MEASURING HEALTH LITERACY AMONG U.S. CHINESE-SPEAKING POPULATIONS WITH LIMITED ENGLISH PROFICIENCY

A Dissertation

by

XUEWEI CHEN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	E. Lisako McKyer
Committee Members,	Adam Barry
	Patricia Goodson
	Sandra Acosta
Head of Department,	Richard B. Kreider

August 2017

Major Subject: Health Education

ABSTRACT

It is important to assess health literacy level among individuals with limited English proficiency (LEP) because assessing their health literacy level is the first step to develop a tailored health education program and reduce health disparities. The purpose of my dissertation is to improve the health literacy measurement and theory among populations with LEP. My dissertation investigates the psychometric properties of two functional health literacy measures and evaluates the adequacy of a modified health literacy survey to elicit valid data among 405 U.S. Chinese-speaking individuals with LEP.

I found researchers assessed health literacy using the non-English Test of Functional Health Literacy in Adults (TOFHLA) with 15 languages in 13 counties with different translation processes. Also, I determined that only 16 of the 74 eligible studies reported reliability coefficients for their data, with a reporting rate of 21.6%.

I also found that the current functional health literacy construct and measurement tools are not applicable among populations with LEP. Most individuals with LEP had adequate functional health literacy when assessed in their native languages. The English TOFHLA was measuring functional health literacy along with language proficiency. The Numeracy items of the Chinese TOFHLA yielded scores with low reliability.

My dissertation results show that the participants earned higher health literacy scores when they encounter health information/situations in Chinese rather than in English. I also found that few participants believed they had a voice in influencing or

reforming U.S. health policy. Further, the theoretical health literacy model had a better fit with the data from the Chinese scenario questions than the data from the English scenario questions.

Based on the dissertation study findings, I have three implications for future health literacy research and practice. First, I recommend researchers use the short form for future non-English TOFHLA instruments application. Second, culturally related constructs such as English language proficiency are key components that should be added to the health literacy measurement and theoretical model for populations with LEP. Last, public health professionals should incorporate health interventions and policy approaches to improve critical health literacy among populations with LEP.

ACKNOWLEDGEMENTS

I would like to thank my committee chair, Dr. McKyer, and my committee members, Dr. Goodson, Dr. Barry, and Dr. Acosta, for their guidance and support throughout the course of this research.

Thanks also go to my friends and colleagues and the department faculty and staff for making my time at Texas A&M University a great experience.

Finally, thanks to my mother and father for their encouragement and to my boyfriend for his patience and love.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor McKyer [advisor] and Professors Goodson and Barry of the Department of Health & Kinesiology and Professor Acosta of the Department of Educational Psychology.

All work for the dissertation was completed by the student, under the advisement of Dr. McKyer of the Department of Health & Kinesiology.

Funding Sources

This work was supported by a dissertation research fellowship from the College of Education and Human Development at Texas A&M University. This study also received funding as graduate student research awards from the College of Education and Human Development and Huffines Institute for Sports Medicine and Human Performance at Texas A&M University.

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

Page

INTRODUCTION

The prevalence of low health literacy in the United States is especially high among populations with limited English proficiency; however, few health literacy instruments have been developed or validated for these linguistic minority groups (McKee & Paasche-Orlow, 2012). Individuals with limited English proficiency (LEP) refers to anyone ages 5 and above who speaks English "less than very well" (Pandya, Batalova, & McHugh, 2011). About 8.6% of the population (26 million people) have LEP (US Census Bureau, 2015). Researchers and health practitioners should not assume universal applicability of health literacy instruments; studies should consider native languages, culture, and health systems before conducting health literacy assessments (Dowse, Lecoko, & Ehlers, 2010). Further, Yip (2012) pointed out the current health literacy instruments were incapable of capturing the complex interaction among populations with LEP. Thus, there is a critical need to develop a tailored health literacy instrument targeting populations with LEP.

Assessing health literacy is the first step in starting a health education program (Thomason & Mayo, 2015). Health literacy measurement tools serve as an essential component of needs assessment and provide the foundation for an intervention plan. A precise tool for assessing health literacy allows clinicians to tailor health services for patients. The tools also help health professionals to shape patient-provider communication and reduce health disparities (Nørgaard, Sørensen, Maindal, & Kayser, 2014; Stonbraker, Schnall, & Larson, 2015; Batterham, Hawkins, Collins, Buchbinder,

& Osborne, 2016). Thus, it is important to assess health literacy level among populations with LEP.

The purpose of my dissertation study is to improve the current health literacy measurement and theory as applied among populations with LEP through investigating the psychometrics properties of three health literacy instruments (two objective tests and one subjective survey). The hypothesis of this dissertation is that the existing health literacy theory and assessment tools can be improved by adding English proficiency and culture components. My dissertation uses the Journal Article Style Format with 3 manuscripts. Manuscript #1 is a systematic literature review and meta-analysis study reviewing the translation process of the non-English Test of Functional Health Literacy in Adults (TOFHLA) and examining which test/sample characteristics impact the score reliability of these instruments. Manuscript #2 investigates the psychometric properties of two functional health literacy assessment instruments (English and Chinese Short-TOFHLA) using Item Response Theory among U.S. native Chinese speakers with limited English proficiency. Manuscript #3 assesses health literacy level among U.S. native Chinese speakers with limited English proficiency using a modified theory-based survey (All Aspects of Health Literacy Scale) and investigates the adequacy of the survey instrument to elicit valid data from this unique population.

DOES TRANSLATION PROCESS INFLUENCE THE SCORE RELIABILITY OF NON-ENGLISH TOFHLA? A SYSTEMATIC LITERATURE REVIEW AND RELIABILITY GENERALIZATION META-ANALYSIS

Introduction

The United States is encountering a widespread problem of low health literacy. Results from the 2003 National Assessment of Adult Literacy show 36% of the U.S. adult population ages 18 and older have basic or below basic health literacy levels, while only 12% have proficient health literacy (Kutner, Greenburg, Jin, & Paulsen, 2006). In other words, more than one third of the U.S. adults have inadequate health literacy. Moreover, nearly half of all American adults (90 million people) have difficulty understanding health information and 40 million cannot read complex texts at all (Institute of Medicine, 2004).

The U.S. is not the only country facing health literacy challenges; the prevalence of low health literacy is a global health problem. For example, the 2003 International Adult Literacy and Lifeskills Survey results show over 40% of the adult populations in Canada, New Zealand, Australia, Hungary, Italy, and Nuevo Leon had limited health literacy skills (Satherley, Lawes, & Sok, 2008). Another project, the European Health Literacy Survey from 2009-2012, indicated 47% adults in Austria, Bulgaria, Germany, Greece, Ireland, the Netherlands, Poland and Spain had limited health literacy (Sørensen et al., 2015).

To enhance health literacy among various populations groups, a necessary first step is to assess people's current levels. A precise health literacy measurement tool is essential for conducting a needs assessment of a population, developing tailored health interventions, and creating program evaluation criteria (McCormack, Haun, Sørensen, & Valerio, 2013). Therefore, a valid health literacy measurement instrument is important for health promotion.

Theoretical Framework

Health literacy represents the ability to "obtain, process, and understand basic health information and services needed to make appropriate health decisions" (Ratzan & Parker, 2000; Institute of Medicine, 2004). As shown from the health literacy definition, functional literacy is the basic dimension of health literacy (Frisch, Camerini, Diviani, & Schulz, 2012). Functional literacy is the "ability to identify, understand, interpret, create, communicate, compute and use printed and written materials associated with varying contexts" (UNESCO, 2004, p. 13). According to Parker and colleagues (1995), functional health literacy refers to the ability to apply literacy skills to health-related context such as prescriptions, medicine labels, and appointment cards. Reading, writing, and numeracy skills are essential components of functional health literacy (Parker, Baker, William, & Nurss, 1995).

More than 50 instruments are available for measuring individual health literacy (Haun, Valerio, McCormack, Sørensen, & Paasche-Orlow, 2014). Among these instruments, the Rapid Estimate of Adult Literacy in Medicine—REALM (Davis et al.,

1993) and the Test of Functional Health Literacy in Adults—TOFHLA (Parker, Baker, Williams, & Nurss, 1995) were widely used to assess functional health literacy among various population groups. Altin and colleagues (2014) noted that almost 30% of existing health literacy measurements were developed based on REALM or TOFHLA. REALM tests word recognition skills but not comprehension skills. Compared to REALM, TOFHLA adds reading comprehension and numeracy assessments.

TOFHLA is considered the gold standard for functional health literacy assessment because it yields strong reliability and validity data in English (Mancuso, 2009). TOFHLA (Parker et al., 1995) was originally developed in English in the U.S. that comprised two subtests with 66 multiple choice items that assess two components of functional health literacy: reading comprehension (measured by 50 items) and numeracy (measured by 16 items). The reading comprehension subtest contains three passages with (1) instructions for the preparation of an upper gastrointestinal series, (2) a Medicaid application form, and (3) a standard hospital informed consent form. These three passages use the Cloze format (Taylor, 1953), which omits every fifth to seventh word in a passage and asks people to select the correct choice from three incorrect choices. The numeracy subtest assesses test-takers' ability to understand instructions for taking medicines, monitoring blood glucose, keeping clinic appointments, and getting financial assistance using actual hospital forms and prescription labels. S-TOFHLA is a short version of the TOFHLA with 40 items (Baker, Williams, Parker, Gazmararian, & Nurss, 1999). The reading comprehension subtest contains 36 items (measured by two passages) and the numeracy section contains 4 items.

There has been a growing number of ethnic-specific translations of TOFHLA, yet these instruments suffer from measurement problems due to language translation and cultural modification process. There is a critical need to review how the non-English TOFHLA instruments were translated/modified and examine the influence of translation process on scale reliability. Nguyen and colleagues (2015) pointed out some items in the original TOFHLA were specific to the culture and healthcare system in the US; therefore, these items might not be applicable in other countries (Nguyen et al., 2015). For example, the reading comprehension passage about Medicaid application contains acronyms (e.g. TANF: Temporary Assistance for Needy Families) unique to the U.S. healthcare system. Another example is one of the numeracy items assessing test-takers' interpretation of a clinic appointment card. However, many countries do not use appointment cards. Thus, the health literacy instrument translation process might affect the measurement errors. To date, no review has systematically examined the translation process of these non-English TOFHLA instruments.

Compared to systematic literature reviews that present findings using a narrative approach, meta-analysis focuses on the aggregation and comparison of the quantitative findings across different studies (Lipsey & Wilson, 2001). Reliability Generalization (RG) is a meta-analytic method that examines the "mean measurement error variance across studies, and also the sources of variability of these variables across studies" (Vacha-Haase, 1998, p. 6).

RG is one type of meta-analysis approaches to examine the influence of instruments and test-takers on the reliability coefficients of the scores (Vacha-Haase, 1998). Tests are neither reliable nor unreliable but rather, reliability is a property of scores (Henson & Thompson, 2002). The same instrument can yield different score reliability parameters when it is administered differently or assessed across various groups of test-takers. To the best of our knowledge, there are no studies investigating how tests and test-takers' characteristics impact the measurement error of a health literacy instrument.

The purpose of this study is to systematically review the translation process of the non-English TOFHLA instruments and examine, employing a meta-analysis approach, which test descriptors (e.g., long or short version, targeted language, time limit, etc.) and sample characteristics (e.g., sample size, gender, age, etc.) impact the score reliability of these instruments. This study is guided through the following questions:

1. In what languages and with which populations have the non-English TOFHLA instruments been tested?
2. How these instruments are translated and modified from English to other languages?
3. How variable is the reliability of the scores across these non-English TOFHLA?
4. What are the test and sample characteristics influencing the reliability of the scores obtained through the translated versions of TOFHLA?

Methods

Guided by the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA, 2010) and Garrard's (2016) matrix method, we conducted a systemic literature review to answer questions 1 and 2. To investigate questions 3 and 4, we applied the Reliability Generalization (RG) meta-analytic approach (Vacha-Haase, 1998). RG highlights the importance of reported reliability coefficients for correct data interpretation (Cousin & Henson, 2000). Therefore, in RG, score reliability coefficients are effect sizes. We listed the detailed process for study selection in Figure 1.

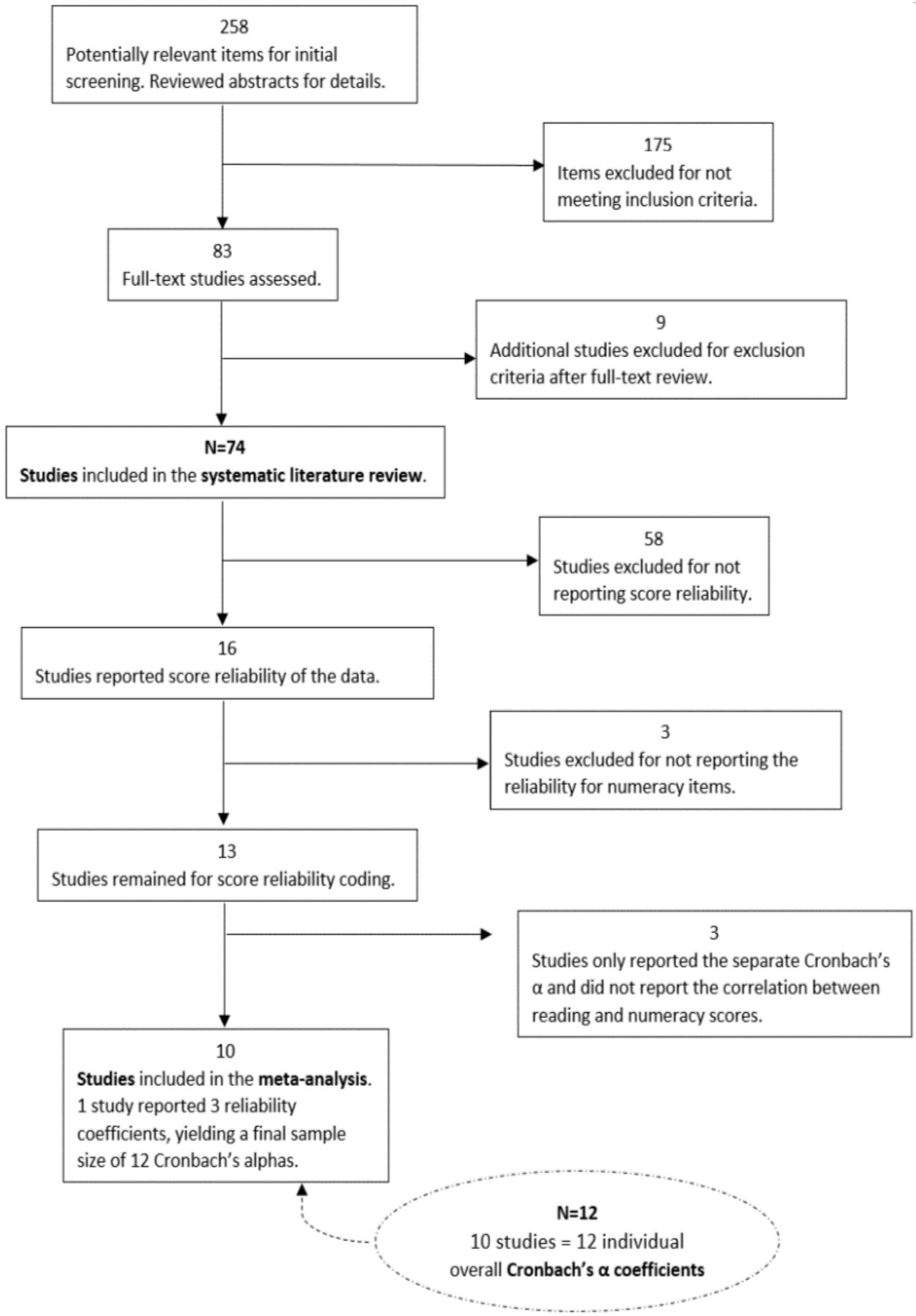

Figure 1. Process Tree for Study Selection

Sample for Systematic Literature Review

We searched four electronic databases: PubMed, Medline, Scopus, and Google Scholar using the search terms "Test of Functional Health Literacy in Adults" or "TOFHLA" which generated 258 hits. Inclusion criteria for the review were: (1) studies measured health literacy using TOFHLA, (2) the TOFHLA instruments were written in languages other than English, (3) and the studies were peer-reviewed publications. Studies were excluded if researchers created new instrument so that the instrument format was not the same as the original TOFHLA. We selected 83 studies for full text review after applying the inclusion criteria. From these 83 studies, nine were excluded. In seven studies, authors created new non-English instruments using TOFHLA as a reference only. Two studies used the Test of Functional Health Literacy in Adolescents instead of the Test of Functional Health Literacy in Adults. Therefore, the total number of publications included in the systematic literature review section was 74 (Figure 1).

Sample for Meta-Analysis

The sample size for our meta-analysis is smaller than that for the systematic literature review because not every study included in the systemic review reported score reliability information. For the meta-analysis, we determined that only 16 of the 74 eligible studies reported reliability coefficients for their data, with a reporting rate of 21.6%. Authors for all 16 studies reported score reliability using Cronbach's α coefficients. Cronbach's α is a measure used to assess the internal consistency of a set of test items (Cronbach, 1951). For our RG meta-analysis, we used Cronbach's α as the effect size.

Of the 16 studies subjected to meta-analysis, three reported the Cronbach's α for the reading comprehension items but did not report the α for numeracy items. Therefore, these three studies were excluded from analysis. For the remaining thirteen studies, three only reported the overall α; six only reported separate α coefficients (numeracy and reading comprehension); and four reported both overall α and separate α coefficients. Of the six studies that only reported separate α coefficients, three reported the correlation coefficients between the raw scores of numeracy and reading comprehension. We were able to calculate the overall α of an instrument when the separate α coefficients and the correlation between the raw scores of the two components were reported (Willson & Reynolds, 1985). Three studies were excluded because we were unable to calculate the overall α coefficients. Thus, a total number of 10 studies were included in the meta-analysis.

One study (Connor, Mantwill, & Schulz, 2013) conducted TOFHLA assessments in three languages (German, Italian, and French) among three sample groups; this particular study contained three sets of reliability coefficients, test descriptions, and sample characteristics. Therefore, ten studies with 12 sets of overall α coefficients were included in our RG meta-analysis (n = 12).

Reliability Transformation and Weight Function

Correlation coefficients are usually transformed using Fisher's r-to-z transformation when used as effect sizes (Lipsey & Wilson, 2001). One of the issues within RG which researchers debate is whether to transform reliability estimates or not (Sánchez-Meca, López-López, & López-Pina, 2013). Henson and Thompson (2002) believe the Fisher r-to-z transformation is not necessary before submitting reliability coefficients to analysis because reliability coefficients are different from correlation coefficients. Therefore, we did not perform the reliability transformation but used the original Cronbach's alpha coefficients as effect sizes in this meta-analysis as suggested by Henson and Thompson (2002).

Effect sizes based on larger samples yield more precise estimate of the population compared to smaller samples (Lipsey & Wilson, 2001). Weighting each effect size by its sample size is a strategy in meta-analysis to reduce the sampling errors of studies with small sample size (Lipsey & Wilson, 2001; Schmidt & Hunter, 2014). We weighted each reliability coefficient based on its corresponding sample size.

Dependent and Independent Variables

The dependent (continuous) variable in this meta-analysis is the effect size—Cronbach's alpha coefficients. The independent variables fall under two categories: sample descriptors and test characteristics.

The sample descriptors include the following continuous variables: sample size, mean age, the gender distribution represented as the percentage of female participants, and the education distribution represented as the percentage of the participants who had an education level of high school and above. All but one (Mantwill & Schulz, 2016) of the studies reported the mean age of the test-takers. Mantwill and Schulz (2016) reported the sample distribution among three age ranges (18-35 years, 36-55 years, and 56-75 years). We utilized the means of each age range as a calculated mean for each group and computed the grand mean age based on this information. There is one binary variable among the sample descriptors—health status, where 0 represents general population with no health conditions, and 1 indicates the samples were recruited from clinical settings with specific health conditions.

Variables as test characteristics included: location (0 = the test was administered outside the US, 1 = in the US), translation modification (0 = no modification, 1 = the test was modified because of culture and system differences), time limit (0 = the test was administrated with no time limit, 1 = time limited), test length (0 = short test version, 1 = long version), translated language (0 = Non-Indo-European languages, 1 = Indo-European languages), and test mode (0 = face-to-face interview, 1 = self-administered paper-pencil test, 0.5 = mixed). Connor and colleagues (2013) administered their three instruments (in German, Italian, and French) with a mixed test mode—self-administered paper-pencil test for the reading comprehension items and face-to-face interview for the numeracy items. The languages were coded as Non-Indo-European or Indo-European according to Ruhlen’s (1991) world languages classification. Compared to Non-Indo-European languages, Indo-European languages are closer to English because they are all nontonal languages, share cognate words and have similar grammatical structures (Barac

& Bialystok, 2012). In the present review, Turkish and Chinese are classified as Non-Indo-European languages; Spanish, Serbian, Albanian, German, Italian, French and Danish are categorized as Indo-European languages. The main author coded the variables twice and compared these two versions of coding to insure the intra-rater reliability.

Data Analysis

We used Vacha-Haase, Henson, and Caruso's excel spread sheet (2002, p. 566) to calculate the 95% confidence intervals for each effect size. We performed homogeneity analysis based on the Q statistic (Hedges & Olkin, 1985) to check whether the effect sizes were significantly different from each other. We applied analog to the analysis of variance (Hedges, 1982) and weighted regression analysis (Hedges & Olkin, 1985) to test the ability of the independent variables (sample descriptors and test characteristics) to explain the effect size variability (Lipsey & Wilson, 2001). For the categorical independent variables (health status, location, translation modification, time limit, test length, and language), we applied the analog to the ANOVA. For the continuous independent variables (sample size, mean age, gender distribution, and education distribution), along with the categorical independent variables, we employed weighted regression analysis.

Results

Systematic Literature Review

Among these 74 studies, researchers assessed health literacy in 13 counties (e.g., US, Puerto Rico, Iran, and Turkey) using non-English TOFHLA in 15 languages (e.g., Spanish, Portuguese, Korean, and Persian) with different translation processes (see details in Table 1). When the original English TOFHLA was developed in the US, the research team (Parker et al., 1995) also translated it into Spanish and pilot tested the Spanish version using a sample of Spanish-speaking Hispanic Americans. The translation process applied translate and back-translate technique. Studies measuring health literacy in the US using the Spanish TOFHLA (*n* = 58) administered the Parker el al. (1995) version. Only one Spanish TOFHLA was tested among populations outside the US — Puerto Rico (Rivero-Méndez et al., 2010). The Korean (Han, Kim, Kim, & Kim, 2011) and Vietnamese (Shaw, Armin, Torres, Orzech, & Vivian, 2012) TOFHLA were employed among immigrants in the US. Other non-English TOFHLA instruments were translated and tested in other countries.

Table 1. Translation Processes of the TOFHLA into Languages other than English

Studies	Language	Population	Translation Process	Translation Guideline	Cultural Modification
Parker et al. (1995)	Spanish	Hispanics in the U.S.	"To develop a Spanish version of TOFHLA, or ROFHLA-S, the reading comprehension passages and numeracy questions were translated into Spanish and back translated into English. Discrepancies were corrected using the consensus of several bilingual staff members and a Spanish literacy expert. The Cloze procedure was then performed on each reading passage to achieve difficulty comparable to that of the English passage instead of using the same word deletions and response options as the English version had" (p. 538).	NA	NA
58 studies (references available upon request)	Spanish	Spanish-speaking populations in the U.S.	Used Parker's version.	NA	NA
Rivero-Méndez et al. (2010)	Spanish	Puerto Ricans living with HIV/AIDS	"The TOFHLA-S was evaluated for cultural equivalence for our population using Beaton et al. (2002) guidelines" (p. 50).	Beaton et al. (2002)	Yes
Carthery-Goulart et al. (2009)	Portuguese	Brazilians	"The English and Spanish versions of S-TOFHLA were translated and adapted to the Brazilian reality, especially the reading comprehension texts as to convey information about the Sistema Único de Saúde (Brazilian Health System – SUS). When sentences needed to be modified for this purpose, the same structure was kept, using stimuli in the alternatives which were either phonetically similar to the target or that belonged to the same grammar class" (p. 633).	NA	Yes

Table 1 Continued

Studies	Language	Population	Translation Process	Translation Guideline	Cultural Modification
Han et al., (2011)	Korean	Korean American women	"S-TOFHLA was translated into Korean and back-translated into English to ensure accuracy of translation. Due to differences in the basic structure of English and Korean, passages in S-TOFHLA were rewritten keeping the consistency in the order of subject-object-verb words, in contrast to English which has a subject-verb-object word order" (p. 255).	NA	NA
Reisi et al., (2012)	Persian	Older adults in Iran	NA	NA	NA
Reisi et al., (2016)	Persian	Diabetes patients in Iran	NA	NA	NA
Shaw et al., (2012)	Vietnamese	Chronic disease patients	NA	NA	NA
Eyüboğlu & Schulz (2016)	Turkish	Diabetes patients in Turkey	"The texts and the options were translated by a philologist and native speaker of Turkish following the standard methodologies for questionnaire translation… Particular emphasis was given to the conceptual equivalent of words and phrases instead of providing a literal translation… 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" (p. 2).	Sperber et al. (1994)	Yes

Table 1 Continued

Studies	Language	Population	Translation Process	Translation Guideline	Cultural Modification
Jovic-Vranes, Bjegovic-Mikanovic, & Marinkovic (2009)	Serbian	Patients in Serbia	"The original English versions of the TOFHLA were translated into the Serbian language by a multidisciplinary team following the standard methods to translate and adapt the questionnaires… The Serbian version of the TOFHLA was administrated to 10 primary care patients. Subsequently, the problematic items were changed (i.e. questions regarding health-care insurance were adapted to the Serbian healthcare insurance system and US dollars were converted to Serbian dinars)" (p. 491).	Sperber et al. (1994)	Yes
Kamberi et al. (2012)	Albanian	Patients in Kosovo	"The original English versions of the TOFHLA were translated into the Albanian language by experts following the standard methods of translation and cross-cultural adaptation of the questionnaires… Of the original version of the TOFHLA, a few not-applicable items were changed. These included questions regarding health care insurance (which are not relevant for Kosovo as there is no healthcare insurance system in place yet). Furthermore, US dollars were converted into Euros" (p. 22).	Sperber et al. (1994)	Yes
Toçi et al. (2015)	Albanian	Individuals in Albania	"The TOFHLA instrument was already translated, back-translated and validated among 54 primary care patients in another Albanian-speaking country, namely in Kosovo. However, because of the changing of currency and health insurance system used in Kosovo and Albania, a panel of experts was invited to agree on the adaptation of the corresponding items of TOFHLA questionnaire in the Albanian context" (p. 483).	Sperber et al. (1994)	Yes
Hæsum et al. (2015)	Danish	Patients with chronic disease in Denmark	"The authors of this paper chose to translate the original American version of the TOFHLA for the Danish setting and culture, also by following the guidelines for cross-cultural adaptation as defined by Beaton et al." (p. 575).	Beaton et al. (2002)	Yes

Table 1 Continued

Studies	Language	Population	Translation Process	Translation Guideline	Cultural Modification
Al-Jumaili et al. (2015)	Arabic	Individuals in Iraq	"When the S-TOFHLA was translated to Arabic, one of the 36 cloze items was dropped from the passages because it did not make cultural sense after translation to Arabic… "Medicaid" was translated as health care assistance for needy people… Each test required translation into formal Arabic. The translation used formal methodology by Wild et al. (2005)… A pilot study with 25 subjects was conducted to assess any ambiguous statements or questions and to make sure the Arabic forms were understandable by pharmacy customers since these were the first Arabic versions of the three health literacy tests. The translation of NVS and S-TOFHLA was refined after the feedback from the pilot study" (p. 804-805).	Wild et al. (2005)	Yes
Connor, Mantwill, & Schulz (2013)	German Italian French	Residents in Switzerland with their preferred languages	"The original English S-TOFHLA was translated by native speakers of German, Italian, and French into the respective languages following the standard methodologies for questionnaire translation (Sperber et al, 1994). Following translation, the questionnaire was back translated by a native English speaker who was fluent in the respective language to see whether differences between the original English and the translated versions would occur. Back translations were systematically reviewed in accordance with predefined grammatical criteria, which had been formulated when translating the original questionnaire into the three target languages, respectively. Furthermore, special attention was paid to the cultural adaptation of the context. Some minor changes were implemented due to differences in the Swiss Health Care System" (p. 13).	Sperber et al. (1994)	Yes
Tang et al. (2008)	Traditional Chinese	Chinese Diabetes patients in Hong Kong	"The four numeric questions were translated into Chinese… We identified two two actual Chinese sets of patient information: (a) on preparation for a colonoscopic examination series and (b) the patient rights and responsibilities section of a Medicaid application normally completed on admission into hospital" (p. 77).	NA	Yes

Table 1 Continued

Studies	Language	Population	Translation Process	Translation Guideline	Cultural Modification
Mantwill & Schulz (2016)	Simplified Chinese	Chinese patients in Mainland China	"A native Mandarin speaker translated the traditional Chinese version (Tang et al., 2008) into simplified Chinese and minimal changes regarding some expressions were made to make it more applicable to the Chinese health care context… Translations were subsequently reviewed by three native Chinese speakers of whom two were medical doctors" (p. 3).	NA	Yes

Note: "NA" indicates that the information was not mentioned in the reviewed studies

As shown in Table 1, few studies (11 out of 74) mentioned what they did to cope with the challenges of linguistic, cultural, and health system differences between the US and the targeted countries. Eleven studies were modified for cultural discrepancies (Al-Jumaili, Al-Rekabi, & Sorofman, 2015; Carthery-Goulart et al., 2009; Connor et al., 2013; Eyüboğlu & Schulz, 2016; Hæsum, Korsbakke, Ehlers, & Hejlesen, 2015; Jovic-Vranes, Bjegovic-Mikanovic, & Marinkovic, 2009; Kamberi, Hysa, Toçi, Jerliu, & Burazeri, 2012; Mantwill & Schulz, 2016; Rivero-Méndez et al., 2010; Tang, Pang, Chan, Yeung, & Yeung, 2008; Toçi, Burazeri, Sørensen, Kamberi, & Brand, 2014). Eight studies (Al-Jumaili et al., 2015; Connor et al., 2013; Eyüboğlu & Schulz, 2016; Hæsum et al., 2015; Jovic-Vranes et al., 2009; Kamberi et al., 2012; Rivero-Méndez et al., 2010; Toçi et al., 2014) followed some cross-culture guidelines (Beaton, Bombardier, Guillemin, & Ferraz, 2002; Sperber, Devellis, & Boehlecke, 1994; Wild et al., 2005) to ensure the translation accuracy and cultural sensitivity of their instruments. The Chinese TOFHLA instruments (Mantwill & Schulz, 2016; Tang et al., 2008) did not mention a specific translation guideline but they modified the instruments according to the Chinese health care context. Some studies focused on developing new health literacy instruments using a non-English TOFHLA for validity evaluation and did not report much information about the non-English TOFHLA (Reisi et al., 2012; Reisi et al., 2016; Shaw et al., 2012).

General Description of Studies

Ten studies (Parker et al., 1995; Aguirre, Ebrahim, & Shea, 2005; Jovic-Vranes, Bjegovic-Mikanovic, & Marinkovic, 2009; Rivero-Méndez et al., 2010; Kamberi et al., 2012; Connor, Mantwill, & Schulz, 2016; Hæsum, Korsbakke, Ehlers, & Hejlesen, 2015; Al-Jumaili, Al-Rekabi, & Sorofman, 2015; Mantwill & Schulz, 2016; Toçi et al., 2015) with twelve score reliability coefficients were included in the RG meta-analysis (Table 2). One study (Connor et al., 2013) reported three α for three language versions of TOFHLA tested among three different sample groups. Three tests were administrated among a general population with no health conditions, while nine tests were assessed among people with specific health issues. Ten tests were used among populations outside the US (all of them were modified for culture or health system differences), while two were in the US (with no modification). Ten tests were translated into Indo-European languages and two were translated into Non-Indo-European languages. Four tests were administrated with time limits (7 minutes for the short version, 22 minutes for the long version) and eight were untimed. Five tests used the long form and seven used the short form.

Table 2. General Description of Studies

	n of Cronbach's α
HEALTH STATUS	
With health condition	9
With no health condition	3
LOCATION	
Inside the US	2
Outside the US	10
TRANSLATION MODIFICATION	
With no modification	2
With modification	10
TEST MODE	
Self-administrated paper-pencil	8
Face-to-face interview	1
Mixed	3
TIME LIMIT	
Yes	4
No	8
TEST LENGTH	
Long form	5
Short form	7
LANGUAGE	
Indo-European languages	10
Non-Indo-European languages	2
Total Number of Reported/Calculated Cronbach's α	**12**

Confidence Intervals for α

We calculated the 95% confidence intervals for these 12 Cronbach's α coefficients (Table 3). As shown in Figure 2, these non-English TOFHLA instruments yielded high score reliability (α > 0.85), and α variability estimates ranged from 0.812 to 1.00.

Table 3. 95% Confidence Intervals (CIs) for α

α	**Sample Size**	**CI Lower Boundary**	**CI Upper Boundary**	**CI Width**
0.98	249	0.976	0.983	0.007
0.97	1066	0.967	0.973	0.005
0.94	105	0.922	0.955	0.033
0.95	30	0.921	0.972	0.052
0.93	54	0.901	0.954	0.054
0.88	249	0.856	0.899	0.043
0.97	273	0.964	0.974	0.011
0.85	137	0.812	0.884	0.072
0.94	42	0.915	0.965	0.050
0.99	93	0.988	0.994	0.005
1.00*	150	1.000	1.000	0.000
0.92	239	0.905	0.934	0.029

*Note: We calculated this α based on the statistics provided in the study.

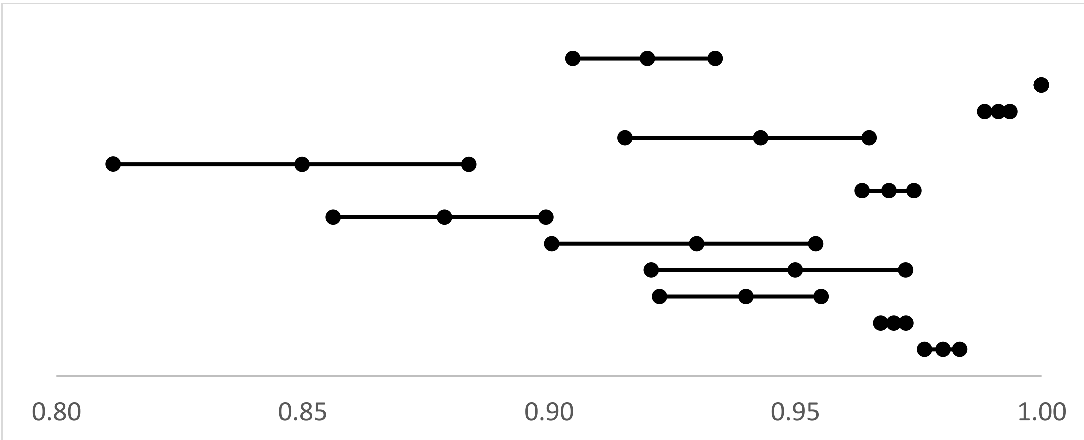

Figure 2. 95% Confidence Intervals for α

Homogeneity Test

We conducted a homogeneity test using Q statistic to examine the effect size variability. We failed to reject the null hypothesis for homogeneity (Q = 4.159, *p* = 0.965), indicating that the variability of the Cronbach's α across this review's samples was not significantly different from zero; in practical terms, these Cronbach's α coefficients did not vary substantially. Lipsey and Wilson (2001) point out fixed effects models are fit for homogenous effect size distributions. Fixed effects models assume the sample of each study comes from a single population with a random subject-level sampling error (Lipsey & Wilson, 2001). Since the effect size distribution was homogeneous in our meta-analysis study, we applied the fixed effect model for our RG meta-analysis.

Analog to the Analysis of Variance

We tested six moderators using analog to the analysis of variance to examine whether the variance among these Cronbach's alpha coefficients was due to the following group differences: (a) the test-takers had certain health conditions (*n* = 9) *vs*. the test-takers were drawn from the general population with no health issues (*n* = 3); (b) the test was administrated inside the US (*n* = 2) *vs*. outside the US (*n* = 10); (c) the test translation was modified to address the culture and health system difference (*n* = 10) *vs*. no translation modification (*n* = 2); (d) the test was administrated with time limits (*n* = 4) *vs*. untimed (*n* = 8); (e) the test used the short version (*n* = 7) *vs*. the long version (*n* = 5); and (f) the test being translated into a Indo-European language (*n* = 10) *vs*. a Non-Indo-European language (*n* = 2). Since all of the tests administrated outside the US have been modified to address culture and health system differences, the statistics were same between the location and translation modification variables. None of the $Q_{between}$ value was statistically significant (Table 4), indicating these moderators had no significant effect on the variability of the Cronbach's α coefficients.

Table 4. Analog to the Analysis of Variance

Moderators	**QB**	***p*(QB)**	**QW**	***p*(QW)**
a. Health Status	1.497	0.221	2.529	0.990
b. Location	1.060	0.303	2.856	0.985
c. Translation Modification	1.060	0.303	2.856	0.985
d. Time limit	0.709	0.400	3.244	0.975
e. Test Length	0.017	0.897	4.122	0.942
f. Language	0.542	0.462	3.241	0.975

Weighted Regression Analysis

Hedges and Olkin's (1985) weighted multiple regression was also applied in our RG meta-analysis to explore possible variables that could predict the Cronbach's α coefficients. Since the sample for the effect size in our RG meta-analysis was small (n=12), putting all the ten predictors in one single multiple regression model would increase estimation bias (Thompson, 2006). Thus, we performed a set of simple linear regression models with one predictor at a time instead of applying commonality analysis that put several predictors simultaneously in one multiple regression model. In Table 5, we listed the standardized regression coefficients, R^2, standard errors, and p values for 10 sets of simple linear regression models with one predictor variable in each model.

Table 5. Weighted Regression Analysis—Simple Linear Regressions Predicting α

Variables	**β**	$\mathbf{R^2}$	**SE**	***p***
1. Sample Size	0.37	14.05%	0.00	0.50
2. Mean Age	-0.20	3.90%	0.00	0.50
3. Gender Distribution	0.44	19.35%	0.15	0.56
4. Health Status	0.60	35.98%	0.04	0.52
5. Education Distribution	-0.15	2.36%	0.15	0.56
6. Test Mode	0.62	38.09%	0.06	0.52
7. Time Limit	-0.41	17.05%	0.04	0.52
8. Test Length	-0.06	0.40%	0.04	0.52
9. Location/Translation Modification	0.50	25.49%	0.04	0.52
10. Language	-0.36	13.03%	0.07	0.53

As shown in Table 5, none of the predictors were statistically significant. However, non-significant *p* values did not mean these predictors were not important. Sample size is a very important determinant of *p* values (Thompson, 2006). Considering the small group of Cronbach's α coefficients in our study ($n = 12$), it was not surprising we did not obtain significant *p* values ($p < 0.05$). We should report and interpret effect sizes (R^2 in this case) along with significance testing to avoid misinterpretation (Thompson, 2006). In this RG meta-analysis, the best three variables for predicting the Cronbach's α coefficients were *Test Mode*, *Health Status*, and *Location/Translation Modification* which predicted 38.09%, 35.89% and 25.49% of the variance in the Cronbach's α coefficients for non-English TOFHLA instruments, respectively. Self-administrated paper-pencil tests were more likely to yield higher reliability than face-to-face interviews. Test-takers who had specific health condition were more likely to yield lower reliability than those with no health condition. Tests administrated within the US were more likely to yield higher reliability than the tests administrated outside the US, even though all the test administrated outside the US were modified to reduce the problems caused by culture and health system differences. On the other hand, the least functional variables in predicting the Cronbach's α coefficients for Non-English TOFHLA instruments was *Test Length* ($R^2 = 0.4\%$), indicating using the short test version instead of the long test version would not reduce the reliability of the score.

Discussion

We believe this review contributes to the literature on health literacy measurement in three ways. First, it identified various non-English TOFHLA instruments tested among patients or general populations in different counties with numerous translation and modification processes. Second, it evaluated the processes undertaken in the translating and modifying of these instruments. Third, it investigated the mean measurement error variance the sources of variability of the Cronbach's α coefficients across these non-English TOFHLA instruments. Our findings indicated these reliability coefficients were not significantly different from each other and using the short version of TOFHLA was not a source of particular concern reducing the reliability of the score.

Limitations

Alongside these contributions, readers should bear in mind our study's limitations. We only reviewed articles published in English in refereed, scholarly journals. Therefore, our study did not include items in other languages or published elsewhere. Also, the main author was the only person coding the data, so there could be some coder bias, too.

Future Study

It is worth noting that the rate of reporting reliability coefficients among these non-English TOFHLA instruments was low (about 21.6%). This low reporting rate was consistent with other studies conducted in the 1980s and 1990s. Too few reliability estimates were reported in peer-review journals (Meier & Davis, 1990; Willson,1980). Therefore, APA recommends authors report reliability coefficients using their own data (instead of referring to test manuals or previous studies), because the same instrument can yield different score reliability coefficients when tested among different populations (Fan & Thompson, 2001). Despite the recommendation, however, the low reporting rate of reliability estimates has not improved much during recent decades. We reiterate the call for social science researchers to report the reliability coefficients of the scores from data they analyze.

Alongside improvements in reporting of reliability, another future direction is to improve the quality of the translation/modification process with a well-illustrated translation standard. Based on the cross-cultural translation guidelines (Beaton et al., 2002; Sperber et al., 1994; Wild et al., 2005) identified in our review, we synthesized following steps for future cross-language health literacy instrument translation: 1) two bilingual health professionals develop two forward translations independently; 2) these two translators and a third native-speaking health professional in the target language, who does not participate in the forward translation, compare the two forward translation versions, discuss the discrepancies, and reach consensus; 3) a bilingual speaker who is not familiar with the health content (to avoid information bias) back translates the target language version into the original language; 4) all translators meet to review and compare the back-translated version with the original one to revise; 5) local community health workers/professionals from target language provide feedback and revise the translation; 6) 5-8 individuals who participate in cognitive interviews further identify and revise the problematic items; and 7) researchers/authors proof-read the final version. Readers should be aware that reporting a specific translation step does not mean that such a step has been completed accurately. We also recommend that future studies apply the Theory of Test Translation Error to critically appraise the translated items of a specific instrument (Solano-Flores, Contreras-Niño, & Backhoff, 2013).

Another possible direction for future study is to apply RG to investigate other health literacy measures since the TOFHLA suffers from several limitations. TOFHLA is a widely used instrument to measure health literacy, but it has been criticized for not reflecting the broader components in health literacy theory (Kindig et al., 2004). TOFHLA focuses on reading and measures health literacy at the individual level while health literacy involves multiple dimensions at different levels (Nutbeam, 2000). We recommend using RG meta-analysis to explore the measurement issues among theory-based health literacy instruments.

Contributions

This study identified the non-English TOFHLA instruments in the literature, synthesized their translation processes, and examined the impact of test and sample characteristics on the Cronbach's α coefficients of these instruments. The translation process varied from study to study. For future TOFHLA implication, we recommend researchers to use the short form rather than the long form. We also call for further studies to improve the language translation quality of non-English health literacy measures.

ASSESSING FUNCTIONAL HEALTH LITERACY AMONG U.S. NATIVE CHINESE SPEAKERS WITH LIMITED ENGLISH PROFICIENCY

Introduction

Populations with language barriers face health disparity challenges. According to the 2015 U.S. Census Bureau, about 8.6% of the population (26 million people) have limited English proficiency (US Census Bureau, 2015). Individual with limited English proficiency (LEP) refers to anyone ages 5 and above who speaks English “less than very well” (Pandya, Batalova, & McHugh, 2011). In clinical settings, patients with LEP suffer higher risk of inadequate informed consent (Lee et al., 2017), lower rates of medical appointment adherences (Andreae et al., 2016), poorer patient-physician interaction (Smith, 2010), and have difficulty explaining their illness as well as understanding their doctor’s instructions and treatment plan (Lindholm, Hargraves, Ferguson, & Reed, 2012). Regarding public health, studies also show that populations with LEP are difficult to recruit for health education programs (Thomson & Hoffman-Goetz, 2011), have limited access to health care (Bruce, Schwei, Park, & Jacobs, 2014), poor overall health status (Sentell & Braun, 2012), and low health literacy (Institute of Medicine, 2004).

Assessing health literacy is the first step in starting a health education program (Thomason & Mayo, 2015). Health literacy refers to “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” (Institute of Medicine, 2004;

Ratzan & Parker, 2000, p. vi; U.S. Department of Health and Human Services, 2000). Health literacy measurement tools serve as an essential component of needs assessment and provide the foundation for an intervention plan. A precise tool for assessing health literacy allows clinicians to tailor health services for patients. The tools also help health professionals to shape patient-provider communication and reduce health disparities (Nørgaard, Sørensen, Maindal, & Kayser, 2014; Stonbraker, Schnall, & Larson, 2015; Batterham, Hawkins, Collins, Buchbinder, & Osborne, 2016). Thus, it is important to assess health literacy level among populations with LEP.

Theoretical Framework

Theory-based assessments and programs are more effective than atheoretical ones (Goodson, 2009). Health literacy measures must be based on sound theory (Pleasant, McKinney, & Rikard, 2011). Although there is no single accepted theory for health literacy (Sørensen et al., 2012), Nutbeam's (2000) health literacy conceptual model is a foundation for developing theory-based instruments (Haun, Valerio, McCormack, Sørensen, & Paasche-Orlow, 2014). His conceptual model has three dimensions: functional health literacy, interactive health literacy, and critical health literacy (Nutbeam, 2000). Functional health literacy represents the ability to understand factual information and use health services. Interactive health literacy represents the ability related to social skills. Critical health literacy represents the ability to empower self, family, and community.

Based on Nutbeam's (2000) conceptual model, health literacy is a multidimensional construct. Within this construct, functional health literacy represents the most basic skills such as reading and writing. Indeed, functional literacy is the basic dimension of many literacy domains (Frisch, Camerini, Diviani, & Schulz, 2012). Functional literacy is the "ability to identify, understand, interpret, create, communicate, compute and use printed and written materials associated with varying contexts" (UNESCO, 2004, p. 13). Functional health literacy refers to the ability to apply literacy skills to health-related context such as prescriptions, medicine labels, and appointment cards (Parker, Baker, William, & Nurss, 1995). Reading, writing, and numeracy skills are essential components of functional health literacy (Parker et al., 1995). Thus, as currently defined and applied in the US, the conceptual model renders functional health literacy to be highly associated with English proficiency.

Due to the close relationship between English language proficiency and functional health literacy, there is a critical need to differentiate low English proficiency from low functional health literacy. In education and psychology field, it is commonly accepted that a test of any content area is, to some extent, a test of the proficiency in the language in which the test is administered (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999; Solano-Flores, 2008). Thus, if this statement holds true in health literacy research, then assessing functional health literacy among populations with LEP using an English measurement tool could generate misleading results. Solano-Flores and Trumbull (2003) noted that using the same items in both English and test-takers' native languages could help researchers better understand the test-takers' content knowledge and the interactions between native and English language proficiency. Therefore, we assessed functional health literacy among people with LEP using a functional health literacy assessment tool with two language versions: English and Chinese.

Health Literacy Assessments

English S-TOFHLA

The Test of Functional Health Literacy in Adults (TOFHLA) is considered the "gold standard" for measuring functional health literacy because of its strong reliability and validity data in English (Mancuso, 2009). The TOFHLA was developed in English (in the US) with 67 multiple-choice items assessing two components of functional health literacy: Reading and Numeracy (Parker et al., 1995). The TOFHLA uses the Cloze procedure (Taylor, 1953) that omits every fifth to seventh word in a passage and asks test-takers to select the correct answer from four possible choices. Incorrect choices contain grammatical or contextual errors.

Later, the same research team (Baker, William, Parker, Gazmararian, & Nurss, 1999) developed a short version the TOFHLA (S-TOFHLA), reducing the number of items to 40. The Reading section now contains 36 multiple-choice items and the Numeracy section contains 4 open-ended questions. In addition, the Reading section contains two passages about (1) instructions on preparation for an upper gastrointestinal series and (2) a Medicaid application form. The Numeracy section assesses test-takers' understanding of instructions for taking medicines, monitoring blood glucose, and keeping clinic appointments. The test items employ actual prescription labels and hospital forms. See Table 6 for example items. The S-TOFHLA score ranges from 0 to 100, with 2 points for each reading item and 7 points for each numeracy item. Cronbach's α was 0.97 for Reading and 0.68 for Numeracy in the original scale development study (Baker et al., 1999).

Table 6. English S-TOFHLA Example Items

Variables	# of items	Original Item Example	Correct Answer
Reading Comprehension	36	Your doctor has sent you to have a ____ X-ray. a. stomach b. diabetes c. stitches d. germs	a
Numeracy	4	Normal blood sugar is 60-150. Your blood sugar today is 160. Question: If this was your score, would your blood sugar be normal today?	No

Chinese S-TOFHLA

Two studies (Tang, Pang, Chan, Yeung, & Yeung, 2008; Mantwill & Schulz, 2016) developed the Chinese S-TOFHLA in Traditional and Simplified Chinese respectively. Tang and colleagues (2008) translated/modified the English S-TOFHLA into Traditional Chinese and applied it in Hong Kong. Similar to the original version, the Reading section contains two passages that commonly used in actual Chinese hospital settings: (1) instructions to patients on how to prepare for a colonoscopic examination and (2) a Medicaid application form including information about patients' rights and responsibilities prior to hospital admission. For the Numeracy section, Tang et al. (2008) translated the original English questions into Traditional Chinese. More recently, Mantwill and Schulz (2016) converted the Traditional Chinese instrument into Simplified Chinese and used it in Mainland China. Cronbach's α was 0.96 for Reading and 0.63 for Numeracy in the Traditional Chinese scale development study (Tang et al., 2008). For the Simplified Chinese study, Cronbach's α was 0.94 for Reading and 0.90 for Numeracy (Mantwill & Schulz, 2016).

Literature Gap

Critical examination of the extant research literature about assessing health literacy reveals a major gap related to measurement reliability and validity. First, few culturally and linguistically appropriate instruments have been developed for minority populations, especially those with LEP (Nguyen et al., 2015). Using an English instrument to assess functional health literacy likely leads to low validity because English proficiency may not be distinguishable from functional health literacy in the assessment for populations with LEP.

Second, few studies reported psychometric properties of their health literacy measures (Haun et al., 2014). For example, Chen and colleagues (cite dissertation article 1) found only 20% of the studies that measured functional health literacy using non-English TOFHLA instruments reported reliability coefficients. Also, Haun and colleagues (2014) pointed out that some health literacy measurement studies applied factor analysis and regression to examine the psychometric properties; yet, most studies did not fully investigate various types of validity such as content and construct validity.

Such lack of reporting critical information raises more questions about the adequacy of health literacy instruments to elicit reliable and valid scores.

Finally, most health literacy instruments have been evaluated for psychometric properties using Classical Test Theory (CTT) instead of Item Response Theory (IRT) for data analysis (Nguyen et al., 2015). CTT is a traditional statistical model for estimating the reliability of test scores. IRT was developed to address and overcome shortcomings of CTT (Hambleton, Swaminathan, & Rogers, 1991). For instance, IRT puts item difficulty and person ability on the same scale so we can predict a test-taker's probability of answering certain items correctly or incorrectly; CTT, on the other hand, provides no information about which tasks can be done successfully by a person (Hambleton et al., 1991).

IRT may be better suited to test the psychometric properties of the TOFHLA because IRT is a more recent statistical model focuses on item level analyses. CTT is item oriented and it does not give much information of each item in constructing a test; however, IRT is item oriented and it models how examinees respond to a particular item (Hambleton et al., 1991). For example, persons A and B received the same TOFHLA scores, but person A answered more difficult items correctly while person B's correct items were from easier questions. Despite receiving identical test scores, their latent ability level should be different because person A was more capable of answering difficult questions than person B. Under CTT, we will conclude these two people had the same level of functional health literacy because they received the same raw score. Under IRT, we will conclude person A had higher functional health literacy level because he/she had higher person ability parameters. Given the potential for IRT approach to improve measurement of the health literacy construct in such situation, there is a critical need for using IRT to investigate the psychometric properties of functional health literacy measures among populations with LEP.

The purpose of this study is to investigate the psychometric properties of two functional health literacy assessment instruments (English and Chinese S-TOFHLA) using IRT among U.S. native Chinese speakers with LEP. The following research questions guided this study:

1. Is there a significant difference between the scores of English and Chinese S-TOFHLA?
2. Does English proficiency impact the accuracy of the English S-TOFHLA assessment?
3. What are the item difficulty and item discrimination parameters yielded by IRT for these two instruments?

Methods

Health Literacy Measures

We have provided detailed information about English S-TOFHLA and Chinese S-TOFHLA in the introduction and have listed the sample items in Table1. To eliminate the guessing parameter for the tests, we added an “I don’t know” choice for each item. Cronbach’s α was 0.99 for the English S-TOFHLA Reading items and 0.94 for the English Numeracy items, with an overall Cronbach’s α of 0.99. Cronbach’s α was 0.83

for the Chinese S-TOFHLA Reading items and 0.54 for the Chinese Numeracy items, with an overall Cronbach's α of 0.84 (see Table 7).

Table 7. Cronbach's α of English and Chinese S-TOFHLA

Instrument	Reading	Numeracy	Overall
English S-TOFHLA	0.99	0.94	0.99
Chinese S-TOFHLA	0.83	0.54	0.84

Participants

A total of 405 participants (158 male and 247 female) completed the study. Their ages ranged from 18 to 96 (M = 51.70, SD = 19.31). Participants' time in the US ranged from 1 month to 74 years (M = 18 years, SD = 14). In our sample, 30.12% considered themselves to speak English "well", 38.27% "fair", 22.47% "not well", and 9.14% "not at all". More than half of the participants (55.31%) had bachelor degrees and above; 20.99% had technique or associate degrees; and 23.70% had high school level and below. About 42.47% of them were not employed (either unemployed, retired, or stay-home mom/dad), 34.07% were employed (either full- time or part-time), 19.75% were students, and 2.47% were business owners. We provide detailed socio-demographic information in Table 8.

Table 8. Participants' Socio-demographic Information

Items	***N***	***%***
English Speaking Proficiency		
Well	122	30.12%
Fair	155	38.27%
Not well	91	22.47%
Not at all	37	9.14%
Gender		
Male	158	39.01%
Female	247	60.99%
Education		
High school and below	96	23.70%
Technique or Associate	85	20.99%
Bachelor	67	16.54%
Master	110	27.16%
PhD and above	47	11.60%
Occupation		
Student	80	19.75%
Full time/Part time employee	138	34.07%
Business owner	10	2.47%
Unemployed	32	7.90%
Retired	118	29.14%
Stay-home mom/dad	22	5.43%
Other	5	1.23%
Total	405	100%

Procedures

We distributed study flyers through health professional organizations, community centers, and churches in an urban region in the southwestern US to recruit potential participants. To be eligible for this study, the participants needed to be: (1) 18 or older, (2) native speakers of Chinese (either Mandarin or Cantonese), (3) literate in Chinese (either simplified or traditional Chinese), and (4) speak English "less than very well". Based on the definition of LEP: any person ages 5 and older who speaks English ''less than very well'' (Pandya, Batalova, & McHugh, 2011), we used one pre-screening question to assess participants' English proficiency. We asked potential participants to self-evaluate their English-speaking ability (a. very well, b. well, c. not well, d. not at all). Only those who chose well, not well, or not at all (b, c, or d) were eligible to participate in this study.

We asked the participants to take the English S-TOFHLA first followed by the Chinese S-TOFHLA. To avoid rest-retest bias and minimize the threats to internal validity, participants could not go back to the English S-TOFHLA questions once they started the Chinese S-TOFHLA. Robinson and colleagues (2011) pointed out that the timed S-TOFHLA could cause false low health literacy, especially among those who need more time to complete the test. Thus, we did not set time limits for returning the questions. We also asked the participants some demographic questions such as age, gender, education, and occupation after the TOFHLA.

Participants chose their preferred Chinese version (either traditional or simplified). We provided both traditional and simplified Chinese versions for participants because most Chinese speakers understand both versions but they have a preference between these two. Compared to the simplified Chinese, the traditional Chinese contains more complicated characters. Before 1949, traditional Chinese was the only written language for Chinese. The Chinese government simplified the characters to promote literacy in 1949. Immigrants from Taiwan, Hong Kong, and Macau generally use the traditional version, but the simplified version is often used among Mainland Chinese immigrants (Zhou & Cai, 2002).

When participants returned the paper-pencil questions, research staff scanned through to make sure there was no accidental missing item. Each participant received a $10 grocery gift card as incentive after completing the questions. Data collection was conducted from June to September 2016. The Institutional Review Board of the authors' institutional affiliation approved this study (IRB2016-0092D).

Data Analysis

We used descriptive analysis and paired t-test to compare the English S-TOFHLA and Chinese S-TOFHLA scores when stratified participants based on English-speaking proficiency and education level. We applied IRT to investigate the psychometric properties (i.e., item difficulty and item discrimination) of these two instruments. All S-TOFHLA items were dichotomous because they were scored either correct (coded as 1) or incorrect (coded as 0). One-, two-, and three-parameter logistic (1PL, 2PL, 3PL) IRT models are appropriate for dichotomous data (Hambleton et al., 1991). Comparing to 1PL, the 2PL model estimates more parameters, where items vary in their difficulty level and discrimination level (Embretson & Reise, 2000). Since we added the "I don't know" choice to eliminate the guessing value, 3PL that including the guessing parameter is not appropriate for our data. We performed descriptive analysis and paired t-test using SPSS 22 and the 2PL IRT model using IRTPRO 3.

2PL Model Fit

Unidimensionality is a key assumption for IRT, which assumes a single latent construct underlie item performance. Therefore, we analyzed the Reading (36 items) and Numeracy (4 items) for the English and Chinese S-TOFHLA items as four separate models (English Reading, English Numeracy, Chinese Reading, and Chinese Numeracy). We tested the 2PL model fit for the four 2PL models using the global fit statistics: the root mean square error of approximation (RMSEA). RMSEA values that are below 0.06 indicate good model fit (Edelen & Reeve, 2007). The global fit statistics indicated that all the four models had good fit for the 2PL IRT: English Reading items (RMSEA < 0.001), Chinese Reading items (RMSEA = 0.02), English Numeracy items (RMSEA < 0.001), and Chinese Numeracy items (RMSEA < 0.001).

Results

The mean score of the English S-TOFHLA was 60.52 (SD = 42.61) and 89.50 (SD = 10.78) points for the Chinese S-TOFHLA. Based on the English S-TOFHLA scores, 141 participants (34.8%) were classified as having inadequate health literacy, 8 participants (2%) as having marginal health literacy, and 256 participants (63.2%) as having adequate health literacy. Based on the Chinese S-TOFHLA scores, 7 participants (1.7%) were classified as having inadequate health literacy, 11 (4.4%) as having marginal health literacy, and 387 (95.6%) as having adequate health literacy.

Research Question #1: Is there a significant difference between the scores of English and Chinese S-TOFHLA?

As shown from the paired t-test results in Table 9, overall, participants' Chinese S-TOFHLA scores were significantly higher than their English S-TOFHLA scores (*p* < .001). As the English-speaking proficiency and education level increased, the gap between English and Chinese S-TOFHLA scores decreased. Participants who spoke English "less than well" had significantly higher Chinese S-TOFHLA scores than the English scores. The scores were not significantly different among the participants who spoke English "well". Participants with education level at Bachelor and below received significantly higher Chinese S-TOFHLA scores than English scores. Scores did not differ among the participants with Master and PhD degrees.

Table 9. Comparing English S-TOFHLA and Chinese S-TOFHLA Scores

Items	***N***	***English S-TOFHLA Mean (SD)***	***Chinese S-TOFHLA Mean (SD)***	***p***
English Speaking Proficiency				
Well	122	88.78 (18.54)	90.79 (8.43)	.218
Fair	155	73.32 (35.13)	88.93 (11.17)	<.001
Not well	91	23.98 (37.81)	88.57 (12.57)	<.001
Not at all	37	3.57 (15.54)	89.97 (11.11)	<.001
Education				
High school and below	96	24.33 (39.95)	89.80 (10.98)	<.001
Technique	7	34.14 (43.71)	88.14 (14.66)	<.05
Associate	78	50.63 (43.69)	88.10 (13.69)	<.001
Bachelor	67	70.93 (38.61)	91.49 (7.56)	<.001
Master	110	85.75 (21.56)	89.52 (9.01)	.075
PhD and above	47	80.87 (29.21)	88.55 (11.94)	.078
Total	405	60.52 (42.61)	89.50 (10.78)	<.001

Research Question #2: Does English proficiency impact the accuracy of the English S-TOFHLA assessment?

Chen and Thissen's (1997) chi-square statistic checks the unidimensionality assumption. Chi-square results showed that the unidimensionality assumption has been violated for the Reading items of the English S-TOFHLA because all the standardized chi-square values for each item pair exceeded 1.96. The unidimensionality assumption has been met for the Reading items of Chinese TOFHLA. In other words, the Reading items of the English S-TOFHLA were affected by more than one latent trait. Therefore, these items were not purely accessing participants' functional health literacy. The unidimensionality assumption was met for the Numeracy items for both the English and Chinese S-TOFHLA. Unidimensionality assumption information for each of the items is displayed in the Appendix A.

Research Question #3: What are the item difficulty and item discrimination parameters yielded by IRT for these two instruments?

English S-TOFHLA Reading Items vs. Chinese S-TOFHLA Reading Items

Difficulty and discrimination information for each of the items is shown in the Appendix B. The item difficulty parameters *b* for the 36 English S-TOFHLA Reading items ranged from 0.02 to 0.67. The easiest items (b_{Item12} = 0.02, b_{Item15} = 0.04, b_{Item13} = b_{Item16} = 0.13) were from the first passage—instructions on preparation for an upper gastrointestinal series. The most difficult items (b_{Item34} = 0.67, b_{Item21} = 0.60, b_{Item31} = 0.56) were from the second passage—a U.S. Medicaid application form. About 67.90%

participants answered the easiest item correctly and 43.46% answered the most difficulty item correctly. The item discrimination parameters *a* for these English Reading items raged from 3.57 to 11.06. The overall test information of the English Reading items (Figure 3) indicated the test was not able to discriminate well across all ability levels but only so at higher middle ability level.

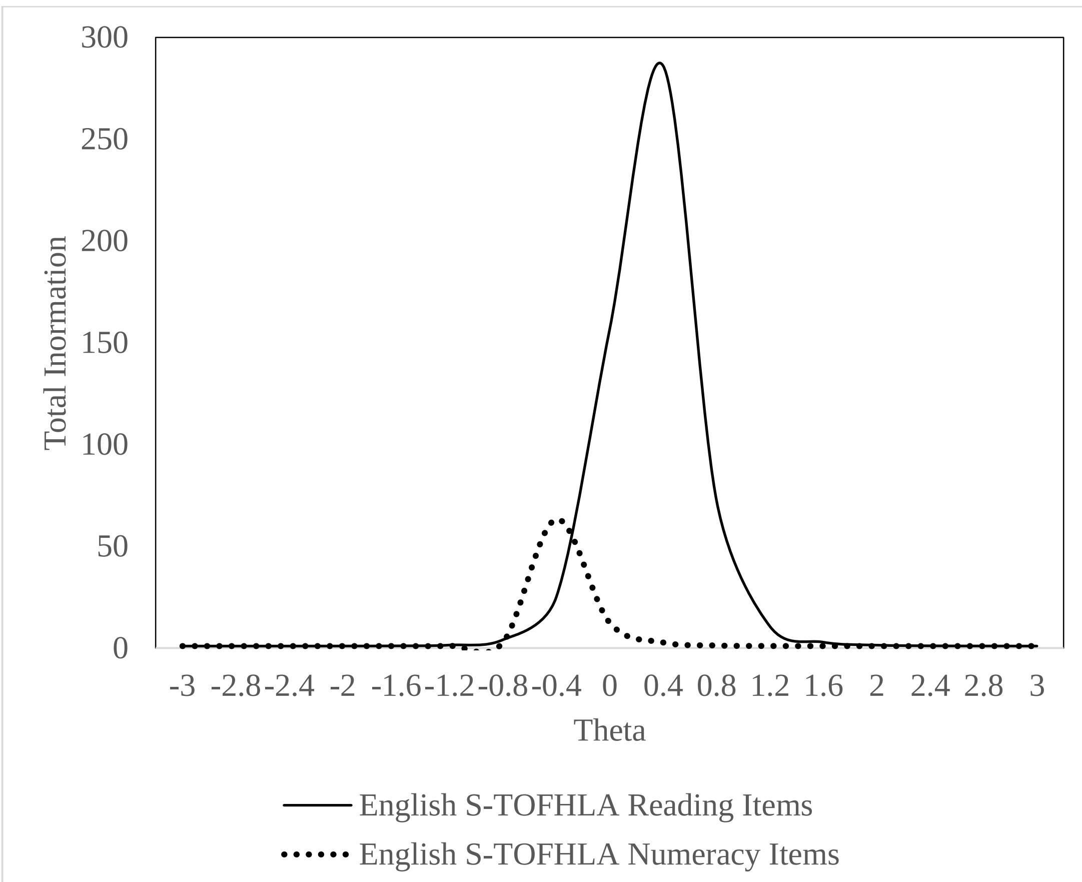

Figure 3. Test Information Curve for the English S-TOFHLA

The item difficulty parameters *b* for the 36 Chinese S-TOFHLA Reading items ranged from -5.09 to 1.15. The easiest items (b_{Item4} = -5.09, b_{Item11} = -3.45, b_{Item8} = -3.17) were from the first passage— instructions on preparation for a colonoscopic examination. This most difficult item (b_{Item29} = 1.15) was from the second passage—a Medicaid application form that used in actual Chinese hospital settings. About 97.28% participants answered the easiest item correctly and 41.23% answered the most difficult item correctly. The item discrimination parameters *a* for these Chinese Reading items raged from 0.31 to 2.82. The overall test information of the Chinese Reading items (Figure 4) indicated the test was able to discriminate well across all ability levels but more so at low ability level.

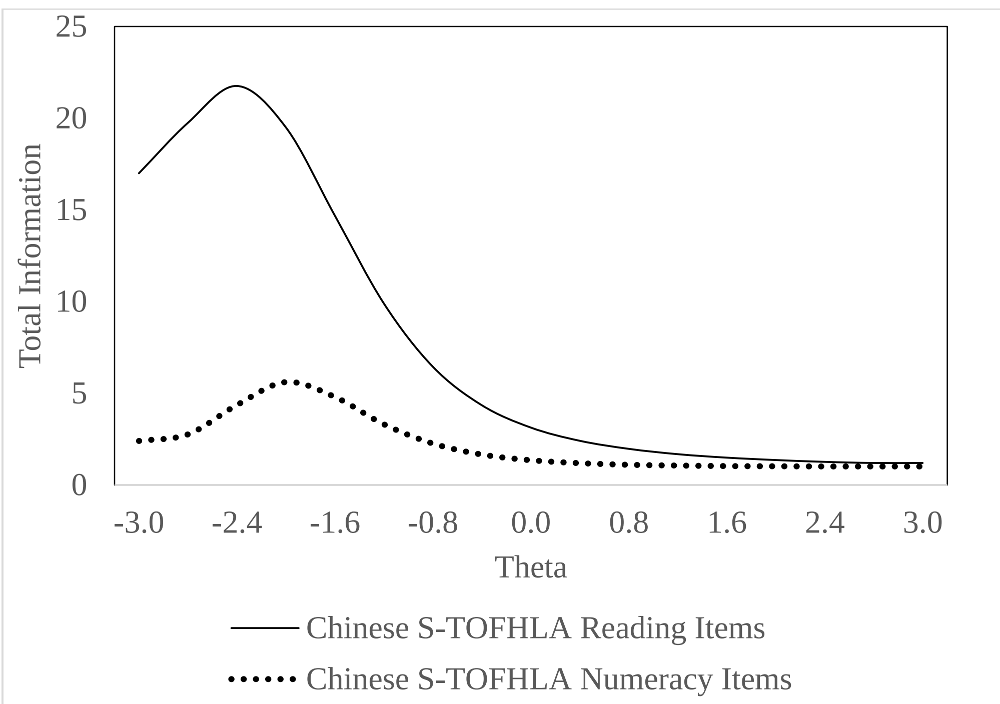

Figure 4. Test Information Curve for the Chinese S-TOFHLA

Compared to the Chinese S-THOFLA Reading items, the English S-TOFHLA Reading items had higher item difficulty and item discrimination parameters. As the test characteristic curve shown in Figure 5, the expected score for 36 Reading items ranged from 0 to 72 because each correct Reading item carries 2 points. For the English Reading items, people with latent ability (theta) lower than -1 tended to receive low scores that were close to 0 point; while people with latent ability (theta) higher than 1 tended to receive high scores that were close to full points. The English S-TOFHLA Reading items were very sensitive to detect the latent ability change among people with middle theta ranged from 0 to 1 because the curve had a steep slope. For the Chinese Reading items, people with theta lower than -3 tended to receive low scores; while people with theta higher than -1 tended to receive high scores. The Chinese S-TOFHLA Reading item were sensitive to detect the latent ability change among people with lower theta ranged from -3 to -1.

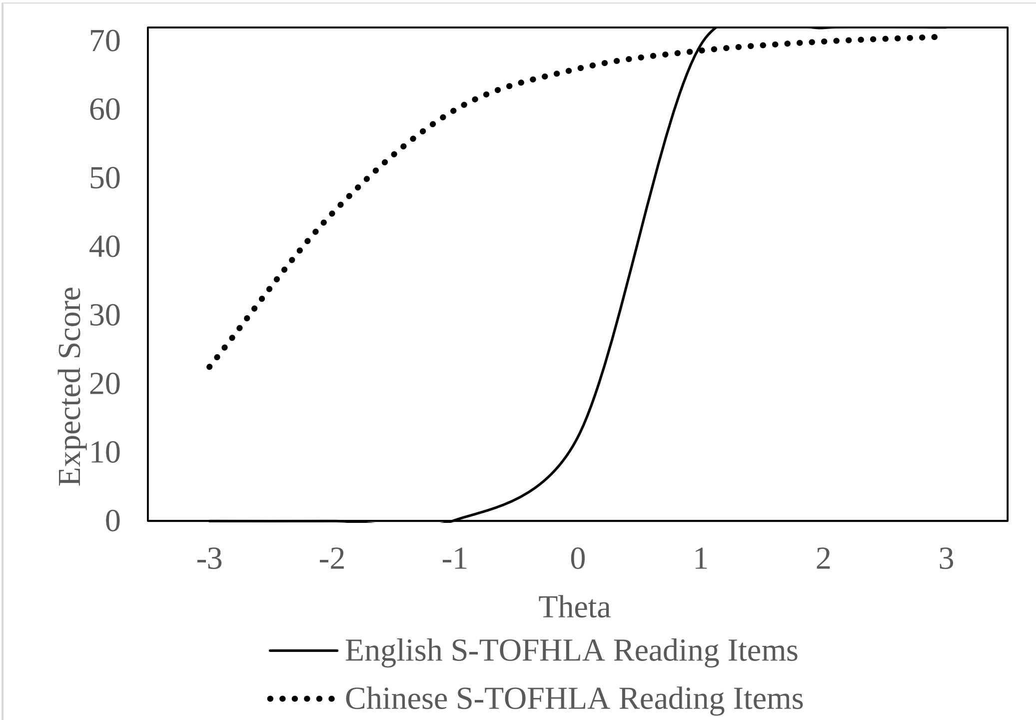

Figure 5. Test Characteristic Curve for the S-TOFHLA Reading Items

English S-TOFHLA Numeracy Items vs. Chinese S-TOFHLA Numeracy Items

The item difficulty parameters (*b*) for the 4 English S-TOFHLA Numeracy items ranged from -0.33 to -0.11. Both the easiest item (b_{Item1} = -0.33) and the most difficult item (b_{Item4} = -0.11) assessed the ability to understand instructions to take medicines. About 62.47% of the participants answered the easiest item correctly and 52.84% of the participants answered the most difficult item correctly. The item discrimination parameters *a* for these English Numeracy items raged from 4.95 to 99.55. The overall test information of the English Numeracy items (Figure 3) indicated the test was not able to discriminate well across all ability levels but only so at lower middle ability level.

The item difficulty parameters *b* for the 4 Chinese S-TOFHLA Numeracy items ranged from -2.11 to -1.88. The easiest item (b_{Item1} = -2.11) tested the ability to understand instructions to take medicines. The most difficult item (b_{Item3} = -1.63) assessed the ability to understand a clinic appointment card. About 92.99% of the participants answered the easiest item correctly and 92.20% of the participants answered the most difficulty item correctly. The item discrimination parameters *a* for these Chinese Numeracy items raged from 1.18 to 3.23. The overall test information of the Chinese Numeracy items (Figure 4) indicated the test was able to discriminate well across all ability levels but more so at low ability level.

Compared to the Chinese S-THOFLA Numeracy items, the English S-TOFHLA Numeracy items had higher item difficulty and item discrimination parameters. As the test characteristic curve shown in Figure 6, the expected score for 4 Numeracy items ranged from 0 to 28 because each correct Numeracy item carries 7 points. For the English Numeracy items, similar to the English Reading item, people with latent ability (theta) lower than -1 tended to receive low scores that were close to 0 point; while people with latent ability (theta) higher than 1 tended to receive high scores that were close to full points. The English S-TOFHLA Numeracy items were very sensitive to detect the latent ability change among people with middle theta ranged from -1 to 0 because the curve had a steep slope. For the Chinese Numeracy items, similar to the Chinese Reading items, people with theta lower than -3 tended to receive low scores; while people with theta higher than -1 tended to receive high scores. The Chinese S-

TOFHLA Reading items were sensitive to detect the latent ability change among people with lower theta ranged from -3 to -1.

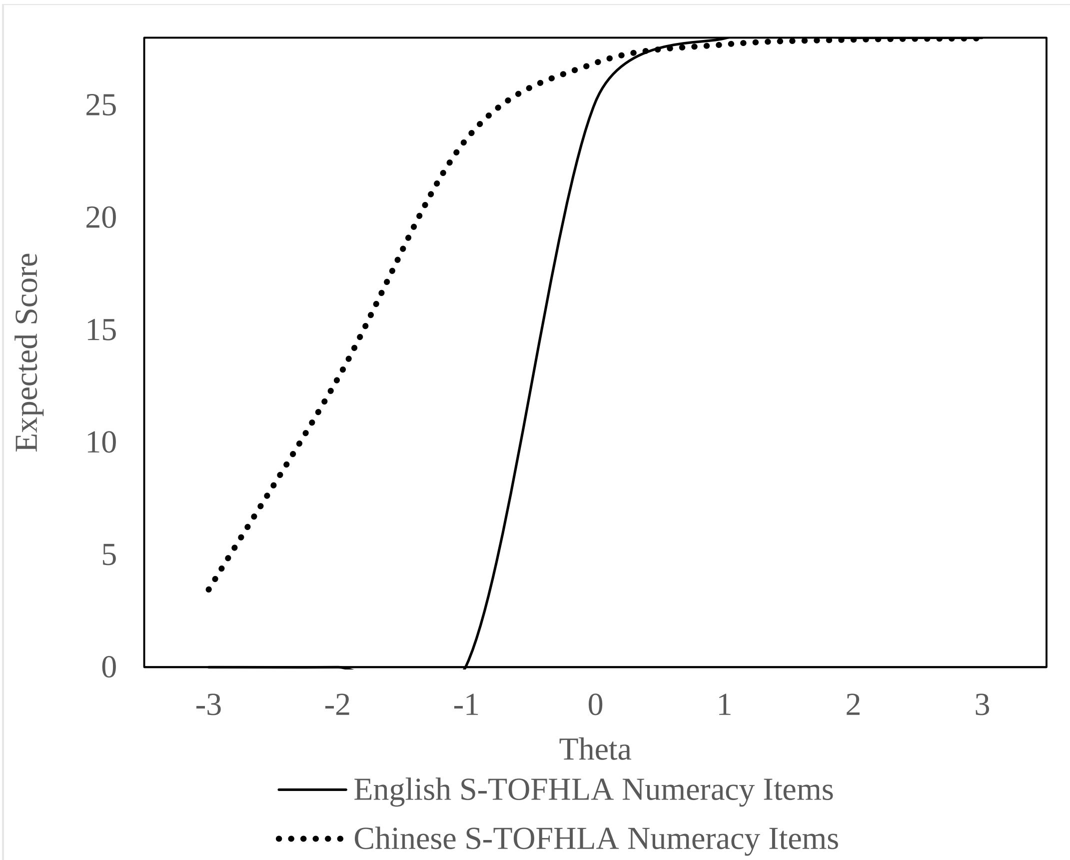

Figure 6. Test Characteristic Curve for the S-TOFHLA Numeracy Items

Discussion

This study found the current functional health literacy construct and measurement tools are not applicable among populations with LEP. Such finding is supported by three key results. First, most individuals with LEP had adequate functional health literacy when measured in their native languages. Second, the English S-TOFHLA was measuring functional health literacy along with language proficiency. Third, the Numeracy items of the Chinese S-TOFHLA exhibited low reliability.

English language proficiency is one of the major causes reducing English functional health literacy levels among populations with LEP. After removing the language barrier, most participants had adequate functional health literacy skills in their native languages. More participants were classified as having adequate functional health literacy based on the scores of Chinese S-TOFHLA than English S-TOFHLA (95.6% vs 63.2%). Thus, researchers should be cautious about using English instrument tools to measure functional health literacy level among individuals with LEP.

The IRT unidimentionality assumption was violated among the English S-TOFHLA Reading items, showing the English S-TOFHLA was not measuring functional health literacy alone but assessing English proficiency at the same time. Therefore, the item difficulty and discrimination parameters of the English S-TOFHLA Reading items might not be estimated precisely. Such assumption violation indicates measuring health literacy among populations with LEP is complex. Our finding further confirmed that the existing functional health literacy measures and conceptual models are insufficient to capture the communication interactions among populations with LEP

(Yip, 2012). Such gap requires more attention to definition, operationalization, measurement and application for future health literacy studies. We believe culturally related constructs such as English language proficiency are key components that should be added to the health literacy model for populations with LEP.

Further, the most difficult items located in the second passage of the English S-TOFHLA Reading section, showing health care system navigating is another important component of health literacy among populations with LEP. The second passage is a U.S. Medicaid application form. Many participants had difficulty understanding and terms related to the Medicaid context such as TANF (Temporary Assistance for Needy Families) because they were not familiar with the Medicaid terminologies. That is one of the reason our participants tended to receive low score on this reading passage.

We found the α coefficient of the Chinese Numeracy items was low. The scores of the English Reading, English Numeracy, and Chinese Reading items exhibited high internal consistency ($\alpha_{\text{English Reading}} = 0.99$, $\alpha_{\text{English Numeracy}} = 0.94$, $\alpha_{\text{Chinese Reading}} = 0.83$). However, the internal consistency level for the Chinese Numeracy items was low ($\alpha_{\text{Chinese Numeracy}} = 0.54$). The α coefficient were not high among the Numeracy items in Baker et al. (1999) original English S-TOFHLA study and Tang et al. (2008) Traditional Chinese S-TOFHLA either ($\alpha_{\text{English Numeracy}} = 0.68$, $\alpha_{\text{Traditional Chinese Numeracy}} = 0.63$). Baker et al. (1999) admitted that the reliability of the Numeracy items was only modest, but they did not mention what might be the cause.

We believe the reliability was low among the Chinese Numeracy items in our study due to the low item difficulty levels of the Chinese measure. The Chinese

Numeracy items were easy for our participants. The scores for the Chinese Numeracy items had a low α because many participants answered them correctly. Therefore, like the original S-TOFHLA study, our study showed the problem of low reliability for the Numeracy items as well.

The high reliability of the English Numeracy items in our study did not mean these items yield high validity for populations with LEP. Among our participants, only half were able to answer the English Numeracy items correctly. The scores for the English Numeracy items had a high α because many participants consistently answered these items incorrectly.

Further, we identified one problematic question in the Chinese Numeracy section that might reduce reliability. The third question assesses individual's understanding of a clinic appointment card. This question asks participants when is the time for the next appointment according to the information on that card. When working on this question, many participants said they were confused by the appointment card. Many of them had never seen a clinic appointment card before. Most of our participants did not know the purpose of this card. Some participants even neglected the information on the card and tried to answer this question based their own health situation. For example, some participants said they should visit the doctor every 3 months because they have diabetes. Shaw and colleagues (2012) also reported that many ethnic minority participants substituted their own illness or health care experience for the S-TOFHLA Numeracy questions. Therefore, the third Numeracy question is not a valid item to measure functional health literacy among immigrant populations who are not familiar with the U.S. healthcare system.

Another reason caused/created problems with the Chinese S-TOFHLA Numeracy items was the instrument translation process. Tang and colleagues (2008) changed the Reading items for the Chinese S-TOFHLA using actual hospital materials in China; however, the Numeracy items were directly translated from English into Chinese (Tang et al, 2008). Thus, the Chinese Numeracy items did not take into account the cultural and health care system differences between the US and China. Previous studies (Stonbraker, Schnall, & Larson, 2015; cite dissertation Article 1) indicate that non-English health literacy measures developed based on original English instruments should pay close attention to the health care system and cultural discrepancies between the country/population where the original instrument was developed and the country/population where the original instrument has been adapted.

Future Study

Both the English and Chinese S-TOFHLA failed to capture the full aspects of health literacy. Health literacy is a multidimensional construct that goes beyond reading and writing (Institute of Medicine, 2004). In fact, reading and numeracy skills comprise one part of health literacy—functional health literacy (Nutbeam, 2000). According to Nutbeams (2000), health literacy also includes: communication and social skills—interactive health literacy and analyzing skills—critical health literacy. Yet, S-TOFHLA only assesses functional health literacy instead of the multidimensional health literacy construct. S-TOFHLA has been criticized for not reflecting the broader components of health literacy (Institute of Medicine, 2004), not being user-friendly for older people because of its Cloze format (Ownby, Acevedo, Waldrop-Valverde, Jacobs, Homs, & Czaja, 2013), and suffering from ceiling effects—most people receive high scores (Baker et al., 1999). Therefore, researchers are calling for the development and use of theory-based measures to capture the entire health literacy construct (Pleasant et al., 2011; Altin, Finke, Kautz-Freimuth, & Stock, 2014).

Moreover, future studies should develop tailored health literacy measurement and interventions targeting linguistic minority populations. Researchers should consider the problems of language and health care system differences when measuring health literacy among populations with LEP. There is a critical need to include both the components of English language instructions and health literacy education in future health literacy interventions among populations with LEP (Chen, Goodson, & Acosta, 2015). Developing such interventions requires transdisciplinary collaborations among health professionals, ESL practitioners, literacy researchers, and adult education experts (Soto Mas, Jacobson, & Olivarez, 2017; Chen et al., 2015). Health education interventions linking native language skills with English language skills could help improve health literacy for the target populations. For example, supplementing English materials with materials in participants' native languages could help transit their previous knowledge from their native languages into English. Research shows that assisting students in transferring their native language skills to the learning of English has a positive effect on English proficiency development (Cummins, 2009; August, Shanahan, & Escamilla, 2009).

Limitations

The present study has some limitations. First, the S-TOFHLA did not fully measure the multidimensional health literacy construct but only functional health literacy. Second, we applied convenience sampling method for the data collection. Moreover, we only targeted U.S. native Chinese speakers with limited English proficiency. Thus, the sample might not be representative of the Chinese American community. Third, the Numeracy items in the Chinese S-TFOHLA were directly translated from the original English Numeracy items. The Chinese Numeracy items might have test-retest bias. The Reading items were different between the English and Chinese S-TOFHLA because the Chinese version substitutes the original passages with materials from actual Chinese hospitals.

Contributions

The present study investigated the psychometric properties of two health literacy measures (English S-TOFHLA and Chinese S-TOFHLA) when administered to U.S. Chinese speakers with limited English proficiency. We believe this study contributes to the literature on measuring health literacy in three ways. First, our study targets vulnerable populations with LEP (e.g., immigrants and linguistic minorities) who live in an English dominant society. Second, our study differentiates language proficiency from functional health literacy level by assessing health literacy with measures in two different languages (English and Chinese). Third, we performed IRT, which was developed to overcome the shortcomings of CTT, for examining psychometric properties of measures. To the best of our knowledge, this is the first study using IRT to compare the item difficulty and discrimination parameters of two functional health literacy instruments among populations with LEP.

ASSESSING HEALTH LITERACY AMONG POPULATIONS WITH LIMITED ENGLISH PROFICIENCY: A CASE OF CHINESE AMERICAN IMMIGRANTS

Introduction

The prevalence of low health literacy in the United States is especially high among populations with limited English proficiency; however, few health literacy instruments have been developed or validated for these linguistic minority groups (McKee & Paasche-Orlow, 2012). Individuals with limited English proficiency (LEP) refers to anyone ages 5 and above who speaks English "less than very well" (Pandya, Batalova, & McHugh, 2011). About 8.6% of the population (26 million people) have LEP (US Census Bureau, 2015). Researchers and health practitioners should not assume universal applicability of health literacy instruments; studies should consider native languages, culture, and health systems before conducting health literacy assessments (Dowse, Lecoko, & Ehlers, 2010). Further, Yip (2012) pointed out the current health literacy instruments were incapable of capturing the complex interaction among populations with LEP. Thus, there is a critical need to develop a tailored health literacy instrument targeting populations with LEP. A precise health literacy measurement tool can provide needs assessment evidence for interventions to reduce health disparities.

Background

Although there is no single accepted definition for health literacy, researchers agree that health literacy is a multidimensional construct (Institute of Medicine, 2004). From a risk factor perspective, health literacy refers to the capability to “obtain, process, and understand basic health information and services needed to make appropriate health decisions” (Institute of Medicine, 2004; Ratzan & Parker, 2000, p. vi; U.S. Department of Health and Human Services, 2000). From an asset perspective, health literacy refers to “the cognitive and social skills which determine the motivation and ability of individuals to gain access to, understand and use information in ways which promote and maintain good health” (Nutbeam, 1998, p. 357). Both definitions demonstrate health literacy is a multidimensional construct goes beyond reading and writing.

Theoretical Framework

Nutbeam’s (2000) health literacy conceptual model has been widely cited in the literature because the model advanced previous understanding of health literacy by including different levels of cognitive, interpersonal, social, and political skills (Chinn, 2011). According to Nutbeam (2000), this conceptual model has three levels: functional, interactive, and critical health literacy. Functional health literacy refers to the capability to understand factual information and use health services. Health-related knowledge, prescription adherence, and health system navigation are examples of functional health literacy skills. Interactive health literacy refers to the capability to act independently in a supportive environment. Examples include communication with others, self-adjustment (e.g. improving motivation, building self-confidence, changing behavior, etc.), and social skills. Critical health literacy refers to the capability to control health-related situations, such as cognitive abilities and skills to act on social, economic, and environmental determinants. Nutbeam (2008) later suggests dividing critical health literacy into three components: the critical analysis of information, an understanding of the social determinants of health, and engagement in collective action.

Chinn (2011) further illustrated the expanded definition of critical health literacy with three constituent domains: information appraisal, understanding the social determinants of health, and collective action. Information appraisal represents the skills of critical analysis of information. The ability to assess the reliability, validity, credibility and applicability of health information are examples of information appraisal skills (Chinn, 2011). Understanding the social determinants of health represents the ability to identify reasons for poor health and health inequalities (Chinn, 2011). Collective action represents the ability to translate the knowledge of social determinants into collective organizing and action (Chinn, 2011).

Health Literacy Assessments

Integrating Nutbeam's (2000, 2008) health literacy conceptual model with Chinn's (2011) expanded critical health literacy concept, Chinn and McCarthy (2013) developed the All Aspects of Health Literacy Scale (AAHLS). The AAHLS is a theory-based health literacy survey with 13 self-reported questions, which takes about 7 minutes to complete (Chinn & McCarthy, 2013). The survey assesses four factors related to individual's health literacy: functional health literacy (3 questions), interactive health literacy (3 questions), information appraisal (4 questions), and empowerment (3 questions).

Populations with LEP

The U.S. Chinese native speakers with LEP are an especially vulnerable group among the LEP populations because these Chinese Americans are more likely to have poor English proficiency and low health literacy. About 60% of the foreign-born population from China in the US are categorized as individuals with LEP who speak English "less than very well" (Gambino, Acosta, & Gireco, 2014). Moreover, Sentell and Braun (2012) conducted a study in California and found 68.3% of the Chinese with LEP had low health literacy; thus, they concluded Chinese had the highest prevalence of low health literacy among the populations with LEP.

Populations with LEP in the US encounter health information both in English and their native languages. Wilson and colleagues (2005) found access to language-concordant healthcare providers (who speak the same language with the LEP patients) could extensively reduce language barriers. Therefore, LEP patients' health care comprehension vary under two language scenarios: when they communicate with language-concordant healthcare providers versus when they communicate with providers who can only speak English (Wilson, Chen, Grumbach, Wang, & Fernandez, 2005). Because limited English proficiency is highly correlated with low health literacy, it is important to assess health literacy among populations with LEP using separate language scenarios.

Study Purpose

This study aims to assess health literacy level among U.S. Chinese native speakers with limited English proficiency using a modified theory-based survey (AAHLS) and investigate the adequacy of the survey instrument to elicit valid data from this unique population. For this purpose, we posed the following questions:

1. What are the health literacy levels for the participants in this study under the English scenario?
2. What are the health literacy levels for the participants in this study under the Chinese scenario?
3. How well does the theoretical health literacy model fit with the data from the English scenario questions?
4. How well does the theoretical health literacy model fit with the data from the Chinese scenario questions?

Methods

Survey Modifications

The original AAHLS survey was developed with English in the U.K. (Chinn & McCarthy, 2013). We modified the original AAHLS to tailor it among U.S. native Chinese speakers with LEP. To separate assessing English language proficiency from health literacy, we created two language scenarios (English and Chinese) for the functional health literacy, interactive health literacy, and information appraisal questions. We kept the original items for the rest of the 3 empowerment items without language scenario separation because they could not be divided according to language situations. Therefore, our modified AAHLS survey contains 23 questions. Table 10 presents the original and the modified example questions.

Table 10. AAHLS Example Items

Variables	**# of Items**	**Original Item Example**	**Modified Item Example**
Functional HL	3	Do you need help to fill in official documents?	Do you need help to fill in English official documents in English? Do you need help to fill in Chinese official documents in Chinese?
Interactive HL	3	When you talk to a doctor or nurse, do you ask the questions you need to ask?	When you talk to a doctor or nurse in English, do you ask the questions you need to ask? When you talk to a doctor or nurse in Chinese, do you ask the questions you need to ask?
Information Appraisal	4	How often do you try to figure out whether information about your health can be trusted?	How often do you try to figure out whether information in English about your health can be trusted? How often do you try to figure out whether information in Chinese about your health can be trusted?
Empowerment	3	Within the last 12 months have you taken action to do something about a health issue that affects your family or community?	

Survey Translation

We translated the original English AAHLS into Chinese (both traditional and simplified) following the cross-cultural translation guidelines (Wild et al., 2005; Beaton et al., 2007). Our translation process contained five steps: 1) two bilingual health professionals who are Chinese native speakers developed two forward translations independently; 2) these two translators, along with a third native-Chinese-speaking health professional who did not participate in the forward translation together compared the two forward translation versions, discussed the discrepancies, and reached consensus; 3) a bilingual native-English speaker who is not familiar with the health content (to avoid information bias) back translated the Chinese version into English; 4) all the translators met to review and compare the back-translated version with the original one to revise the Chinese-AAHLS; 5) received feedback from ten bilingual community health workers/professionals and further revise the translation.

Two versions of the translated survey were available because written Chinese has two versions: traditional and simplified. Before 1949, traditional Chinese was the only written language for Chinese. The Chinese government simplified the characters to promote literacy in 1949. Most Chinese speakers understand both versions but they have a preference between these two. Immigrants from Taiwan, Hong Kong, and Macau generally use the traditional version, but the simplified version is often used among Mainland Chinese immigrants (Zhou & Cai, 2002). Because the US and the United Nations provide both versions when communicating with Chinese speakers, we also furnished both versions for this study's participants.

Cognitive Interview

After the translation, we conducted cognitive interviews among ten Chinese speakers with LEP to identify potential sources of measurement error related to the survey questions. We applied think-aloud and probing techniques during the cognitive interviews (Collins, 2003). Regarding the think-aloud approach, we asked our interviewees to speak out their cognitive process while taking the instruments item by item. Regarding the probing approach, we asked specific questions at the end of each interview. For example, we asked our interviewees: Are these questions easy to understand? Do you notice any confusing words or phrases? I noticed you hesitated before you answered that question—what were you thinking about?

Survey Revision

After the cognitive interview, we further revised the Chinese AAHLS questions to clarify the wording. We also added detail instructions at the beginning of the survey to avoid ambiguity. The original English AAHLS survey has a three-category response scale (rarely, sometimes, and often) for the functional health literacy, interactive health literacy, information appraisal questions, and the first question of the empowerment factor. We changed the three-category response scale to a five-category response scale (never, rarely, sometimes, often, and always). Such response scale change was based on the feedback from our cognitive interviewees. We coded these questions' response as 0, 1, 2, 3, and 4 with 0 indicating the lowest health literacy level while 4 indicating the highest health literacy level. We kept the binary response scale of the last two empowerment questions. We coded these two binary items as 0 and 1. See Table 1 for example questions. The Cronbach's α was 0.80, showing a good internal consistency for using the modified Chinese AAHLS survey among our study participants.

Data Collection Procedures

This study was approved by the Institutional Review Board (IRB2016-0092D) of the authors' university affiliation. We distributed study flyers through health professional organizations, community centers, and churches in an urban region in the southwestern US to recruit potential participants. Interested individuals were encouraged to contact the researchers and talk to their community staff/leaders. Some communities invited us into their community centers or churches to recruit and collect data. An electronic version of the survey was also available upon request.

To be eligible for this study, the participants must be: 18 years or older, native speakers of Chinese (either Mandarin or Cantonese), literate in Chinese (either simplified or traditional Chinese), and self-report as speaking English "less than very well". We assessed the eligibility of our potential participants using four pre-screening questions asking about their age, native language, Chinese literacy, and English-speaking proficiency (see detail in Chen et al., 20XX cite dissertation article 2). When participants returned the survey, the research staff scanned through their answers to make sure there were no accidental missing items. Each participant received a $10 grocery gift card as incentive after completing the survey. Data were collected from June to September 2016.

Participants

A total of 405 participants (158 male and 247 female) completed the study. Their ages ranged from 18 to 96 (M = 51.70, SD = 19.31). Participants' time in the US ranged from 1 month to 74 years (M = 18 years, SD = 14). In our sample, 30.12% considered themselves to speak English "well", 38.27% "fair", 22.47% "not well", and 9.14% "not at all". More than half of the participants (55.31%) had bachelor degrees and above; 20.99% had technique or associate degrees; and 23.70% had high school level and below. About 42.47% of them were not employed (either unemployed, retired, or stay-home mom/dad), 34.07% were employed (either full- time or part-time), 19.75% were students, and 2.47% were business owners. We provided our participants' socio-demographic information in Table 11.

Table 11. Participants' Socio-demographic Information

Items	N	%
English Speaking Proficiency		
Well	122	30.12%
Fair	155	38.27%
Not well	91	22.47%
Not at all	37	9.14%
Gender		
Male	158	39.01%
Female	247	60.99%
Education		
High school and below	96	23.70%
Technique or Associate	85	20.99%
Bachelor	67	16.54%
Master	110	27.16%
PhD and above	47	11.60%
Occupation		
Student	80	19.75%
Full time/Part time employee	138	34.07%
Business owner	10	2.47%
Unemployed	32	7.90%
Retired	118	29.14%
Stay-home mom/dad	22	5.43%
Other	5	1.23%
Total	405	100%

Data Analysis

We used descriptive analysis and paired t-test to examine the distributions and differences between participants' health literacy levels under the English and Chinese scenarios. Confirmatory factor analysis (CFA) assesses whether items correlate in a way that is consistent with the hypothesized theoretical structure (Long, 1983). We performed CFA using Mplus 7 to examine the construct validity of the Chinese AAHLS by evaluating the 4-factor health literacy model fit (Chinn & McCarthy, 2013). Since our data did not exhibit normal distribution, we chose the diagonally weighted least squares estimation (WLSMV). WLSMV does not assume normal distribution and provides the best option for modelling categorical and ordinal data (Brown, 2006). Compared to the maximum likelihood estimation (ML) and the robust ML estimation (MLR), WLSMV is less bias and more accurate in estimating factor loadings for categorical and ordinal data (Li, 2016).

Model Fit Criteria

The Chi-Square value is a traditional measure for evaluating overall model fit (Barrett, 2007). When the Chi-Square is large enough to generate a significant *p* value, it indicates the data does not have a perfect fit with the model (Kline, 2010). One limitation of Chi-Square is that its statistical significance test is sensitive to sample size; therefore, the Chi-Square criteria nearly always rejects the model when the sample size are large (Bentler & Bonnet, 1980). Due to such limitation, Chi-Square measure is not the sole index to assess model fit (Barrett, 2007). Other model fit indices include the root mean square error of approximation (RMSEA), comparative fir index (CFI), Tucker-Lewis index (TLI), and weighted root mean square residual (WRMR) (Barrett, 2007; Muthén, 2004). The model is considered as "good fit" when RMSEA < 0.06, CFI > 0.95, TIL > 0.95, and WRMR < 1.0 (Hu & Bentler, 1999; Yu, 2002). We used these indices to evaluate and compare the model fit degree between the English scenario questions and the Chinese scenario questions.

Results

Table 12 presents information about the response distributions of each question. There were no missing data because data collection protocol was set up to prevent collection of incomplete questionnaires. Compared to the English language scenario, besides the item assessing if participants would question healthcare providers (item Info4), participants had statistically higher health literacy level when they were immersed in communications using Chinese ($p < 0.01$). Interestingly, participants had similar response patterns for English and Chinese scenarios ($p = 0.515$), on the questions indicating they were not likely to question their healthcare providers' advice based on their own research (items Info4a and Info4b). Few participants chose "often" or "always" for these two items (21.7% for the English scenario question and 16.3% for the Chinese scenario question). Further, many participants had limited empowerment capabilities at the level of community and social engagement. About 62.7% chose "never" or "rarely" for the question asking whether they believed they had the right to influence the U.S. government's action on health issues (item Emp1). Also, only 30.6% of the participants had ever take actions on health issues that affect their family or community within the last 12 months (item Emp2). Most participants (75.6%) prioritized individual lifestyle choices and behaviors rather than social infrastructure as factors influencing health (item Emp3).

Table 12. Response Distributions of the Modified Chinese AAHLS Questions

Items	**Mean (SD)**	**Never**	**Rarely**	**Sometimes**	**Often**	**Always**	***T-test P***
Functional Health Literacy							
F1a. How often do you need help when you are given information in English to read by your doctor, nurse or pharmacist?	1.78 (1.17)	6.9%	20.5%	35.1%	18.8%	18.8%	<0.01
F1b. How often do you need help when you are given information in Chinese to read by your doctor, nurse or pharmacist?	2.80 (1.15)	33.3%	32.3%	21.7%	6.7%	5.9%	
F2a. When you need help to read the given information in English, can you easily get hold of someone to assist you?	2.40 (1.05)	1.2%	19.5%	37.5%	22.0%	19.7%	<0.01
F2b. When you need help to read the given information in Chinese, can you easily get hold of someone to assist you?	2.77 (1.22)	2.0%	18.0%	22.7%	15.3%	41.9%	
F3a. Do you need help to fill in English official documents in English?	1.87 (1.15)	6.4%	24.0%	36.5%	16.3%	16.8%	<0.01
F3b. Do you need help to fill in Chinese official documents in Chinese?	2.85 (1.15)	34.8%	33.3%	20.2%	4.9%	6.7%	

Table 12 Continued

Items	Mean (SD)	Never	Rarely	Sometimes	Often	Always	*T-test P*
Interactive Health Literacy							
I1a. When you talk to a doctor or nurse in English, do you give them all the information they need to help you?	2.73 (1.11)	4.7%	8.6%	24.2%	33.8%	28.6%	<0.01
I1b. When you talk to a doctor or nurse in Chinese, do you give them all the information they need to help you?	2.97 (1.19)	6.4%	5.9%	15.6%	28.6%	43.5%	
I2a. When you talk to a doctor or nurse in English, do you ask the questions you need to ask?	2.66 (1.11)	5.2%	9.4%	25.7%	34.1%	25.7%	<0.01
I2b. When you talk to a doctor or nurse in Chinese, do you ask the questions you need to ask?	2.86 (1.03)	3.2%	5.9%	24.0%	35.1%	31.9%	
I3a. When you talk to a doctor or nurse in English, do you make sure they explain anything that you do not understand?	2.61 (1.12)	5.7%	10.1%	25.4%	34.8%	24.0%	<0.01
I3b. When you talk to a doctor or nurse in Chinese, do you make sure they explain anything that you do not understand?	3.00 (1.01)	2.2%	5.9%	19.5%	34.1%	38.3%	
Information Appraisal							
Info1a. Are you someone who like to find out lots of different information in English about health?	1.98 (1.18)	11.1%	25.9%	27.4%	24.9%	10.6%	<0.01
Info1b. Are you someone who like to find out lots of different information in Chinese about health?	2.61 (1.02)	2.5%	11.6%	29.4%	35.1%	21.5%	

Table 12 Continued

Items	Mean (SD)	Never	Rarely	Sometimes	Often	Always	*T-test* ***P***
Information Appraisal *continued*							
Info2a. How often do you think carefully about whether health information in English makes sense in your or your family's particular situation?	2.27 (1.12)	7.4%	16.3%	32.1%	29.9%	14.3%	<0.01
Info2b. How often do you think carefully about whether health information in Chinese makes sense in your or your family's particular situation?	2.71 (1.01)	2.2%	9.4%	27.7%	36.5%	24.2%	
Info3a. How often do you try to work out whether information in English about your or your family's health can be trusted?	2.30 (1.20)	8.9%	17.0%	26.7%	30.1%	17.3%	<0.01
Info3b. How often do you try to work out whether information in Chinese about your or your family's health can be trusted?	2.76 (0.99)	1.7%	8.6%	27.9%	35.6%	26.2%	
Info4a. Under the English scenario, are you the sort of person who might question your doctor or nurse's advice based on your own research?	1.83 (0.99)	7.9%	29.1%	41.2%	15.8%	5.9%	.515
Info4b. Under the Chinese scenario, are you the sort of person who might question your doctor or nurse's advice based on your own research?	1.79 (0.93)	6.2%	31.1%	46.4%	9.9%	6.4%	

Table 12 Continued

Items	Mean (SD)	Never	Rarely	Sometimes	Often	Always	*T-test P*
Empowerment							
Emp1. Do you think that there plenty of ways to have a say in what the U.S. government does about health?	1.32 (1.13)	26.4%	36.3%	20.7%	11.6%	4.9%	NA (not divided by language scenarios)
Emp2. Within the last 12 months have you taken action to do something about a health issue that affects your family or community?		☐Yes 30.6%			☐No 69.4%		
Emp3. What do you think matters most for everyone's health? (tick one answer only)	☐information and encouragement to lead healthy lifestyles 75.6%			☐good housing, education, decent jobs and good local facilities 24.4%			

CFA Model Fit

As shown in Table 13, the 4-factor model (functional health literacy, interactive health literacy, information appraisal, and empowerment) of the English scenario questions had poor fit (Chi-Square = 315.932, $p < 0.01$, RMSEA = 0.104, CFI = 0.958, TLI = 0.944, WRMR = 1.473). The 4-factor model of the Chinese scenario questions had adequate fit (Chi-Square = 177.069, $p < 0.01$, RMSEA = 0.070, CFI = 0.971, TLI = 0.962, WRMR = 1.110). Comparing to the English scenario, the Chinese scenario model had a better fit.

Modified CFA Model

Based on the model modification indices information from the analysis output, we added the correlation path between the reading (F1) and writing (F3) items of the functional health literacy to improve the model fit for both English and Chinese scenario questions. Both modified model fit indices improved compared to the original versions (Table 4). The modified English model exhibited adequate fit (Chi-Square = 196.851, $p < 0.01$, RMSEA = 0.077, CFI = 0.977, TLI = 0.969, WRMR = 1.098). The modified Chinese model had good fit (Chi-Square = 148.902, $p < 0.01$, RMSEA = 0.062, CFI = 0.978, TLI = 0.970, WRMR = 1.001).

Table 13. CFA Model Fit Indices

Model Fit Indices	**English Scenario Model**		**Chinese Scenario Model**	
	4-factor	*4-factor modified*	*4-factor*	*4-factor modified*
Chi-Square	315.932	196.851	177.069	148.902
p	< 0.01	< 0.01	< 0.01	< 0.01
RMSEA	0.104	0.077	0.070	0.062
CFI	0.958	0.977	0.971	0.978
TLI	0.944	0.969	0.962	0.970
WRMR	1.473	1.098	1.110	1.001

Figure 7 presents the visual depiction of the English and Chinese scenario models with standardized regression coefficients for all the paths. Other than paths related to the empowerment factor, all others under the both language scenarios had statistically significant coefficients ($p < 0.01$). In other words, the higher score of each functional, interactive, and information appraisal item indicated higher level of the corresponding ability. The factors of functional health literacy, interactive health literacy, and information appraisal were significantly associated with each other (p <

0.01). The relationship between reading (F1) and writing (F3) were also significant ($p <$ 0.01).

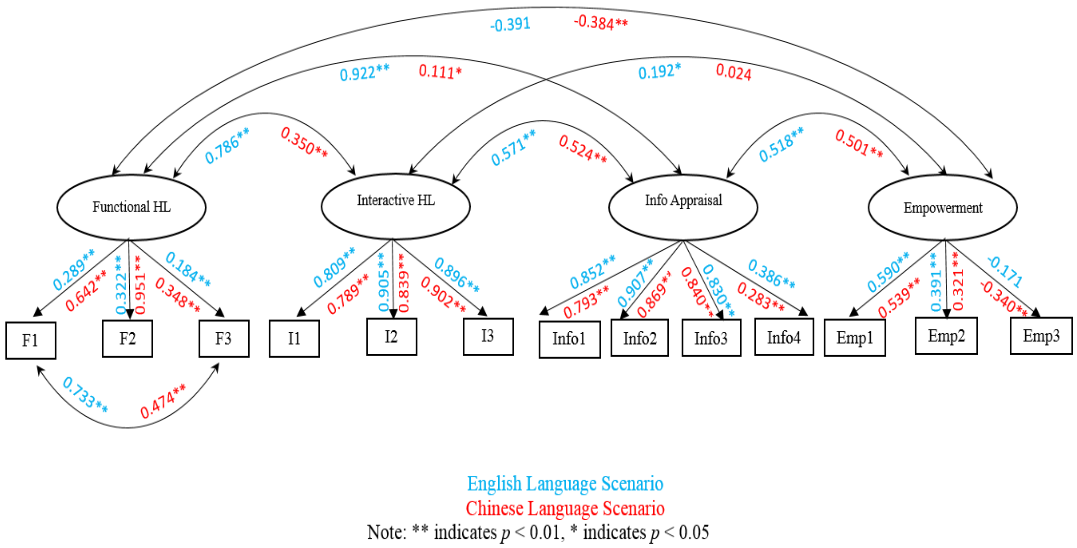

Figure 7. CFA Measurement Models with Standardized Regression Coefficients

English Scenario

Functional Health Literacy

We presented the R^2 variance of each item in Table 14. Under the English scenario, about 8.3% of the variance ($p < 0.05$) in reading (item F1a), 10.4% of the variance ($p < 0.05$) in seeking for help (item F2a), and 3.4% of the variance ($p = 0.097$) in writing (item F3a) was explained by the functional health literacy factor.

Table 14. R-Square Variance of the Modified 4-factor Models

Factors	Items	English Scenario		Chinese Scenario	
		R^2	*p*	R^2	*p*
Functional Health	F1	0.083	0.028	0.412	0.000
Literacy	F2	0.104	0.029	0.904	0.000
	F3	0.034	0.097	0.121	0.001
Interactive Health	I1	0.655	0.000	0.622	0.000
Literacy	I2	0.819	0.000	0.704	0.000
	I3	0.803	0.000	0.813	0.000
Information Appraisal	Info1	0.725	0.000	0.629	0.000
	Info2	0.822	0.000	0.755	0.000
	Info3	0.689	0.000	0.706	0.000
	Info4	0.149	0.000	0.080	0.001
Empowerment	Emp1	0.348	0.012	0.291	0.005
	Emp2	0.153	0.025	0.103	0.079
	Emp3	0.029	0.393	0.115	0.069

Interactive Health Literacy

About 65.5% of the variance ($p < 0.01$) in the question assessing how often people give information to their healthcare providers (item I1a), 81.9% of the variance ($p < 0.01$) in the question assessing how often people ask questions to their healthcare providers (item I2a), and 80.3% of the variance ($p < 0.01$) in the question assessing how often people make sure their healthcare providers explain anything that they do not understand (item I3a) was explained by the interactive health literacy factor.

Information Appraisal

About 72.5% of the variance ($p < 0.01$) in the question asking whether individuals like to find health-related information (item Info1a), 82.2% of the variance($p < 0.01$) in the question assessing how often they critically evaluated the relevance of the health information to their health situations (item Info2a), 68.9% of the variance ($p <$

0.01) in the question asking how often they critically evaluated the reliability of the health information (item Info3a), and 14.9% of the variance ($p < 0.01$) in the question asking how often they question their healthcare providers' advice based on their own research (item Info4a) was explained by the information appraisal factor.

Empowerment

About 34.8% of the variance ($p < 0.05$) in the question asking whether they believe they have the right in the U.S. health policies (item Emp1), 15.3% of the variance ($p < 0.05$) in the question assessing whether they have acted upon a health issue within the last 12 months (item Emp2), and 2.9% of the variance ($p = 0.393$) in the question asking which aspect (individual vs. social) they believe have more impact on health (item Emp3) was explained by the empowerment factor.

Chinese Scenario

Functional Health Literacy

Under the Chinese scenario, about 41.2% of the variance ($p < 0.01$) in reading (item F1b), 90.4% of the variance ($p < 0.01$) in seeking for help (item F2b), and 12.1% of the variance ($p < 0.01$) in writing (item F3b) was explained by the functional health literacy factor.

Interactive Health Literacy

About 62.2% of the variance ($p < 0.01$) in the question assessing how often people give information to their healthcare providers (item I1b), 70.4% of the variance ($p < 0.01$) in the question assessing how often people ask questions to their healthcare providers (item I2b), and 81.3% of the variance ($p < 0.01$) in the question assessing how often people make sure their healthcare providers explain anything that they do not understand (item I3b) was explained by the interactive health literacy factor.

Information Appraisal

About 62.9% of the variance ($p < 0.01$) in the question asking whether individuals like to find health-related information (item Info1b), 75.5% of the variance ($p < 0.01$) in the question assessing how often they critically evaluated the relevance of the health information to their health situations (item Info2b), 70.6% of variance ($p < 0.01$) in the question asking how often they critically evaluated the reliability of the health information (item Info3b), and 8% of the variance ($p < 0.01$) in the question asking how often they question their healthcare providers' advice based on their own research (item Info4b) was explained by the information appraisal factor.

Empowerment

About 29.1% of the variance ($p < 0.01$) in the question asking whether they believe they have the right in the U.S. health policies (item Emp1), 10.3% of the variance ($p = 0.079$) in the question assessing whether they have acted upon a health issue within the last 12 months (item Emp2), and 11.5% of the variance ($p = 0.069$) in the question asking which aspect (individual vs. social) they believe have more impact on health (item Emp3) was explained by the empowerment factor.

Discussion

This study used a modified theory-based survey (AAHLS) to assess health literacy level among U.S. native Chinese speakers who have limited English proficiency. Generally, participants earned higher health literacy scores when they encountered health information/situations in Chinese rather than in English. We found participants were unlikely to question their healthcare providers' advice, irrespective of language scenarios. This is consistent with results from Wang and colleagues (2012), who also found Chinese immigrants were less likely to challenge physicians or express their needs to physicians compared to US-born Chinese and non-Hispanic Whites. This might be given that physicians are highly respected in Chinese culture because they represent the authority of medical knowledge; further, Chinese culture does not encourage people to question authority (Wang, Liang, Schwartz, Lee, Kreling, & Mandelblatt, 2008).

We also found few participants believed they had a voice in influencing or reforming the U.S. health policy. Most had not even engaged with the U.S. health care system within the last 12 months. Similarly, other research shows Chinese immigrants in the US do not interact with the American political system, and Chinese Americans are less likely to vote compared to other Asian American peers (Arts, 2015; Wray-Lake & Tang, 2017). This belief might be related to the Chinese political system and cultural values leading to political disengagement among Chinese American immigrants. For example, the Chinese government operates behind closed doors and does not encourage Chinese people to question authority (Pye, 1993), which creates distance between the people and the government (Arts, 2015). Also, education in China is centered on

Confucianism, which further emphasizes respect for authority (Chen & Lu, 2011). Such educational and cultural values tend to separate people actively engaging in activities related to policy influence and reform.

Most participants in our study believed individual lifestyle choices and behaviors had a greater impact on health compared to social infrastructure. This finding aligns with previous studies that many people believe individual behaviors have more effects on health compared to structural and material factors (Robert & Booske, 2011; Davidson, Kitzinger, & Hunt, 2006). However, people's ability to engage in healthy behaviors are affected by the social, economic, environmental, and political conditions (Robert & Booske, 2011). Public health professionals should incorporate health interventions and policy approaches to improve such narrow understanding of the social determinants of health (Collins, Abelson, & Eyles, 2007).

Our study also tested the appropriateness of the modified AAHLS survey to elicit valid data. The CFA results showed that the Chinese scenario questions had better model fit than the English scenario questions. Further, compared to the Chinese scenario model, the English scenario model had lower R^2 variances of the functional health literacy items. Such low variances indicate participants' functional health literacy skill is overlapping with English language proficiency when they encountering health information in English. Therefore, measuring functional health literacy without differentiating English language from participants' native languages fails to yield precise assessment. Chen and colleagues (cite dissertation article 2) pointed out an English instrument was not able to assess functional health literacy accurately among populations with LEP because such instrument was assessing their English proficiency at the same time.

Future Study

The R^2 variances of the item about questioning clinicians (Info4) and the empowerment questions were low under both English and Chinese scenarios. In other words, these questions only assessed a small portion of the information appraisal and empowerment factors. We believe these questions are culturally sensitive to minorities, especially among Chinese Americans with LEP; therefore, these questions did not effectively measure participants' information appraisal skill and empowerment capability. Future research should develop more precise questions to measure critical health literacy among populations with LEP.

Our results also showed Chinese Americans with LEP had low levels of empowerment and were unlikely to engage in collective actions. Thus, another direction for future study is to investigate their perceptions of community empowerment and health-promoting activism. This is a preliminary step for developing effective interventions because empowerment is an essential concept of critical health literacy.

Further, for future health literacy interventions among populations with LEP, health professionals, literacy researchers, and ESL practitioners should work collaboratively to combine the English language instruction with the health education components into one program (Chen, Goodson, & Acosta, 2015). As McKee and Paasche-Orlow (2012, p. 7) pointed out:

> It is critical for health literacy and limited English proficiency researchers to work together to understand how culture, language, literacy, education, and disabilities influence health disparities and health outcomes. It is important to ensure that research is collaborative and inclusive in order to broaden the reach of future interventions to smaller linguistic minority populations.

Limitations

Besides the above contributions, this study has limitations. We only targeted U.S. native Chinese speakers with limited English proficiency using convenience sampling approach. Thus, the sample might not be representative for the Chinese American community. Also, we used a self-report pre-screening question to identify individuals with LEP. This pre-screen question ask potential participants to rate their English-speaking proficiency. Such self-assessment might contain some bias. Last, the self-report survey might also have some health literacy measurement bias compared to objective tests.

Contributions

We believe this study contributes to the literature on assessing health literacy among populations with LEP in three ways. First, our study targets vulnerable populations with LEP (e.g., immigrants and linguistic minorities) who live in an English dominant society. Second, our study develops a tailored theory-based health literacy assessment tool for native Chinese speakers with LEP. Third, this study evaluates how well the health literacy conceptual model fit with the survey data, which provides evidence for future health literacy interventions among populations with LEP.

CONCLUSION

A large number of health literacy measurement tools exist in the current literature, but few of them were developed based on health literacy theory (Haun et al., 2014). Further, few studies designed and implemented a tailored health literacy measure for minority populations, especially for populations with limited English proficiency (Nguyen, et al., 2015). However, it is very important to assess health literacy level among populations with limited English proficiency. Assessing health literacy is the first step in starting a health education program (Thomason & Mayo, 2015). A precise tool for assessing health literacy allows clinicians to tailor health services for patients. The tools also help health professionals to shape patient-provider communication and reduce health disparities (Nørgaard, Sørensen, Maindal, & Kayser, 2014; Stonbraker, Schnall, & Larson, 2015; Batterham, Hawkins, Collins, Buchbinder, & Osborne, 2016). Therefore, the overarching research question of my dissertation study is: Does the current health literacy theory and measurement fit for individuals with limited English proficiency?

My dissertation study aims to improve the current health literacy theory and measurement tools as applied among populations with limited English proficiency (LEP). The dissertation has two specific aims: (1) to clarify the translation process of the non-English assessment tools and its impact on score reliability of these scores and (2) to investigate the psychometric properties of health literacy measures applied to the U.S. Chinese speaking individuals who have limited English proficiency.

To achieve these specific aims, my dissertation uses the manuscript format with three papers. Manuscript #1 is a systematic literature review and meta-analysis study to examine the translation process of the non-English Test of Functional Health Literacy in Adults (TOFHLA) and its impact on score reliability of these tests. Manuscript #2 is a quantitative study to investigate the psychometric properties of the English and Chinese versions of the TOFHLA. Manuscript #3 is a quantitative study to evaluate the adequacy of a modified All Aspects of Health Literacy Scale (AAHLS) survey to elicit valid data among the U.S. Chinese speaking individuals with limited English proficiency.

Manuscript #1 identified the non-English TOFHLA instruments in the literature, synthesized their translation processes, and examined the impact of test and sample characteristics on the Cronbach's α coefficients of these instruments. The translation process varied from study to study. For future TOFHLA implication, I recommend researchers to use the short form rather than the long form. I also call for further studies to improve the language translation quality of non-English health literacy measures.

Manuscript #2 using Item Response Theory to examine and compare the item difficulty and discrimination parameters of two functional health literacy instruments among populations with LEP. I found the current functional health literacy construct and measurement tools are not applicable among populations with LEP. Most individuals with LEP had adequate functional health literacy when assessed in their native languages. The English S-TOFHLA was measuring functional health literacy along with language proficiency. The Numeracy items of the Chinese S-TOFHLA yielded scores with low reliability.

Manuscript #3 developed a tailored theory-based health literacy assessment tool for native Chinese speakers with LEP and evaluated how well the health literacy conceptual model fit with the survey data, which provides evidence for future health literacy interventions among populations with LEP. I found participants earned higher health literacy scores when they encountered health information/situations in Chinese rather than in English. I also found few participants believed they had a voice in influencing or reforming the U.S. health policy. Further, the theoretical health literacy model had a better fit with the data from the Chinese scenario questions than the data from the English scenario questions.

In sum, my dissertation study found differentiating English proficiency from functional health literacy and adding components of language proficiency and cultural difference could improve the health literacy measurement and theory among populations with LEP. This study confirms the current health literacy instruments and theory were incapable of capturing the complex interaction among populations with LEP (Yip, 2012). Culturally related constructs such as English language proficiency are key components that should be added to the health literacy measurement and theoretical model for populations with LEP.

Future research studies should develop tailored health literacy measurement and interventions targeting linguistic minority populations. Researchers should consider the problems of language and health care system differences when measuring health literacy among populations with LEP. There is a critical need to include both the components of

English language instructions and health literacy education in future health literacy interventions among populations with LEP.

My dissertation study also provides implications for future health education practice. Public health professionals should incorporate health interventions and policy approaches to improve critical health literacy among populations with LEP. Also, health professionals, literacy researchers, and ESL practitioners should work collaboratively to combine the English language instruction with the health education components into one program (Chen, Goodson, & Acosta, 2015).

My dissertation study has some limitations. First, I applied convenience sampling method for the data collection. Second, I only targeted U.S. native Chinese speakers with limited English proficiency. Thus, the sample might not be representative for the general Chinese American community.

REFERENCES

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

Al-Jumaili, A. A., Al-Rekabi, M. D., & Sorofman, B. (2015). Evaluation of instruments to assess health literacy in Arabic language among Iraqis. *Research in Social and Administrative Pharmacy*, *11*(6), 803-813.

Altin, S. V., Finke, I., Kautz-Freimuth, S., & Stock, S. (2014). The evolution of health literacy assessment tools: a systematic review. *BMC Public Health*, *14*(1), 1207.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Andreae, M. H., White, R. S., Chen, K. Y., Nair, S., Hall, C., & Shaparin, N. (2016). The effect of initiatives to overcome language barriers and improve attendance: A cross-sectional analysis of adherence in an inner city chronic pain clinic. *Pain Medicine, 18*(2), 265-274.

Arts, H. B. O. (2015). Asian American political participation: An analysis of the influences on immigrant political efficacy in the United States. Doctoral dissertation, The University of Michigan.

August, D., Shanahan, T., & Escamilla, K. (2009). English language learners: Developing literacy in second-language learners—Report of the National Literacy Panel on Language-Minority Children and Youth. *Journal of Literacy Research, 41*(4), 432-452.

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.

Barac, R., & Bialystok, E. (2012). Bilingual effects on cognitive and linguistic development: Role of language, cultural background, and education. *Child Development*, *83*(2), 413-422.

Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences, 42*(5), 815-824.

Batterham, R. W., Hawkins, M., Collins, P. A., Buchbinder, R., & Osborne, R. H. (2016). Health literacy: Applying current concepts to improve health services and reduce health inequalities. *Public Health, 132*, 3-12.

Beaton, D., Bombardier, C., Guillemin, F., & Ferraz, M. B. (2002). Recommendations for the cross-cultural adaptation of health status measures. *American Academy of Orthopaedic Surgeons*, 1-9.

Bentler, P.M., & Bonnet, D.C. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606.

Brown, T. (2006). Confirmatory factor analysis for applied research. New York: Guildford.

Bruce, K. H., Schwei, R. J., Park, L. S., & Jacobs, E. A. (2014). Barriers and facilitators to preventive cancer screening in Limited English Proficient (LEP) patients: Physicians' perspectives. *Communication & Medicine*, *11*(3), 235-247.

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

Chen, J., & Lu, C. (2011). Democratization and the middle class in China: The middle class's attitudes toward democracy. *Political Research Quarterly, 64*(3), 705-719.

Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265-289.

Chen, X., Goodson, P., & Acosta, S. (2015). Blending health literacy with an English as a second language curriculum: A systematic literature review. *Journal of Health Communication, 20*(sup2), 101-111.

Chinn, D. (2011). Critical health literacy: A review and critical analysis. *Social Science & Medicine, 73*(1), 60-67.

Chinn, D., & McCarthy, C. (2013). All Aspects of Health Literacy Scale (AAHLS): Developing a tool to measure functional, communicative and critical health literacy in primary healthcare settings. *Patient Education and Counseling, 90*(2), 247-253.

Collins, D. (2003). Pretesting survey instruments: An overview of cognitive methods. *Quality of Life Research, 12*(3), 229-238.

Collins, P. A., Abelson, J., & Eyles, J. D. (2007). Knowledge into action? Understanding ideological barriers to addressing health inequalities at the local level. *Health Policy, 80*(1), 158-171.

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.

Cousin, S. L., & Henson, R. K. (2000). What is reliability generalization, and why is it important? Retrieved from ERIC database. (ED44507).

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334.

Cummins, J. (2009). Fundamental psycholinguistic and sociological principles underlying educational success for linguistic minority students. *Social Justice through Multilingual Education*, 19-35.

Davidson, R., Kitzinger, J., & Hunt, K. (2006). The wealthy get healthy, the poor get poorly? Lay perceptions of health inequalities. *Social Science & Medicine, 62*(9), 2171-2182.

Davis, T. C., Long, S. W., Jackson, R. H., Mayeaux, E., 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.

Dowse, R., Lecoko, L., & Ehlers, M. S. (2010). Applicability of the REALM health literacy test to an English second-language South African population. *Pharmacy World & Science, 32*(4), 464-471.

Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation, and refinement. *Quality of Life Research, 16*(1), 5-18.

Embretson, S. E. & Reise, S. P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates.

Eyüboğlu, E., & Schulz, P. J. (2016). Validation of Turkish health literacy measures. *Health Promotion International, 31*(2), 355-362.

Fan, X., & Thompson, B. (2001). Confidence intervals for effect sizes confidence intervals about score reliability coefficients, please: An EPM guidelines editorial. *Educational and Psychological Measurement, 61*(4), 517-531.

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.

Gambino, C. P., Acosta, Y. D., & Grieco, E. M. (2014). English-speaking ability of the foreign-born population in the United States: 2012. *Population, 10*, 11.

Garrard, J. (2016). *Health sciences literature review made easy: The matrix method.* Burlington, MA: Jones & Bartlett.

Goodson, P. (2009) *Theory in health promotion research and practice: Thinking outside the box.* Burlington, MA: Jones & Bartlett.

Graham, J. M., & Unterschute, M. S. (2015). A reliability generalization meta-analysis of self-report measures of adult attachment. *Journal of Personality Assessment, 97*(1), 31-41.

Hæsum, E., Korsbakke, L., Ehlers, L., & Hejlesen, O. K. (2015). Validation of the test of functional health literacy in adults in a Danish population. *Scandinavian Journal of Caring Sciences*, *29*(3), 573-581.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage.

Han, H.-R., Kim, J., Kim, M. T., & Kim, K. B. (2011). Measuring health literacy among immigrants with a phonetic primary language: A case of Korean American women. *Journal of Immigrant and Minority Health, 13*(2), 253-259.

Haun, J.N., Valerio, M. A., McCormack, L. A., Sørensen, K., & Paasche-Orlow, M. K. (2014). Health Literacy Measurement: An Inventory and Descriptive Summary of 51 Instruments. *Journal of Health Communication, 19*, 302-333.

Hedges, L. V. (1982). Fitting categorical models to effect sizes from a series of experiments. *Journal of Educational Statistics, 7*, 119–137.

Hedges, L. V., & Olkin, I. (1985). *Statistical Method for Meta-Analysis.* Orlando, FL: Academic Press.

Henson, R. K., & Thompson, B. (2002). Characterizing measurement error in scores across studies: Some recommendations for conducting" reliability generalization" studies. *Measurement and Evaluation in Counseling and Development, 35*(2), 113.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1-55.

Institute of Medicine. (2004). *Health literacy: A prescription to end confusion.* Washington, DC: National Academies Press.

Jovic-Vranes, A., Bjegovic-Mikanovic, V., & Marinkovic, J. (2009). Functional health literacy among primary health-care patients: Data from the Belgrade pilot study. *Journal of Public Health, 31*(4), 490-495.

Kamberi, H., Hysa, B., Toçi, E., Jerliu, N., & Burazeri, G. (2012). Functional health literacy in primary health care users in Kosovo: A validation study. *Albanian Medical Journal, 4*, 21-25.

Kline, R. (2010). *Principles and practice of structural equation modeling (3rd Ed.).* New York: Guilford.

Kutner, M., Greenburg, E., Jin, Y., & Paulsen, C. (2006). The Health Literacy of America's Adults: Results from the 2003 National Assessment of Adult Literacy. NCES 2006-483. *National Center for Education Statistics.*

Lane, G. G., White, A. E., & Henson, R. K. (2002). Expanding reliability generalization methods with KR-21 estimates an RG study of the Coopersmith Self-Esteem Inventory. *Educational and Psychological Measurement*, *62*(4), 685-711.

Lee, J. S., Pérez-Stable, E. J., Gregorich, S. E., Crawford, M. H., Green, A., Livaudais-Toman, J., & Karliner, L. S. (2017). Increased access to professional interpreters in the hospital improves informed consent for patients with limited English proficiency. *Journal of General Internal Medicine*, 1-8.

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936-949.

Lindholm, M., Hargraves, J. L., Ferguson, W. J., & Reed, G. (2012). Professional language interpretation and inpatient length of stay and readmission rates. *Journal of General Internal Medicine, 27*(10), 1294-1299.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis.* Thousand Oaks, CA: Sage publications.

Mancuso, J. M. (2009). Assessment and measurement of health literacy: An integrative review of the literature. *Nursing Health Science*, *11*(1), 77-89.

Mantwill, S., & Schulz, P. J. (2016). Health literacy in mainland China: Validation of a functional health literacy test in simplified Chinese. *Health Promotion International, 31*(4), 742-748.

McCormack, L., Haun, J., Sørensen, K., & Valerio, M. (2013). Recommendations for advancing health literacy measurement. *Journal of Health Communication, 18*(sup1), 9-14.

McKee, M. M., & Paasche-Orlow, M. K. (2012). Health literacy and the disenfranchised: The importance of collaboration between limited English proficiency and health literacy researchers. *Journal of Health Communication, 17*(sup3), 7-12.

Meier, S. T., & Davis, S. R. (1990). Trends in reporting psychometric properties of scales used in counseling psychology research. *Journal of Counseling Psychology*, *37*(1), 113.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery*, *8*(5), 336-341.

Muthén, B. O. (2004). *Mplus technical appendices.* Los Angeles, CA: Muthén & Muthén.

Nguyen, T. H., Park, H., Han, H.-R., Chan, K. S., Paasche-Orlow, M. K., Haun, J., & Kim, M. T. (2015). State of the science of health literacy measures: Validity implications for minority populations. *Patient Education and Counseling*, *98*(12), 1492-1512.

Nørgaard, O., Sørensen, K., Maindal, H. T., & Kayser, L. (2014). Measuring health literacy can improve communication in health care. *Ugeskrift for Laeger*, *176*(1), 37-39.

Nutbeam, D. (1998). Health promotion glossary. *Health Promotion International, 13*(4), 349-364.

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.

Nutbeam, D. (2008). The evolving concept of health literacy. *Social Science & Medicine*, *67*(12), 2072-2078.

Ownby, R. L., Acevedo, A., Waldrop-Valverde, D., Jacobs, R. J., Homs, A. M., & Czaja, S. J. (2013). Development and initial validation of a computer-administered health literacy assessment in Spanish and English: FLIGHT/VIDAS. *Patient Related Outcome Measures, 4*, 1-15.

Pandya, C., Batalova, J., & McHugh, M. (2011). *Limited English proficient individuals in the United States: Number, share, growth, and linguistics diversity*. Washington, DC: Migration Policy Institute.

Parker, R. M., Baker, D. W., Williams, M. V., & Nurss, J. R. (1995). The Test of Functional Health Literacy in Adults: A new instrument for measuring patients' literacy skills. *Journal of General Internal Medicine*, *10*, 537–541.

Pleasant, A., McKinney, J., & Rikard, R. V. (2011). Health literacy measurement: a proposed research agenda. *Journal of Health Communication, 16*(sup3), 11-21.

Ratzan, S. C., Parker, R. M. (2000). Introduction. In C. R. Selden, M. Zorn, S. C. Ratzan, & R. M. Parker (Eds.), *National Library of Medicine current bibliographies in medicine: Health literacy* (pp. v–vii). Bethesda, MD: National Institutes of Health.

Reisi, M., Javadzade, S. H., Mostafavi, F., Sharifirad, G., Radjati, F., & Hasanzade, A. (2012). Relationship between health literacy, health status, and healthy behaviors among older adults in Isfahan, Iran. *Journal of Education and Health Promotion, 1*(1), 31.

Reisi, M., Mostafavi, F., Javadzede, H., Mahaki, B., Sharifirad, G., & Tavassoli, E. (2016). The functional, communicative, and critical health literacy (FCCHL) scales:

Cross-Cultural adaptation and the psychometric properties of the Iranian version. *Iranian Red Crescent Medical Journal, 19*(1): e29700. DOI: 10.5812/ircmj.29700

Rivero-Méndez, M., Suárez, E., Solís-Báez, S. S., Hernández, G., Cordero, W., Vázquez, I., . . . Bonilla, J. L. (2010). Internal consistency of the Spanish health literacy test (TOFHLA-SPR) for Puerto Rico. *Puerto Rico Health Sciences Journal, 29*(1), 49.

Robert, S. A., & Booske, B. C. (2011). US opinions on health determinants and social policy as health policy. *American Journal of Public Health, 101*(9), 1655-1663.

Robinson, S., Moser, D., Pelter, M. M., Nesbitt, T., Paul, S. M., & Dracup, K. (2011). Assessing health literacy in heart failure patients. *Journal of Cardiac Failure, 17*(11), 887-892.

Rouse, S. V. (2007). Using reliability generalization methods to explore measurement error: An illustration using the MMPI–2 PSY–5 scales. *Journal of Personality Assessment*, *88*(3), 264-275.

Ruhlen, M. (1991). *A guide to the world's languages: Classification.* Stanford, CA: Stanford University Press.

Sánchez-Meca, J., López-López, J. A., & López-Pina, J. A. (2013). Some recommended statistical analytic practices when reliability generalization studies are conducted. *British Journal of Mathematical and Statistical Psychology*, *66*(3), 402-425.

Satherley, P., Lawes, E., & Sok, S. (2008). *The Adult Literacy and Life Skills (ALL) survey: Overview and international comparisons.* Comparative Education Research Unit, Research Division, Ministry of Education.

Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings.* Thousand Oaks, CA: Sage publications.

Sentell, T., & Braun, K. L. (2012). Low health literacy, limited English proficiency, and health status in Asians, Latinos, and other racial/ethnic groups in California. *Journal of Health Communication, 17*(sup3), 82-99.

Shaw, S. J., Armin, J., Torres, C. H., Orzech, K. M., & Vivian, J. (2012). Chronic disease self-management and health literacy in four ethnic groups. *Journal of Health Communication, 17*(sup3), 67-81.

Smith, D. L. (2012). Health care disparities for persons with limited English proficiency: Relationships from the 2006 Medical Expenditure Panel Survey (MEPS). *Journal of Health Disparities Research and Practice, 3*(3), 57-67.

Solano-Flores, G., & Trumbull, E. (2003). Examining language in context: The need for new research and practice paradigms in the testing of English-language learners. *Educational Researcher, 32*(2), 3-13.

Solano-Flores, G. (2008). Who is given tests in what language by whom, when, and where? The need for probabilistic views of language in the testing of English language learners. *Educational Researcher, 37*(4), 189-199.

Solano-Flores, G., Contreras-Niño, L. A., & Backhoff, E. (2012). The measurement of translation error in PISA-2006 items: An application of the theory of test translation error. In M. Prenzel, M. Kobarg, K. Schöps, & S. Rönnebeck (Eds.), *Research in the context of the programme for international student assessment* (pp. 71–85). Germany: Springer Verlag.

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., Pelikan, J. M., Röthlin, F., Ganahl, K., Slonska, Z., Doyle, G., . . . Uiters, E. (2015). Health literacy in Europe: comparative results of the European health literacy survey (HLS-EU). *European Journal of Public Health, 25*(6), 1053-1058.

Soto Mas, F., Jacobson, H. E., & Olivárez, A. (2017). Adult Education and the Health Literacy of Hispanic Immigrants in the United States. *Journal of Latinos and Education,* 1-9.

Sperber, A. D., Devellis, R. F., & Boehlecke, B. (1994). Cross-cultural translation: methodology and validation. *Journal of Cross-Cultural Psychology, 25*(4), 501-524.

Stonbraker, S., Schnall, R., & Larson, E. (2015). Tools to measure health literacy among Spanish speakers: an integrative review of the literature. *Patient Education and Counseling, 98*(12), 1513-1523.

Tang, Y. H., Pang, S., Chan, M. F., Yeung, G., & Yeung, V. (2008). Health literacy, complication awareness, and diabetic control in patients with type 2 diabetes mellitus. *Journal of Advanced Nursing, 62*(1), 74-83.

Taylor, W. L. (1953). Cloze procedure: A new tool for measuring readability. *Journalism Quarterly, 30*, 415–433.

Thomason, T. R., & Mayo, A. M. (2015). A Critique of the Short Test of Functional Health Literacy in Adults. *Clinical Nurse Specialist, 29*(6), 308-312.

Thompson, B. (2006). *Foundations of behavioral statistics: An insight-based approach.* New York, NY: Guilford Press.

Thomson, M. D., & Hoffman-Goetz, L. (2011). Challenges of recruiting ESL immigrants into cancer education studies: reflections from practice notes. *Journal of Cancer Education, 26*(1), 170-174.

Toçi, E., Burazeri, G., Sørensen, K., Kamberi, H., & Brand, H. (2014). Concurrent validation of two key health literacy instruments in a South Eastern European population. *European Journal of Public Health, 25*(3), 482-486.

U.S. Census Bureau. (2015). Language spoken at home: 2015 American community survey 1-year estimates. Retrieved from https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_13_1YR_S1601&prodType=table

U.S. Department of Health and Human Services. (2009). *Healthy people 2010: understanding and improving health.* Washington, DC: U.S. Department of Health and Human Services.

UNESCO. (2004). *The plurality of literacy and its implications for policies and programmes.* UNESCO Education Sector Position Paper. UNESCO, Paris.

Vacha-Haase, T. (1998). Reliability generalization: Exploring variance in measurement error affecting score reliability across studies. *Educational and Psychological Measurement, 58*(1), 6-20.

Vacha-Haase, T., Henson, R. K., & Caruso, J. C. (2002). Reliability generalization: Moving toward improved understanding and use of score reliability. *Educational and Psychological Measurement, 62*(4), 562-569.

Wang, J. H. Y., Adams, I., Huang, E., Ashing-Giwa, K., Gomez, S. L., & Allen, L. (2012). Physical distress and cancer care experiences among Chinese-American and non-Hispanic White breast cancer survivors. *Gynecologic Oncology, 124*(3), 383-388.

Wang, J. H., Liang, W., Schwartz, M. D., Lee, M. M., Kreling, B., & Mandelblatt, J. S. (2008). Development and evaluation of a culturally tailored educational video: Changing breast cancer–related behaviors in Chinese women. *Health Education & Behavior, 35*(6), 806-820.

Wild, D., Grove, A., Martin, M., Eremenco, S., McElroy, S., Verjee-Lorenz, A., & Erikson, P. (2005). Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: Report of the ISPOR task force for translation and cultural adaptation. *Value in Health, 8*(2), 94-104.

Willson, V. L. (1980). Research techniques in AERJ articles: 1969 to 1978. *Educational Researcher, 9*(6), 5-10.

Willson, V. L., & Reynolds, C. R. (1985). Constructing short forms from composite tests: Reliability and validity. *Educational and Psychological Measurement, 45*(3), 469-476.

Wilson, E., Chen, A. H., Grumbach, K., Wang, F., & Fernandez, A. (2005). Effects of limited English proficiency and physician language on health care comprehension. *Journal of General Internal Medicine, 20*(9), 800-806.

Wray-Lake, L., Tang, J., & Victorino, C. (2017). Are they political? Examining Asian American college students' civic engagement. *Asian American Journal of Psychology, 8*(1), 31.

Yip, M. P. (2012). A health literacy model for limited English speaking populations: Sources, context, process, and outcomes. *Contemporary Nurse, 40*(2), 160-168.

Yu, C. Y. (2002). Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes. Doctoral dissertation, University of California Los Angeles.

Zhou, M., & Cai, G. (2002). Chinese language media in the United States: Immigration and assimilation in American life. *Qualitative Sociology, 25*(3), 419-440.

English S-TOFHLA Reading Items

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3	4	5	6	7	8	9	10
1	scorE1	54.7										
2	scorE2	70.2	67.2									
3	scorE3	68.8	65.9	80.6								
4	scorE4	75.0	68.1	72.6	70.4							
5	scorE5	99.5	77.8	84.5	79.3	92.6						
6	scorE6	80.2	67.0	74.3	70.9	79.1	79.8					
7	scorE7	104.3	75.7	78.3	79.1	78.0	84.0	88.5				
8	scorE8	76.3	69.9	73.8	70.6	75.8	85.1	76.7	78.0			
9	scorE9	65.8	65.8	69.7	68.8	69.2	75.2	71.6	76.0	73.2		
10	scorE10	125.0	89.6	89.1	91.6	90.7	92.5	89.6	92.4	92.2	89.5	
11	scorE11	95.4	76.3	77.9	77.2	79.5	85.0	79.7	82.8	82.1	75.2	100.7
12	scorE12	57.3	62.5	70.5	65.2	74.7	80.1	77.3	76.8	72.2	68.5	90.8
13	scorE13	69.7	69.7	70.8	71.0	76.8	79.6	76.0	78.3	75.0	71.8	94.2
14	scorE14	75.3	71.8	72.1	70.7	76.1	78.6	76.9	78.3	77.9	73.4	92.6
15	scorE15	56.4	60.4	71.2	67.6	70.3	78.4	71.2	75.9	71.4	70.6	94.0
16	scorE16	73.7	68.3	73.9	70.7	78.3	82.0	79.0	79.7	75.8	75.5	91.6
17	scorE17	79.7	66.9	72.9	69.2	71.9	83.8	75.6	79.5	76.3	68.1	89.5
18	scorE18	89.6	81.0	79.6	78.4	83.7	85.1	81.7	81.0	82.8	77.9	96.3
19	scorE19	53.2	49.7	62.3	60.7	60.1	72.3	65.4	75.1	64.5	56.5	88.1
20	scorE20	99.0	72.9	81.7	76.7	76.5	85.1	80.3	82.5	79.8	75.0	93.3
21	scorE21	132.3	93.0	93.7	94.3	96.5	96.8	94.7	97.6	94.9	97.0	101.5
22	scorE22	111.6	82.5	87.1	83.0	85.9	87.8	86.5	88.0	85.5	81.8	98.0
23	scorE23	92.2	71.3	78.8	75.0	74.9	86.1	75.4	82.6	80.3	74.1	92.5
24	scorE24	97.9	75.5	76.1	76.6	77.5	83.0	79.0	82.5	81.4	73.9	92.6
25	scorE25	80.7	61.0	64.7	65.6	66.7	76.7	75.9	78.9	72.7	63.6	89.3
26	scorE26	89.6	67.8	70.8	75.0	71.0	77.8	76.4	80.5	74.1	70.3	91.8
27	scorE27	104.8	77.7	82.6	83.7	79.4	86.9	81.2	85.7	88.2	79.3	95.7
28	scorE28	119.7	87.3	89.5	92.2	88.6	89.7	88.8	90.5	90.6	90.4	100.9
29	scorE29	123.5	87.6	88.4	88.7	88.9	91.3	90.7	92.2	88.0	87.0	98.3
30	scorE30	110.3	78.3	79.7	80.3	83.4	87.6	81.0	86.4	83.5	79.8	95.7
31	scorE31	64.1	54.5	60.9	59.1	61.3	74.6	67.8	78.4	63.0	57.6	88.3
32	scorE32	109.3	81.3	84.6	82.6	80.4	87.0	81.2	86.7	83.8	80.1	94.9
33	scorE33	116.8	88.7	91.0	91.4	90.3	92.2	88.6	91.7	90.0	88.3	98.5
34	scorE34	45.3	47.2	54.9	57.4	58.0	72.7	59.5	74.7	62.0	52.0	88.0
35	scorE35	84.4	63.8	71.3	67.0	70.7	80.7	73.8	80.5	70.5	66.6	89.7
36	scorE36	134.8	95.2	95.1	95.9	95.0	100.8	95.7	97.4	96.0	94.7	103.8

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	scorE11	95.4										
12	scorE12	57.3	77.5									
13	scorE13	69.7	81.9	73.4								
14	scorE14	75.3	84.6	74.4	89.8							
15	scorE15	56.4	77.3	66.8	69.9	71.2						
16	scorE16	73.7	79.3	74.1	77.2	78.0	73.2					
17	scorE17	79.7	75.7	70.0	75.4	74.9	71.7	73.4				
18	scorE18	89.6	86.8	75.8	85.1	82.9	75.6	80.8	80.3			
19	scorE19	53.2	69.6	57.2	60.7	63.5	55.1	62.3	64.1	74.6		
20	scorE20	99.0	85.8	79.9	77.9	78.7	77.3	80.6	77.2	84.1	73.8	
21	scorE21	132.3	96.6	----	93.9	94.0	93.3	93.9	94.2	95.6	97.2	96.4
22	scorE22	111.6	89.4	85.3	89.2	87.5	85.5	86.2	82.5	88.2	81.6	89.3
23	scorE23	92.2	82.7	73.6	74.9	75.8	77.2	77.6	75.6	80.1	68.5	82.6
24	scorE24	97.9	81.3	77.5	79.4	80.2	79.0	78.9	80.3	82.5	70.2	81.6
25	scorE25	80.7	75.9	66.8	67.3	68.2	68.1	69.4	68.9	74.1	62.0	76.1
26	scorE26	89.6	78.3	70.6	76.1	76.0	71.0	72.2	74.5	79.7	65.6	77.8
27	scorE27	104.8	90.1	80.0	81.2	81.5	84.2	85.2	79.4	86.1	75.4	88.3
28	scorE28	119.7	95.1	89.4	89.5	88.0	91.6	89.7	88.1	92.7	87.6	93.4
29	scorE29	123.5	94.8	88.7	89.9	90.1	89.5	89.3	90.4	90.8	86.9	91.5
30	scorE30	110.3	86.6	80.4	80.9	81.5	82.6	81.8	80.7	86.0	78.1	84.9
31	scorE31	64.1	71.2	56.4	60.7	62.3	57.0	62.2	65.9	70.7	54.7	72.3
32	scorE32	109.3	89.0	80.7	85.4	87.6	82.2	84.6	84.5	89.0	77.3	89.8
33	scorE33	116.8	93.6	87.3	90.6	90.2	90.3	88.8	89.9	93.7	84.1	92.4
34	scorE34	45.3	68.8	51.1	56.7	61.2	49.9	60.1	59.2	68.1	46.9	70.1
35	scorE35	84.4	75.3	66.3	68.1	70.8	66.6	72.3	72.7	77.7	62.6	77.2
36	scorE36	134.8	97.5	94.9	95.5	94.9	96.8	95.1	95.1	96.6	95.2	97.1

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29	30
21	scorE21	132.3										
22	scorE22	111.6	98.6									
23	scorE23	92.2	96.4	92.3								
24	scorE24	97.9	96.1	88.3	84.0							
25	scorE25	80.7	97.5	82.0	73.2	76.0						
26	scorE26	89.6	98.1	87.8	76.4	77.5	72.9					
27	scorE27	104.8	98.7	92.0	86.9	88.6	79.7	83.8				
28	scorE28	119.7	101.7	95.5	92.5	90.4	92.3	92.6	99.2			
29	scorE29	123.5	100.9	96.4	90.1	91.6	88.2	92.0	96.3	102.4		
30	scorE30	110.3	101.0	92.2	84.3	88.1	81.0	83.0	91.4	92.9	93.5	
31	scorE31	64.1	94.6	80.7	69.3	71.9	63.9	67.4	77.4	85.7	87.4	78.9
32	scorE32	109.3	98.1	92.7	86.2	87.7	80.5	83.3	91.0	95.1	96.8	88.1
33	scorE33	116.8	100.2	92.2	92.6	92.6	88.2	94.8	97.3	101.7	97.8	96.5

34	scorE34	45.3	96.6	79.4	68.8	72.0	59.4	64.9	75.3	85.0	87.3	77.5
35	scorE35	84.4	99.6	83.5	74.3	75.9	70.2	72.7	81.1	86.6	88.5	81.7
36	scorE36	134.8	104.2	98.7	99.0	98.6	95.1	95.7	100.3	101.5	100.6	99.2

Item	Label	Marginal X^2	31	32	33	34	35
31	scorE31	64.1					
32	scorE32	109.3	82.2				
33	scorE33	116.8	86.7	98.9			
34	scorE34	45.3	52.8	79.1	85.8		
35	scorE35	84.4	65.7	81.4	90.3	61.3	
36	scorE36	134.8	94.7	98.8	104.7	95.3	95.6

Chinese S-TOFHLA Reading Items

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3	4	5	6	7	8	9	10
1	scorC1	0.0										
2	scorC2	0.1	3.3									
3	scorC3	0.0	-0.7	-0.5								
4	scorC4	0.0	0.3	2.0	0.8							
5	scorC5	0.0	-0.1	1.4	3.8	-0.1						
6	scorC6	0.0	0.4	-0.2	1.4	2.5	4.2					
7	scorC7	0.0	0.0	-0.6	-0.6	-0.3	0.2	0.9				
8	scorC8	0.1	-0.1	0.6	2.1	3.6	3.4	7.4	1.1			
9	scorC9	0.0	1.0	0.1	-0.5	-0.4	-0.4	-0.6	-0.5	-0.6		
10	scorC10	0.0	0.4	1.9	-0.6	1.4	-0.4	0.6	0.5	3.7	-0.2	
11	scorC11	0.0	-0.5	3.0	-0.4	1.6	1.8	0.9	-0.7	5.2	-0.7	2.3
12	scorC12	0.0	0.8	-0.5	1.3	0.3	1.2	1.9	-0.5	2.1	-0.7	2.6
13	scorC13	0.0	-0.7	-0.6	-0.6	-0.6	0.3	-0.3	-0.7	1.4	0.4	0.9
14	scorC14	0.1	-0.4	-0.6	0.6	4.3	-0.4	1.2	-0.6	1.4	0.1	2.2
15	scorC15	0.0	-0.7	1.3	1.4	-0.2	-0.2	1.9	-0.6	3.3	-0.1	3.1
16	scorC16	0.0	-0.0	-0.1	-0.6	-0.1	-0.6	-0.3	-0.7	0.9	-0.4	3.6
17	scorC17	0.0	0.5	-0.6	-0.6	0.2	-0.7	-0.5	-0.5	-0.5	-0.7	-0.7
18	scorC18	0.0	0.6	-0.4	1.1	-0.7	-0.2	0.2	-0.3	-0.2	1.6	2.3

19	scorC19	0.0	-0.6	-0.4	-0.4	-0.7	-0.7	-0.7	-0.6	0.8	-0.7	-0.3
20	scorC20	0.0	0.2	2.6	-0.7	-0.6	-0.6	-0.7	-0.6	-0.1	-0.0	-0.6
21	scorC21	0.1	4.2	-0.5	-0.6	1.1	0.1	1.9	-0.4	-0.6	-0.6	-0.6
22	scorC22	0.0	-0.2	-0.5	-0.7	-0.7	-0.6	-0.1	-0.4	-0.6	-0.7	-0.6
23	scorC23	0.0	-0.7	0.4	-0.5	0.1	-0.6	-0.5	1.5	0.2	-0.6	-0.7
24	scorC24	0.0	-0.6	-0.5	-0.6	-0.6	-0.6	1.2	-0.5	-0.0	-0.6	-0.1
25	scorC25	0.0	0.1	0.0	-0.4	-0.7	0.3	0.4	1.3	0.5	-0.6	-0.6
26	scorC26	0.0	-0.7	-0.7	-0.3	-0.4	0.2	-0.5	0.3	0.0	-0.3	0.1
27	scorC27	0.0	2.2	-0.6	-0.6	0.4	-0.4	0.2	-0.7	-0.3	0.2	-0.6
28	scorC28	0.0	-0.6	2.2	-0.4	-0.0	-0.3	2.0	-0.5	0.2	-0.6	1.4
29	scorC29	0.0	-0.7	-0.5	0.7	-0.4	-0.2	-0.6	-0.7	-0.3	-0.7	-0.4
30	scorC30	0.0	-0.5	0.2	0.1	-0.4	-0.7	0.5	0.5	0.0	3.4	-0.3
31	scorC31	0.0	1.9	-0.3	-0.6	-0.6	-0.5	-0.6	-0.5	-0.6	-0.7	1.1
32	scorC32	0.0	-0.6	-0.6	-0.5	-0.3	0.4	-0.2	-0.2	-0.5	-0.7	-0.6
33	scorC33	0.0	-0.7	-0.4	-0.7	0.5	-0.5	-0.3	-0.4	-0.6	-0.1	-0.5
34	scorC34	0.0	-0.7	-0.1	-0.7	-0.7	0.1	-0.4	0.7	0.3	0.2	-0.7
35	scorC35	0.0	-0.0	-0.5	-0.5	-0.5	1.0	-0.5	-0.4	0.1	0.6	-0.6
36	scorC36	0.0	-0.1	1.0	-0.7	0.4	-0.7	-0.7	-0.2	-0.6	1.4	-0.0

Item	Label	Marginal X^2	11	12	13	14	15	16	17	18	19	20
11	scorC11	0.0										
12	scorC12	0.0	3.5									
13	scorC13	0.0	-0.6	-0.3								
14	scorC14	0.1	0.9	3.1	-0.3							
15	scorC15	0.0	-0.0	1.1	3.0	0.8						

16	scorC16	0.0	2.0	6.6	1.8	1.2	3.1					
17	scorC17	0.0	-0.5	-0.5	-0.7	0.1	-0.7	0.8				
18	scorC18	0.0	-0.5	5.5	-0.7	2.8	-0.3	1.2	0.3			
19	scorC19	0.0	0.8	-0.6	-0.2	-0.5	-0.3	-0.7	3.8	0.1		
20	scorC20	0.0	-0.7	-0.7	-0.5	0.3	-0.2	-0.7	-0.2	-0.6	0.6	
21	scorC21	0.1	-0.6	-0.1	-0.6	0.7	-0.6	-0.6	3.0	-0.2	3.5	-0.3
22	scorC22	0.0	-0.4	-0.2	-0.0	-0.6	-0.5	-0.6	-0.7	-0.3	-0.4	0.6
23	scorC23	0.0	0.1	-0.7	1.0	-0.3	-0.7	-0.6	0.7	-0.6	-0.5	-0.2
24	scorC24	0.0	-0.6	-0.0	0.5	1.1	-0.7	1.1	-0.4	0.9	-0.7	-0.7
25	scorC25	0.0	0.7	-0.5	-0.7	0.5	-0.4	-0.5	0.8	-0.7	0.3	0.4
26	scorC26	0.0	0.9	-0.6	-0.5	-0.6	-0.7	-0.6	0.4	0.0	1.2	-0.6
27	scorC27	0.0	-0.6	-0.1	-0.6	0.2	-0.0	2.4	0.4	-0.3	1.4	-0.7
28	scorC28	0.0	-0.0	-0.3	-0.6	1.0	1.0	0.7	-0.7	-0.7	-0.2	-0.2
29	scorC29	0.0	-0.1	1.6	-0.7	-0.6	-0.7	-0.7	1.9	0.0	1.7	3.3
30	scorC30	0.0	-0.5	-0.4	0.1	-0.6	-0.5	-0.5	3.4	0.0	4.5	-0.2
31	scorC31	0.0	-0.7	-0.7	-0.7	1.7	-0.6	-0.7	-0.2	-0.1	-0.7	-0.2
32	scorC32	0.0	-0.3	-0.6	-0.7	-0.6	-0.7	-0.6	-0.7	2.1	0.2	-0.5
33	scorC33	0.0	-0.3	-0.4	0.2	-0.5	-0.6	-0.7	-0.5	2.5	0.9	-0.5
34	scorC34	0.0	-0.5	-0.4	-0.4	-0.6	-0.6	-0.7	-0.6	-0.7	-0.6	-0.1
35	scorC35	0.0	0.5	1.2	-0.6	0.1	0.8	-0.7	-0.3	-0.2	-0.5	-0.6
36	scorC36	0.0	-0.3	-0.1	-0.7	2.9	-0.7	-0.6	-0.7	-0.5	-0.7	-0.1

Item	Label	Marginal X^2	21	22	23	24	25	26	27	28	29	30
21	scorC21	0.1										
22	scorC22	0.0	0.0									

23	scorC23	0.0	-0.6	-0.2								
24	scorC24	0.0	-0.5	-0.7	1.1							
25	scorC25	0.0	0.2	-0.7	-0.5	-0.1						
26	scorC26	0.0	0.1	-0.6	-0.5	-0.6	1.7					
27	scorC27	0.0	0.4	-0.7	-0.7	0.3	0.8	-0.3				
28	scorC28	0.0	0.3	0.6	-0.3	-0.6	1.0	0.6	3.2			
29	scorC29	0.0	-0.5	0.2	-0.6	-0.4	-0.7	1.9	-0.7	1.0		
30	scorC30	0.0	2.2	-0.6	0.7	-0.7	-0.3	-0.7	0.0	-0.1	-0.5	
31	scorC31	0.0	2.6	0.3	3.1	0.8	2.5	-0.7	3.3	-0.6	-0.5	0.2
32	scorC32	0.0	0.6	0.4	-0.6	-0.6	-0.4	-0.6	-0.6	-0.7	-0.0	1.3
33	scorC33	0.0	-0.1	0.2	-0.7	-0.6	-0.3	0.1	0.1	-0.7	-0.3	0.6
34	scorC34	0.0	-0.1	-0.2	-0.5	-0.4	-0.7	2.9	0.4	0.2	0.3	-0.3
35	scorC35	0.0	-0.6	-0.5	0.2	-0.4	-0.2	0.2	-0.1	-0.7	-0.4	-0.6
36	scorC36	0.0	0.5	-0.4	-0.7	1.1	-0.1	-0.6	-0.0	-0.7	-0.6	-0.6

Item	Label	Marginal X^2	31	32	33	34	35
31	scorC31	0.0					
32	scorC32	0.0	2.7				
33	scorC33	0.0	-0.7	-0.2			
34	scorC34	0.0	0.2	-0.3	-0.1		
35	scorC35	0.0	-0.7	-0.2	-0.7	0.4	
36	scorC36	0.0	0.3	-0.7	-0.7	3.7	-0.7

English S-TOFHLA Numeracy Items

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3
1	NscorE1	0.0			
2	NscorE2	0.0	-0.5		
3	NscorE3	0.0	0.2	-0.6	
4	NscorE4	0.2	-0.4	0.1	-0.3

Chinese S-TOFHLA Numeracy Items

Marginal fit (X^2) and Standardized LD X^2 Statistics for Group 1 **(Back to TOC)**

Item	Label	Marginal X^2	1	2	3
1	NscorC1	0.0			
2	NscorC2	0.0	-0.3		
3	NscorC3	0.0	-0.7	-0.6	
4	NscorC4	0.0	-0.7	-0.7	-0.6

APPENDIX B. ITEM DIFFICULTY AND ITEM DISCRIMINATION

English S-TOFHLA Reading Items

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ **(Back to TOC)**

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	scorE1	2	3.57	0.42	1	-0.67	0.28	0.19	0.06
2	scorE2	4	5.07	0.69	3	-0.84	0.41	0.17	0.07
3	scorE3	6	4.60	0.60	5	-0.89	0.37	0.19	0.06
4	scorE4	8	4.82	0.66	7	-1.10	0.39	0.23	0.06
5	scorE5	10	6.49	1.21	9	-1.91	0.71	0.29	0.06
6	scorE6	12	5.63	0.80	11	-1.16	0.44	0.21	0.06
7	scorE7	14	5.93	0.89	13	-2.77	0.58	0.47	0.04
8	scorE8	16	6.23	0.93	15	-0.96	0.48	0.15	0.06
9	scorE9	18	4.32	0.54	17	-0.85	0.33	0.20	0.06
10	scorE10	20	8.42	2.04	19	-3.32	1.19	0.39	0.05
11	scorE11	22	5.82	0.93	21	-1.84	0.55	0.32	0.05
12	scorE12	24	8.04	1.41	23	-0.19	0.54	0.02	0.07
13	scorE13	26	5.70	0.80	25	-0.74	0.42	0.13	0.06
14	scorE14	28	6.37	0.95	27	-0.92	0.48	0.14	0.06
15	scorE15	30	5.92	0.81	29	-0.23	0.40	0.04	0.06
16	scorE16	32	6.47	0.98	31	-0.85	0.48	0.13	0.06
17	scorE17	34	4.78	0.67	33	-1.35	0.43	0.28	0.06
18	scorE18	36	8.17	1.48	35	-1.50	0.66	0.18	0.06
19	scorE19	38	3.08	0.41	37	-1.53	0.31	0.50	0.05
20	scorE20	40	5.82	0.87	39	-2.07	0.53	0.36	0.05
21	scorE21	42	11.06	2.02	41	-6.61	1.34	0.60	0.03
22	scorE22	44	7.55	1.86	43	-2.46	1.06	0.33	0.07
23	scorE23	46	5.46	0.83	45	-1.78	0.51	0.33	0.05
24	scorE24	48	5.79	0.89	47	-2.01	0.56	0.35	0.05
25	scorE25	50	4.44	0.72	49	-2.24	0.50	0.50	0.05
26	scorE26	52	4.92	0.84	51	-2.16	0.59	0.44	0.06
27	scorE27	54	6.82	1.11	53	-2.15	0.63	0.32	0.05
28	scorE28	56	8.60	1.63	55	-2.92	0.87	0.34	0.05
29	scorE29	58	7.86	1.42	57	-3.40	0.87	0.43	0.04
30	scorE30	60	6.42	1.15	59	-2.90	0.76	0.45	0.05
31	scorE31	62	3.66	0.66	61	-2.07	0.53	0.56	0.05
32	scorE32	64	6.69	1.53	63	-2.48	0.95	0.37	0.06
33	scorE33	66	10.00	4.33	65	-2.89	2.14	0.29	0.09
34	scorE34	68	2.91	0.42	67	-1.94	0.34	0.67	0.05
35	scorE35	70	4.61	0.74	69	-2.10	0.53	0.46	0.05
36	scorE36	72	9.43	1.68	71	-4.42	1.01	0.47	0.03

Chinese S-TOFHLA Reading Items

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	scorC1	2	1.47	0.35	1	4.20	0.48	-2.85	0.50
2	scorC2	4	1.86	0.49	3	5.50	0.79	-2.96	0.50
3	scorC3	6	2.12	0.52	5	5.58	0.82	-2.63	0.40
4	scorC4	8	0.76	0.32	7	3.86	0.39	-5.09	1.88
5	scorC5	10	2.09	0.46	9	4.96	0.61	-2.37	0.34
6	scorC6	12	1.04	0.25	11	2.88	0.27	-2.76	0.53
7	scorC7	14	0.80	0.17	13	1.13	0.13	-1.41	0.28
8	scorC8	16	1.85	0.50	15	5.86	0.88	-3.17	0.53
9	scorC9	18	1.09	0.19	17	0.61	0.13	-0.56	0.13
10	scorC10	20	1.62	0.43	19	4.88	0.63	-3.01	0.55
11	scorC11	22	1.25	0.34	21	4.32	0.48	-3.45	0.71
12	scorC12	24	1.47	0.36	23	4.19	0.48	-2.86	0.50
13	scorC13	26	1.96	0.47	25	4.06	0.50	-2.07	0.33
14	scorC14	28	2.82	0.70	27	6.88	1.09	-2.44	0.34
15	scorC15	30	1.98	0.48	29	4.73	0.59	-2.39	0.37
16	scorC16	32	1.82	0.46	31	4.52	0.57	-2.49	0.41
17	scorC17	34	1.94	0.50	33	4.96	0.67	-2.56	0.42
18	scorC18	36	1.45	0.34	35	4.09	0.46	-2.83	0.46
19	scorC19	38	1.20	0.28	37	3.52	0.35	-2.92	0.52
20	scorC20	40	1.78	0.43	39	4.05	0.48	-2.27	0.36
21	scorC21	42	2.28	0.62	41	5.98	0.92	-2.63	0.40
22	scorC22	44	1.67	0.44	43	4.72	0.59	-2.82	0.49
23	scorC23	46	1.88	0.37	45	3.28	0.37	-1.74	0.23
24	scorC24	48	2.09	0.57	47	4.47	0.66	-2.14	0.32
25	scorC25	50	0.64	0.15	49	0.65	0.11	-1.01	0.26
26	scorC26	52	0.54	0.14	51	0.94	0.12	-1.76	0.47
27	scorC27	54	0.48	0.14	53	0.96	0.12	-1.98	0.58
28	scorC28	56	1.01	0.18	55	0.68	0.12	-0.68	0.14
29	scorC29	58	0.31	0.13	57	-0.36	0.10	1.15	0.56
30	scorC30	60	1.00	0.21	59	2.40	0.21	-2.41	0.42
31	scorC31	62	1.49	0.36	61	4.49	0.53	-3.01	0.49
32	scorC32	64	2.33	0.63	63	4.36	0.70	-1.87	0.25
33	scorC33	66	1.61	0.28	65	2.65	0.28	-1.65	0.19
34	scorC34	68	2.10	0.56	67	4.21	0.65	-2.01	0.28
35	scorC35	70	1.72	0.44	69	3.86	0.51	-2.24	0.34
36	scorC36	72	1.15	0.25	71	2.45	0.23	-2.12	0.34

English S-TOFHLA Numeracy Items

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	NscorE1	2	12.64	6.08	1	4.11	2.19	-0.33	0.04
2	NscorE2	4	99.55	5.32	3	26.90	1.29	-0.27	0.02
3	NscorE3	6	11.25	4.75	5	3.63	1.69	-0.32	0.04
4	NscorE4	8	4.95	0.90	7	0.53	0.37	-0.11	0.07

Chinese S-TOFHLA Numeracy Items

2PL Model Item Parameter Estimates for Group 1, logit: $a\theta + c$ or $a(\theta - b)$ (Back to TOC)

Item	Label		*a*	*s.e.*		*c*	*s.e.*	*b*	*s.e.*
1	NscorC1	2	1.59	0.45	1	3.35	0.48	-2.11	0.37
2	NscorC2	4	3.23	1.27	3	6.51	1.89	-2.02	0.26
3	NscorC3	6	2.20	0.68	5	3.58	0.72	-1.63	0.23
4	NscorC4	8	1.18	0.32	7	2.22	0.26	-1.88	0.36