553-5_2

Ha, O.-R., Lim, S.-L., Bruce, J. M., & Bruce, A. S. (2019). Unhealthy foods taste better among children with lower self-control. *Appetite, 139*, 84–89. https://doi.org/10.1016/j.appet.2019.04.015

Hernández-Gómez, H. J., Canul-Reich, J., Hernández-Ocaña, B., & de la Cruz Hernández, E. (2023). An agglomerative hierarchical clustering approach to identify coexisting bacteria in groups of bacterial vaginosis patients. *Intelligent Data Analysis, 27*(3), 583–611. https://doi.org/10.3233/IDA-216488

Lance, G. N., & Williams, W. T. (1966). Computer programs for hierarchical polythetic classification ("similarity analyses"). *The Computer Journal, 9*(1), 60–64. https://doi.org/10.1093/comjnl/9.1.60

Lance, G. N., & Williams, W. T. (1967). Mixed-data classificatory programs I—Agglomerative systems. *Australian Computer Journal, 1*(1), 15–20.

Lane, M. M., Gamage, E., Du, S., Ashtree, D. N., McGuinness, A. J., Gauci, S., Baker, P., Lawrence, M., Rebholz, C. M., Srour, B., Touvier, M., Jacka, F. N., O'Neil, A., Segasby, T., & Marx, W. (2024). Ultra-processed food exposure and adverse health outcomes: Umbrella review of epidemiological meta-analyses. *BMJ, 384*, e077310. https://doi.org/10.1136/bmj-2023-077310

Liem, D. G., & Zandstra, L. H. (2009). Children's liking and wanting of snack products: Influence of shape and flavour. *International Journal of Behavioral Nutrition and Physical Activity, 6*, Article 38. https://doi.org/10.1186/1479-5868-6-38

Little, R. J. A., & Rubin, D. B. (2019). *Statistical analysis with missing data* (3rd ed.). Wiley. https://doi.org/10.1002/9781119482260

Manterola, C., & Otzen, T. (2014). Estudios observacionales: Los diseños utilizados con mayor frecuencia en investigación clínica. *International Journal of Morphology, 32*(2), 634–645. https://doi.org/10.4067/S0717-95022014000200042

Pacheco, E. R. (2015). *Unsupervised learning with R*. Packt Publishing.

Pearce, A. L., Adise, S., Roberts, N. J., White, C., Geier, C. F., & Keller, K. L. (2020). Individual differences in the influence of taste and health impact successful dietary self-control: A mouse tracking food choice study in children. *Physiology & Behavior, 223*, Article 112990. https://doi.org/10.1016/j.physbeh.2020.112990

Rigo, M., Mohebbi, M., Keast, R., Harrison, P., Kelly, M., Olsen, A., Bredie, W. L. P., & Russell, C. G. (2023). An investigation into food choices among 5–12 years children in relation to sensory, nutritional, and healthy product cues. *Food Quality and Preference, 111*, Article 104990. https://doi.org/10.1016/j.foodqual.2023.104990

Rubin, D. B. (1976). Inference and missing data. *Biometrika, 63*(3), 581–592. https://doi.org/10.1093/biomet/63.3.581

Schafer, J. L. (1999). Multiple imputation: A primer. *Statistical Methods in Medical Research, 8*(1), 3–15. https://doi.org/10.1177/096228029900800102

Serrano-Gonzalez, M., Herting, M. M., Lim, S.-L., Sullivan, N. J., Kim, R., Espinoza, J., Koppin, C. M., Javier, J. R., Kim, M. S., & Luo, S. (2021). Developmental changes in food perception and preference. *Frontiers in Psychology, 12*, Article 654200. https://doi.org/10.3389/fpsyg.2021.654200

Smith, A. D., Fildes, A., Forwood, S., Cooke, L., & Llewellyn, C. (2017). The individual environment, not the family is the most important influence on preferences for common non-alcoholic beverages in adolescence. *Scientific Reports, 7*, Article 16822. https://doi.org/10.1038/s41598-017-17020-x

van Meer, F., van der Laan, L. N., Eiben, G., Lissner, L., Wolters, M., Rach, S., Herrmann, M., Erhard, P., Molnar, D., Orsi, G., Viergever, M. A., Adan, R. A. H., Smeets, P. A. M., & I.Family Consortium. (2019). Development and body mass inversely affect children's brain activation in dorsolateral prefrontal cortex during food choice. *NeuroImage, 201*, Article 116016. https://doi.org/10.1016/j.neuroimage.2019.116016

von Hippel, P. T. (2013). Should a normal imputation model be modified to impute skewed variables? *Sociological Methods & Research, 42*(1), 105–138. https://doi.org/10.1177/0049124112464866