BMC Public Health

CrossMark

Limitations of the S-TOFHLA in measuring poor numeracy: a cross-sectional study

Ashley J. Housten[1*] , Lisa M. Lowenstein[1], Diana S. Hoover[2], Viola B. Leal[1], Geetanjali R. Kamath[3] and Robert J. Volk[1]

Abstract

Background: Although the Short Test of Functional Health Literacy in Adults (S-TOFHLA) is widely used, misidentification of individuals with low health literacy (HL) in specific HL dimensions, like numeracy, is a concern. We examined the degree to which individuals scored as "adequate" HL on the S-TOFHLA would be considered as having low HL by two additional numerical measures.

Methods: English-speaking adults aged 45–75 years were recruited from a large, urban academic medical center and a community foodbank in the United States. Participants completed the S-TOFHLA, the Subjective Numeracy Scale (SNS), and the Graphical Literacy Measure (GL), an objective measure of a person's ability to interpret numeric information presented graphically. Established cut-points or a median split classified participants and having high and low numeracy.

Results: Participants ($n = 187$), on average were: aged 58 years; 63% female; 70% Black/African American; and 45% had a high school degree or less. Of those who scored "adequate" on the S-TOFHLA, 50% scored low on the SNS and 40% scored low on GL. Correlation between the S-TOFHLA and the SNS Total was moderate ($r = 0.22$, $n = 186$, $p = 0.01$), while correlation between the S-TOFHLA and the GL Total was large ($r = 0.53$, $n = 187$, $p \leq 0.01$).

Conclusions: Findings suggest that the S-TOFHLA may not capture an individuals' HL in the dimension of numeracy. Efforts are needed to develop more encompassing and practical strategies for identifying those with low HL for use in research and clinical practice.

Trial registration: NCT02151032 (retrospectively registered: May 30, 2014).

Keywords: Health literacy, Numeracy, Decision-making, Health disparities

Background

Health literacy (HL) represents a complex intersection of skills needed to "obtain, process, understand, and communicate about health-related information needed to make informed health decisions." [1–3] The 2003 National Assessment of Adult Literacy (NAAL), a representative survey of 19,000 adults in the United States (US), found that approximately half of all adults demonstrate HL related difficulties, and over one-third (36%) have basic or below basic HL [4]. Compared to individuals with higher HL, those with limited HL are found to use fewer preventive services (e.g., cancer screening) and are more likely to engage in unhealthy behaviors (e.g., poor

medication adherence), resulting in increased risk for hospitalization and diminished health outcomes [5, 6]. Furthermore, estimates suggest that low HL costs the US economy between $106–$238 billion annually, and accounts for between 7%–17% of personal healthcare expenses [7]. Due to the common occurrence of limited HL, and its corresponding social and economic impact on population health, it is a top public health priority [5, 8]. With a recent shift in healthcare practice to prioritize patient involvement in medical decision making, measuring HL in order to evaluate patient abilities, develop patient-centered interventions, and promote patient empowerment in the healthcare setting continues to gain support [8, 9]. Given the importance of HL, our aim was to look at the performance of the most commonly used HL measure, The Short Test of Functional Health Literacy in Adults (S-TOFHLA) [9–11].

* Correspondence: ajhousten@mdanderson.org
[1]Department of Health Services Research, Division of Cancer Prevention & Population Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd.Unit 1444, Houston, TX 77030, USA
Full list of author information is available at the end of the article

Housten *et al. BMC Public Health* (2018) 18:405

Page 2 of 9

HL measures are useful for evaluating and classifying patient abilities so that information can be presented in a way to meet patients' skills and needs. Yet, existing measures of HL may lack the specificity to accurately assess patients' ability to comprehend numeric information, providing a limited view of patients' abilities [8, 9, 12–14]. The comprehensive measurement of HL is challenging within clinical settings because HL includes multiple elements, such as print literacy, speaking and listening (oral and aural literacy), cultural knowledge, social skills, and numeracy [2, 9, 13, 15, 16]. Numeracy, defined as one's aptitude with probabilities, fractions and ratios [16, 17], is of primary interest among those focused on developing risk communication strategies to promote patient engagement in healthcare decisions [14, 18]. Risk estimates and numerical information designed to depict probabilities, percentages, frequencies and trade-offs are widely used in patient decision support materials such as decision aids, but are often poorly understood even among those with higher HL [8, 9, 13, 18–20]. Objective numeracy measures provide insight into individuals' ability to understand numerical and quantitative information; yet, individuals may be reluctant to objective test questions (e.g. math test questions, probability test questions) and more amenable to subjective measures (e.g., self-reported comfort with numbers, preference for numerical information), without compromising clinical utility [12, 17, 21]. While there is general consensus about the importance of evaluating HL and its associated dimensions, there is no agreed upon "gold-standard" measure, and there is limited agreement about which dimensions of HL can be measured while maintaining clinical feasibility [9, 22]. Moreover, over half of commonly used measures of health literacy have limited psychometric properties and often lack reporting on critical types of validity (e.g., content, construct, criterion, internal, predictive) [9, 12]. As a result, acceptable strategies are needed that address the limitations of existing HL measures, particularly in the numeracy related dimensions [9].

The S-TOFHLA is the most frequently used measure of HL, used in over half of all published papers measuring HL [9, 11]. However, it measures reading fluency, leaving out key domains in HL [8–10, 23], and is often not feasible to use in clinical settings due to limited time and resources for administering and scoring the measure [24]. Prior research has questioned established S-TOFHLA scoring and categories [9, 25–31]. Thus, the purpose of this study was to look at the performance of the S-TOFHLA in identifying those with limited numerical HL when compared to a subjective and an objective numerical HL measure.

Methods
Study design
This study was part of a randomized controlled trial (clinicaltrials.gov: NCT02151032) designed to investigate the use of decision aids in colorectal cancer screening. Eligible participants were English-speaking, aged 45–75 years, and had no history of colorectal cancer. Participants were recruited in person between November 2012 and January 2013 in the Greater Houston Metropolitan area from: [1] a large academic cancer center and [2] a nonprofit community foodbank, to engage those with varying HL levels.

After providing written informed consent, participants completed a battery of questionnaires, including a measure of demographic characteristics and three measures that assess HL related competencies: the S-TOFHLA, the Subjective Numeracy Scale (SNS), and the Graphical Literacy Measure (GL). Questionnaires were paper-based and completed in-person. The research assistant was present during the completion of the questionnaires and answered questions as needed. We included these four HL related measures because they represent a breadth of HL related constructs that may be of interest to researchers and of importance in clinical settings. The cut-points used to categorize HL levels are described below. These cut-points were primarily based on the S-TOFHLA since that is the most commonly used HL measure. This study was approved by the Institutional Review Board of the sponsor institution.

Measures
Demographic characteristics
Participants reported sociodemographic characteristics, such as age, gender, race/ethnicity, education level, and self-rated general health status (5-point Likert scale with scores ranging from 0 = poor to 4 = excellent).

Short test of functional health literacy in adults (S-TOFHLA)
The S-TOFHLA is a short version of the Test of Functional Health Literacy in Adults (Table 1) [10, 32]. This objective measure was designed to evaluate general HL by assessing individuals' ability to read and understand health-related information [9, 10, 31]. For this investigation, we used the 36-item reading comprehension subscale, which is used in approximately 71% of papers using the S-TOFHLA [9–11]. We used the 36-item version of the S-TOFHLA because of its wide acceptance and use in HL research [9–11]. Reading passages are written at 4th and 10th grade levels, and using a modified Cloze procedure, the fifth and seventh words are removed, tasking the reader to choose the best response from four choices [10, 32].

Subjective numeracy scale (SNS)
Numeracy is a vital element of HL and is a priority for optimizing risk communication [13, 17]. The SNS is a self-report measure of one's subjective ability to execute math related tasks and preferences for numbers versus

Housten *et al. BMC Public Health* (2018) 18:405

Page 3 of 9

Table 1 Health Literacy Measures and Cut-points

Measure	Scores	Cut-points	Rationale for chosen cut-point*
Short Test of Functional Health Literacy in Adults (S-TOFHLA)			
S-TOFHLA Total Score	0–36 Range	≤16 Low 17–22 Marginal ≥23 Adequate	The S-TOFHLA has a 7 min time limit and scores range from 0 to 36 with items worth 1 point each [10]. Based on their scores, participants are classified as having "inadequate" (0–16), "marginal" [17–22], or "adequate" [23–36] HL [10, 32]. The S-TOFHLA has good validity when compared to the full Test of Functional Health Literacy (TOFHLA) and internal consistency (Cronbach's alpha = 0.97), yet its correlation with the Rapid Estimate of Adult Literacy in Medicine (REALM) is unclear [10, 54]
Subjective Numeracy Scale (SNS)			
SNS Total Score	1–6 Mean	≤4 Low > 4 High	SNS cores range from 1 to 6 and we used the average SNS total score for analysis [17, 21, 33]. There is no established optimal cut-point for the SNS. Various methods have been used [27, 34–36]. We selected to use the median split score of 4 (median score on total and subscales) to be the cut-point for this investigation. This cut-point was chosen to compare those who are scoring above the median score to those who score below. The SNS is a reliable (Cronbach's alpha = 0.82) and valid numeracy measure when compared to Objective Numeracy Scale items (r = 0.63–0.68) [17, 21].
SNS Ability Subscale	1–6 Mean	≤4 Low > 4 High	
SNS Preference Subscale	1–6 Mean	≤4 Low > 4 High	
Graphical Literacy Measure (GL)			
GL Total Score	0–13 Range	≤6 Low > 6 High	Each item is worth 1 point and scores range from 0 to 13, with higher scores suggesting higher graphical literacy [37]. There is no established optimal cut-point for the GL. The median cut-point for the total score and for the subscales (levels) was chosen to compare those who have some difficulty with graphical literacy (low) to those who do not have less difficulty with graphical literacy (high) [37, 43–46]. The GL is an adequately reliable (Cronbach's alpha = 0.79) and valid (Construct validity = 0.54; Convergent validity = 0.50) measure of graphical literacy [37].
GL 1, Reading	0–4 Range	≤3 Low > 3 High	
GL 2, Between	0–4 Range	≤1 Low > 1 High	
GL 3, Beyond	0–5 Range	≤2 Low > 2 High	

*References reported in the table are the sources used for the cut-points

prose [17, 21]. It does not contain math questions and there are no correct or incorrect answers [17, 21]. The SNS contains eight items in two 4-item subscales: Ability and Preference [17, 21]. Response options are on a 6-point Likert scale and scores range from 1 to 6 [17, 21]. The average of each participant's responses to all eight items is calculated to create their subjective numeracy score, and higher scores indicating higher subjective numeracy [17, 21]. Average scores are also calculated for the Ability and Preference subscales [33]. Since there is no universal SNS score cut-point to identify those with limited HL, we opted to use a median split to categorize participants as having either high or low HL for the total score and for each subscale [27, 34–36].

Graphical literacy measure (GL)

The GL is a 13-item objective measure that assesses how individuals understand graphically-presented quantitative information [37]. Understanding graphical information is an important dimension of numeracy [38]. Thus, quantitative and graphically-presented information is a critical component of HL and making health decisions [39–41]. Using graphical images, the GL assesses abilities related to graph comprehension by increasing level of difficulty: (GL 1) "the ability to read the data," or "find specific information in the graph," (GL 2) "the ability to read between the data," or "find relationships in the data as shown on the graph," and (GL 3) "the ability to read beyond the data," or "make inferences and predictions from the data" [37, 42]. For example, one line graph shows years on the x-axis and percent of people with a fictional disease, "Adeolitis," on the y-axis, with three questions: [1] "Approximately what percentage of people had Adeolitis in the year 2000?," [2] "When was the increase in the percentage of people with Adeolitis higher?," and [3] "According to your best guess, what will the percentage of people with Adeolitis be in the year 2010?" [37] Because no verified optimal score cut-point exists, scores were categorized into low and high graphical literacy based on a median split [37, 43–46].

Description of health literacy measure cut-points

As reported in the previous section, no optimal score cut-points exist for the SNS and the GL. Thus, we used three complementary scoring strategies. First, we used the median score from the total score to identify low and high literacy groups, based on a median split. This approach is justified based on the study setting, where more than half of the participants came from a community-based organization serving underserved groups. Second, a more conservative threshold was used, where the lowest quartile of participants from the total scores were classified as having low HL, an approach common in educational literature [47, 48]. We would

Housten et al. BMC Public Health (2018) 18:405

Page 4 of 9

expect few of these participants to score within the "adequate" range on the S-TOFHLA. Third, we used the population level estimate (36th percentile) for American adults who have basic or below HL, from the 2003 National Assessment of Adult Literacy (NAAL) [4].

Statistical analysis

Demographic characteristics, frequencies, and descriptive statistics were generated to summarize our participant population using IBM SPSS Version 23. The three S-TOFHLA HL categories ("inadequate", "marginal", and "adequate") were used for scoring. For the other HL measures, we categorized the variables into high and low categories (Table 1). These categories were based on existing score cut-points or using the median split. For categorical variables, single proportion confidence intervals were generated to assess the proportion of the S-TOFHLA score levels and the other HL measures. Pearson's Correlations were used to assess the direction and strength of correlation between total scores on the HL measures (small ≥0.10; moderate ≥0.30; large ≥0.50) [49].

Results

Participants

One hundred eighty-nine individuals enrolled in the study (67 from the cancer prevention center and 122 from the foodbank) and completed the questionnaires; however, two participants were excluded from the analyses due to missing data, for a total of 187 participants (Table 2). One hundred eighty-seven participants completed the 36-item S-TOFHLA, but one participants did not complete all other HL measures, which is noted in the tables. Participants' mean age was 58 years. Over two-thirds (70%) of participants identified as Black or African American and about one-third (37%) reported some college or trade school. More than two-thirds (71%) reported good, very good, or excellent health.

Health literacy measures

Almost three-quarters of participants (71.7%; $n = 134$) scored "adequate" on the S-TOFHLA, while only 10.7% ($n = 20$) scored in the "marginal" category, and 17.6% ($n = 33$) scored "inadequate" (Table 3). The three strategies (median, 25th percentile, 36th percentile) to identify optimal thresholds for determining low HL for the SNS and GL reinforced each other and, therefore, we report the median and lowest 25th percentile scores to describe the results (Figs. 1 and 2). Notably, of those who scored "adequate" on the S-TOFHLA, about half scored low on SNS measures and over one third scored low on GL measures (Fig. 2).

Of those who scored "adequate" on the S-TOFHLA, 26% ($n = 35$) of participants were in the bottom 25th

Table 2 Demographic Characteristics.

Demographic Characteristic	Total ($n = 187$) n (%)
Age in years, Mean (SD)	58 (8.1)
45-59	110 (58.8)
60-75	77 (41.2)
Sex (female)	118 (63.1)
Race/ethnicity	
Black/African American	131 (70.1)
Hispanic/Latino	7 (3.7)
White	38 (20.3)
Mixed Race/Other	11 (5.9)
Education	
Less than high school graduate	36 (19.5)
High school graduate	49 (26.5)
Some college/trade school	68 (36.8)
College degree	32 (17.3)
Health Status	
Excellent	11 (5.9)
Very good	40 (21.4)
Good	81 (43.3)
Fair	43 (21.4)
Poor	12 (5.9)

percentile of the SNS Total score. Moreover, for the Ability and Preference Subscales, 30% ($n = 40$) and 35% ($n = 47$) of participants were categorized as "adequate" by the S-TOFHLA but in the bottom 25th percentile of their respective scale scores.

Forty percent ($n = 53$) of those classified as "adequate" on the S-TOFHLA were low on GL Total score. Furthermore, 18% ($n = 24$) of those scoring in the bottom 25th percentile on GL Total, scored "adequate" on the S-TOFHLA. For the GL Subscales, 13% ($n = 18$) of those classified as "adequate" on the S-TOFHLA were in the bottom 25th percentile of the GL1 Subscale. For the GL2 subscale, 21% ($n = 28$) of those classified as "adequate" on the S-TOFHLA were in the bottom 25th percentile. Lastly, 30% ($n = 40$) were classified as "adequate" on the S-TOFHLA and in the bottom 25th percentile on the GL3 Subscale.

Correlations between S-TOFHLA scores and SNS Total ($r = 0.22$, $n = 186$, $p = 0.01$; Table 4) and SNS Ability scores ($r = 0.23$, $n = 187$, p = 0.01). The correlation between the S-TOFHLA and SNS Prefer scores was not significant ($r = 0.14$, $n = 186$, $p ≤ 0.06$). The S-TOFHLA and GL Total and GL 1 scores showed large, positive correlations ($r = 0.53$, $n = 187$, $p ≤ 0.01$ and $r = 0.52$, $n = 187$, $p ≤ 0.01$, respectively). The correlation for the S-TOFHLA and GL3 was moderate in magnitude ($r = 0.37$, $n = 187$, $p ≤ 0.01$).

Housten *et al. BMC Public Health* (2018) 18:405

Page 5 of 9

Table 3 Health Literacy Measure Scores[*]

Health Literacy Measure Scores (n = 187)	n (%)
S-TOFHLA	
Total Score (Median (Mean [SD]))	31 (26.8 [10.0])
Inadequate	33 (17.6)
Marginal	20 (10.7)
Adequate	134 (71.7)
Subjective Numeracy (SNS)	
SNS Total Score[†] (Median (Mean [SD]))	4 (3.9 [1.0])
Low (≤4)	100 (53.8)
High (> 4)	86 (46.2)
SNS Ability Subscale (Median (Mean [SD]))	4 (3.8 [1.3])
Low (≤4)	108 (57.8)
High (> 4)	79 (42.2)
SNS Preference Subscale[†] (Median (Mean [SD]))	4 (4.1 [1.1])
Low (≤4)	101 (54.3)
High (> 4)	85 (45.7)
Graphical Literacy (GL)	
GL Total Score (Median (Mean [SD]))	6 (6 [3.3])
Low (≤6)	94 (50.3)
High (> 6)	93 (49.7)
GL 1, reading the data (Median (Mean [SD]))	3 (2.6 [1.5])
Low (≤2)	114 (61.0)
High (> 2)	73 (39.0)
GL 2, reading between the data (Median (Mean [SD]))	1 (1.6 [1.4])
Low (≤1)	94 (50.3)
High (> 1)	93 (49.7)
GL 3, reading beyond the data (Median (Mean [SD]))	2 (1.7 [1.1])
Low (≤2)	141 (75.4)
High (> 2)	46 (24.6)

[*]Numbers are shown as n and percentages unless noted otherwise
[†]n = 186

Discussion

This study raises concerns about the 36-item S-TOFHLA, a commonly used measure that has been used to identify individuals with low HL, in identifying individuals with limited numeracy. Results indicated that a large proportion of participants whose scores characterize them with "adequate" HL based on the S-TOFHLA scored low on measures of individuals' ability to understand and interpret quantitative information.

Our results suggest that individuals categorized as having low HL on quantitative HL measures will be misclassified as having "adequate" HL with the S-TOFHLA. This is critical, as individuals with HL difficulties are at-risk for slipping through the cracks and may not receive the numerical support they need if they are screened with the S-TOFHLA. The S-TOFHLA only assesses limited aspects of HL, and yet, it persists as the most commonly-used HL measure in both research and clinical contexts [9, 11]. The current results support previous findings that participants are over-classified with "adequate" HL on the S-TOFHLA when compared to other HL measures [9, 11, 25–30, 50]. Moreover, our findings build on existing literature by adding evidence for the notable numeracy and graphical deficits of the widely used 36-item S-TOFHLA, challenging the utility of the S-TOFHLA and its use as a general HL measure.

The SNS and GL Total and Subscale score findings highlight the deficits of the S-TOFHLA for assessing basic and advanced numeracy skills, such as understanding risk, probabilities, percentages, and frequencies. While the 36-item S-TOFHLA was not designed specifically to assess numeracy, it is being used to assess general HL, of which numeracy is a critical component. Additionally, objective and subjective measures may capture different skills associated with HL and using both types of questions may be needed to reduce participant burden without compromising clinical utility. Correlations between the S-TOFHLA and the SNS were small
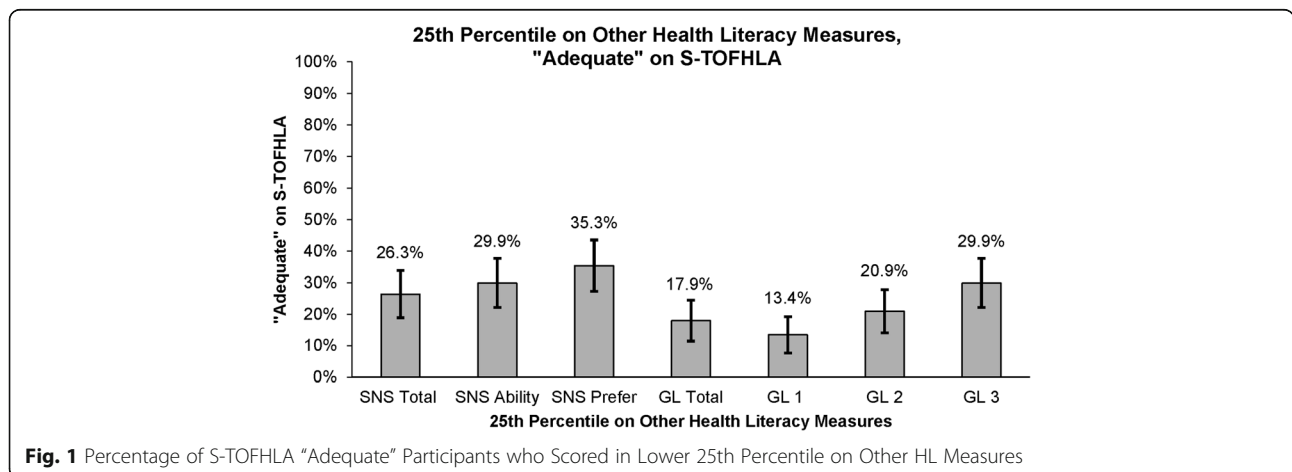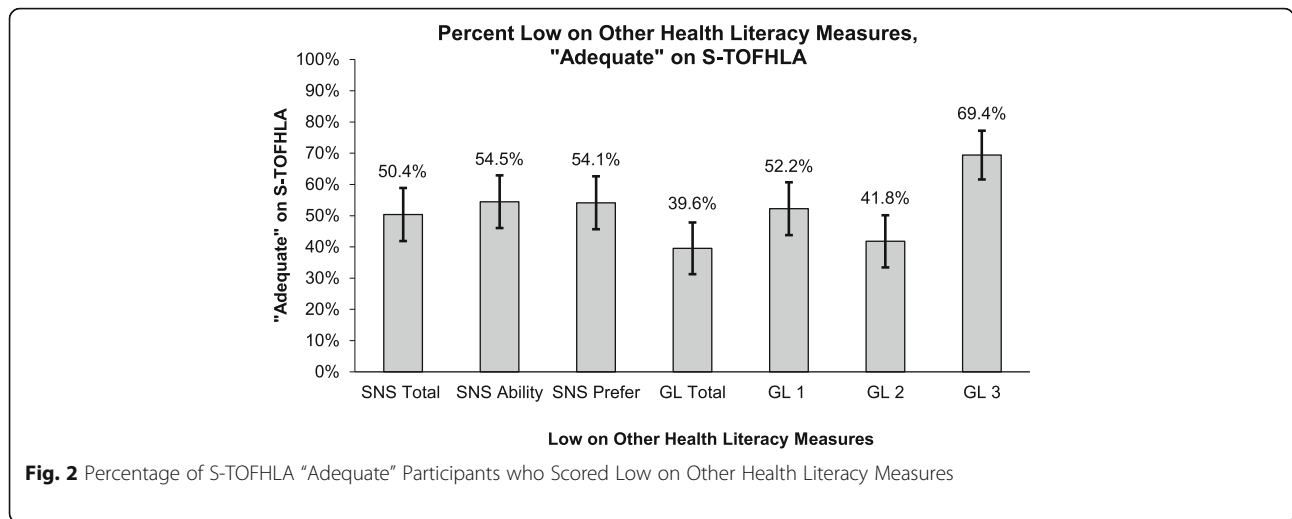

Fig. 1 Percentage of S-TOFHLA "Adequate" Participants who Scored in Lower 25th Percentile on Other HL Measures

Housten et al. BMC Public Health (2018) 18:405

Page 6 of 9

Fig. 2 Percentage of S-TOFHLA "Adequate" Participants who Scored Low on Other Health Literacy Measures

to moderate, while correlations between the S-TOFHLA and the GL scales were moderate to large. The latter correlations between the S-TOFHLA and the GL scales may be due to both being objective measures. Despite these associations, the S-TOFHLA still misclassified many individuals based on the numeracy scores. Our findings question the broad acceptance and use of the S-TOFHLA as a universal measure of HL. A more systematic approach that provides supports for those who have deficits in HL may be a better intervention strategy rather than over-relying on limited, individual HL measures.[51–54]

Our findings add to the understanding of challenges associated with HL measurement. In order to make informed choices, patients must understand the likelihood of achieving a benefit or a harm from a treatment. Approaches that identify those with limited HL and numeracy are needed to ensure that patients receive support (if needed) to engage in these types of decisions. HL is a complex construct, and existing literature identifies and describes multiple ways of measuring it. For example, Duell et al. identified three levels for HL measurement:

basic, communicative/interactive, and critical HL [9]. These levels are similar to the three GL Levels: reading the data, reading between the data, and reading beyond the data [37, 42]. In the current study, over half of those who scored "adequate" on the S-TOFHLA scored low on the GL1 subscale (reading the data). Additionally, about two-thirds of those who scored "adequate" on the S-TOFHLA scored low on the GL3 subscale (reading beyond the data). This can be observed in the correlations between the S-TOFHLA score and the GL subscales scores decrease as the GL level increases, suggesting that the S-TOFHLA may not adequately capture these more advanced level numeracy skills. For promoting patient involvement in medical decision making, numeracy is a primary skill needed to understand risk, probabilities, percentages, frequencies and trade-offs [14, 18]. Results highlight how those scoring "adequate" on the S-TOFHLA lack not only the advanced skills, but the basic HL skills needed to function in healthcare settings which may inhibit patient engagement in medical decision making.

Table 4 Pearson Correlations Between Health Literacy Measures (n = 187 [unless otherwise noted])

	S-TOFHLA	SNS Total[b]	SNS Ability	SNS Prefer[b]	GL Total	GL 1	GL 2
SNS Total[b]	0.22[a]						
SNS Ability	0.23[a]	0.89[a]					
SNS Prefer[b]	0.14	0.83[a]	0.47[a]				
GL Total	0.53[a]	0.43[a]	0.41[a]	0.33[a]			
GL 1	0.52[a]	0.35[a]	0.34[a]	0.26[a]	0.91[a]		
GL 2	0.43[a]	0.40[a]	0.38[a]	0.31[a]	0.88[a]	0.73[a]	
GL 3	0.37[a]	0.34[a]	0.33[a]	0.26[a]	0.73[a]	0.50[a]	0.44[a]

[a]Correlation is significant at the 0.01 level (2-tailed)
[b]n = 186

Housten et al. BMC Public Health (2018) 18:405

Page 7 of 9

There are various approaches to help address the challenges associated with measuring HL. First, the assumption that a single HL measure is adequate may not be the case. The HL measures included in our investigation show the need for capturing the complex skills that make-up HL. While previous studies have provided evidence to push back against commonly-used measures, such as the S-TOFHLA and REALM, a continued effort to challenge the expected use of one of these tools as a way to definitively identify those with low HL is needed [11, 50]. Furthermore, simply challenging the existing score cut-points employed by the S-TOFHLA may not be enough to identify those with limited HL skills as our findings show the discordance compared to objective and subjective numerical HL measures using multiple score cut-points. Second, there is a need for the development of a feasible strategy to capture patients' ability to interpret and apply quantitative information in clinical and research settings. Developing strategies that incorporate subjective and objective factors critical to assessing HL, such as graphical literacy, culture, physiological condition, and relevance to disease type, are to be considered in updated measurement strategies [23, 51]. Third, is the priority for incorporating HL principles and strategies to support patient-centered care [51–53]. Strategies such as narratives, engaging storytelling and other visual supports may reduce patient burden and promote engagement for those with both high and low HL.

This research study has potential limitations. This study was conducted in a large urban area using a convenience sample. Thus, the sample is diverse and matches the makeup of large urban centers, but rural patients may have not been well represented which may impact generalizability. Second, we used the 36-item S-TOFHLA measure, which does not assess numeracy. Although this measure is broadly accepted [10], including the additional four numeracy items may have provided more detailed numeracy information. We used cut-points to categorize HL levels, which is consistent with research and clinical use of the S-TOFHLA and enabled comparisons between measures. Optimal score cut-points did not exist for the SNS and GL. To address this limitation, we used a median split approach, and more generous score cut-points of the 25th and 36th percentiles. With this strategy, we were able to present different measure score cut-points and compare them to the S-TOFHLA categories.

Conclusions

The S-TOFHLA is limited in its measure of HL. Though widely used, those with inadequate HL skills may be over-classified as having "adequate" HL, based on their S-TOFHLA scores. Use of HL measures that include numeracy, graphical, and preference questions is essential when considering measuring HL. Existing numeracy or graphical measures may be too cumbersome for the clinical setting but have higher utility when categorizing those with high and low numerical skills, which may be of particular interest to those developing decision support tools. Developing a brief tool with both subjective and objective quantitative questions may augment HL numeracy measurement.

Abbreviations
GL: Graphical Literacy Measure; HL: Health Literacy; NAAL: National Assessment of Adult Literacy; SNS: Subjective Numeracy Scale; S-TOFHLA: Short Test of Functional Health Literacy in Adults; TOFHLA: Test of Functional Health Literacy in Adults; US: United States

Authors' contributions
AJH drafted this manuscript; AJH, LML, DSH, RJV developed the conceptual framework informing this study; VBL and GRK participated in data acquisition and management; AJH, LML, DSH, GRK, and RJV participated in data analysis and interpretation; All authors participated in essential critical appraisal and editing of the manuscript. All authors read and approved the final manuscript

Ethics approval and consent to participate
The Institutional Review Board of The University of Texas MD Anderson Cancer Center approved this protocol and participant informed written consent process.

Consent for publication
Not applicable.

Competing interests
The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Department of Health Services Research, Division of Cancer Prevention & Population Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd.Unit 1444, Houston, TX 77030, USA. [2]Department of

Housten *et al. BMC Public Health* (2018) 18:405

Page 8 of 9

Health Disparities Research, Division of Cancer Prevention & Population Sciences, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd.Unit 1440, Houston, TX 77030, USA. [3]School of Public Health, The University of Texas Health Science Center at Houston, 1200 Pressler St., Houston, TX 77030, USA.

References

1. US Department of Health and Human Services. Healthy people 2010. Washington, DC: US government printing office. Originally developed for Ratzan SC, Parker RM. introduction. In National Library of medicine current bibliographies in medicine: health literacy; 2000. Selden CR, Zorn M, Ratzan SC, Parker RM, eds. NLM pub. No. CBM 2000–1. Bethesda, MD: National Institutes of Health, U.S: Department of Health and Human Services; 2000.
2. Institute of Medicine. Health literacy: a prescription to end confusion. Washington, DC: National Academies Press; 2004.
3. Berkman ND, Davis TC, McCormack L. Health literacy: what is it? J Health Commun. 2010;15(Suppl 2):9–19. https://doi.org/10.1080/10810730.2010. 499985.
4. Kutner M, Greenburg E, Jin Y, Paulsen C. The health literacy of America's adults: results from the 2003 National Assessment of adult literacy (NCES 2006–483). U.S. Department of Education. Washington, DC: National Center for Education Statistics; 2006. https://nces.ed.gov/pubs2006/2006483.pdf. Accessed 1 Apr 2017
5. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011;155(2):97–107. https://doi.org/10.7326/0003-4819-155-2-201107190-00005.
6. Sayah FA, Qiu W, Johnson JA. Health literacy and health-related quality of life in adults with type 2 diabetes: a longitudinal study. Qual Life Res. 2016; 25(6):1487–94. https://doi.org/10.1007/s11136-015-1184-3.
7. Vernon JA, Trujillo A, Rosenbaum SJ, DeBuono B. Low health literacy: implications for national health policy. Department of Health Policy, School of Public Health and Health Services: The George Washington University; 2007. https://hsrc.himmelfarb.gwu.edu/cgi/viewcontent.cgi?referer= &httpsredir=1&article=1173&context=sphhs_policy_facpubs. Accessed 1 Apr 2017
8. Nutbeam D. Health literacy as a public health goal: a challenge for contemporary health education and communication strategies into the 21st century. Health Promot Int. 2000;15(3):259–67. https://doi.org/10.1093/ heapro/15.3.259.
9. Duell P, Wright D, Renzaho AMN, Bhattacharya D. Optimal health literacy measurement for the clinical setting: a systematic review. Patient Educ Couns. 2015;98(11):1295–307. https://doi.org/10.1016/j.pec.2015.04.003.
10. Baker DW, Williams MV, Parker RM, Gazmararian JA, Nurss J. Development of a brief test to measure functional health literacy. Patient Educ Couns. 1999; 38(1):33–42. https://doi.org/10.1016/S0738-3991(98)00116-5.
11. Sayah FA, Williams B, Johnson JA. Measuring health literacy in individuals with diabetes: a systematic review and evaluation of available measures. Health Educ Behav. 2012;40(1):42–55. https://doi.org/10.1177/1090198111436341.
12. Haun JN, Valerio MA, McCormack LA, Sørensen K, Paasche-Orlow MK. Health literacy measurement: an inventory and descriptive summary of 51 instruments. J Health Commun. 2014;19(Suppl 2):302–33. https://doi.org/10. 1080/10810730.2014.936571.
13. Rothman RL, Montori VM, Cherrington A, Pignone MP. Perspective: the role of numeracy in health care. J Health Commun. 2008;13(6):583–95. https://doi.org/10.1080/10810730802281791.
14. Frosch DL, Kaplan RM. Shared decision making in clinical medicine: past research and future directions. Am J Prev Med. 1999;17(4):285–94. https://doi.org/10.1016/S0749-3797(99)00097-5.
15. Rudd RE. Health literacy skills of US adults. Am J of Health Behav. 2007; 31(Suppl 1):S8–S18.
16. Zarcadoolas C, Pleasant A, Greer DS. Understanding health literacy: an expanded model. Health Promot Int. 2005;20(2):195–203. https://doi.org/10. 1093/heapro/dah609.
17. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the subjective numeracy scale. Med Decis Mak. 2007;27(5):672–80. https://doi.org/10.1177/ 0272989X07304449.
18. Lipkus IM, Samsa G, Rimer BK. General performance on a numeracy scale among highly educated samples. Med Decis Mak. 2001;21(1):37–44. https:// doi.org/10.1177/0272989X0102100105.
19. Nutbeam D. The evolving concept of health literacy. Soc Sci Med. 2008; 67(12):2072–8. https://doi.org/10.1016/j.socscimed.2008.09.050.
20. Schwartz LM, Woloshin S, Black WC, Welch HG. The role of numeracy in understanding the benefit of screening mammography. Ann Intern Med. 1997; 127(11):966–72. https://doi.org/10.7326/0003-4819-127-11-199712010-00003.
21. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the subjective numeracy scale: effects of low numeracy on comprehension of risk communications and utility elicitations. Med Decis Mak. 2007;27(5):663–71. https://doi.org/10.1177/0272989X07303824
22. Sørensen K, Van den Broucke S, Fullam J, et al. Health literacy and public health: a systematic review and integration of definitions and models. BMC Public Health. 2012;12(80) https://doi.org/10.1186/1471-2458-12-80.
23. Baker DW. The meaning and the measure of health literacy. J Intern Med. 2006;21:878–83. https://doi.org/10.1111/j.1525-1497.2006.00540.x.
24. Stewart DW, Vidrine JI, Shete S, et al. Health literacy, smoking, and health indicators in African American adults. J Health Commun. 2015;20(Suppl 2): 24–33. https://doi.org/10.1080/10810730.2015.1066465.
25. Apter AJ, Cheng J, Small D, et al. Asthma numeracy skill and health literacy. J Asthma. 2006;43(9):705–10. https://doi.org/10.1080/02770900600925585.
26. Haun J, Luther S, Dodd V, Donaldson P. Measurement variation across health literacy assessments: implications for assessment selection in research and practice. J Health Commun. 2012;17(Suppl 3):141–59. https:// doi.org/10.1080/10810730.2012.712615.
27. McNaughton C, Wallston KA, Rothman RL, Marcovitz DE, Storrow AB. Short, subjective measures of numeracy and general health literacy in an adult emergency department. Acad Emerg Med. 2011;18(11):1148–55. https://doi. org/10.1111/j.1553-2712.2011.01210.x.
28. Macek MD, Haynes D, Wells W, Bauer-Leffler S, Cotten PA, Parker RM. Measuring conceptual health knowledge in the context of oral health literacy: preliminary results. J Public Health Dent. 2010;70(3):197–204. https:// doi.org/10.1111/j.1752-7325.2010.00165.x.
29. Weld KK, Padden D, Ricciardi R, Garmon Bibb SC. Health literacy rates in a sample of active duty military personnel. Mil Med. 2009;174(11):1137–43.
30. Chew LD, Griffin JM, Partin MR, et al. Validation of screening questions for limited health literacy in a large VA outpatient population. J Gen Intern Med. 2008;23(5):561–6. https://doi.org/10.1007/s11606-008-0520-5.
31. Wolf MS, Feinglass J, Thompson J, Baker DW. In search of 'low health literacy': threshold vs. gradient effect of literacy on health status and mortality. Soc Sci Med. 2010;70(9):1335–41. https://doi.org/10.1016/j. socscimed.2009.12.013.
32. Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults. J Gen Intern Med. 1995;10:537–41. https://doi.org/10.1007/BF02640361.
33. Anderson BL, Obrecht NA, Chapman GB, Driscoll DA, Schulkin J. Physicians' communication of down syndrome screening test results: the influence of physician numeracy. Genet Med. 2011;13(8):744–9. https://doi.org/10.1097/ GIM.0b013e31821a370f.
34. Galesic M, Garcia-Retamero R. Statistical numeracy for health: a cross-cultural comparison with probabilistic national samples. Arch Intern Med. 2010; 170(5):462–8. https://doi.org/10.1001/archinternmed.2009.481.
35. Hanoch Y, Miron-Shatz T, Himmelstein M. Genetic testing and risk interpretation: how do women understand lifetime risk results? Judgm Decis Mak. 2010;5(2):116.
36. Hawley ST, Zikmund-Fisher B, Ubel P, Jancovic A, Lucas T, Fagerlin A. The impact of the format of graphical presentation on health-related knowledge and treatment choices. Patient Edu Couns. 2008;73(3):448–55. https://doi. org/10.1016/j.pec.2008.07.023.
37. Galesic M, Garcia-Retamero R. Graph literacy a cross-cultural comparison. Med Decis Mak. 2011;31(3):444–57. https://doi.org/10.1177/ 0272989X10373805.
38. Lipkus I, Hollands J. The visual communication of risk. J Natl Cancer Inst Monogr. 1999;25:149–63. https://doi.org/10.1093/oxfordjournals. jncimonographs.a024191.
39. Lipkus IM. Numeric, verbal, and visual formats of conveying health risk: suggested best practices and future recommendations. Med Decis Mak. 2007;27(5):696–713. https://doi.org/10.1177/0272989X07307271.
40. Ancker JS, Senathirajah Y, Kukafka R, Starren JB. Design features of graphs in health risk communication: a systematic review. J Am Med Inform Assoc. 2006;13(6):608–18. https://doi.org/10.1197/jamia.M2115.

Housten *et al. BMC Public Health* (2018) 18:405

Page 9 of 9

41. Friel SN, Curcio FR, Bright GW. Making sense of graphs: critical factors influencing comprehension and instructional implications. J Res Math Educ. 2001;32(2):124–58. https://doi.org/10.2307/749671.

42. Gaissmaier W, Wegwarth O, Skopec D, Muller AS, Broschinski S, Politi MC. Numbers can be worth a thousand pictures: individual differences in understanding graphical and numerical representations of health-related information. Health Psychol. 2012;31(3):286–96. https://doi.org/10.1037/a0024850.

43. Okan Y, Garcia-Retamero R, Cokely ET, Maldonado A. Individual differences in graph literacy: overcoming denominator neglect in risk comprehension. J Behav Decis Mak. 2012;25(4):390–401. https://doi.org/10.1002/bdm.751.

44. Garcia-Retamero R, Galesic M. Who profits from visual aids: overcoming challenges in people's understanding of risks. Soc Sci Med. 2010;70(7):1019–25. https://doi.org/10.1016/j.socscimed.2009.11.031.

45. Peters E, Västfjäll D, Slovic P, Mertz CK, Mazzocco K, Dickert S. Numeracy and decision making. Psychol Sci. 2006;17(5):407–13. https://doi.org/10.1111/j.1467-9280.2006.01720.x.

46. Betts JR, Grogger J. The impact of grading standards on student achievement, educational attainment, and entry-level earnings. Econ Educ Rev. 2003;22(4):343–52. https://doi.org/10.1016/S0272-7757(02)00059-6.

47. Stumpf H, Stanley JC. Group data on high school grade point averages and scores on academic aptitude tests as predictors of institutional graduation rates. Educ Psychol Meas. 2002;62(6):1042–52. https://doi.org/10.1177/0013164402238091.

48. Cohen J. Statistical power analysis for the behavioral sciences. L. Erlbaum Associates: Hillsdale; 1988.

49. Jordan JE, Osborne RH, Buchbinder R. Critical appraisal of health literacy indices revealed variable underlying constructs, narrow content and psychometric weaknesses. J Clin Epidemiol. 2011;64(4):366–79. https://doi.org/10.1016/j.jclinepi.2010.04.005.

50. Koh HK, Brach C, Harris LM, Parchman ML. A proposed 'health literate care model' would constitute a systems approach to improving patients' engagement in care. Health Aff. 2013;32(2):357–67. https://doi.org/10.1377/hlthaff.2012.1205.

51. DeWalt DA, Broucksou KA, Hawk V, et al. Developing and testing the health literacy universal precautions toolkit. Nurs Outlook. 2011;59(2):85–94. https://doi.org/10.1016/j.outlook.2010.12.002.

52. Koh HK, Berwick DM, Clancy CM, et al. New federal policy initiatives to boost health literacy can help the nation move beyond the cycle of costly 'crisis care'. Health Aff. 2012;31(2): 434–443; doi :https://doi.org/10.1377/hlthaff.2011.1169.

53. McCormack L, Thomas V, Lewis MA, Rudd R. Improving low health literacy and patient engagement: a social ecological approach. Patient Educ Couns. 2017;100(1):8–13. https://doi.org/10.1016/j.pec.2016.07.007.

54. Griffin JM, Partin MR, Noorbaloochi S, et al. Variation in estimates of limited health literacy by assessment instruments and non-response bias. J Gen Intern Med. 2010;25(7):675-681. doi: https://doi.org/10.1007/s11606-010-1304-2.