



User experience of urban public transport in Mazatlán: A preliminary review of a scale

Experiencia del usuario del transporte público urbano de Mazatlán: revisión preliminar de una escala

Génesis Cristal Guzmán Oros¹  
Víctor Manuel Martínez García¹  
Yennifer Díaz Romero¹  
Pedro Antonio Valdez Lizárraga¹  
José de Jesús Reyes Machain¹  
Sergio Armando Morán Sánchez¹  

¹ Universidad Autónoma de Sinaloa, Vicerrectoría. Mazatlán, México.

Submitted: 08-07-2026 | Revised: 11-09-2026 | Accepted: 16-10-2026 | Published: 17-10-2026

Corresponding Author: Víctor Manuel Martínez García 

ABSTRACT

Urban public transport is part of the experience of living in the city, and travel quality is related to user well-being. This instrumental, cross-sectional and exploratory study carried out a preliminary review of a ten-item scale with two dimensions, perceived transport conditions and tension during the trip, administered in the pilot survey of project FZ-HAB-0023 of the Factoriza UAS platform to urban public transport users in Mazatlán, Sinaloa, Mexico. A preliminary cut of 6 cases, out of a target of 100, was analyzed with descriptive statistics, Cronbach's alpha, McDonald's omega and inter-item correlations; factor analysis was not interpretable given the size of the cut. Tension during the trip showed high internal consistency ($\alpha = .855$; $\omega = .863$), with worry about arriving late as the highest-scoring item, whereas perceived transport conditions showed a negative alpha ($\alpha = -.3359$), suggesting that it groups heterogeneous service attributes such as frequency, comfort, cleanliness, crowding and shade at the stop. Respondents rated bus comfort, cleanliness and crowding unfavorably. The results do not estimate the properties of the scale, but they guide its revision before the final data cut.

Keywords: public transport; user experience; commuting stress; psychometric evidence.

RESUMEN

El transporte público urbano forma parte de la experiencia de habitar la ciudad, y la calidad del traslado se relaciona con el bienestar de las personas usuarias. Este estudio instrumental, transversal y exploratorio realizó una revisión preliminar de una escala de diez reactivos con dos dimensiones, Condiciones percibidas del transporte y Tensión durante el traslado, aplicada en el levantamiento piloto del proyecto FZ-HAB-0023 de Factoriza UAS a personas usuarias del transporte público urbano de Mazatlán, Sinaloa. Se analizó un corte preliminar de 6 casos, de una meta de 100, con estadísticos descriptivos, alfa de Cronbach, omega de McDonald y correlaciones entre reactivos; el análisis factorial no fue interpretable por el tamaño del corte. Tensión durante el traslado mostró una consistencia interna alta ($\alpha = .855$; $\omega = .863$), con la preocupación por llegar tarde como reactivo de media más alta, mientras que Condiciones percibidas del transporte obtuvo un alfa negativo ($\alpha = -.3359$), indicio de que reúne atributos heterogéneos del servicio, como la frecuencia, la comodidad, la limpieza, la ocupación y la sombra en la parada. Las personas encuestadas valoraron de manera poco favorable la comodidad, la limpieza y la ocupación de los camiones. Los resultados no estiman las propiedades de la escala, pero orientan su revisión antes del corte final.

Palabras clave: transporte público; experiencia del usuario; tensión durante el traslado; evidencias psicométricas.

INTRODUCTION

Urban public transport is part of the everyday experience of inhabiting the city. For those who depend on it, the journey between home, work, school, and services is an additional component of housing and neighborhood conditions, and the quality of that journey—its regularity, comfort, and safety—is related to the well-being of users. A critical review of the literature on commuting and well-being found that mood during the commute to work is lower than during other daily activities, that congestion, crowding, and unpredictability can trigger stress, and that public transport users are among those least satisfied with their commute (Chatterjee et al., 2020). This work, which predates the 5-year period prioritized by the search, is included because it is the reference review on the topic.

The quality of bus service has mainly been studied from the perspective of user satisfaction. A systematic review of 104 studies on bus systems in developing countries identified safety, comfort, reliability, and accessibility as the attributes with the greatest weight in the perception of service quality and in satisfaction, and noted that these attributes are especially important while waiting at the stop, on board, and during the walk to the destination (Sogbe et al., 2025). In warm climates, waiting at the stop also involves heat exposure: in Austin, Texas, trees around stops were associated with a smaller decline in passenger numbers during high temperatures (Lanza y Durand, 2021).

Most of this literature comes from large metropolitan areas in Asia, Europe, North America, and developing countries with large-scale bus systems (Abu Bakar et al., 2022; Sogbe et al., 2025), and the instruments used in these studies usually measure satisfaction with the service or the general well-being associated with travel. The literature reviewed for this study did not identify any brief instruments in Spanish that jointly measure the assessment of urban bus service conditions and the stress states that users experience during travel in a medium-sized warm-climate city in northwestern Mexico. This absence justifies the design of a dedicated scale, but it also requires gathering evidence of its functioning before using it to describe the experience of the user population or to relate it to other dimensions of habitability.

The FZ-HAB-0023 project of the Factoriza UAS platform was originally proposed as a study of the psycho-environmental habitability of urban public transport in Mazatlán and was reformulated to focus on the experience of users. To this end, a 10-item scale was designed, organized into 2 dimensions, Perceived transport conditions and Tension during travel, and a pilot survey was initiated on the city's routes and stops. This article presents the results of preliminary wave 1 of that survey. The general objective is to conduct a preliminary review of the scale's functioning. The specific objectives are to describe the responses to the items, estimate the internal consistency of each dimension, examine the correlations among the items, and identify the adjustments that the instrument requires before the final wave.

Public transport as part of urban habitability

Habitability is not limited to housing: it includes access to the places of everyday life and the conditions under which that access occurs. A systematic review of the factors associated with residential satisfaction grouped these factors into categories that interact with each other, ranging from functional and technical to psychological and public policy (Saleem and Alchalabi, 2025). For people who do not own an automobile, public transport is the condition for accessing the city, and its quality is part of the experience of inhabiting it.

Studies on bus service quality agree on a set of attributes. A systematic review of 41 articles on bus performance in Asia, most of them conducted in China, identified frequency, punctuality, reliability, comfort, safety, and travel time as attributes linked to service performance and passenger satisfaction, and noted that passenger and operator perceptions are related to trust in the service (Abu Bakar et al., 2022). The review by Sogbe et al. (2025) added that the connection between the stop and the origin or destination of the trip, the so-called first and last mile, is a pending challenge in developing countries.

Tension, crowding, and well-being during commuting

The relationship between commuting and well-being has been examined in detail. A systematic review of 45 observational studies found that long commutes and congestion were consistently associated with lower satisfaction and lower well-being, and it brought together studies in which perceived crowding was related to passenger stress and exhaustion, even when objective crowding was low (Liu et al., 2022). In an experiment with 360 subway users, perceived crowding and environmental comfort mediated the relationship between the visual design of stations and

momentary affective well-being (Zhu and Yu, 2026), indicating that the perception of crowding also depends on the environment and not only on the number of passengers.

Safety is another source of tension during travel. The systematic review by te Braak and van Tienoven (2025) on women's perception of safety in urban public space proposed a framework in which personal, social, environmental, and temporal factors interact, such as the difference between day and night, factors that also play a role in waiting for and traveling on public transport. In warm climates, finally, heat exposure has been associated with mood: a review with meta-analysis found that heat waves and extremely high temperatures were associated with a higher risk of mental and behavioral disorders (Li et al., 2023), which underscores the relevance of shade at stops and the thermal comfort of vehicles.

Service attributes and states during transfer: two measurement objects

The reviewed literature makes it possible to distinguish two types of content in the measurement of public transport experience. The first comprises assessments of external service attributes, such as frequency, punctuality, comfort, cleanliness, occupancy, or stop equipment, which service quality reviews treat as differentiated components (Abu Bakar et al., 2022; Sogbe et al., 2025). The second comprises the states that a person experiences during travel, such as bad mood, worry, calmness, or fatigue, which the literature on travel and well-being studies as affective dimensions of experience (Chatterjee et al., 2020; Liu et al., 2022).

The distinction has implications for analysis. The affective states of a single experience tend to vary together, as they reflect the same perception of the journey, and it is reasonable to expect high internal consistency among them. Service attributes, on the other hand, may vary independently: frequency depends on route operation, cleanliness on vehicle maintenance, and shade on the equipment at each stop. That a person rates one of these attributes highly does not imply that they will rate the others highly as well, so the internal consistency of a set of attributes is not as clear a requirement as it is for affective states. This distinction guides the interpretation of the results for the two dimensions of the scale.

Validity evidence in the initial stage of a scale

The internal consistency of a dimension is usually estimated with Cronbach's alpha (Cronbach, 1951), which assumes equivalent loadings among items, and with McDonald's omega (McDonald, 1999/2013), which does not assume them; simulation studies have compared the performance of both coefficients under different data conditions (Malkewitz et al., 2023). Internal structure is explored with exploratory factor analysis, which requires first verifying data adequacy and sufficient sample size (Ferrando et al., 2022; Goretzko et al., 2021). Scales with items in both directions can show wording effects, because a proportion of people respond inconsistently to positive and negative items of the same dimension (Steinmann et al., 2022), and identifying these responses can reduce such effects (García-Batista et al., 2021). The project scale includes a reverse-scored item in each dimension.

METHODS

Design

An instrumental, quantitative, non-experimental, cross-sectional study with a descriptive and exploratory scope was conducted. The data correspond to survey LEV-0028, "Mazatlán Pilot · Transport User Experience," from project FZ-HAB-0023 of the Factoriza UAS platform, and to the frozen dataset DST-0027, with a cutoff date of June 14 2026. The survey constitutes one independent sample and is the only one in the project.

Population, sample, and data cut-off

The population was defined as individuals aged 18 years or older who use the urban public transport in Mazatlán, Sinaloa, and the survey was conducted on routes and stops selected by the project lead. According to the Factoriza log, 15 individuals opened the questionnaire and 6 completed it; no cases were excluded for quality control, so 6 cases were analyzed (Table 2).

The analyzed cutoff is preliminary: it includes 6 cases out of a target of 100, and the survey remained open, in field status, as of the cutoff date. With 6 cases for 10 items, the ratio is 0.6 cases per item, so that the correlation matrix among the 10 items is not invertible.

Instrument

A ten-item scale was applied with a five-point response format, ranging from Strongly disagree (1) to Strongly agree(5), organized into two dimensions of five items each: Perceived transport conditions (R1 to R5) and Tension during travel (R6 to R10). Two items are worded in the reverse direction, R4 and R9, and are recoded in the data matrix, so that a higher score indicates better perceived conditions in the first dimension and greater tension in the second. Table 1 presents the code, dimension, text, and direction of each item. In Factoriza's item dictionary, the instrument appears as the Public Transport User Experience Scale.

Table 1. Structure of the instrument.

Code	Dimension	Reagent	Inverse
R1	Conditions	The truck I use runs regularly	No
R2	Conditions	I travel comfortably inside the truck	No
R3	Conditions	The trucks I use are clean	No
R4	Conditions	I travel squeezed in among other passengers in the truck	Yes
R5	Conditions	The bus stop where I wait has shade	No
R6	Tension	Traveling by truck puts me in a bad mood	No
R7	Tension	I worry about arriving late when I use the truck	No
R8	Tension	I feel unsafe when traveling by bus	No
R9	Tension	I travel by truck with peace of mind	Yes
R10	Tension	When I get off the truck, I feel tired	No

Note. Conditions = Perceived transport conditions; Tension = Tension during transfer. 5-point response scale (Strongly disagree, Disagree, Neither agree nor disagree, Agree, Strongly agree). Reverse-coded items are already recoded in the delivered matrix. Source: Factoriza UAS, project FZ-HAB-0023, survey LEV-0028, dataset DST-0027.

Procedure

Responses were collected using the questionnaire of the survey administered through Factoriza UAS, which records explicit online informed consent before the first item and does not collect identifying data. Factoriza froze the DST-0027 dataset with a cryptographic fingerprint, generated the matrix with the reverse-scored items recoded, and produced the descriptive, internal consistency, and within-dimension correlation analyses (runs RUN-00145 to RUN-00149). The complementary analyses were performed using the data file and item dictionary exported by the platform.

Data processing

The 6 analyzed cases answered the 10 items, so there were no missing values and it was not necessary to impute responses. The reverse-coded items, R4 and R9, were analyzed in the recoded version provided by the platform, and the supplementary program did not re-invert any item. The values in the tables were copied as produced by Factoriza UAS and the analysis program, without recalculation or further rounding; when 2 tables from the package were combined into a single table or a column was omitted, the table note indicates this. The program output files, execution log, and parameters used are compiled in the reproducibility appendix of the article.

Because the cut-off is preliminary, the coefficients are interpreted as information for reviewing the instrument rather than as estimates of its properties in the user population. For the same reason, p values and confidence intervals are reported to document the uncertainty of each estimate, not to support conclusions about the presence or absence of relationships.

Data analysis

In a first stage, Factoriza UAS calculated the mean, standard deviation, skewness, kurtosis, and response percentages at the extreme values of each item; the internal consistency of each dimension using Cronbach's alpha (Cronbach, 1951), including the corrected item-total correlation and alpha if the item was deleted; and Pearson correlations among the items of each dimension, with 95% confidence intervals via Fisher's transformation. In a second stage, the program analisis_psicometrico.py, version 1.0, was run with Python 3.13.16 (Python Software Foundation, 2026), with

NumPy 2.5.3 (Harris et al., 2020), SciPy 1.18.1 (Virtanen et al., 2020), and pandas 3.0.5. The program uses fixed and declared procedures, without manual intervention in the calculations.

The program estimated McDonald's total omega (McDonald, 1999/2013) for each dimension from a single-factor model, without reverse-scoring items. It also attempted an exploratory factor analysis of the 10 items, including sampling adequacy, a sphericity test, the number of factors via parallel analysis with 100 simulations, minimum residual extraction, and oblimin rotation. Finally, it estimated the relationship between the mean scores of the 2 dimensions using Spearman's coefficient and its 95 % confidence interval, via Fisher transformation with the Bonett and Wright standard error (2000). Reference values for psychometric indices are treated as guidelines and not as decision rules (Ferrando et al., 2022). No item was eliminated, and confirmatory factor analysis was not performed.

Ethical considerations

Participation was voluntary and anonymous. Factoriza UAS records explicit online informed consent and does not collect names, email addresses, or other identifying data.

Statement on the use of artificial intelligence

During the preparation of this manuscript, the Factoriza UAS platform was used for data management and for descriptive, internal consistency, and inter-item correlation analyses; a Python analysis program with fixed and validated procedures was used for omega, exploratory factor analysis, and correlation between dimensions; and the Claude Opus language model (Anthropic) was used to run that program, assist with the literature search, drafting, and document assembly based on the results. The authors reviewed, corrected, and approved the entire text, verified each reference, and assume full responsibility for the content of the article.

Results

Sample composition

Table 2 summarizes the case flow. Of the 15 people who opened the questionnaire, 6 completed it and were analyzed, with no exclusions due to quality control. The 6 analyzed cases answered the 10 items.

Perceived transport conditions and tension during transfer

Table 3 and Figure 1 present the descriptive statistics. For the Perceived transport conditions dimension, all means were at or below the midpoint of the scale. The regularity with which the truck passes (R1: $M = 3.17$, $DE = 0.75$) and shade at the waiting stop (R5: $M = 3.17$, $DE = 1.47$) had the highest means in the dimension, the latter showing the greatest dispersion in the instrument. Truck cleanliness had a mean of 2.67 (R3: $DE = 0.52$), and comfort inside the truck (R2: $M = 2.33$, $DE = 0.52$) and the absence of crowding among passengers, in the recoded version of the reverse-scored item (R4: $M = 2.33$, $DE = 0.82$), had the lowest means in the instrument; R4 also showed a skewness of 2.45 and a kurtosis of 6.00.

Table 2. Sample composition.

Indicator	Value
Individuals who opened the questionnaire	15
Completed questionnaires	6
Cases excluded by quality control	0
Cases analyzed	6
Items	10

Note. Survey «Mazatlán Pilot - Transport User Experience» (LEV-0028); population: people aged 18 years or older who use urban public transport in Mazatlán; territory: Mazatlán, Sinaloa, on routes and stops selected by the project lead. Dataset DST-0027, fingerprint 655cab507f5183cb. Source: Factoriza UAS. Preliminary cut: $n = 6$ of a target of 100, with the survey open.

For Tension during transfer, concern about arriving late had the highest mean on the instrument (R7: $M = 4.17$, $DE = 0.98$). Bad mood when traveling by truck (R6) and the feeling of insecurity (R8) both had a mean of 3.50 ($DE = 0.84$), and tiredness when getting off the truck (R10: $M = 3.33$, $DE = 0.82$) and lack of calmness in the recoded version of the reverse-coded item (R9: $M = 3.33$, $DE = 1.03$) had a mean of 3.33. Overall, respondents rated truck comfort, cleanliness, and occupancy unfavorably and reported moderate tension during transfer, with punctuality as the main concern.

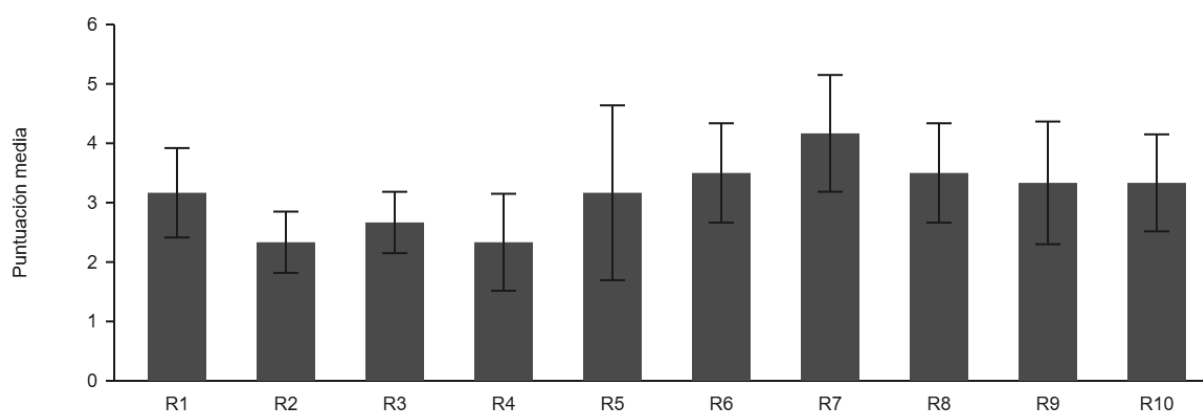
The shape of the distributions reflects the size of the subset. Bad mood (R6) and the feeling of insecurity (R8) had the same positive skewness (1.54) and the same kurtosis (1.43), and worry about being late had the most negative kurtosis on the instrument (R7: -2.39), which is characteristic of responses distributed over few values. The standard deviations for comfort and cleanliness (R2 and R3: $DE = 0.52$) were the lowest, indicating that responses to these items were concentrated at two adjacent values of the scale. With 6 cases, these shape indices are very sensitive to a single response, and they are reported as a description of the subset and not as evidence about the distribution of the items in the user population.

Table 3. Descriptive statistics for the items.

Item	M	DE	Skewness	Kurtosis	Floor	Ceiling
R1	3.17	0.75	-0.31	-0.10	17%	33%
R2	2.33	0.52	0.97	-1.88	67%	33%
R3	2.67	0.52	-0.97	-1.88	33%	67%
R4	2.33	0.82	2.45	6.00	83%	17%
R5	3.17	1.47	-0.42	-0.86	17%	17%
R6	3.50	0.84	1.54	1.43	67%	17%
R7	4.17	0.98	-0.46	-2.39	33%	50%
R8	3.50	0.84	1.54	1.43	67%	17%
R9	3.33	1.03	0.67	0.59	17%	17%
R10	3.33	0.82	-0.86	-0.30	17%	50%

Note. $N = 6$. Floor and ceiling indicate the percentage of responses at the minimum and maximum values, respectively. The item-total correlation column of the Factoriza descriptive table is omitted; its values do not match those in Table 4 because they are not calculated over the total for each dimension. Factoriza UAS execution: RUN-00145. Preliminary cutoff: $n = 6$ out of a target of 100, with data collection open.

Figure 1. Mean scores per item, with bars denoting standard deviations.



Descriptivos · $N = 6$ · ejecución RUN-00145

Note. $N = 6$. Scale from 1 to 5; reverse-scored items R4 and R9 are shown recoded. RUN-00145 was executed by Factoriza UAS. Preliminary cutoff: $n = 6$ of a target of 100, with data collection still open.

Perceived transportation conditions had an alpha of -3.359 . The five items of the dimension had negative corrected item-total correlations, between $-.739$ (R5) and $-.200$ (R4), and eliminating any of them, except for the shadow at the stop, would also leave a negative alpha; after eliminating R5, the alpha of the four remaining items would be $.094$, still close to zero. A negative alpha indicates that the average covariance among the items is negative; that is, in this sample, the items do not covary as indicators of the same attribute. Therefore, the software did not report omega for this dimension.

Table 4. Item analysis by dimension.

Item	Corrected item-total r	α if the item is removed
R1	-.675	-1.029
R2	-.533	-2.000
R3	-.354	-2.741
R4	-.200	-5.200
R5	-.739	.094
R6	.831	.785
R7	.660	.828
R8	.457	.874
R9	.778	.794
R10	.650	.830

Note. $N = 6$. R1 to R5 = Perceived transport conditions; R6 to R10 = Stress during transfer. The dimension column is omitted to preserve table legibility. The corrected item-total correlation excludes the item itself from its dimension total. Cronbach's α for each complete dimension (5 items each): Conditions = -3.359 (execution RUN-00146); Stress = $.855$ (RUN-00148). A negative alpha indicates that the average covariance among the items in the dimension is negative. Source: Factoriza UAS, Tables 4 and 6 from the package, combined without modifying values. Preliminary snapshot: $n = 6$ of a target of 100, with data collection open.

Table 5. Internal consistency by dimension.

Dimension	Items	n	α	ω
Conditions	5	6	-3.359	n/a
Tension	5	6	.855	.863

Note. α = Cronbach's alpha; ω = McDonald's total omega from a minimum-residual single-factor model without reverse-scoring items. Complete cases per dimension. Conditions = Perceived transport conditions; Stress = Stress during transfer. For Conditions, the program does not report ω because the five items correlate negatively with the rest of the dimension; this observation is transferred to the note for the corresponding column. Source: own elaboration with analysis_psicometrico.py 1.0 (Python 3.13.16), dataset DST-0027_FZ_HAB_0023. Preliminary cutoff: $n = 6$ out of a target of 100, with data collection open.

Table 6 presents all correlations among the items of each dimension. For Conditions, two negative correlations had confidence intervals that excluded 000: the correlation between truck regularity and cleanliness (R1 and R3: $r = -.857$, 95% CI $[-.984, -.151]$, $p = .029$) and that between on-board comfort and shade at the stop (R2 and R5: $r = -.877$, 95% CI $[-.986, -.227]$, $p = .022$); by contrast, comfort was positively related to the absence of jostling (R2 and R4: $r = .632$) and to cleanliness (R2 and R3: $r = .500$), with intervals that included 000. For Tension, correlations were positive or null, and only the correlation between worry about arriving late and fatigue had an interval that excluded 000 (R7 and R10: $r = .914$, 95% CI $[.394, .991]$, $p = .011$); the feeling of insecurity was not related to fatigue (R8 and R10: $r = .000$).

Internal structure and relationships among dimensions

The planned exploratory factor analysis could not be estimated in an interpretable manner, and its results are not presented in tables. With 6 cases for 10 items, the correlation matrix is not invertible; therefore, the measure of sampling adequacy and the test of sphericity could not be calculated. Parallel analysis suggested 2 factors, which in the software solution explained 65.35% of the variance, but with fewer cases than items, this solution is not supported. These results are included in the reproducibility appendix but do not provide evidence on the dimensionality of the instrument.

Table 6. Correlations among items for each dimension.

Dimension	Variable pair	r	IC 95%	p
Conditions	R1 and R2	-.171	[-.863, .744]	.745
Conditions	R1 and R3	-.857	[-.984, -.151]	.029
Conditions	R1 and R4	-.108	[-.846, .771]	.838
Conditions	R1 and R5	-.211	[-.873, .725]	.689
Conditions	R2 and R3	.500	[-.524, .933]	.313
Conditions	R2 and R4	.632	[-.368, .954]	.178
Conditions	R2 and R5	-.877	[-.986, -.227]	.022
Conditions	R3 and R4	.316	[-.666, .897]	.541
Conditions	R3 and R5	-.175	[-.864, .742]	.740
Conditions	R4 and R5	-.388	[-.912, .618]	.447
Tension	R6 and R7	.608	[-.402, .951]	.201
Tension	R6 and R8	.714	[-.231, .966]	.111
Tension	R6 and R9	.694	[-.269, .963]	.126
Tension	R6 and R10	.586	[-.431, .947]	.222
Tension	R7 and R8	.122	[-.766, .849]	.819
Tension	R7 and R9	.525	[-.499, .937]	.285
Tension	R7 and R10	.914	[.394, .991]	.011
Voltage	R8 and R9	.694	[-.269, .963]	.126
Tension	R8 and R10	.000	[-.812, .812]	1.000
Tension	R9 and R10	.553	[-.469, .942]	.255

Note. N = 6. Conditions = perceived transport conditions; Stress = stress during transfer. Pearson correlation. Confidence intervals were obtained using Fisher's z transformation. Cross-sectional design: the reported relationships are associations. Runs RUN-00147 and RUN-00149 from Factoriza UAS (Tables 5 and 7 of the package, combined without modifying values). Preliminary cutoff: n = 6 out of a target of 100, with data collection ongoing.

The Spearman correlation between mean scores for Conditions and Tension, based on 6 cases, was negative and moderate, $\rho = -.525$, 95% CI [-.946, .554], $p = .285$. The sign is as expected, since better perceived conditions would be accompanied by lower tension, but the interval, calculated using the standard error of Bonett and Wright (2000), is so wide that the result does not support asserting the existence of this relationship. Because the analysis comprises a single pair of dimensions, the result is presented in the text and not in a table.

Discussion of the results

The purpose of this study was to conduct a preliminary examination of the functioning of a user experience scale for urban public transport in Mazatlán. With 6 cases, the coefficients cannot be interpreted as estimates of the instrument's properties: the simulation literature on alpha and omega was developed with much larger samples (Malkewitz et al., 2023), and factor analysis requires a sample size that this subset does not reach (Ferrando et al., 2022; Goretzko et al., 2021). The subset does, however, allow a clear contrast between the 2 dimensions to be identified.

Tension during travel showed high internal consistency, with items describing bad mood, lack of calm, worry about punctuality, and tiredness. This set aligns with the affective dimensions of travel documented in the literature, which associates unpredictability and crowding with passenger stress and exhaustion (Chatterjee et al., 2020; Liu et al., 2022). The finding that worry about being late had the highest mean and the highest correlation with tiredness is consistent with the importance of reliability and punctuality in the service experience (Abu Bakar et al., 2022; Sogbe et al., 2025). The feeling of insecurity was the least coherent item, suggesting that personal safety, which depends on social, environmental, and temporal factors (te Braak and van Tienoven, 2025), could constitute a separate facet of the travel experience.

Perceived transport conditions, by contrast, did not behave as a dimension. Its items describe distinct attributes of the service—frequency, on-board comfort, cleanliness, occupancy, and shade at the stop—that the literature treats as separate quality attributes (Abu Bakar et al., 2022; Sogbe et al., 2025) and that do not necessarily coincide on the same route: a bus may pass regularly and be dirty, or a stop may have shade and serve an overcrowded route. Moreover, with 6 cases, several

items had very low variability, which makes the correlations unstable. The negative alpha should therefore be interpreted as an indication that the dimension combines heterogeneous attributes, not as an estimate of its reliability.

This heterogeneity has implications for measurement. If service conditions are considered attributes that together describe the quality of the experience, rather than manifestations of a single disposition, their measurement need not meet the requirement of internal consistency; it is advisable to evaluate them as a profile of attributes or to group them into more homogeneous subdimensions, for example, on-board conditions, which combine comfort, cleanliness, and occupancy, and waiting conditions, which combine regularity and shade at the stop. The positive, albeit uncertain, correlation between comfort and the absence of jostling points in that direction is in line with studies that link perceived crowding with on-board comfort (Liu et al., 2022; Zhu y Yu, 2026).

Finally, shade at the stop showed the greatest dispersion in the instrument, which indicates that respondents differ in the waiting conditions they face. In a city with a warm climate such as Mazatlán, shade at the stop is a condition for comfort and health, as suggested by studies on heat and transport use (Lanza y Durand, 2021) and on the association between high temperatures and mood (Li et al., 2023). Its inclusion in the scale is relevant, although it may be advisable to measure it together with other attributes of the stop rather than as part of a general dimension of conditions.

Implications for item review

The revision of the instrument can be informed by the attributes that the literature consistently identifies. Reviews of bus service quality include punctuality, reliability, travel time, and safety among the most relevant attributes (Abu Bakar et al., 2022; Sogbe et al., 2025), whereas the current scale captures only the regularity of bus arrivals. A revised version could incorporate items on waiting time, punctuality relative to expectations, and the walking route between the stop and the trip origin or destination—the first and last kilometer indicated by Sogbe et al. (2025)—so that each subdimension of conditions has several items.

The wording of the items also merits attention. Two items in the tension dimension are worded in the masculine, “I feel unsafe when traveling by bus” and “When getting off the bus I feel tired,” which may seem foreign to the women surveyed; given that women’s perception of safety in urban public space depends on personal, social, environmental, and temporal factors, such as the difference between day and night (te Braak and van Tienoven, 2025), it is advisable to use one neutral wording and to record the time of the trip. Reverse-worded items, for their part, can introduce wording effects (Steinmann et al., 2022), and with a larger sample it will be advisable to examine whether R4 and R9 behave differently from the directly worded items in their dimension. Finally, the feeling of insecurity could be developed with several items that distinguish waiting at the stop, the journey on board, and the walking route.

These proposals do not derive from the cut-off coefficients, which lack the precision necessary to justify changes in specific items, but rather from the joint reading of the content of the scale, of the reviewed literature and of the patterns that the cut-off reveals. Their relevance should be examined with the complete sample and, if possible, with a review of the content by experts and by users of public transport in Mazatlán.

CONCLUSIONS

The study met its objective to the extent allowed by a preliminary sample of 6 cases: it does not provide estimates of the psychometric properties of the scale, but it identifies the adjustments required. Tension during transport showed high internal consistency ($\alpha = .855$; $\omega = .863$), whereas Perceived transport conditions obtained a negative alpha ($\alpha = -.3359$), indicating that it brings together heterogeneous attributes of the service. Surveyed individuals rated the comfort, cleanliness, and occupancy of the trucks unfavorably, reported moderate tension during transport, and identified punctuality as their main concern. The factor analysis was not interpretable, and the relationship between the dimensions, which was negative, was highly uncertain.

The study also has strengths. The analyzed dataset is frozen with a cryptographic fingerprint, the analyses were run using fixed and declared procedures, and the complete program output is compiled in a reproducibility appendix, so that any reader can repeat the calculations with the same data. The early review of the instrument, before closing the survey, allows its structure to be corrected while this is still possible without losing the fieldwork.

The limitations are substantial. The sample of 6 cases, compared with the goal of 100 and with data collection still open, is insufficient to estimate internal consistency accurately and does

not allow factor analysis. The respondents answered on routes and stops selected by the project manager, so the results cannot be generalized to transport users in Mazatlán. The study did not measure the characteristics of the respondents or of their trips, such as route, frequency of use, or trip duration. The cross-sectional design does not allow the direction of the relationships to be established. Confirmatory factor analysis was not performed.

These results give rise to decisions for the authors. The first is to reorganize the conditions dimension into more homogeneous subdimensions, such as on-board conditions and waiting conditions, each with several items, or to treat it as a service attribute profile. The second is to decide whether the feeling of insecurity is retained within the tension dimension or developed as a separate personal security facet. The third is to record the route, frequency of use, and travel duration of each surveyed person. None of these decisions was made in the present analysis, which retains the 10 items.

As future directions, it is proposed to complete the survey up to the planned target, with users from different routes and schedules; to repeat the internal consistency analysis and perform the exploratory factor analysis with the final data collection; to confirm the structure in one independent sample through confirmatory factor analysis; and to relate the travel experience to the housing and neighborhood habitability of the users, in line with the original project approach. A scale with sufficient validity evidence would make it possible to incorporate the public transport experience into the study of urban habitability in Mazatlán.

REFERENCES

1. Abu Bakar, M. F., Norhisham, S., Katman, H. Y., Fai, C. M., Azlan, N. N. I. M., y Samsudin, N. S. S. (2022). Service quality of bus performance in Asia: A systematic literature review and conceptual framework. *Sustainability*, 14(13), Artículo 7998. <https://doi.org/10.3390/su14137998>
2. Bonett, D. G., y Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika*, 65(1), 23–28. <https://doi.org/10.1007/BF02294183>
3. Chatterjee, K., Chng, S., Clark, B., Davis, A., De Vos, J., Ettema, D., Handy, S., Martin, A., y Reardon, L. (2020). Commuting and wellbeing: A critical overview of the literature with implications for policy and future research. *Transport Reviews*, 40(1), 5–34. <https://doi.org/10.1080/01441647.2019.1649317>
4. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–334. <https://doi.org/10.1007/BF02310555>
5. Ferrando, P. J., Lorenzo-Seva, U., Hernández-Dorado, A., y Muñoz, J. (2022). Decalogue for the factor analysis of test items. *Psicothema*, 34(1), 7–17. <https://doi.org/10.7334/psicothema2021.456>
6. García-Batista, Z. E., Guerra-Peña, K., Garrido, L. E., Cantisano-Guzmán, L. M., Moretti, L., Cano-Vindel, A., Arias, V. B., y Medrano, L. A. (2021). Using constrained factor mixture analysis to validate mixed-worded psychological scales: The case of the Rosenberg Self-Esteem Scale in the Dominican Republic. *Frontiers in Psychology*, 12, Artículo 636693. <https://doi.org/10.3389/fpsyg.2021.636693>
7. Goretzko, D., Pham, T. T. H., y Bühner, M. (2021). Exploratory factor analysis: Current use, methodological developments and recommendations for good practice. *Current Psychology*, 40(7), 3510–3521. <https://doi.org/10.1007/s12144-019-00300-2>
8. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernández del Río, J., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
9. Lanza, K., y Durand, C. P. (2021). Heat-moderating effects of bus stop shelters and tree shade on public transport ridership. *International Journal of Environmental Research and Public Health*, 18(2), Artículo 463. <https://doi.org/10.3390/ijerph18020463>
10. Li, D., Zhang, Y., Li, X., Zhang, K., Lu, Y., y Brown, R. D. (2023). Climatic and meteorological exposure and mental and behavioral health: A systematic review and meta-analysis. *Science of the Total Environment*, 892, Artículo 164435. <https://doi.org/10.1016/j.scitotenv.2023.164435>
11. Liu, J., Ettema, D., y Helbich, M. (2022). Systematic review of the association between commuting, subjective wellbeing and mental health. *Travel Behaviour and Society*, 28, 59–74. <https://doi.org/10.1016/j.tbs.2022.02.006>
12. Malkewitz, C. P., Schwall, P., Meesters, C., y Hardt, J. (2023). Estimating reliability: A comparison of Cronbach's α , McDonald's ω and the greatest lower bound. *Social Sciences & Humanities Open*, 7(1), Artículo 100368. <https://doi.org/10.1016/j.ssaho.2022.100368>
13. McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press. <https://doi.org/10.4324/9781410601087> (Obra original publicada en 1999)
14. Python Software Foundation. (2026). Python (Versión 3.13.16) [Software]. <https://www.python.org/>
15. Saleem, S., y Alchalabi, O. (2025). An in-depth analysis of factors impacting housing satisfaction: A systematic review and post-occupancy evaluation. *International Journal of Sustainable Development and Planning*, 20(1), 119–131. <https://doi.org/10.18280/ijstdp.200113>
16. Sogbe, E., Susilawati, S., y Tan, C. P. (2025). Scaling up public transport usage: A systematic literature review of service quality, satisfaction and attitude towards bus transport systems in developing countries. *Public Transport*, 17(1), 1–44. <https://doi.org/10.1007/s12469-024-00367-6>
17. Steinmann, I., Strietholt, R., y Braeken, J. (2022). A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales. *Psychological Methods*, 27(4), 667–702. <https://doi.org/10.1037/met0000392>
18. te Braak, P., y van Tienoven, T. P. (2025). Navigating the city: A systematic literature review on women's perceived safety in urban public space. *Cities*, 162, Artículo 105907. <https://doi.org/10.1016/j.cities.2025.105907>
19. Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, 17(3), 261–272. <https://doi.org/10.1038/s41592-019-0686-2>
20. Zhu, W., y Yu, Y. (2026). Nature-inspired visual interventions in metro stations and commuters' affective well-being: The mediating roles of perceived crowding and environmental comfort. *Frontiers in Psychology*, 17, Artículo 1901802. <https://doi.org/10.3389/fpsyg.2026.1901802>

FUNDING

The authors received no funding for the development of this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTIONS

Conceptualization: Génesis Cristal Guzmán Oros
Data curation: Pedro Antonio Valdez Lizárraga
Formal analysis: Víctor Manuel Martínez García
Research: Génesis Cristal Guzmán Oros
Methodology: Yennifer Díaz Romero
Project administration: Pedro Antonio Valdez Lizárraga
Software: Pedro Antonio Valdez Lizárraga
Supervision: Víctor Manuel Martínez García
Validation: Víctor Manuel Martínez García
Visualization: Yennifer Díaz Romero
Writing – original draft: Génesis Cristal Guzmán Oros
Writing – review and editing: Yennifer Díaz Romero