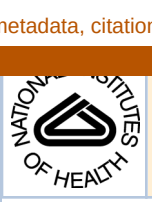

Published in final edited form as:

Community Dent Oral Epidemiol. 2011 October ; 39(5): 474–480. doi:10.1111/j.1600-0528.2011.00619.x.

Development of the Two Stage Rapid Estimate of Adult Literacy in Dentistry (TS-READ)

Brian D. Stucky, MA[Graduate Student],
The University of North Carolina at Chapel Hill, L.L. Thurstone Psychometric Laboratory, Department of Psychology, bstucky@email.unc.edu

Jessica Y. Lee, DDS, MPH, PhD[Associate Professor],
The University of North Carolina at Chapel Hill, Department of Pediatric Dentistry

Shoou-Yih Daniel Lee, PhD[Associate Professor], and
The University of Michigan, Department of Health Management and Policy

R. Gary Rozier, DDS, MPH[Professor]
The University of North Carolina at Chapel Hill, Department of Health Policy and Management

Abstract

This work proposes a revision of the 30 item Rapid Estimate of Adult Literacy in Dentistry (READ-30), into a more efficient and easier-to-use two-stage scale. Using a sample of 1,405 individuals (primarily women) enrolled in a Special Supplemental Nutrition Program for Women, Infants, and Children (WIC), the present work utilizes principles of item response theory and multi-stage testing to revise the READ-30 into a two-stage test of oral health literacy, named Two-Stage READ or TS-READ, which maximizes score precision at various levels of participant ability. Based on the participant's score on the 5-item first-stage (i.e., routing test), one of three potential stage-two tests is administered: a 4-item Low Literacy test, a 6-item Average Literacy test, or a 3-item High Literacy test. The reliability of scores for the TS-READ is greater than .85 for a wide range of ability. The TS-READ was found to be predictive of perceived impact of oral conditions on well-being, after controlling for educational level, overall health, dental health, and a general health literacy measure. While containing approximately one-third of the items on the original scale, the TS-READ was found to maintain similar psychometric qualities.

Keywords

Dental Health Literacy; Dental Care; Oral Health Quality of Life; Health Literacy; Psychometrics

INTRODUCTION

Dental health literacy is defined by the NIDCR Working Group on Functional Health Literacy as "the degree to which individuals have the capacity to obtain, process, and understand basic oral health information and services needed to make appropriate health decisions" (1). This working group also concluded that in-depth dental studies on literacy and its effects on dental health will be difficult without instruments for quantifying people's dental health literacy. A dental health literacy instrument could have many practical uses. For example, it could be used to screen for individual dental health literacy in clinic settings.

Correspondence: Brian D. Stucky. MA, L.L. Thurstone Psychometric Laboratory, Department of Psychology, University of North Carolina at Chapel Hill, CB 3270 Davie Hall, Chapel Hill, NC 27599.

Results of the screening could improve the communication between dental health care providers and patients (e.g., improving patient's understanding of dental health information, treatment, and intervention). Moreover, researchers and public health practitioners could use the instrument to assess the level of dental health literacy in a group of patients or a community. Used in that way, the instrument is critical for identifying factors and outcomes of dental health literacy and for devising interventions to effectively improve dental health and quality of life.

Since the release of the 2005 NIDCR report, at least five instruments have been developed to measure dental health literacy: 1) Rapid Estimate of Adult Literacy in Dentistry-30 (2), 2) Rapid Estimate of Adult Literacy in Dentistry-99 (3), 3) Test of Functional Health Literacy in Dentistry (4), 4) Oral Health Literacy Instrument (5) and 5) Rapid Estimate of Adult Literacy in Medicine and Dentistry (6). One additional instrument measures oral health knowledge in the context of dental health literacy (7). Although all of the instruments have promising characteristics, testing has been limited to specific populations and has not taken advantage of recent advancements in psychometric testing. Given the growing interest in measuring dental health literacy and modern methods for assessing measurement properties, on-going evaluation and refinement of these instruments is important.

Improving a Dental Health Literacy Scale

The present work is an extension of initial work done in developing the REALD-30 (Rapid Estimate of Adult Literacy in Dentistry; (5)), a 30-item scale consisting of words chosen based on etiology, anatomy, prevention, and treatment. The words were designed to be read aloud by the person whose dental health literacy is being assessed, and scored based on correctness of pronunciation. The original scale was found to produce reliable scores (Cronbach's alpha = 0.87), while high scores were predictive of positive oral health-related quality of life as measured by the Oral Health Impact Profile (OHIP-14; (6)) after controlling for a number of covariates, including dental visits, gender, and education level (2).

Recent confirmatory studies have found REALD-30 to be predictive of various risk indicators of self-rated oral health (7,8) and a clinical measure of dental health status (9). However, the scale has not undergone a detailed psychometric review. The present work takes a model-based approach to determine the properties of the items and to shorten the administration of the scale. Initially, a one-factor confirmatory factor analysis model is fit to determine the appropriateness of a single factor in explaining the covariance between the items comprising the REALD-30. Next, after setting aside poorly fitting items, the remaining items are calibrated using item response theory (IRT). Finally, based on the IRT item parameters, a two-stage test is developed (Two-Stage REALD) which maintains the score reliability of the original scale while substantially reducing test length.

Item Response Theory

This investigation is based on an item response theory analysis of the REALD-30. IRT is a model-based psychometric technique used to examine the relationship between item responses and the underlying latent ability (in this case, dental health literacy). The relationship between an item response and the latent ability is represented by an item characteristic curve (ICC), which is typically the logistic distribution. For binary items, the two-parameter logistic model (2PL) ICC for the probability of correctly answering an item is:

$$P_i(\theta)=\frac{1}{1+\exp\{-Da_i(\theta-b_i)\}}, \quad (1)$$

where $P_i(\theta)$ is the probability that an examinee with ability θ (i.e., dental health literacy) answers item *i* correctly; a_i is the discrimination parameter indicating the strength of relationship between the item and θ; b_i is the difficulty parameter and indicates the ability level associated with a .50 probability of answering item *i* correctly; and *D* is a scaling constant of 1.7 used to transform the metric from logistic to normal with a mean of 0 and standard deviation of 1.

Like many scale development applications of IRT, the present study makes use of *item information* to indicate the characteristics of REALD-30 items. Based on equation 1, item information reflects how precisely a given item measures ability across the continuum:

$$I(\theta)=a_i^2 P_i(\theta)Q_i(\theta). \quad (2)$$

The amount of item information across θ is the product of the item's squared discrimination parameter *a*, and the probabilities of correct (P_i) and incorrect response (Q_i). Item information indicates the utility of a particular item at various locations along the ability continuum. Hence item information is useful in comparing the performance of individual items at various levels of dental health literacy. Finally, to consider the measurement properties of a set of items (i.e., a scale or test), item information is summed over all items to provide *test information*.

Item and test information are most interpretable when illustrated. For example, consider two items with the following parameters: *item 1* has a discrimination parameter of 2 and difficulty parameter of −1 (i.e., $a_1 = 2$, $b_1 = -1$), and *item 2* has a discrimination parameter of 3 and difficulty parameter of 1 (i.e., $a_2 = 3$, $b_2 = 1$). Figure 1 provides the item information and test information functions from both items.

Figure 1 illustrates the concepts of item and test information which will later be used to develop the Two-Stage REALD. Recall that *item 1* was less discriminating than *item 2* (*a* = 2 and 3, respectively). This property is reflected in the heights of the item information functions. The magnitude of item information reflects the measurement precision along the ability continuum (i.e., dental health literacy). Note that the maximum of the information function is located at θ = −1 and 1, for items 1 and 2, respectively. While *item 2* provides a greater maximum in information, it is the more difficult item and therefore out-performs *item 1* only from about θ = 0 to about 3. For lower levels of ability, *item 1* is a better performing item. When both items are considered simultaneously, the resulting test information indicates that scores for the overall scale are more precise for higher levels of ability. Finally, information may be translated into reliability by taking one less the inverse of information. Hence, the scale indicates that reliability is .5 when information is 2.0 (1-1/2.0 = 0.5).

Multi-Stage Tests

Multi-stage tests (MSTs), commonly used in educational settings, contain sets of items that are administered adaptively based on the ability of the respondent (10,11). In the first stage, a routing test, comprised of a small number of items covering a broad range of difficulty, is administered to obtain an initial estimate of ability. Based on the routing test score, a second stage (i.e., stage-two) of the test is administered which contains items with a difficulty level

similar to the ability of the test-taker. Thus, the second-stage provides a more refined estimate of ability.

The adaptive nature of MSTs improves measurement efficiency as fewer items are needed to achieve score reliability at a given level of ability. Because MSTs administer items relevant to the ability of the test-taker, the predictive and concurrent validity of MSTs has been demonstrated to be at least equal to traditional fixed-length tests (12).

METHODS

Sample Characteristics

A sample of 1405 English-speaking adults, recruited from North Carolina Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) clinics, participated in the present study. The majority of participants were female (96.5%). The sample varied in educational level: 23.7% did not finish high school, 38.5% finished high school or received a GED, 24.6% attended or completed community college, and 13.1% attended or completed college. The sample reflected a diverse ethnic background: 42.2% self-identified as White, 40.5% as African American, 19.0% as American Indian, and 0.85% Asian.

Analysis Plan

We first assessed the factor structure of the REALD-30 using confirmatory factor analysis (CFA) of the inter-item polychoric correlation matrix. The analysis was performed using weighted least squares mean and variance adjusted (WLSMV) estimation in the software program M*plus* (13). We examined the fit of the single-factor model as well as the presence of local dependence (i.e., multi-dimensionality).

Next, the software program MULTILOG (14) was used to obtain the IRT item parameters. Once obtained, item information was computed at a variety of score locations between low and high dental health literacy (see eq. 2). The resulting two-stage test, named Two-Stage REALD or TS-REALD, was constructed based on the magnitude of item information at various levels of dental health literacy.

Finally, we assessed the concurrent validity of the TS-REALD by evaluating its correlation with the Newest Vital Sign (NVS), a scale designed to identify individuals at risk for low health literacy (15), on the assumption that there should be a high correlation between dental health literacy and general health literacy. We also assessed the predictive validity of the TS-REALD by regressing the oral health impact profile (OHIP), a 14-item measure of the perceived impact of oral health conditions on overall well-being (6), on the TS-REALD, holding constant educational level, overall health, dental health, and health literacy (NVS).

RESULTS

The Dimensionality of the REALD-30

Prior to conducting factor analyses of the inter-item polychoric correlation matrix, items *sugar* and *smoking* were set aside because they had few incorrect responses (both items were pronounced correctly by 99.5% of the sample). Additionally, *apicoectomy* was set aside because only 1.4% of the sample responded correctly.

After removing the three items, a 27-item CFA model was fit to the data (including a residual correlation between the item pair *brush* and *floss*). The resulting model was found to closely fit the data according to commonly used assessments of model fit (16,17,18): $\chi^2(114) = 613$, CFI = 0.95, TLI = 0.97, RMSEA = 0.056. To ensure unidimensionality, the

item *brush* was set aside from the single item pair, resulting in 26 unidimensional items to be calibrated.

Item Calibration and Development of the Two-Stage REALD

IRT calibration of the remaining 26 items resulted in slope parameters ranging from 2.91 to 1.09 (mean = 2.12, *SD* = 0.44) and difficulty parameters ranging from 1.97 to −2.93 (mean = −0.01, *SD* = 1.42). Item information was then calculated for each of the 26 items. Items with the greatest magnitude of information at five dental health literacy score locations (−1.5, −0.75, 0, .75, and 1.5 standard deviations around mean) were selected as the routing test. From easiest to most difficult, those items included *denture, abscess, restoration, fistula*, and *temporomandibular*.

Next, the three stage-two tests were developed by computing the information provided by each item at three score locations of dental health literacy (−1.5, 0, and 1.5 standard deviations around the mean). The development of the stage-two tests capitalized on the additive nature of item information. Specifically, items were sorted by the magnitude of information at each of the three score locations and then iteratively added until the total test information reached the designated level of 6.67, equivalent to a score reliability of 0.85. Table 1 contains items and IRT item parameters for the routing and all stage-two items. Note that because items at the mean did not discriminate as well as those on the Low and High Literacy second-stage tests, more items were required for the Average Literacy test to compensate for the lack of information.

The raw score of the routing test determines which stage-two test to administer. Respondents with raw scores of 0 or 1 on the routing test receive the “Low Literacy” stage-two test; those with scores of 2 or 3 receive the “Average Literacy” stage-two test; and those with scores of 4 or 5 receive the “High Literacy” stage-two test. Our analysis showed that about 20% of the sample would receive the Low Literacy test, 62% the Average Literacy test, and 19% the High Literacy test. Thus, dividing the sample into the three stage-two tests approximates what would be expected from normally distributed data, suggesting that the routing test directs the expected number of test-takers to the correct stage-two test.

Figure 2 provides the test information functions for each of the three stage-two tests based on a standardized metric with a mean of 50 and a standard deviation of 10. The figure illustrates the efficiency of the two-stage test. From nearly two standard deviations below the mean to two standard deviations above the mean (i.e., from 30 to 70), scores maintain a reliability greater than 0.85, which was achieved by using the three tests with a minimum number of items for a given score. In other words, each of the three stage-two tests maximizes information at a particular score location, thus saving the administration of items not appropriate for a given dental health literacy level.

Scoring the Two-Stage REALD

Table 2 illustrates the scoring of the new two-stage scale, TS-REALD, which requires translating the raw summed score (column “summed score” in Table 2) to the IRT-scaled score (column “scaled score” in Table 2). Using principles of IRT, the score translation places the different stage-two tests on the same standardized IRT metric (19), allowing the comparison scores for individuals taking different stage-two tests. Thus, while each stage-two test contains a different set of items, the scaled scores for individuals remain comparable between tests. For ease of score interpretation, the standardized metric was transformed to have a mean of 50 and a standard deviation of 10. Using an example from Table 2, a summed score of 3 on the Low Literacy test (the most likely score for this stage-

two test) is associated with a scaled score of 38, indicating that an individual with this score is a little more than one standard deviation (1.2) below the mean.

Validity Evaluation

Initially, the appropriateness of the new scoring procedure was confirmed by considering a correlation of 0.96 between the TS-READ and the original READ-30. Next, the correlation between TS-READ and NVS was 0.51 ($p < .05$) in the study sample, providing evidence of convergent validity. Finally, holding constant the subjects' educational level, overall health, dental health, and NVS, the TS-READ remained a statistically significant predictor for OHIP in a multiple regression model ($\beta = .10$, $se = .04$, $p < .05$). This result confirms the predictive validity of TS-READ, beyond the contribution of general health literacy to oral health-related quality of life.

DISCUSSION

The present work proposes a revision to the READ-30. The revised Two-Stage READ capitalizes on a strength of the original scale—i.e., that it contains items of widely varying difficulty. The new two-stage test format contains a routing test, which serves to define the respondent's initial dental health literacy level, and three stage-two tests, which refine or more accurately measure the respondent's dental health literacy score. Because the scoring is based on a standard metric, the resulting scores are comparable and easily interpretable despite the administration of different test items to different individuals.

The TS-READ allows tailoring the administration of the scale to a respondent's dental health literacy level. Unlike the original READ-30 scale, which requires the administration of an entire battery of test items, the new scale uses only a subset of test items that are adequately suited to the respondent's dental health literacy level. This unique feature of the TS-READ has several advantages. First, the scale is approximately one-third the length of the original READ-30 and therefore takes a shorter amount of time to administer. Second, the new scale may be less threatening and therefore more receptive to low literacy individuals because overly difficult words are avoided. Third, participant response rate may improve because the scale is shorter and tailored to the respondent's dental health literacy level.

Limitations

Despite these advantages, our analysis of the TS-READ is limited in two key ways. First, our sample was predominately female, and while prior work investigating the READ-30 identified no gender differences in dental health literacy (2), future researchers using the TS-READ should not assume that the measure operates equivalently across gender. Second, the TS-READ achieves only in part the objectives of an ideal measure of health literacy as previously identified (20). Baker recognizes an ideal measure as one that assesses reading fluency, vocabulary, prior and conceptual knowledge of health, and potential difficulties in understanding written health materials and comprehending health care professionals' speech. While the development of such a comprehensive assessment remains a challenging goal, these concepts underscore the success and limitations of the TS-READ in measuring a broad description of dental health literacy. Clearly, the context-depended nature of patient-provider communication leaves ample room for future investigation.

Conclusions

Prior work using the READ-30 has established a relationship between dental health literacy and a variety of health outcomes, including poor oral health-related quality of life (2), as well as other risk indicators of poor oral health (7,8,9). Consistent with these

findings, our analysis of the TS-READ showed a significant relationship between dental health literacy and the impact of oral health conditions on overall well-being, after controlling for differences in education, levels of overall and dental health, and risk for low health literacy. Together, these results suggest the importance and the unique contribution of dental health literacy to oral health and oral health-related life quality.

Acknowledgments

This study was supported by grants from the National Institute of Dental and Craniofacial Research (NIDCR) Award Numbers RO1DE018045 and R01DE018236. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIDCR or the National Institutes of Health (NIH).

REFERENCES

1. National Institute of Dental and Craniofacial Research. The invisible barrier: literacy and its relationship with oral health. A report of a workgroup sponsored by NIDCR, USPHS, DHHS. J Public Health Dent. 2005; 65:174–182. [PubMed: 16171263]
2. Lee JY, Rozier G, Lee S, Bender D, Ruiz RE. Development of a word recognition instrument to test health literacy in dentistry: the REALD-30 - a brief communication. J Public Health Dent. 2007; 67:94–98. [PubMed: 17557680]
3. Richman JA, Lee JY, Rozier RG, Gong D, Pahel BT, Vann WF Jr. Evaluation of a word recognition instrument to test health literacy in dentistry: the REALD-99. J Pub Health Dent. 2007; 67(2):99–104. [PubMed: 17557681]
4. Gong DA, Lee JY, Rozier G, Pahel BT, et al. Development and testing of the Test of Functional Health Literacy in Dentistry (TOFHLiD). J Public Health Dent. 2007; 67:105–112. [PubMed: 17557682]
5. Sabbahi D, Lawrence HP, Limeback H, Rootman I. Development and evaluation of an oral health literacy instrument for adults Community Dent Oral Epidemiol. 2009; 37:451–462.
6. Slade GD. Derivation and validation of a short-form oral health impact profile. Community Dent Oral Epidemiol. 1997; 25:284–290. [PubMed: 9332805]
7. Atchison KA, Gironda MW, Messadi D, Der-Martirosian C. Screening for oral health literacy in an urban dental clinic. J Public Health Dent. (In press).
8. Macek MD, Haynes D, Wells W, Bauer-Leffler S, Cotton PA, Parker RM. Measuring conceptual health knowledge in the context of oral health literacy: preliminary results. J Public Health Dent. (In press).
9. Miller EK, Lee JY, DeWalt DA, Vann WF Jr. Impact of caregiver health literacy on children's oral health outcomes. Pediatrics. (In press). NIHMCID: 196223.
10. Betz, NE.; Weiss, DJ. Psychometric Methods Program, Research Rep. No. 73-4. Minneapolis: Department of Psychology, University of Minnesota; 1973. An empirical study of computer administered two-stage ability testing.
11. Lord, FM., editor. Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum; 1983.
12. Weiss DJ. Improving measurement quality and efficiency with adaptive testing. Appl Psych Meas. 1982; 6:473–492.
13. Muthén, LK.; Muthén, BO. Mplus user's guide. 5th Ed. Los Angeles, CA: Muth0én & Muthén; 2007. [Computer Software]
14. Thissen, D. MULTILOG 7.03. Lincolnwood, IL: Scientific Software International; 2003. [Computer Software]
15. Weiss BD, et al. Quick assessment of literacy in primary care: the Newest Vital Sign. Ann Fam Med. 2005; 3:514–522. [PubMed: 16338915]
16. Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Modeling. 1999; 6:1–55.
17. Browne, MW.; Cudeck, R. Alternative ways of assessing model fit. In: Bollen, KA.; Long, JS., editors. Testing structural equation models. Newbury Park, CA: Sage; 1993. p. 136-162.

18. Bentler PM. Comparative fit indexes in structural models. Psychol Bull. 1990; 107:238–246. [PubMed: 2320703]
19. Thissen, D.; Orlando, M. Item response theory for items scored in two categories. In: Thissen, D.; Wainer, H., editors. Test Scoring. Mahwah, NJ: Lawrence Erlbaum Associates; 2001. p. 73-140.
20. Baker DW. The meaning and the measure of health literacy. J Gen Intern Med. 2006; 21:878–883. [PubMed: 16881951]

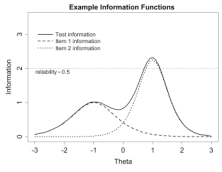

Figure 1.
For this illustration θ, the latent variable, is on a scale with a mean of 0 and standard deviation of 1. Hence, scores greater than 0 on the θ indicate ability levels greater than the mean.

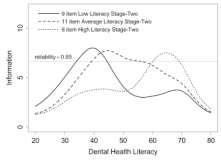

Figure 2. TS-REALD information functions

For this illustration θ, dental health literacy, is on a scale with a mean of 50 and standard deviation of 10. Note that each Stage-Two test reaches information levels of at least 6.67, or reliability of 0.85.

Table 1

IRT item parameters for the routing test and stage-two tests.

	Item Parameters	
Item Stem	***a***	***b***
Routing test items:		
Denture	2.40	−1.57
Abscess	1.58	−0.83
Restoration	2.47	−0.55
Fistula	2.27	1.07
Temporomandibular	2.91	1.97
Low Literacy stage-two items:		
Braces	2.05	−2.18
Plaque	1.83	−1.59
Pulp	2.96	−1.03
Extraction	2.18	−0.82
Average Literacy stage-two items:		
Enamel	2.03	−0.69
Genetics	2.25	−0.67
Sealant	1.87	−0.49
Halitosis	1.89	0.36
Cellulitis	1.86	0.61
Incipient	2.44	0.85
High Literacy stage-two items:		
Hyperemia	2.48	1.17
Hypoplasia	2.49	1.21
Analgesia	2.44	1.75

Table 2

Score translation table for the Two-stage REALD

Low Literacy		Average Literacy		High Literacy	
Summed Score	**Scaled Score**	**Summed Score**	**Scaled Score**	**Summed Score**	**Scaled Score**
0	27	---	---	---	---
1	31	---	---	---	---
2	35	2	39	---	---
3	38	3	43	---	---
4	41	4	45	4	56
5	45	5	48	5	61
---	---	6	51	6	64
---	---	7	54	7	68
---	---	8	57	8	73
---	---	9	61	---	---

Note: Dashes indicate unattainable scores based on the Two-stage REALD scoring rules. The Low Literacy test has 6 possible scores, ranging from 0 to 5 (a score of 0 or 1 from the routing test in addition to 4 possible points from the four-item stage-two test). The Average Literacy test has 8 possible scores, ranging from 2 to 9 (a score of 2 or 3 from the routing test in addition to 6 possible scores from the six-item stage-two test). The High Literacy test has only 5 possible scores, ranging from 4 to 8 (a score of 4 or 5 from the routing test in addition to 3 possible scores from the three-item stage-two test). Scaled scores are on a metric with a mean of 50 and a standard deviation of 10.