

OPEN ACCESS

EDITED BY

Fernanda Loureiro,
Catholic University of Portugal, Portugal

REVIEWED BY

Mariana Moreira,
Polytechnical Institute of Coimbra,
Portugal
Cilmara Levy,
Centro Especializado Paulista, Brazil

*CORRESPONDENCE

Pietro Alfano
✉ pietro.alfano@cnr.it

[†]These authors have contributed equally to this work and share first authorship

[†]These authors have contributed equally to this work and share last authorship

RECEIVED 24 July 2026

REVISED 28 August 2026

ACCEPTED 03 September 2026

PUBLISHED 21 September 2026

CITATION

Pandolfo A, Perri A, Carlino G, Alfano P, Torta S, Montalbano AM, Albano GD, Bonomolo A, Muggeo VM, La Grutta S and Malizia V (2026) Psychometric properties of the Italian Place Standard Tool for Children: validation in primary school children within the PNC Clima national programme. *Front. Public Health* 14:1947358. doi: 10.3389/fpubh.2026.1947358

COPYRIGHT

© 2026 Pandolfo, Perri, Carlino, Alfano, Torta, Montalbano, Albano, Bonomolo, Muggeo, La Grutta and Malizia. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Psychometric properties of the Italian Place Standard Tool for Children: validation in primary school children within the PNC Clima national programme

Alessandra Pandolfo^{1,2†}, Andrea Perri^{1,3†}, Giustino Carlino^{1,4}, Pietro Alfano^{1*}, Silvia Torta², Angela Marina Montalbano¹, Giusy Daniela Albano¹, Anna Bonomolo¹, Vito Michele Rosario Muggeo², Stefania La Grutta^{1†} and Velia Malizia^{1†}

¹Institute of Translational Pharmacology (IFT), National Research Council (CNR), Palermo, Italy,

²Department of Economics, Business and Statistics (dSEAS), University of Palermo (UNIPA), Palermo, Italy, ³Department of Earth and Marine Sciences (DiStEM), University of Palermo (UNIPA), Palermo, Italy, ⁴Department of Culture and Society, University of Palermo (UNIPA), Palermo, Italy

Background: The Place Standard Tool Children's Version (PST-C) lets children evaluate the quality of places where they live. Developed in Scotland and adapted in over a dozen countries, it had no validated Italian version. This study evaluated the psychometric properties of the Italian PST-C within a case study of the national research programme PNC E.1 "Salute, Ambiente, Biodiversità e Clima" (PNC Clima).

Methods: A sample of 131 children attending the last 2 years of primary school and the first 2 years of middle school (69 boys, 62 girls, aged 9 to 12 years) completed the 14-item instrument on a three-point ordinal scale. Psychometric evaluation included floor and ceiling effects, internal consistency (Cronbach's α and McDonald's ω), structural validity from polychoric correlations, item functioning under Samejima's graded response model, known-groups validity. Translation equivalence was evaluated using both conventional expert review procedures and multilingual sentence embeddings.

Results: Internal consistency was good: $\alpha = 0.83$ (95% CI [0.77, 0.87]) and $\omega = 0.86$ (95% CI [0.82, 0.89]). Parallel analysis and confirmatory factor analysis (CFI = 0.904, RMSEA = 0.058) supported a unidimensional structure, with all polychoric loadings ≥ 0.42 . Semantic equivalence between versions was high (mean cosine similarity = 0.86). Two findings warrant caution: all items showed ceiling effects above the 15% threshold, and no known-group differences survived correction for multiple testing.

Conclusion: Combining COSMIN-based validation, item response theory and AI-based semantic equivalence, the Italian PST-C provides a reliable measure of perceived place quality in children. Future studies should evaluate an expanded response scale and confirm these findings in larger, more diverse populations.

KEYWORDS

children, item response theory, Place Standard Tool, polychoric factor analysis, semantic equivalence, validation

Introduction

The quality of the everyday environments in which children grow up shapes their health, autonomy and wellbeing (1–3). Children develop meaningful relationship with the places and natural environment in which they live, and these connections support coping and development (4). Article 12 of the United Nations Convention on the Rights of the Child also establishes children's right to express views on matters that affect them and to have those views taken seriously. Initiatives such as the Child Friendly Cities programme translate this principle into local planning practice (5).

What is known about the environmental features that matter to children comes from public health, environmental psychology and urban planning? On the physical side, green and natural space, safe and attractive places to play, walkable and cyclable streets, low exposure to motorised traffic and the upkeep of public space have all been associated with children's physical activity, independent mobility, cognitive and socio-emotional development and mental wellbeing (1, 2). The social side matters as much. Perceived safety, neighbourhood cohesion, opportunities for informal contact with peers and adults, and the sense of being listened to determine whether children actually use the spaces available to them, and they are among the strongest correlates of independent mobility and outdoor play (3, 6). The two interact: a well-equipped playground that children do not perceive as safe, or a green area that adults do not allow them to reach unaccompanied, contributes little to their daily experience. Most of this evidence, however, rests on objective or adult-reported measures. Children's own appraisal of their neighbourhood is documented far less systematically (2, 7).

Direct measurement of children's perceptions has usually relied on youth adaptations of adult questionnaires, such as the Neighborhood Environment Walkability Scale for Youth (8), or on item sets built for a single study and reported without formal psychometric evaluation. Participatory and mapping-based approaches, including softGIS methods in which children locate and rate meaningful places in their neighbourhood, have produced rich context-specific evidence (6, 9). They are also demanding to administer at scale, and the data they yield are hard to compare across settings. Two gaps run through this literature. Instruments are rarely validated with children of primary-school age, and they rarely cover physical and social qualities of place within one framework. A third problem is specific to adaptation: when an instrument is translated, semantic equivalence between source and target versions is usually assumed rather than demonstrated. Urban planners and public health professionals therefore require instruments that let children evaluate the places where they live in a structured, age-appropriate and psychometrically defensible way (5, 7, 8).

The Place Standard Tool (PST), developed in Scotland by the Scottish Government, NHS Health Scotland, and Architecture and

Design Scotland (10), is one of the most widely adopted frameworks for the participatory assessment of place quality. It organizes the physical and social dimensions of a place into fourteen domains, ranging from movement and transport to safety, identity and influence, and asks respondents to rate each one. The tool has been adopted well beyond Scotland. Published applications include a national internet survey in Cyprus (11); a Spanish adaptation whose psychometric properties were evaluated in a municipality in Valencia and subsequently applied to compare urban and rural settings (12, 13); an implementation in North Macedonia, where the Macedonian version demonstrated good internal consistency (Cronbach's α of 0.89) (14); and an analysis of more than 8,000 responses from northern England examining the tool's capacity to capture place-related inequalities (15). A version specifically designed for children of approximately 6 to 12 years, featuring simplified wording, illustrated prompts and a reduced three-point response format, was released in 2022 (16); a distinct version for young people aged 12 and over, which retains a seven-point response scale, was released alongside it (16). Despite its widespread adoption, the Place Standard Tool has primarily been used as a participatory community engagement tool rather than as a psychometrically validated measurement instrument, and evidence regarding its measurement properties remains limited (11, 15). For the Children's Version (PST-C) such evidence is particularly scarce, and no validated Italian version has yet been published.

The present validation was conducted within an objective case study of PNC Clima, the programme E.1 "Salute, Ambiente, Biodiversità e Clima" of the Italian National Plan for Complementary Investments to the National Recovery and Resilience Plan. The programme supports applied research on the relationships between climate, environment and health within a One Health perspective. One of its case studies focuses on how children perceive and evaluate their everyday urban environments in the context of climate change. The school-based administration of the PST-C reported here forms part of that case study, and the availability of a validated Italian version is a prerequisite for the programme's subsequent monitoring activities.

Validating the PST-C in Italian raises two methodological problems. The first concerns the response format: items are ordinal and include only three response categories. Under these conditions, Pearson correlation-based factor analysis tends to underestimate inter-item associations; therefore, polychoric correlations provide a more appropriate basis for structural analyses (17, 18). The second challenge relates to translation quality, which is typically assessed through qualitative procedures involving independent forward translation and expert reconciliation (19). Recent advances in multilingual sentence embeddings (20, 21) offer a quantitative complement to these procedures, allowing semantic similarity between the source and target versions of each item to be computed directly in a shared vector space.

Following the COSMIN methodological standards for evaluating measurement properties (22, 23), we examined: (a) item functioning, including floor and ceiling effects; (b) internal consistency; (c) structural validity, through parallel analysis, exploratory and confirmatory factor analysis on polychoric correlations, and Samejima's graded response model (24); (d) known-groups validity across sex, school and grade; and (e) the semantic equivalence of the Italian translation.

Abbreviations: PST-C, Place standard tool-children's version; COSMIN, Consensus-based Standards for the selection of health measurement instruments; IRT, Item response theory; GRM, Graded response model; EFA, Exploratory factor analysis; CFA, Confirmatory factor analysis; CFI, Comparative fit index; RMSEA, Root mean square error of approximation; CI, Confidence interval; SD, Standard deviation.

Materials and methods

Study context, participants and procedure

Data were collected in the schools participating in the PNC Clima case study on children's perception of urban environments. Participants were 131 primary school children (69 boys, 52.7%; 62 girls, 47.3%) from four school buildings belonging to four different state comprehensive institutes (istituti comprensivi statali), attending grades 4 and 5 of Italian Primary School and grades 1 and 2 of Italian Middle School, age range of 9 to 12 years. The questionnaire was administered in paper-based format during regular school hours. No private or state-subsidised (paritarie) schools were involved. The four institutes are all located in the municipality of Palermo, Sicily, and serve different districts of the same urban area. Participation was voluntary. Written informed consent was obtained from the parents or legal guardians of every child before administration. A member of the research team then introduced the study to each class in age-appropriate language, explaining what it was about, what the children would be asked to do and how long it would take. The children were told that there were no right or wrong answers, that their answers would be used only for research and reported only in aggregate form, and that their teachers would not see individual responses. They were invited to ask questions, and verbal assent was obtained individually from each child before the questionnaire was distributed. We also made clear that they could decline to take part, leave any question unanswered, or stop at any moment, without giving a reason and without consequence. Some did stop before reaching the end, which accounts for part of the item-level missing data reported below. These procedures follow the principles of the International Charter for Ethical Research Involving Children (25). The study was approved by the Ethics Committee of *Comitato Etico Locale Palermo 1* (Approval No. 25, date: 30/10/2024). Item-level missing data ranged from 1.5 to 9.2%. The listwise-complete dataset included 98 children for analyses requiring complete responses on all 14 items and 121 children for the confirmatory factor analysis.

Instrument

The instrument used here is the Children's Version of the Place Standard Tool (PST-C), released in 2022 for children of roughly 6 to 12 years (16). It is not the same as the Young People's Version, which is intended for respondents aged 12 and over and keeps the seven-point response scale of the adult tool (16). The version translated and validated in this study is the Children's Version, with its three-point scale. The PST-C asks children to evaluate fourteen domains of their living environment: streets and active movement, natural spaces, places for play, walking and cycling, public transport, traffic and parking, feeling heard, care and cleanliness of public space, perceived safety, identity and pride, services and amenities, work and the local economy, housing and neighbours, and social contact. The questionnaire contains one scored question per domain, so there are fourteen closed items in total, no sub-items, and a total score ranging from 14 to 42. Each domain is rated on a three-point scale. The verbal anchors in the source tool are 1 = "not good, lots of improvement needed," 2 = "okay, some improvement needed" and 3 = "great, it's good or as good as it can be." Each domain also carries two open-ended prompts, for twenty-eight open questions overall. The first is What is good now? The second appears

in the source tool as What would you use your magic wand on? and was rendered in the Italian version as Cosa si potrebbe migliorare? (What could be improved?), a wording judged more transparent for classroom administration during the reconciliation step. Illustrated examples and supplementary probing questions accompany each domain; they support reflection and are not scored. Table 1 gives, for each of the fourteen domains, the domain heading and the scored question as worded in the English source tool. For access to the full original scale and the accompanying facilitator guidance, please refer to reference (16). Translation and cross-cultural adaptation

The validation followed established procedures for the translation and cross-cultural adaptation of measurement instruments (19). Two independent bilingual experts performed forward translations of the English source items, adapting wording and examples to the Italian primary-school context. A third independent expert compared the two translated versions, resolved discrepancies and verified that each reconciled item remained conceptually and semantically equivalent to the original while remaining understandable to young children. The reconciled version was subsequently pilot-tested for comprehensibility with a group of 10 children of the target age range before administration. No expert committee was convened beyond the three bilingual experts described above.

To complement the qualitative cross-cultural adaptation process, we conducted an objective quantitative assessment of semantic equivalence. This last analysis was carried out by the study authors, and not by the bilingual experts, whose involvement was limited to the translation and reconciliation steps. The English and Italian wording of each item was encoded using the multilingual sentence-embedding model *paraphrase-multilingual-MiniLM-L12-v2* (20, 21), which maps sentences from more than 50 languages into a shared 384-dimensional vector space. For each item we computed the cosine similarity between the English source and its Italian translation. As a benchmark, we also calculated the mean similarity between each Italian item and the English source items of the other thirteen domains. A faithful translation would be expected to show substantially greater similarity to its own source than to conceptually unrelated items.

Statistical analysis

Analyses were conducted in Python 3.11 using the packages *factor_analyzer* (0.5.1 or later), *pingouin*, *semopy*, *girth* and *statsmodels*. Given the three-category ordinal response format, analyses were based primarily on polychoric correlation matrix (estimated by pairwise maximum likelihood) wherever appropriate instead of Pearson correlations, which tend to underestimate associations among coarsely categorized ordinal variables (17, 18). The analytic flow followed the COSMIN standards for evaluating measurement properties (22, 23). Descriptive statistics included floor and ceiling effects using the conventional 15% threshold. Internal consistency was estimated with Cronbach's α and McDonald's total ω ; ω is generally preferred when items are congeneric rather than tau-equivalent (26, 27). Ninety-five percent confidence intervals were estimated using non-parametric bootstrap with 1,000 resamples (28). Factorability was assessed using the Kaiser-Meyer-Olkin index and Bartlett's test of sphericity (29). The number of factors was determined primarily through parallel analysis based on 500 randomly generated Gaussian datasets, one of the most accurate factor-retention methods under realistic

TABLE 1 The fourteen domains of the children's version of the place standard tool, with the domain heading and the scored question as worded in the English source tool (16).

Item label	PST-C domain (source heading)	Scored question (English source wording)
Streets and movement	Streets, squares and buildings	What are the streets, squares and buildings like in my place?
Natural spaces	Nature—parks, trees, animals, plants	How much nature is there around my place and is it easy to get to?
Places for play	Places to play	What spaces are there for me to play and have fun in and around my place?
Walking and cycling	Walking, wheeling and cycling	How easy is it to walk, scoot, wheel or cycle to places you need and want to go?
Public transport	Buses, trains, trams, boats	Is it easy for me to get where I need to go to by bus, train (or other public transport)?
Traffic and parking	Traffic and parking	How does traffic in my place affect me when I'm out and about?
Feeling heard	Having our say and being listened to	When things happen in my place do people include me in decisions and listen to what I say?
Care and cleanliness	Fixed, clean and looking nice	Is my place clean and are things well cared for?
Feeling safe	Feeling safe	Do I feel safe in my place?
Identity and pride	Feeling proud and part of my place	Do I feel proud and a part of my place?
Services and amenities	Schools, libraries, shops and other services	What are the schools, libraries, shops and places to get what I need like?
Work and economy	Jobs and places to work	Are there places for people to work and get a good job in my place?
Housing and neighbours	Homes, friends and neighbours	How well do the homes in my place meet my needs and my family's needs?
Social contact	Meeting and talking with people	How nice are the nearby places where I meet with other children and trusted adults?

Each domain contributes one rating on a three-point scale and two open-ended prompts. Short item labels used in the remaining tables and figures are given in the first column.

conditions (30). The eigenvalue-greater-than-one criterion was used only as a secondary reference, because of its tendency to over extract factors (31). Exploratory factor analysis (EFA) was performed using minimum residual extraction with oblique Promax rotation, and was fitted to both the polychoric and the Pearson matrices for comparison (29). A congeneric one-factor confirmatory factor analysis (CFA) model was estimated using *semopy* under maximum likelihood and evaluated using conventional goodness-of-fit indices, while acknowledging the known limitations of these indices when applied to ordered categorical data (32). Samejima's graded response model (GRM) was estimated using marginal maximum likelihood to obtain discrimination and threshold parameters (24, 33). Known-groups validity was examined across sex (Mann–Whitney U), school and grade level (Kruskal–Wallis) with Holm correction applied for multiple testing across the 14 questionnaire domains.

The reference values used to interpret each measurement property were set in advance and applied throughout (22, 23, 27, 29, 32, 33). Floor and ceiling effects were treated as problematic above 15% of respondents in an extreme category. Internal consistency was read as adequate for group-level use from α and $\omega \geq 0.70$ and as good from ≥ 0.80 . Corrected item-total correlations had to exceed 0.30, and values between 0.30 and 0.40 were flagged as weak. Sampling adequacy was acceptable at $KMO \geq 0.60$ and good at $KMO \geq 0.80$, and factor loadings were treated as salient from 0.40. Model fit was mixed but broadly supportive of the one-factor solution: CFI met the prespecified criterion and RMSEA was acceptable, whereas TLI fell slightly below the 0.90 threshold. Graded response model discrimination parameters were read as low below 1.0, moderate between 1.0 and 1.7, and high above 1.7. For the semantic equivalence analysis, cosine similarities of 0.80 or above between matched pairs were taken to indicate close correspondence between source and translation. Statistical significance was set at a two-sided α of 0.05, with Holm correction across the 14 domain-level comparisons; because the known-groups analysis was exploratory, the direction and size of the differences were considered alongside the corrected p values.

Results

Sample and item descriptive statistics

Table 2 summarizes the sample and Table 3 presents the item-level descriptive statistics. Item means ranged from 2.05 (*Traffic and parking*) to 2.36 (*Housing and neighbours*). The median response was 2 for all item, and skewness was mildly negative throughout (between -0.48 and -0.05), indicating a general tendency for children to evaluate their living environment positively. All 14 items exceeded the conventional 15% ceiling-effect threshold, with ceiling rates ranging from 24.0 to 44.0%. Four items showed ceiling effects exceeding 40%: *Identity and pride* (44.0%) *Housing and neighbours* (43.1%) *Services and amenities* (42.9%) *Natural spaces* (40.6%). Floor effects were less pronounced (7.3 to 24.8%) and exceeded the threshold for five items (Walking and cycling, Public transport, Traffic and parking, Feeling heard, Care and cleanliness), with the highest value observed for Traffic and parking (24.8%). Figure 1 illustrates both floor and ceiling effects.

TABLE 2 Characteristics of the validation sample ($N = 131$ primary school children): sex, school grade and school of recruitment, with absolute frequencies and percentages.

Variable	<i>n</i>	%
Sex: female	62	47.3
Sex: male	69	52.7
Grade 4—elementary school	20	15.3
Grade 5—elementary school	19	14.5
Grade 1—middle school	54	41.2
Grade 2—middle school	38	29.0
School 1	39	29.8
School 2	20	15.3
School 3	55	42.0
School 4	17	13.0

No third-grade classes were sampled.

TABLE 3 Descriptive statistics for the 14 PST-C items: number of valid responses (N), mean (M) and standard deviation (SD) on the 1–3 response scale, and percentage of children at the lowest (floor) and highest (ceiling) response category.

Item	Domain	N	M	SD	Floor %	Ceiling %
Streets and movement	Streets and active movement	129	2.10	0.61	14.0	24.0
Natural spaces	Natural spaces	128	2.26	0.70	14.8	40.6
Places for play	Places for play	125	2.22	0.65	12.8	34.4
Walking and cycling	Walking and cycling	129	2.22	0.71	16.3	38.0
Public transport	Public transport	126	2.14	0.72	19.8	34.1
Traffic and parking	Traffic and parking	125	2.05	0.74	24.8	29.6
Feeling heard	Feeling heard	124	2.15	0.73	20.2	34.7
Care and cleanliness	Care and cleanliness	119	2.06	0.69	21.0	26.9
Feeling safe	Perceived safety	121	2.23	0.69	14.9	38.0
Identity and pride	Identity and pride	125	2.34	0.65	9.6	44.0
Services and amenities	Services and amenities	126	2.33	0.66	10.3	42.9
Work and economy	Work and local economy	122	2.26	0.59	7.4	33.6
Housing and neighbours	Housing and neighbours	123	2.36	0.62	7.3	43.1
Social contact	Social contact and meeting	125	2.24	0.66	12.8	36.8

Item labels are the Italian short names; the corresponding domain is given in English.

Internal consistency

Internal consistency was good: Cronbach's $\alpha = 0.829$ (95% bootstrap CI [0.769, 0.870]) and McDonald's $\omega = 0.863$ (95% bootstrap CI [0.824, 0.893]), both exceeding the 0.80 thresholds commonly recommended for applied use (27, 34). Corrected item-total correlations ranged from 0.33 to 0.59 (Table 4), all above the conventional minimum criterion of 0.30. Four items fell below the more stringent 0.40 criterion: *Feeling heard* (0.33), *Identity and pride* (0.34), *Services and amenities* (0.35) and *Housing and neighbours* (0.39). Deleting any individual item would leave α essentially unchanged (0.81 to 0.83); therefore, no item removal was justified on reliability grounds. All corrected item-total correlations thus exceeded the conventional 0.30 minimum, and the four values below 0.40 identify the weaker items noted above.

Polychoric correlations and factorability

Polychoric correlations among the 14 items ranged from -0.01 to 0.58 (mean 0.33) and were consistently positive (Figure 2). This pattern already suggests the presence of a dominant common dimension. Sampling adequacy was good (KMO = 0.814), all item-level KMO values ≥ 0.71 (29) and Bartlett's test of sphericity was significant, $\chi^2 = 333.6$, $p < 0.001$, indicating the data were suitable for factor analysis.

Dimensionality

Parallel analysis and inspection of the scree plot supported a uni-dimensional solution (Figure 3). The first observed eigenvalue was 5.37 on the polychoric matrix (4.43 on the Pearson matrix), substantially exceeding the 95th percentile of eigenvalues obtained from random data (1.88). The second observed eigenvalue (1.46 and 1.44 respectively) already fell below the corresponding random-data threshold (1.64). Although the Kaiser criterion suggested retaining three factors, this method is known to overestimate dimensionality (31). Accordingly, greater weight was given of the results of the parallel analysis (30).

Exploratory factor analysis

The one-factor solution based on the polychoric matrix explained 33.8% of the variance. All factor loadings were positive and exceeded the conventional 0.40 benchmark (29), ranging from 0.42 (*Feeling heard*) to 0.70 (*Feeling safe*) (Table 5; Figure 4). Factor loadings estimated from the Pearson matrix were on average 0.067 lower in absolute value, as expected when analyzing ordinal items with only three responses categories (18).

Confirmatory factor analysis

The congeneric one-factor model estimated by maximum likelihood (listwise $N = 121$) yielded $\chi^2 (77) = 102.5$, $p = 0.025$; CFI = 0.904 ; TLI = 0.887 ; RMSEA = 0.058 . Overall model fit was acceptable, although only marginally so. Given the modest sample size and the three-category response format, these results were considered satisfactory for an initial validation study. Nevertheless, replication using a categorical estimator such as WLSMV is warranted (18, 32).

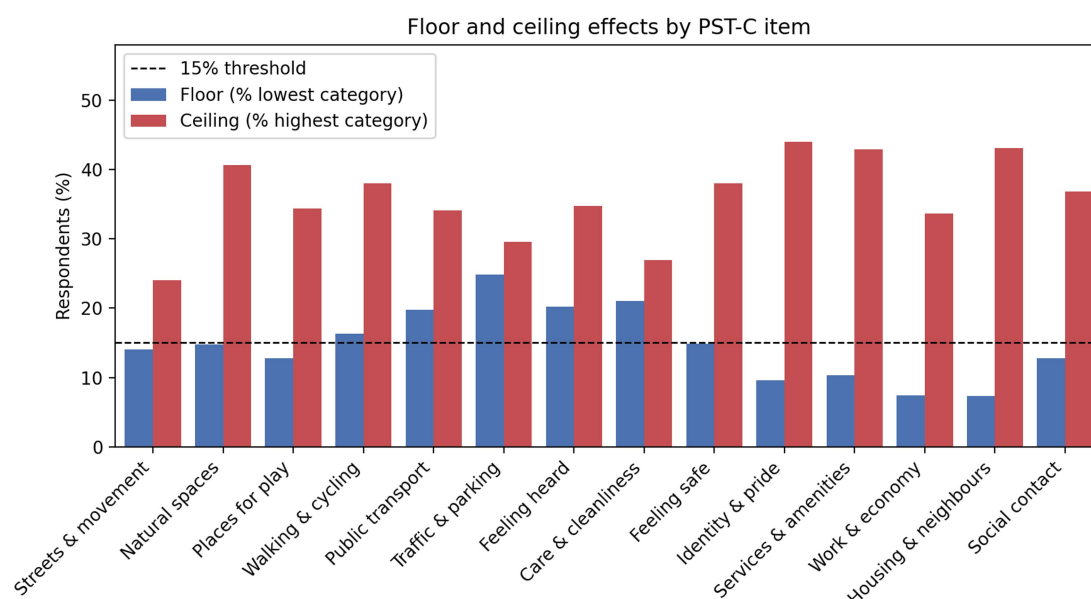


FIGURE 1

Floor and ceiling effects for the 14 PST-C items, shown as the percentage of children choosing the lowest (floor, blue bars) and highest (ceiling, red bars) response category. The dashed line marks the conventional 15% criterion.

TABLE 4 Item-level reliability indices ($N = 98$ listwise-complete cases): corrected item-total correlation, computed between each item and the sum of the remaining 13 items, and the value Cronbach's α would take if the item were deleted.

Item	r item-total (corrected)	α if item deleted
Streets and movement	0.50	0.82
Natural spaces	0.45	0.82
Places for play	0.54	0.81
Walking and cycling	0.55	0.81
Public transport	0.46	0.82
Traffic and parking	0.51	0.81
Feeling heard	0.33	0.83
Care and cleanliness	0.59	0.81
Feeling safe	0.57	0.81
Identity and pride	0.34	0.82
Services and amenities	0.35	0.83
Work and economy	0.41	0.82
Housing and neighbours	0.39	0.82
Social contact	0.46	0.82

Item response theory: graded response model

GRM discrimination parameters ranged from 0.77 to 1.75 (mean 1.26), spanning low to high discrimination, with most items in the

moderate range (33) (Table 6). First threshold parameters were strongly negative for all items (-2.94 to -1.07), suggesting that the lowest response category did not present comprehension or endorsement difficulties. Second threshold parameters were mildly positive (0.05 to 1.06), indicating that the transition from “average” to “good” occurred slightly above the trait mean. Total test information peaked at approximately 7.1 slightly above the trait mean, with a minimum standard error close to 0.37. This corresponds to a marginal reliability of roughly 0.86 in that region (Figure 5).

Known-groups validity

After Holm correction for multiple testing across the 14 comparisons, no statistically significant differences were observed by sex (Mann–Whitney U), school or grade level (Kruskal–Wallis). Possible interpretations of these null findings are considered in the Discussion.

Cross-language semantic equivalence

Cosine similarities between the English source items and their Italian translations averaged 0.86 (range 0.65 to 0.94; Table 7, Figure 6). The two open-ended prompts achieved similarity scores of 0.86 and 0.90, respectively. Matched item pairs were approximately two to three times more similar than the benchmark similarities obtained from non-matching items (0.31 to 0.38), indicating that each Italian item was semantically much closer to its corresponding English source item than to any other domain. The lowest value was observed for *Feeling heard* (0.65), whose English wording is the most idiomatic and conceptually abstract among the fourteen domains.

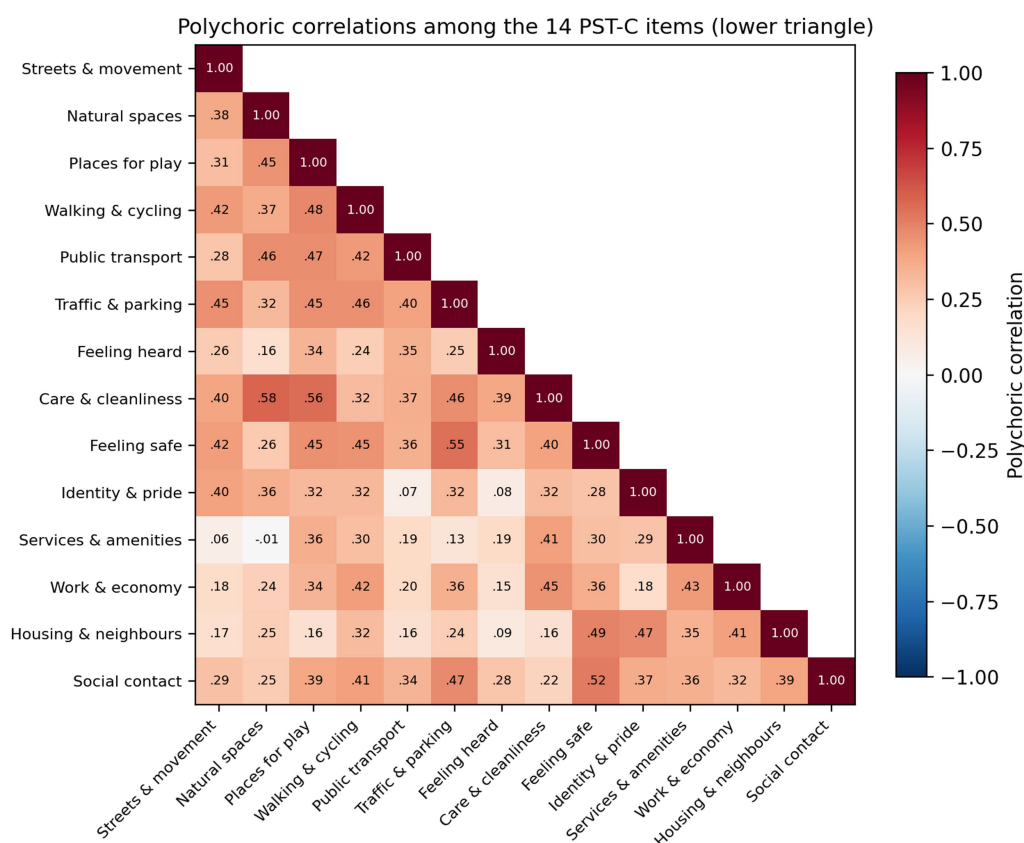


FIGURE 2

Polychoric correlation matrix among the 14 PST-C items (lower triangle shown). Correlations range from -0.01 to 0.58 with a mean of 0.33 and are almost uniformly positive.

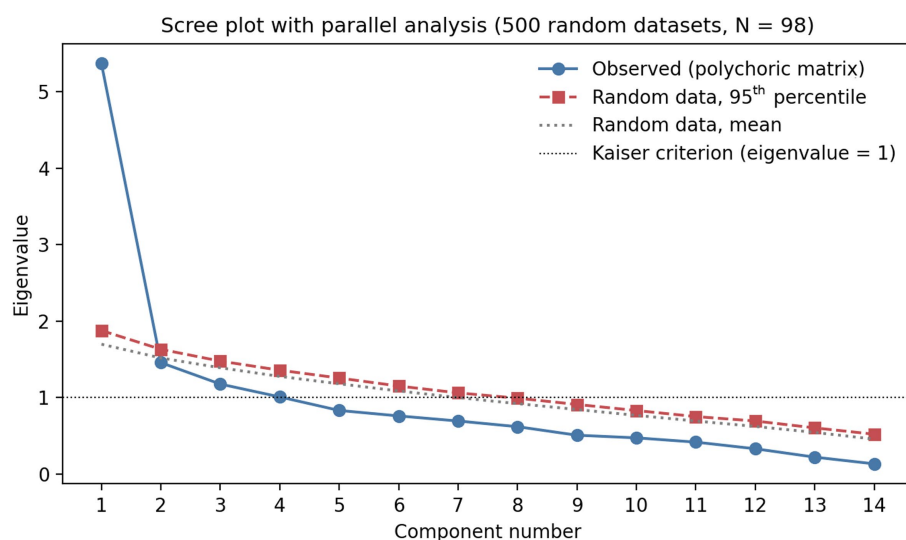


FIGURE 3

Scree plot with parallel analysis on the polychoric correlation matrix. Observed eigenvalues (blue circles) are compared with the mean (grey dotted line) and the 95th percentile (red squares) of eigenvalues from 500 random Gaussian datasets of the same size ($N = 98$, 14 variables). The dotted line at eigenvalue = 1 marks the Kaiser criterion.

Discussion

To our knowledge, this is one of the first validation studies of a child-centered environmental assessment instrument to integrate

COSMIN standards, item response theory (Samejima's graded response model), and AI-based semantic evaluation using multilingual sentence embeddings. This integrated framework extends conventional validation procedures by combining psychometric

evaluation with objective assessment of cross-language semantic equivalence. The present study was conducted within the PNC Clima programme, which seeks to investigate the relationships between environment, climate, and health from a One Health perspective. In this context, the availability of a validated Italian PST-C is particularly relevant because it enables the systematic inclusion of children's

perceptions within broader environmental-health monitoring activities.

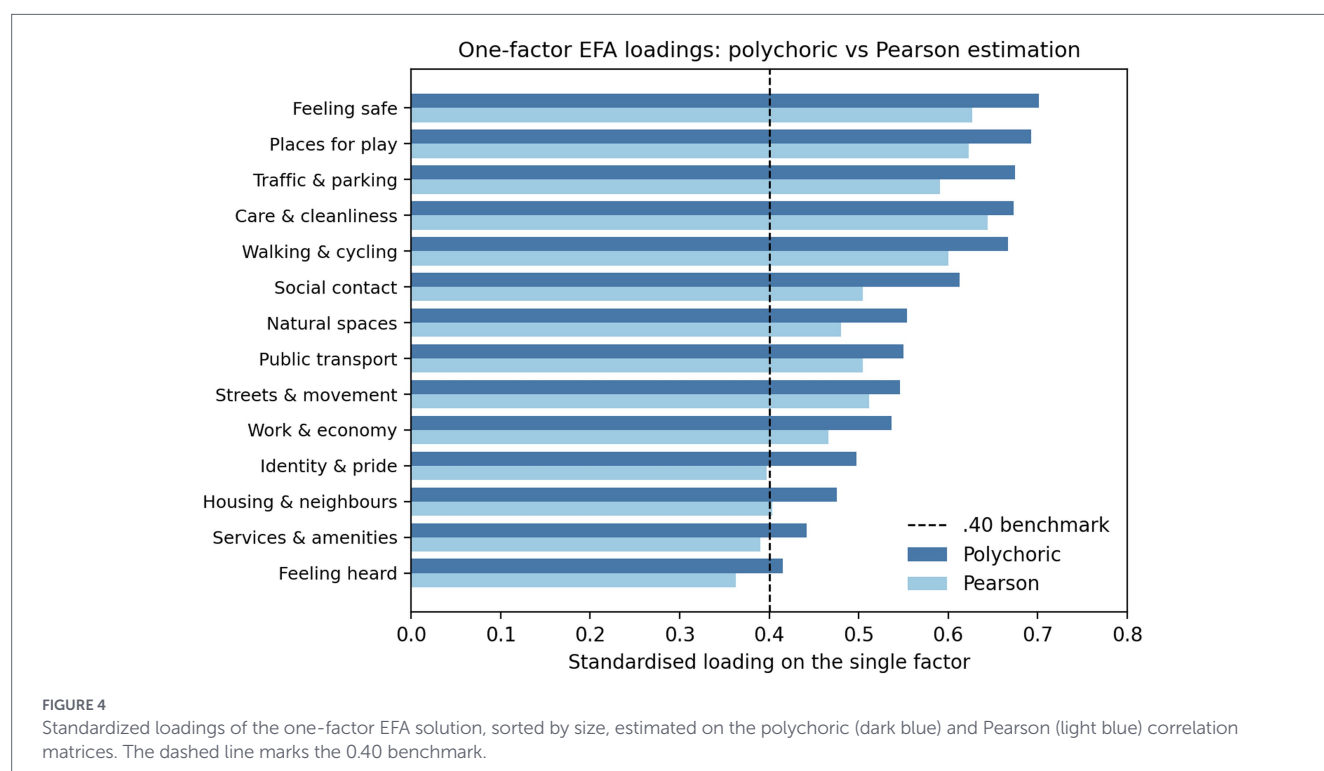
The instrument demonstrated an essentially unidimensional structure. Parallel analysis, scree plot inspection, exploratory factor analysis, and confirmatory factor analysis all supported the presence of a single general factor of perceived place quality underlying the fourteen theoretical domains. Moreover, all polychoric factor loadings were positive and substantively meaningful. Although the single factor explained approximately one-third of the total variance, the convergence of evidence across multiple psychometric approaches supports the interpretation of the PST-C as measuring a dominant latent construct. Internal consistency was good ($\alpha = 0.83$; $\omega = 0.86$), with bootstrap confidence intervals remaining above the conventional 0.70 threshold, indicating that the total scores are sufficiently reliable for applied participatory research and practice. The modest advantage of ω over α is expected, given that ω does not assume tau-equivalence among items (26, 27). The embedding-based analysis provides an additional source of evidence that cannot be obtained through qualitative translation procedures alone: the Italian items showed high semantic similarity to their English source versions, and the one domain displaying the greatest divergence (*Feeling heard*) was clearly identified as a potential target for future refinements. These reliability estimates are comparable to those reported for adult PST adaptations, including the Cronbach's α of 0.89 reported for the Macedonian version (14).

The item response theory findings further refine this interpretation. Measurement precision was highest slightly above the mean of the latent trait, where the instrument most effectively distinguishes children who rate their living environment as “average” from those who rate it as “good.” The consistently negative first thresholds suggest that the lowest response category contributes relatively little information across the trait continuum. Four

TABLE 5 Standardized loadings of the one-factor EFA solution (minimum residual extraction, promax rotation), estimated on the polychoric and on the Pearson correlation matrix.

Item	λ (polychoric)	λ (Pearson)
Streets and movement	0.55	0.51
Natural spaces	0.55	0.48
Places for play	0.69	0.62
Walking and cycling	0.67	0.60
Public transport	0.55	0.50
Traffic and parking	0.67	0.59
Feeling heard	0.42	0.36
Care and cleanliness	0.67	0.64
Feeling safe	0.70	0.63
Identity and pride	0.50	0.40
Services and amenities	0.44	0.39
Work and economy	0.54	0.47
Housing and neighbours	0.48	0.40
Social contact	0.61	0.50

Polychoric loadings are the reference values; Pearson loadings are reported as a robustness check.



items (*Feeling heard*, *Identity and pride*, *Services and amenities* and *Housing and neighbours*) emerged as comparatively weaker indicators, showing corrected item-total correlations below 0.40, GRM discrimination parameters below 1.0 and polychoric loadings below 0.50. Notably, these items represent the more affective, relational, or interpretive dimensions of place experience, including feeling heard, pride, access to services and neighborhood relations. By contrast, the strongest items were associated with more concrete environmental characteristics, such as opportunities for play, traffic conditions, orderliness and safety. Their removal

would not meaningfully improve reliability, and would reduce the conceptual breadth of the construct being measured.

Beyond its psychometric properties, the validated Italian PST-C has potential relevance for environmental health research and participatory urban governance. Contemporary approaches to child-friendly cities, climate adaptation and One Health increasingly recognize that children's subjective experiences of neighbourhood environments complement objective indicators such as green space availability, walkability or environmental exposures. The PST-C provides a structured mechanism through which children's perspectives can be systematically collected and incorporated into planning and evaluation processes. It complements the other ways in which children's experience of place can be documented rather than replacing them. What it adds is comparability: ratings that can be set side by side across children, classes and neighbourhoods. Used on its own, though, it cannot capture the relational, affective and contextual dimensions of how children inhabit the places they use every day, and we treat it as one component of a broader participatory repertoire rather than a stand-alone assessment. Within the PNC Clima programme, the instrument may support future efforts to integrate perceptions of place quality with environmental, climatic and health data, contributing to a more comprehensive understanding of how children experience and respond to their everyday environments.

TABLE 6 Item parameters of Samejima's graded response model, estimated by marginal maximum likelihood: discrimination (a) and the two category thresholds on the latent trait metric (b_1 separates category 1 from categories 2–3; b_2 separates categories 1–2 from category 3).

Item	a	b_1	b_2
Streets and movement	1.21	−1.94	0.87
Natural spaces	1.10	−2.16	0.33
Places for play	1.73	−1.73	0.44
Walking and cycling	1.59	−1.72	0.19
Public transport	1.18	−1.45	0.69
Traffic and parking	1.58	−1.07	0.62
Feeling heard	0.77	−1.81	1.06
Care and cleanliness	1.72	−1.07	0.76
Feeling safe	1.75	−1.57	0.32
Identity and pride	0.88	−2.94	0.05
Services and amenities	0.86	−2.59	0.22
Work and economy	1.14	−2.70	0.60
Housing and neighbours	0.89	−2.92	0.16
Social contact	1.21	−2.12	0.53

Limitations

Several limitations should be acknowledged. First, ceiling effects were observed for all items, with four items exceeding 40%. This likely reflects the restricted three-point response scale, the generally positive response tendencies of younger children and the relatively homogeneous environmental characteristics of the participating schools, all of which are located within the same municipality. Future studies should evaluate whether a five-point response scale with clearly defined verbal anchors improves score dispersion while preserving comprehensibility.

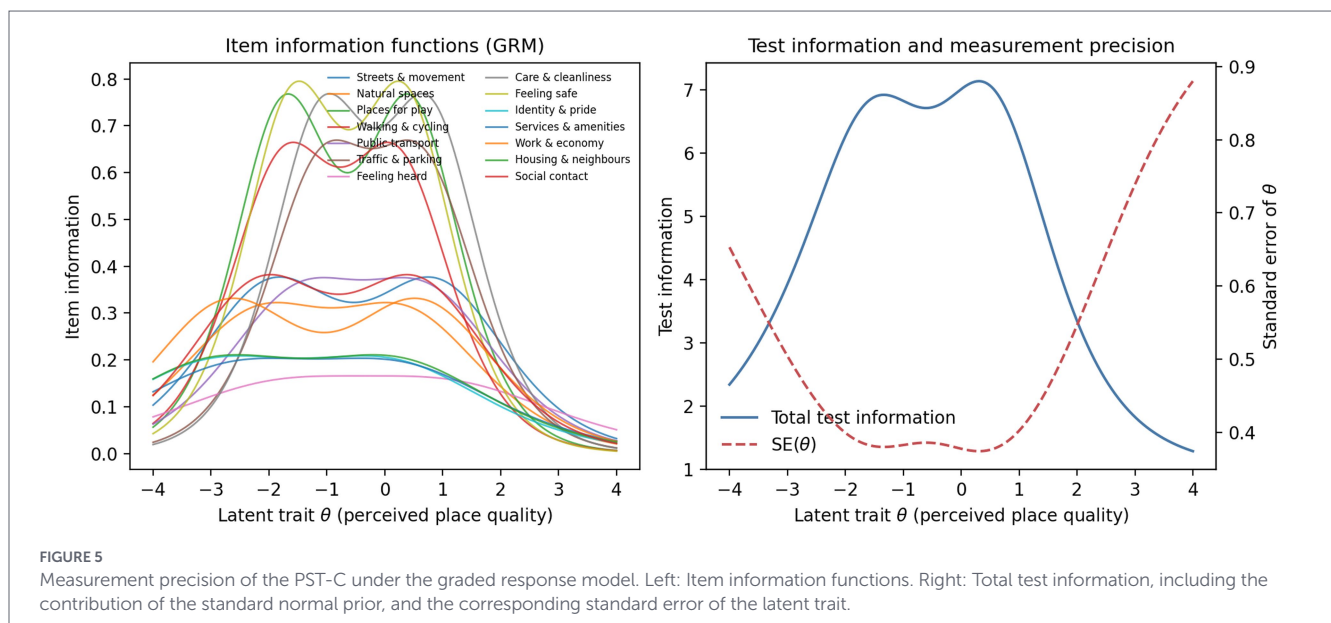
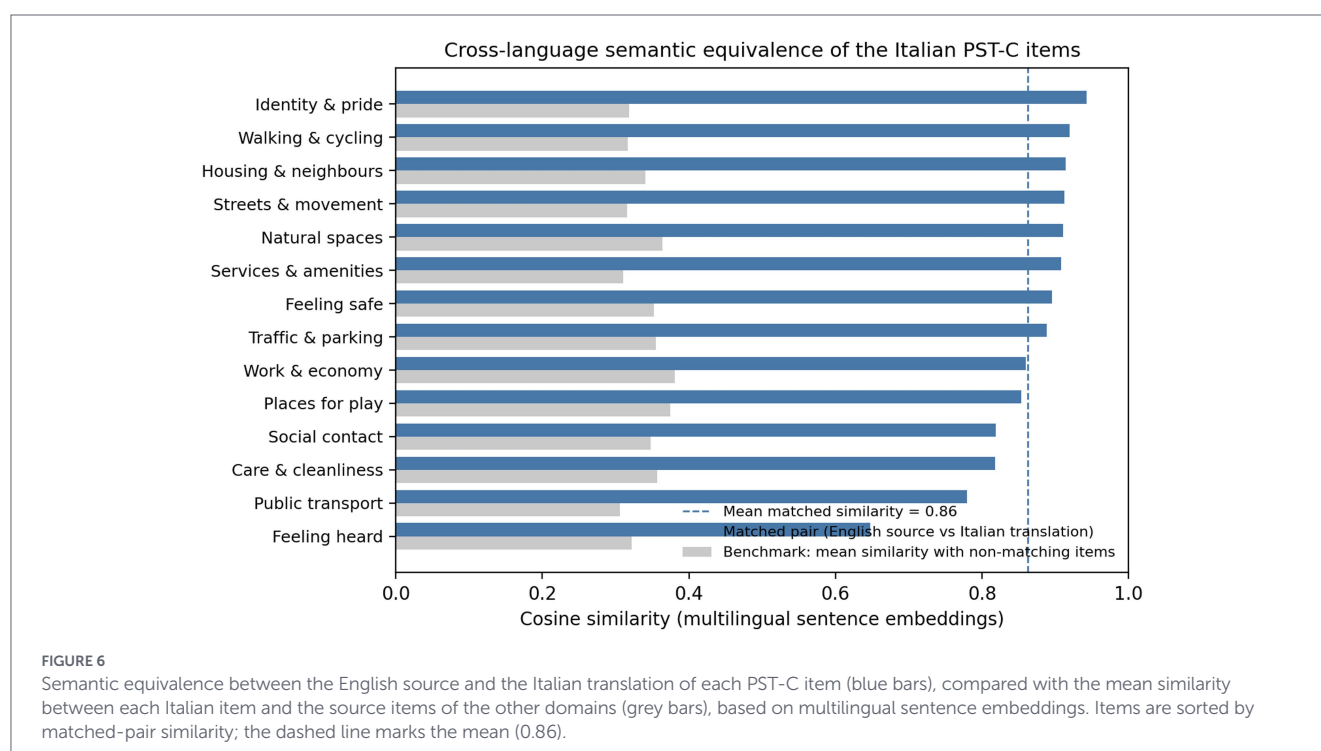


TABLE 7 Cross-language semantic equivalence of the Italian PST-C, computed with the paraphrase-multilingual-MiniLM-L12-v2 sentence-embedding model.

Item	Cosine similarity, matched pair	Mean similarity, non-matching items
Streets and movement	0.91	0.32
Natural spaces	0.91	0.36
Places for play	0.85	0.37
Walking and cycling	0.92	0.32
Public transport	0.78	0.31
Traffic and parking	0.89	0.36
Feeling heard	0.65	0.32
Care and cleanliness	0.82	0.36
Feeling safe	0.90	0.35
Identity and pride	0.94	0.32
Services and amenities	0.91	0.31
Work and economy	0.86	0.38
Housing and neighbours	0.91	0.34
Social contact	0.82	0.35
Open prompt: "What is good now?"	0.86	n/a
Open prompt: "What could be improved?"	0.90	n/a

For each item: Cosine similarity between the English source and its Italian translation (matched pair), and mean similarity between the Italian item and the English source items of the other 13 domains (benchmark). The last two rows refer to the open-ended prompts, for which no benchmark applies.



Second, known-groups validity was not supported after correction for multiple testing. Although this may reflect genuine similarities in perceived place quality across sex, schools, and grade levels, the modest sample size, limited environmental heterogeneity, and conservative multiplicity correction are more likely to have reduced statistical power. Larger and more diverse samples are therefore needed to further evaluate this measurement property.

Third, the confirmatory model was estimated using maximum likelihood despite the ordinal nature of the data. Although model fit was acceptable, future studies should replicate the analyses using weighted least squares with mean and variance adjustment (WLSMV), the preferred estimator for ordinal items (18, 32). Fourth, the sample came from a single urban context and is not nationally representative. The single-city sampling frame (four

state comprehensive institutes in Palermo) and the item-level missing data (up to 9.2%), part of which follows from children exercising their right to skip questions or to stop, limit how far the findings generalize. Multi-site studies covering northern, central and southern Italy, and both urban and rural settings, are needed before the Italian PST-C can be recommended for national use.

A last limitation is conceptual rather than statistical. The PST-C gives a structured, comparable and quickly administered picture of how children rate their environment, but a rating on a three-point scale is a compressed record of a complex experience. It can tell us that a child scores traffic poorly. It cannot tell us which street, at which time of day, or what it feels like to cross it. Qualitative and participatory methods reach that layer of meaning better: individual and group interviews, focus groups, and visual methods such as photo elicitation, child-led walking tours, drawing and mapping activities, all of which let children take a more active part in producing and interpreting knowledge about the places they use every day (35, 36). We see the PST-C and these approaches as complementary rather than alternative. Mixed-methods designs, in which PST-C scores identify the domains that matter most in a given place and qualitative work then asks why, seem to us the direction worth pursuing with the Italian version.

Conclusion

The Italian PST-C provided a reliable measure of perceived place quality among primary school children, with evidence supporting a clear unidimensional structure and overall high semantic equivalence to the English version. The instrument appears suitable for the participatory assessment of children's living environments in Italian schools, both within the Italian PNC Clima programme and in a broader research and planning context. However, interpretation of total scores should take into account the ceiling effects associated with the current three-point response format.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by Ethics Committee of Comitato Etico Locale Palermo 1 (Approval No. 25, date: 30/10/2024). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

APa: Conceptualization, Writing – original draft, Software, Methodology, Data curation, Writing – review and editing, Investigation,

Formal analysis. APe: Formal analysis, Writing – original draft, Data curation, Investigation, Conceptualization, Writing – review and editing, Software, Methodology. GC: Formal analysis, Writing – review and editing, Methodology, Data curation, Conceptualization, Investigation, Writing – original draft, Software. PA: Conceptualization, Writing – review and editing, Methodology, Writing – original draft, Supervision. ST: Writing – original draft, Visualization, Data curation, Validation, Conceptualization. AM: Writing – original draft, Data curation, Visualization, Validation. GA: Supervision, Validation, Writing – original draft, Investigation. AB: Writing – original draft, Visualization, Project administration, Validation, Conceptualization. VMu: Writing – review and editing, Formal analysis, Validation, Writing – original draft, Methodology, Supervision. SG: Supervision, Project administration, Writing – review and editing, Investigation. VMa: Supervision, Conceptualization, Writing – review and editing, Investigation, Visualization, Project administration, Validation.

Funding

The author(s) declared that financial support was received for this work and/or its publication. This work was carried out within the programme PNC E.1 “Salute, Ambiente, Biodiversità e Clima” (PNC Clima) of the Italian National Plan for Complementary Investments to the PNRR [CUP B5I22001840001; project code PREV-A-2022-12376994]. The funding sources had no role in study design; collection, analysis and interpretation of data; writing of the report; or the decision to submit the article for publication.

Acknowledgments

We are thankful to the NHS Scotland and the Planning and Architecture Scottish Government for authorizing the use of the Place Standard Tool for its Italian adaptation and to the JAHEE for facilitating the implementation. The authors wish to express their sincere gratitude to the school institution that generously supported this research, for its valuable collaboration and the availability shown throughout the course of this work. Its contribution was instrumental to the successful completion of this study.

Conflict of interest

The author(s) declared that this work was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declared that Generative AI was not used in the creation of this manuscript.

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Christian H, Zubrick SR, Foster S, Giles-Corti B, Bull F, Wood L, et al. The influence of the neighborhood physical environment on early child health and development: a review and call for research. *Health Place*. (2015) 33:25–36. doi: 10.1016/j.healthplace.2015.01.005
- Villanueva K, Badland H, Kvalsvig A, O'Connor M, Christian H, Woolcock G, et al. Can the neighborhood built environment make a difference in children's development? Building the research agenda to create evidence for place-based children's policy. *Acad Pediatr*. (2016) 16:10–9. doi: 10.1016/j.acap.2015.09.006
- Broberg A, Kyttä M, Fagerholm N. Child-friendly urban structures: Bullerby revisited. *J Environ Psychol*. (2013) 35:110–20. doi: 10.1016/j.jenvp.2013.06.001
- Chawla L. Childhood nature connection and constructive hope: a review of research on connecting with nature and coping with environmental loss. *People Nat*. (2020) 2:619–42. doi: 10.1002/pan3.10128
- UNICEF. *Child Friendly Cities and Communities Handbook*. New York: UNICEF (2018).
- Kytä M, Broberg A, Tzoulas T, Snabb K. Towards contextually sensitive urban densification: location-based softGIS knowledge revealing perceived residential environmental quality. *Landsc Urban Plan*. (2013) 113:30–46. doi: 10.1016/j.landurbplan.2013.01.008
- Nordbø ECA, Nordh H, Raanaas RK, Aamodt G. GIS-derived measures of the built environment determinants of mental health and activity participation in childhood and adolescence: a systematic review. *Landsc Urban Plan*. (2018) 177:19–37. doi: 10.1016/j.landurbplan.2018.04.009
- Rosenberg D, Ding D, Sallis JF, Kerr J, Norman GJ, Durant N, et al. Neighborhood environment walkability scale for youth (NEWS-Y): reliability and relationship with physical activity. *Prev Med*. (2009) 49:213–8. doi: 10.1016/j.ypmed.2009.07.011
- Broberg A, Salminen S, Kyttä M. Physical environmental characteristics promoting independent and active transport to children's meaningful places. *Appl Geogr*. (2013) 38:43–52. doi: 10.1016/j.apgeog.2012.11.014
- Scottish Government NHS Health Scotland Architecture and Design Scotland. Place Standard tool (2017). Available online at: <https://www.placestandard.scot>. Accessed 3 Jul 2026.
- Kleopa D, Panayiotou A, Kouta C, Kaifa C, Middleton N. Profiling the variability and inequity in the residential environment in Cyprus according to citizens' ratings: a cross-sectional internet-based "place standard" survey. *BMC Public Health*. (2022) 22:22. doi: 10.1186/s12889-022-12706-y
- Ocaña A, Gea-Caballero V, Peiró R, Juárez-Vela R, Pérez-Sanz E, Corchón S, et al. Properties of the Spanish version of the place standard tool. *Int J Environ Res Public Health*. (2022) 19:9395. doi: 10.3390/ijerph19159395
- Ocaña-Ortiz A, Gea-Caballero V, Juárez-Vela R, Peiró R, Pérez-Sanz E, Santolalla-Arnedo I, et al. Health equity in urban and rural settings: implementation of the place standard tool in Spain. *Front Public Health*. (2024) 12:1292032. doi: 10.3389/fpubh.2024.1292032
- Gjorgjev D, Dimovska M, Morris G, Howie J, Borota Popovska M, Topuzovska Latkovikj M. How good is our place: implementation of the place standard tool in North Macedonia. *Int J Environ Res Public Health*. (2020) 17:194. doi: 10.3390/ijerph17010194
- Powell L, McCluskey Z, Bewsher H, Brown P, Gomersall T, Retzler J, et al. An evaluation of the place standard tool as a means of examining inequalities in relation to place. *Local Environ*. (2024) 29:1135–49. doi: 10.1080/13549839.2023.2284940
- Scottish Government Public Health Scotland Architecture and Design Scotland Play Scotland A Place in Childhood. The Place Standard tool: a version for children Edinburgh (2022). Available online at: <https://www.ourplace.scot/place-standard-childrens-version>. Accessed 20 Aug 2026.
- Flora DB, Flake JK. The purpose and practice of exploratory and confirmatory factor analysis in psychological research: decisions for scale development and validation. *Can J Behav Sci*. (2017) 49:78–88. doi: 10.1037/cbs0000069
- Li CH. Confirmatory factor analysis with ordinal data: comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods*. (2016) 48:936–49. doi: 10.3758/s13428-015-0619-7
- Tsang S, Royse CF, Terkawi AS. Guidelines for developing, translating, and validating a questionnaire in perioperative and pain medicine. *Saudi J Anaesth*. (2017) 11:80–9. doi: 10.4103/sja.SJA_203_17
- Reimers N, Gurevych I. Sentence-BERT: sentence embeddings using Siamese BERT-networks proceedings of the 2019 conference on empirical methods in natural language processing (EMNLP) (2019) 1:3982–92. doi: 10.18653/v1/D19-1410, emnlp-main.365,
- Reimers N, Gurevych I. Making monolingual sentence embeddings multilingual using knowledge distillation proceedings of the 2020 conference on empirical methods in natural language processing (EMNLP) (2020) 1:4512–25. doi: 10.18653/v1/2020.emnlp-main.365,
- Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Qual Life Res*. (2018) 27:1171–9. doi: 10.1007/s11136-017-1765-4
- Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, et al. COSMIN guideline for systematic reviews of patient-reported outcome measures. *Qual Life Res*. (2018) 27:1147–57. doi: 10.1007/s11136-018-1798-3
- Zanon C, Hutz CS, Yoo H, Hambleton RK. An application of item response theory to psychological test development. *Psicol Reflex Crit*. (2016) 29:18. doi: 10.1186/s41155-016-0040-x
- Graham A, Powell M, Taylor N, Anderson D, Fitzgerald R. *Ethical Research Involving Children*. Florence: UNICEF Office of Research—Innocenti (2013).
- Hayes AF, Coutts JJ. Use omega rather than Cronbach's alpha for estimating reliability. *But...*. *Commun Methods Meas*. (2020) 14:1–24. doi: 10.1080/19312458.2020.1718629
- McNeish D. Thanks coefficient alpha, we'll take it from here. *Psychol Methods*. (2018) 23:412–33. doi: 10.1037/met0000144
- Rousseelet GA, Pernet CR, Wilcox RR. The percentile bootstrap: a primer with step-by-step instructions in R. *Adv Methods Pract Psychol Sci*. (2021) 4:881. doi: 10.1177/2515245920911881
- Watkins MW. Exploratory factor analysis: a guide to best practice. *J Black Psychol*. (2018) 44:219–46. doi: 10.1177/0095798418771807
- Auerswald M, Moshagen M. How to determine the number of factors to retain in exploratory factor analysis: a comparison of extraction methods under realistic conditions. *Psychol Methods*. (2019) 24:468–91. doi: 10.1037/met0000200
- Goretzko D, Pham TTH, Bühner M. Exploratory factor analysis: current use, methodological developments and recommendations for good practice. *Curr Psychol*. (2021) 40:3510–21. doi: 10.1007/s12144-019-00300-2
- Xia Y, Yang Y. RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: the story they tell depends on the estimation methods. *Behav Res Methods*. (2019) 51:409–28. doi: 10.3758/s13428-018-1055-2
- Baker FB, Kim SH. *The Basics of Item Response Theory using R*. Cham: Springer (2017).
- Flora DB. Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Adv Methods Pract Psychol Sci*. (2020) 3:484–501. doi: 10.1177/2515245920951747
- Clark A, Moss P. *Listening to young children: the Mosaic Approach*. London: National Children's Bureau (2011).
- Harper D. Talking about pictures: a case for photo elicitation. *Vis Stud*. (2002) 17:13–26. doi: 10.1080/14725860220137345