Validation of Turkish health literacy measures

Ezgi Eyüboğlu¹,* and Peter J. Schulz²

¹Public Relations and Publicity Department, Maltepe University, Marmara Eğitim Köyü, Maltepe, Istanbul 34857, Turkey, and ²Institute of Communication and Health, Università Della Svizzera Italiana, Lugano, Switzerland

*Corresponding author. E-mail: ezgieyuboglu@maltepe.edu.tr

Summary

This study aimed to validate a Turkish version of the Short Test of Functional Health Literacy (S-TOFHLA) (Baker et al., Development of a brief test to measure functional health literacy. Patient Educ Counsel 1999;38:33–42) and a Turkish version of the Chew self-report scale (Chew et al., Brief questions to identify patients with inadequate health literacy. Family Med, 2004;36:588–94) for measuring functional health literacy. The original English version of the S-TOFHLA and the Chew items were translated by applying standardized translation methods and cultural adaptations, and both were administered to a sample of diabetes patients (N=302) in two diabetes clinics in one of the major cities in Turkey. Self-administered paper–pencil questionnaires were distributed to eligible outpatients who had a clinic appointment. In addition to the S-TOFHLA measurement and the Chew screening questions, gender, age, educational attainment, income, marital status and diabetes knowledge were obtained. The Turkish version of S-TOFHLA showed high internal consistency. Both S-TOFHLA and the Chew screening scale correlated significantly with known predictors of health literacy: age, education and income. The Chew scale was also related weakly but significantly with general diabetes knowledge. It is expected that the Turkish versions of S-TOFHLA and the Chew scale will be used in Turkey as well as in other countries with large Turkish communities.

Key words: health literacy, measurement, validity, S-TOFHLA

INTRODUCTION

Research indicates that inadequate health literacy is associated with many negative outcomes such as poor health status, less medication adherence, lack of knowledge about disease, worse medical conditions and earlier death (Baker et al., 1997; Williams et al., 1998; Kalichman et al., 1999; Gazmararian et al., 2003; Zarcadoolas et al., 2005). People with chronic diseases (e.g. diabetes, heart diseases, asthma and high blood pressure) and limited health literacy have less understanding about their disease and experience more negative outcomes than do individuals with higher health literacy. Based on such findings, the improvement of health literacy is expected to have profound effects on the health of a group of people who need it most: patients with chronic diseases and low health literacy. The importance of health literacy has been recognized not only in the USA and Canada, where initial research originated, but also in other countries such as Australia (Jordan et al., 2010), Korea (Lee and Kang, 2013), Japan (Ishikawa et al., 2008; Inoue et al., 2013), the UK (Bostock and Steptoe, 2012; Chinn and McCarthy, 2013), the Netherlands (Van der Heide et al., 2013; Nijman et al., 2014) and Switzerland (Wang and Schmid, 2007; Connor et al.,...
Although the USA has produced more than three times as much health literacy research than Europe in the period of 1991–2005 (Kondilis et al., 2008), the importance of the issue has been increasingly recognized in Europe, too.

The situation is different in Turkey, where research on health literacy is really rare. One reason for this dearth of research is that the standard measures of health literacy such as the Short Test of Functional Health Literacy in Adults (S-TOFHLA) have never been translated into Turkish and validated for use in Turkey, with its population of 75.6 million and an 1.2% annual population growth rate (TUIK, 2012), or among the Turkish communities living in other countries, comprising more than five million people (TCDIB, 2013).

The most commonly used instruments in the literature to measure health literacy are the Rapid Estimate of Adults’ Literacy (REALM) and the Test of Functional Health Literacy in Adults (TOFHLA; Rudd and Keller, 2009), p. 243). Both are available as shortened or adapted versions (e.g. S-TOFHLA, S-REALM, REALM-R; Mancuso, 2009). S-TOFHLA is one of the most widely used test of functional literacy due to its strong reliability and validity data in English. Also it has been translated and validated in several languages such as Spanish (Baker et al., 1999), Chinese (Cantonese; Tang et al., 2007), Brazilian Portuguese (Carthey-Goulart et al., 2009), German, French and Italian (Connor et al., 2013), Serbian (Jovic-Vranes et al., 2009) and Hebrew (Baron et al., 2007).

The increased attention to health literacy and the evolution of shorter screening instruments has fostered interest in developing screening programs in the clinical context (Seligman et al., 2005), p. 1006). Therefore, many studies have been either using or developing screening questions, for they are effective and easy to use (Williams et al., 1995; Bennett et al., 2003; Chew et al., 2004; Lee et al., 2013).

Instruments such as REALM and TOFHLA have some major disadvantages: they take much time to fill in, even in their short versions, and they may make patients feel ashamed as their low health literacy becomes apparent. Screening questions are based on self-report measures, which quickly and accurately screen patients for limited health literacy. It would be helpful to develop rapid and inexpensive measures to assess the level of health literacy in clinical context or for studies with a great number of patients (Chew et al., 2008). Thus, efficient intervention strategies could be designed and implemented focusing on health literacy.

Measures discussed in this article are measures of functional health literacy. In addition to this concept, Nutbeam (Nutbeam, 2000) defines the dimensions of communicative/interactive and critical health literacy. Communicative/interactive health literacy refers to the skills of patients to be able to obtain health information and derive meaning from different forms of communication [(Nutbeam, 2000), p. 263]. Critical health literacy is the integration of social and cognitive skills to be able to analyze information critically and so to have control over life events.

This paper analyzes the reliability and validity of a Turkish translation of S-TOFHLA and the Chew screening questions for measuring functional health literacy. By using both, this study will allow comparing screening questions with a gold standard measurement and will find out whether there will be a strong correlation between both instruments.

METHODS
Forward-translation and back-translation of the S-TOFHLA
In order to achieve a Turkish version conceptually equivalent to the English S-TOFHLA, we used a forward and back-translation procedure (Sperber et al., 1994). The original S-TOFHLA consists of four questions tapping numeracy and 36 items testing the understanding of medical terms (Baker et al., 1999). It presents two texts in which important meaning-carrying words are left blank and offers to respondents four options to choose from for each blank. The texts and the options were translated by a philologist and native speaker of Turkish following the standard methodologies for questionnaire translation. All 36 items were translated. In some cases, the places of the blanks had to be changed owing to the fact that language structures are completely different. English sentence structure is subject–verb–object, while Turkish follows a subject–object–verb pattern. However, the number of blanks did not change. Particular emphasis was given to the conceptual equivalent of words and phrases instead of providing a literal translation. The translation aimed to be as clear and concise as possible, aiming to address the average patient who is able to understand basic expressions. Technical terms as well as any jargon were avoided. The questionnaire was back-translated by another philologist fluent in English to see whether differences between the original English and the Turkish version would arise. Special attention was paid to the back-translation in terms of grammatical criteria and semantics. Discrepancies regarding single words and expressions between the two translations were discussed by the translators and clarified. Furthermore, cultural adaptation of the context was taken into consideration during the whole translation process, and some idiomatic expressions regarding health issues were adapted to Turkish. Some minor changes
were implemented owing to differences in the Turkish healthcare system. The translation excluded the four numeracy items of the original S-TOFHLA. Once the Turkish version was realized, a pre-test with 120 participants was conducted, using cognitive interviewing. As a result, some wording was changed resulting in the final version of the instrument.

**Study population and questionnaire**

Data collection took place between May 30th and November 25th 2013 in two diabetes clinics in one of the major cities in Turkey. This study was conducted in collaboration with the Turkish Diabetes Foundation, so the two foundation-owned clinics were used to collect the data. Self-administered paper-pencil questionnaires were distributed to eligible outpatients who had a clinic appointment. All the participants had to be diagnosed with type 1 or type 2 diabetes, and also be 18 years or older. In total, 302 diabetes outpatients participated in the study. Patients with impaired vision were excluded (20 patients). The study was conducted in collaboration with TUBITAK (Scientific and Technological Research Council of Turkey) and also approved by a committee from this institution.

Questionnaires were distributed and collected by two collaborators trained in confidentiality, recruitment and data collection procedures. The purpose of the study was explained to the patients and they were asked for their consent. After obtaining oral consent, two collaborators asked the patients to fill in the questionnaire including S-TOFHLA, the Chew screening questions, a diabetes knowledge test and demographics questions before the doctoral meeting. Some patients were permitted to answer the questionnaire after seeing the doctor.

The reading section of S-TOFHLA includes two parts with altogether 36 items. One part tests the reading comprehension with an example of getting prepared for an upper gastrointestinal screening; the second is about the patient’s rights and responsibilities. The former is written for a 4th grade level (normally a child aged 10), the latter for 10th grade level (a teenager aged 16).

The S-TOFHLA questionnaire has a 7-min time limit, which, however, was not enforced for part of the sample. Otherwise, people who overran the time were stopped, but many of the patients mistook the questionnaire as a test and asked for their score. Some of them stated that the questionnaire was very much like a quiz game and they wanted to go on. They were allowed to do so, but the items completed after the 7-min limit were, as a rule, not noted and not counted for the S-TOFHLA score. Due to exceptions from that rule and the decision not to enforce the time limit in a part of the sample, 43% of our respondents completed the S-TOFHLA without an effective time limit, while for the rest the 7-min limit was applied. The two ways of applying S-TOFHLA of course produce different score distributions and have to be analyzed separately. Patients who started to fill in the questionnaire but changed their mind afterwards were excluded.

After filling in S-TOFHLA, patients answered the four Chew screening questions:

1. How often do you have someone help you read hospital materials?
2. How confident are you filling out medical forms by yourself?
3. How often do you have problems learning about your medical condition because of difficulty understanding written information?
4. How often do you have problems understanding what is told to you about your medical condition?

The Chew screening questions were translated using a forward and back-translation procedure similar to the S-TOFHLA translation process. Diabetes Knowledge Test (Fitzgerald *et al.*, 1998) was also translated into Turkish with the same translation process. It consists of 20 questions for diabetes patients who use and do not use insulin. There are three options for each question (yes/no/don’t know) and patients obtain one point for each correct answer and no point for the ‘don’t know’ and false answers.

Patients completed demographic inventories last and delivered the questionnaire to the collaborator. A total score for the S-TOFHLA was formed by summing up the correct answers. It can run theoretically from 0 to 36, and the distribution covered that total range. The mean score was $M = 30.2$ (SD = 6.8, $N = 135$) for the subsample without time limit and $M = 17.3$ (SD = 11.2, $N = 167$) for the subsample that stopped after 7 min. For the Chew measure, the second item was reversed and an average score computed for every respondent. The theoretical and empirical range is 1–5 ($M = 4.1$, SD = 0.74).

For validation, these measures were correlated with education, age, years since diagnosis, income and condition-specific knowledge. Age was coded in real values; education, years since diagnosis and income were measured in five, respectively, four (income) ordered categories, which are treated in the analysis as ordinal scales. Diabetes knowledge was measured with 10 true/false items; the number of correct answers was summed up. Four of these items were related to insulin therapy; they were asked only of patients receiving insulin therapy. Thus, there are two knowledge measures, one ranging from 0 to 6, covering general aspects of diabetes and...
based on all respondents, and the other one ranging from 0 to 10, additionally covering insulin knowledge and based only on patients who took insulin.

In a second step, patients were categorized into three mutually exclusive groups, similar to the original S-TOFHLA: inadequate health literacy with scores of 0–16, marginal health literacy with scores of 17–22 and finally adequate health literacy with scores of 23–36.

**Data analysis**

Data were analyzed using IBM SPSS 21. Internal consistency for the 36 items in the S-TOFHLA was determined using Cronbach’s alpha. In a second step, the newly translated measures were related to some known predictors, mostly socio-demographic. The correlations were assessed by Kendall’s tau-b correlation coefficient. Pearson’s r was not used because many of the data are not normally distributed and most do not reach beyond the ordinal level of measurement. Spearman’s rho was not used because there were very few ranks for the many cases. Results, however, hold for either of these two coefficients. Moreover tau-b is smaller than the other two in all analyses; using it is therefore a conservative method. Thirdly group classifications according to S-TOFHLA scores and predictors were related by ANOVA or cross tabs with $\chi^2$-test.

**RESULTS**

The sample is diverse and spreads well across gender, age groups, education groups and income. Table 1 shows the distribution of these variables in the two subsamples, showing that the subsample with an effective time limit in applying S-TOFHLA had women overrepresented.

The Turkish S-TOFHLA showed high internal consistency. A reliability test of all 36 items revealed a high internal consistency (alpha = 0.97), no single item was below 0.968. This is the same value as for the original S-TOFHLA version (Baker et al., 1999).

Comparing the two newly translated instruments with known predictors of health literacy provides hints about construct validity. The two measures themselves are correlated at a fair and highly significant level of tau = 0.21, relating to the subsample with a 7-min time limit. Both measures correlate as expected with age (negatively), education and income.

Correlations of the health literacy measures with knowledge are also signed positively, but reveal a rather complicated picture. General diabetes knowledge correlates with the screening measure only, while the combined measure of diabetes and insulin knowledge as applied to patients who take insulin shows a significant correlation only with S-TOFHLA. The only assumed predictor that works against expectations is years since diagnosis. We had expected that a longer time of living with the condition improved patients’ communicative abilities, but that is obviously not the case.

The S-TOFHLA measures correlate more strongly than the Chew screening index with age and education, while the screening measure appears to correlate more strongly with income. Chew et al. found that the predictive power of the single items was as large as that of the total scale. This is reason enough to look at the correlations of predictors with a single items (also in Table 2). Coefficients for single items tend to be a somewhat smaller than for the

**Table 1: Socio-demographic characteristics of sample**

<table>
<thead>
<tr>
<th></th>
<th>Subsample without time limit in S-TOFHLA measure (n = 131)</th>
<th>Subsample with 7-min time limit in S-TOFHLA measure (n = 161)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age, mean (SD)</td>
<td>47.4 (13.01)</td>
<td>51.6 (14.24)</td>
</tr>
<tr>
<td>Gender, female</td>
<td>49.0%</td>
<td>65.3%</td>
</tr>
<tr>
<td>Education (%)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5th grade (elementary school)</td>
<td>13.3</td>
<td>8.4</td>
</tr>
<tr>
<td>6th–8th grade (secondary school)</td>
<td>8.9</td>
<td>9.0</td>
</tr>
<tr>
<td>9th–11th grade (high school)</td>
<td>34.1</td>
<td>41.9</td>
</tr>
<tr>
<td>University</td>
<td>42.2</td>
<td>40.7</td>
</tr>
<tr>
<td>Marital status, percent married</td>
<td>76.3</td>
<td>79.6</td>
</tr>
<tr>
<td>Income in Turkish Lira</td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;775</td>
<td>5.2</td>
<td>8.4</td>
</tr>
<tr>
<td>776–1500</td>
<td>26.7</td>
<td>24.0</td>
</tr>
<tr>
<td>1501–2500</td>
<td>26.7</td>
<td>27.5</td>
</tr>
<tr>
<td>&gt;2500</td>
<td>32.6</td>
<td>37.7</td>
</tr>
</tbody>
</table>
whole scale, but many are significantly related with the predictors.

For practical purposes, S-TOFHLA scores are conventionally categorized into three groups: inadequate (0–16), marginal (17–22) and adequate (23–36) health literacy (Baker et al., 1999). According to this categorization, our sample of respondents with the time limit in operation produces 49% people with inadequate, 14% with marginal, and 37% with adequate health literacy. This means a considerably higher share of persons with inadequate health literacy than found in most other studies from different countries. Looking at the same predictors as above, we find that age is again linked with health literacy; those with inadequate literacy are significantly older with an average of 56.6 years than the other two literacy groups ($F = 11.245$, df = 2, 161, $p < 0.001$). The post-hoc Scheffe test revealed that the significant differences were between the inadequate and the other two groups, while the marginal and adequate groups do not differ from one another.

A cross tabulation of education and health literacy groups shows a steady decline of the share of inadequately health literate persons from 93% among persons with only elementary school to 25% among those with a postgraduate degree. The relationship is highly significant ($\chi^2 = 26.373$, df = 9, $p < 0.001$). The significant correlation between income and health literacy cannot be repeated in this analysis ($\chi^2 = 9.686$, df = 6, $p = 0.14$) although the differences are in the expected direction. Gender ($\chi^2 = 0.096$, df = 3, $p = 0.97$) and marital status ($\chi^2 = 1.174$, df = 2, $p = 0.56$) do not have any influence on health literacy. By and large the results of the group comparisons therefore confirm the earlier correlation analysis.

DISCUSSION

The Turkish translation of S-TOFHLA was shown to be a reliable instrument. The links between the instrument and known predictors of health literacy provide evidence of its construct validity. The translation of the Chew screening questions also emerges as a valid instrument. Both can be applied in Turkey or among the Turkish population in other countries.

Using the conventional definitions of inadequate and marginal health literacy by cutoffs in the S-TOFHLA score produces a higher share of people with inadequate health literacy than in most other studies from other countries (Baker et al., 1999; Baron et al., 2007; Tang et al., 2007; Carthery-Goulart et al., 2009; Jovic-Vranes et al., 2009; Connor et al., 2013). That can be explained by the different education system in Turkey compared with most western industrialized countries. Until 1997, children in Turkey were obliged to take but 5 years of education in Turkey, far less than obligatory education in most western countries. Education reforms since extended compulsory education to 8 years in 1997 and to 12 years in 2012 (Milli Eğitim Bakanlığı, 2012). It will of course take time until these reforms improve the education level of the total population. Most participants in our study grew up at a time when obligatory education and health education levels were low, which might explain the large share of people with inadequate health literacy and especially the high average age of the group with inadequate health literacy (56.6 years). Furthermore, current data about education level show that 28% of people are primary school graduates and just 11% have higher

### Table 2: Correlation of Turkish S-TOFHLA and screening questions with known predictors

<table>
<thead>
<tr>
<th></th>
<th>Chew screening measure</th>
<th></th>
<th></th>
<th></th>
<th>S-TOFHLA</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Complete scale</td>
<td>Range among four single items</td>
<td>Number of items reaching p &lt; 0.05</td>
<td>Without time limit</td>
<td>With 7-min limit</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chew screening scale</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
<td>3</td>
<td>0.15*</td>
<td>0.21***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>−0.12**</td>
<td>−0.06/−0.12</td>
<td>3</td>
<td>$−0.27^*$</td>
<td>$−0.25^{***}$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Years since diagnosis</td>
<td>−0.09*</td>
<td>−0.03/−0.11</td>
<td>3</td>
<td>$−0.06$ (ns)</td>
<td>$−0.08^a$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education</td>
<td>0.18***</td>
<td>0.08/0.20</td>
<td>3</td>
<td>0.41**</td>
<td>0.29**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Income</td>
<td>0.23***</td>
<td>0.14/0.26</td>
<td>4</td>
<td>0.15*</td>
<td>0.12*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes knowledge</td>
<td>0.08*</td>
<td>0.05/0.07</td>
<td>–</td>
<td>0.02 (ns)</td>
<td>0.02 (ns)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diabetes and insulin knowledge</td>
<td>0.08 (ns)</td>
<td>0.06/0.07</td>
<td>–</td>
<td>0.19*</td>
<td>0.14^a</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: Coefficients are Kendall’s tau-b.

`p < 0.10.

^p < 0.05.

**p < 0.01.

***p < 0.001.`
education in Turkey (TUIK, 2012). Considering that the level of health literacy is strongly correlated with education, lower health literacy in Turkey would not be a surprising result.

These considerations support the assumption that health literacy is indeed lower in Turkey than in other, mostly western industrialized countries. But our results can also mean that the scale and especially the cutoff values need to be calibrated for use in Turkey. With the data at hand, it cannot be decided which interpretation is correct. Defining new cutoff points in order to achieve a more realistic classification of Turkish people into the groups of inadequate, marginal and adequate health literacy would shield the Turkish findings from comparison with other countries. That cannot be the intention of work on health literacy, and it is not the intention of the research reported in this article. But still we need to know whether the conventional cutoff points produce meaningful classifications in Turkey before using them in actual research.

The unexpected results with regard to years since diagnosis deserve a second look. Years since diagnoses were not correlated positively with health literacy as expected. It is relatively strongly correlated with age, and age has a negative impact on health literacy in our measures as well as other studies (Baron-Epel et al., 2007; Von Wagner et al., 2007; Jovic-Vranes et al., 2009; Xining et al., 2013). But even if the effect of age is controlled and partial correlations of years since diagnosis and health literacy are computed, the coefficients, all non-significant, remain negative. Moreover, there is no correlation between years since diagnosis and general diabetes knowledge, meaning that the experience patients make with diabetes does not translate into better knowledge of the condition and neither in improved health literacy. If experience did have this effect, it would be stronger the longer a patient had to deal with the condition. As no such correlations emerged, we can reject the idea of an effect of experience on knowledge or literacy. One explanation of the absence of effects of long experience with the condition on health literacy or condition-specific knowledge could be that Turkish patients are not as much involved in health decision making as is usual in other countries, maybe because the ideals of shared decision-making and patient empowerment are not rooted as strongly in Turkey as in other countries where relevant research was conducted. Some literature seems to support this (Glanz et al., 2009; Jovic-Vranes et al., 2009; Omer et al., 2009; Smith et al., 2011).

The different measures of S-TOFHLA in the two subsamples correlate rather similarly with the predictors listed in Table 2. This indicates that all conclusions with regard to construct validity pertain to both measures, the standard one with a 7-min time limit as well as the same measure without a time limit. This indicates that the S-TOFHLA questionnaire can also be used without a time limit provided the results are not compared with other studies that enforced the time limit.

CONCLUSIONS
This paper reports on the validation of Turkish versions of S-TOFHLA and the Chew scale. Both measures can be applied in Turkey and other countries which have large Turkish population. However, the conventional cutoff points of S-TOFHLA scale need to be reconsidered; to date it is unclear whether they represent a meaningful classification in Turkey.

LIMITATIONS
The open questions discussed above are among the limitations of this study. This refers to the question of whether the conventional cutoff points for defining health literacy groups are applicable in Turkey and to the extent to which recent ideals in healthcare such as shared decision-making and patient empowerment are rooted within the Turkish medical system. That we did not validate the translated instruments to an external measure, like for example a reading comprehension test, constitutes another limit of this study.

PRACTICAL IMPLICATIONS
The validated Turkish S-TOFHLA will allow conducting studies in Turkey in medical care and prevention, comparing results with other countries and identifying people with limited literacy whose disease management skills might be affected (e.g. HIV medication management or glucose monitoring for patients with diabetes). In countries with large Turkish communities, it will become possible to accurately assess the level of health literacy among Turkish minorities and develop strategies to improve health outcomes, decrease health disparities, enhance health status and achieve a high quality of life.

REFERENCES


Smith P. J., Humiston S. G., Marcuse E. K., Zhao Z., Dorell C. G., Howes C., et al. (2011) Parental delay or refusal of vaccine doses, childhood vaccination coverage at 24 months of age,
and the Health Belief Model. Public Health Reports, 126, 135–181.