Adult Oral Health-Related Quality Of Life Instruments: A Systematic Review.

Running title: OHRQoL instruments

FEDRICO RIVA¹, MARIANA SEOANE CAMPOMAR¹, MICHAEL EDUARDO REICHENHEIM², GEORGIOS TSAKOS³, ROGER KELLER CELESTE⁴

1. Faculty of Dentistry, University of the Republic, Montevideo, Uruguay
2. Institute of Social Medicine, State University of Rio de Janeiro, Rio de Janeiro, Brazil
3. Department of Epidemiology, University College London, London, United Kingdom
4. Faculty of Dentistry, Federal University of Rio Grande do Sul, Porto Alegre, Brazil.

Contributors: RKC, FR, MSC, GT and MER conceptualized the manuscript and FR and MSC carried data collection. FR wrote a draft of the manuscript. All authors interpreted results and revised critically the manuscript. All authors approved final version.

Conflict of interest: The authors report no conflicts of interest.

Acknowledgement: MER and RKC have a PQ Fellowship from the Brazilian National Research Council (CNPq).

Postal address:
Roger Keller Celeste
Faculdade de Odontologia, Universidade Federal do Rio Grande do Sul
Rua Ramiro Barcelos 2492, 3º andar
Porto Alegre – RS - CEP 90035-003
Phone: (0xx51) 3308-5015.
E-mail: roger.keller@ufrgs.br

Counter
Number of words in the abstract: 250
Number of words in the text: 2676 from Introduction to Acknowledgements
Number of references: 47
Number of Tables and Figures (max=6): 2 (3 supplementary)
ABSTRACT

To identify the existing OHRQoL instruments for adults, describe their scope (generic or specific), theoretical background, validation type, and cross-cultural adaptation. **Methods:** A systematic search was conducted and articles presenting validation of OHRQoL instruments in adults were included. Data were collected about the validation type: external validation (correlations/associations); or internal validation (Factor Analysis/Principal Components Analysis, Item Response Theory); and cross-cultural adaptation. **Results:** Of 3730 references identified, 326 were included reporting 392 studies. Forty-two original instruments were found among 74 different versions, 39 generic and 35 condition-specific. Locker’s theoretical framework was the predominant model. The Oral Health Impact Profile (OHIP) presented 20 versions, with OHIP-14 being the most frequent (26.8%), followed by Geriatric Oral Assessment Index (GOHAI) (14.0%), OHIP-49 (11.7%) and Oral Impacts on Daily Performances (OIDP) (9.7%). Most studies focused on external validation (65.3%), while internal validation was reported in 24.8% (n=26) of OHIP-14 studies, 50.9% (n=28) of GOHAI, and 21.1% (n=8) of OIDP studies. Most internal validation studies were conducted in English-speaking countries (n=33), and cross-cultural adaptation mostly in non-English-speaking European countries (n=40). **Conclusions:** Many generic and condition-specific instruments were found, but few have gone through a rigorous internal validation process, neither have undergone cross-cultural adaptation, making it difficult for researchers to choose based on psychometric properties. OHIP-14, OIDP and GOHAI seem to be the most widely validated instruments. Equalising measurement properties for comparability is challenging due to theoretical heterogeneity. Future studies should assess psychometric properties, explore the factorial structure, and work towards a consensus on critical issues.

**Keywords:** Oral Health-Related Quality of Life, Factor Analysis, Validity, Patient-Reported Outcome Measures.
INTRODUCTION

Oral Health-Related Quality of Life (OHRQoL) studies date back to conceptual models in early ‘80s\(^1,2\) - based on the International Classification of Impairments, Disabilities and Handicaps (ICIDH) - and have been growing fast since the mid-90s\(^3\). Locker et al. defined OHRQoL as the extent to which oral disorders affect the functioning and psychosocial well-being\(^1\). Similarly, the World Health Organization (WHO) defines Health-Related Quality of Life (HRQoL) as “individuals' perception of their position in life in the context of the culture and value systems in which they live and in relation to their goals, expectations, standards and concerns”. It is a broad and complex concept influenced by a person’s physical health, psychological state, level of independence, social relationships, personal beliefs and relationship to salient features of their environment\(^4\).

Many measurement tools have been developed and validated to assess the degree of impact that oral health has on Quality of Life (QoL). Differently from normative clinical measures, they include the subjective perceptions of individuals about their oral health. QoL instruments have been used in epidemiological surveys and clinical trials; guided health policies to incorporate patient-centred approaches, and assessed treatment needs\(^5,6\). It can be understood as a latent variable and, as such, can only be evaluated indirectly through composite measures\(^7\). Furthermore, it is a culturally sensitive concept, reflecting perceptions and norms that vary in different contexts\(^8\). Instruments, sometimes called measures or questionnaires, may be developed for specific age groups. Children, adults, and older adults have usually been focused on separately. Such measures may also be classified according to their scope, either generic or specific\(^9\).

The development of an instrument is a long process, and validating it is part of this process\(^10\). Initially, theoretical validation includes assessing the instrument concept and background, followed by face and content validation of proposed items. Then, internal validity assesses the dimensional structure of items, usually using factor analysis and related techniques in addition to commonly used internal consistency indicators (e.g., Cronbach’s alpha). Finally, external validity is assessed with constructs theoretically related to the instrument, based on a strong a prior hypothesis.
about how they covary, with the purpose to assess whether the instrument measures what is intended from a conceptual stance.

OHRQoL instruments have been extensively used in dental research to assess the impact of different oral conditions on daily life, beyond the setting in which they were initially developed. When an instrument requires use in a different culture, it needs to undergo a rigorous cross-cultural adaptation process\textsuperscript{11,12}.

To date, to the best of the authors' knowledge, there is no systematic review of OHRQoL instruments, much less concerning their respective developmental histories, be them theoretical or empirical. Such information would help researchers in selecting the most appropriate one in a specific setting and context. Therefore, this study aimed to identify the existing OHRQoL instruments for adults and describe their scope (generic or specific), theoretical background, validation type, and cross-cultural adaptation.

**METHOD**

Two research questions were stated: "Which are the available oral health-related quality of life instruments for the adult population?" and "Which validation methods have been mostly used?" A search strategy was developed combining two groups of strings: 1) OHRQoL terms and 2) a high-sensitivity filter to retrieve validation studies proposed by the COSMIN initiative\textsuperscript{13}. This strategy was developed using PubMed controlled vocabulary (MeSH terms) and then adapted for Scopus (see supplemental file). To include grey literature, a Google Scholar search was run, and references of two books\textsuperscript{9,14} and two previous revisions\textsuperscript{15,16} and included articles were scrutinised to detect additional papers not retrieved in the search. Besides, authors of identified instruments were contacted by e-mail if further information was needed.

**Selection criteria**

Articles assessing psychometric properties of OHRQoL instruments in the adult population were included without language or year limits until April 2021, the review is registered in Prospero (CRD42018110341). Psychometric information was also extracted from studies in which the primary purpose was not the validation of an
OHRQoL instrument but presented results about it. We excluded studies during any selection step: a) in which the whole sample was under 18 years old; b) that did not include psychometric analysis; c) not involving a QoL instrument; and d) that were review, animal, or laboratory articles. Instruments with fewer than three items were dropped\textsuperscript{17}.

**Data extraction and study variables**

Once potential studies were identified, two researchers (FR and MCS) read all the titles and abstracts, if there was insufficient information for a decision, the article was selected for full text reading. In case of disagreement, a third author (RKC or GS) was consulted, but only 45 cases remained unclear out of 3730 titles/abstracts screened. Subsequently, the following information was extracted: (i) on the development of the instrument (original instrument or a new version); (ii) on the scope of the instrument (generic oral health or condition-specific); (iii) on the main psychometric properties assessed (external or internal validation); and (iv) whether or not the study involved a cross-cultural adaptation process. In this regard, the eligibility criteria were if the authors explicitly mentioned cross-cultural adaptation as the study aim or if a pre-established guideline or necessary steps for a translation was employed. If the background of the instrument was not clear, the original reference was consulted. Also, the following information was sought: (i) first author; (ii) journal of publication; (iii) year of publication; (iv) country of the study; (v) on whether validation of the OHRQoL instrument was one of the objectives; (vi) instrument name; (vii) the number of items; (viii) aim of the instrument when developed; (ix) type of validation performed; and (x) data on the process of cross-cultural adaptation.

Several psychometric properties were reported in the included studies, and non-exclusive categories were created based on available information. We classified as addressing internal validation if an article reported results from a Principal Component Analysis, an Exploratory or Confirmatory Factor Analysis; an Item Response Theory model; or a Structural Equation Model. External validation was considered when any association was identified between the instrument and another construct.

**RESULTS**
The initial search identified 3730 references. Reading titles and abstracts excluded 3340 publications not fulfilling the eligibility criteria. The remaining 390 articles were read in full, 64 of which were subsequently excluded according to the specified eligibility criteria. As this study objective was to describe the process of instrument validation, and many articles covered more than one at a time and sometimes in different populations, the selected 326 articles effectively covered 392 validation studies (Figure 1).

Based on the data, the selected instruments were then classified into five features, namely, (i) name and the number of items; (ii) year of publication (1990-2000, 2001-2005, 2006-2010, 2011-2015, and 2016-2021); (iii) journal of publication; (iv) language and country or group of countries of publication (a single country, group of countries, or multi-country according to cultural and language similarities); and (v) validation as an objective (yes or no).

**Characteristics of the studies and the retrieved instruments**

A total of 74 OHRQoL instruments were identified, all derived from 42 original versions. Regarding their aims, 39 were generic (18 original versions), and 35 were condition-specific (33 original versions) OHRQoL questionnaires, the most frequent among the latter being aesthetic-, prosthetic-, and surgical-related instruments (Table 1). Their theoretical model was often difficult to establish because it is usually not explicitly specified\(^1,18\textsuperscript{–}21\). Locker's framework was the most widely used, with four generic instruments (Oral Health Impact Profile - OHIP, Oral Impact on Daily Performances - OIDP, Geriatric Oral Assessment Index – GOHAI, Subjective Oral Health Status Indicators - SOHSI) and ten specific ones (OHIP-Temporomandibular disorders [two instruments], OHIP-Masticatory efficiency, OHIP-Edentulism, OHIP-Aesthetic, OHIP-Prosthodontics, OHIP-Periodontitis [two instruments], OPMDQoL [Oral Potentially Malignant Disorders], OHIP-sign language). Out of 18 generic and 33 condition-specific instruments, respectively, 4 and 9 are based on Locker's framework (Figure S1 and Table 1).

The OHIP presented the highest number of variants with 20 versions. The instrument was also the most frequently validated; there were 105 validation studies (26.8%) about OHIP-14 and 46 (11.7%) about OHIP-49. The second most frequently assessed
measurement tool was the GOHAI, with 55 studies (14.0%), followed by the OIDP with 38 studies (9.7%) (Table S1).

The group of journals with the highest number of publications were Dental Journals, with 50.8% (Table S1). The number of publications has continuously been increasing over time, peaking in the 2011-2015 period with 32.4% of the cases. The majority of the studies were from English-speaking countries (USA, UK, Canada, Ireland, Australia, New Zealand) with 25.6%; while other European countries represented 18.4%, and Brazil-Portuguese was the second-highest single-language group (8.9%) (Table S1).

**Type of validation**

A larger number of studies carried out external (65.3%) over internal validation (34.7%). The GOHAI was the only single instrument subjected to more internal than external validation. Internal validation was reported in 24.8% (n=26) of the OHIP-14 studies, 21.1% of the OIDP (n=8) studies, and in 50.9% (n=28) of the GOHAI studies (Table S1).

The number of internal validation studies has been increasing over the years, peaking in the 2016-2021 period (n=44), but external validity was more frequent than internal validation in all periods (Table S1). A similar situation was found regarding the journals, were all groups presented some form of external validation as the most common focus. Dental Journals had the highest number of both internal (n=67) and external (n=132) validity studies (Table S1). Differences were found regarding the target populations for internal validation (Table S1), the English-speaking countries having most studies (n=33) followed by Brazil (n=15).

**Cross-cultural adaptation**

Cross-cultural adaptation was reported in 99 studies (Table S1), 25.3% of all studies. GOHAI presented the highest percentage (34.6%) among all instruments. Cross-cultural adaptation studies comprised 8.3% of validation studies in 1990-2000, increasing over time with the highest percentage in 2015-2021 with 35.6%. The category of non-English-speaking European countries (Scandinavian,
Germany/Netherlands and all others) had the highest number of studies (n=40). Because most instruments were developed in English, English-speaking countries presented only one cross-cultural validation study each (Table S1).

**DISCUSSION**

This review compiled, presumably, all currently available OHRQoL instruments and described their theoretical background and validation type. A variety of theoretical models were described, and the most frequent was Locker’s framework. Internal validation was performed in 34.7% of the studies, and 25.3% published some kind of cross-cultural adaptation. The number of internal validation studies was low in the early periods, showing that this methodology was not popular during the development of current generic instruments. Nevertheless, there were many external validation studies, which is an essential aspect for a comprehensive evaluation of their scope and performance.

Some limitations should be highlighted. This review should not imply that any instrument is better validated than the others or that any instrument is fully validated because it was tested in several studies. This issue is an important point since only the number of validation studies was addressed, yet without any detailing on the psychometric properties and related statistical methods. For example, some studies reported that several factors emerged from factor analysis of OHIP-14, while others only one. Although a systematic search was conducted, a few publications might not have been retrieved; even so, it is likely we identified all existing instruments.

Several well-known instruments have already undergone internal and external validation process. For a start, the number of factors in an instrument should mirror its theoretical dimensions; nonetheless, a rapid assessment shows a plethora of different factorial solutions for the same instruments. It is unclear if this is due to the different methodological approach. While most studies use classical theory (e.g., factor analysis), a few have assessed the performance over the latent trait score, using item response theory.

A larger number of instruments (n=42) was identified in the current study than in previous ones (n=14 and n=17). Probably, this reflects the rising tendency to use...
condition-specific instruments (most of them developed in the last decade), together with the demand for shorter and easily applicable versions\textsuperscript{36} in large surveys. Many condition-specific instruments were found, and some authors suggested their use in addition to generic ones to address clinically relevant factors\textsuperscript{23}. It should be noted that many specific instruments are derived from generic ones; therefore, some overlap is likely to exist either in items or the purpose. Our categorization was based on the authors’ recommendations, but additional assessment of their properties is warranted.

In the current study, an attempt was made to understand the relationship between the instruments and their theoretical models (Figure S1). \textit{A recent scoping review found nine models used in OHRQoL research}\textsuperscript{37}, however it was no clear how much those models were used for instrument development. Ideally, such development should start from a theoretical model towards the generation of items, but this process is not always clear. Sometimes, an instrument concept and dimensional structure are refined and clarified after exploratory analysis. This aspect is evident in Table 1, where only a few explicitly stated their specific theoretical models\textsuperscript{2, 38–42}. The most frequent is the Locker’s model, which is based on ICIDH, considered superseded by the International Classification of Functioning, Disability and Health (ICF)\textsuperscript{9}. Previous model follows from disease to impairment, but a shifted towards health and function has been advocated and current instruments should be assessed under new theoretical models\textsuperscript{9}.

Factor analysis provides information for internal validity by testing a postulated model (dimensional structure) to evaluate whether it explains the observed data\textsuperscript{7}. The present work identified fewer publications addressing internal validation than external validation, perhaps because researchers in the field are used to the latter type of studies. Although factor analysis is not a new method, internal validation studies were relatively scarce in the first analysed periods. Interpreting OHRQoL scores are very important\textsuperscript{43}, and factor analysis studies\textsuperscript{24–26, 28, 29} can help in this matter, refining scale properties and comparing results in different cultures, promoting understanding of the underlying constructs that the instruments cover.

Since OHRQoL is a culturally and dynamically defined concept\textsuperscript{44}, more cross-cultural adaptations are desirable in different socio-cultural and linguistic domains and across periods. Perceptions about what constitutes quality of life may change over time\textsuperscript{45}. A
relatively small number of cross-cultural adaptations assessing psychometric properties may also be an issue\textsuperscript{46}. However, some studies may not have required cross-cultural adaptation because we could not define when each instrument was used in a setting different from where it was initially developed. Therefore, the percentage of cross-cultural adaptations reported here should not be interpreted as low or high. A key aspect for further analysis will be to investigate when those instruments were used for the first time in a new setting/culture and assess in more detail the adaptation process.

Instruments described here have been applied in countries with languages and cultures different from the original versions. The target population is an important aspect to consider with assessing the validity of a measure, and a universalist approach has usually been adopted in the course of cross-cultural adaptation\textsuperscript{11,12,47}. Accordingly, qualitative studies could be considered part of the theoretical equivalence process and contribute to incorporating cultural differences in the item pool. Unfortunately, this has been scarcely reported in our findings.

In conclusion, the present study is a step towards a more comprehensive analysis of OHRQoL instruments and their theoretical background. The historical and current high number of instruments offers a broad range of measurement options for different settings. However, few have gone through a rigorous internal validation process or cross-cultural adaptation, making it difficult for researchers to choose based on psychometric properties. Although instruments are conceptually different, they may have good psychometric properties; While OHIP-14, OIDP and GOHAI seem to be the most widely validated instruments, their specific psychometric properties need to be scrutinized. Equalising the measurement properties, and therefore allowing comparisons may be challenging because of the lack of theoretical comparability. Perhaps, instead of improving any specific instrument, work towards an international forum may help develop consensus on critical issues.

ACKNOWLEDGEMENTS
There are no conflicts of interest to report. The authors would like to thank the Faculty of Dentistry librarians of UDELAR to retrieve articles for this review. Respectively, RKC and MER hold a PQ-2 and a PQ-1C Fellowships from the National Research Council (CNPq, Brazil).
REFERENCES


31. Åstrøm AN, Mtaya M. Factorial structure and cross-cultural invariance of the


47. Guillemin F, Bombardier C, Beaton D. Cross-cultural adaptation of health-