STATISTICAL ANALYSIS PLAN

REDUCING ASSESSMENT BARRIERS FOR PATIENTS WITH LOW LITERACY

Version 4.2

Sites: Northwestern University and Boston University Medical Center
Version 1 Date: 21 May 2018
Version 2 Date: 27 March 2019
Version 3 Date: 27 February 2020
Version 3.1 Date: 02 March 2020
Version 3.2 Date: 09 June 2020
Version 4.0 Date: 09 November 2020
Version 4.1 Date: 10 March 2021
Version 4.2 Date: 02 June 2021

Principal Investigators: James W Griffith, PhD and Michael Paasche-Orlow, MD

NCT#: NCT03584490

Table of Contents

Introduction

Study Background

The concept of *health literacy* is the degree to which an individual can obtain, process, and understand basic health information,[1-6] which includes the ability to understand and respond to questionnaires. Questionnaires are susceptible to being inappropriate for the respondent (e.g., containing overly complex or confusing questions), so it is important to examine questionnaire properties so that participants can express their symptoms in clinical care and research.

In recent years, the NIH has invested heavily in the creation of the Patient-Reported Outcomes Measurement Information System (PROMIS),[7-13] which is a large collection of state-of-the-art questionnaires designed to measure symptoms and other health-related concepts. PROMIS has advanced health measurement and the ability to incorporate patients' perspectives into clinical decision-making. But PROMIS, like almost all other questionnaires, has not been validated for use by people with low health literacy, and may work poorly for individuals with low levels of health literacy. Thus, people with low literacy may be inaccurately assessed in all research and clinical settings due to unrecognized measurement bias of commonly used tools. The goal of this study is to determine how health literacy affects selected PROMIS and other questionnaires, so that these tools can be improved and health assessment can be precise for all participants.

At the time of this version of the analysis plan (Version 4.0), we have made several modifications in response to the COVID-19 pandemic, which has necessitated some changes to our protocol and analysis plan. None of the main, *a priori* analyses have been conducted, but we plan to carry out some analyses of the existing data using modified statistical procedures (see below). Analyses so far include 1) a trial run of downloading the data from REDCap and merging constituent files, 2) exploratory analyses of cognition (measured by the MOCA-B) and PROMIS Depression, and 3) exploratory analyses of screening data (Single Item Literacy Screener, demographics) as a predictor of health literacy; these latter analyses were done to help better stratify our participants. Our protocol and analysis plan will be updated as needed before the main *a priori* analyses are run. All analyses that were not pre-registered will be labelled as post-hoc, exploratory analyses. This document is meant to supplement the detailed study protocol and focuses on the analytic methodology. The goal of this document is to describe *a priori* statistical and methodological decisions, which will be adhered to rigorously through the analysis process.

Study Objectives

Using a large sample, we will determine the psychometric properties of health questionnaires across levels of health literacy. Testing will be conducted in English and Spanish, across two modes of administration (paper-and-pencil vs. talking touchscreen). Each participant will have three visits (baseline, 3-month, and 6-month follow-up), described below, in order to determine changes in psychometric properties over time. We will use *differential item functioning* (DIF) and other psychometric analyses to identify questionnaire items with varying characteristics across levels of health literacy. Before the COVID-19 pandemic, our original plan was to recruit N = 1216 participants for in-person assessments. At the time of this plan our N is 706 completed baseline assessments (note: one person dropped out and requested not to use their data, leaving N = 705). In-person recruitment is on hold and we are shifting activities to a phone-based protocol. For all in-person data, we will update this plan on an analysis-by-analysis basis before conducting the analyses.

- **Develop alternative modes of test administration that can better accommodate people with low health literacy.** National statistics indicate that 14% of adults in the United States have below basic reading skills, and over 75 million Americans have below basic or basic health literacy skills.[14] These data strongly indicate that many individuals seeking healthcare will have difficulties with written questionnaires, even those that have been carefully crafted such as those in the NIH PROMIS. We will test an alternative mode of administration that has been specifically designed to aid people of low health literacy – a computer-assisted "talking" version that

uses speech technology that participants can privately access as needed to have items and response options read to them. We will analyze the talking touchscreen groups and paper-based groups separately. We hypothesize that the talking touchscreen will reduce health literacy-related test bias by level of health literacy, relative to traditional paper-based methods. This will be evidenced by fewer items being flagged for DIF in the group that receives the talking touchscreen. In terms of descriptive statistics to display the magnitude of effects, we will report effect sizes (McFadden's R^2) of DIF items, as well as test characteristic curve (TCC) for DIF items. Among participants with low health literacy, we expect that the discrimination parameters of the touchscreen condition will be higher than the paper-based condition.

- **Determine the degree of differential item functioning (DIF) of tools relevant to primary care across different levels of health literacy.** DIF analyses will be conducted to identify test bias by level of health literacy. This will ensure that individuals with low literacy are assessed fairly in clinical and research settings. We will assess a broadly applicable instrument to primary care – the PROMIS Profile, which measures anxiety, depression, fatigue, pain interference, pain intensity, physical function, sleep disturbance, and social roles and activities. Other questionnaires included in the study are shown in Table 1. We will stratify participants into low versus adequate literacy levels using HealthLiTT (groups will be formed according to T scores of < 55 vs. 55+).[15] A secondary analysis will use the NUMI, a measure of health numeracy, in three groups (low = 0-3, medium = 4-6, high = 7 or greater). Then, we will determine whether psychometric properties of each questionnaire scale are consistent for individuals with low health literacy versus adequate health literacy. We hypothesize that DIF will be observed for multiple scales using paper-and-pencil administration, suggesting a need to improve test items, and a need to examine alternative ways to administer these tools.

Statistical methodology

Overview

Data will be summarized by group and in total. In summary tables of continuous variables, the minimum, maximum, mean, and standard deviation will be presented with one decimal place. We will use counts and percentages in table for categorical variables. The denominator for each percentage will be the number of participants within the group unless otherwise specified. All hypothesis testing will be carried out at a two-tailed Type I error rate of $\alpha = .05$ unless otherwise specified. P-values will be rounded to three decimal places with a minimum of < 0.001 in tables. If statistical methods change during analysis, this will be documented in all reports with justification. Additional analyses of these data will inevitable, but any exploratory analyses will be clearly marked as such.

Statistical power analysis and sample size selection

IRT and DIF analyses require large samples due to the large number of free parameters to be estimated. The proposed sample size herein is larger than other successful implementations of these same analyses carried out in other IRT studies.[16-19] Moreover, we have carried out computer simulations to determine if our proposed sample sizes will be sufficient to create IRT models. We assumed a one-factor model across two groups at n = 100 each with standardized factor loadings set to .7 for eight questionnaire items. We used two groups for this simulation to ensure that any pairwise model comparisons would have adequate power. For this simulation, we specified that a non-significant model $\chi 2$ was a good fit, which is a conservative decision because this particular test is highly sensitive relative to other measures of model fit, especially with large samples. Across 500 iterations, a good-fitting model was identified across groups 90% of the time. Even dropping the sample size to as low as n = 50 would identify a good-fitting model 84% of the time.

Missing data

We will describe the rate of missing data on an item-by-item basis using frequency tables in R. We do not anticipate high rates of missing data, but we will conduct subgroup analyses by health literacy, site, and administration mode as need to help diagnose any problems with missing data. For DIF analyses, IRT models

Table 1	
Visit 1 (Baseline)	**Description of instrument**
Language Assessment	To determine testing in English or Spanish
Single Item Health Literacy Screener	To screen for possible literacy status
Short Portable Mental Status Questionnaire (SPMSQ)	To determine cognitive status to ensure the possible participant has sufficient mental status to participate
Demographics	To assess participant characteristics
Health Literacy Assessment (Health LiTT)	Assessment of health literacy.
NUMI	Assessment of numeracy. That is, how well the person can process numerical information
PROMIS Profile	To assess health-related quality of life.
PHQ-9	Self-report measure of depression.
Berlin Questionnaire (for sleep)	Self-report measure of sleep quality
Purpose and Meaning (NIH Toolbox)	Self-report measure of the degree of meaning and purpose a person feels in their life.
Ruminative Responses Scale (RRS)	Self-report measure of negative, repetitive thoughts
PTSD Checklist	Self-report measure of posttraumatic stress disorder symptoms
* C-SSRS Lifetime	Self-report measure of recent history of suicidal ideation and behavior. (*This is not a variable of interest that will be used only to ensure patients' safety*).
*C-SSRS Last Contact	Self-report measure of suicidal ideation and behavior since the individual's most recent assessment. (*Used only as necessary to ensure patients' safety*).
Visit 1, but if missing at Visit 1, then given at Visit 2 or 3	
Screening tools for Alzheimer's disease and related dementias (ADRD)	This brief battery contains the Montreal Cognitive Assessment-Basic (MOCA-B), and selected items from the RUDAS. * Note: These items were added to the study after recruitment began, so we give them at Visit 1, but if they were not given at Visit 1, they are then administered at Visit 2 or 3
Visit 2 (3-months)	
PROMIS Profile	To assess health-related quality of life
PHQ-9	Self-report measure of depression
Berlin Questionnaire (for sleep)	Self-report measure of sleep quality
Purpose and Meaning (NIH Toolbox)	Self-report measure of the degree of meaning and purpose a person feels in their life.
RRS	Self-report measure of negative, repetitive thoughts.
PTSD Checklist	Self-report measure of posttraumatic stress disorder symptoms.
*C-SSRS Lifetime	Self-report measure of recent history of suicidal ideation and behavior. (*This is not a variable of interest that will be used only to ensure patients' safety*).
*C-SSRS Last Contact	Self-report measure of suicidal ideation and behavior since the individual's most recent assessment. (*Used only as necessary to ensure patients' safety*).
Visit 3 (6-months)	
PROMIS Profile	To assess health-related quality of life.
PHQ-9	Self-report measure of depression.
Berlin Questionnaire (for sleep)	Self-report measure of sleep quality.
Purpose and Meaning (NIH Toolbox)	Self-report measure of the degree of meaning and purpose a person feels in their life.
RRS	Self-report measure of negative, repetitive thoughts.
PTSD Checklist	Self-report measure of posttraumatic stress disorder symptoms.
*C-SSRS Lifetime	Self-report measure of recent history of suicidal ideation and behavior. (*This is not a variable of interest that will be used only to ensure patients' safety*).
*C-SSRS Last Contact	Self-report measure of suicidal ideation and behavior since the individual's most recent assessment. (*Used only as necessary to ensure patients' safety*).
Exit Interview	

are estimated using full-information approaches that use all available data and can accommodate missing responses. We do not foresee a need for an imputation strategy, but if that were to become necessary, we will update this plan accordingly before running any analyses.

Study Design

This is a randomized experiment aimed at determining whether talking touchscreen technology can reduce DIF and provide a better testing experience, relative to paper-based testing. The original study design is shown below in Table 2, although the COVID-19 pandemic has put the in-person data collection on hold.

Table 2: Study Design			
Northwestern		BUMC	
304 English speaking 152 with low health literacy 76 women g1 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g2 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 152 with adequate health literacy 76 women g3 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized talking touchscreen 19 Order 1 19 Order 2 76 men g4 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized talking touchscreen 19 Order 1 19 Order 2	304 Spanish speaking 152 with low health literacy 76 women g5 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g6 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 152 with adequate health literacy 76 women g7 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized talking touchscreen 19 Order 1 19 Order 2 76 men g8 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized talking touchscreen 19 Order 1 19 Order 2	304 English speaking 152 with low health literacy 76 women g9 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g10 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 152 with adequate health literacy 76 women g11 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g12 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2	304 Spanish speaking 152 with low health literacy 76 women g13 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g14 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 152 with adequate health literacy 76 women g15 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2 76 men g16 38 randomized to paper-based 19 Order 1 19 Order 2 38 randomized to talking touchscreen 19 Order 1 19 Order 2

Study Timepoints

Each participant will be asked to complete a baseline assessment (Visit 1), three-month follow-up (Visit 2), and a six-month follow-up (Visit 3). Longitudinal data will be analyzed, within language and within mode of administration (paper-based vs. talking touchscreen), on a scale-by-scale basis using multiple-group factor analysis to compare in Mplus software (see Table 3 schematic below). All items will be treated as categorical indicators with response categories collapsed if there are fewer than five responses for a given response option.

Table 3: Schematic of multiple-group confirmatory factor analysis to be used on a scale-by-scale basis			
	Visit 1	Visit 2	Visit 3
Adequate health literacy	*Factor loadings/ IRT parameters*	*Factor loadings/ IRT parameters*	*Factor loadings/ IRT parameters*
Low health literacy	*Factor loadings/ IRT parameters*	*Factor loadings/ IRT parameters*	*Factor loadings/ IRT parameters*

We hypothesize that a metric invariable model will fit well for the talking touchscreen method. Acceptable fit will be defined as CFI >= .90, TLI >= .90, and RMSEA <= .10. We hypothesize that unacceptable fit indices will be obtained for the paper-based testing. For descriptive statistics, we will report all fit indices produced by Mplus as well as factor loadings for each item. Longitudinal data will also be used to develop new longitudinal DIF tools in R, but these results will be considered exploratory until they can be tried and tested in other data sets.

Study Populations

The study population includes medical patients and community member, all might which complete questionnaires in other clinical and research contexts, ages 18 and older.

Study Variables

Demographics

All demographic variables are described in the protocol. We will report descriptive statistics of all demographics within language group as well as in total. We do not have specific hypotheses about demographics, but may conduct and report analyses of them, as well as presented additional analyses of demographics based on requests from manuscript reviewers.

Health literacy and numeracy

HealthLiTT and the NUMI will be used to determine health literacy groups using the above-described cutoffs. HealthLiTT is our main health literacy variable. All analyses of the NUMI are secondary.

Health questionnaires

The following questionnaires will be subjected to DIF analyses: the PROMIS Profile (physical function, anxiety, depression, fatigue, sleep disturbance, ability to participate in social roles and activities, and pain interference), the Berlin Questionnaire (for sleep), the NIH Toolbox scale for purpose and meaning, the PTSD checklist (20 symptom items assessing overall PTSD symptoms), the Ruminative Responses Scale (total score and brooding and reflection subscales), and the PHQ-9.

Screening tools for Alzheimer's disease and related dementias (ADRD)

Based on supplemental funding, a brief battery of screening tools for ADRD was added to the project. These screening tools include the Montreal Cognitive Assessment Test–Basic (MOCA-B),[20] the Visuospatial/Executive, Attention, and Language subtests of the MOCA 8.1,[21] and the Body Orientation, Fist/Palm, Judgement, and Language Generativity – Animal Naming subtests of the Rowland universal dementia assessment scale.[22] The first participant completed the ADRD screening tools on 10 July 2019. These tools are given at the baseline visit, but if the baseline visit has already occurred, the tools are given at one of the follow-up visits (Visit 2 or Visit 3).

Interim analyses to ensure data quality

In order to ensure data integrity and quality, we will conduct limited interim data analyses after 150 patients have enrolled. The goal of this analysis will be to conduct a "dry run" of data management procedures. We will import, merge, clean, and examine the data quality. Merging is necessary because data will be collected from multiple sites using different data collection platforms. Data merging will require uniformity of file format, data variable names, and the value coding within each variable. Once merged, we will examine the data for any out-of-range, non-numerical, unexpected or missing values that may represent technological or experimental errors. We will also compose statistical software syntax based on existing scoring rules; these rules will name and define composite variables at the questionnaire (scale) score level. Health variables will be correlated and examined to confirm the correct coding of the variables (e.g., re-coding of responses or reversing). We will document these data management steps and produce software syntax necessary for successful merger, any challenges encountered, and remedies we put in place. The overarching goal of these interim analyses are to prepare for the analyses of the final data set.

Primary analyses

DIF by health literacy

We hypothesize that DIF will be observed across adequate versus low levels of health literacy as measured by HealthLiTT.[15] This hypothesis was formulated before any data were examined. In terms of direction, we predict that low-health-literacy groups will show lower item slopes. All questionnaires and subscales mentioned in the "Core battery for DIF Testing" section of protocol will undergo DIF testing. To examine DIF across health literacy status, we will conduct separate analyses within language and within testing format (paper-based vs. talking touchscreen). The overall N for these analyses, within strata, will be 304. We will use the *lordif* package in R for these analyses. This involves several steps. Four models are fit to the data (see Box 1). The models are regressions that link the probability of response type (e.g., "never", "rarely", "sometimes", "often", "always") to a latent trait. If an item is free of DIF, then group (e.g., adequate vs. low health literacy) should have no effect. Thus, Model 2 and Model 3 should not fit significantly better than Model 1. It is well known that statistical significance is influenced, in part, by sample size. Thus, lordif uses effect-size measures to detect DIF, and to flag individual items where the effect of group exceeds a particular threshold. Effect sizes indicative of DIF will be chosen for each scale based on a Monte Carlo simulation carried out in lordif, which returns an effect size that would be unlikely given the null hypothesis of no DIF.

Box 1: Models fit by lordif

Model 0: $\text{logit}[P(Y \geq k)] = \alpha_k$
Model 1: $\text{logit}[P(Y \geq k)] = \alpha_k + \beta_1 (\text{ability})$
Model 2: $\text{logit}[P(Y \geq k)] = \alpha_k + \beta_1 (\text{ability}) + \beta_2 (\text{group})$
Model 3: $\text{logit}[P(Y \geq k)] = \alpha_k + \beta_1 (\text{ability}) + \beta_2 (\text{group}) + \beta_3 (\text{ability} \times \text{group})$

Note: α and β are logistic regression parameters. k refers to the number of thresholds needed for each item.

The following *a priori* decisions are important. We will use a Type I error rate of $\alpha = .01$ to generate the McFadden ΔR^2 effect size for each item that would be considered unusual if the null hypothesis of no DIF were true. Within each scale, we will select the smallest McFadden R^2 across all of the items. The minimum effect size we will consider is .001. This effect size will then be used in lordif to detect items that show DIF. It should be noted that the simulation results from lordif would be a minimum, in which case we would need to choose larger (i.e., more conservative) effect sizes if lordif identifies a large number of items. In the event that more than 50% of items are flagged for DIF, we will re-run lordif with successive increases in .005 increments in the DIF detection criterion. This will allow us to identify which items produce the highest R^2 values consistently, regardless of the *a priori* R^2 DIF detection criterion

We will compile a list of items flagged for uniform (Model 1 vs. 2 from Box 1), non-uniform (Model 2 vs. 3), and total DIF (Model 1 vs. 3). Our a priori hypothesis is that DIF will be observed across levels of health literacy for paper-based testing. We further hypothesize that DIF will also be found for the talking touchscreen, but for fewer items. In addition, we hypothesize that the magnitude of each DIF item's ΔR^2 value will be larger in the paper-and-pencil group relative to the touch-screen group.

In addition to testing these specific hypotheses, we will create reports of descriptive statistics. For each item flagged for DIF, we will, 1) report the ΔR^2 values associated with uniform and non-uniform DIF, 2) summarize all IRT parameters across groups, 3) inspect the magnitude of θ versus a DIF-purified θ (see reference[23] for computational details) on a participant-by-participant basis to determine the practical effects of DIF, and 4) evaluate test characteristics curves (TCCs) to determine the effect of DIF on the overall test score.

In addition to scale-by-scale analyses, we will conduct exploratory analyses of questionnaires together (e.g., PROMIS depression together with PHQ-9, PROMIS sleep and Berlin), to explore the power of DIF analyses using multiple scales of the same construct.

Prediction of depression: Health literacy as a moderator of rumination, pain, fatigue, and sleep disturbance

We will determine the degree to which health literacy moderates the relationship between depression and associated constructs of rumination, pain intensity, pain interference, sleep disturbance, and fatigue. Depression will be measured using both the Patient Health Questionnaire (PHQ-9), a nine-item scale with each item scored between 0 and 3 providing a 0-27 severity score with higher scores representing greater depression severity. The NIH Patient-Reported Outcomes Measurement Information System (PROMIS Profile 57 v 2.0) will also be used to measure depression, as well as pain intensity, pain interference, sleep disturbance, and fatigue. PROMIS is a self-report measure with subscales producing a T-Score; higher T-scores representing a higher level of the construct. Rumination will be measured using the Ruminative Responses Scale (RRS) short form, a 10-item form that measures two factors, brooding and reflection. The RRS uses a four-point scale for each item with a total score derived from the sum of the individual items. Higher scores indicate higher levels of rumination. All variables will be examined in a continuous manner, rather than using cut scores.

Multiple regression analyses will be conducted using the *lm()* function in base R to test a moderation model for the PHQ-9 and PROMIS depression scale separately with health literacy (a continuous score from Health LiTT) acting as the proposed moderator between depression symptoms and pain intensity and interference, sleep disturbance, fatigue, and rumination. PROMIS profile scale scores will be entered into the regression model together. Health literacy will then be added as a moderator. In accord with our hypotheses, the addition of health literacy into the model should explain a significantly greater portion of the variance than symptom scores alone. PROMIS profile scale scores will then be removed from the model in a step-down procedure to determine the contribution of each scale to the model. We hypothesize that there will be a significant interaction between health literacy and symptoms scores. Specifically, we hypothesize that having higher health literacy will result in a stronger association between rumination and depression, whereas individuals with lower health literacy will evidence a stronger association between somatic symptoms (pain, fatigue, and sleep disturbance) and depression. Additionally, we predict that there will be a greater effect of health literacy for analyses involving the PHQ-9 than the PROMIS. We will also compare standardized beta coefficients to determine the strength of the effect for each variable. As we anticipate few instances of missing data due to our computerized administration, listwise deletion will be used in the presence of missing data.

Meaning and Purpose Analyses

DIF Analysis. In addition to the lordif approach described above, we will conduct DIF analysis using the moderated nonlinear factor analysis (MNLFA) approach, which allows health literacy to be continuous, rather than categorical (high versus low). Our analyses will examine the impact of health literacy on DIF as well as possible moderation by administration type (pencil-and-paper versus talking touchsceen). The MNLFA enables the examination of both variables simultaneously. Our a priori hypothesis is that DIF will be observed across levels of health literacy for paper-and-pencil based testing. We further hypothesize that DIF will also be found for the talking touchscreen, but for fewer items. Mplus permits IRT analyses with constraints imposed such that each item slope is linearly related to external variables – in our case, this would be:

$$\text{item slope}_j = a_j + b1_j * \text{Health LiTT} + b2_j * \text{Administration Type} + b3_j * \text{Interaction}$$

where j is an index for the k items (8 items in this case).

We will use Mplus in conjunction with the aMNLFA[24] package of R. Following the MNLFA procedure outlined by Curran et al.,[25] we will test models allowing item parameters (intercept and slope) to vary as a function of

hypothesized DIF variable, one item at a time, in order to arrive at a final model. We will use the final model to generate item characteristics curves (ICCs, which will show the estimated probabilities of endorsement for each item across different levels of meaning and purpose. Among non-invariant items, we will consider differences (due to DIF variables) in probability endorsement of 10% to be potentially meaningful scientifically. Using the final model, we will also plot meaning and purpose factor scores by healthy literacy and administration type to understand the relationship among these variables. For these analyses, we will use all available baseline data collected through March of 2020 (N = 566 English speakers).

Longitudinal analyses of Meaning and Purpose Predicting Quality of Life. In addition to examining DIF by health literacy in the NIH Toolbox Meaning and Purpose measure (as described above), cross-lagged panel analyses will be used to examine whether meaning and purpose predicts health-related quality of life (i.e., anxiety, depression, fatigue, pain intensity and interference, physical function, sleep disturbance, and disruption of social roles and activities) over time across the three study assessments. We will also identify covariates including sociodemographics associated with of meaning and purpose (i.e., age, gender, financial strain, education, race/ethnicity, language). We hypothesize that greater meaning and purpose will prospectively predict greater quality of life at subsequent study assessments. We will only include participants who could have provided complete data before we ceased recruitment due to COVID-19. Thus, we will exclude participants who are 1) were enrolled on or after 16 September 2020, and 3) missing their 6-month assessment. In this way, data will not be systemically missing due to censoring from the COVID-19 pandemic (i.e., to be eligible, the person would have had sufficient time to complete all three assessments before the COVID-19 pandemic).

Exit Interview

Our exit interview serves multiple purposes. One purpose is to assess the testing experience for the participant. We hypothesize differences between the talking touchscreen and paper-based testing in terms of the testing experience. We will calculate Spearman correlations with 99% confidence intervals (CIs) for the three 5-point ordinal questions (see Exit Interview) and whether testing was talking touchscreen or not (coded 0/1). This nonparametric approach will determine which method of testing is more appropriate for people of low health literacy. CIs that do not contain zero will be regarded as statistically significant.

A second purpose of the exit interview is to gather data on how health literacy influences other areas of health. Specifically, we have included four questions about colorectal cancer screening to determine a person's level of adherence to published guidelines.

Meaning and purpose in life, defined as the pursuit of worthwhile goals and an accompanying sense of fulfilment and coherence in one's life,[26-30] is recognized as a fundamental facet of psychological wellbeing and important determinant of both physical and mental health. Individuals who report greater meaning and purpose are more motivated to engage in their healthcare and improve their overall health and wellbeing. Research demonstrates that individuals who report greater meaning and purpose also endorse greater self-efficacy, or perceived confidence in the ability to manage one's health, and are more likely to participate in preventive health services, including cholesterol screenings, colonoscopies, mammograms, Pap smears, and prostate exams.[31-33] In the current study, we will examine the impact of meaning and purpose in life and health-related quality of life (i.e., anxiety, depression, fatigue, pain intensity and interference, physical function, sleep disturbance, and disruption of social roles and activities) on adherence to national guidelines for colorectal cancer screening. Colorectal cancer screening was assessed using the National Cancer Institute Health Interview National Trends Survey (HINTS).[34] Adherence will be defined as receipt of screening within recommended timeframes according to guidelines from the United States Preventive Services Task Force (USPSTF) in effect during the study years. For colorectal cancer, men and women ages 50 to 75 years old will be classified as adherent if they report receipt of either a fecal occult blood test (FOBT) within 1 year, colonoscopy within 10 years, sigmoidoscopy within 5

years, or both a colonoscopy and sigmoidoscopy within 10 years. Correlations will be conducted in order to examine bivariate associations among study variables, including meaning and purpose, health-related quality of life, and other relevant factors (i.e., age, gender, financial strain, education, race/ethnicity, language, and medical comorbidities). Primary analyses will be conducted using logistic regression to examine the relationship between meaning and purpose and health-related quality of life with adherence to national colorectal cancer screening guidelines (0 = non-adherence, 1 = adherence) when adjusting for relevant factors.

Exploratory analyses will include the calculation of a non-adherence risk score. That is, based on meaning and purpose and other covariates, we will determine a cutoff from (receiver operating characteristic) ROC curves derived from the logistic regression. Then, we will cross validate the cutoff using a subset of the data set aside and not used in the modelling. In both the training and testing data, we will calculate sensitivity and specificity and other prediction indices as a means to better identify individuals who may be at risk for not adhering to screening guidelines.

Interim and Final Analyses: Sleep as a mediator between depression and chronic pain

The follow analyses are an interim analysis to be conducted for the 2020 meeting of the International Society for Quality of Life Research (ISOQOL). These analyses will be conducted twice, once using all available data before the ISOQOL 2020 conference (21-24 October 2020, if the presentation is accepted and if the conference takes place), and again when data collection for the parent study has been completed. Thus, there is a planned interim analysis and a planned final analysis. The following analyses will be conducted within participants completing measures in English.

Hypothesis 1: Depression predicts pain at 6-month follow-up, mediated by sleep disturbance at 3-month follow-up: We hypothesize depression will be related to ***more*** sleep problems at 3-month follow up, which in turn will lead to ***more*** pain at 6-months. Depression at baseline will be operationalized by the PROMIS depression scale within the 57-item Profile. Note that we made an *a priori decision* to use PROMIS rather than the PHQ-9 for depression because the PHQ-9 contains questions about sleep whereas the PROMIS depression scale does not. We will assess sleep at 3-month follow-up using the PROMIS sleep disturbance subscale of the 57-item Profile. We will use two different PROMIS Profile scales to assess pain at 6-month follow-up: pain intensity and pain interference. These hypotheses were formulated before any data were examined.

To carry out these analyses, we will use the *mediate* package in R for these analyses. The form of the mediation is depression at baseline ➔ worse sleep at 3mo ➔ more pain at 6mo. We will first create the regression models for depression predicting sleep, and sleep predicting pain. Sleep as a mediator will be tested in the *mediate* package based on the 95% bca confidence intervals.

Hypothesis 2: Pain predicts depression at 6-month follow-up, mediated by sleep disturbance at 3-month follow-up: To examine bidirectional causality, we will analyze the mediation from Hypothesis 1 in the opposite direction, specifically Pain at baseline ➔ worse sleep at 3mo ➔ more depression at 6mo. The same mediation procedures will be used, as in Hypothesis 1, with two different pain measures as the initial covariate: pain intensity and pain interference.

Hypothesis 3: Sex will moderate mediation relationships found in Hypotheses 1 & 2. Given that depression disproportionately affects females when compared to males, we plan to conduct moderated mediation for sex. We hypothesize sex will moderate the mediation, with female sex being association with stronger associations among depression and sleep, and sleep and pain.

Exploratory analyses. We anticipate that age influences depression, sleep, pain, but have no directional hypothesis for age. Thus, age will be included as an exploratory variable in the above-mentioned mediations.

Statistical Analysis Plan Addendum: Vaccine Confidence Analysis from phone-based substudy (section added 10 March 2021)

Objective and Background: To determine if health literacy mediates the relationship between race/ethnicity and vaccine confidence.

Reducing Assessment Barriers for patients with low literacy has been conducted during the COVID-19 global pandemic, necessitating changes in the study protocols with transition to phone-based recruitment and administration of tests. To understand the needs of people with low health literacy during the COVID-19 global pandemic, we added several additional survey tools, administered remotely by phone to participant; questions were read by a coordinator and participants made verbal responses. Participants completed the Vaccine Confidence Index (VCI)[35], in order to explore sentiments relating to vaccination confidence among a diverse sample population. Although not used in this analysis, participants also completed the PROMIS Profile-29,[36,37] the 8-item Meaning and Purpose Scale,[38,39]and a shortened version of the Epidemic-Pandemic Impacts Inventory,[40] a survey tool aimed at measuring social, economic, and housing impacts of the pandemic.

The COVID-19 global pandemic is the worst pandemic in over 100 years. Currently there are three highly efficacious vaccines with impressive safety data, but also initial observational data suggesting a high proportion of the eligible population reluctant to get the vaccine. There have been a few studies published exploring vaccine hesitancy in the setting of the COVID-19 pandemic. A French study published in Lancet Infectious Disease found individuals with lower education levels were less likely want COVID-19 vaccination.[41] In contrast, a study from an Italian sample of students identified many university students with COVID-19 vaccine hesitancy.[42] In Israel, those who perceived themselves to be at higher risk of severe infection had the greatest likelihood of vaccination, but healthcare workers with high educational attainment were no more likely than general public to accept vaccination.[43]

The literature exploring the relationship between overall vaccine hesitancy (not specific to COVID-19 vaccine) has also found conflicting directional relationships between vaccine hesitancy and educational status. Although some studies have found greater vaccine refusal or hesitancy in lower education groups,[44,45] others have not.[46] The relationship between health literacy and vaccine views is also not clear, with studies finding conflicting results.[47] Health literacy is a potential modifiable factor for vulnerable populations with uncertain vaccination views. We seek to explore the relationship between COVID-19 vaccine hesitancy and health literacy in a diverse participant sample of $N = 300$, who are a subset of the parent study.

Hypothesis: We hypothesize that health literacy is a mediator of the relationship between race/ethnicity and vaccine confidence, such that race predicts health literacy, which is in turn inversely associated with vaccine confidence. Using mediation analysis, we will determine how much of the effect of race/ethnicity on vaccine confidence is mediated by health literacy (i.e., the indirect effect) while also estimating the direct and total effect. A sample of 300 participants will be included in this baseline analysis exploring the relationship between race/ethnicity, health literacy (measured by Health LiTT[48] obtained at the baseline of the parent study), and the VCI which was administered at follow-up during a period of COVID-19 vaccine approval and early administration. This hypothesis was formulated before looking at any of the data from the phone-based protocol.

Measures:

Demographic Covariates

Demographic information was collected at the baseline interview. We collected data on age, sex/gender, race and ethnicity, primary language, and education level. The covariates/confounders assessed in our model include age, sex/gender, and education, chosen based on the conceptual model exploring the relationship between health literacy and various health outcomes.[49] We will report descriptive statistics of demographic variables. To ease interpretation, **Education** will be treated as a dichotomous variable with two categories: high school/GED level

education or less versus more than high school education. **Gender** will be treated as dichotomous variable (male/female), because no participants identified as transgender or non-binary in this sample (although our assessment procedures were designed to accommodate non-binary participants). **Age** (in years) will be treated as continuous variable.

Primary Covariate

We will focus on race and ethnicity as primary covariate in our mediation analysis exploring the relationship between race/ethnic status and vaccine confidence. **Race/ethnicity status** will be dummy coded in the following way using white (non-Hispanic) as the reference category:

- X1 =1 if Black/African American (Non-Hispanic); X1=0 otherwise
- X2 = 1 if Latin X/Hispanic Ethnicity; X2=0 otherwise
- X3 = 1 if Other Race/Ethnicity (Non-white); X3 =0 otherwise

The other race category will include participants of mixed race and individuals who identified as Asian, Pacific Islander, or American Indian, categories with fewer participants.

Health literacy: Hypothesized Mediator

Health literacy was measured at baseline using Health Literacy Assessment Using Talking Touchscreen Technology (Health LiTT),[50] a computerized assessment that does not require an interviewer. This measure has been validated in English and Spanish. We will define **Health Literacy** as a continuous T score obtained from Health LiTT where a higher score is associated with greater degree of health literacy.

Vaccine Confidence Index: Outcome

The **Vaccine Confidence Index (VCI)** is a tool developed in 2015 that has been used in global samples to measure sentiments correlated with individual confidence around vaccination.[35,51] The VCI is measured as a ratio of scores on questions that suggest confidence in vaccines divided by scores that indicate suspicion and hesitancy toward vaccination. A larger numerical total score and ratio greater than 1.0 on VCI reflects a greater confidence in pursuing vaccination **Vaccine Confidence Index** (VCI) will be treated as a continuous variable in the model, with higher numerical values being associated with greater degrees of vaccine confidence.

Primary analysis: Exploring the relationship between vaccine confidence and race/ethnicity status during the COVID-19 pandemic

The analysis plan was formulated and written before any data were analyzed from the phone-based assessment of VCI. We may conduct additional analyses beyond what is described in this plan, but these additional analyses be considered exploratory. We hypothesize that racial or ethnic minority status will be associated with lower levels of vaccine confidence as measured with VCI, controlling for gender, age, and education level. In a theoretical causal pathway for the relationship between race/ethnic demographic status and vaccine confidence we consider race/ethnic minority status as the covariate of interest, and health literacy as a partial mediator of the relationship between race/ethnicity and vaccine confidence. We will use the *lavaan* package in R using regression analyses to compare the estimates of individual, total, direct, and indirect effect estimates for the variables in the model and calculated estimates based on 95% bootstrap confidence intervals for the indirect effect (i.e., X1(race) ➔ lower Health LiTT ➔ lower VCI.

Limitations and Potential sources of Bias

This analysis will have several limitations. 1. The sample may not be generalizable as it was restricted to urban and English- or Spanish-speaking participants who consented to participation in a research study. 2. This is a baseline analysis that does not take into the account the potential for changing views on vaccination in the setting of an ongoing pandemic and public health education campaigns. We plan longitudinal analyses in the future to

address this limitation. 3. Although we have hypothesized a causal pathway that includes race/ethnicity and health literacy as a partial mediator, we are unable to evaluate alternative pathways for factors that were not measured as part of this study. 4. Mediator-outcome confounding or collider bias is a possibility in this analysis in reference to the variable, educational level, which is related to health literacy and potentially related to the outcome of vaccine confidence. 5. In addition, the possibility of measurement error in the assessment of health literacy not only affects the estimate of health literacy's relationship with vaccine confidence, but also the estimate of the relationship of race or ethnicity with VCI.

Despite these limitations, we feel this analysis adds to the available knowledge on vaccine confidence during the COVID-19 pandemic and opportunities to improve public uptake of vaccination through targeted campaigns and interventions.

Cognitive Correlates of Rumination

Version notes. This plan was added in Version 4.2, 02 June 2021. At the time of the plan, adjust MOCA-B scores had been examined for distributional characteristics and differential item functioning for depression items (DIF; i.e., different measurement properties for PHQ-9 and PROMIS Depression by high-low MOCA-B). Some other analyses of rumination and depression had also been carried out as described in the above section *Prediction of depression: Health literacy as a moderator of rumination, pain, fatigue, and sleep disturbance*.

Sample. The sample will contain all participants (N = 478) who completed cognitive measures (see below) recruited into the parent health literacy study before recruitment was stopped due to the COVID-19 pandemic (March 2021). The cognitive measures were administered at one of three time points in the study; corresponding rumination and PROMIS depression measures will be used from the same time point.

Overarching goal

We will determine the relationship between multiple facets of cognition (represented as latent factors) and rumination (divided into brooding and reflection). We hypothesize that higher levels of rumination will be associated with lower levels of cognition.

Descriptive statistics

Our brief battery contains the Montreal Cognitive Assessment-Basic (MOCA-B),[21] the SPMSQ,[52] selected items from the MOCA 8.1, and selected items from the RUDAS.[22] Rumination will be assessed using the 10 items from the RRS. Descriptive analyses of the cognitive measures will include frequency tables for each item (e.g., number correct/incorrect on a per variable basis). Items that are sparse (i.e., with fewer than 5 responses) will be collapsed, or if this is not possible the item will be dropped because of insufficient variability (note: for the dichotomous cognitive items, they are mostly binary, which will necessitate dropping sparse items). We will also inspect percentages of missing data; we do not anticipate the need to drop any variables with too much missingness, but a post-hoc decision may be needed if there are unanticipated levels of missingness. We plan on using the tableone package of R for descriptive statistics.[53,54]

Determining the number of factors

We will first conduct analyses to determine the number of exploratory cognitive factors to extract, which later will be used in *exploratory structural equation modelling (ESEM)*[55] to link the cognitive factors to rumination (brooding and reflection). Because all of the items in the cognitive measures are categorical, we will use the method of Lubbe (2019)[56] to determine the number of factors, which compares observed eigenvalues to simulated eigenvalues with a correction for item distributions (i.e., based on the proportions of response type within each category). Within the approach of Lubbe, we will use 1000 replications, seeking eigenvalues that exceed the 99th percentile of the simulated distribution. If needed, we will supplement this approach with the nFactors and psych packages of R. Factors with fewer than three items with loadings ≥ .4 will not be retained. We are using an exploratory approach to forming the factors, but we will be guided by research on the known structure of cognition in extracting and naming the factors (e.g., "memory" vs. ""executive functioning").

Mean centering. Because we plan to test interaction effects (i.e., moderation), we will mean-center brooding, reflection, PROMIS depression, age, and Health LiTT. If any additional continuous covariates are used in exploratory analyses, these will also be centered around the mean.

Exploratory Structural Equation Modelling

All items in the cognition measures (as latent variables) and brooding and reflection (as scores) will be included (see above). Based on the parallel analysis to determine the number of factors, multiple exploratory factors from

the cognitive measures will be regressed on our two covariates of interest – brooding and reflection. We anticipate that brooding and reflection will be inversely related with cognition, in particular items related to memory and executive function. In other words, higher levels of rumination will be associated with lower levels of cognition.

We will conduct these analyses using Mplus. We will report all fit indices, unstandardized associations among variables, corresponding SEs, and significance tests at a Type I error rate of .05.

Scoring rumination

Rumination will be measured using the Ruminative Responses Scale (RRS) short form, a 10-item form that includes two subscales, brooding and reflection. The RRS uses a four-point response scale for each item with a total score derived from the sum of the individual items. Higher scores indicate higher levels of rumination. All variables will be examined in a continuous manner, rather than using cut scores. If 80% or more of the items are present, we will score the subscale using prorating.

Hypotheses and testable predictions

We hypothesize that rumination will have a significant association with cognition. Specifically, higher scores in both domains of rumination (brooding and reflection) will be related to lower scores on cognitive factors. Overall, we expect cognition to be more strongly related to brooding than reflection. In past work (Griffith & Raes, 2014), we found that the brooding factor was more robust than the reflection factor. Thus, it is possible that the more reliable brooding scale will have larger associations with cognition. However, the literature contains mixed findings with regard to the relative contributions of brooding versus reflection (e.g.,[57,58].

In our model, our cognition factors will be the dependent variables and our rumination factors will be the independent variables. These analyses will be associative; we do not seek to make inferences about causality.

Missing data. Participants who did not complete the cognition measures will not be included in these analyses. For individual missing items and/or tests, the default options of Mplus (i.e., WLSMV estimator) will be used to carrying out the analyses, which will include all cases.

Supplementary analyses

Alternative approaches to deal with model identification. If we have difficultly identifying the ESEM model (see above), we will create factor scores for the cognitive variables using EFA and explore them with multiple regression analyses with the *lm()* function in base R to test the association between cognition and rumination. For each of the anticipated cognitive factors we will test associations between cognitive factors and rumination. We will create separate regression models for each of our three anticipated cognition factors X each of the two subscales of rumination (brooding and reflection), thus, we will have a total of 3 (α = .05) models (depending on EFA), starting with two subscales.

Additional covariates. Given that cognition is known to be related to other variables, such as age, we will include covariates to see if rumination is significant above and beyond these covariates. These additional variables will include age, gender, Health LiTT (continuous), language (English vs. Spanish), depressive symptoms, and survey delivery method ("talking touchscreen" vs. paper-and-pencil). If Health LiTT and survey delivery method are not significant, they will be deleted from the model and not considered further.

Moderation analysis. We will test depression as a moderator of the associations between cognition and brooding and reflection. A product term will be created depression, reflection, and brooding. Before creating the product term, each variable will be mean-centered (see above). We will then regress the cognitive factors

(dependent variables) on brooding, reflection, depression, and the two product terms (depression × brooding, depression × reflection); depression will be the moderator. The statistical significance of these two product terms will be evaluated at .05 and effect sizes will be reported. We will then conduct addition moderation analyses with age, gender, and language in the same fashion as described above. Depending on the results, moderation will be summarized using simple-slope analyses and visualization.

Other Approaches (Exploratory)

We will explore these associations in other software (e.g., lavaan, mirt)[59,60] for consistency across software. Supplementary results will be made available in an open-science repository.

Backup plan for factor analyses. If we have difficulty identifying latent variable models, we will reframe the analyses of cognition around general levels of cognition using the MOCA-B total score.

Other Interim and Exploratory Analyses

Differential Item Functioning of Depression by Cognitive Ability (MoCA-B)

With funding from a National Institute on Aging (NIA) Supplement, we have collected data on the Montreal Cognitive Assessment Test – Basic (MoCA-B). Data from the current project provide an opportunity for preliminary data for an upcoming grant proposal on the relationship between cognitive function and health-related quality of life. Thus, we will conduct DIF analyses of the depression scale of the PROMIS Profile and the PHQ-9 using high vs. low scorers on the MOCA-B as determined by a median split. The procedures of the DIF analyses will be the same as described above. We will also conduct visual analysis of the relationship between the MoCA-B and depression using the *ggplot2* package of R, as well as quantify the associations among these scales using Pearson correlations along with 95% CIs.

These analyses will be interim, before all data have been collected, as they will be needed for a grant submission in June of 2020. The results of this interim analysis will not affect any planned hypotheses presented elsewhere in this plan. The data analyzed for this interim analysis will be all data collected before 01 March 2020.

REFERENCES CITED

1. Affairs HlrotCoS. Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. JAMA. 1999;281(6):552-7.
2. Health UDo, Services H. Healthy People 2010. Washington, DC: US Government Printing Office. Originally developed for Ratzan SC, Parker RM. 2000. Introduction. National Library of Medicine Current Bibliographies in Medicine: Health Literacy Selden CR, Zorn M, Ratzan SC, Parker RM, Editors NLM Pub No CBM. 2000;1.
3. Parker RM, Ratzan SC, Lurie N. Health Literacy: A Policy Challenge For Advancing High-Quality Health Care. Health Affair. 2003;22(4):147-53. doi: 10.1377/hlthaff.22.4.147.
4. Parker R, Ratzan SC. Health Literacy: A Second Decade of Distinction for Americans. Journal of Health Communication. 2010;15(sup2):20-33. doi: 10.1080/10810730.2010.501094.
5. Ratzan SC, Parker RM. Health Literacy—Identification and Response. Journal of Health Communication. 2006;11(8):713-5. doi: 10.1080/10810730601031090.
6. Parker RM, Gazmararian JA. Health Literacy: Essential for Health Communication. Journal of Health Communication. 2003;8(sup1):116-8. doi: 10.1080/713851963.
7. Cella D, Reeve B, Amtmann D, Hays R, Gershon R, editors. The patient reported outcomes management information system (PROMIS); a cooperative effort to advance and enable clinical research. International Society for Quality of Life Research; 2006 October 10-14; Lisbon, Portugal.
8. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, Dewalt D, Fries JF, Gershon R, Hahn EA, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R, Lai JS, PROMIS Cooperative Group. The Patient Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. J Clin Epidemiol. 2010;63(11):1179-94. doi: S0895-4356(10)00173-3 [pii]10.1016/j.jclinepi.2010.04.011. PubMed PMID: 20685078.
9. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap Cooperative Group During its First Two Years. Med Care. 2007;45(5 Suppl 1):S3-S11. doi: 10.1097/01.mlr.0000258615.42478.55. PubMed PMID: 51345; PMCID: PMCID: PMC2829758.
10. Fries JF, Bruce B, Cella D. The promise of PROMIS: Using item response theory to improve assessment of patient-reported outcomes. Clinical and Experimental Rheumatology. 2005;23(S38):S33-S7. PubMed PMID: 10811.
11. Hays RD, Bjorner J, Revicki DA, Spritzer K, Cella D. Development of physical and mental health summary scores from the Patient Reported Outcomes Measurement Information System (PROMIS) global items. Quality of Life Research. 2009;18(7):873-80. PubMed PMID: 52412.
12. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, Hays RD. Representativeness of the PROMIS Internet Panel. J Clin Epidemiol. 2010;63(11):1169-78. PubMed PMID: 52416.
13. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D. Psychometric Evaluation and Calibration of Health-Related Quality of Life Item Banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). Med Care. 2007;45(5 Suppl 1):S22-S31. PubMed PMID: 51348.
14. U.S. Department of Education NCfES. 1992 National Adult Literacy Survey (NALS) and 2003 National Assessment of Adult Literacy (NAAL), A First Look at the Literacy of America's Adults in the 21st Century; and supplemental data retrieved July 6, 2006, from http://nces.ed.gov/naal/Excel/2006470_DataTable.xls. U.S. Department of Education, National Center for Education Statistics, 2006 July 2006. Report No.
15. Hahn EA, Choi SW, Griffith JW, Yost KJ, Baker DW. Health Literacy Assessment Using Talking Touchscreen Technology (Health LiTT): A New Item Response Theory-Based Measure of Health Literacy. Journal of Health Communication. 2011;16:150-62. doi: Doi 10.1080/10810730.2011.605434. PubMed PMID:

WOS:000299952500014.
16. Griffith JW, Kleim B, Sumner JA, Ehlers A. The factor structure of the Autobiographical Memory Test in recent trauma survivors. Psychol Assessment. 2012;24(3):640.
17. Griffith JW, Raes F. Factor Structure of the Ruminative Responses Scale. Eur J Psychol Assess. 2014. doi: 10.1027/1015-5759/a000231.
18. Griffith JW, Sumner JA, Debeer E, Raes F, Hermans D, Mineka S, Zinbarg RE, Craske MG. An item response theory/confirmatory factor analysis of the Autobiographical Memory Test. Memory. 2009;17(6):609-23. doi: Pii 912937501
Doi 10.1080/09658210902939348. PubMed PMID: WOS:000267767200001.
19. Kupst MJ, Butt Z, Stoney CM, Griffith JW, Salsman JM, Folkman S, Cella D. Assessment of stress and self-efficacy for the NIH toolbox for neurological and behavioral function. Anxiety, Stress, & Coping. 2014(just-accepted):1-28.
20. Julayanont P, Tangwongchai S, Hemrungrojn S, Tunvirachaisakul C, Phanthumchinda K, Hongsawat J, Suwichanarakul P, Thanasirorat S, Nasreddine ZS. The montreal cognitive assessment—basic: a screening tool for mild cognitive impairment in illiterate and low-educated elderly adults. J Am Geriatr Soc. 2015;63(12):2550-4.
21. Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. J Am Geriatr Soc. 2005;53(4):695-9.
22. Storey JE, Rowland JT, Conforti DA, Dickson HG. The Rowland universal dementia assessment scale (RUDAS): a multicultural cognitive assessment scale. Int Psychogeriatr. 2004;16(1):13-31.
23. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques: DIFdetect and difwithpar. Med Care. 2006;44(11):S115-S23.
24. Gottfredson NC, Cole VT, Giordano ML, Bauer DJ, Hussong AM, Ennett ST. Simplifying the implementation of modern scale scoring methods with an automated R package: Automated moderated nonlinear factor analysis (aMNLFA). Addict Behav. 2019;94:65-73.
25. Curran PJ, McGinley JS, Bauer DJ, Hussong AM, Burns A, Chassin L, Sher K, Zucker R. A moderated nonlinear factor model for the development of commensurate measures in integrative data analysis. Multivariate behavioral research. 2014;49(3):214-31.
26. McKnight PE, Kashdan TB. Purpose in life as a system that creates and sustains health and well-being: An integrative, testable theory. Review of General Psychology. 2009;13(3):242-51.
27. Ryff CD. Psychological well-being revisited: Advances in the science and practice of eudaimonia. Psychother Psychosom. 2014;83(1):10-28.
28. Frankl VE. Man's search for meaning: Simon and Schuster; 1985.
29. Park CL. Making sense of the meaning literature: an integrative review of meaning making and its effects on adjustment to stressful life events. Psychol Bull. 2010;136(2):257.
30. Reker GT. Theoretical perspective, dimensions, and measurement of existential meaning. Exploring existential meaning: Optimizing human development across the life span. 2000:39-55.
31. Kim ES, Strecher VJ, Ryff CD. Purpose in life and use of preventive health care services. Proceedings of the National Academy of Sciences. 2014;111(46):16331-6.
32. Wells JNB, Bush HA. Purpose-in-life and breast health behavior in Hispanic and Anglo women. Journal of Holistic Nursing. 2002;20(3):232-49.
33. Holahan CK, Suzuki R. Motivational factors in health promoting behavior in later aging. Activities, Adaptation & Aging. 2006;30(1):47-60.
34. Nelson D, Kreps G, Hesse B, Croyle R, Willis G, Arora N, Rimer B, Vish Viswanath K, Weinstein N, Alden S. The health information national trends survey (HINTS): development, design, and dissemination. Journal of health communication. 2004;9(5):443-60.
35. Larson HJ, Schulz WS, Tucker JD, Smith DM. Measuring vaccine confidence: introducing a global vaccine

confidence index. PLoS Curr. 2015;7. Epub 2015/03/20. doi: 10.1371/currents.outbreaks.ce0f6177bc97332602a8e3fe7d7f7cc4. PubMed PMID: 25789200; PMCID: PMC4353663.
36. Fischer F, Gibbons C, Coste J, Valderas JM, Rose M, Leplège A. Measurement invariance and general population reference values of the PROMIS Profile 29 in the UK, France, and Germany. Qual Life Res. 2018;27(4):999-1014. Epub 2018/01/20. doi: 10.1007/s11136-018-1785-8. PubMed PMID: 29350345.
37. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. Medical care. 2007;45(5 Suppl 1):S3-s11. Epub 2007/04/20. doi: 10.1097/01.mlr.0000258615.42478.55. PubMed PMID: 17443116; PMCID: PMC2829758.
38. Salsman JM, Lai JS, Hendrie HC, Butt Z, Zill N, Pilkonis PA, Peterson C, Stoney CM, Brouwers P, Cella D. Assessing psychological well-being: self-report instruments for the NIH Toolbox. Qual Life Res. 2014;23(1):205-15. Epub 2013/06/19. doi: 10.1007/s11136-013-0452-3. PubMed PMID: 23771709; PMCID: PMC3883907.
39. Salsman JM, Butt Z, Pilkonis PA, Cyranowski JM, Zill N, Hendrie HC, Kupst MJ, Kelly MA, Bode RK, Choi SW, Lai JS, Griffith JW, Stoney CM, Brouwers P, Knox SS, Cella D. Emotion assessment using the NIH Toolbox. Neurology. 2013;80(11 Suppl 3):S76-86. Epub 2013/04/23. doi: 10.1212/WNL.0b013e3182872e11. PubMed PMID: 23479549; PMCID: PMC3662334.
40. Grasso DJ, Briggs-Gowan MJ, Ford JD, Carter AS. The Epidemic-Pandemic Impacts Inventory (EPII). In: Medicine. UoCSo, editor. 2020.
41. Coconel Group. A future vaccination campaign against COVID-19 at risk of vaccine hesitancy and politicisation. The Lancet Infectious diseases. 2020;20(7). doi: 10.1016/S1473-3099(20)30426-6. PubMed PMID: 32445713.
42. Barello S, Nania T, Dellafiore F, Graffigna G, Caruso R. 'Vaccine hesitancy' among university students in Italy during the COVID-19 pandemic. Eur J Epidemiol. 2020;35(8):781-3. Epub 2020/08/08. doi: 10.1007/s10654-020-00670-z. PubMed PMID: 32761440; PMCID: PMC7409616.
43. Dror AA, Eisenbach N, Taiber S, Morozov NG, Mizrachi M, Zigron A, Srouji S, Sela E. Vaccine hesitancy: the next challenge in the fight against COVID-19. Eur J Epidemiol. 2020;35(8):775-9. Epub 2020/08/14. doi: 10.1007/s10654-020-00671-y. PubMed PMID: 32785815.
44. Facciolà A, Visalli G, Orlando A, Bertuccio MP, Spataro P, Squeri R, Picerno I, Di Pietro A. Vaccine hesitancy: An overview on parents' opinions about vaccination and possible reasons of vaccine refusal. J Public Health Res. 2019;8(1):1436. Epub 2019/04/19. doi: 10.4081/jphr.2019.1436. PubMed PMID: 30997357; PMCID: PMC6444379.
45. Wagner AL, Shotwell AR, Boulton ML, Carlson BF, Mathew JL. Demographics of Vaccine Hesitancy in Chandigarh, India. Front Med (Lausanne). 2020;7:585579. Epub 2021/02/02. doi: 10.3389/fmed.2020.585579. PubMed PMID: 33521011; PMCID: PMC7844137.
46. Wagner AL, Masters NB, Domek GJ, Mathew JL, Sun X, Asturias EJ, Ren J, Huang Z, Contreras-Roldan IL, Gebremeskel B, Boulton ML. Comparisons of Vaccine Hesitancy across Five Low- and Middle-Income Countries. Vaccines (Basel). 2019;7(4). Epub 2019/10/23. doi: 10.3390/vaccines7040155. PubMed PMID: 31635270; PMCID: PMC6963484.
47. Lorini C, Santomauro F, Donzellini M, Capecchi L, Bechini A, Boccalini S, Bonanni P, Bonaccorsi G. Health literacy and vaccination: A systematic review. Hum Vaccin Immunother. 2018;14(2):478-88. Epub 2017/10/20. doi: 10.1080/21645515.2017.1392423. PubMed PMID: 29048987; PMCID: PMC5806657.
48. Hahn EA, Choi SW, Griffith JW, Yost KJ, Baker DW. Health literacy assessment using talking touchscreen technology (Health LiTT): a new item response theory-based measure of health literacy. Journal of health communication. 2011;16 Suppl 3(Suppl 3):150-62. Epub 2011/09/29. doi: 10.1080/10810730.2011.605434. PubMed PMID: 21951249; PMCID: PMC3269120.
49. Paasche-Orlow MK, Wolf MS. The causal pathways linking health literacy to health outcomes. Am J

Health Behav. 2007;31 Suppl 1:S19-26. Epub 2007/10/20. doi: 10.5555/ajhb.2007.31.supp.S19. PubMed PMID: 17931132.
50. Davidson PJ, Ochoa KC, Hahn JA, Evans JL, Moss AR. Witnessing heroin-related overdoses: the experiences of young injectors in San Francisco. Addiction. 2002;97(12):1511-6. Epub 2002/12/11. PubMed PMID: 12472634.
51. Larson HJ, de Figueiredo A, Xiahong Z, Schulz WS, Verger P, Johnston IG, Cook AR, Jones NS. The State of Vaccine Confidence 2016: Global Insights Through a 67-Country Survey. EBioMedicine. 2016;12:295-301. Epub 2016/10/26. doi: 10.1016/j.ebiom.2016.08.042. PubMed PMID: 27658738; PMCID: PMC5078590.
52. Pfeiffer E. A short portable mental status questionnaire for the assessment of organic brain deficit in elderly patients. J Am Geriatr Soc. 1975;23(10):433-41. Epub 1975/10/01. PubMed PMID: 1159263.
53. Team RC. R-project. Vienna, Austria: R Foundation for Statistical Computing; 20162018.
54. Yoshida K, Bohn J, Yoshida MK. Package 'tableone'. R Foundation for Statistical Computing, Vienna, Austria (30 November 2016). 2020.
55. Asparouhov T, Muthen B. Exploratory Structural Equation Modeling. Struct Equ Modeling. 2009;16(3):397-438. doi: Doi 10.1080/10705510903008204. PubMed PMID: WOS:000270094800001.
56. Lubbe D. Parallel analysis with categorical variables: Impact of category probability proportions on dimensionality assessment accuracy. Psychological methods. 2019;24(3):339.
57. Satyshur MD, Layden EA, Gowins JR, Buchanan A, Gollan JK. Functional connectivity of reflective and brooding rumination in depressed and healthy women. Cognitive, Affective, & Behavioral Neuroscience. 2018;18(5):884-901. doi: 10.3758/s13415-018-0611-7.
58. Joormann J, Nee DE, Berman MG, Jonides J, Gotlib IH. Interference resolution in major depression. Cognitive, Affective, & Behavioral Neuroscience. 2010;10(1):21-33.
59. Chalmers RP. mirt: A multidimensional item response theory package for the R environment. J Stat Softw. 2012;48(6):1-29.
60. Rosseel Y. Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). J Stat Softw. 2012;48(2):1-36.