Patient Education and Counseling 97 (2014) 88–95

Contents lists available at ScienceDirect

Patient Education and Counseling

journal homepage: www.elsevier.com/locate/pateducou

Assessment

Measurement properties and comparative performance of health literacy screening questions in a predominantly low income African American population with diabetes

Fatima Al Sayah [a], Sumit R. Majumdar [b], Leonard E. Egede [c], Jeffrey A. Johnson [a,*]

[a] Department of Public Health Sciences, University of Alberta, Edmonton, Canada
[b] Department of Medicine, University of Alberta, Edmonton, Canada
[c] Department of Medicine, Medical University of South Carolina, Charleston, USA

ARTICLE INFO

Article history:
Received 7 February 2014
Received in revised form 18 June 2014
Accepted 3 July 2014

Keywords:
Health literacy
Measure
Measurement properties
Measurement model
Diabetes

ABSTRACT

Objective: To examine the measurement properties of the 16 screening questions (16-SQ) of inadequate health literacy (HL) and their briefer version (3-SQ), and identify the best screen for inadequate HL in non-white populations.
Methods: Sample included 378 individuals with type-2 diabetes. We computed sensitivity, specificity, positive and negative likelihood ratios, and C-indices, using the s-TOFHLA as a reference measure. We also conducted exploratory factor analysis, and used structural equation modeling (SEM) for confirmatory purposes.
Results: Mean age was 56.1 years, 69% were female, and 83% were African–American. 10% had limited HL (s-TOHFLA scores <23). Six questions (6-SQ) were identified and included in the final item-reduced factor analysis, which showed good fit in confirmatory SEM (chi-square = 9.5; $P = 0.305$; RMSEA = 0.023). Weighted summative score of the 6-SQ and the item "difficulty understanding written information" performed better than the 3-SQ in identifying patients with inadequate HL (C-indices 0.67 versus 0.75).
Conclusion: The weighted summative score of the 6-SQ and the item "difficulty understanding written information" performed better than the other items or combinations of these items in identifying individuals with inadequate HL.
Practice implications: The proposed weighting of scores could be applied in studies using these screening questions for better classification of inadequate HL.

1. Background

Health literacy (HL) is "the ability to obtain, process, understand, and communicate basic health information needed to make informed health care decisions" [1]. Inadequate HL has been found to adversely influence health outcomes, especially in low-income patients with chronic diseases [2–5]. Although routine screening for inadequate HL in clinical settings is still controversial, it is high (46%) prevalence in the US population [6], and its association with poor health outcomes [7,8], has led to an increased interest in HL assessments. However, common measures of HL, such as the test of functional health literacy in adults—short form (s-TOFHLA) and the rapid estimate of adult literacy in medicine (REALM), are time consuming, require face-to face interviews, might introduce discomfort and embarrassment particularly in clinical settings, especially for those with inadequate HL, cannot be administered by telephone, and they are not feasible in large surveys [9–12].

Chew and colleagues developed 16 self-reported HL screening questions (16-SQ) [13], then identified a briefer version of three questions (3-SQ) [14]. They subsequently reported that out of the 3-SQ, a single item about "confidence with completing forms" with a response cut-point of "somewhat," may be sufficient to detect patients with inadequate HL; this item did not, however, perform as well in identifying patients with limited (i.e. inadequate plus marginal) HL [13,14]. This single item was also reported by others to perform best in identifying patients with inadequate HL at a university-based primary clinic [15]. Chew and colleagues also found that a scale combining the three questions offered no

* Corresponding author at: Department of Public Health Sciences, 2-040 Li Ka Shing Centre for Health Research Innovation, University of Alberta, Edmonton, Canada T6G 2E1. Tel.: +1 780 492 9266; fax: +1 780 492 7455.
E-mail address: jeff.johnson@ualberta.ca (J.A. Johnson).

http://dx.doi.org/10.1016/j.pec.2014.07.008
0738-3991/

additional benefit to the one question about confidence with forms [14]. The question about needing help to read hospital materials was predictive of inadequate HL in sample of patients at a university based vascular surgery clinic [16]. This same question, which is known as the single item literacy screener (SILS), performed reasonably well in ruling out inadequate HL in adults [17]. Sarkar & colleagues later evaluated the performance of the 3-SQ among Spanish and English-speaking individuals with type 2 diabetes, and found that one of the 3 items "confidence with forms" or a summative score of the three items were both useful in identifying inadequate HL in this population [18].

Overall, the evidence on the utility of these screening questions in identifying inadequate HL, as single items or a combination of these items, is inconsistent. This evidence is based on studies that validated these questions among predominantly white, English speaking populations drawn from academic practices. Additionally, the original identification of the 3-SQ from the 16-SQ was based only on receiver operating characteristic (ROC) analysis in one patient population [13], while no investigation of the factor structure of the 16-SQ has been undertaken. Furthermore, in all of the studies that used a summative score of these items [13–15,18,19], it was done by simple summation of the item scores assuming that all these items equally contribute to the total score. Therefore, we sought to examine the factor structure and the measurement properties of the 16-SQ and the 3-SQ in greater detail, and examine whether there is a better set of items to screen for inadequate HL in a predominantly lower income non-white population.

2. Methods

2.1. Design and data source

This validation study used cross-sectional data from a study conducted in South Carolina, USA that has been previously described in detail [20]. Briefly, patients were recruited at two adult primary care clinics, and were included if they were 18 years or older with a diagnosis of type 2 diabetes in their medical record and a clinic appointment between June and August 2010. Patients were ineligible if they did not speak English or if the research assistants determined that they were too ill or cognitively impaired to participate. Ethics approval of this study was obtained from the Institutional Review Board (IRB) at the University of South Carolina.

2.2. Self-reported health literacy measure

Participants completed the 16-SQ assessing HL, including the 3-SQ on "difficulty understanding written information" (HL12), "confidence with forms" (HL14), and "needing help in reading hospital materials" (HL16). All questions were scored on a five-point Likert scale (always = 1, often = 2, sometimes = 3, occasionally = 4, never = 5) with higher scores indicating lower HL. The scores of items HL5–HL13 and item HL16 were reversed so that higher scores indicate lower self-reported HL.

2.3. Reference health literacy measure

The s-TOFHLA was administered to all subjects; we considered the reference measure as it is the most frequently used HL measure in the literature [9], and was used in the validation studies of these questions. The s-TOFHLA is the short form of the TOFHLA, which was developed in the United States to measure "functional health literacy", defined as assessing reading, writing, and numeracy skills in relation to health [21]. The s-TOFHLA includes 36 reading comprehension and four numeracy items, and uses the modified cloze procedure, where every fifth to seventh word in passage is omitted, and the respondent selects a response from four options [10]. The s-TOFHLA scores range from 0 to 36, with higher scores indicating better reading comprehension, and thus higher functional HL. We used standard cut-offs where scores from 0 to 16 represent inadequate HL, 17–22 marginal HL, and 23–36 adequate HL [10]. s-TOFHLA scores of 0–22 are collectively referred to as limited HL. We assessed the performance of the self-reported questions and their summative score compared to the s-TOFHLA categories of inadequate (scores 0–16) and limited HL (scores 0–22) [10].

2.4. Other measures

Data on self-reported age, sex, race/ethnicity (black; white), years of formal education (grade school to post-secondary), employment status (employed; unemployed), health insurance (insured; uninsured), and annual household income (<$10,000; <$25,000; >$25,000) were collected. Diabetes knowledge was assessed with the diabetes knowledge questionnaire (DKQ) [22], and diabetes-specific self-efficacy with the perceived diabetes self-management scale (PDSMS) [23]. Most recent A1c results were obtained from patients' medical records.

2.5. Statistical analysis

2.5.1. Reliability and validity

Descriptive statistics and estimates of reliability and validity were computed. Internal consistency reliability was assessed using Cronbach's alpha coefficient. Criterion validity was assessed by examining correlations of each of the 16-SQ with the s-TOFHLA. Construct validity was assessed using the hypothesis-testing approach. Specifically, we established a priori hypotheses about the direction and magnitude of correlations of the 16-SQ with related traits and constructs based on evidence on the relationships between HL and these constructs [2,7]. Based on previous research, we hypothesized that the 16-SQ would be positive and moderately correlated with education, positive and strongly correlated with diabetes knowledge, and positive and moderately correlated with self-efficacy [2,24,25]. We used the following criteria for the strength of correlation: <0.3 "weak"; 0.3–0.5 "moderate"; and >0.5 "strong" [26].

2.5.2. ROC analysis

We calculated C-indices (the area under the ROC curve) for each question for the HL categories of inadequate (comparing s-TOFHLA scores of 0–16 versus 17–36) and limited (comparing s-TOFHLA scores of 0–22 versus 23–36). We considered a C-index greater than 0.6 to be useful; this is higher than the 0.5 cut-off that reflects discrimination no better than chance [27]. We also calculated sensitivity, specificity, and positive (LR+) and negative (LR−) likelihood ratios for each question.

2.5.3. Factor analysis

We conducted exploratory factor analysis of the 16-SQ using the principal factor analysis method and oblique Promax rotation. The eigenvalues from the factor analysis were used to determine the number of factors in the optimum solution. The 16 questions were assigned to the factor on which they loaded most heavily in the rotated solution. Next, to minimize redundancy and simplify the 16-item model, we eliminated redundant items. To do so, we grouped the items based on their content and measurement scope, and selected the best item within each set based on factor loadings and C-indices. Among the 16 questions, four items assess "ease of reading" (set 1: HL1–HL4), four assess "difficulty in understanding" (set 2: HL5–HL8), five assess "problems due to difficulty in understanding" (set 3: HL9–HL13), and the remaining miscellaneous