



Published in final edited form as:

*Acad Emerg Med.* 2014 February ; 21(2): 137–146. doi:10.1111/acem.12315.

## Feasibility and Diagnostic Accuracy of Brief Health Literacy and Numeracy Screening Instruments in an Urban Emergency Department

Christopher R. Carpenter, MD, MSc, Kimberly A. Kaphingst, ScD, Melody S. Goodman, PhD, Margaret J. Lin, MD, Andrew T. Melson, and Richard T. Griffey, MD, MPH

Division of Emergency Medicine, Barnes-Jewish Hospital, Washington University School of Medicine, (CRC, RTG) St. Louis, MO; Division of Public Health Sciences, Department of Surgery, Washington University in St. Louis, School of Medicine, (KAK, MSG) St. Louis, MO; Washington University Institute for Public Health, (KAK, RTG) St. Louis MO; Department of Emergency Medicine, Beth Israel Deaconess Medical Center, Harvard Affiliated Emergency Medicine Residency, (MJL) Boston, MA; Saint Louis University, School of Medicine, (ATM) St. Louis, MO

### Abstract

**Objectives**—To evaluate the diagnostic accuracy of five health literacy screening instruments in emergency department (ED) patients: the Rapid Evaluation of Adult Literacy in Medicine-Revised (REALM-R), the Newest Vital Sign (NVS), Single Item Literacy Screens (SILS), health numeracy, and physician gestalt. A secondary objective was to evaluate the feasibility of these instruments as measured by administration time, time on task, and interruptions during test administration.

**Methods**—This was a prospective observational cross-sectional study of a convenience sampling of adult patients presenting during March 2011 and February 2012 to one urban university-affiliated ED. Subjects were consenting non-critically ill, English-speaking patients over the age of 18 years without aphasia, dementia, mental retardation, or inability to communicate. The diagnostic test characteristics of the REALM-R, NVS, SILS, health numeracy, and physician gestalt were quantitatively assessed by using the short Test of Functional Health Literacy in Adults (S-TOHFLA). A score of 22 or less was the criterion standard for limited health literacy (LHL).

**Results**—Four hundred thirty-five participants were enrolled, with mean age of 45 years (SD  $\pm 15.7$  years) and 18% had less than a high school education. As defined by an S-TOHFLA score of 22 or less, the prevalence of LHL was 23.9%. In contrast, the NVS, REALM-R, and physician gestalt identified 64.8%, 48.5%, and 35% of participants as LHL, respectively. A normal NVS screen was the most useful test to exclude LHL, with a negative likelihood ratio of 0.04 (95% CI = 0.01 to 0.17). When abnormal, none of the screening instruments, including physician gestalt,

---

Corresponding author contact information: Christopher R. Carpenter, MD, MS, Campus Box 8072, 660 S. Euclid Avenue, St. Louis MO 63110, Telephone: (314) 362-7979; FAX: (314) 362-0419, carpenterc@wusm.wustl.edu.

**Presentations:** Portions of this manuscript were presented at the 2011 American College of Emergency Physicians Scientific Assembly Research Forum (San Francisco, CA) and the 2012 Society for Academic Emergency Medicine Annual Meeting (Chicago IL).

significantly increased the post-test probability of LHL. The NVS and REALM-R require 3 and 5 minutes less time to administer than the S-TOHFLA. Administration of the REALM-R is associated with less test interruptions.

**Conclusions**—One-quarter of these ED patients had marginal or inadequate health literacy. Among the brief screening instruments evaluated, a normal Newest Vital Sign result accurately reduced the probability of limited health literacy, although it will identify two-thirds of ED patients as high-risk for limited health literacy. None of the brief screening instruments significantly increases the probability of limited health literacy when abnormal.

---

## INTRODUCTION

Health literacy is defined by the Institute of Medicine (IOM) as “the degree to which individuals can obtain, process, and understand basic health information and services needed to make appropriate health decisions.”<sup>1</sup> Limited health literacy (LHL) is widely recognized as a major determinant of important health outcomes, and is estimated by the IOM to cost \$73 billion annually.<sup>2</sup> According to the American Medical Association, poor health literacy is “a stronger predictor of a person’s health than age, income, employment status, education level, and race.”<sup>3-6</sup> Inadequate health literacy is associated with poorer health status, less knowledge about chronic disease self-management, lower rates of medication adherence, and higher rates of acute health care utilization in patients with chronic diseases, as well as with increased hospitalization rates and mortality.<sup>1,5,7-9</sup> The IOM has ranked addressing LHL among the first quartile of research priorities. Estimates of the number of adults with LHL vary greatly depending upon the population and screening instrument. The best available estimate from a nationally representative sample is that nearly half of all American adults can be categorized as having LHL,<sup>10</sup> although this estimate was generated using an instrument that is not publicly available. Other estimates of health literacy in patient populations use various publicly available screening instruments.<sup>11-17</sup>

The effect of health literacy is due to an interaction between patients’ health literacy skills and the demands that the health context, such as the emergency department (ED), place on these skills.<sup>1</sup> Discrepancies observed between patients’ reading abilities and ED discharge materials were noted in the emergency medicine literature nearly 20 years ago.<sup>18-20</sup> Recently, the importance of assessing patients’ health literacy skills in the ED has garnered increasing attention, with estimates of the prevalence of LHL ranging from 10.5% to 88% depending on the screening instrument used and the geographic locale.<sup>11-17</sup> Adequate health literacy can potentially influence multiple aspects of ED care, including the ability to comprehend and incorporate verbal information provided by health care personnel, understand numeric risk information, comprehend written materials and forms, give informed consent, access health care services, and adhere to post-discharge follow-up and medication recommendations.<sup>1</sup> Despite recent interest in this area, there is a paucity of ED-based research investigating the association between LHL and health care outcomes.<sup>17</sup>

A major barrier to high-quality studies of health literacy in the ED is the lack of health literacy measures validated for use in busy clinical settings like the ED. Most studies evaluating the diagnostic accuracy of screening instruments have not been ED-based, but

rather have been performed in controlled clinical research environments where factors often accompanying ED visits, such as anxiety, time pressures, acute illness, noise, and interruptions are not present.<sup>21-32</sup> It is known that a tool that has been validated for use in one environment may not be valid in others.<sup>33,34</sup>

A related gap in knowledge and potential barrier to translation of research findings into clinical practice relates to the feasibility of use of health literacy instruments for routine screening in the ED. We are not aware of any published data regarding the feasibility of use of these instruments for screening in the ED.

The primary objectives of this study were: 1) to evaluate and compare the diagnostic accuracy of five brief objective and subjective screening instruments for LHL in urban ED patients, and 2) to evaluate the feasibility of these instruments based upon time burden of administration and number of interruptions relative to performance. A secondary objective of this study was to assess the accuracy of emergency physician (EP) gestalt for LHL.

## **METHODS**

### **Study Design**

This was a prospective, cross-sectional study using convenience sampling. We sought to minimize design-related bias by using the Standards for Reporting of Diagnostic Accuracy (STARD) criteria, which consist of a 25-item checklist that serves to ensure consistent and reproducible quality in the design, conduct, and reporting of diagnostic trials.<sup>35,36</sup> Hospital institutional review board approval was granted and written informed consent was obtained from all subjects prior to enrollment.

### **Study Setting and Population**

This study was conducted in an urban academic ED with over 97,000 total annual visits. Trained research assistants recruited study participants from the ED between March 1, 2011 and February 29, 2012 at different times of the day on different days of the week. All ED patients aged at least 18 years were prospectively identified for enrollment by review of the electronic medical record dashboard by research assistants. Exclusion criteria included undue patient distress as judged by the attending physician, altered mental status, aphasia, mental handicap, previously diagnosed dementia or insurmountable communication barrier as judged by family or the screener, non-English speaking, sexual assault victims, acute psychiatric illness, or corrected visual acuity worse than 20/100 using both eyes. Exclusion criteria were meant to minimize confounding variables that could bias estimates of diagnostic accuracy of health literacy screening instruments.

### **Study Protocol**

Deidentified age, race, and sex data were recorded for patients declining to participate for comparison with consenting patients. For each eligible and consenting patient, a research assistant administered all health literacy tests during the patient's ED visit. All screeners received standardized training on the administration of screening instruments, consisting of an in-person presentation, review of a pre-recorded training presentation, practice sessions

administering the instruments to mock patients, and monitored screening of the first patient. The screeners were paid research assistants and medical students. Training included specific instruction to avoid language that might lead to feelings of shame or embarrassment among patients with LHL.<sup>37</sup> Screeners read participants standardized instructions for each of the health literacy screening instruments and recorded the start and stop times for all screening tests other than the single-item literacy tests (SILS). Instructions were given that screening should not interfere with patient care. Stop and restart times were also recorded for interruptions that occurred during screening tests such as nursing care, laboratory or imaging studies, etc. When present during screening, family members and friends were asked not to assist with responses. The order of presentation of screening instruments was varied based on even or odd days of the week, aimed at altering whether the Short Test of Functional Health Literacy in Adults (S-TOFHLA), which was the lengthiest test, was presented first or last. The screeners discretely recorded all participant responses. Demographic data elements were collected during the interview and from the electronic medical record. Attending EPs were not informed of the health literacy-screening test results. After enrollment of the first 112 patients, the study protocol was amended to solicit from either the resident or attending treating physician, their subjective estimate of the patient's level of health literacy in advance of testing (see the data collection instrument in Data Supplement 1). This change did not alter the patient component of the protocol. All data were entered into a Microsoft Access database (Microsoft Corp., Redmond, WA).

**Screening Tools**—The screening tools administered included the abbreviated S-TOFHLA,<sup>23,24</sup> which was used as the criterion standard; the Rapid Evaluation of Adult Literacy in Medicine – Revised (REALM-R);<sup>25</sup> the Newest Vital Sign (NVS);<sup>27</sup> and three SILS<sup>29</sup> that were considered as single items, and as an index. The thresholds for LHL for the S-TOFHLA,<sup>23,24</sup> REALM-R,<sup>21,22</sup> and NVS<sup>27,28</sup> are based on published scoring rules for the instruments. The cutoff for the SILS was not previously established in the literature. We derived the threshold for the SILS based on the value that optimized both sensitivity and specificity in our population. We also evaluated physician gestalt. We describe each of these screening instruments in detail in Data Supplement 2. In addition, Data Supplement 3 illustrates the REALM-R and Data Supplement 4 demonstrates the NVS.

## Outcome Measures

The criterion standard for LHL was an abbreviated S-TOFHLA score  $\geq 22$ , which is the sum of inadequate health literacy (S-TOFHLA score 0 to 16) and marginal health literacy (S-TOFHLA score 17 to 22). The primary outcome measures in the evaluation of diagnostic accuracy consisted of 1) test characteristics for each instrument, including positive likelihood ratios (LR+) and negative likelihood ratios (LR-) compared with the S-TOFHLA criterion standard, and 2) the correlation of each test with the S-TOFHLA. The primary outcome measures in the evaluation of feasibility were 1) the time on average that it took to administer the “time on test,” (i.e., total time minus time during interruptions), and 2) the number and duration of interruptions during administration of the tests.

Depending on the goals of LHL detection, it may be alternately preferable to use tests that optimize sensitivity versus specificity. Because of this, in determining which tests are the

most feasible, we defined feasibility based on maximizing measures of ease and efficiency (time on test and interruptions) while achieving the best performance, using LR+ (when the emphasis is on identifying LHL) and the LR- (when the emphasis is on excluding LHL). Although the risk of interruptions naturally increases with an increasing time of test administration, we considered interruptions in addition to time, because interruptions are felt to lead to worse task performance that could affect accuracy.<sup>38</sup>

## Data Analysis

Analysis was conducted according to STARD criteria with SAS (version 9.2, SAS Institute, Cary, NC).<sup>36</sup> Demographic data were compared between excluded and enrolled patient groups using t-tests for normally distributed continuous variables, and chi-square for categorical variables. Each instrument was dichotomized with commonly used cutoffs in order to report LR+ and LR- to interpret clinically with individual patients. A summary SILS score was computed by assigning a value of “5” for the response indicating the least difficulty using or interpreting medical information independently, and a value of “1” for the response indicating the most difficulty. For example, for the first SILS question inquiring how often patients have somebody else read hospital materials, an “always” response was scored a “1” and a “never” response was scored a “5.” The summed SILS score ranged from 3 to 15. The SILS demonstrated reasonable internal consistency reliability. The standardized Cronbach’s alpha for the three-item SILS measure was 0.78.

Standard operating characteristics of diagnostic tests were computed for each health literacy screening instrument, including sensitivity, specificity, LRs, receiver operating characteristic (ROC) curves with area under the ROC curve (AUC), and 95% confidence intervals (CI). We compared the AUCs for statistically significant differences.<sup>39</sup> To evaluate the validity of each instrument, we examined Spearman’s correlation coefficient against the abbreviated S-TOFHLA. Verifying a sensitivity of 90% with 5% range of error, and a baseline prevalence of LHL of 40%,<sup>17</sup> would require 346 subjects to be enrolled with complete data collection. Assuming 19% of subjects would have incomplete data collection based on our previous experience, we planned to enroll 430 subjects.<sup>40</sup>

Time was analyzed in increments of whole minutes. Times of less than one minute were rounded up to the nearest whole minute; this was felt to best reflect the level of precision reliably achievable. Mean and median total test time and interruption data were calculated for each individual screening tool.

## RESULTS

We approached 588 patients, excluded 142, and enrolled 446 patients (Figure 1). We excluded from analysis 11 patients who did not complete the abbreviated S-TOFHLA, leaving 435 subjects in this analysis. The mean age of patients was 45 years (SD  $\pm$ 15.7 years), and 18% had less than a high school education. Additional demographics are summarized in Table 1. The age, sex, and race of enrolled patients did not differ significantly from those of excluded patients or the 93,476 total patient visits to the ED in 2011.

## Diagnostic Accuracy

As summarized in Table 2, 23.9% of patients had LHL as identified by the S-TOFHFLA (13.1% with inadequate and 10.8% with marginal functional health literacy). The REALM-R identified 48.5% of patients as having LHL, and the NVS identified 64.8% as having LHL (29.7% with a high likelihood of limited literacy, and another 35.1% as possible limited literacy). Physician gestalt identified 35.0% of patients as having LHL, including 8.1% with inadequate health literacy and an additional 26.9% with marginal health literacy.

Table 3 demonstrates the relative performance of the screening tools tested. The NVS had the best LR- to exclude LHL at 0.04 (95% CI = 0.01 to 0.17). The LR+ estimates ranged from 1.61 to 2.78. Although the summed SILS scores at a cutoff of 12 only had a LR+ 2.78 (95% CI = 1.92 to 4.02), some individual SILS question responses would significantly increase the probability of LHL. Specifically, a response of “always” need help reading (LR + 7.42, 95% CI = 1.13 to 48.89), “not at all” on confidence with medical forms (LR+ 9.93, 95% CI = 1.9 to 52.03), and “poor” on ability to read (LR+ 16.56, 95% CI = 2.01 to 136.46) would each significantly increase the probability of LHL in an individual patient. Two-by-two contingency tables for each instrument are reproduced in Data Supplement 5. Unaided EP impression of health literacy was inadequate to significantly increase or decrease the probability of LHL. AUCs are presented in Figure 2, demonstrating that the distributions for NVS and REALM-R were widely overlapping.

The NVS and REALM-R had the highest correlations with the abbreviated S-TOFHFLA (Spearman correlation 0.602 and 0.540 respectively,  $p < 0.001$ ). The SILS (correlation range 0.239-0.485,  $p < 0.001$ ) and physician gestalt (correlation 0.285,  $p < 0.001$ ) were weakly correlated with the S-TOFHFLA.

## Feasibility

The S-TOFHFLA had the longest time administration burden (time on test), followed by the NVS, which took on average 2.9 minutes less than the S-TOFHFLA, and then the REALM-R, which took on average 5.0 minutes less than the S-TOFHFLA and almost a third of the time of the NVS (Table 4). The REALM-R had the fewest interruptions, with 0.5% of tests being interrupted compared to the NVS and S-TOFHFLA, which were interrupted 6.0% and 13.1% of the time, respectively. On average, patients with lower health literacy levels took longer to complete tests than those with adequate health literacy levels.

## DISCUSSION

Limited health literacy is a relatively silent epidemic in today’s ED, threatening patient safety and with profound implications including access to quality care.<sup>41,42</sup> LHL also potentially affects informed consent, prudent layperson policies, shared decision-making, and the consideration of results of prior studies that did not stratify by health literacy, among other areas. Examining levels of health literacy remains a challenge for researchers, policy-makers, and emergency providers. Screening tools developed for use in clinical settings have been studied in the ED, but none of these studies measured health literacy using an instrument validated to do so.<sup>18-20,43-46</sup> For any instrument used routinely in the ED setting,

rather than just for research, simplicity and efficiency in training and administration are critical for adoption and reliability, particularly as these are included alongside other regulatory screening demands and patient care needs.<sup>47</sup> We envision a test that can be administered routinely to quickly and accurately assess a patient's health literacy level without disrupting patient flow. It should be noted that routine screening in clinical settings is an area of current controversy among health literacy experts, in that clinical environments are presently not geared to tailor therapies or interventions based on determination of health literacy level.<sup>48</sup>

Though feasibility of health literacy screening tools has not been studied in the ED, validation studies in the ED are not entirely missing. Recently, McNaughton et al. published an ED-based observational investigation assessing the diagnostic accuracy of the REALM-R, S-TOFHLA, Wide Range Achievement Test 4, and subjective literacy screen questions, reporting on the validity of these instruments.<sup>32</sup> McNaughton et al. provided a needed first step in assessment of these tools in the ED. However, the study did not report on sensitivity, specificity, or LRs, and did not use the STARD criteria in the conduct or reporting of the study, threatening its internal validity.<sup>36</sup> That study's primarily white demographic also limits the external validity of its findings to similar populations until additional validation trials are available. Compared to that of McNaughton et al., our population had a higher proportion of LHL, and the ROC AUC for the REALM-R was higher in our study (0.80 versus 0.72). Our population sample had a lower education level and was more ethnically diverse, reflective of many urban ED settings. In our sample, we obtained similar estimates of internal reliability (i.e. Cronbach's alpha) to that of McNaughton et al. for the screening instruments evaluated with significantly higher correlations.

Using the S-TOFHLA as the criterion standard, our results demonstrate a prevalence of LHL in our population varying from approximately 24% to 65% depending on the screening test used. The wide range of estimates of LHL between instruments is probably related to the specific health literacy domains tested by each instrument.<sup>49</sup> The NVS resulted in the highest prevalence estimate of LHL. This tool tests the health literacy domains of document literacy and numeracy in addition to print literacy, while the REALM-R and S-TOFHLA are more based on word pronunciation, reading comprehension, and print literacy.<sup>50</sup> Based on these differential performance results, health literacy tasks that require basic numeracy skills may therefore be particularly problematic for this patient population, suggesting the possible importance of screening for limited numeracy.<sup>51</sup> Furthermore, the abbreviated S-TOFHLA is an imperfect criterion standard. It does not assess all domains of the health literacy construct, was originally validated on the population of one clinic, and did not use available published diagnostic accuracy methods.<sup>24,36,49,50</sup> Nonetheless, it is perhaps the most widely accepted criterion standard for distinguishing LHL from adequate health literacy, commonly used in studies of this nature, and we are not aware of a better approach for assessment that is publicly available.<sup>31</sup> In our study, however, the abbreviated S-TOFHLA was also the least feasible screening tool for routine use, taking on average nearly six minutes to administer and incurring the most interruptions.

The NVS was highly correlated with the abbreviated S-TOFHLA, and was the most accurate screening instrument to rule out LHL with a LR- of 0.04. The NVS did not fare as well on

feasibility, with an average “time on test” of 3.13 minutes. Incidentally, our data provide the first estimates of sensitivity and specificity for the NVS in an ED population. When stratified by health literacy level, for both the S-TOFHLA and the NVS, patients in the lowest categories had times on test that were shorter than those in the marginal or possible categories were. It is not clear if this might be because those in the lowest categories provided answers quickly without as much deliberation.

The REALM-R was highly correlated with the abbreviated S-TOFHLA, but using the S-TOFHLA as the criterion standard identified almost half of patients as LHL. The REALM-R LR+ of 2.14 (95% CI = 1.64 to 2.80) is among the better values but still fails to achieve a significant discriminatory level. The LR- of 0.30 for the REALM-R (95% CI = 0.19 to 0.46) is also respectable. Although we did not specifically study training or test for ease of use, on its face, the REALM-R is simpler to administer compared to the S-TOFHLA and NVS. The REALM-R had the shortest time burden and fewest interruptions of the tests for which we measured time of administration, while still achieving excellent sensitivity for identifying LHL, making it the most feasible of these tests.

The ability to exclude LHL with the composite SILS, using a threshold score of 12, was roughly equivalent to that for the individual SILS items and the REALM-R, with a LR- of 0.42. The summed SILS score had a clinically inconsequential, but superior LR+ (2.78; 95% CI = 1.92 to 4.02), although the CIs were overlapping with the LR+ confidence intervals for the other screening instruments. The single question “How would you rate your ability to read?” performed competitively with the three other screening tools in identifying LHL. Although we did not measure time for completion of the SILS, anecdotally the single questions consistently yielded patient response times less than one minute.

It is known that physicians in non-ED settings overestimate their patients’ health literacy,<sup>52,53</sup> but this has never been assessed among EPs. Our findings were consistent with an accumulating body of evidence that physician gestalt is an inaccurate estimate for LHL.<sup>52,53</sup> Physician gestalt cannot exclude inadequate health literacy (LR- of 0.70, 95% CI = 0.55 to 0.89), and physician gestalt for identifying LHL had a LR+ of 1.72 (95% CI = 1.20 to 2.47), which is in the same range as the tested health literacy screening instruments except for some of the SILS. However, using the guide that LR+ exceeding 10 are optimal diagnostic or screening tests, none of the screening instruments, including physician gestalt, is sufficiently accurate to diagnose LHL without further testing.<sup>54</sup> Another interpretation of our data is that physician gestalt is as accurate as any of the currently available validated health literacy tests and does not require additional time for testing. The development of a more comprehensive measure of health literacy that is feasible for administration in clinical settings would further advance this area of inquiry.

Several high-yield research opportunities for assessing and intervening on health literacy exist. One important construct is understanding the optimal personnel, situation, and time during the ED course of care in which health literacy screening should occur. Another important question for future research is whether screening patients for LHL allows clinicians to target interventions and improve post-ED health outcomes. Another research question is understanding the most meaningful definition of LHL among ED populations.



Less resource-intensive and inexpensive interventions justify the broadest definition of LHL, whereas more resource-intensive and expensive interventions favor a more narrow definition of LHL.

Interventions such as “teachback” for discharge instructions have been proposed in the context of “universal precautions” for LHL. In this case, screening is foregone and the intervention is applied to all.<sup>55</sup> Although few would argue that discharge instructions should be more clear and understandable, there are little data evaluating the effectiveness, reliability, or time effect of teachback, particularly as a ‘universal precaution.’<sup>56,57</sup> Moreover, discharge instructions are but one area where identifying those with LHL may be important. Those interventions that are more costly and resource-intensive, such as individualized counseling, or use of discharge materials that may benefit from additional instruction, such as pictorial representations, may warrant a more restricted selection for those found to be at highest risk. Research is needed to examine whether different types of interventions targeting patients with LHL, defined broadly or narrowly, improve comprehension of information, and ultimately, outcomes for ED patients. Assuming that screening is felt to be of value, use of a validated instrument is of particular importance to insure that a construct one desires to measure is in fact addressed by the tool being used, and that this can be employed reliably.

## LIMITATIONS

This was a single-center observational trial and excluded several groups of previously described patients, limiting the external validity in dissimilar populations. Specifically, we cannot extrapolate the estimates of diagnostic accuracy for the health literacy screening instruments to patients with undue distress, sexual assault victims, acute psychiatric illness, altered mental status, aphasia, mental or visual handicap, dementia, non-English speaking individuals, or those with communication barriers. We did not assess these populations and each instrument could be more or less accurate in these groups. Although we selected the most commonly used tools to evaluate, there are several other brief screening instruments in lesser use that we did not evaluate, including the Medical Term Recognition Test.<sup>30</sup> Even though we did not assess inter- or intra-rater reliability of these screening instruments, the ability to do so when the interval between tests is short as it would be in ED settings is largely limited by test-retest phenomena.<sup>58</sup> To our knowledge, the reliability of the health literacy screening instruments has yet to be described in any research or clinical setting. Although convenience sampling includes the potential for selection bias and spectrum bias, which can falsely increase sensitivity and specificity,<sup>59</sup> nearly all ED-based health literacy studies, including the only other validation study, have used convenience sampling. Ours is the only study that compares enrolled and declining patients along with the general ED population, demonstrating lack of differences among these populations in terms of basic demographics.

Contrary to STARD recommendations to blind the outcome assessors to the new test(s) being evaluated and vice versa, our screeners collected both the index tests and the abbreviated S-TOFHLA criterion standard. This increases the risk for incorporation bias, which can also falsely increase sensitivity and specificity.<sup>59,60</sup> However, both the index tests

and the criterion standard were not scored real-time. Instead, responses were entered into a database that then computed the score, which was used to provide the health literacy category for each patient. These methods minimize but do not completely eliminate the potential for incorporation bias.

## CONCLUSIONS

In an urban ED population, the prevalence of patients with limited health literacy as determined by the abbreviated S-TOFHLA was 23.9%, and ranged as high as 64.8% when using the NVS. When compared to this criterion standard, the NVS performed best at excluding limited health literacy. However, none of the instruments that we evaluated significantly increased the post-test probability of limited health literacy. The S-TOHFLA and the NVS are less feasible for use in the ED, taking on average approximately six and three minutes respectively to administer, and incurring frequent interruptions. The REALM-R was the most feasible of the tools for which we measured administration time, and this test performed reasonably well in reducing the probability of limited health literacy. If the SILS questions can be administered in approximately one minute, then this is the most feasible of the instruments with the best performance for identifying lower health literacy. Physician gestalt does not accurately identify or exclude limited health literacy. Selection of the optimal screening tool should consider these diagnostic test characteristics, including likelihood ratios, in conjunction with the goals and objectives that are intended by limited health literacy screening efforts.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

The authors would like to acknowledge the assistance of the research and screening staff: Meng-Ru Cheng, Sarah Lyons, Ralph O'Neil, Emma Dwyer, Ian Ferguson, Mallory Jorif, Matthew Kemperwas, Jasmine Lewis, Darain Mitchell, and John Schneider.

### DISCLOSURES:

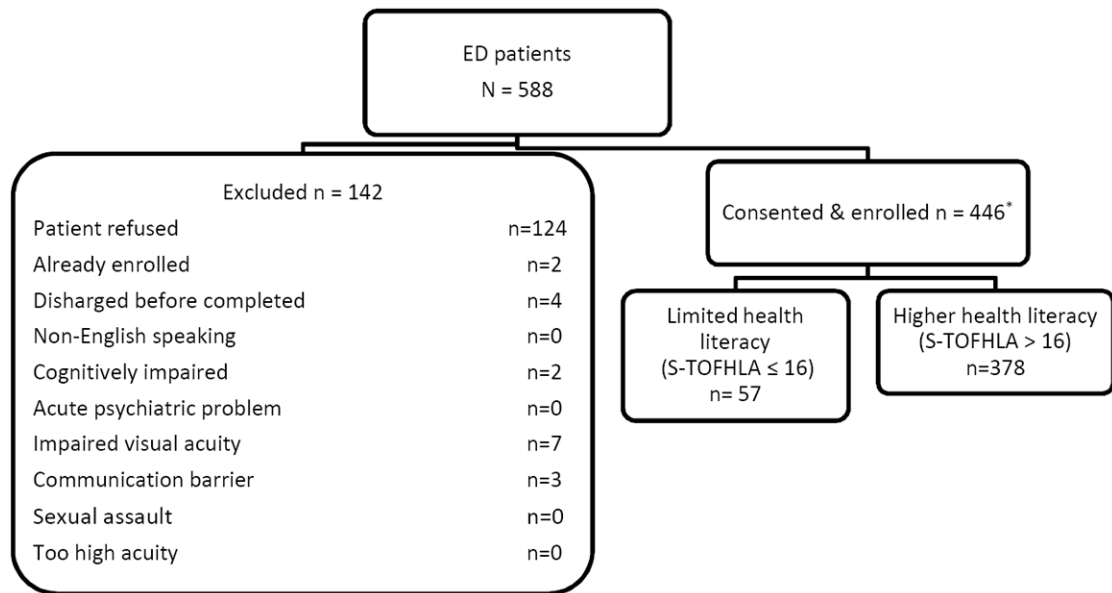
Dr. Carpenter was supported by an institutional KM1 Comparative Effectiveness Award (1KM1CA156708-01). Dr. Kaphingst is supported by R21 HS020309 from the Agency for Healthcare Research and Quality, P50 CA95815 and P30 DK092950 from the National Institutes of Health, U58 DP0003435 from the Centers for Disease Control and Prevention and R01 CA168608 and 3U54CA153460-03S1 from NCI. Dr. Goodman is supported by funding from the Barnes-Jewish Hospital Foundation. Dr. Griffey is supported by an institutional KM1 Comparative Effectiveness Award Number KM1CA156708 through the National Cancer Institute (NCI) at the National Institutes of Health (NIH) and Grant Numbers UL1 RR024992, KL2 RR024994, TL1 RR024995 through The Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources and the *National Center for Advancing Translational Sciences* at the National Institutes of Health. Dr. Griffey is also supported through the Emergency Medicine Foundation/ Emergency Medicine Patient Safety Foundation Patient Safety Fellowship. Andrew Melson was Supported by the Clinical and Translational Science Award (CTSA) program of the National Center for Research Resources (NCRR) at the National Institutes of Health (NIH) Grant Numbers UL1 RR024992, TL1 RR024995. The content is solely the responsibility of the authors and does not necessarily represent the official views of the supporting societies and foundations or the funding agencies. Neither Dr. Carpenter nor Dr. Griffey, both of whom are associate editors for this journal, had a role in the peer review or publication decision for this paper.

## References

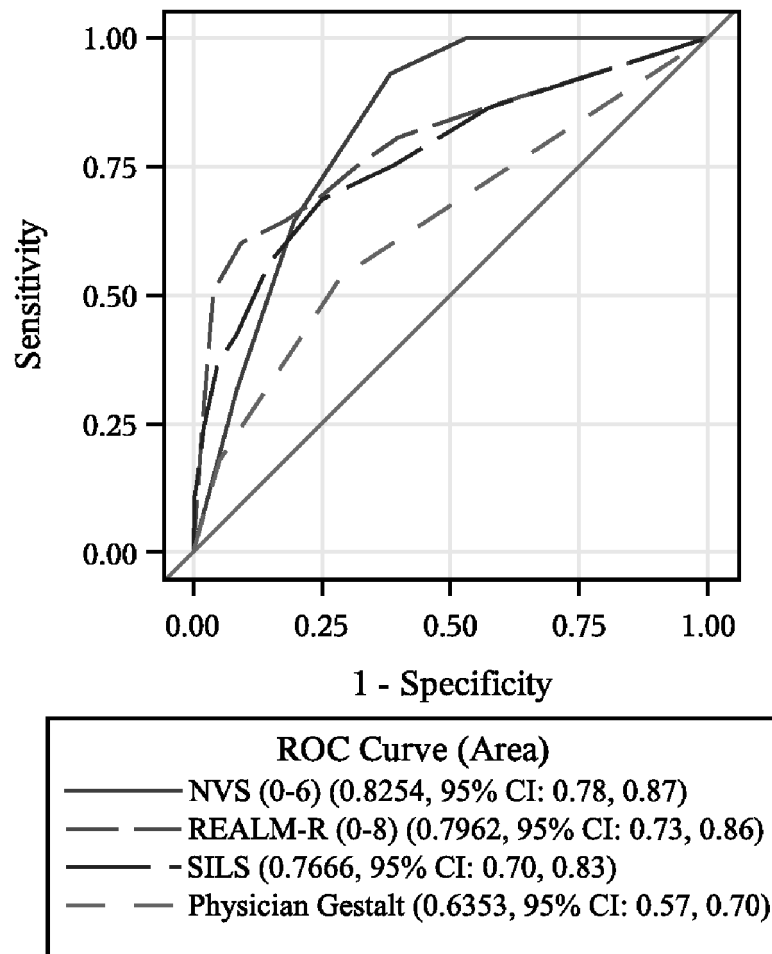
1. Nielson-Bohlman, L.; Panzer, A.; Kindig, D. Health Literacy: A Prescription to End Confusion. Washington DC: National Academies Press; 2004.
2. American Medical Association. Health literacy: report of the Council on Scientific Affairs. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA. 1999; 281:552–7. [PubMed: 10022112]
3. Weiss BD, Hart G, McGee DL, D'Estelle S. Health status of illiterate adults: relation between literacy and health status among persons with low literacy skills. J Am Board Fam Pract. 1992; 5:257–64. [PubMed: 1580173]
4. Baker DW, Parker RM, Williams MV, Clark WS, Nurss JR. The relationship of patient reading ability to self-reported health and use of health services. Am J Public Health. 1997; 87:1027–30. [PubMed: 9224190]
5. Sudore RL, Yaffe K, Satterfield S, et al. Limited literacy and mortality in the elderly: the health, aging, and body composition study. J Gen Intern Med. 2006; 21:806–12. [PubMed: 16881938]
6. Weiss, BD. Manual for Clinicians. 2. Chicago IL: American Medical Association; 2007. Health Literacy and Patient Safety: Help Patients Understand.
7. Wolf MS, Gazmararian JA, Baker DW. Health literacy and functional status among older adults. Arch Intern Med. 2005; 165:1946–52. [PubMed: 16186463]
8. Baker DW, Wolf MS, Feinglass J, Thompson JA, Gazmararian JA, Huang J. Health literacy and mortality among elderly persons. Arch Intern Med. 2007; 167:1503–9. [PubMed: 17646604]
9. Berkman ND, Sheridan SL, Donahue KE, Halpern DJ, Crotty K. Low health literacy and health outcomes: an updated systematic review. Ann Intern Med. 2011; 155:97–107. [PubMed: 21768583]
10. Kutner, M.; Greenberg, E.; Jin, Y.; Paulsen, C.; White, S. The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. Washington DC: U.S. Department of Education; 2006.
11. Williams MV, Parker RM, Baker DW, et al. Inadequate functional health literacy among patients at two public hospitals. JAMA. 1995; 274:1677–82. [PubMed: 7474271]
12. Hayes KS. Literacy for health information of adult patients and caregivers in a rural emergency department. Clin Excell Nurse Pract. 2000; 4:35–40. [PubMed: 11858294]
13. Trifiletti LB, Shields WC, McDonald EM, Walker AR, Gielen AC. Development of injury prevention materials for people with low literacy skills. Patient Educ Couns. 2006; 64:119–27. [PubMed: 16723205]
14. Tran TP, Robinson LM, Keebler JR, Walker RA, Wadman MC. Health literacy among parents of pediatric patients. Western J Emerg Med. 2008; 9:130–4.
15. Ginde AA, Clark S, Goldstein JN, Camargo CA Jr. Demographic disparities in numeracy among emergency department patients: evidence from two multicenter studies. Patient Educ Couns. 2008; 72:350–6. [PubMed: 18462915]
16. Olives T, Patel R, Patel S, Hottinger J, Miner JR. Health literacy of adults presenting to an urban ED. Am J Emerg Med. 2011; 29:875–82. [PubMed: 20627212]
17. Herndon JB, Chaney M, Carden D. Health literacy and emergency department outcomes: a systematic review. Ann Emerg Med. 2011; 57:334–45. [PubMed: 21035902]
18. Jolly BT, Scott JL, Feied CF, Sanford SM. Functional illiteracy among emergency department patients: a preliminary study. Ann Emerg Med. 1993; 22:573–8. [PubMed: 8442547]
19. Williams DM, Counselman FL, Caggiano CD. Emergency department discharge instructions and patient literacy: a problem of disparity. Am J Emerg Med. 1996; 14:19–22. [PubMed: 8630148]
20. Duffy MM, Snyder K. Can ED patients read your patient education materials? J Emerg Nurs. 1999; 25:294–7. [PubMed: 10424958]
21. Davis TC, Crouch MA, Long SW, et al. Rapid assessment of literacy levels of adult primary care patients. Fam Med. 1991; 23:433–5. [PubMed: 1936717]
22. Davis TC, Long SW, Jackson RH, et al. Rapid estimate of adult literacy in medicine: a shortened screening instrument. Fam Med. 1993; 25:391–5. [PubMed: 8349060]

23. Parker RM, Baker DW, Williams WV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J Gen Intern Med.* 1995; 10:537–41. [PubMed: 8576769]
24. Baker DW, Williams DM, Parker RM, Gazmararian JA, Nurss JR. Development of a brief test to measure functional health literacy. *Patient Educ Couns.* 1999; 38:33–42. [PubMed: 14528569]
25. Bass PF, Wilson JF, Griffith CH. A shortened instrument for literacy screening. *J Gen Intern Med.* 2003; 18:1036–8. [PubMed: 14687263]
26. Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam Med.* 2004; 36:588–94. [PubMed: 15343421]
27. Weiss BD, Mays MZ, Martz W, et al. Quick assessment of literacy in primary care: the newest vital sign. *Ann Fam Med.* 2005; 3:514–22. [PubMed: 16338915]
28. Osborn CY, Weiss BD, Davis TC, et al. Measuring adult literacy in health care: performance of the newest vital sign. *Am J Health Behav.* 2007; 31:S36–S46. [PubMed: 17931135]
29. 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; 23:561–6. [PubMed: 18335281]
30. Rawson KA, Gunstad J, Hughes J, et al. The METER: a brief, self-administered measure of health literacy. *J Gen Intern Med.* 2010; 25:67–71. [PubMed: 19885705]
31. Powers BJ, Trinh JV, Bosworth HB. Can this patient read and understand written health information? *JAMA.* 2010; 304:76–84. [PubMed: 20606152]
32. 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; 18:1148–55. [PubMed: 22092896]
33. Carpenter CR. The San Francisco Syncope Rule did not accurately predict serious short-term outcome in patients with syncope. *Evid Based Med.* 2009; 14:25. [PubMed: 19181957]
34. Perry JJ, Sharma M, Sivilotti ML, et al. Prospective validation of the ABCD2 score for patients in the emergency department with transient ischemic attack. *CMAJ.* 2011; 183:1137–45. [PubMed: 21646462]
35. Smidt N, Rutjes AW, van der Windt DA, et al. The quality of diagnostic accuracy studies since the STARD statement: has it improved? *Neurology.* 2006; 67:792–7. [PubMed: 16966539]
36. Bossuyt PM, Reitsma JB, Bruns DE, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. *Ann Intern Med.* 2003; 138:W1–12. [PubMed: 12513067]
37. Wolf MS, Williams MV, Parker RM, Parikh NS, Nowlan AW, Baker DW. Patients' shame and attitudes toward discussing the results of literacy screening. *J Health Commun.* 2007; 12:721–32. [PubMed: 18030638]
38. Bailey BP, Konstan JA. On the need for attention-aware systems: measuring effects of interruption on task performance, error rate, and affective state. *Comp Human Behav.* 2006; 22:685–708.
39. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology.* 1983; 148:839–43. [PubMed: 6878708]
40. Obuchowski NA. Sample size calculations in studies of diagnostic accuracy. *Stat Methods Med Res.* 1998; 7:371–92. [PubMed: 9871953]
41. Clark B. Using law to fight a silent epidemic: the role of health literacy in health care access, quality, & cost. *Ann Health Law.* 2011; 20:253–327. [PubMed: 21910318]
42. Miller MJ, Abrams MA, Earles B, Phillips K, McCleary EM. Improving patient-provider communication for patients having surgery: patient perceptions of a revised health literacy-based consent process. *J Patient Saf.* 2011; 7:30–8. [PubMed: 21921865]
43. Spandorfer JM, Karras DJ, Hughes LA, Caputo C. Comprehension of discharge instructions by patients in an urban emergency department. *Ann Emerg Med.* 1995; 25:71–4. [PubMed: 7802373]
44. Engel KG, Heisler M, Smith DM, Robinson CH, Forman JH, Ubel PA. Patient comprehension of emergency department care and instructions: are patients aware of when they do not understand? *Ann Emerg Med.* 2009; 53:454–61. [PubMed: 18619710]

45. Han JH, Bryce SN, Ely EW, et al. The effect of cognitive impairment on the accuracy of the presenting complaint and discharge instruction comprehension in older emergency department patients. *Ann Emerg Med.* 2011; 57:662–71. [PubMed: 21272958]
46. Hastings SN, Barrett A, Weinberger M, et al. Older patients' understanding of emergency department discharge information and its relationship with adverse outcomes. *J Patient Saf.* 2011; 7:19–25. [PubMed: 21921863]
47. Diner BM, Carpenter CR, O'Connell T, et al. Graduate medical education and knowledge translation: role models, information pipelines, and practice change thresholds. *Acad Emerg Med.* 2007; 14:1008–14. [PubMed: 17967963]
48. Paasche-Orlow MK, Wolf MS. Evidence does not support clinical screening of literacy. *J Gen Intern Med.* 2008; 23:100–2. [PubMed: 17992564]
49. Baker DW. The meaning and the measure of health literacy. *J Gen Intern Med.* 2006; 21:878–83. [PubMed: 16881951]
50. Mancuso JM. Assessment and measurement of health literacy: an integrative review of the literature. *Nurs Health Sci.* 2009; 11:77–89. [PubMed: 19298313]
51. Griffey RT, Melson AT, Lin MJ, Carpenter CR, Goodman MS, Kaphingst KA. Does numeracy correlate with measures of health literacy in the emergency department? *Acad Emerg Med.* 2014 this issue.
52. Bass PF, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. *Acad Med.* 2002; 77:1039–41. [PubMed: 12377684]
53. Kelly PA, Haidet P. Physician overestimation of patient literacy: a potential source of health care disparities. *Patient Educ Couns.* 2007; 66:119–22. [PubMed: 17140758]
54. Worster A, Carpenter C. A brief note about likelihood ratios. *CJEM.* 2008; 10:441–2. [PubMed: 18826732]
55. DeWalt DA, Broucksou KA, Hawk V, et al. Developing and testing the health literacy universal precautions toolkit. *Nurs Outlook.* 2011; 59:85–94. [PubMed: 21402204]
56. Baker DW, DeWalt DA, Schillinger D, et al. "Teach to goal": theory and design principles of an intervention to improve heart failure self-management skills of patients with low health literacy. *J Health Commun.* 2011; 16:73–88. [PubMed: 21951244]
57. Schenker Y, Fernandez A, Sudore RL, Schillinger D. Interventions to improve patient comprehension in informed consent for medical and surgical procedures: a systematic review. *Med Decis Making.* 2011; 31:151–73. [PubMed: 20357225]
58. Galasko D, Abramson I, Corey-Bloom J, Thal LJ. Repeated exposure to the Mini-Mental State Examination and the Information-Memory-Concentration Test results in a practice effect on Alzheimer's disease. *Neurology.* 1993; 43:1559–63. [PubMed: 8351011]
59. Newman, TB.; Kohn, MA. *Evidence-Based Diagnosis.* New York, NY: Cambridge University Press; 2009.
60. Worster A, Carpenter C. Incorporation bias in studies of diagnostic tests: how to avoid being biased about bias. *Can J Emerg Med.* 2008; 10:174–5.



**Figure 1.** Enrollment Flow Diagram \*11 patients excluded after enrollment due to missing S-TOFHLA scores. S-TOFHLA = Short Test of Functional Health Literacy in Adults.



**Figure 2.** ROC curve. NVS = Newest Vital Sign; REALM-R = Rapid Evaluation of Adult Literacy in Medicine-Revised; ROC = receiver operating characteristic; SILS = Single Item Literacy Screens.

**Table 1**

## Patient Demographics (N=435)

<b>Demographics</b>	<b>n (%)</b>
Female	241 (55.5)
Race/ethnicity	
White	133 (30.8)
Black	292 (67.6)
Asian	1 (0.2)
Other	6 (1.4)
Education Level Attained	
Less than high school	77 (17.7)
High school	218 (50.1)
Some college or higher	140 (32.2)
Primary Insurance	
Private	146 (33.3)
Self-pay	115 (26.4)
Medicaid	89 (20.5)
Medicare	73 (16.8)
Private + Medicare	2 (0.5)
Other insurance	10 (2.3)
Age in yrs, mean ( $\pm$ SD)	45 ( $\pm$ 15.7)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Table 2**

## Health Literacy Interpretations (N=435)

Instrument	Number of Participants (%)
Abbreviated S-TOFHLA	
Inadequate health literacy (score 0-16)	57 (13.1)
Marginal health literacy (score 17-22)	47 (10.8)
Adequate health literacy (score 23-36)	331 (76.1)
REALM-R, n=433	
Lower health literacy ( 6 items correct)	220 (48.5)
Higher health literacy (>6 items correct)	223 (51.5)
NVS, n=428	
High likelihood of limited literacy (0-1 items correct)	127 (29.7)
Possible limited literacy (2-3 items correct)	150 (35.1)
Adequate literacy (4-6 items correct)	151 (35.3)
Numeracy score	
0	79 (18.2)
1	171 (39.3)
2	118 (27.1)
3	49 (11.3)
4	18 (4.1)
Summed SILS, n=433	
3	3 (0.7)
4	5 (1.2)
5	4 (0.7)
6	4 (0.9)
7	5 (1.2)
8	11 (2.5)
9	19 (4.4)
10	16 (3.7)
11	36 (8.3)
12	49 (11.3)
13	54 (12.5)
14	79 (18.2)
15	149 (34.4)
Physician gestalt, n=309	
Inadequate	25 (8.1)
Marginal	83 (26.9)
Adequate	207 (65.1)

S-TOFHLA = Short Test of Functional Health Literacy in Adults; REALM-R = Rapid Estimate of Adult Literacy in Medicine-Revised; NVS = Newest Vital Sign; SILS = Single Item Literacy Screens

**Table 3**

Diagnostic Test Characteristics of NVS, REALM-R, SILS, and Numeracy Tests (N=435)

<b>Instrument</b>	<b>Sensitivity % (95% CI)</b>	<b>Specificity % (95% CI)</b>	<b>Positive Likelihood Ratio (95% CI)</b>	<b>Negative Likelihood Ratio (95% CI)</b>
NVS 3, n=428	98.0 (93.1-99.8)	45.7 (40.3-51.3)	1.81 (1.51-2.17)	0.04 (0.01-0.17)
REALM-R 6, n=433	80.8 (73.2-88.3)	61.7 (56.5-67.0)	2.14 (1.64-2.80)	0.30 (0.19-0.46)
Numeracy 1	80.8 (73.2-88.3)	50.2 (44.8-55.5)	1.61 (1.30-2.01)	0.38 (0.24-0.60)
SILS 12, n=433	68.0 (59.0-77.07)	75.5 (70.8-80.1)	2.78 (1.92-4.02)	0.42 (0.31-0.57)
<b>Help Reading, n=433</b>				
1 – Always	9.7 (4.0-15.4)	98.8 (96.9-99.7)	7.42 (1.13-48.89)	0.92 (0.86-0.98)
2 – Often	16.5 (9.3-23.7)	97.0 (95.1-98.8)	5.30 (1.59-17.64)	0.87 (0.79-0.95)
3 – Sometimes	42.7 (33.2-52.3)	85.5 (81.7-89.3)	2.81 (1.65-4.78)	0.69 (0.58-0.83)
4 – Rarely	49.5 (39.9-59.2)	66.1 (61.0-71.2)	1.49 (1.06-2.08)	0.75 (0.59-0.96)
5 – Never	100	0.00	1.00	--
Dichotomized*	42.7 (33.2-52.3)	85.5 (81.7-89.3)	2.81 (1.65-4.78)	0.69 (0.58-0.81)
<b>Medical Form Help</b>				
1 – Not at all	14.4 (7.7-21.2)	98.5 (97.2-99.8)	9.93 (1.90-52.03)	0.86 (0.79-0.94)
2 – A little bit	30.8 (21.9-39.6)	94.6 (92.1-97.0)	5.69 (2.38-13.59)	0.73 (0.64-0.84)
3 – Somewhat	54.8 (45.2-64.4)	80.4 (76.1-84.6)	2.78 (1.80-4.29)	0.57 (0.45-0.72)
4 – Quite a bit	74.0 (65.6-82.5)	55.9 (50.5-61.2)	1.69 (1.32-2.17)	0.45 (0.31-0.66)
5 – Extremely	1.00	0.00	1.00	--
Dichotomized*	54.8 (45.2-64.4)	80.4 (76.1-84.6)	2.78 (1.80-4.29)	0.57 (0.45-0.72)
<b>Ability To Read</b>				
1 – Very poor or terrible	6.7 (1.9-11.6)	100.0 (98.9-100.0)		
2 – Poor	14.4 (7.7-21.2)	99.1 (97.4-99.8)	16.56 (2.01-136.46)	0.86 (0.79-0.93)
3 – Okay	40.4 (31.0-49.8)	95.5 (93.2-97.7)	8.58 (3.36-21.87)	0.64 (0.55-0.76)
4 – Good	72.1 (63.5-80.7)	75.2 (70.6-80.0)	2.83 (1.97-4.08)	0.40 (0.29-0.55)
5 – Excellent or very good	1.00	0.00	1.00	--
Dichotomized*	40.4 (31.0-49.8)	95.5 (92.2-97.7)	8.58 (3.36-21.87)	0.64 (0.55-0.76)
Physician gestalt, n=309: inadequate or marginally adequate	52.7 (41.3-64.1)	70.6 (64.8-76.5)	1.72 (1.20-2.47)	0.70 (0.55-0.89)

\* Dichotomization used the following groupings: help reading (always/often as adequate, other responses as inadequate), medical form help (not at all/a little bit as adequate, other responses as inadequate), ability to read (excellent/very good/good as adequate, other responses as inadequate).

REALM-R = Rapid Estimate of Adult Literacy in Medicine-Revised; NVS = Newest Vital Sign; SILS = Single Item Literacy Screens

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

## Feasibility Assessments

Table 4

Health Literacy	S-TOFHLA (n=434)				REALM-R (n=433)				NVS (n=428)			
	Total Time n=434	% Interrupted n=435	Average Length of Interruption n=57	Time on Test n=433	Total Time n=433	% Interrupted n=433	Average Length of Interruption n=2	Time on Test n=433	Total Time n=428	% Interrupted n=428	Average Length of Interruption n=26	Time on Test n=429
Total	6.55 (7)	13.1	3.77 (2)	6.07 (6)	1.06 (1)	0.46	1.5 (1.5)	1.06 (1)	3.31 (3)	5.98	2.85 (2)	3.13 (3)
Adequate	6.29 (6)	14.20	3.53 (2)	5.79 (6)	1.03 (1)	0.45	2.00 (2)	1.03 (1)	3.07 (2)	5.96	3.11 (2)	2.89 (2)
Possible/Limited <sup>‡</sup>	7.40 (7)	9.62	5.00 (1)	6.98 (7)	1.09 (1)	0.48	1.00 (1)	1.09 (1)	3.43 (3)	6.14	2.71 (2)	3.27 (3)

Total time = total time for the test in minutes; Time on test = (Total time – Interrupted time)

<sup>‡</sup> For REALM-R, only have inadequate and adequate health literacy

Data are reported as mean (median) unless otherwise noted

S-TOFHLA = Short Test of Functional Health Literacy in Adults; REALM-R = Rapid Estimate of Adult Literacy in Medicine-Revised; NVS = Newest Vital Sign