Journal of Health Communication

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uhcm20

Applicability of Internationally Available Health Literacy Measures in the Netherlands

M. P. Fransen a, T. M. Van Schaik a, T. B. Twickler b, c & M. L. Essink-Bot a

a Department of Public Health, Amsterdam Medical Centre, University of Amsterdam, Amsterdam, The Netherlands
b Department of Internal Medicine, Amsterdam Medical Centre, University of Amsterdam, Amsterdam
c Department of Internal Medicine, Lievensberg Hospital, Bergen op Zoom, The Netherlands

Available online: 27 Sep 2011


To link to this article: http://dx.doi.org/10.1080/10810730.2011.604383

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan, sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Applicability of Internationally Available Health Literacy Measures in the Netherlands

M. P. FRANSEN AND T. M. VAN SCHAICK
Department of Public Health, Amsterdam Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

T. B. TWICKLER
Department of Internal Medicine, Amsterdam Medical Centre, University of Amsterdam, Amsterdam and Department of Internal Medicine, Lievensberg Hospital, Bergen op Zoom, The Netherlands

M. L. ESSINK-BOT
Department of Public Health, Amsterdam Medical Centre, University of Amsterdam, Amsterdam, The Netherlands

Health literacy measures for use in clinical-epidemiological research have all been developed outside Europe. In the absence of validated Dutch measures, we evaluated the cross-cultural applicability of the Rapid Estimate of Adult Literacy in Medicine (REALM), the Newest Vital Sign (NVS), the Set of Brief Screening Questions (SBSQ), and the measure of Functional Communicative and Critical Health Literacy (FCCHL). Each measure was translated into Dutch following standardized procedures. We assessed feasibility, internal consistency, and construct validity among patients with coronary artery disease (n = 201) and patients with diabetes type 2 (n = 88). Patients expressed most problems in responding to the NVS-D. They were not familiar with the type of food label and had difficulties calculating in portions instead of grams. The FCCHL-D items seemed too theoretical for many patients. Cronbach’s alpha was acceptable for all measures. Correlation patterns between the measures were moderately coherent with a priori hypotheses. All translated measures were able to distinguish between high- and low-educated groups of patients, with the NVS-D performing best. Despite reasonable psychometric properties as demonstrated so far, these measures need to be further developed in order to increase applicability for assessing health literacy in clinical-epidemiological research in the Netherlands.

Individual responsibility for health and self-management of disease are promoted and relied on in modern society. Access to information and knowledge on health and disease increasingly determine health, and people are expected to adequately use this knowledge and information to remain in good health. This active patient role requires adequate health literacy. The Institute of Medicine defined health...
literacy as the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions (Ratzan & Parker, 2000). People with low health literacy tend to be less likely to successfully manage chronic diseases (DeWalt, Berkman, Sheridan, Lohr, & Pignone, 2004; DeWalt & Hink, 2009). They generally have less knowledge about their disease, exhibit inadequate self-management, and have poor health outcomes (DeWalt et al., 2006; Gazmararian, Williams, Peel, & Baker, 2003; Rothman, DeWalt et al., 2004; Rothman, Malone, Bryant, DeWalt, & Pignone, 2002; Rothman, Malone et al., 2004; Schillinger et al., 2002).

In order to investigate the impact of low health literacy on people’s health, use of health care, and effectiveness of interventions, appropriate and valid measurement of health literacy in medical-epidemiological research is essential. Adequate measures of health literacy facilitate the development of effective ways to handle low health literacy in health care and to improve health outcomes. All instruments to measure health literacy have been developed and validated outside Europe. Objective health literacy measures involve assessment of a subject’s actual skills, such as reading and calculating. For example, the respondent completes a questionnaire to allow for a judgment of his/her skills based on external norm scores. Subjective measures involve assessment of a subject’s perceived skills, asking, for example, if he or she experiences limitations in obtaining and understanding information. Objective measures include the Rapid Estimate of Adult Literacy in Medicine (REALM) and the Newest Vital Sign (NVS), whereas Chew’s Set of Brief Screening Questions (SBSQ) and Ishikawa’s Japanese measure of Functional Communicative and Critical health literacy (FCCHL), represent subjective measures (Chew, Bradley, & Boyko, 2004; Davis et al., 1991; Davis et al., 1993; Ishikawa, Takeuchi, & Yano, 2008; Weiss et al., 2005). In most European countries, such as the Netherlands, the prevalence of low health literacy is unknown. Since no health literacy measures suitable for use in medical-epidemiological research are currently available in Dutch, our aim was to evaluate the applicability of the REALM, NVS, SBSQ, and the FCCHL in the Netherlands.

**Methods**

**Selection of Health Literacy Measures**

We performed a literature search and used two literature reviews to select the most valid, reliable, and feasible objective and subjective health literacy measures for adaptation into Dutch (Jordan, Osborne, & Buchbinder, 2010; Mancuso, 2009). A selection of measures were listed and discussed among the authors of this paper. We decided to include the REALM, because it has proven to provide a reliable estimate of individual reading skills (Ibrahim et al., 2008; Jordan et al., 2010; Mancuso, 2009). Although the REALM merely measures reading skills and not the broader concept of health literacy, it has been widely used in the United States to validate other health literacy measures. We considered the translation of the REALM as an important first step in health literacy research in the Netherlands, because it enables us to compare correlation patterns with newly developed health literacy measures and bring health literacy research in line with international developments. The original REALM consists of 125 words and was developed in the United States to assist physicians in the identification of adults with limited reading skills in a primary
care setting (Davis et al., 1991). Later, a shorter version was developed, consisting of 66 words. Subjects are asked to read these 66 words aloud and receive a score of one point if they pronounce the word correctly. This results in a total score range of 0–66, which is converted to a United States school-grade estimate of reading ability. Scores below 18 indicate that patients might not be able to read most low-literacy materials, scores between 19 and 44 indicate that patients need low-literacy materials, scores between 45 and 60 indicate that patients may have problems in reading most patient education materials, and scores above 60 indicate that patients are probably able to read most patient education materials (Davis et al., 1993). The REALM is highly correlated with standardized general reading tests (Slosson Oral Reading Test, \( r = 0.96 \)) and has demonstrated a test-retest reliability of 0.98 (Jordan et al., 2010; Mancuso, 2009). The content validity of the REALM was based on the selection of health-related words (Davis et al., 1993).

The other objective measure that we decided to adapt is the NVS. The NVS has been developed in the United States and consists of six items based on a nutrition label from an ice cream container. The NVS not only measures reading skills, but also numeracy skills and the ability to apply information. For each correctly answered item, a score of one point is granted. A score between 0 and 1 suggests a likelihood of ≥50% of limited literacy, 2–3 indicates the possibility of limited literacy, and 4–6 almost always indicates adequate literacy (Weiss et al., 2005). The correlation of NVS and REALM scores was only 0.41, which is in accordance with expectations, because the NVS measures not only reading skills, but also calculating skills. The NVS items showed adequate internal consistency (\( \alpha = 0.76 \)) (Jordan et al., 2010; Mancuso, 2009).

In addition to these objective measures, we included Chew’s SBSQ as a subjective measure. This tool consists of three statements. Responses are scored on a 5-point Likert scale from 0 to 4, added, and averaged. The response of ‘somewhat’ or less provided optimum sensitivity and specificity and is considered as an optimal screening threshold in most studies (Chew et al., 2004; Chew et al., 2008). This means that an average score of ≤2 indicates inadequate health literacy, and a score >2 indicates adequate health literacy.

Because these objective and subjective tools measure only functional health literacy, we decided to include Ishikawa’s subjective FCCHL for measuring functional, communicative, and critical health literacy (Ishikawa et al., 2008). Communicative health literacy refers to the cognitive and literacy skills which, together with social skills, can be used to actively participate in everyday activities, to extract information and derive meaning from different forms of communication, and to apply new information to changing circumstances. Critical health literacy refers to the more advanced cognitive skills which, together with social skills, can be applied to critically analyze information, and to use this information to exert greater control over life events and situations (Nutbeam, 2000). The FCCHL measures these three constructs by 14 statements using 4-point Likert scales (1–4) as response options. The total score is obtained by summing item scores and dividing by the total number or items. The FCCHL has been applied only in Japan. Because other health literacy measures were not available in Japanese, demonstration of construct validity was limited. The internal consistency of the functional health literacy and communicative health literacy scales appeared to be sufficient for group comparisons (\( \alpha > 0.7 \)). Cronbach’s alpha of the critical health literacy scale was 0.65 (Ishikawa et al., 2008).
Translation and Adaptation of the Measures

The development of the REALM, NVS, SBSQ, and FCCHL into Dutch followed standardized formal forward-backward procedures for health status measures (Streiner & Norman, 2008).

Step 1: The measures were independently forward translated by two native Dutch speakers with an epidemiological or clinical background and fluency in English. They both were aware of the study objectives.

Step 2: Comprehensiveness and grammar were checked and the discrepancies were arbitrated. Solutions were reached by consensus between the two native speakers.

Step 3: The merged Dutch versions were backward translated by a native English speaker fluent in Dutch. The native speaker was blinded for the original English version.

Step 4: The backward translations were compared with the original English version. There were several discrepancies, but these were mainly related to the wording and not to the specific meaning of the items. We discussed discrepancies with the native English speaker and reached a consensus.

Step 5: The translated versions (REALM-D, NVS-D, SBSQ-D and FCCHL-D) were reviewed.

Step 6: The translated versions were pretested qualitatively to determine practical feasibility and applicability among 34 diabetic patients before starting the patient study described below.

Step 7: Based on the results of the pretest, the REALM-D, NVS-D, SBSQ-D and FCCHL-D were slightly adapted in order to enhance feasibility and applicability. Point of departure in present adaptations was to stick as closely as possible to the original measures in an attempt to maintain international comparability. In the quantitative stage of our study we assessed psychometric properties of the translated measures and gained more information on what cross-cultural adaptations would be needed. As an example, we deliberately did not at that point try to change the NVS food label into a more “European” type of food label. We adjusted only those items that appeared unfeasible and inapplicable for most pretest respondents.

Step 8: We cross-culturally adapted the measures and tested them again. In the FCCHL-D, respondents did not understand what was meant by “various sources of information”, for example. We therefore divided this item into 5 items describing specific sources (e.g., health care professionals, books). This brought the total number of items in the Dutch version to 18. Because some of the items in the FCCHL refer to a diabetes diagnosis, we changed “diabetes” into a Dutch word representing the more general term “disease.”

Study Population

The translated measures were tested among patients with coronary artery disease (CAD) and patients with type 2 diabetes mellitus (T2DM). The CAD population consisted of patients participating in a Dutch multicenter randomized controlled trial investigating the effect of secondary prevention in patients admitted for an acute coronary syndrome (Jørstad et al., 2009). During the 12-month follow-up visit, we recruited patients from two of the participating centers. The T2DM-population
consisted of patients from one general practice in the southeast of Amsterdam, participating in a qualitative study on diabetes self-management. They were recruited after their 3-month visit to the nurse practitioner. Patients were excluded if they were younger than 18 years of age or if they were unable to communicate in Dutch.

Data Collection

In the CAD population, the REALM-D, NVS-D, SBSQ-D, and FCCHL-D were assessed by telephone or face-to-face interviews. The procedures for the telephone and face-to-face interviews were essentially similar. For the telephone interviews, the patients received an introductory letter about the study, along with the REALM-D and the NVS-D in a sealed envelope labeled with the instruction not to open the envelope until the start of the interview. This was to ensure that the patient did not study the REALM-D or the NVS-D in advance. In the T2DM population health literacy was assessed in face-to-face interviews as part of an interview study on diabetes self-management. This study included only the REALM-D and NVS-D to assess health literacy levels in this population.

Background characteristics were measured by assessing respondents’ date of birth, gender, ethnic origin, and educational attainment level. Ethnic origin was assessed by country of birth of the respondent and his/her parents. Educational attainment level was categorized (Statistics Netherlands, 2011) as low (primary school), medium (first and second stage secondary education), or high (vocational college or university).

Analysis

Feasibility of the REALM-D, NVS-D, SBSQ-D and FCCHL-D was evaluated by patient burden and difficulties experienced in performing the tests. Quantitative data were used to assess score distributions (including floor and ceiling effects), internal consistency and construct validity. Ceiling effects refer to the level above which variance in an independent variable is no longer measured or estimated. A ceiling effect occurs when a high proportion of subjects in a study have maximum scores on the observed variable. This makes discrimination among subjects among the top end of the scale impossible. For this reason, examination of test results for a possible ceiling effect, and the converse floor effect, is often built into the analysis of psychometric properties of instruments such as those used for measuring quality of life (Cramer & Howitt, 2005; Essink-Bot, Krabbe, Bonsel, & Aaronson, 1997). Floor or ceiling effects causing an abnormal score distribution were considered to be present when >15% of the respondents had the lowest or the highest possible health literacy score, respectively. This criterion is arbitrary, but often used in the analysis of psychometric properties of, for example, quality of life measures (Bregnballe et al., 2008; Ponto et al., 2011; Valero et al., 2007).

The internal consistency of a measure refers to the extent to which the different items assess the same characteristic. Internal consistency was estimated by calculating Cronbach’s alpha coefficients for each measure. We calculated Cronbach’s alpha for the three subscales of FCCHL and for the total scale.

Construct validity refers to whether a scale measures or correlates with the underlying theoretical construct. It was assessed by correlation analysis, where Pearson correlations coefficients with \( r > 0.3 \) were considered relevant, and by
known groups comparisons. p-Values ≤.01 were considered statistically significant. This significance level was chosen in order to avoid the side effects of multiple testing and to decrease the risk of unjustified rejection of true null hypotheses. Our hypotheses were as follows:

1. The REALM-D scores would correlate stronger with the NVS-D scores than with the SBSQ-D and FCCHL-D scores, because REALM-D and NVS-D are both objective measures.
2. The correlation between REALM-D scores and NVS-D scores would not be perfect, because the NVS measures a broader concept of health literacy.
3. The SBSQ-D scores would correlate stronger with the FCCHL-D scores, because both are subjective measures.
4. The correlation between SBSQ-D and FCCHL-D would not be perfect, because FCCHL-D is aimed at measuring a broader concept of health literacy.
5. All adapted health literacy measures would be able to significantly differentiate between patients with low and high educational level.
6. Objective measures would be more capable to differentiate between low- and high-educated patients, because objective measures are tests of skills, whereas patients easily overestimate or underestimate their own health literacy with the subjective measures.
7. The NVS-D is most capable to distinguish between patients with low and high health literacy, because it involves the most difficult assignment.

Results

Study Population

In total, 269 CAD patients were asked to participate in the study, and of these, 201 were interviewed (response rate 75%). Most patients who did not want to participate reported that they did not have time or they suffered from fatigue. In total, six CAD patients participated in face-to-face interviews, 195 were interviewed by telephone. The exact percentage of nonresponse in the T2DM-population is unknown by design. In total, 88 T2DM-patients who were recruited by the nurse practitioner participated in a face-to-face interview.

We could not assess the REALM and/or the NVS among four T2DM patients and two CAD patients. Two T2DM patients lacked sufficient Dutch reading skills, two forgot their reading glasses. Two CAD patients were unable to read out loud because they had suffered a stroke. Table 1 presents the background characteristics of the study population. The CAD patients were more often male, from Dutch ethnic origin, and were a little bit younger than the T2DM patients.

Health Literacy Scores

Table 2 presents the scores per health literacy measure. According to the REALM-D scores, 19% of the patients had difficulty reading. The mean score was 61.9. The REALM-D showed a ceiling effect with 23% of the patients exhibiting the maximum score of 66. In total, 56% of the patients scored one or no items correctly in the NVS-D, which suggests a high likelihood of low health literacy. The mean NVS-D score was 1.8. In total, 31% of the patients did not answer any of the six items correctly, indicating a floor effect. In total, 72% of the patients scored 3 points or less on
the FCCHL-D, indicating low subjective health literacy. In total, 5% of the patients scored low on the SBSQ-D, indicating low subjective health literacy as defined by this measure. The SBSQ-D showed a ceiling effect with 42.5% of the patients exhibiting the maximum score.

Feasibility of the Adapted Health Literacy Measures

Patients expressed problems primarily in responding to the NVS-D. Especially the diabetic patients were confused, because eating ice cream is not in keeping with their dietary recommendations and they did not understand why they had to answer these questions on this food item. We frequently noticed that patients did not recognize the food label, because Dutch food labels differ from U.S. labels in layout and content. For example, patients had difficulties calculating in portions instead of the usual grams that are reported on Dutch food labels. Moreover, many patients did not understand that they were supposed to apply the information that was given in the ice cream label; this can be regarded as a sign of low health literacy. Patients experienced fewer difficulties with the REALM-D. As with the American version, the REALM-D was easy to administer. Patients easily understood the instructions. The feasibility of the SBSQ-D seemed relatively high as well. However, patients did not always understand the item “How often do you have problems learning about your medical condition because of difficulty understanding written information?” A possible explanation is that this item is difficult to disentangle for some patients, because it actually involves three questions: (1) “Do you have problems learning about your medical condition?” (2) “Do you have difficulty understanding information?” (3) “Does your difficulty understanding information cause problems learning about your medical condition?”

Patients experienced several problems in answering the FCCHL-D items that seemed to be too theoretical. Patients, for example, expressed difficulty in answering the items, “Since being diagnosed with diabetes, have you applied the obtained

Table 1. Background characteristics of the study population

<table>
<thead>
<tr>
<th>Gender n (%)</th>
<th>Total (n = 289)</th>
<th>CAD-patients (n = 201)</th>
<th>T2DM-patients (n = 88)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>191 (66)</td>
<td>158 (80)</td>
<td>33 (37.5)</td>
</tr>
<tr>
<td>Age mean (SD)</td>
<td>59.7 (10.5)</td>
<td>58.6 (9.4)</td>
<td>62.5 (12.3)</td>
</tr>
<tr>
<td>Educational level n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low</td>
<td>40 (14)</td>
<td>24 (12)</td>
<td>16 (18)</td>
</tr>
<tr>
<td>Medium</td>
<td>187 (65)</td>
<td>127 (64)</td>
<td>60 (68)</td>
</tr>
<tr>
<td>High</td>
<td>60 (21)</td>
<td>48 (24)</td>
<td>12 (14)</td>
</tr>
<tr>
<td>Ethnic origin n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dutch</td>
<td>210 (73)</td>
<td>168 (84)</td>
<td>42 (48)</td>
</tr>
<tr>
<td>Afro-Caribbean</td>
<td>33 (11)</td>
<td>5 (2)</td>
<td>28 (32)</td>
</tr>
<tr>
<td>Other</td>
<td>30 (15)</td>
<td>18 (9)</td>
<td>12 (2)</td>
</tr>
</tbody>
</table>

*2 missing.
<table>
<thead>
<tr>
<th>Table 2. Health literacy scores</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>N (%)</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>REALM-D*</td>
</tr>
<tr>
<td>Score 0-18</td>
</tr>
<tr>
<td>Score 19-44</td>
</tr>
<tr>
<td>Score 45-60</td>
</tr>
<tr>
<td>Score 61-66</td>
</tr>
<tr>
<td>NVS**</td>
</tr>
<tr>
<td>Score 0-1</td>
</tr>
<tr>
<td>Score 2-3</td>
</tr>
<tr>
<td>Score 4-6</td>
</tr>
<tr>
<td>SBSQ</td>
</tr>
<tr>
<td>Score ≤2</td>
</tr>
<tr>
<td>Score &gt;2</td>
</tr>
<tr>
<td>FCCHL</td>
</tr>
<tr>
<td>Score ≤3</td>
</tr>
<tr>
<td>Score &gt;3</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*2 missing among CAD patients, 3 missing among T2DM patients.

**2 missing among CAD patients, 4 missing among T2DM patients.

***Not assessed in this group.
information to your daily life?” and “Since being diagnosed with diabetes, have you considered whether the information was applicable to your situation?” Their questions for clarification indicated that they did not actually understand what was meant by applying information to their daily life or what was meant by considering whether information was applicable.

Internal Consistency of the Adapted Health Literacy Measures

The Cronbach’s alpha coefficient for the REALM-D was 0.91 (Table 2). Cronbach’s alpha was 0.78 for the NVS-D. Both coefficients are regarded as sufficient for group comparisons. The Cronbach’s alpha coefficient for the SBSQ-D was 0.69, which indicates an acceptable internal consistency. The FCCHL-D also showed an acceptable internal consistency of 0.68. The Cronbach’s alpha coefficients for the first and third FCCHL-D scale were acceptable, but the second scale showed a poor internal consistency.

Construct Validity of the Adapted Health Literacy Measures

Tables 3 and 4 show the correlations between the adapted health literacy measures. In contrast to our first hypothesis, we found that the correlation of REALM-D scores was strongest with the SBSQ-D scores ($r = 0.59$, $p = .00$). Correlation with the FCCHL-D was weak ($r = 0.15$, $p = .04$). In line with our second hypothesis, the correlation between REALM-D and NVS-D was moderate, both in the CAD and in the T2DM population ($r = 0.32$, $p = .00$/$r = 0.22$, $p = .04$). In contrast to our third hypothesis, the SBSQ-D was not correlated to the FCCHL-D. In fact, it showed a very weak correlation ($r = 0.07$, $p = .35$). As stated in the fourth hypothesis, we did not expect a very high correlation, because the FCCHL is aimed at measuring a broader concept of health literacy.

Table 3. Correlations between scores in the CAD-population ($n = 201$)

<table>
<thead>
<tr>
<th></th>
<th>REALM-D</th>
<th>NVS-D</th>
<th>SBSQ-D</th>
<th>FCCHL-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>REALM-D</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVS-D</td>
<td>0.32 (.00)</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SBSQ-D</td>
<td>0.59 (.00)</td>
<td>0.22 (.00)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>FCCHL-D</td>
<td>0.15 (.04)</td>
<td>0.18 (.01)</td>
<td>0.07 (.35)</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4. Correlations between scores in the T2DM-population ($n = 88$)

<table>
<thead>
<tr>
<th></th>
<th>REALM-D</th>
<th>NVS-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>REALM-D</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NVS-D</td>
<td>0.22 (.04)</td>
<td>1</td>
</tr>
</tbody>
</table>
Known group comparisons confirmed the fifth hypothesis. Except for the FCCHL-D, all translated measures were able to significantly ($p < .01$) differentiate between low and high educated patients (see Table 5). Patients with a low educational level scored lowest on all health literacy measures, those with a high educational level scored highest. In line with our sixth hypothesis, we found that the objective measures were most capable to differentiate between low- and high-educated groups. In contrast to the other measures, NVS-D scores significantly differed among all three educational levels. This confirms the seventh hypothesis that the NVS-D is most capable to distinguish between patients with low and high health literacy.

**Discussion**

Despite reasonable psychometric properties as demonstrated so far, the adaptation of international health literacy measures need to be further developed to increase applicability for assessing health literacy in a clinical-epidemiological setting in the Netherlands.

In the following sections, most important findings, possible explanations, and scientific comparisons will be discussed for each health literacy measure, followed by strengths and limitations of this study, conclusions, and implications.

**NVS-D**

The NVS-D was most capable to distinguish among patients with low, intermediate, and high educational levels, but showed most difficulties in cross-cultural applicability. The problems had to do with the type of food label and with ice cream as the product used in the test. Food labels are not standardized internationally, and U.S. food labels differ in layout and content from Dutch food labels. Patients did not recognize the food label and had difficulties in calculating in portions instead of grams. Ice cream in the Netherlands is much less a product for daily use than...
it is in the United States. The level of difficulty of the NVS-D caused anger and irritation; this is inherent to the test and probably reflects low health-literacy skills on the side of the respondent, but it limits the applicability of the NVS-D. The fact that 31% of the patients did not answer any of the items correctly, questions the specificity of the NVS-D. Low specificity may result in overestimating the percentage of patients with limited health literacy. Since Weiss and colleagues also reported low specificity in the NVS, the low specificity might not be subscribed to the cross-cultural applicability of the NVS-D (Weiss et al., 2005).

The U.S. version of the NVS has also been applied in Australia and Singapore (Adams et al., 2009; Ko, Lee, Toh, Tang, & Tan, 2011). In the Australian population, 24% of the respondents had an NVS score between 2 and 3, 21% scored 0–1. This means that 55% had adequate health literacy, which is relatively high compared to the 21% adequate health literacy that we found. In the UK, Rowlands and colleagues recently adapted and validated the NVS for use in the UK population. The U.S. food label was adapted to fit in a UK format of food labeling through a Delphi procedure. Preliminary analysis showed acceptable reliability and validity of the NVS-UK (Rowlands, 2011). While developing the NVS, Weiss and colleagues simultaneously developed a Spanish version and tested it among Spanish-speaking primary care patients in the United States. Compared with the English version, the psychometric qualities of the Spanish version were lower, but sufficient (Weiss et al., 2005). Ozdemir and colleagues translated the NVS-US and REALM-US into Turkish in order to assess health literacy levels in Turkey, using similar translation procedures as we did in the present study (cross-cultural adaptation of the U.S. food label is not mentioned in their paper). They found that 28% of the family medicine clinic patients had an NVS score between 4 and 6, which is higher than 21% in our population. The difference cannot be explained by differences in educational level, because our population was more highly educated than the Turkish population. As in our study, the Turkish NVS had sufficient internal consistency (0.70) and was significantly related to educational level (Ozdemir, Alper, Uncu, & Bilgel, 2010).

**REALM-D**

Patients experienced few problems in responding to the REALM-D. However, cross-cultural translation of the REALM seemed complicated. In the translation of the REALM, we tried to maintain the same number of syllables, meaning, and level of difficulty in the words. This had a positive effect on the comparability between the REALM-D and the original REALM. However, the words only partially represent the words that are most often used in the Dutch medical information materials, limiting the content validity and international comparability of scores.

An attempt to translate the REALM into Spanish was reported unsuccessful, because of the phonetic structure of the Spanish language (Mancuso, 2009). Because Dutch words are not phonetic, this problem was irrelevant in our translation process. Ozdemir and colleagues also translated the REALM into Turkish. They used translations of the same medical terms as in the original, but rearranged them in order of pronunciation difficulty in the Turkish language. Probably the same drawbacks regarding international comparability of scores as found in our study apply to this Turkish study. They found a mean REALM score of 60.29. Only 58.7% had a
score above 60; this was 81% in our study (Ozdemir et al., 2010). A possible explanation is that Turkish respondents were not asked to read the words on the paper, but were asked to say all the words they know. Although the Turkish respondents scored higher on the NVS, they scored lower on REALM compared with Dutch respondents. This seems contradictory and warrants further research. Dowse and colleagues evaluated the appropriateness of the REALM in an English second-language South African population. On average, 46 out of 66 words were adequately pronounced and less than half this number adequately comprehended. The authors concluded that the original REALM appears to be unsuitable for use in its current form in an English second-language population (Dowse, Lecoko, & Ehlers, 2010). Ibrahim and colleagues evaluated the REALM for use in the UK against a UK general literacy screening tool and found that the REALM had face, criterion, and construct validity for use as a health literacy screening tool in the UK. They did not report on content validity (Ibrahim et al., 2008).

**SBSQ**

The cross-cultural applicability of the SBSQ seemed relatively high. However, we noticed that patients did not always understand what was meant by “Do you have problems learning about your medical condition, because of difficulty understanding written information?” This problem could probably not be attributed to restrictions in the cross-cultural applicability of this item, but to the complicated wording that is used in the original item. A validation study in 2008 also showed that only one item performed best (“Confident with forms”) in detecting patients with inadequate health literacy. The capacity to detect a broader group of patients with inadequate or marginal health literacy was weaker for all three items (Chew et al., 2008). To our knowledge, the SBSQ has not been applied in other countries nor has it been translated into other languages.

**FCCHL**

Our study showed several problems in the application of the FCCHL. These problems could not directly be attributed to the cross-cultural applicability of the items, but to the complexity of the original items. Independent of the translation, the items require high abstraction skills. We therefore think the measure seems less suitable for patients with low health literacy, irrespective of their language and cultural background. Except for Ishikawa and colleagues (2008), no other studies have reported on the applicability of the FCCHL.

**Strengths and Limitations**

A strength of our study is that the findings contribute to the few results on cross-cultural applicability of health literacy measures. This evaluation is not only essential for further development of health literacy research in the Netherlands, but also offers a starting point for cross-cultural evaluation in other languages. One of the limitations of our study refers to the fact that we were unable to compare the translated measures with a standardized Dutch literacy screening instrument, because no valid and practical instrument yet exists in the Netherlands. Furthermore, the requirement to read and provide written informed consent for all patients
participating in the randomized controlled trial potentially may have caused selection bias, leading to an overestimation of literacy in comparison to the general (patient) population. However, this does not impede the conclusion of this study, because we aimed at psychometric analysis and not at establishing health literacy levels. Because we measured health literacy in CAD patients at a follow-up moment of the clinical trial, we had no health literacy data from patients who withdrew consent or became deceased. Another limitation is that we combined personal and telephone interviews to assess health literacy, causing potential mode effects. In general, telephone interviews are suitable to assess patient-reported outcomes, but it is unknown to what extent they are suitable to assess objective HL. Although we took measures to avoid that respondents would prepare themselves, and we did not receive any signs that respondents had already opened the envelope, we cannot exclude that some respondents read the REALM and NVS label beforehand. However, respondents did not know the exact assignment, because the REALM assignment and NVS items were not added to the envelope.

**Conclusion and Implications**

Health literacy measures need further cross-cultural adaptation in order to adequately assess health literacy in the Netherlands. Cross-cultural adaptation of health literacy measures is more complicated than of, for example, health status measures, because language abilities are an important element of the object of measurement in measuring health literacy (Koller et al., 2007; van Baar et al., 2006).

The promising concept of the NVS is encouraging further efforts toward the cross-cultural adaptation of this measure in the Netherlands. The development of a highly applicable Dutch NVS requires changing the NVS to a European food label, preferably relating to a product other than ice cream. This in fact means developing a whole new measure, which in turn may reduce the cross-cultural comparability of the scores. In further development of the REALM in other languages, it should be recognized that a conceptually comparable REALM requires similar decoding and vocabulary (lexicon) skills. Since decoding is heavily affected by word length; the comparability of the decode component should be assessed by comparing word lengths in letters and syllables. The comparability of the lexicon component should be assessed by comparing word frequency—referring to the frequency of a word in a collection of texts, such as spoken language, fiction, popular magazines, and newspapers. Because the problems that we identified in assessing the SBSQ-D and the FCCHL-D were probably not related to cross-cultural differences but rather to the content and construct of the items, our recommendations are aimed at general improvement of these items (Streiner & Norman, 2008).

In the process of developing and adapting health literacy measures, we have to keep in mind that current measures only measure the ability to read, calculate, and understand. In order to measure the broader concept of health literacy, new measures need to be developed that measure not only functional health literacy, but also interactive and critical health literacy. Recent publications report on promising initiatives being taken in the United States and the UK to measure these various domains of health literacy (Chinn, 2011; Clayman et al., 2010; Gazmararian, Beditz, Pisano, & Carreon, 2010; Yost et al., 2010). In order to increase cross-cultural applicability of such measures, we recommend international collaboration in the further development of these measures.
References


Shriners Hospital for Children Burn Outcomes Questionnaire (5–18 years of age). *Journal of Burn Care & Research*, 27, 790–802.
