





ORIGINAL ARTICLE OPEN ACCESS

Refining Scale Measurement: Reassessing Oral Impacts on Daily Performances Properties With Item Response Theory

Roger Keller Celeste^{1,2}  | Matheus de França Perazzo^{3,4}  | Georgios Tsakos⁵  | Michael Reichenheim⁶ 

¹Department of Preventive and Social Dentistry, Federal University of Rio Grande Do Sul, Porto Alegre, Brazil | ²Aging Research Center, Karolinska Institute, Solna, Sweden | ³Faculty of Dentistry, Federal University of Goiás, Goiânia, Brazil | ⁴Faculty of Dentistry, Evangelical University of Goiás, Anápolis, Brazil | ⁵Department of Epidemiology and Public Health, University College London, London, UK | ⁶Institute of Social Medicine Hesio Cordeiro, State University of Rio de Janeiro, Rio de Janeiro, Brazil

Correspondence: Roger Keller Celeste (roger.keller@ufrgs.br)

Received: 17 June 2025 | **Revised:** 12 September 2025 | **Accepted:** 3 October 2025

Keywords: factor analysis | item response theory | oral health-related quality of life | patient-reported outcome measures

ABSTRACT

Objectives: Many oral health-related quality of life instruments have been developed but few have undergone a comprehensive psychometric assessment. One commonly used measure is the Oral Impact on Daily Performance (OIDP). This study revised the configural and metric properties as well as the performance of items based on Item Response Theory (IRT) of a dichotomous-item version of OIDP in Brazil.

Methods: The nine-item dichotomous version of the OIDP was analysed using data from a nationally representative sample from the Oral Health Survey (SBBrasil 2010). It consisted of 30064 individuals aged 12 to 75 and was split into two partitions comprising $n_1 = 20040$ and $n_2 = 10024$, respectively. Confirmatory factor analyses (CFA) were conducted on the larger partition and cross-validated on the smaller to assess configural and metric properties. The item performance was evaluated using a 2-parameter item response theory (IRT) model. Sampling weights were used in all analyses.

Results: The unidimensional model presented two violations of conditional independence, one between items i5 (practising sports) and i4 (going out) and another between items i6 (trouble in speaking) and i7 (shame of speaking or smiling). A CFA of the most parsimonious model (removing i5, i6 and i7) yielded a RMSEA = 0.02, WRMR = 1.42, CFI = 0.99 and TLI = 0.99. The IRT analyses showed that three pairs of items had very similar levels of difficulty and discrimination suggesting redundancy.

Conclusions: A shorter dichotomous version of the OIDP scale has acceptable configural and metric properties. Being more concise and thus efficient, it may be better suited for large-scale population surveys than the version currently in use.

1 | Introduction

The relevance of Oral Health-Related Quality of Life (OHRQoL) has long been established in oral health because tooth and mouth problems impact daily activities, disrupting work and study and having a high cost to individuals and society [1, 2]. Since the 1990s, research in this area has shown a steep and continuous growth [3]. In addition to

the traditional generic OHRQoL instruments, more recent condition-specific ones have been developed to tap this concept [4]. Nonetheless, the most used ones are short versions of generic scales, abridged to be applied in large epidemiological oral health surveys. Studies undertaken in several countries have supported construct validity and reliability as minimum psychometric properties for using some of the older generic instruments [5]. However, there are criticisms for using mostly

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2025 The Author(s). *Community Dentistry and Oral Epidemiology* published by John Wiley & Sons Ltd.

external validation methods [4]. The current consensus highlights the importance of well-conducted internal validation studies as the first stage [6], for example, using confirmatory factor analysis.

One of the oldest and most widely used instruments is the Oral Impacts on Daily Performances (OIDP), developed in 1996 [7, 8]. The instrument was developed concurrently in the UK and Thailand. It assesses the impact of oral health issues on daily activities by evaluating both the frequency and severity of these impacts. Each item's score is calculated by multiplying the frequency (on a 5-point scale) by the severity (also on a 5-point scale) of the impact caused by oral disorders on daily life. The scale consists of eight items probing three hypothesised theoretical dimensions: (a) physical performances (eating, speaking, cleaning); (b) psychological performances (sleeping, smiling, emotional state); and (c) social performances (working, contact with people) [7]. In the initial version, the authors included 'light physical activities' (e.g., sports and domestic work) as a ninth item in the physical performances dimension but removed it in a later version [8]. The nine-item version has been translated and used in Brazil [9, 10] and it was included in the national oral health survey (SBBrazil) in a dichotomous item response version. Although the OIDP was developed for adults, it has also been used in several age groups; indeed, the Child-OIDP [11, 12] has been applied to children over 11 years using the same items as the adult version [13]. The OIDP dimensionality and psychometric properties have been reported in countries such as Brazil, Peru, Norway, Spain, Sweden, Turkey, Madagascar, Tanzania, Uganda, Pakistan and Sri Lanka [12, 14–23]. Overall, it has been translated to more than 80 countries in Africa, Asia, Europe, Latin America, and North America [4, 24].

Valid instruments require fine-tuning. However, essential measurement properties are not always examined in depth, and this seems to be the case for many OHRQoL measures, including the OIDP [4]. The proposed configural 3-factor model of OIDP has not been confirmed in studies, and there is reasonable evidence supporting one factor. First, violations of the original configural model have been reported, including a lack of item specificity to their respective dimension and item ambiguity, i.e., items loading into a different factor presenting unexpected cross-loadings [12, 14–17, 22]. Second, most studies described good reliability for the overall score, implying one dimension, with Cronbach's alpha values from 0.62 to 0.77. Third, each OIDP item is measured with a 25-point scale to capture a fine grain of the construct, but it is possible that such long item scale captures some random error, explaining why a study using dichotomous items showed that the one-factor solution had the best statistical fit [18] and with a ratio between the first and second eigenvalue higher than four, which strongly suggests unidimensionality [25]. Indeed, it has been suggested that using the OIDP frequency score could improve simplicity and efficiency [7]. Pilotto et al.'s study [18] also showed that the version using dichotomous items had good item reliabilities and no redundancies, presenting loadings above 0.50 and the absence of substantial residual correlations.

The state of the art of psychometric properties of OIDP indicates the need to review the configural and metric properties, given

the changes in the Brazilian version. In spite of the mounting evidence from Confirmatory Factor Analyses (CFA), scalar properties of the unidimensional model have not been assessed so far. In fact, the instrument seems to adequately order items over the latent trait according to their impact levels of OHRQoL. For example, physical impacts are more frequently endorsed than those affecting social function—the item about avoiding contact with other people due to problems with mouth and teeth had only 2.4% of endorsement, while difficulties eating had 49.7% [7]. Items in a unidimensional scale are expected to discriminate the difficulty levels to avoid redundant items and to measure the intensity of the latent trait efficiently. Therefore, this study reviewed the configural and metric properties as well as the performance of items based on Item Response Theory (IRT) of a dichotomous-item version of the OIDP in Brazil.

2 | Methods

2.1 | Design and Sample

This study used publicly available data from the Brazilian National Oral Health Survey 2010 (SBBrazil 2010), a nationally representative survey of the urban population. Data from four age groups were analysed: children 12 years old ($n = 7328$), adolescents 15–19 years old ($n = 5445$), adults 35–44 years old ($n = 9779$) and older individuals 65–74 years old ($n = 7619$). Amongst adults and older individuals, there were 2.2% and 1.3% exam refusals, respectively. However, due to the household sampling design, only 61% and 93% of the targeted sample size were reached for adults and older individuals, respectively, as households contained fewer adults than estimated. This study employed the complex sampling used in the survey, combining stratified and cluster sampling processes, and further details can be found elsewhere [26].

2.2 | Instrument and the Item Map

The OIDP version applied in SB2010 has some differences from the original version; it included nine items (Table 1) in a standard questionnaire, administered via electronic devices during household interviews conducted by trained examiners. The questionnaire began with the introductory statement: 'Some people have problems that can be caused by their teeth. Given the situations below, which applied to you in the last six months? Did (you) [...]:' followed by nine items: i1—[Did you] have difficulty eating because of your teeth or feel pain while drinking cold or hot beverages?; i2—Did your teeth bother you when brushing?; i3—Did your teeth make you nervous or irritated?; i4—Did you avoid going out, having fun, or attending parties because of your teeth?; i5—Did you stop doing sports because of your teeth?; i6—Did you have difficulty speaking because of your teeth?; i7—Did your teeth make you feel ashamed while smiling or speaking?; i8—Did your teeth cause trouble studying or working, or completing tasks at school or work?; i9—Did you stop sleeping or sleep poorly because of your teeth? The response to each question was either 'yes' (code 1) or 'no' (code 0).

The Wright Item Map [27] with the presumed order of intensity of impacts is described in Table 1, having some reference

TABLE 1 | Item map of oral impact on daily performances.

Original hypothesised dimension	SBBrazil version items content	Original version items content	Frequency of endorsement
Physical performances	i1-Difficulty Eating	i1 – Eating	Less intense ↓ More Intense
	i2-Toothbrushing	i3 – Cleaning teeth	
	i6-Trouble speaking	i2 – Speaking	
Psychological performances	i3-Making nervous or irritated	i6 – Being irritable	
	i9-Sleeping poorly	i4 – Sleeping and relaxing	
	i7-Shame smiling or speaking	I5 – Smiling without embarrassment	
Social performances	i5-Doing sport	i9 – Light Physical activities	
	i4-Avoid going out	i7 – Contact with people (Leisure)	
	i8-Trouble studying or working	i8 – Major work or social role	

points (physical, psychological and social impact items). Everyday activities that individuals carry out on their own should be theoretically the first affected/refrained items because they can be avoided/alterd/replaced without much impact. For example, this includes eating and cleaning one's teeth. On the other hand, activities that can hardly be avoided will be less affected/refrained. For example, carrying out major social roles such as going to school or work. The order of items endorsement within each dimension was obtained from a previous study [18].

2.3 | Statistical Analysis

The sample was randomly split into two groups for CFA, with partition one being 2/3 of the total sample and partition two having the remaining 1/3. The unidimensional structure was evaluated in partition one and replicated in partition two. The initial model was a strict one, without residual correlations. The goodness-of-fit of the model to the data was evaluated using the following parameters: Comparative Fit Index (CFI), Tukey-Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and Weighted Root Mean Square Residual (WRMR). The root mean square error of approximation (RMSEA) incorporates a penalty function for poor model parsimony, and values under 0.06 suggest close approximate (adequate) fit. The comparative fit index (CFI) and the Tucker-Lewis index (TLI) represent incremental fit indices, and values > 0.95 are indicative of adequate fit [28], 70. WRMR values below or close to 1 indicate good fit; although it can increase with misspecification and sample size, it is mostly informative if the *p*-value is not significant. The chi-square test was used to compare the changes in χ^2 between nested models (DIFFTEST command). A significant *p*-value ($p < 0.05$) suggests that the imposed constraints were overly restrictive, and the constrained model should therefore be rejected.

The metric structure was evaluated assessing item loadings magnitude, coefficients sign (positive/negative), and residual

correlations. High item loadings are expected, and they are related to items residuals, which are expected to be low. Residuals >|0.70| indicate that an item may be unreliable and may be considered for exclusion [29]. Potential reasons for the residual correlations were inspected individually and addressed in alternative models. Additional parameters were added and freely estimated if modification indexes were > 10 and the standardised magnitude of the coefficient was > 0.30.

Finally, the scaling properties were assessed separately in both partitions. A 2-parameter probit item response theory (2PP IRT) was estimated to assess the items' difficulty (intensity) and discrimination (slope) abilities. Additionally, the item characteristic curve of each item was plotted in a graph. Items with close difficulty levels may be considered redundant, and all items are expected to have a similar high discrimination ability, and a formal t-test (ttesti Stata command) of the difference between pairwise difficulty parameters was carried out using the standard errors provided by the software. CFAs were carried out in Mplus 7.11 and other analyses in Stata 19.

3 | Results

The initial sample comprised 30171 individuals, of which 107 (0.35%) had missing data for all OIDP items. The number of missing responses varied by item, ranging from 0.91% ($n = 274$ missing in item 5) to 0.51% ($n = 155$ missing in item 6). The weighted mean OIDP score was 1.35 (95% Confidence interval [95% CI]: 1.24–1.45), ranging from zero (56.9% of respondents) to nine (0.74% of respondents), with a median value equal to zero (interquartile range: 0–2). Table S1 shows that 58.3% of respondents were female, 44.8% had > 8 years of education, and the largest age group was adults 35–44 years old, with 40.3%. Most individuals were residents of the Southeast Brazilian region (59.8%). The item with the highest proportion of endorsement was i1 (around 25%, see Table S2) in any group by sex or age, while i5 was the item with the lowest proportion of endorsement (about 4%).

3.1 | Assessment of Configural and Metric Structure

A comparison of fit indices of the initial unidimensional (1-factor) model and different modified versions of that model is reported in Table 2 for the sample partition one (see Table S3 for the sample partition two). All models had an acceptable fit, despite violating item independence assumptions (items' residual error correlations). Two additional parameters were included and significantly improved the fit (Mplus chi-square test (DIFFTEST) of nested models: initial vs. modified initial $p < 0.01$, Table 2).

All factor loadings were > 0.70 and residuals below 0.50 (Tables 2 and S2 and Figures S2–S4). The initial unidimensional model presented two large residual correlations (> 0.30 , $p < 0.01$). The first residual correlation was between items i4–i5 (contact with

others during leisure time and physical activities/sports) and the second i6–i7 (trouble speaking and shame smiling or speaking). After detailed inspection, i5 was removed from all alternative models because it was deemed theoretically non-universal; that is, not all individuals practice sports in general; hence they cannot stop doing it. This was also supported by the low endorsement of the item (only 4%). The other residual correlation may have been caused by the wording of items, given that both of them included the idea of 'speaking problems,' either related to shame or trouble. Five alternative models were further tested: (M1) a new item was generated merging individuals who endorsed either i6 or i7, (M2) a new item was generated merging individuals who endorsed both i6 and i7, (M3) removing i6, (M4) removing i7, and (M5) removing both i6 and i7. All alternative models presented very good metric properties, and the most parsimonious was model 5 (see Figures S2–S4).

TABLE 2 | OIDP item factor loadings (standardised λ) and uniqueness (δ) 1-factor models with and without modifications (larger sub-sample $n = 20047$).

F1 BY	Initial model		Modified initial model		Alternative Model 1		Alternative Model 2		Alternative Model 3		Alternative Model 4		Alternative Model 5	
	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ	λ	δ
OIDP1	0.76	0.47	0.77	0.45	0.76	0.42	0.76	0.42	0.75	0.43	0.77	0.41	0.75	0.43
OIDP2	0.70	0.51	0.71	0.49	0.72	0.48	0.73	0.47	0.71	0.49	0.72	0.48	0.74	0.46
OIDP3	0.83	0.31	0.84	0.29	0.85	0.28	0.86	0.27	0.87	0.25	0.85	0.28	0.86	0.26
OIDP4	0.87	0.29	0.84	0.34	0.84	0.30	0.83	0.32	0.83	0.31	0.82	0.32	0.81	0.34
OIDP5	0.84	0.31	0.78	0.39										
OIDP6	0.84	0.29	0.81	0.37							0.81	0.34		
OIDP7	0.78	0.36	0.75	0.42					0.74	0.46				
OIDP8	0.84	0.30	0.85	0.27	0.85	0.28	0.83	0.31	0.87	0.25	0.83	0.31	0.83	0.32
OIDP9	0.81	0.34	0.82	0.32	0.81	0.34	0.83	0.32	0.80	0.36	0.83	0.31	0.84	0.30
OIDP6 and 7							0.81	0.35						
OIDP6 or 7					0.82	0.32								
Residual Correlations														
OIDP4 WITH OIDP5			0.52											
OIDP7 WITH OIDP6			0.41											
Fit Indices:														
CFI	0.98		0.99		0.99		0.99		0.99		0.99		0.99	
TLI	0.97		0.98		0.98		0.99		0.98		0.99		0.99	
RMSEA	0.02		0.01		0.02		0.02		0.02		0.02		0.02	
WRMR	1.89		1.44		1.61		1.38		1.68		1.41		1.42	

3.2 | Assessment of Scaling Structure

The order of intensity by endorsement of the items in Table 1 was confirmed only after removing three items (i5, i6 and i7), with i1 being the most common impact and i5 the least (Table S2). Merging i6 and i7, or alternatively using i6 or i7, proved to be redundant to other items tapping the same position in the latent trait (theta value in the x-axis that shows the difficulty parameter). Indeed, in the 2-parameter probit IRT (Tables S4 and S5), i6 was located in close position to i4 (not significant difference in difficulty level, p -value=0.63); likewise, i7 was located in a position close to i3 (not significant difference in difficulty level, p -value=0.41). The remaining 6 items (after removing i5, i6 and i7) spread almost evenly along the latent trait continuum ranging from about -1 to $+4$ of the horizontal axis (see Item Information Function [IIF], Figure S1), showing a lack of low-intensity items, as most items were located in the positive part of the latent trait.

Despite a good fit, the item characteristic curve showed that i3 (nervous or irritated) had a high discriminating slope, crossing i2 and i1 (Figure 1); while i2 (toothbrushing) had a low discriminating slope, crossing i3 and i9. Those situations violated an important assumption as items' order should not change over the continuum, unless it is located in the extreme upper or lower s-shaped curves.

4 | Discussion

The current study confirmed that a unidimensional model of a dichotomous scoring item version of the OIDP showed appropriate configural properties, as argued by Pilotto and colleagues [18]. This model also presented fewer violations in the metric evaluations, and alternative models were tested to deal with redundancy between items i4–i5 and i6–i7. Finally, an alternative model with six items performed better than the one with eight items, based on IRT results. This is the most comprehensive assessment of OIDP to the authors' knowledge.

This study has some important limitations. Validation is context-dependent; therefore, the current findings may not be generalisable to other countries without proper cross-cultural validation. This issue is sensitive because OIDP has several versions with different item content without a rigorous psychometric assessment [4]. The Brazilian version has a couple of items with different content from the original, and a change to a binary response alternative (yes/no) in each item. This new item-scoring system needs further investigation. On one side, dichotomous items might lose the ability to discriminate different levels of the intensity of quality of life; on the other side, if individuals are not able to discriminate small degrees of intensity, a dichotomous response may reduce random error [18, 30].

The initial and alternative models in the current study provided evidence for a unidimensional structure. This finding confirmed another Brazilian study that assessed the structural dimension in a CFA, showing low discriminant validity of a three-factor model with high correlations ($r > 0.90$) between factors [18]. Two previous studies reported moderate factor correlations when testing a multidimensional structure in Spain [17] and Madagascar [20], perhaps due to a different item-scoring system (frequency times severity). Violations of the initial model also include conditional dependency between items. In this regard, two studies [16, 17] reported residual correlations between pairs of items different from the current work and no clear conclusion may be drawn. In the current study, there was a content overlap of items i6 and i7 tapping on speaking difficulties, both thus requiring revision. It is worth noting that the wording in item 7 is a feature of the OIDP version used in SBBrazil, referring to 'shame' that can be linked to 'smiling' and 'speaking,' but not in the original OIDP where the respective item refers to 'embarrassment with smiling'. Moreover, item 5 was (re)introduced in the SBBrazil version based on an older preliminary version of the original OIDP, but 'stop doing sports' may not be conceptually universal (not everyone practices sports); hence, we opted to remove it. The low proportion of endorsement (around 5% in

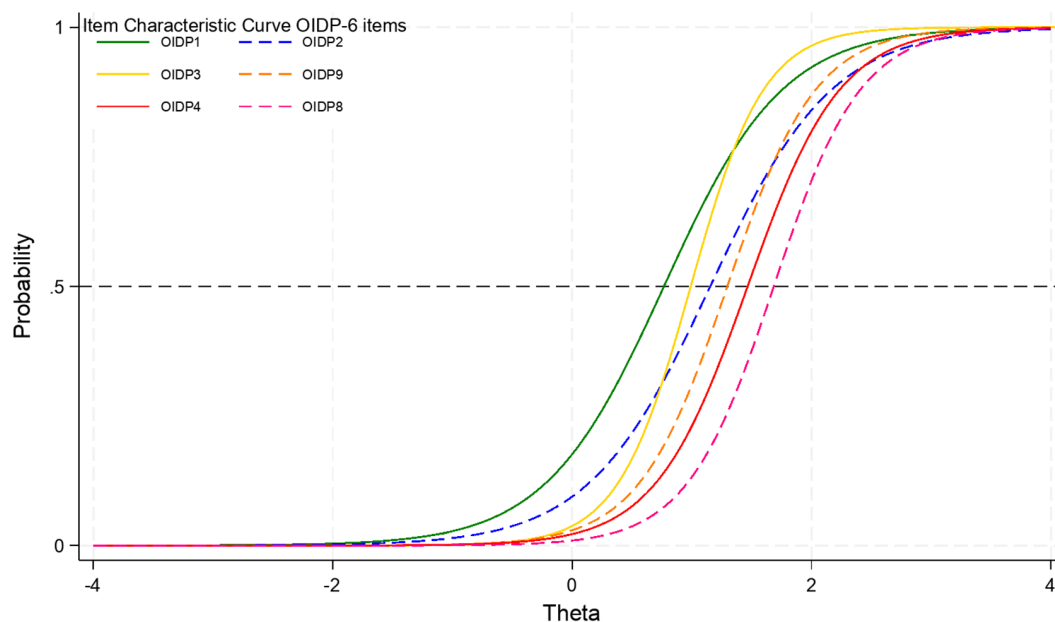


FIGURE 1 | Item characteristic curve of a six-item OIDP version (removing i5, i6 and i7).

all age groups) suggests that people either do not practice sports or they did not stop doing it because of the impact of teeth and mouth problems. Although the original item 'light physical activity' can be universally applicable, its psychometric performance cannot be anticipated and should be evaluated in future studies.

The scaling structure assessment suggests that the OIDP items can be ordered along a latent trait, though mostly with items that tap theta values higher than zero. The somewhat low proportion of endorsement (43.1% of respondents with a zero score in all items) in many items can be explained because surveys include a high proportion of healthy participants, while most instruments, such as the OIDP, have been developed to measure impacts amongst individual affected by oral diseases. The most endorsed item was 'difficulty in eating' (item 1), which has been reported as the most endorsed item in other studies [8, 12, 14, 20, 23]. Although this is the first study to assess OIDP using IRT, the order of endorsement has been similarly published elsewhere [7, 12, 14, 20, 23, 31]. It is important to highlight that raw scores and scores obtained from factor analysis are strongly related, they are different and further analysis for scalar properties must be pursued with non-parametric methods. In the current study, a 2-parameter IRT identified a few issues (after removal of i5), most notably, the theta values (difficulty coefficient) of i6 and i7 were close to i4 and i2, respectively, and this suggests redundancy. This could be due to a differential item functioning, given that people with fewer teeth (mostly older adults) may have less trouble cleaning their teeth than fully dentate ones (younger). However, this hypothesis should be further explored in future studies investigating invariance of item functioning.

In conclusion, this study indicated that a dichotomous scoring item version of the OIDP without redundant items has acceptable configural, metric, and scaling properties, and would be much easier to use in large population surveys. The use of dichotomous items may be easier in practice, and a shorter unidimensional instrument can be logistically better because it demands less cognitive effort from respondents and facilitates data collection [18, 30]. Further studies are warranted to assess whether the same items could or should be removed based on redundancy. It is important to confirm this in different contexts/cultures before implementing a final modification. Additionally, it is suggested to examine item functioning and effect invariance (using IRT), for instance across age groups.

Author Contributions

R.K.C., M.F.P., G.T. and M.R. conceptualised the manuscript. R.K.C. carried out statistical analysis and wrote a draft of the manuscript. All authors interpreted results and revised critically the manuscript. All authors approved final version.

Acknowledgements

M.R. and R.K.C. have a PQ Fellowship from the Brazilian National Research Council (CNPq number 308901/2022-3). The Article Processing Charge for the publication of this research was funded by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) (ROR identifier: 00x0ma614).

Ethics Statement

This study was performed in line with the principles of the Declaration of Helsinki. The SBBrazil 2010 Survey was approved by the National Research Ethics Committee (number 15498).

Consent

Written informed consent was obtained from study participants.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study are available in Ministério da Saúde do Brasil at <https://www.gov.br/saude/pt-br/composicao/saps/brasil-sorridente/sb-brasil/dados>. These data were derived from the following resources available in the public domain:—Ministério da Saúde—Projeto SBBrazil, <https://www.gov.br/saude/pt-br/composicao/saps/brasil-sorridente/sb-brasil/dados>.

References

1. A. Sheiham, J. E. Maizels, and A. M. Cushing, "The Concept of Need in Dental Care," *International Dental Journal* 32, no. 3 (1982): 265–270.
2. S. Listl, J. Galloway, P. A. Mossey, and W. Marcenes, "Global Economic Impact of Dental Diseases," *Journal of Dental Research* 94, no. 10 (2015): 1355–1361.
3. R. K. Celeste, J. M. Broadbent, and S. J. Moyses, "Half-Century of Dental Public Health Research: Bibliometric Analysis of World Scientific Trends," *Community Dentistry and Oral Epidemiology* 44, no. 6 (2016): 557–563.
4. F. Riva, M. Seoane, M. E. Reichenheim, G. Tsakos, and R. K. Celeste, "Adult Oral Health-Related Quality of Life Instruments: A Systematic Review," *Community Dentistry and Oral Epidemiology* 50, no. 5 (2022): 333–338, <https://doi.org/10.1111/CDOE.12689>.
5. G. Tsakos and F. Allen, "Oral Health-Related Quality of Life," in *Oral Epidemiology: A Textbook on Oral Health Conditions, Research Topics and Methods*, ed. M. A. Peres, J. L. F. Antunes, and R. G. Watt (Springer International Publishing, 2021), 319–332, https://doi.org/10.1007/978-3-030-50123-5_20.
6. L. B. Mokkink, H. C. W. De Vet, C. A. C. Prinsen, et al., "COSMIN Risk of Bias Checklist for Systematic Reviews of Patient-Reported Outcome Measures," *Quality of Life Research* 27, no. 5 (2018): 1171–1179, <https://doi.org/10.1007/S11136-017-1765-4>.
7. S. Adulyanon, J. Vourapukjaru, and A. Sheiham, "Oral Impacts Affecting Daily Performance in a Low Dental Disease Thai Population," *Community Dentistry and Oral Epidemiology* 24, no. 6 (1996): 385–389.
8. S. Adulyanon and A. Sheiham, "Oral Impact on Daily Performances," in *Measuring Oral Health and Quality of Life*, ed. G. Slade (University of North Carolina, Dental Ecology, 1997), 152–160.
9. P. S. A. de Goes, *The Prevalence, Severity and Impact of Dental Pain in Brazilian Schoolchildren and Their Families (PhD)*, ed. A. Sheiham (University College London, 2001).
10. M. I. Cortes, W. Marcenes, and A. Sheiham, "Impact of Traumatic Injuries to the Permanent Teeth on the Oral Health-Related Quality of Life in 12-14-Year-Old Children," *Community Dentistry and Oral Epidemiology* 30, no. 3 (2002): 193–198.
11. R. A. Castro, M. I. Cortes, A. T. Leao, et al., "Child-OIDP Index in Brazil: Cross-Cultural Adaptation and Validation," *Health and Quality of Life Outcomes* 6 (2008): 68.

12. K. Peker, E. Eden, A. T. Ak, Ö. Uysal, and G. Bermek, "Psychometric Evaluation of the Child Oral Impacts on Daily Performances (C-OIDP) for Use in Turkish Primary School Children: A Cross Sectional Validation Study," *BMC Oral Health* 20, no. 1 (2020): 1–12.
13. F. Gilchrist, H. Rodd, C. Deery, and Z. Marshman, "Assessment of the Quality of Measures of Child Oral Health-Related Quality of Life," *BMC Oral Health* 14, no. 1 (2014): 40, <https://doi.org/10.1186/1472-6831-14-40>.
14. M. Mtaya, A. N. Åström, and G. Tsakos, "Applicability of an Abbreviated Version of the Child-OIDP Inventory Among Primary Schoolchildren in Tanzania," *Health and Quality of Life Outcomes* 5, no. 1 (2007): 1–11.
15. E. Bernabé, A. Sheiham, and G. Tsakos, "A Comprehensive Evaluation of the Validity of Child-OIDP: Further Evidence From Peru," *Community Dentistry and Oral Epidemiology* 36, no. 4 (2008): 317–325.
16. A. N. Åström and M. Mtaya, "Factorial Structure and Cross-Cultural Invariance of the Oral Impacts on Daily Performances," *European Journal of Oral Sciences* 117, no. 3 (2009): 293–299.
17. J. Montero, M. Bravo, M. P. M.-P. P. M. P. Vicente, et al., "Dimensional Structure of the Oral Health-Related Quality of Life in Healthy Spanish Workers," *Health and Quality of Life Outcomes* 8 (2010): 24.
18. L. M. Pilotto, G. P. C. Scalco, C. Abegg, and R. K. Celeste, "Factor Analysis of Two Versions of the Oral Impacts on Daily Performance Scale," *European Journal of Oral Sciences* 124, no. 3 (2016): 272–278.
19. U. Amilani, P. Jayasekara, I. R. Perera, H. E. Carter, S. Senanayake, and S. Kularatna, "Oral Impact on Daily Performance (OIDP) Scale for Use in Sri Lankan Adolescents: A Cross Sectional Modification and Validation Study," *BMC Oral Health* 20, no. 1 (2020): 16, <https://doi.org/10.1186/s12903-020-1006-z>.
20. N. Razanamihaja and E. Ranivoharilanto, "Assessing the Validity and Reliability of the Malagasy Version of Oral Impacts on Daily Performance (OIDP): A Cross-Sectional Study," *BioPsychoSocial Medicine* 11, no. 1 (2017): 1–9.
21. A. N. Åström, G. Ekback, and S. Ordell, "Factor Structure of a Conceptual Model of Oral Health Tested Among 65-Year Olds in Norway and Sweden," *Community Dentistry and Oral Epidemiology* 38, no. 2 (2010): 110–119.
22. A. N. Astrom, G. Ekback, S. Ordell, and L. Unell, "Socio-Behavioral Predictors of Changes in Dentition Status: A Prospective Analysis of the 1942 Swedish Birth Cohort," *Community Dentistry and Oral Epidemiology* 39, no. 4 (2011): 300–310, <https://doi.org/10.1111/J.1600-0528.2010.00594.X>.
23. F. A. Chaudhary, A. Iqbal, M. D. Khalid, et al., "Validation and Reliability Testing of the Child Oral Impacts on Daily Performances (C-OIDP): Cross-Cultural Adaptation and Psychometric Properties in Pakistani School-Going Children," *Children* 9, no. 5 (2022): 631, <https://doi.org/10.3390/CHILDREN9050631>.
24. G. Tsakos, P. F. Allen, J. G. Steele, and D. Locker, "Interpreting Oral Health-Related Quality of Life Data," *Community Dentistry and Oral Epidemiology* 40, no. 3 (2012): 193–200, <https://doi.org/10.1111/j.1600-0528.2011.00651.x>.
25. M. Reichenheim and J. L. Bastos, "What, What for and How? Developing Measurement Instruments in Epidemiology," *Revista de Saúde Pública* 55 (2021): 1–17, <https://doi.org/10.11606/S1518-8787.2021055002813>.
26. A. G. Roncalli, N. N. da Silva, P. C. Narvai, et al., "Relevant Methodological Issues From the SBBrazil 2010 Project for National Health Surveys," *Cadernos de Saúde Pública* 28 (2012): s40–s57, <https://doi.org/10.1590/S0102-311X2012001300006>.
27. M. Wilson, "Constructing Measures: An Item Response Modeling Approach," in *Constructing Measures: An Item Response Modeling Approach, Second Edition*, 2nd ed. (Taylor and Francis, 2023), <https://doi.org/10.4324/9781003286929/CONSTRUCTING-MEASURES-MARK-WILSON/ACCESSIBILITY-INFORMATION>.
28. B. M. Byrne, *Structural Equation Modeling With Mplus: Basic Concepts, Applications, and Programming* (Routledge Academic, 2012).
29. M. E. Reichenheim, Y. H. M. Hökerberg, and C. L. Moraes, "Assessing Construct Structural Validity of Epidemiological Measurement Tools: A Seven-Step Roadmap," *Cadernos de Saúde Pública* 30 (2014): 927–939.
30. P. C. Beatty, D. Collins, L. Kaye, J.-L. Padilla, G. B. Willis, and A. Wilmot, *Advances in Questionnaire Design, Development, Evaluation and Testing* (John Wiley & Sons, 2020).
31. K. G. Peres, A. M. Cascaes, A. T. T. Leão, M. I. d. S. Côrtes, and M. V. Vettore, "Sociodemographic and Clinical Aspects of Quality of Life Related to Oral Health in Adolescents," *Revista de Saúde Pública* 47, no. Suppl.3 (2013): 19–28.

Supporting Information

Additional supporting information can be found online in the Supporting Information section. **Data S1:** cdoe70034-sup-0001-TableS1-S5-FigureS1-S4.docx.