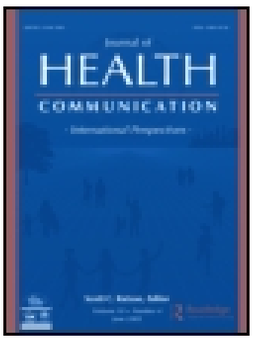

Taylor & Francis
Taylor & Francis Group

Journal of Health Communication
International Perspectives

ISSN: 1081-0730 (Print) 1087-0415 (Online) Journal homepage: http://www.tandfonline.com/loi/uhcm20

Validity of Three Brief Health Literacy Screeners to Measure Functional Health Literacy – Evidence from Five Different Countries

Sarah Mantwill, Ahmed Allam, Anne-Linda Camerini & Peter J. Schulz

To cite this article: Sarah Mantwill, Ahmed Allam, Anne-Linda Camerini & Peter J. Schulz (2018): Validity of Three Brief Health Literacy Screeners to Measure Functional Health Literacy – Evidence from Five Different Countries, Journal of Health Communication, DOI: 10.1080/10810730.2017.1417515

To link to this article: https://doi.org/10.1080/10810730.2017.1417515

Published online: 10 Jan 2018.

Journal of Health Communication, 00: 1–9, 2018
Published with license by Taylor & Francis
ISSN: 1081-0730 print/1087-0415 online
DOI: https://doi.org/10.1080/10810730.2017.1417515

Validity of Three Brief Health Literacy Screeners to Measure Functional Health Literacy – Evidence from Five Different Countries

SARAH MANTWILL[1], AHMED ALLAM[2], ANNE-LINDA CAMERINI[3], and PETER J. SCHULZ[3]

[1]*Department of Health Sciences & Health Policy, University of Lucerne,Lucerne Switzerland*
[2]*Department of Pathology, Yale University School of Medicine, New Haven, CT, USA*
[3]*Institute of Communication and HealthUniversità della Svizzera italiana,Lugano Switzerland*

Self-reported health literacy measures have seen increased application throughout the last years, among those are the brief health literacy screeners (BHLS) developed by Chew and colleagues (2004). There has been little systematic research on the performance of these measures across different contexts, including countries and languages, to draw conclusions about their predictive power outside of the United States.

This study aimed at replicating the original validation of the BHLS. Receiver operating characteristic (ROC) analysis was applied to data from Hungary, Italy, Lebanon, Switzerland, and Turkey. In addition, logistic regression models incorporating ROC analysis using BHLS as predictors were compared to models using socio-demographics only to identify individuals with inadequate and inadequate or marginal health literacy as measured with the Short Test of Functional Health Literacy in Adults.

Analyses showed that in all cases the BHLS were not sufficiently able to identify individuals with different health literacy levels. Logistic regression models using socio-demographics only as predictors outperformed models using the BHLS.

The findings highlight the limitations of using the BHLS outside the United States. Further, they question in how far self-reported health literacy measures are comparable across different contexts and whether thresholds for different health literacy levels are universally applicable.

In 2004, Chew and colleagues introduced three Brief Health Literacy Screeners (BHLS) to identify people with inadequate levels of functional health literacy in clinical settings (Chew, Bradley, & Boyko, 2004; Chew et al., 2008). To that date, existing measures either took a lot of time for patients to fill in, such as the Test of Functional Health Literacy in Adults (TOFHLA) and its short version (Baker, Williams, Parker, Gazmararian, & Nurss, 1999; Parker, Baker, Williams, & Nurss, 1995), or required an interviewer for assessment, as in the case of the Rapid Estimate of Adult Literacy in Medicine (REALM) (Davis et al., 1993). Chew and colleagues (Chew et al., 2004) aimed to develop a more practical screening measure that allowed identifying individuals with inadequate functional health literacy more easily, while not taking up too much time and reducing the burden on the patient´s and the provider´s side. Starting with originally 16 items (Chew et al., 2004), the authors eventually identified three items that each individually was able to sufficiently predict inadequate health literacy (Chew et al., 2008).

Address correspondence to Sarah Mantwill, Frohburgstrasse 3, P.O. Box 4466, Lucerne 6002, Switzerland. Tel: +41 41 229 5945; E-mail: sarah.mantwill@unilu.ch

Originally developed for the clinical setting, the three BHLS have gained increasing popularity in population-based surveys (Bishop et al., 2016; Fernandez, Larson, & Zikmund-Fisher, 2016; Sentell, Baker, Onaka, & Braun, 2011). Moreover, their application has spilled over from the United States, where they were originally developed, to other countries and language regions (Altin & Stock, 2015; Fransen, Schaik, Van, Twickler, & Essink-Bot, 2011; Geboers, Reijneveld, Jansen, & De Winter, 2016). This has raised the question of whether the BHLS are indeed valid measures of functional health literacy in other countries than the United States, and in particular whether they have the same potential to identify people with inadequate health literacy. This study intends to answer this question by replicating the original validation procedures used by Chew and colleagues (Chew et al., 2004, 2008) using data from five different countries in Europe and the Middle East.

Validation of the Brief Health Literacy Screeners in the United States

Chew and colleagues (Chew et al., 2004) initially developed 16 self-reported items that inquired about patients´ confidence or frequency in handling different situations during the medical

encounter. Response options were scaled from 0 ("extremely" when asked to rate confidence, and "always" when asked to report on frequency) to 4 ("not at all", respectively "never"). Thematically, the items addressed issues such as doctor–patient communication, navigation of healthcare institutions, as well as the handling and understanding of medical forms or medication regimens (Chew et al., 2004). The original 16 screening items were tested among English-speaking veterans (n = 332) in a VA preoperative clinic in Seattle, United States. After the identification of three items that showed to be most predictive of inadequate and marginal health literacy, a second study was conducted to validate those items. For the second study 1796 face-to-face interviews at four VA medical centers in the United States were conducted (Chew et al., 2008).

The authors assessed the validity of the items by testing how well they predicted cases of inadequate, respectively the combination of inadequate or marginal health literacy, as determined by the short version of the TOFHLA (S-TOFHLA) (Baker et al., 1999; Parker et al., 1995) and the REALM (Davis et al., 1993). For the purpose of this study, we will only focus on the description and reported results for the S-TOFHLA.

The S-TOFHLA consists of two reading-comprehension parts, with overall 36 cloze items and respondents are asked to choose one out of four words to fill in the respective gaps. Additionally, the test includes four numeracy items. Participants are shown four prompts and have to respond to related questions. Whereas the reading-comprehension parts have a time limit of altogether 7 min, there is no time limit for the numeracy items (Baker et al., 1999). The original scoring of the S-TOFHLA ranges from 0 to 100 but many studies have adopted a simple scoring frame that only includes the two reading-comprehension parts and assigns a score of one to each correct answer (0–36) (Aguirre, Ebrahim, & Shea, 2005; Cho, Lee, Arozullah, & Crittenden, 2008; Friedman, Corwin, Dominick, & Rose, 2009).

Both, the S-TOFHLA and the BHLS differ from each other not only in their administration procedure but also, and primarily, in their informative value. The S-TOFHLA is a performance-based measure, while the BHLS are self-reported. In order to assess if and to what extent people understand the meaning of medical expressions and instructions, performance-based measures record observed reactions as respondents apply their health literacy skills to a particular task. In the case of the S-TOFHLA, for example, this includes filling in close items and responding to numeracy items in the medical context. In contrast, self-reported measures ask respondents to report on themselves on, for example, the frequency of a specific situation they might come across during a medical encounter (Kiechle, Bailey, Hedlund, Viera, & Sheridan, 2015). For instance, one of the three BHLS asks respondents how often they need assistance in filling out medical forms. One of the advantages of self-reported measures is that they are considered to create less embarrassment for the respondent, as they do not directly inquire about knowledge and behavior (Chew et al., 2008; Kiechle et al., 2015; Wolf et al., 2007). In addition, self-reported measures tend to be overall shorter than performance-based measures and provide the means for more efficient screening, in particular in the clinical setting (Kiechle et al., 2015). Further, the S-TOFHLA is thematically limited by the subject matter of the two reading-comprehension parts (Passage A: instructions and information about an X-ray; Passage B: medical informed consent), whereas the BHLS do not have such thematic limitation.

In order to evaluate the validity and discriminative power of the original 16 screening items, Chew and colleagues (Chew et al., 2004, 2008) conducted a receiver operating characteristic (ROC) analysis (see below). Based on the analysis, the authors (Chew et al., 2004, 2008) selected three items that performed best to identify people with inadequate health literacy according to their score on the S-TOFHLA. The selected items also performed comparatively well in predicting the combination of people with either inadequate or marginal health literacy, although this combination was generally less well predicted than inadequate health literacy only. The three retained items were: "How confident are you filling out forms by yourself?" (abbreviated as "Confident with Forms"), "How often do you have someone (like a family member, friend, hospital/clinic worker, or caregiver) help you read hospital materials?" ("Help Read"), and "How often do you have problems learning about your medical condition because of difficulty reading hospital materials?" ("Problems Reading") (Chew et al., 2004).

Study Objectives

In order to test for the potential validity of the three BHLS in other contexts than the United States, we performed secondary data analysis on datasets from studies conducted in five different countries (Hungary, Italy, Lebanon, Switzerland, and Turkey), comprising overall nine different languages. All studies included the S-TOFHLA and the BHLS in other languages than English.

Method

Samples and Recruitment Procedure

In Hungary 302 participants were recruited in different locations with the aim to collect a fairly representative sample of the Hungarian population. Among others, data were collected in public places, such as hospitals, retirement homes, and coffee shops (Papp-Zipernovszky et al., 2016).

Data for the study in Italy were collected among diabetes patients in an inner city hospital in Milan. 218 participants were included for final data analysis.

In Lebanon study participants were recruited among patients who had a clinical appointment at the Division of Gastroenterology at the Medical Center of the American University of Beirut. Overall, 230 participants were included for final data analysis (Fadda et al., 2016).

For the Swiss study data were collected among the general population in the German- and Italian-speaking parts of Switzerland. Besides the inclusion of German- and Italian-speaking native Swiss, also first generation Albanian-, Portuguese- and Serbian-speaking immigrants were included in the study. Using snowball sampling 1146 participants were included in the final data analysis (Mantwill and Schulz, 2017).

Data in Turkey were collected in a sample of 302 diabetes patients in two diabetes clinics in Istanbul. Participants were outpatients diagnosed with diabetes type 1 or type 2 who came for a clinic appointment. For this study data were analyzed for 161 participants who had a valid time recording (Eyüboğlu and Schulz, 2015).

In the case of the Hungarian, Lebanese, and Turkish samples, data came from the original validation studies that aimed at validating the S-TOFHLA in the respective languages. On the other hand, data for the Italian and Swiss samples came from follow-up studies, using already validated tools. In all cases the S-TOFHLA had been translated from the original English version into the respective languages and back-translated into English as part of the linguistic validation process (see e.g. Beaton, Bombardier, Guillemin, & Ferraz, 2000). Further, all translations were reviewed by expert committees to account for differences in culture and healthcare systems. Issues were resolved through consensus finding (Eyüboğlu and Schulz,2015; Fadda et al., 2016; Mantwill and Schulz, 2017; Papp-Zipernovszky et al., 2016). In all cases the S-TOFHLA significantly correlated with common predictors of health literacy, including age, education, and income. Further, all measures showed high reliability with Cronbach's alphas of above .90.

Table 1 provides an overview of the demographic structure of the samples, including the samples used in the studies conducted by Chew et al. (2004) and Chew et al. (2008).

Measures

For the purpose of this study, we only included the three BHLS and the S-TOFHLA, as well as comparable measures on socio-demographics, such as age, gender, and educational background.

Brief Health Literacy Screeners

For each dataset the translated versions of the three BHLS were used in form of single items, where higher scores indicated lower functional health literacy. The wording of the items reported in Chew and colleagues' publication in 2004 slightly varied from the study published in 2008. Whereas the wording for "Help Read" remained the same, the two other items were slightly changed. "Problems Reading" was changed in so far as it was specified for what type of written material help was needed ("hospital material"). On the other hand, the item "Confident with Forms" became more general by changing the wording from "medical forms" to "forms" only. Accordingly wording of items included in this analysis was slightly different from one dataset to another, depending on which publication was used as a reference by the authors of the respective studies. In the case of the Turkish study, for instance, the authors referred to the original study (Chew et al., 2004), whereas in the case of the Hungarian study authors referred to the study from 2008. Further, in the case of the Swiss study, items for participants with an immigration background were formulated slightly different. In order to account for the Swiss healthcare system in comparison to immigrants' home countries' systems, it was specified that information would be either in German or Italian, depending on in which part of Switzerland participants lived (ex. "How often do you have someone (like a family member, friend, hospital/clinic worker, or caregiver) help you read hospital materials in German?").

S-TOFHLA

In their original study with 16 items, Chew and colleagues (Chew et al., 2004) did not include the numeracy items of the S-TOFHLA and used the scoring system from 0 to 36. According to this scoring a score of 16 or lower is interpreted as inadequate health literacy, and a score of 17 to 22 as marginal health literacy. Any score above is considered to be adequate health literacy. However, in their second study the authors also included the numeracy items and used the scoring system from 0 to 100 (Chew et al., 2008). For the purpose of this study, the original scoring from 0 to 36 (excluding the numeracy items) was used. The distribution of participants across the three levels of functional health literacy is reported in Table 1.

Data Analysis

Data for this study were analyzed using R version 3.0.2 (R Core Development Team, 2015) using GLM and pROC packages (Robin et al., 2011). Data analysis consisted of three steps. In

Table 1. Sample characteristics

	US Chew et al., 2004 (*N* = 332)		US Chew et al., 2008 (*N* = 1796)		Hungary (*N* = 302)		Italy (*N* = 218)		Lebanon (*N* = 230)		Switzerland (N = 1146)		Turkey (N = 161)	
	N	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*	*N*	*%*
Age group 1	60	*18*	537	*30*	145	*48*	47.5 (15.9)		37.5 (12.5)		40.4 (13.8)		51.6 (14.2)	
Age group 2	191	*49*	748	*42*	112	*37*								
Age group 3	111	*33*	511	*28*	45	*15*								
Female	18	*5*	246	*14*	160	*53*	128	*59*	122	*53*	606	*53*	105	*65*
> High School	158	*48*	1174	*65*	85	*28*	58	*27*	169	*74*	227	*24*	66	*41*
S-TOFHLA														
Inadequate health literacy	15	*4.5*	123	*6.8*	25	*8.3*	16	*7.3*	22	*9.6*	93	*8.1*	79	*49*
Marginal health literacy	25	*7.5*	132	*7.4*	18	*6.0*	11	*5.0*	22	*9.6*	113	*9.9*	23	*14*
Adequate health literacy	292	*88*	1408	*78.4*	257	*85.7*	191	*87.6*	186	*80.8*	940	*82*	59	*37*

Note. Percentages for socio-demographic characteristics rounded to the nearest integer; Age groups US 2004: 1 "18–45", 2 "46–64", 3 "65+"; US 2008: 1 "< 50", 2 "50–75", 3 "> 75"; Hungary: 1 "18–45", 2 "46–65", 3 "> 65"; Italy, Lebanon, Switzerland, Turkey: 1 sample mean age (SD); S-TOFHLA Turkey: sample size for each category calculated based on reported overall *N* and percentages.

a first step, to gain a better understanding of the similarity between the scales, correlations between the three BHLS and the S-TOHLA were calculated for each dataset.

In a second step, ROC analysis was performed. ROC analysis was originally developed to evaluate the accuracy in differentiating signal from noise in radar detection (Hajian-Tilaki, 2013; Swets, 1986). Nowadays, in the medical and clinical field, the analysis plays an important role in evaluating the performance of screening and diagnostic tests to depict a preexisting dichotomy, such as the presence or absence of a medical condition (Lloyd, 1998; Pepe, 2003; Zou, O'Malley, & Mauri, 2007). ROC analysis allows to determine the quality of a test to sort positive and negative cases on a condition or feature, which, in turn, is determined by another established test. In our case, the quality of the three BHLS was determined by its ability to recognize cases of inadequate, respectively inadequate or marginal health literacy according to the established performance-based S-TOFHLA. ROC analysis works by plotting the true positive rate (sensitivity) of the test (i.e. BHLS) against its false positive rate (1-specificity) for different thresholds/cutoff points. Sensitivity is the share of de facto included cases among all that should be included, and specificity is the share of de facto excluded cases among all that should be excluded. Sensitivity is the capability to recognize positive and specificity the capability to recognize negative cases. The quality indicator is the area under the curve (AUC). For a good test the curve rises steeply close to the y-axis, and then bends and runs close to the upper margin of the graph, and the area under the curve approaches a value of 1. For a bad test, the curve runs close to the diagonal and the area under the curve gets close to 0.5. Therefore an AUC of 0.5 or close to it indicates poor predictive power of the test. To be certain that the test has any predictive power, the 95% confidence interval should not reach lower than 0.5. Values for AUC larger than 0.7 are considered acceptable (Pines, Carpenter, Raja, & Schuur, 2012).

The third step consisted of fitting logistic regression models using all three BHLS as independent variables to predict participants with inadequate, respectively inadequate or marginal health literacy. By fitting a logistic regression using the BHLS we aimed at identifying the optimal coefficients to better predict participants' health literacy level according to the S-TOFHLA score. Again we assessed the performance of the logistic regression models using the area under the ROC curve by employing ROC analysis where the decision boundary of the model was varied from 0 to 1 with increments of 0.1. In both stages, the classifiers were built for every dataset separately and their performance was compared to the ones reported in Chew et al. (2004).

In a last step, we built classifiers by fitting logistic regression models using participants' socio-demographic characteristics (i.e., age, gender, and educational background) as a set of independent variables. Following this approach, we were able to compare the performance of socio-demographic characteristics as screeners to the performance of the BHLS.

In summary it was expected that if the BHLS were successful screeners, we would find (1) an acceptable AUC level (>0.7) or at least a comparable performance to the ones reported in Chew et al. (2004) and (2) a better or at least equal performance of the BHLS compared to a model that would only use socio-demographic characteristics as classifiers.

Results

Correlations

Overall, the correlations between the BHLS and the S-TOFHLA scores were relatively low (Table 2). Moreover, with regard to the item "Help Read", the data from Switzerland and Lebanon showed inconsistencies with regard to the expected directionality of the correlations.

ROC Analysis

We assessed the performance of each BHLS separately by employing ROC analysis to identify participants with inadequate, as well as inadequate or marginal health literacy combined as determined by the S-TOFHLA (Table 3).

Comparing the AUC values of all datasets showed that the raw screening scores performed considerably worse outside the United States, both in identifying people with inadequate and inadequate or marginal health literacy combined. Best AUC values were found for the Hungarian dataset.

Logistic Regressions Incorporating ROC Analysis

In a next step, all three BHLS were used as independent variables in a logistic regression model to test if this model was a better predictor of performance-based health literacy levels. As can be seen in Table 4, this procedure achieved comparable and

Table 2. Correlations between S-TOFHLA and brief health literacy screeners

	Hungary (N = 302)	Italy (N = 218)	Lebanon (N = 230)	Switzerland (N = 1146)	Turkey (N = 167)
BHLS 1: How confident are you filling out forms by yourself? (Confident with Forms)	−0.31***	−0.08	−0.25***	−0.13***	−0.27***
BHLS 2: How often do you have someone help you read hospital materials? (Help Read)	−0.21***	−0.14*	0.36***	0.14***	−0.24**
BHLS 3: How often do you have problems learning about your medical condition because of difficulty reading written information? (Problems Reading)	−0.19**	−0.24***	−0.17**	−0.15***	−0.17***

Note. BHLS1 Switzerland, Turkey: How confident are you filling out medical forms by yourself?; BHLS3 Hungary: How often do you have problems learning about your medical condition because of difficulty understanding hospital material?; * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 3. Areas under the receiver operating characteristic curve and 95% CI for BHLS raw scores

	US Chew et al. (2004) (N = 332)	Hungary (N = 302)	Italy (N = 218)	Lebanon (N = 230)	Switzerland (N = 1146)	Turkey (N = 167)
Identification of persons with inadequate health literacy (S-TOFHLA 0–16)						
BHLS 1 (Confident with Forms)	0.80 (0.67–0.93)	0.69 (0.58–0.81)	0.50 (0.34–0.66)	0.51 (0.39–0.63)	0.61 (0.55–0.67)	0.64 (0.57–0.72)
BHLS 2 (Help Read)	0.87 (0.78–0.96)	0.65 (0.52–0.77)	0.44 (0.28–0.61)	0.40 (0.30–0.51)	0.61 (0.55–0.67)	0.63 (0.55–0.71)
BHLS 3 (Problems Reading)	0.76 (0.62–0.90)	0.61 (0.49–0.73)	0.56 (0.37–0.75)	0.53 (0.42–0.64)	0.58 (0.52–0.64)	0.60 (0.52–0.68)
Identification of persons with inadequate or marginal health literacy (S-TOFHLA 0–22)						
BHLS 1 (Confident with Forms)	0.66 (0.57–0.76)	0.67 (0.58–0.76)	0.55 (0.43–0.66)	0.54 (0.44–0.63)	0.57 (0.52–0.61)	0.56 (0.48–0.64)
BHLS 2 (Help Read)	0.68 (0.60–0.77)	0.61 (0.51–0.71)	0.47 (0.34–0.60)	0.37 (0.28–0.45)	0.57 (0.53–0.62)	0.58 (0.50–0.67)
BHLS 3 (Problems Reading)	0.60 (0.51–0.69)	0.61 (0.51–0.71)	0.63 (0.50–0.75)	0.53 (0.44–0.61)	0.58 (0.54–0.62)	0.57 (0.49–0.66)

Table 4. Areas under the receiver operating characteristic curve and 95% CI for logistic regression with BHLS

	Hungary (n = 302)	Italy (n = 218)	Lebanon (n = 230)	Switzerland (n = 1146)	Turkey (n = 167)
Identification of persons with inadequate health literacy (S-TOFHLA 0–16)					
$logit(IHL) = \beta_0 + \beta_1 * BHLS1 + \beta_2 * BHLS2 + \beta_3 * BHLS3$	0.68 (0.55–0.81)	0.56 (0.38–0.74)	0.63 (0.52–0.75)	0.63 (0.57–0.69)	0.67 (0.59–0.75)
Identification of persons with inadequate or marginal health literacy (S-TOFHLA 0–22)					
$logit(IMHL) = \beta_0 + \beta_1 * BHLS1 + \beta_2 * BHLS2 + \beta_3 * BHLS3$	0.70 (0.60–0.80)	0.63 (0.51–0.76)	0.63 (0.54–0.72)	0.59 (0.55–0.63)	0.60 (0.51–0.68)

Note. IHL: Inadequate Health literacy: coded 1 if person had an S-TOFHLA score (0–16) and 0 otherwise; IMHL: Inadequate or Marginal Health Literacy: coded 1 if person has an S-TOFHLA score and 0 otherwise; Capital letter "literacy": Inadequate Health Literacy; Inadequate of Marginal Health Literacy.
β_0: intercept; β_1: parameter/coefficient associated with independent variable (BHLS 1 Confident with Forms);
β_2: parameter/coefficient associated with independent variable (BHLS 2 Help Read); β_3: parameter/coefficient associated with independent variable.
(BHLS 3 Problems Reading).

even better results (i.e. higher AUC values), compared to using the raw scores of the BHLS as classifiers[1]. However, overall the performance of the classifiers had an upper bound below .70, except for the dataset collected in Hungary.

In a final step, logistic regression models were fitted using baseline socio-demographics (age, gender, education) as independent variables to predict inadequate, respectively inadequate, or marginal health literacy (Table 5).

The performance of the models using socio-demographics as independent classifiers clearly outperformed the models using the three BHLS. Given these results, the three BHLS showed overall poor utility in classifying people according to their health literacy levels based on the S-TOFHLA.

Discussion

The present study aimed to evaluate whether the three BHLS developed by Chew et al. (2004) and Chew et al. (2008) are valid measures of functional health literacy in other contexts than the United States. Item validity was evaluated by applying ROC and logistic regression analysis incorporating ROC analysis to data from five different countries (Hungary, Italy, Lebanon, Switzerland, and Turkey). The results of the analyses are exceptionally clear: the BHLS items, developed as a quick assessment of health literacy in clinical practice, are not (or even inversely) related with one of the most commonly used measures of health literacy, the S-TOFHLA. Not related here means that (a) within each country the two measures did not sufficiently correlate with each other, and (b) the BHLS were not able to single out different individuals as members of the actual problem group as defined by the S-TOFHLA, that is, individuals with inadequate and individuals with inadequate or marginal health literacy. The absence of relationships between the two measures in our datasets stands in contrast to the validation

1.Before including all three BHLS items as classifiers in the logistic regression model, we built three models including each BHLS item as sole classifier. These models performed worse and, therefore, results are not reported here.

Table 5. Areas under the receiver operating characteristic curve and 95% CI for logistic regression with socio-demographics

	Hungary (n = 302)	Italy (n = 218)	Lebanon (n = 230)	Switzerland (n = 1146)	Turkey (n = 167)
Identification of persons with inadequate health literacy (S-TOFHLA 0–16)					
$logit(IHL) = \beta_0 + \beta_1 * Age + \beta_2 * Gender + \sum_{i=1}^{M-1} \beta_{i+2} * Education_i$	0.94	0.89	0.75	0.74	0.78
	(0.89–0.98)	(0.79–0.98)	(0.65–0.85)	(0.69–0.79)	(0.70–0.85)
Identification of persons with inadequate or marginal health literacy (S-TOFHLA 0–22)					
$logit(IMHL) = \beta_0 + \beta_1 * Age + \beta_2 * Gender + \sum_{i=1}^{M-1} \beta_{i+2} * Education_i$	0.91	0.86	0.64	0.71	0.72
	(0.87–0.95)	(0.78–0.93)	(0.55–0.73)	(0.67–0.75)	(0.64–0.80)

Note. IHL: Inadequate Health literacy: coded 1 if person has an S-TOFHLA score (0–16) and 0 otherwise;IMHL: Inadequate or Marginal Health Literacy: coded 1 if person has an S-TOFHLA score and 0 otherwise; Capital letter "literacy": Inadequate Health Literacy; Inadequate of Marginal Health Literacy.

β_0: intercept; β_1: parameter/coefficient associated with independent variable (Age); β_2: parameter/coefficient associated with independent variable (Gender); $\sum_{i=1}^{M-1} \beta_{i+2}$: set of parameters/coefficients associated with the categories/values of the independent variable (Education) where M is the total number of categories/values.

studies presented by Chew and colleagues (Chew et al., 2004, 2008), who concluded that both measures broadly identified the same people with regard to their health literacy levels. Given the results of the present study, we, therefore, conclude that the BHLS are not easily transferable to other countries and might not be reliable measures of health literacy in other cultural and linguistic contexts.

However, we do not believe that our results imply that the BHLS should be easily disposed. First of all, the screening items have practical value as they might help health professionals to identify patients with comprehension problems, at almost no cost. This argument pushes back the screening items to the realm they were originally intended for: the clinical setting. On top of that, the screeners have seen increased application in studies outside of the United States and have found significant associations similar to those that have been reported for the S-TOFHLA or other functional health literacy measures (Fransen et al., 2011; Franzen et al., 2014; Joshi et al., 2013; Lin et al., 2014; Lupattelli, Picinardi, Einarson, & Nordeng, 2014; Tokuda, Doba, Butler, & Paasche-Orlow, 2009; Verkissen et al., 2014; Zeguers et al., 2012). All this has enabled fruitful discussions, which have led to stimulating and interesting new research questions.

A potential explanation for the variability found across studies might be related to the underlying conceptualization and subsequent operationalization of the BHLS. The BHLS have been mostly interpreted as a measure of abilities. Yet, it is also plausible to assume that the items essentially assess an individual's perception of the latitude of her/his own mastery of health decisions (Frisch et al., 2012). This interpretation suggests that the BHLS are not necessarily a measure of health literacy per se but a measure of a different, yet related concept, namely patient empowerment. In the field of health literacy patient empowerment, defined as the belief that oneself possesses mastery over one's health decisions (Schulz and Nakamoto, 2013), has received conceptually and empirically relatively little attention. Yet, Schulz and Nakamoto (2013), for instance, has argued that both concepts should be considered together, as health literacy without empowerment (likewise empowerment without health literacy) may result in diverse health behaviors and outcomes. The assumption that the BHLS tell us more about empowerment than about health literacy might even reach further and apply to other self-reported measures of health literacy, such as the European Health Literacy Survey for example (Sørensen et al., 2013). Further, the literature has widely agreed that health literacy is a multi-dimensional concept (Mårtensson & Hensing, 2012; Sørensen et al., 2012), including dimensions such as critical or communicative health literacy (Nutbeam, 2000), which encompass a variety of different skills (Sørensen et al., 2012). It might well be that the BHLS items under consideration in this study do indeed address different dimensions of health literacy beyond functional skills as measured with the S-TOFHLA, and that these dimensions correlate differently with one another in different contexts.

The latter argument leads back to the more traditional task of discussing research results and to find meaning in them. In the case of this study, this would involve the systematic listing of differences between the different samples and the attempt to find explanations of why and how these differences affect self-reported and performance-based measures similarly or differently. Issues to look at would include, for instance, the composition of the sample populations (i.e. veterans in the case of the United States vs. diabetes patients in Italy and Turkey). Further, one would have to tease out the influence of language and culture, such as the quality of translations and the applicability of the measures to different realities. However, there are several barriers to such an attempt. One is that the examination of these issues would be widely restricted to speculations, as comparable data to draw systematic comparisons are not (yet) available. An additional barrier is the problem of having to explain the non-existence of an expected pattern, in the case of this study the missing correlations between the BHLS and the S-TOFHLA. Since there are numerous factors that might affect the correlation (i.e. the conceptualization and operationalization of health

literacy pertaining to both measures) further in-depth analyses, including qualitative explorations, might be necessary.

Beyond these more general conceptual issues, there might be some methodological issues that pertain directly to the measures themselves. One issue is the lack of correspondence between the two measures. The S-TOFHLA, as a performance-based measure, assumes that it is unlikely that someone would cheat on purpose on it. To be precise: one could only cheat in terms of a negative directionality, meaning someone could hide her/his actual level of health literacy by trying to purposefully score lower. Self-reported (perception-based) measures, on the other hand, can be considered to be much more susceptible to individual and cultural influences, such as social desirability (Bernardi, 2006; Lee, Tsai, & Tsai, 2013) or beliefs about health and illness (Shaw, Huebner, Armin, Orzech, & Vivian, 2009). If perception-based measures are indeed more susceptible to cultural influences, it might well be that they correlate with a performance-based measure in one culture and fail to do so in another, which is what we have found in our secondary data analysis.

Another aspect to consider is the origin of the thresholds that define inadequate or marginal health literacy. Our analysis relied on the cutoff points that were originally established for the S-TOFHLA. These cutoff points in turn were determined by looking at age- and education-stratified frequency distributions of inadequate and marginal health literacy as identified by the original TOFHLA (Baker et al., 1999; Parker et al., 1995). This suggests that there might be some arbitrariness in the definition of inadequate or marginal health literacy. The more arbitrariness is recognized in the definition of health literacy levels, the more this would weaken the relevance of Chew and colleagues' (Chew et al., 2008) claim that the three BHLS identify the same groups as the S-TOFHLA. In turn, it would also weaken the relevance of the doubts we cast on the transferability of the BHLS to other contexts than the United States. To put it bluntly: if the definition of inadequate health literacy according to the standard measure is arbitrary, then it does not matter whether a new measure labels the same individuals as having inadequate or marginal health literacy, whether it is in the same or a different cultural or linguistic context. This argument advances another possible explanation for the different results found. That is, it might well be that the measures discussed here are indeed valid, but the cutoff points according to the S-TOFHLA are culturally specific. That possibility has not been considered in our analyses, but should be in future research.

Results of this study should be considered in light of its limitations. As partly acknowledged, one important limitation concerns the composition of the different samples that were included in this analysis. Firstly, all samples were convenience samples. Secondly, recruitment procedures differed from one country to another in that at times participants were sampled from the general population and at other times from specific patient populations. Thus, differences in the study design and reporting of sample characteristics limit comparability across the different countries, including the United States. However, it is important to highlight that even though the argument of limited comparability is an important one, the analyses showed throughout all datasets similar relationships. Meaning that *despite* limited comparability, patterns held true across different samples.

This study has important implications for research in health literacy. There is an increased interest in developing standardized health literacy measures that would allow to compare the distribution of health literacy levels and its influence on health (-related) outcomes across different countries and languages (Connor et al., 2013; Duong et al., 2017; Sørensen et al., 2015). Depending on the scope of a study, meaning whether the objective is to understand the impact of low health literacy in the clinical setting or its impact on a population-based level, researchers need to critically assess the validity of the chosen measure and its potential to systematically compare scores across different countries and languages. Besides taking into account issues such as the translation itself or the applicability of different thresholds, it might also include a sound evaluation of possible predictive models that would be able to sufficiently capture the concept of health literacy. There is considerable evidence on the association between different socio-demographic variables, such as educational level, income, or age, and health literacy (Baker, Gazmararian, Sudano, & Patterson, 2000; Carthery-Goulart et al., 2009; Gazmararian et al., 1999; Sørensen et al., 2015; Williams et al., 1995; Wolf et al., 2007). Finding the right combination of socio-demographic predictors to develop models that are able to predict health literacy (Martin et al., 2009; Van Der Heide et al., 2016) and health literacy-related outcomes might provide a good starting point to better understand if and, if so, how literacy affects health outcomes across countries where comparable measures are lacking.

Acknowledgments

The authors would like to thank all of the research coordinators and staff members at the study sites that coordinated the data collection efforts for the original studies.

References

Aguirre, A. C., Ebrahim, N., & Shea, J. A. (2005). Performance of the English and Spanish S-TOFHLA among publicly insured medicaid and medicare patients. *Patient Education and Counseling*, *56*(3), 332–339.

Altin, S. V., & Stock, S. (2015). Impact of health literacy, accessibility and coordination of care on patient's satisfaction with primary care in Germany. *BMC Family Practice*, *16*, 148.

Baker, D. W., Gazmararian, J. A., Sudano, J., & Patterson, M. (2000). The association between age and health literacy among elderly persons. *The Journals of Gerontology: Series B*, *55*(6), S368–S374.

Baker, D. W., Williams, M. V., Parker, R. M., Gazmararian, J. A., & Nurss, J. (1999). Development of a brief test to measure functional health literacy. *Patient Education and Counseling*, *38*(1), 33–42.

Beaton, D. E., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2000). Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*, *25*(24), 3186–3191.

Bernardi, R. A. (2006). Associations between Hofstede's cultural constructs and social desirability response bias. *Journal of Business Ethics*, *65*(1), 43–53.

Bishop, W. P., Craddock Lee, S. J., Skinner, C. S., Jones, T. M., McCallister, K., & Tiro, J. A. (2016). Validity of single-item screening for limited health literacy in English and Spanish Speakers. *American Journal of Public Health*, *106*(5), 889–892.

Carthery-Goulart, M. T., Anghinah, R., Areza-Fegyveres, R., Bahia, V. S., Brucki, S. M. D., Damin, A., … Nitrini, R. (2009). Performance of a Brazilian population on the test of functional health literacy in adults. *Revista De Saúde Pública*, *43*(4), 631–638.

Chew, L. D., Bradley, K. A., & Boyko, E. J. (2004). Brief questions to identify patients with inadequate health literacy. *Family Medicine*, *36*(8), 588–594.

Chew, L. D., Griffin, J. M., Partin, M. R., Noorbaloochi, S., Grill, J. P., Snyder, A., … VanRyn, M. (2008). Validation of screening questions for limited health literacy in a large VA outpatient population. *Journal of General Internal Medicine*, *23*(5), 561–566.

Cho, Y. I., Lee, S.-Y. D., Arozullah, A. M., & Crittenden, K. S. (2008). Effects of health literacy on health status and health service utilization amongst the elderly. *Social Science & Medicine*, *66*(8), 1809–1816.

Connor, M., Mantwill, S., & Schulz, P. J. (2013). Functional health literacy in Switzerland–Validation of a German, Italian, and French health literacy test. *Patient Education and Counseling*, *90*(1), 12–17.

Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E. J., George, R. B., Murphy, P. W., & Crouch, M. A. (1993). Rapid estimate of adult literacy in medicine: A shortened screening instrument. *Family Medicine*, *25*(6), 391–395.

Duong, T. V., Aringazina, A., Baisunova, G., Nurjanah, Pham, T. V., Pham, K. M., … Chang, P. W. (2017). Measuring health literacy in Asia: Validation of the HLS-EU-Q47 survey tool in six Asian countries. *Journal of Epidemiology*, *27*(2), 80–86.

Eyüboğlu, E., & Schulz, P. J. (2015). Validation of Turkish health literacy measures. Health Promotion International, *31*(2), 355–362.

Fadda, M., Kanj, M., Kabakian-Khasholian, T., & Johannes Schulz, P. (2016). Validation of three Arabic health literacy assessment tools in Lebanon. *Health Promotion International*. doi:10.1093/heapro/daw079

Fernandez, D. M., Larson, J. L., & Zikmund-Fisher, B. J. (2016). Associations between health literacy and preventive health behaviors among older adults: Findings from the health and retirement study. *BMC Public Health*, *16*, 596.

Fransen, M. P., Schaik, T. M., Van, Twickler, T. B., & Essink-Bot, M. L. (2011). Applicability of internationally available health literacy measures in The Netherlands. *Journal of Health Communication*, *16* (sup3), 134–149.

Franzen, J., Mantwill, S., Rapold, R., & Schulz, P. J. (2014). The relationship between functional health literacy and the use of the health system by diabetics in Switzerland. *The European Journal of Public Health*, *24* (6), 996–1002.

Friedman, D. B., Corwin, S. J., Dominick, G. M., & Rose, I. D. (2009). African American Men's understanding and perceptions about prostate cancer: Why multiple dimensions of health literacy are important in cancer communication. *Journal of Community Health*, *34*(5), 449–460.

Frisch, A.-L., Camerini, L., Diviani, N., & Schulz, P. J. (2012). Defining and measuring health literacy: How can we profit from other literacy domains?. *Health Promotion International*, *27*(1), 117–126.

Gazmararian, J. A., Baker, D. W., Williams, M. V., Parker, R. M., Scott, T. L., Green, D. C., … Koplan, J. P. (1999). Health literacy among medicare enrollees in a managed care organization. *JAMA : the Journal of the American Medical Association*, *281*(6), 545–551.

Geboers, B., Reijneveld, S. A., Jansen, C. J. M., & De Winter, A. F. (2016). Health literacy is associated with health behaviors and social factors among older adults: Results from the lifelines cohort study. *Journal of Health Communication*, *21*(sup2), 45–53.

Hajian-Tilaki, K. (2013). Receiver Operating Characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine*, *4*(2), 627–635.

Joshi, A., Mehta, S., Grover, A., Talati, K., Malhotra, B., & Puricelli Perin, D. M. (2013). Knowledge, attitude, and practices of individuals to prevent and manage metabolic syndrome in an indian setting. *Diabetes Technology & Therapeutics*, *15*(8), 644–653.

Kiechle, E. S., Bailey, S. C., Hedlund, L. A., Viera, A. J., & Sheridan, S. L. (2015). Different measures, different outcomes? A systematic review of performance-based versus self-reported measures of health literacy and numeracy. *Journal of General Internal Medicine*, *30*(10), 1538–1546.

Lee, S.-Y. D., Tsai, T.-I., & Tsai, Y.-W. (2013). Accuracy in self-reported health literacy screening: A difference between men and women in Taiwan. *BMJ Open*, *3*(11), e002928.

Lin, X., Wang, M., Zuo, Y., Li, M., Lin, X., Zhu, S., … Lamoureux, E. L. (2014). Health literacy, computer skills and quality of patient-physician communication in chinese patients with cataract. *Plos One*, *9*(9), e107615.

Lloyd, C. J. (1998). Using smoothed receiver operating characteristic curves to summarize and compare diagnostic systems. *Journal of the American Statistical Association*, *93*(444), 1356–1364.

Lupattelli, A., Picinardi, M., Einarson, A., & Nordeng, H. (2014). Health literacy and its association with perception of teratogenic risks and health behavior during pregnancy. *Patient Education and Counseling*, *96*(2), 171–178.

Mantwill, S., & Schulz, P. J. (2017). Does acculturation narrow the health literacy gap between immigrants and non-immigrants—An explorative study. *Patient Education and Counseling*. *100*(4), 760–767.doi:10.1016/j.pec.2016.10.021

Mårtensson, L., & Hensing, G. (2012). Health literacy – A heterogeneous phenomenon: A literature review. *Scandinavian Journal of Caring Sciences*, *26*(1), 151–160.

Martin, L. T., Ruder, T., Escarce, J. J., Ghosh-Dastidar, B., Sherman, D., Elliott, M., … Lurie, N. (2009). Developing predictive models of health literacy. *Journal of General Internal Medicine*, *24*(11), 1211–1216.

Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International*, *15*(3), 259–267.

Papp-Zipernovszky, O., Náfrádi, L., Schulz, P. J., & Csabai, M. (2016). "So that each patient may comprehend": Measuring health literacy in hungary. *Orvosi Hetilap*, *157*(23), 905–915.

Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The test of functional health literacy in adults. *Journal of General Internal Medicine*, *10*(10), 537–541.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction* (1st ed.). Oxford, United Kingdom: Oxford University Press..

Pines, J. M., Carpenter, C. R., Raja, A. S., & Schuur, J. D. (2012). *Evidence-based emergency care: diagnostic testing and clinical decision rules*. Oxford, United Kingdom: Wiley-Blackwell Publishing, Wiley.

R Core Development Team. (2015). *R: A language and environment for statistical computing, 3.2.1*. Vienna, Austria: R Foundation for Statistical Computing. Document Freely Available on the Internet at: *http://Www. r-Project. Org*. doi:10.1017/CBO9781107415324.004.

Robin, X., Turck, N., Hainard, A., Tiberti, N., Lisacek, F., Sanchez, J.-C., & Müller, M. (2011). pROC: An open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics*, *12*(1), 77.

Schulz, P. J., & Nakamoto, K. (2013). Health literacy and patient empowerment in health communication: The importance of separating conjoined twins. *Patient Education and Counseling*, *90*(1), 4–11.

Sentell, T., Baker, K. K., Onaka, A., & Braun, K. (2011). Low health literacy and poor health status in asian americans and pacific Islanders in Hawai'i. *Journal of Health Communication*, *16*(sup3), 279–294.

Shaw, S. J., Huebner, C., Armin, J., Orzech, K., & Vivian, J. (2009). The role of culture in health literacy and chronic disease screening and management. *Journal of Immigrant and Minority Health*, *11*(6), 460–467.

Sørensen, K., Pelikan, J. M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., … Brand, H. (2015). Health literacy in Europe: Comparative results of the European health literacy survey (HLS-EU). *The European Journal of Public Health*, *25*(6), 1053–1058.

Sørensen, K., Van Den Broucke, S., Fullam, J., Doyle, G., Pelikan, J., Slonska, Z., & Brand, H. (2012). Health literacy and public health: A systematic review and integration of definitions and models. *BMC Public Health*, *12*, 80.

Sørensen, K., Van Den Broucke, S., Pelikan, J. M., Fullam, J., Doyle, G., Slonska, Z., … Brand, H. (2013). Measuring health literacy in populations:

Illuminating the design and development process of the European Health Literacy Survey Questionnaire (HLS-EU-Q). *BMC Public Health*, *13*, 948.

Swets, J. A. (1986). Indices of discrimination or diagnostic accuracy: Their ROCs and implied models. *Psychological Bulletin*, *99*(1), 100–117.

Tokuda, Y., Doba, N., Butler, J. P., & Paasche-Orlow, M. K. (2009). Health literacy and physical and psychological wellbeing in Japanese adults. *Patient Education and Counseling*, *75*(3), 411–417.

Van Der Heide, I., Uiters, E., Sørensen, K., Röthlin, F., Pelikan, J., Rademakers, J., & Boshuizen, H. (2016). Health literacy in Europe: The development and validation of health literacy prediction models. *European Journal of Public Health*, *26*(6), 906–911.

Verkissen, M. N., Ezendam, N. P. M., Fransen, M. P., Essink-Bot, M.-L., Aarts, M. J., Nicolaije, K. A. H., … Husson, O. (2014). The role of health literacy in perceived information provision and satisfaction among women with ovarian tumors: A study from the population-based PROFILES registry. *Patient Education and Counseling*, *95*(3), 421–428.

Williams, M. V., Parker, R. M., Baker, D. W., Parikh, N. S., Pitkin, K., Coates, W. C., & Nurss, J. R. (1995). Inadequate functional health literacy among patients at two public hospitals. *JAMA : The Journal of the American Medical Association*, *274*(21), 1677–1682.

Wolf, M. S., Davis, T. C., Osborn, C. Y., Skripkauskas, S., Bennett, C. L., & Makoul, G. (2007). Literacy, self-efficacy, and HIV medication adherence. *Patient Education and Counseling*, *65*(2), 253–260.

Wolf, M. S., Williams, M. V., Parker, R. M., Parikh, N. S., Nowlan, A. W., & Baker, D. W. (2007). Patients' shame and attitudes toward discussing the results of literacy screening. *Journal of Health Communication*, *12* (8), 721–732.

Zeguers, M., De Haes, H. C. J. M., Zandbelt, L. C., Ter Hoeven, C. L., Franssen, S. J., Geijsen, D. D., … Smets, E. M. A. (2012). The information needs of new radiotherapy patients: How to measure? Do they want to know everything? And if not, why? *International Journal of Radiation Oncology*Biology*Physics*, *82*(1), 418–424.

Zou, K. H., O'Malley, A. J., & Mauri, L. (2007). Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation*, *115*(5), 654–657.