Author’s response to reviews

Title: Development of a Greek Oral Health Literacy Measurement Instrument: GROHL

Authors:

konstantina taoufik (ktaoufik@dent.uoa.gr)
Kimon Divaris (kimon_divaris@unc.edu)
Katerina Kavvadia (katerina.kavvadia@louisville.edu)
Haroula Koletsi-Kounari (hkounari@dent.uoa.gr)
Argy Polychronopoulou (argypoly@dent.uoa.gr)

Version: 1 Date: 18 Oct 2019

Author’s response to reviews:

See attached point-by-point response letter uploaded as supplemental material. The unformatted text of our response is also pasted below.

Reviewers’ comments; authors’ responses in blue; specific manuscript changes in italics and highlighted yellow in-text.

Reviewer #1

1.1- Introduction A critical review of the background of the development of oral health literacy instrument is needed. What are the limitations and recommendations found in the systematic reviews? Systematic reviews, such as "Efficacy of instruments measuring oral health literacy: a systematic review. Oral Health Prev Dent. 2014;12(3):201-7", are suggested to be included in the Introduction section.

Authors’ response: This is a great suggestion and we have incorporated the mention of this key systematic review. We reference it, as well as the strengths and recommendation of the OHL instruments identified in the report, as follows:

A recent systematic review of OHL measurement tools reported by Parthasarathy and colleagues [21] illustrated the strengths and weaknesses of the available instruments. The authors suggest that additional work to refine and validate most tools is needed, and further research is warranted to measure broader conceptualizations of OHL, address cross-cultural and language adaptation issues. It was also noted that 23 of 29 OHL reviewed studies were conducted in the United States.

Reference
1.2- Methods 1. Sample size calculation is missing in this section. Authors’ response: We did not conduct a sample size calculation. As we note in the methods section, we drew from the experience of other OHL instrument development studies and aimed to substantially surpass their sample size by 50% of more. Our instrument was based on the REALD-99 pool of items and resembled REALD-30. To provide a comparison, REALD-99 was developed among 102 adults and REALD-30 was developed among 202 adults. In this study, we initially recruited 300 individuals and our final analytical sample size was 282. Specifically, previous reports of the development of REALD-99 [13] and REALD-30 were based upon sample sizes of 102 and 202 adults [14], respectively. Based on this information, we did not conduct a formal sample size estimation or power calculation and sought to enroll a sample of over 200 individuals. Thus, we recruited a consecutive convenience sample of 300 adults who were seeking care at two private practice dental clinics in Athens, Greece.

1.3- "Participants were to read each work aloud and state whether they knew what the word meant — they were instructed not to guess." How about if the participant hesitated, or read the word very slowly? Would this be counted as correct or not? Authors’ response: This is a good question. From a methodological standpoint, if a participant hesitated or read the word slowly, s/he would be reminded that one should only read the words associated with dentistry that s/he knows the meaning of and not guess. If a participant was positive that s/he knew the word, this would be considered as ‘correct’ and we would proceed to the comprehension quiz. Of note, no such ambiguous events took place during our study, likely owing to the phonetic nature (i.e., there is a direct correlation between the spelling and the sound of each word) of the Greek language). We have made note of this feature in our methods section as follows:

1.4- How long did the participants complete the test? If this was recorded, what analyses were performed on these data? Authors’ response: As we state in the methods section, the entire interview lasted approximately 40 minutes, but included a wide array of administered instruments. The administration of our new index was not timed separately, there was no intention to record interviews of analyze any meta-data of that nature. We have added this information in our methods section as follows:
1.5- The Discussion section is relatively short. The first part of this section repeats what were stated in the Introduction, Methods and Results. More and deeper interpretation, discussion and comparison are expected in this section.

Authors’ response: According the reviewer’s comment, we have revised the first paragraph of the discussion section and further enriched several other parts of the discussion, as follows:

\[\ldots\] The scale development and testing were done among nearly 300 adults and followed common procedures and practices. Overall, we found that the GROHL demonstrated favorable psychometric properties and we recommend its further application and evaluation in additional populations, in clinical and public health settings. Additional properties of interest that can be studied in the future include, among others, its responsiveness to change (i.e., after educational interventions), its association with oral health care-seeking (i.e., our sample was limited to dental care-seeking individuals), and its potential for further item reduction (i.e., the development of a short-form GROHL version).\[\ldots\]

\[\ldots\] but it includes an additional, comprehension component, and enables a double-scoring method; i.e., upon both word pronunciation and recognition. GROHL scores were normally distributed, and were positively correlated with a wide array of other variables and constructs, including education, dental knowledge and oral hygiene behaviors. Although comparisons with literacy estimates from other studies and populations cannot be directly made, the mean score of 12/20 in this dental clinic-recruited population is within the low-end of the theorized range, in comparison to scores of 16/30 in a community setting [14], 21/30 in a University clinic [27] and 24/30 in a private dental clinic setting in the U.S. [30]. Based upon the score and information distribution of index score in this sample, a preliminary expectation would be that GROHL scores below 9 may indicate ‘insufficient’ OHL; this arbitrary threshold will certainly need to be empirically verified and interpreted separately in each application context.

\[\ldots\] Our iterative IRT approach resulted in a high-information content set of items, that contribute to the differentiation of test takers across a wide spectrum of OHL—the double-scoring method employed, similar to the Spanish-version of REALD-30 (OHLA) [15] accounts for both pronunciation and recognition, which is, in our opinion, essential for the Greek language context.

\[\ldots\] Further, word recognition and pronunciation are used here as proxies of functional health literacy—which pertains to actual skills (i.e., interpreting a prescription, recognizing signs of dental disease, following instructions to use a dental product or device, or performing oral hygiene tasks) were not tested as part of this investigation. Arguably, tests of functional literacy are in principle superior because they assess such actions or tasks of relevance to the domain of interest. For instance, the Hong Kong Oral Health Literacy Assessment Task for Pediatric Dentistry (HKOHLAT-P) [31] and the Test of Functional Health Literacy in Dentistry (TOFHLiD) [17] are based upon assessments of applied, oral health-related tasks and abilities, while word-recognition and comprehensions tests serve as proxies of these task-performing abilities. With this in mind, the development of a functional (oral) health literacy instrument in the Greek language would be a welcome and actually necessary addition. In spite of this limitation, we support that the GROHL is a step in the right direction, has favorable
psychometric properties and was found to be significantly and positively associated with important health literacy screening questions that are demonstrative of daily, applied, healthcare-related skills. In sum, we foresee that the index will enable valid measurements of OHL, for the first time, in the Greek language.

References


1.6- "...word recognition and pronunciation are used here as proxies of functional health literacy—which pertains to actual skills (i.e., interpreting a prescription, recognizing signs of dental disease, following instructions to use a dental product or device) were not tested as part of this investigation." In order to acquaint the readers or researchers who are not in the field of health literacy, the "functional health literacy" should be explained more so that it is clarified in the Limitation section that GROHL is only a quick and simple test, and more comprehensive tools in Greek are warranted. TOFHLID (reference 17) and HKOHLAT-P (Int J Paediatr Dent. 2013;23(5):366-75) can be compared for this purpose.

Authors’ response: This is a great suggestion and we have expanded upon this point in our discussion, as follows:

Further, word recognition and pronunciation are used here as proxies of functional health literacy—which pertains to actual skills (i.e., interpreting a prescription, recognizing signs of dental disease, following instructions to use a dental product or device, or performing oral hygiene tasks) were not tested as part of this investigation. Arguably, tests of functional literacy are in principle superior because they assess such actions or tasks of relevance to the domain of interest. For instance, the Hong Kong Oral Health Literacy Assessment Task for Pediatric Dentistry (HKOHLAT-P) [30] and the Test of Functional Health Literacy in Dentistry (TOFHLiD) [17] are based upon assessments of applied, oral health-related tasks and abilities, while word-recognition and comprehensions tests serve as proxies of these task-performing abilities. With this in mind, the development of a functional (oral) health literacy instrument in the Greek language would be a welcome and actually necessary addition. In spite of this limitation, we support that the GROHL is a step in the right direction, has favorable psychometric properties and was found to be significantly and positively associated with important health literacy screening questions that are demonstrative of daily, applied, healthcare-related skills. In sum, we foresee that the index will enable valid measurements of OHL, for the first time, in the Greek language.
References


Reviewer #2
2.1- Abstract must to be improved, as the results does not include all relevant information, especially those relate with IRT and validity measures.
Authors’ response: We have revised our abstract to include more information from the analysis, without attempting to include all numeric results from the narrative section of the results all Tables. Specifically, our added and revised section reads as follows:

From an initial item pool of 44 items that were carried forward to IRT, 12 were excluded due to no or little variance, 10 were excluded due to low item-test correlation, and 2 due to insignificant contribution (i.e., difficulty parameter estimate p>0.05) to the scale. The twenty remaining items composed the final index which showed favorable internal consistency (alpha=0.80) and test-retest reliability (ICC=0.95).

2.2- Introduction includes many relevant facts about OHL but fails to support the election of REALD-30 as a basis for the development of GROHL.
Authors’ response: There is always some judgment in instrument selection when multiple instruments aiming to measure the same construct exist. In our case, as we note in the introduction [lines 79-80], we selected the REALD family of indices to be based upon, because it is the most commonly used instrument in dentistry and it has been adapted to at least 5 more languages—a unique feature for OHL instruments. Also, at the time of the study conduct, other scales such as HELD and OHL-AQ had not been developed yet.

The most commonly used OHL instrument is REALD-30—a word recognition scale that has been adapted for use in several languages other than English [23-27], but not Greek.

2.3- Sample selection is clearly described but is not clear why they choose private dental clinics as a setting.
Authors’ response: We chose two different private dental clinics as recruitment venues in an effort to recruit and enroll individuals who were seeking ‘mainstream’ dental care versus specialized or referral-based care in an academic dental center/University. This is similar to the report by Jones et al. 2007 [reference: 30] in our main text. To address the reviewer’s comment, we have supplemented our methods section as follows:

The choice of two private dental clinics as recruitment venues was motivated by an effort to capture individuals who were seeking ‘mainstream’ dental care versus specialized care at an academic center.
2.4- Development of GROHL is not well described. It is not clear why they choose 44 words, nor why two dental professionals was enough to determine face validity, nor how they determine the pronunciation quality. Moreover, is quite confuse why they mixed IRT and CTT approaches in this study.

Authors’ response: We have made concerted efforts to explain these points better in our revised manuscript. Specifically: the choice of 44 words, departed from the original ‘30’ of the REALD-30 and was supplemented with additional 14 words from the REALD-99 pool of items, including words strongly associated with oral health (e.g., calculus, avulsion) and excluding: a) words used in daily routine without specific association with dentistry (e.g., diet, habits, snacking, approval), b) words having a stronger association with medicine (e.g., cancer, diabetes) and c) extremely familiar common words even if they are related to dentistry (e.g., tongue, dentist). In sum, we strived to be maximally inclusive of initial pool items that could serve the scale purpose, while excluding items that we determined upfront that would not perform well. We considered that two individuals were sufficient to determine face validity and the interviewer (first author) made the determination of correct pronunciation using her judgement based on her use and knowledge of the Greek. Importantly, participants were instructed not to guess if they didn’t know the meaning of a word and not read it. This is an important feature in our study, because Greek is a phonetic language, meaning that words sound as they are written. Further, the follow-up comprehension quiz further validated the respondents’ knowledge of the word or not. With regard to the combination (vs. mixing) of IRT and CTT approach in our study, we posit that this is strength. It is not uncommon for both approaches to be used complementary (see reference Bechger et al. 2013 among others). For instance, we used IRT to determine the optimal set of items that we could carry forward and retain in our index based on their individual performance in terms of contributing information to the overall test score. Moreover, we conducted more ‘classical’ tests of scale reliability (i.e., test-retest, internal consistency, etc.) and overall performance as we believe complement well the array of information available for this new index, and allow opportunities for its comparison with other, similarly-purposed scales.

Reference


2.5- Discussion is fair, but fails to explore the implications of all statistical measure they got, especially those related to IRT approach.
Authors’ response: We have avoided repeating statistical measures and crude results in the discussion, but we do appreciate the need to emphasize their relevance in our GROHL development work. The essence of the IRT process is that we obtained a ‘solution’ with 20 high-information content items that, in the present sample, can operate optimally to differentiate individuals with different OHL levels. In other words, we eliminated items that were contributing nothing or little to the predictive and discriminant ability of the index, resulting to a more robust and informative final collection of items. We explain and elaborate on this point in the new and augmented section of our discussion, as follows:

As expected, several words used in the REALD family of indices were not included in the GROHL-20; these 24 words did not add to the information content of the index for various reasons, ultimately due to the differential meaning and pronunciation between the English and Greek languages. Our iterative IRT approach resulted in a high-information content set of items, that contribute to the differentiation of test takers across a wide spectrum of OHL—the double-scoring method employed, similar to the Spanish-version of REALD-30 (OHLA) [15] accounts for both pronunciation and recognition, which is, in our opinion, essential for the Greek language context.

Reference