

Review Article

Assessment and measurement of health literacy: An integrative review of the literature

Josephine M. Mancuso, MSc, ANP-BC, PhD(cand)
College of Nursing, Marquette University, Milwaukee, Wisconsin, USA

Abstract Research linking health literacy to health knowledge, health behaviors, health outcomes, health disparity, health status, and increasing health-care costs is prevalent around the globe. Given the importance of health literacy, it is prudent to examine the tools available to assure that patients are health-literate. This article provides an integrative review in order to investigate what has been developed to evaluate health literacy in the health-care setting. The research questions considered include: (i) Which instruments or screening tools are available to assess or measure health literacy in the clinical setting?; and (ii) What are the psychometric properties, advantages, and limitations of the identified tools? A number of databases are utilized to locate research specific to this topic. The research is analyzed, the findings are summarized, and the limitations are mentioned. The implications, recommendations, and the need for future research are discussed.

Key words health literacy assessment, health literacy measures, health literacy tools.

INTRODUCTION

A health-literate individual can apply the basic skills of reading, writing, and numeracy to health-related materials and activities within the health-care setting (Speros, 2005). Being health-literate means that one has the ability to understand health-care providers regarding health conditions and treatment options and that one knows where to go and who to seek out if help is needed. It also means being able to take medication correctly and to properly use home medical devices (Barrett & Puryear, 2006).

Research has consistently demonstrated far-reaching consequences for individuals and society if health literacy is not attained. Poor health status, defined as reported physical illness or perceptions of illness, disease, or impairment (Nielsen-Bohlman *et al.*, 2004), adverse health outcomes due to a lack of knowledge (Beckman *et al.*, 2004), health disparity, and a higher risk of disease and disability (Edmunds, 2005) are associated with health illiteracy. Those with inadequate health literacy use more health-care services, have a greater risk for hospitalization, and have a higher utilization of expensive services, such as emergency care and inpatient admissions (Nielsen-Bohlman *et al.*, 2004). When health literacy is lacking, individuals have less knowledge of their diseases and treatments, fewer self-management skills, poor compliance, and more medical or medication treatment errors. Health-illiterate individuals also lack the skills needed

to successfully negotiate the health-care system, have less access to health-care services, and incur increased health-care costs (American Medical Association, 1999).

Given the importance of health literacy to health status, health disparity, medical costs, issues of access to care, comprehension of health information, and decision-making in health care, it would be prudent to examine what is being done to evaluate patients' health literacy. The purpose of this article is to perform an integrative review of the literature to explore what has been developed to assess and measure health literacy in the everyday health-care setting. The research questions to be considered include the following: (i) Which instruments or screening tools are available to assess or measure health literacy in the health-care setting?; and (ii) What are the psychometric properties, advantages, and limitations of the identified tools?

HEALTH LITERACY

The term "health literacy" was first published in 1974 during the proceedings of a health education conference discussing health education as a social policy issue affecting the health-care system, mass communication, and the education system (Simonds, 1974). It was not until ~25 years later that definitions of health literacy began to appear in the health-care literature. In the USA, physicians denote health literacy as "a constellation of skills, including the ability to perform basic reading and numerical tasks required to function in the health care environment" (American Medical Association, 1999: 553). The National Library of Medicine defines health literacy as "the degree to which individuals have the capacity to obtain, process, and understand basic health

Correspondence address: Josephine M. Mancuso, Marquette University College of Nursing, PO Box 1881, 530 N. 16th Street, Milwaukee, WI 53201-1881, USA. Email: jmurphy1@wi.rr.com

Received 6 May 2008; accepted 8 June 2008

information and services needed to make appropriate health decisions" (Selden *et al.*, 2000). Internationally, the World Health Organization defines health literacy as "the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health" (World Health Organization, 1998). Researchers in Canada note health literacy as "people's ability to find, understand, appraise and communicate information to engage with the demands of different health contexts to promote health across the life-course" (Rootman & El-Bihbety, 2008: 11).

As one can conceptualize, health literacy has originated from the necessary skills of reading and numeracy to one of critical thinking, problem-solving, decision-making, information-seeking, and communication, along with a multitude of social, personal, and cognitive skills that are imperative in order to function in the health-care system. In addition, health literacy has expanded into the realm of culture, context, and language.

With such a vast body of core essentials, measuring and assessing health literacy in the clinical setting is difficult. Comprehensive, yet practical, measures to identify individuals with limited or absent health literacy are needed. A number of studies that purport to determine ways to assess health literacy and a number of instruments that have been developed to measure health literacy are reported in the literature.

METHODS

A review of the literature was completed using the PubMed, CINAHL, Web of Science, and Health Sciences in ProQuest databases. The search terms entered were "measurement of health literacy," "assessment of health literacy," and "health and literacy". Each term yielded from 764 to > 1000 entries. The entries were scanned for the appropriate terminology, that is, studies with a tool, question, or method to assess or measure health literacy. From this search, a total of 42 articles were retrieved.

Once retrieved, the articles were examined. The criteria for inclusion into the literature review were that: the article was research-based regarding an instrument or method utilized to assess or measure health literacy; the research study was based upon an explicit aim, method, sample, data collection, data analysis, and findings being reported; the study measured patient literacy related to health-care materials or in the health-care setting; the research was inclusive regarding the measurement of adult health literacy; and the research was related to health literacy of print-based, not web-based, materials. From these criteria, 14 research studies were analyzed. All of the studies found were based in the USA. Of note, although research on health literacy is available internationally, patient tools specific to the assessment and measurement of health literacy in the medical setting were lacking in Europe (Kondilis *et al.*, 2006), Canada (Rootman & Ronson, 2005), Australia (Keleher & Hagger, 2007), and other countries.

Research studies

The research studies examined were dated from 1991 to 2006. The majority of the studies involved the development of instruments to measure health literacy (57%; $n = 8$). The two main instruments were the Rapid Estimate of Adult Literacy in Medicine (REALM) (Table 1) and the Test of Functional Health Literacy in Adults (TOFHLA) (Table 2). Each of these instruments were then developed into shortened versions (Short-TOFHLA) and (Short-REALM) or brief versions (REALM-Revised) and (Brief-TOFHLA). Other instruments were the Medical Achievement Reading Test (MART), the Newest Vital Sign (NVS), and the Short Assessment of Health Literacy for Spanish-Speaking Adults (SAHLSA) (Table 3). The remainder of the studies were screening questions to assess the health literacy of adults in outpatient health clinics (29%; $n = 4$) and of parents caring for young children (14%; $n = 2$) (Table 4). To compare the research findings, the instruments and screening questions were grouped into sections and the psychometric properties were reviewed. A summary of the psychometric properties with their meanings is provided in Appendix 1.

Instruments

Rapid Estimate of Adult Literacy in Medicine

In 1991, Davis *et al.* conceived the first screening instrument to estimate patient literacy in the primary care, patient education, and medical research settings (Davis *et al.*, 1991). The REALM is a 125 word recognition test developed to identify patients with low literacy levels and to provide a reading grade range for those with limited literacy skills. The patients are asked to read all words aloud and are scored with a plus for correct pronunciation and a check for incorrect pronunciation. The raw scores are converted to grade ranges corresponding with lower elementary, upper elementary, junior high, and senior high school levels. The administration and scoring takes ~ 3–5 min.

The test-retest reliability was excellent, at 0.98. The validity was established by highly positive correlation of the REALM to two widely recognized and utilized standardized reading recognition tests, the Slosson Oral Reading Test (SORT; $r = 0.95$) and the Peabody Individual Achievement Test-Revised (PIAT-R; $r = 0.98$). The content validity and face validity were based on the selection of health-related words.

In 1993, Davis *et al.* developed a shortened version of the REALM (Davis *et al.*, 1993). The instrument was decreased to a 66 word recognition test. The retention of the items was based on a psychometric estimate of the item's difficulty and discrimination and the frequency of the retained words in the written material given to patients. The administration method and scoring now takes ~ 1–2 min. The test-retest reliability remained excellent, at 0.99. The validity was established by highly positive correlations of the REALM to the SORT ($r = 0.96$) and the PIAT-R ($r = 0.97$). An additional standardized reading recognition test utilized in medical settings was added for correlation, the Wide Range Achievement Test-Revised (WRAT-R; $r = 0.88$), and the validity remained acceptable.

Table 1. Rapid Estimate of Adult Literacy in Medicine (REALM) instrument

Characteristic	REALM	Shortened REALM	REALM-R
Author	Davis <i>et al.</i> (1991)	Davis <i>et al.</i> (1993)	Bass <i>et al.</i> (2003)
Purpose	A rapid-screening instrument designed to assess how well patients read common medical and lay terms that adult primary care patients are expected to recognize; to assist physicians in identifying patients with limited reading skills and in estimating patients' reading levels so that the appropriate level of patient education materials or oral instructions can be used	Same as REALM	Same as REALM
Design	A reading recognition test to assess an adult patient's ability to read and pronounce common medical terminology and lay terms for body parts and illnesses in ascending order of difficulty	Same as REALM	Same as REALM
Description	Consists of 125 words arranged in four columns according to the number of syllables and item difficulty	Consists of 66 words arranged in three columns of 22 words each according to the number of syllables and item difficulty	Consists of eight words instead of 66 words
Scoring	Scoring: 0–18 (below 3rd grade): might not be able to read most low-literacy materials 19–44 (4th–6th grade): might need low-literacy materials; might not be able to read prescription labels 45–60 (7th–8th grade): might struggle with most patient education materials; might not be offended by low-literacy materials 61–66 (high school): might be able to read most patient education materials	Same as REALM	Those with a score of ≤ 6 should be considered to be at risk for poor health literacy
Reliability	Test-retest reliability = 0.98	Test-retest reliability = 0.99	Cronbach's $\alpha = 0.91$
Validity	Content and face: established Criterion validity SORT: $r = 0.95$; PIAT-R: $r = 0.98$	Content and face: established Criterion validity SORT: $r = 0.96$; PIAT-R: $r = 0.97$; WRAT-R: $r = 0.88$	Criterion validity WRAT-R: $r = 0.64$
Instrument development	All words on the instrument were chosen from patient education materials and patient intake forms that were used in university-based primary care clinics		

PIAT-R, Peabody Individual Achievement Test-Revised; SORT, Slosson Oral Reading Test; WRAT-R, Wide Range Achievement Test-Revised.

Table 2. Test of Functional Health Literacy in Adults (TOFHFLA) instrument

Characteristic	TOFHFLA	TOFHFLA-S	S-TOFHFLA (Brief)	S-TOFHFLA (Short)
Author	Parker <i>et al.</i> (1995)	Parker <i>et al.</i> (1995)	Baker <i>et al.</i> (1999)	Baker <i>et al.</i> (1999)
Purpose	To measure a patient's ability to perform health-related tasks that require reading and numerical skills	Same as TOFHFLA	Same as TOFHFLA	Only reading comprehension is measured
Design	A timed, 12 min reading comprehension and a timed, 10 min numeracy section	Same as TOFHFLA, except in Spanish	A timed, 7 min reading comprehension and a timed, 5 min numeracy section	A timed, 7 min reading comprehension section
Description	A 50 item reading comprehension section from the instructions to prepare for an upper GI series, the patient rights and responsibilities section of a Medicaid application form, and the standard hospital informed consent form; a 17 item numerical ability test that assesses a patient's ability to use numerical skills that are necessary to comprehend directions on a prescription bottle, monitor blood glucose, keep clinic appointments, and obtain financial assistance	Same as TOFHFLA, except that the instrument is in Spanish	A 36 item reading comprehension section from instructions to prepare for an upper GI series and the patient rights and responsibilities section of a Medicaid application form; a four-item numerical ability test that assesses a patient's ability to use numerical skills that are necessary to comprehend directions on a prescription bottle, monitor blood glucose, and keep clinic appointments	A 36 item reading comprehension section from passages of the instructions to prepare for an upper GI series and the patient rights and responsibilities section of a Medicaid application form
Scoring	0-59 (inadequate health literacy): unable to read and interpret health texts 60-74 (marginal health literacy): difficulty reading and interpreting health texts 75-100 (adequate health literacy): can read and interpret most health texts Cronbach's $\alpha = 0.98$	Same as TOFHFLA	0-53: inadequate health literacy 54-66: marginal health literacy 67-100: adequate health literacy	0-16: inadequate health literacy 17-22: marginal health literacy 23-36: adequate health literacy
Reliability	Cronbach's $\alpha = 0.98$	Cronbach's $\alpha = 0.98$	Cronbach's $\alpha = 0.97$; reading Cronbach's $\alpha = 0.68$; numeracy	Cronbach's $\alpha = 0.97$
Validity	Criterion validity REALM: $r = 0.84$; WRAT-R: $r = 0.74$ Content: established	No criterion Content: established	Criterion validity REALM: $r = 0.80$ (total); REALM: $r = 0.61$ (numeracy); REALM: $r = 0.81$ (reading) Content: established	Criterion validity REALM: $r = 0.81$ Content: established
Instrument development	Review of > 30 examples of commonly used hospital texts, including patient education materials, instructions for diagnostic tests, prescription bottle labels and of varying difficulty			

GI, gastrointestinal; WRAT-R, Wide Range Achievement Test-Revised.

Table 3. Medical Achievement Reading Test (MART), the Newest Vital Sign (NVS), and the Short Assessment of Health Literacy for Spanish-Speaking Adults (SAHLSA) instruments

Characteristic	MART	NVS	SAHLSA
Author	Hanson-Divers (1997)	Weiss <i>et al.</i> (2005)	Lee <i>et al.</i> (2006)
Purpose	To develop a terminology literacy test that is easily and quickly administered and can accurately assess individual reading levels	To develop an English and Spanish screening tool that identifies patients at risk for low health literacy and will elicit information about the patient, allowing providers to adapt their communication practices in an effort to achieve better health outcomes	To develop an easy-to-use health literacy test for Spanish speakers
Design	Medical word recognition test based on words from prescription labels and patient education materials	A nutrition label from an ice cream container: patients are given the label and then asked questions about how they would interpret and act on the information contained on the label	Medical word recognition test with a comprehension component
Description	Composed of 42 medically related words typed in small letters, but at the top of the words are 15 capital letters; a glossy sheet covers the tool	The patient is handed a copy of the nutrition label and then asked a series of six questions about it; patients can and should retain the label so they can refer to it while answering questions	Consists of 50 words, each on a laminated 5 × 6.25 cm flash card with a medical term printed in boldface on the top and two association words, key and distracter, at the bottom
Scoring	By the number of letters and words correctly pronounced that correlates with an exact grade level; the raw score is converted to a grade equivalent Cronbach's $\alpha = 0.98$	0–1 suggests a high likelihood of $\geq 50\%$ of limited literacy 2–3 indicates the possibility of limited literacy 4–6 almost always indicates adequate literacy	A score between 0 and 37 is suggestive of inadequate health literacy
Reliability		Cronbach's $\alpha = 0.76$: English Cronbach's $\alpha = 0.69$: Spanish	Cronbach's $\alpha = 0.92$ Test-retest reliability = 0.86
Validity	Content: claims established Criterion: not established	Criterion validity TOFHLA: $r = 0.59$ (English); TOFHLA: $r = 0.49$ (Spanish)	Criterion validity TOFHLA-S: $r = 0.65$
ROC curve	Not done for MART	AUROC curve = 0.88 (95% CI = 0.84–0.93) for predicting NVS/TOFHLA-English AUROC curve = 0.72 (95% CI = 0.66–0.79) for predicting NVS/TOFHLA-Spanish	Not done for SAHLSA
Sensitivity and specificity	Not done for MART	English Score < 2: sensitivity = 72%; specificity = 87% Score < 4: sensitivity = 100%; specificity = 64% Spanish Score < 2: sensitivity = 77%; specificity = 57% Score < 4: sensitivity = 100%; specificity = 19%	Not done for SAHLSA
Instrument development	500 words were chosen from 119 prescription labels and from a medical dictionary; the words were evaluated according to difficulty, classified into groups, and then selected	Developed from a series of scenarios with questions that involved both reading and numeracy skills. Five scenarios were tested and the NVS was selected on the basis of its psychometric properties	Based on the 66-item REALM. The medical terms were translated into Spanish taking into account the dictionary definition and the commonality of usage in daily conversations with selection of a key and distracter to assess comprehension

AUROC, area under the receiver operating characteristic; CI, confidence interval.

Table 4. Screening questions to assess health literacy

Author	Study	Purpose	Measure of health literacy	Screening questions utilized or found to detect inadequate/marginal health literacy with results	Question development
Williams <i>et al.</i> (1995)	Inadequate functional health literacy among patients at two public hospitals	To ask patients about their perceived reading ability in an attempt to develop a simple screening tool for identifying those with inadequate health literacy	TOFHLA	<ol style="list-style-type: none"> 1. Can you read a newspaper? Sensitivity = 16.7%; specificity = 99.4% 2. Can you read forms and written materials obtained from the hospital? Sensitivity = 19.8%; specificity = 99.3% 3. Do you usually ask somebody to help you read materials you receive from the hospital? Sensitivity = 51.4%; specificity = 88.6% 	Questions were part of the data collection regarding self-reported reading difficulties when interviewing patients for a study measuring health literacy with the TOFHLA
Chew <i>et al.</i> (2004)	Brief questions to identify patients with inadequate health literacy	To identify clinically useful questions that might be effective for detecting inadequate or marginal health literacy among an adult population	Short TOFHLA	<ol style="list-style-type: none"> 1. How often do you have someone help you read hospital materials? (always, often, sometimes, occasionally, never): AUROC curve = 0.87 (95% CI = 0.78–0.96) 2. How confident are you filling out medical forms yourself? (extremely, quite a bit, somewhat, a little bit, not at all): AUROC curve = 0.80 (95% CI = 0.67–0.93) 3. How often do you have problems learning about your medical condition because of difficulty understanding written information? (always, often, sometimes, occasionally, never): AUROC curve = 0.76 (95% CI = 0.62–0.90) 	Sixteen screening questions based on the five domains identified with limited health literacy: navigating the health-care system, completing medical forms, following medication instructions, interacting with providers, and reading appointment slips
Wallace <i>et al.</i> (2006)	Screening items to identify patients with limited health literacy skills	To evaluate three candidate questions to determine their accuracy in identifying patients with limited or marginal health literacy	REALM	<ol style="list-style-type: none"> 1. How confident are you filling out medical forms yourself? (extremely, quite a bit, somewhat, a little bit, not at all) Detection of limited health literacy: AUROC curve = 0.82 (95% CI = 0.77–0.86); “Somewhat” response: sensitivity = 83%, specificity = 65% Detection of limited/marginal health literacy: AUROC curve = 0.79 (95% CI = 0.74–0.83); “Somewhat” response: sensitivity = 77%, specificity = 74% 	Used the three health literacy screening questions of Chew <i>et al.</i> (2004), with five possible responses

<p>Morris <i>et al.</i> (2006)</p>	<p>The Single Item Literacy Screener (SILS): evaluation of a brief instrument to identify limited reading ability</p>	<p>To assess the diagnostic accuracy of the SILS in order to identify limited reading ability: the SILS is a single item question intended to identify adults in need of help with printed health material</p>	<p>Short TOFHLA</p>	<p>1. How often do you need to have someone help you when you read instructions, pamphlets, or other written material from your doctor or pharmacy? (1, never; 2, rarely; 3, sometimes; 4, often; 5, always) Scores > 2 are considered to be positive for limited health literacy AUROC curve = 0.73 (95% CI = 0.69–0.78); sensitivity = 54%; specificity = 83%</p>	<p>Modified questions from Chew <i>et al.</i> (2004) to develop a single item literacy screener that would efficiently identify patients who have difficulty with a central aspect of health literacy: reading health related material</p>
<p>Bennett <i>et al.</i> (2003)</p>	<p>Screening for low literacy among adult caregivers of pediatric patients</p>	<p>To identify adults with an increased risk of low literacy skills among the primary caregivers of preschool children with questions in a simple screening tool</p>	<p>REALM</p>	<p>Identifies parents likely to have literacy skills at or below the 6th grade level: 1. How many years of school have you completed? (if the answer is < 12) 2. Is your child's other parent living with you now? (if the answer is no) 3. Do you ever read books for fun? (if the answer is no) AUROC curve = 0.76 (95% CI = 0.66–0.86) with two-question cut-off; sensitivity = 84%; specificity = 56%</p>	<p>Seventeen screening questions to explore three domains: literacy activity at home, the literacy skill of the respondent, and parental literacy skill</p>
<p>Sanders <i>et al.</i> (2004)</p>	<p>Number of children's books in the home: an indicator of parental health literacy</p>	<p>To determine which of seven screening questions is most useful for identifying parents with adequate health literacy</p>	<p>Short TOFHLA</p>	<p>Indicator of adequate parent health literacy: 1. How many adult's books do you have in your home? (> 10; positive predictive value = 94%) 2. How many children's books do you have in your home? (> 10; positive predictive value = 91%) Together, both have a positive predictive value of 89% and a negative predictive value of 33%</p>	<p>Seven adult health literacy questions garnered from studies documenting the impact of an adult's educational achievement, educational ambition, and home literacy environment</p>

AUROC, area under the receiver operating characteristic; CI, confidence interval; REALM, Rapid Estimate of Adult Literacy in Medicine; TOFHLA, Test of Functional Health Literacy in Adults.

To condense the REALM even further, Bass *et al.* (2003) designed the Rapid Estimate of Adult Literacy in Medicine-Revised (REALM-R). The REALM was now decreased to an eight-word recognition test, with an administration time of 1–2 min. In actuality, the test has 11 words. The first three words, “fat, flu, and pill” are not scored, but were left at the start of the REALM-R to decrease test anxiety and to enhance patients’ confidence. The development of the REALM-R was a pilot study of 157 adults, ranging in age from 18–93 years, and 85% of them were Caucasian. The reliability was excellent ($\alpha = 0.91$). The validity, established by correlation of the REALM-R to the WRAT-R, was questionable ($r = 0.64$).

The advantages of the REALM, shortened REALM, and REALM-R are many. All are quick and easy to administer and score and are acceptable to patients in a health-care setting because they use health-related words. The REALM and shortened REALM have been used widely in the literature and are highly reliable estimates of an individual’s ability to read words commonly used in medical settings. The criterion validity is well established by highly positive correlation with widely recognized and utilized standardized reading recognition tests: the SORT, the PIAT-R, and the WRAT-R. The content validity and face validity were based on the selection of health-related words. In addition, both may be used by researchers to identify potential validity problems with self-reported reading level data gathered from patients (Davis *et al.*, 1993).

However, all three forms of the REALM do not measure the patient’s understanding of the words, they only measure sight-reading ability. The REALM and shortened REALM assign only grade range equivalents and were specifically designed to identify patients who read at levels below ninth grade. The REALM-R only identifies those at risk for poor health literacy. The REALM R also has not been tested beyond the pilot study. Further study establishing the validity and reliability of the REALM-R is needed. In addition, all three forms of the REALM are available only in English. An attempt to translate the REALM to Spanish was reported in the literature as unsuccessful due to the phonetic structure of the Spanish language (Nurss *et al.*, 1995).

Test of Functional Health Literacy in Adults

The TOFHLA was the next instrument to appear in the literature and was specifically identified as a measure of “functional health literacy”. Developed by Parker *et al.* (1995), the TOFHLA tests a patient’s ability to read passages and phrases containing numbers using real materials from the health-care setting. The test consists of two timed parts: (i) numeracy, which assesses a patient’s ability to use numerical skills to comprehend directions; and (ii) reading comprehension at the 4th grade level, the 10th grade level, and the 19th grade level. The reading comprehension section uses the modified cloze procedure, where every fifth to seventh word in a passage is omitted. The reader selects a word to fill in the space from four possible choices, one of which is correct and three of which are similar but contextually or grammatically incorrect. It takes ~22 min to administer and is scored by the

sum of the two sections to yield scores that range from 0–100. The reliability was excellent ($\alpha = 0.98$). The content validity was enhanced by using actual hospital medical texts for both sections. The validity, established by correlation of the TOFHLA to the REALM ($r = 0.84$) and WRAT-R ($r = 0.74$), was good.

The TOFHLA also has been developed into a Spanish version (TOFHLA-S) (Parker *et al.*, 1995). To develop the Spanish version, the reading comprehension passages and numeracy questions were translated into Spanish and back-translated into English. The discrepancies were corrected using the consensus of several bilingual staff and a Spanish literacy expert. However, the criterion validity was not established because the REALM and WRAT-R do not have Spanish versions for correlation. The reliability remained excellent ($\alpha = 0.98$). The content validity was unchanged.

The TOFHLA is considered to be the “gold standard” of health literacy testing and one will find it used extensively in research on health literacy. The advantages of the TOFHLA are its strong reliability and validity data in English, its availability in Spanish and in a 14 point font print version for patients with poor vision, and it measures reading and numeracy skills, as well as comprehension skills. Another positive is that the reading passages measure a wide range of reading levels. Those who complete passage 1, but have difficulty completing passage 2, are comprehending between the 4th and 9th grade levels, the middle grades of 5–8, which the literature suggests patient education materials should be written at (Safeer & Keenan, 2005). The limitations of the TOFHLA include the time involved in the administration of the test (22 min) and the frustration for the participant of taking a timed test. There is also no validity data for the Spanish version.

In 1999, abbreviated versions of the TOFHLA were developed by Baker *et al.* (1999). These authors developed two abbreviated versions: the brief-TOFHLA and the short-TOFHLA. Both use the abbreviation, S-TOFHLA, in research studies and one must look at the description of the instrument to determine which test is being administered. The brief TOFHLA consists of two reading passages and four numeracy items. It takes ~ 12 min to administer and is scored by the sum of the two sections to yield scores that range from 0–100. The reliability of the reading comprehension section was excellent ($\alpha = 0.97$), while that of the numeracy items was questionable ($\alpha = 0.68$). The content validity was unchanged. The criterion validity, determined by correlation with the REALM, was good for the reading comprehension section ($r = 0.81$) and questionable for the numeracy section ($r = 0.64$). The short TOFHLA consists of two reading passages, takes ~7 min to administer, and is scored by the sum of correct answers to yield scores that range from 0–36. The reliability of the reading comprehension section ($\alpha = 0.97$) and criterion validity with the REALM ($r = 0.81$) were unchanged.

An advantage of both tests is the reduced time of administration. The training for and timing of the tests are the same. The reliability for the comprehension section remained strong. The reliability and validity for the numeracy section were diminished and further study is needed. Both have

versions in Spanish. However, these versions have no reported reliability or validity data. It is also noted that the short-TOFHLA is only a test of reading comprehension and might prove useful as a screening instrument to identify patients with very limited reading ability rather than health literacy. One must be cautious when reviewing research using the abbreviation, S-TOFHLA. If it is the short-TOFHLA, reading comprehension is measured, not health literacy.

Medical Achievement Reading Test

Following the TOFHLA, Hanson-Divers (1997) developed the MART. Like the REALM, it is a word recognition test. The 42 word MART is based on the Wide Range Achievement Test (WRAT). The WRAT is a standardized reading test, with strong validity and reliability, and it is utilized to determine grade-level reading ability by the pronunciation of words. Unlike the REALM, the MART was designed with three reasons why the individual might not be able to read the instrument. The “three excuses” were the use of medical or medically related words that are commonly seen on prescriptions or patient education leaflets, small print similar to that on prescription labels, and a glossy cover that creates a glare, which makes reading the words difficult. These difficulties allow patients to feel less intimidated by furnishing a pretext for not being able to read the words. The sample consisted of 405 participants with an average age of 36 years. The majority of the participants was Caucasian (56%) and female (56%). The reliability was excellent ($\alpha = 0.98$). The content validity was said to be established by the word placement used to formulate the test, based on the WRAT, which already had content validity. The author also concluded that, as the WRAT previously had been validated as an effective determinant of literacy status, it could be concluded that the MART was an equally effective determinant of literacy; thus, no correlations were completed.

The advantages of the MART are its reliability, quick administration, scoring in 3–5 min, and its unthreatening appearance. The MART also can be claimed to be more precise in its grade-level placement because it places the respondents into exact grade levels. But, the scoring and grade-level placement were not elaborated upon. Furthermore, many limitations exist. The MART only measures the recognition of words by sight and not by understanding. The sample size was small and not generalizable to the rest of the population. Finally, a major limitation was the assumption of content and criterion validity. Future studies need to address correlation with other instruments. Of note, the MART has not been used or studied in the literature since its publication.

The Newest Vital Sign

Over the next 8 years, no new instruments to measure health literacy were published. Then, Weiss *et al.* (2005) developed the NVS. The authors touted the NVS as a quick and accurate screening test for limited literacy, available in English and Spanish. The NVS is a six-question assessment based on an ice cream nutrition label. One point is given for each correct

answer, with literacy determined by a range. The sample consisted of 500 participants (250 English-speaking and 250 Spanish-speaking adults). The gender was mentioned, but the specifics were not provided. The mean age of both groups was 41 years and 43% of the English group was Caucasian. For the English version of the NVS, the reliability was acceptable ($\alpha = 0.76$). The criterion validity was determined with correlation to the TOFHLA and was poor ($r = 0.59$). The reliability of the Spanish version of the NVS was questionable ($\alpha = 0.69$) and the criterion validity with the TOFHLA was unacceptable ($r = 0.49$). The area under receiver operating characteristic (AUROC) curves were plotted for predicting the TOFHLA English (0.88) and Spanish (0.72) scores and were good. The sensitivity and specificity were reported for the various scores, with the sensitivity (English and Spanish: score $< 2 = 72\%$ and 77% , respectively; score $< 4 = 100\%$ for both) being predominately better than the specificity (English and Spanish: score $< 2 = 87\%$ and 57% , respectively; score $< 4 = 64\%$ and 19% , respectively).

The major advantage of the NVS is its ease in administration and scoring in 3–5 min, with Johnson and Weiss (2008) reporting an average of 2.9 min in an outpatient primary care clinic. Its sensitivity in both English and Spanish is noteworthy. However, there are still many limitations that need attention. The scoring descriptives lack precision, that is, “suggests a high likelihood of”, “indicates the possibility of”, and “almost always health-literate”. The sample size was small and mostly Hispanic, limiting generalizability. Gender also might be an issue. With high sensitivity, the NVS might misclassify patients with adequate health literacy, while the specificity might result in overestimating the percentage of patients with limited literacy. This point was clarified in a recently published study undertaken to determine the performance of the NVS (Osborn *et al.*, 2007). In addition, the psychometric properties of the NVS Spanish version were not as good as those of the English version and the reliability and validity were not as robust as other instruments published in the literature. Further studies need to be conducted to better examine and determine its reliability and validity data.

Short Assessment of Health Literacy for Spanish-speaking Adults

The last instrument identified through the literature search was the SAHLSA. Lee *et al.* (2006) developed the SAHLSA to screen for health literacy in the Spanish-speaking population. The SAHLSA is a word recognition test, based on the 66 item REALM, which incorporates a comprehension test. It requires the examinees to read out loud from a list of 50 medical terms and associate each term to another word similar in meaning to demonstrate comprehension. The answer is deemed correct only when the respondent accurately pronounces the word and makes the right association. It is easy to administer and scores < 37 indicate inadequate health literacy. The sample consisted of 201 Spanish-speaking and 202 English-speaking, primarily female (56%) adults, with a mean age of 34 years. The main purpose of including an English-speaking sample was to verify the design of the

association words in the SAHLSA, using the correlation between the REALM score and the SAHLSA association score. The criterion validity indicated a correlation of $r = 0.76$. This suggests that the design of the association questions was adequate. The reliability was excellent ($\alpha = 0.92$; test-retest score = 0.86). The criterion validity, established by correlation with the TOFHLA-S, was questionable ($r = 0.65$).

The advantages of the SAHLSA are its minimal training, ease of administration, scoring in 3–6 min, and strong reliability data. However, limitations exist. The validity was not convincing. The sample size was small and mainly female. The USA is host to populations from many different Latin American backgrounds that use different idiomatic expressions, making further testing of the instrument's application in different Latino subpopulations necessary. Furthermore, the SAHLSA measures only word recognition and comprehension. There is no numeracy section.

Screening questions

The current validated instruments that are discussed in order to assess health literacy have been criticized as being either too long or potentially embarrassing for individuals and have been noted to need further validation prior to being integrated into clinical care settings (Chew *et al.*, 2004; Davis *et al.*, 2005). Therefore, some researchers have attempted to evaluate health literacy with simple screening questions. These researchers measured health literacy with a form of the REALM or the TOFHLA and utilized ROC curves, sensitivity, and specificity, or predictive values, in order to compare the performance of the screening questions in relation to the measured health literacy.

Williams *et al.* (1995) asked patients three questions about their perceived reading ability in an attempt to develop a simple screening tool for identifying those with inadequate functional health literacy. The sample included a total of 2659 individuals at two outpatient hospital sites, consisting of primarily African American women and a Spanish component of 767 Latino women. The participants' health literacy was assessed with the TOFHLA. The sensitivity to the questions was low (16.7%, 19.8%, and 51.4%) while the specificity was high (99.4%, 99.3%, and 88.6%). Although the specificity was good, the sensitivity might result in underestimating the percentage of patients with limited literacy. Thus, the authors concluded that individuals' self-reported reading ability did not adequately screen for functional health literacy. Even though the sample was not diverse, the conclusion might be accurate because individuals tend to report higher reading levels than literacy levels (Doak *et al.*, 1996).

Chew *et al.* (2004) asked patients 16 screening questions based upon five domains identified in a qualitative study of patients with limited health literacy, followed by the short TOFHLA to determine health literacy. The sample consisted of 332 predominantly Caucasian male Veterans Administration patients. The results, according to the AUROC curves (0.87, 0.80, and 0.76) indicated that only three of the questions were effective in detecting limited health literacy. These questions were weaker for identifying the patients with marginal health literacy. The limitations of the study were that the

sample was small and predominately composed of Caucasian, male, veterans and, thus, was not generalizable. The cut-off points for sensitivity and specificity were not reported and one question if that might have been a better or worse predictor of the screening questions.

Wallace *et al.* (2006) expanded upon the work of Chew *et al.* (2004) by using the same three screening questions in a patient population that was demographically different from the veteran population. However, the REALM was used instead of the Short TOFHLA to measure health literacy. The sample consisted of 305 predominantly Caucasian females (67.5%). The results indicated that only one of the questions was effective in detecting limited and limited/marginal health literacy. The recommended cut-off point was the "somewhat" response, producing reasonable AUROC curves (0.82 and 0.79, respectively), sensitivity (83% and 77%, respectively), and specificity (65% and 74%, respectively). The limitations of the study include the small, predominantly Caucasian female, convenience sample, affecting generalizability. The REALM instead of the short TOFHLA was used to assess patients' health literacy and one inquires if this difference in health literacy measurement might have affected the results. Furthermore, the AUROC curve, sensitivity, and specificity for the other questions were not reported, leaving one to speculate regarding its comparison with the results of Chew *et al.* (2004).

In 2006, Morris *et al.* developed the Single Item Literacy Screener (SILS) as part of a larger project, the *Vermont Diabetes Information System*, by modifying the three questions identified by Chew *et al.* in 2004 into one question (Morris *et al.*, 2006). The short TOFHLA was used to measure health literacy. The sample consisted of 999 diabetic, predominantly female (54%), Caucasian adults. The results indicated a fair AUROC curve (0.73%), using a cut-off score of > 2 . The sensitivity (54%) and specificity (83%) were acceptable, indicating that the SILS performed reasonably well in detecting limited health literacy. However, the SILS did not perform well in detecting marginal health literacy (sensitivity of 34%). The limitations of the study include the sample of diabetics that were predominantly female and Caucasian, which might not represent the general population. Also, a higher sensitivity and AUROC curve closer to 1 would be preferable.

Besides screening adults, some researchers have screened the adult caregivers of children to ascertain their health literacy. Bennett *et al.* (2003) sought to identify simple screening items that could be used in clinical practice to identify parental literacy. They asked parents 17 screening questions based upon three domains hypothesized to be related to parental literacy, followed by the REALM to determine health literacy. The sample of 98 consisted of mainly African American mothers. The results indicated that three items were identified as candidate-screening items by virtue of their statistically significant associations with low literacy. The ROC curve analysis indicated that the three-item instrument had a fair AUROC curve (0.76), with good sensitivity (84%) and specificity (56%) at a two-question cut-off point. This correlated with less than a 6th grade reading level on the REALM. The advantage was in the sensitivity and specificity of the screening questions. These numbers appropriately

balanced the costs of false-positives with the risks of false-negatives. The two-question cut-off point resulted in the detection of 84% of adults with low literacy and 44% of the low-literacy risk group being incorrectly screened (actually being classified as having higher literacy).

To continue with adults who care for children, Sanders *et al.* (2004) conducted a study asking parents seven adult health literacy screening questions to identify those with adequate health literacy. Health literacy was measured with the short TOFHLA. The sample of 163 parents consisted of predominately African American mothers. The results yielded positive predictive values of 94% and 91% for two of the questions, with logistic regression identifying only two questions as being independently associated with adequate caregiver literacy. Together, these two had adequate positive (89%) and negative (33%) predictive values.

The limitations of both studies include the small, homogenous sample, which might not represent the general population. The REALM and S-TOFHLA were designed for use in internal medicine settings, not pediatrics. Lastly, both studies utilized screening questions versed in adult literacy research and not by a more formative qualitative research process with caregivers of children.

IMPLICATIONS AND RECOMMENDATIONS

Many constraints exist to the assessment of health literacy. All of the tools focused on medical terms or material found in medical settings and did not represent the broad spectrum of health literacy materials and processes that occur outside the clinical setting. This limits the conclusions that could be drawn regarding health literacy (Nielsen-Bohlman *et al.*, 2004). Health literacy includes more than word recognition, reading comprehension, and numeracy. The existing measures and screenings do not fully grasp the concept of health literacy in terms of language, context, culture, communication, or technology. Thus, we do not yet possess a measure that takes into account the full set of skills and knowledge associated with health literacy. Furthermore, potential confounders, such as test anxiety, distress due to illness, or cognitive deficits secondary to disease, are not taken into account when measuring health literacy. Another constraint is that, despite the diverse population in the USA and abroad, health literacy testing is possible in only two languages – English and Spanish – which might or might not be the patients' first or preferred language.

What then needs to be done to develop a measure that incorporates the vast expanse of health literacy? Should an assessment of health literacy be generalized and entail just the fundamentals of reading, comprehension, and numeracy, as previously discussed? Or, should an assessment of health literacy be specific to the medical population addressed, that is, literacy regarding diabetes, cancer, or heart disease? For instance, it stands to reason that if people suffer from a certain disease, they would exert an interest and want to develop a better understanding of that disease process and the self-care that it encompasses. Therefore, patients with hypertension might be successful at reading and understanding material pertaining to hypertension, but might not under-

stand other topics written at the same reading level. Accordingly, disease-specific measures of literacy could be one agreeable option. The literature reveals a few examples of instruments based on this rationale.

Diamond (2007) detailed the development and psychometric characteristics of the Nutritional Literacy Scale, offered as a measure of an adult's ability to comprehend nutritional information. Nath *et al.* (2001) developed a tool for assessing literacy specific to diabetes and Agre *et al.* (2006) reported on a literacy assessment tool specific to cancer patients, which is currently being tested in a research study. Other measures related to health literacy that are context-specific or age-specific include the e-Health Literacy Scale, designed to assess consumers' perceived skills at using information technology for health (Norman & Skinner, 2006), and the Rapid Estimate of Adolescent Literacy in Medicine, a tool to screen adolescent literacy in the health-care setting (Davis *et al.*, 2006).

There are a number of issues to consider when selecting and administering health literacy screening tools in the clinical setting. Clinical settings must consider the cost of the test, the training and time required to administer and score the test, the validity and reliability of the test, and whether the test is appropriate for the target population and language. The screening needs to be conducted in a private setting and the rationale for the screening explained to the patient. Before screening, health-care entities need to consider what the patient will be told about the results, how the results will be kept, and how the results will affect care (Davis *et al.*, 2005). Privacy and confidentiality must be assured and communicated to the patient. Principally, the patient's perspective should always be considered. One must take into account anxiety, stressors, or possible shame or embarrassment. The individual's language, age, eyesight, hearing, and mental status are important. The timing is equally important and screening might not be appropriate if the patient is too sick to respond, feels ill, or has just been given bad news (Davis *et al.*, 1998). It also has been advocated that the screening of patients not occur unless specific health education interventions are being individually tailored for the patient (Davis *et al.*, 2002).

Finally, it is necessary to recognize that this entire literature review and a majority of the citations are from the medical literature. Nurses are at the forefront of educating patients and are vocal advocates for vulnerable groups, yet very little substantive research exists in the nursing literature about health literacy, its screening, or its components. Given the consequences, nurses need to be educated and competent in the assessment of health literacy. Unfortunately, there are few curricular standards for undergraduate nursing education that currently address the need for health literacy. At the graduate level, the National Organization of Nurse Practitioner Faculties includes a section on health literacy to give added emphasis for the need to incorporate it into nurse practitioner curricula (Weill, 2006). This is an important beginning. However, nursing must continue to evolve and incorporate health literacy into the core components of nursing curricula. The development of competency-based programs with continuing education credit is necessary and

should incorporate health literacy and be made available to all nursing specialties. Nurses must be educated to become active participants in and contribute to health literacy research.

FUTURE RESEARCH

Although there is an increased awareness of the effects of health literacy, it is unknown how common health literacy screening is in the clinical setting (Davis *et al.*, 2005) or how nurses assess health literacy. A few studies that report on residents' and physicians' ability to identify patients with low health literacy have been published in the literature (Bass *et al.*, 2002; Powell & Kripalani, 2005; Kelly & Haidet, 2007), but no research studies are found that determine how nurses assess the health literacy of patients in the hospital or clinical setting. The nursing literature views health literacy through only one component: ensuring that patient education is at the appropriate grade level. Future research needs to explore how health literacy is assessed by nurses in the clinical setting and even if health literacy is being assessed in health-care settings at all.

The concept of health literacy is not entirely straightforward and the term is defined broadly and in a variety of ways in the literature. The instruments reviewed here are not strongly connected to the broader definitions of health literacy. Future research with integrative reviews of the literature and concept/dimensional analysis need to be undertaken in order to develop a definition that represents homogeneous significance for all disciplines in order to facilitate instrument development. Finally, to obtain health literacy for non-English-speaking populations, research is needed to develop tools to measure health literacy in the native language and within the culture of the patients or to determine whether existing English-based tools yield valid and reliable results.

CONCLUSION

The assessment of health literacy is an emerging field of research and study. The instruments currently utilized are the ground-breaking tools that have made it possible to advance the study of measuring health literacy among many patient populations. However, this is the groundwork that enables us to further our understanding and promotion of health literacy screening as a vital sign in the health-care setting. As the field of health literacy evolves, increased research and investigation are needed to progress health literacy screening from bench research to translational research applied with general populations in clinical and hospital settings. Accurately assessing health literacy will assist to improve health outcomes, decrease health disparity, and increase health status, leading to enhanced quality of life.

REFERENCES

- Agre P, Stieglitz E, Milstein G. The case for development of a new test of health literacy. *Oncol. Nurs. Forum* 2006; **33**: 283–289.
- Altman DG, Bland JM. Statistics notes: diagnostic tests 2: predictive values [electronic version]. *BMJ* 1994; **309**. [Cited 27 Apr 2008.] Available from URL: <http://www.bmj.com/cgi/content/full/309/6947/102>.
- American Medical Association. Health literacy: Report of the Council on Scientific Affairs. *JAMA* 1999; **281**: 552–557.
- 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**: 33–42.
- Barrett SE, Puryear JS. Health literacy: Improving quality of care in primary care settings. *J. Health Care Poor Underserved* 2006; **17**: 690–697.
- Bass PF, Wilson JF, Griffith CH. A shortened instrument for literacy screening. *J. Gen. Intern. Med.* 2003; **18**: 1036–1038.
- Bass PF, Wilson JF, Griffith CH, Barnett DR. Residents' ability to identify patients with poor literacy skills. *Acad. Med.* 2002; **77**: 1039–1041.
- Beckman ND, Dewalt DA, Pignone MP *et al.* Literacy and health outcomes: summary, evidence report/technology assessment no. 87. Rockville, MD: Agency for Healthcare Research and Quality, 2004. No. 04-E007-1. [Cited 27 Apr 2008.] Available from URL: <http://www.ahrq.gov/clinic/epcsums/litsum.pdf>.
- Bennett IM, Robbins S, Al-Shamali N, Haecker T. Screening for low literacy among adult caregivers of pediatric patients. *Fam. Med.* 2003; **35**: 585–590.
- Chew LD, Bradley KA, Boyko EJ. Brief questions to identify patients with inadequate health literacy. *Fam. Med.* 2004; **36**: 588–594.
- Davis TC, Crouch MA, Long SW *et al.* Rapid assessment of literacy levels of adult primary care patients. *Fam. Med.* 1991; **23**: 433–435.
- Davis T, Kennen EM, Gazmararian JA, Williams MV. Literacy testing in health care research. In: Schwartzberg JG, Vangeest JB, Wang CC (eds). *Understanding Health Literacy: Implications for Medicine and Public Health*. Chicago, IL: American Medical Association, 2005; 157–179.
- Davis TC, Long SW, Jackson RH *et al.* Rapid Estimate of Adult Literacy in Medicine: a shortened screening instrument. *Fam. Med.* 1993; **25**: 391–395.
- Davis TC, Michielutte R, Askov EN, Williams MV, Weiss BD. Practical assessment of adult literacy in health care. *Health Educ. Behav.* 1998; **25**: 613–624.
- Davis TC, Williams MV, Marin E, Parker RM, Glass J. Health literacy and cancer communication. *CA Cancer J. Clin.* 2002; **52**: 134–149.
- Davis TC, Wolf MS, Arnold CL *et al.* Development and validation of the Rapid Estimate of Adolescent Literacy in Medicine (REALM-Teen): a tool to screen adolescents for below-grade reading in health care settings. *Pediatrics* 2006; **118**: 1707–1714.
- Diamond JJ. Development of a reliable and construct valid measure of nutritional literacy in adults. *Nutr. J.* 2007; **6**. [Cited 20 Apr 2008.] Available from URL: <http://www.nutritionj.com/content/pdf/1475-2891-6-5.pdf>.
- Doak CC, Doak LG, Root JH. *Teaching Patients with Low Literacy Skills* (2nd edn). Philadelphia, PA: JB Lippincott, 1996.
- Edmunds M. Health literacy a barrier to patient education. *Nurse Pract.* 2005; **30**: 54.
- Hanson-Divers EC. Developing a medical achievement reading test to evaluate patient literacy skills: a preliminary study. *J. Health Care Poor Underserved* 1997; **8**: 56–69.
- Johnson K, Weiss BD. How long does it take to assess literacy skills in clinical practice? *J. Am. Board Fam. Med.* 2008; **21**: 211–214.
- Keleher H, Hagger V. Health literacy in primary health care. *Aust. J. Primary Health* 2007; **13**: 24.
- Kelly PA, Haidet P. Physician overestimation of patient literacy: a potential source of health care disparities. *Patient Educ. Couns.* 2007; **66**: 119–122.

- Kondilis BK, Soteriades ES, Falagas ME. Health literacy research in Europe: a snapshot. *Eur. J. Public Health* 2006; **16**: 113.
- Lee SY, Bender DE, Ruiz RE, Cho YI. Development of an easy-to-use Spanish health literacy test. *Health Serv. Res.* 2006; **41**: 1392–1412.
- Morris NS, MacLean CD, Chew LD, Littenberg B. The single item literacy screener: Evaluation of a brief instrument to identify limited reading ability. *BMC Fam. Pract.* 2006; **7**. [Cited 12 Apr 2008.] Available from URL: <http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1435902&blobtype=pdf>.
- Nath CR, Sylvester ST, Yasek V, Gunel E. Development and validation of a literacy assessment tool for persons with diabetes. *Diabetes Educ.* 2001; **27**: 857–864.
- Nielsen-Bohlman L, Panzer AM, Kindig DA. *Health Literacy: A Prescription to End Confusion*. Washington, DC: National Academies Press, 2004.
- Norman CD, Skinner HA. eHEALS: The eHealth Literacy Scale. *J. Med. Internet Res.* 2006; **8**. [Cited 20 Apr 2008.] Available from URL: <http://www.jmir.org/2006/4/e27>.
- Nurss JR, Baker DW, Davis TC, Parker RM, Williams M. Difficulties in functional health literacy screening in Spanish-speaking adults. *J. Read.* 1995; **38**: 632–637.
- 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** (Suppl. 1): 36–46.
- Parker RM, Baker DW, Williams MV, Nurss JR. The test of functional health literacy in adults: a new instrument for measuring patients' literacy skills. *J. Gen. Intern. Med.* 1995; **10**: 537–541.
- Polit DF, Beck CT. *Nursing Research: Principles and Methods* (7th edn). Philadelphia, PA: Lippincott, Williams & Wilkins, 2004.
- Powell CK, Kripalani S. Brief report: resident recognition of low literacy as a risk factor in hospital readmission. *J. Gen. Intern. Med.* 2005; **20**: 1042–1044.
- Rootman I, El-Bihbety DGA. *Vision for a Health Literate Canada: Report of the Expert Panel on Health Literacy*. Ottawa, ON: Canadian Public Health Association, 2008.
- Rootman I, Ronson D. Literacy and health research in Canada: where have we been and where should we go? *Can. J. Public Health.* 2005; **96** (Suppl. 2): 62–77.
- Safer RS, Keenan J. Health literacy: the gap between physicians and patients. *Am. Fam. Physician* 2005; **72**: 463–468.
- Sanders LM, Zacur G, Haecker T, Klass P. Number of children's books in the home: an indicator of parent health literacy. *Ambul. Pediatr.* 2004; **4**: 424–428.
- Selden CR, Zorn M, Ratzan S, Parker R. Current bibliographies in medicine 2000–1: health literacy. 2000. [Cited 20 Apr 2008.] Available from URL: <http://www.nlm.nih.gov/archive//20061214/pubs/cbm/hliteracy.html>.
- Simonds SK. Health education as social policy. *Health Educ. Monogr.* 1974; **21**: 1–10.
- Speros C. Health literacy: concept analysis. *J. Adv. Nurs.* 2005; **50**: 633–640.
- Tape TG. The area under an ROC curve. n.d. [Cited 13 Apr 2008.] Available from URL: <http://gim.unmc.edu/dxtests/ROC3.htm>.
- Wallace LS, Rogers ES, Roskos SE, Holiday DB, Weiss BD. Brief report: screening items to identify patients with limited health literacy skills. *J. Gen. Intern. Med.* 2006; **21**: 874–877.
- Weill V. Health literacy in nurse practitioner education: The NONPF perspective. 2006. [Cited 13 Apr 2008.] Available from URL: <http://www.iom.edu/File.aspx?ID=37044>.
- Weiss BD, Mays MZ, Martz W *et al.* Quick assessment of literacy in primary care: the Newest Vital Sign. *Ann. Fam. Med.* 2005; **3**: 514–522.
- Williams MV, Parker RM, Baker DW *et al.* Inadequate functional health literacy among patients at two public hospitals. *JAMA* 1995; **274**: 1677–1682.
- World Health Organization. Health promotion glossary. 1998. [Cited 13 Apr 2008.] Available from URL: http://www.who.int/hpr/NPH/docs/hp_glossary_en.pdf.

APPENDIX I

Psychometric properties of the instruments and screening questions

Instruments

Validity: verifies that one is in fact measuring what one intends to measure (Polit & Beck, 2004).

Criterion validity: involves determining the relationship between an instrument and an external measure through correlation. Correlations are designated by the lower case, *r*, and range in value from -1 to $+1$, with $+1$ being ideal (Polit & Beck, 2004).

Reliability: the extent to which an instrument is dependable, precise, predictable, and consistent. Cronbach's alpha (α) is a commonly used index of reliability and anything > 0.7 is considered to be acceptable (Polit & Beck, 2004).

Test-retest reliability: administering the same measure to a sample on two occasions and then comparing the scores. Values of 0.7 and 0.8 are considered to be satisfactory or good, respectively (Polit & Beck, 2004).

Screening questions

Sensitivity: the ability of the screen to identify a case correctly, yielding a true positive – screen in those with (Polit & Beck, 2004).

Specificity: the ability to identify non-cases correctly, yielding a true negative – screen out those without (Polit & Beck, 2004).

Receiver operating characteristic (ROC) curve analysis: used to evaluate the screening questions' discriminating capacity to accurately identify those individuals in the low and high literacy groups. The area under the ROC (AUROC) curve represents an overall measurement of the performance of the screening test, with 1.0 being a perfect test and 0.5 representing a test with no discriminating capacity (Tape, n.d.).

Positive predictive value: the proportion of patients with positive results who are correctly identified (Altman & Bland, 1994).

Negative predictive value: the proportion of patients with negative results who are correctly identified (Altman & Bland, 1994).