Updated: June 15, 2020

This article is the Author's Accepted Manuscript version accepted by *Journal of Nursing Measurement*, a Spring Publishing journal. This Accepted Manuscript is shared under a CC-BY-NC-ND license and should not be used as a substitute for the final published version of the journal. When citing this article, please use the following:

James, T. G., Miller, M. D., Nicolette, G., & Cheong, J. (In press.) Psychometric evaluation of a 10-item health insurance knowledge scale. *Journal of Nursing Measurement*.

An updated version of this document will be posted once a DOI is assigned to the publication.

Psychometric Evaluation of a 10-Item Health Insurance Knowledge Scale

Tyler G. James, MS, CHES[1], M. David Miller, PhD[2], Guy Nicolette, MD[3,4], and JeeWon Cheong, PhD[1]

[1] Department of Health Education and Behavior, University of Florida, Gainesville, FL.

[2] Department of Research and Evaluation Methodology, University of Florida, Gainesville, FL.

[3] (former) Student Health Care Center, Department of Community Health and Family Medicine, University of Florida, Gainesville, FL.

[4] University Health Services, University of California – Berkeley, Berkeley, CA.

Author Notes

Tyler G. James, ORCID: 0000-0002-0694-4702, tjames95@ufl.edu

M. David Miller, dmiller@coe.ufl.edu

Guy Nicolette, gnicolette@berkeley.edu

JeeWon Cheong, ORCID: 0000-0001-6033-7180, jwcheong@ufl.edu

Correspondence concerning this article should be addressed to Tyler G. James, Department of Health Education and Behavior, PO Box 118210, Florida Gym Room 5, University of Florida, Gainesville, Florida, 32611. Email: tjames95@ufl.edu

Author's Accepted Manuscript

Disclosure: Parts of this study were presented as a student poster at the American Public Health Association 2019 Annual Meeting.

Abstract

Background: College students are a priority population for health insurance literacy interventions. Yet, there are few psychometric studies on measuring health insurance knowledge – a core construct of health insurance literacy.
Methods: We administered a health insurance survey to 2,250 college students. We applied Classical Test Theory and Item Response Theory methods to estimate psychometric properties of the Kaiser Family Foundation's 10-item health insurance knowledge quiz.
Results: The scale is unidimensional, and data best fit a two-parameter logistic model. IRT estimates indicated varying item discriminations (*a* range: 0.717 to 2.578) and difficulties (*b* range: -0.913 to 1.790). Precision of measurement was maximized for students half a standard deviation below the mean ($\theta = -0.686$) health insurance knowledge ability.
Conclusions: This scale can be used to identify gaps in health insurance knowledge among college students and be applied in clinical and community health education practice.

Author's Accepted Manuscript

Patients in the United States who have health insurance still report experiencing financial barriers to accessing healthcare (Allen, Call, Beebe, McAlpine, & Johnson, 2017; Vujicic, Buchmueller, & Klein, 2016). One contributing financial barrier is the complexity of the U.S. healthcare system which may cause patients to have limited knowledge and skills to use health insurance plans; that is, patients likely have limited health insurance literacy. Health insurance literacy is defined as a person's knowledge, ability, and confidence to find, use, and choose health insurance plans (Quincy, 2012). Among the U.S. adult population, health insurance literacy is low due to lower general health literacy (Berkman, Sheridan, Donahue, Halpern, & Crotty, 2011), disparities in understanding health insurance terms (Norton, Hamel, & Brodie, 2014), and difficulty with estimating costs of using healthcare (Paez et al., 2014). Further, there are specific sub-populations in the U.S. which may be at higher risk. College students (age 18 to 24) are a particularly important population when considering health insurance literacy. Provisions in the Patient Protection and Affordable Care Act allow college students to remain dependents on their parents'/guardians' health insurance plans until age 26; thus, it is likely they have infrequent exposure to health insurance terminology and little (if any) experience with choosing their own health insurance. Further, recent research indicates that health insurance literacy is inadequate for college student populations and that this has implications for healthcare utilization, which calls for action to increase college student health insurance literacy (James et al., in press; Nobles, Curtis, Ngo, Vardell, & Holstege, 2018; Yang, 2016). However, despite various available scales assessing health insurance literacy, psychometric properties of those scales have rarely been examined. Psychometric studies of the measurement properties of health literacy scales will enable health education specialists and healthcare providers to accurately

Author's Accepted Manuscript

assess health insurance literacy and identify health education needs when developing programs to improve health insurance literacy and utilization.

Background: Measurement of Health Insurance Literacy

To date, health education specialists and researchers have developed their own tools or used one of three measures available. The 42-item *Health Insurance Literacy Measure* (HILM; Paez et al., 2014) has the strongest evidence of comprehensive and systematic development leading to strong evidence of validity: scale content was developed based on literature review, key informant interviews, and stakeholder meetings, there is evidence of validity based on response processes (e.g., testing to ensure participants adequately understand the scale's items), and its internal structure was supported by exploratory and confirmatory factor analysis and Rasch analyses. The HILM conceptualizes health insurance literacy as including health insurance knowledge, information seeking, document literacy, and cognitive skills, but operationalizes this construct as a self-efficacy measure (i.e., ability to choose and use health insurance; Paez et al., 2014); thus, there is no separate measurement of knowledge and self-efficacy. However, health behavior theory (e.g., Social Cognitive Theory) would posit that personal cognitive factors – such as self-efficacy and knowledge – represent separate constructs that should be separately measured (Bandura, 2004; Glanz, Rimer, & Viswanath, 2015); this should also apply to the measurement of a multidimensional construct such as health insurance literacy.

To our knowledge, there are two published health insurance knowledge measures available in the literature. One tool is a 10-item quiz developed by the Kaiser Family Foundation (KFF; a non-profit organization focusing on health issues in the U.S.) which measures cost-calculation knowledge and familiarity with common insurance terms (Norton et al., 2014). This tool was developed by health insurance experts at KFF and administered to a nationally

Author's Accepted Manuscript

representative sample of U.S. Americans to describe health insurance knowledge at a population level. A second, 20-item, tool was developed by Nobles and colleagues (2018) to measure health insurance vocabulary knowledge among college students. Nobles and colleagues' (2018) tool was developed by aligning knowledge constructs to recommendations from HealthCare.gov, then reviewed by experts prior to a pilot test. Neither the KFF nor Nobles' scale has been subject to in-depth psychometric study, with the exception of review by experts and the survey study's population prior to administration (James et al., in press; Nobles et al., 2018).

When considering health insurance literacy, we must accurately measure both self-efficacy and knowledge; this requires providing substantial evidence of the measurement structure of tools we use to gather data. The HILM has multiple studies describing the psychometric nature of its measurement of health insurance self-efficacy. However, we lack measurement studies focusing on health insurance knowledge scales: neither the knowledge scale developed by Nobles and colleagues (2018) nor the KFF have published validity studies. Thus, the goal of this study was to conduct a psychometric investigation of the internal structure of the KFF's 10-item health insurance knowledge quiz among college students by applying classical and modern test theory techniques to provide information from item response data on the reliability, dimensionality, and precision of scores.

Methods

Procedure

Upon receiving Institutional Review Board approval, we received a random sample of 10,000 students attending a large university in the southeast U.S. (student population approximately 52,000) to take a survey about health insurance. Inclusion criteria for the sample were that students must be aged 18 years or older and be attending classes on-campus. Students

Author's Accepted Manuscript

were recruited in March 2017 via email which described the purpose of the survey, estimated time to complete the survey, and the incentive (i.e., the first, middle, and last 10 respondents received a $20 gift card). A total of 2,430 students started the survey (24.3% response rate). The majority of students were white (64.0%), female (61.3%), undergraduate (66.0%), and insured on their parents'/guardians' health insurance plan (61.3%). In comparison, the sampled university is majority white (53.1%), female (55.1%), and undergraduate (65.2%). Although demographic characteristics differ between the institution and the present study, our sample is more diverse in gender, race, and student status than those of other health surveys administered at this university using similar sampling and incentive strategies (Emmereé, James, Varnes, & Arceneaux, 2015; James, Emmereé, & Brantley, 2015; James & Ryan, 2018). In the current analysis, we performed listwise deletion and include 2,250 students who answered all items of the KFF quiz (i.e., 92.6% of respondents). Of the 180 participants excluded due to missing responses on the KFF quiz, the vast majority (86%) also had missing values on race, gender, and undergraduate status. However, those excluded did not differ from the analytic sample in the proportion of being insured (66.0% of the analytic sample and 62.4% of the excluded sample).

Measure

The measure of health insurance knowledge used in this study is the Kaiser Family Foundation's (KFF) *Health Insurance Knowledge Quiz*. At the time of data collection, the KFF tool was the only health insurance knowledge scale available, as Nobles and colleagues' findings were not published until 2018.

The KFF quiz was developed by health care experts at KFF to assess U.S. adults' familiarity with health insurance terms (Norton et al., 2014; see Table 1). The 10-item measure includes true/false and three- to five-option multiple choice questions; the quiz is scored by

Author's Accepted Manuscript

summing the number of correct responses (range: 0 to 10). During the quiz's development, it was administered to a sample of 1,292 adults (age 18 and older) living in the U.S. The items were developed based on content domains identified by KFF experts on the health insurance marketplace and polling teams, in addition to being reviewed by KFF's contracted data collection firm, Growth from Knowledge. Items cover common health insurance topics including health insurance premiums and deductibles (e.g., Item 1: Which of the following is the best definition of the term "health insurance premium"?).

Prior to administering the present study's survey, a pool of experts also evaluated the survey. The expert pool included seven individuals: two students in the study sampling frame; a Certified and Master Certified Health Education Specialist (CHES®/MCHES®) with a collective 14 years in college health promotion and survey development; a student health center (SHC) finance administrator who is a Certified Professional Coder and Certified Health Data Analyst; a sports medicine board certified physician and (then) director of the SHC at the sampled university; and a PhD-level scientist trained in social and quantitative psychology. Minor changes were made to the survey, but no changes to the KFF scale citing that it adequately covered the content domain of health insurance knowledge. Further evidence of sufficient construct representation was based on review of the KFF scale in comparison to the SHC's health insurance education materials; the 10-items had almost perfect overlap, with the exception of minor wording or calculation details.

Data Analysis

Our analysis took place through two frameworks: Classical Test Theory (CTT) and modern test theory, known as Item Response Theory (IRT). CTT provides crucial information about the performance of an overall test, and insights for IRT model selection. We first estimated

Author's Accepted Manuscript

item difficulties (i.e., proportion responding correct to each item) and corrected item-total correlations (i.e., discriminations) using the package *CTT* (Willse, 2018) in the R statistical environment. We then confirmed dimensionality of the knowledge quiz by estimating a one-factor categorical CFA in M*plus* 8.0 (Muthén & Muthén, 2011). We assessed model fit using multiple fit indices including χ^2 test, root mean square approximation (RMSEA), the Comparative Fit Index (CFI), and Tucker Lewis Index (TLI). Acceptable model fit was specified as RMSEA ≤ 0.06, and CFI and TLI > 0.95 (Hu & Bentler, 1999). In addition to the IRT assumption of unidimensionality, we tested the assumption of local independence using the approximately standardized chi-square index (Chen & Thissen, 1997) using the R package *mirt* (Chalmers, 2012).

We then fit a one-, two-, and three-parameter logistic IRT model to the item response data of the 10-item quiz. In our case, these models simultaneously estimate person ability (θ) and item parameters, depending on the type of model. The most simplistic of these models is the one-parameter logistic model (1PL):

$$P(Y_{is} = 1|\theta_s) = \frac{\exp\left(1.7a(\theta_s - b_i\right)}{1 + \exp\left(1.7a(\theta_s - b_i)\right)} \tag{1}$$

P is probability, *Y* is a participant's response to an item, θ is the person's latent trait ability (in our case, health insurance knowledge), *b* is item difficulty, *a* is item discrimination; *s* represents a person-parameter, while *i* represents an individual item. In the 1PL model, item discrimination (*a*) parameters are held constant; thus, items are assumed to have the same relation to the underlying construct. An additional model relevant to the current study is the two-parameter logistic model (2PL), shown below using similar notation to Equation 1:

$$P(Y_{is} = 1|\theta_s) = \frac{\exp\left(1.7a_i(\theta_s - b_i)\right)}{1 + \exp\left(1.7a_i(\theta_s - b_i)\right)} \tag{2}$$

Author's Accepted Manuscript

The 2PL is an extension of the 1PL model that allows the *a*-parameter to differ across items; that is, the relation between an item and the underlying construct of the test may differ across items. This allows an item's slope to change, thus creating different item characteristic curves and item information functions. The 3PL is an extension of the 2PL model, which incorporates a pseudo-guessing, *c*, parameter to create a lower asymptote. We compared the 1PL, 2PL, and 3PL models on item and overall fit measures estimated in the R package *ltm* (i.e., Q1 statistic with simulated *p*-values and loglikelihood ratio tests [LRT], respectively).

All item and person (θ) parameters were estimated on a z-score scale; θ was estimated using expected a posteriori (EAP) Bayes methodology implemented in *ltm* in R (Rizopoulos, 2007). The *mirt* package in R was also used to graphically display item fit, test information and standard error, and item residual plots and person-fit was examined using the Zh statistic (Drasgow, Levine, & Williams, 1985). Lastly, given the potential application of the quiz scores for health education/promotion practice, we examined the correlation between IRT-derived scores (θ) and CTT raw summated scores.

Author's Accepted Manuscript

Results

Classical Test Theory

The average summated quiz score was 5.425 (SD = 2.504), with a median score of a 6. Corrected item-total correlations, to provide information about item discrimination, were calculated: higher correlations indicate stronger relation between an item's response and the summated score of the items on the quiz, not including the item of interest. Correlations ranged from 0.269 to 0.525, with a mean of 0.414 (Table 2). Item difficulty (*p*; i.e., the proportion of respondents scoring correct on an item) ranged from 0.172 to 0.764, with a mean difficulty of

0.543. The most difficult item was Item 6 and Item 8 was the least discriminating, while Item 1 was most the discriminating.

Item Response Theory

Dimensionality. A CFA model with categorical indicators was fit to the data to assess the assumption of unidimensional structure. A single factor CFA fit the data well [χ^2 (df=35) = 297.028, $p < 0.05$; RMSEA = 0.058 (95% CI: 0.052 to 0.064), CFI = 0.961, TLI = 0.950], supporting the IRT assumption of unidimensionality. Item factor loadings ranged from 0.400 to 0.829, with all standard errors less than 0.032. The single factor scale had acceptable internal consistency reliability ($\alpha = 0.752$).

Model selection. Due to the range of corrected item-total correlations in CTT analyses, a 1PL model was hypothesized to have poorer fit because item discriminations (*a*) are held constant across items. We fit a 1PL, 2PL, and 3PL model and estimated item fit (Q1). Each model had at least one item with significant Q1 statistics, suggesting poorer item fit: 1PL had three items, 2PL one item (see Table 2), and 3PL two items. When comparing overall model fit using the LRT with -2 log likelihood values, the 2PL model had significantly better fit than the 1PL model [LRT = 303.13(df=9), $p < 0.001$] while the 2PL model and 3PL model did not differ significantly [LRT = 16.93(df=10), $p = 0.076$]. Thus, the 2PL model was selected for its parsimonious nature (i.e., not estimating an additional, *c*, parameter) and better overall and item fit.

Prior to examining item and person parameters of the 2PL model, we checked the assumption of local independence. Only two of forty-five (4%) item pairs had standardized chi-square values above |0.100| (Chen & Thissen, 1997): Item 1 and 2 ($r = 0.105$), and Item 3 and 4 ($r = 0.140$); this indicates potential local dependence in the response data. Given the reasonable

Author's Accepted Manuscript

overlap in relevant construct domains between those pairs (i.e., insurance premium for Items 1 and 2; deductible, Items 3 and 4) and the relatively lower standardized chi-square values, we determined there was sufficient evidence to support the assumption of local independence.

Item parameters. Discrimination (a_i) values ranged from 0.717 to 2.578. Higher values indicate a stronger relation between the item response and latent trait (i.e., health insurance knowledge), and thus, items with higher discriminations can differentiate individuals with smaller differences in the latent trait and provide more information. Difficulty (b_i) parameters represent a point where a respondent has a 0.50 probability of scoring correct to an item. The KFF scale's *b*-parameters ranged from -0.913 to 1.790 (see Table 2 and Figure 1). Positive values of b-parameters indicate more difficult items. For example, an item with b_i = -1.0 would indicate that a person with health insurance knowledge ability 1 standard deviation below the sample mean would have a 0.50 probability of getting the item correct; an item b_i = 1.0, a student with a knowledge ability 1 standard deviation above the sample mean would have a 0.50 probability of scoring correct.

Item 1 (definition of health insurance premium) had the highest discrimination (a_1 = 2.578) and was relatively easy (b_1 = -0.865). Analysis of item residual plots indicate that Item 3's observed data were slightly outside the 95% CI of the expected data for persons with higher than average θ; potentially leading to the item's slight statistical misfit. When comparing ranks of CTT difficulty (*p*) and 2PL *b*-parameters, there was consensus among the majority of items (i.e., Items 3, 4, 5, 6, 7, 8, and 9).

Person theta estimates. Person fit (Zh) statistics indicated good person fit for 97.6% of the sample; only 54 cases had Zh values outside of the range of -1.96 to 1.96. Person estimates ranged from -1.874 to 1.722, with a mean of -0.332. The minimum standard error of

Author's Accepted Manuscript

measurement occurred at $\theta = -0.686$ (Figure 2); thus, the KFF quiz has more precision in estimating health insurance knowledge when a student's ability is a slightly less than half a standard deviation below the average of students at the sampled university. Standard error was below 0.50 throughout the ability range -1.673 to 0.369; this indicates good precision of measuring student health insurance knowledge across the range of theta. However, the further a person's estimate was from -0.686, the less precision we have when estimating their ability. Raw summated and IRT estimated scores were strongly correlated ($r = 0.986$).

Discussion

The purpose of this study was to examine psychometric properties of the KFF's 10-item health insurance knowledge quiz to provide evidence of validity based on internal structure for use among college students. Our findings indicate that the KFF quiz measures a unidimensional construct and elicits reliable responses from college students in our sample. A 2PL IRT model best fits the data, indicating that items differ in their relation to the latent trait of health insurance knowledge. In total, the most precision in measuring health insurance knowledge ability for college students in our sample occurred slightly less than half a standard deviation below the average students' health insurance knowledge. This level of precision is beneficial in identifying knowledge deficits, which is useful for health promotion practitioners and health services researchers who are largely interested in identifying gaps in knowledge.

Item 3 (definition of annual health insurance deductible) was identified as an item with poor fit in the 2PL model and was also flagged for local dependence concerns with Item 4 (calculate out-of-pocket costs for hospital stay with deductible and copay). Given these potential issues, paired with the limited information provided by Item 3 relative to other items, researchers and health education specialists may consider removing Item 3 if using IRT estimates; however,

Author's Accepted Manuscript

removing this item should be done with caution. Item 3 had the third-highest *a*-parameter, and the item is more strongly related with the latent trait of health insurance knowledge than the other items. In addition, removal would likely lead to content underrepresentation of health insurance deductibles. Although Item 3 and Item 4 overlap by including deductible-based information, the definition of a health insurance deductible (found in Item 3) is likely foundational to health insurance knowledge and, therefore, important to measure. Finally, removing this item when using CTT estimates would likely decrease reliability of the score and, thus, increase standard error of measurement [in this sample: α (Item 3 deleted) = 0.720 vs. α (Item 3 included) = 0.752].

In both CTT and IRT analyses, the items displayed a range of difficulty and discriminatory power. The hardest item in both CTT and IRT analyses was Item 6 (definition of health insurance formulary; $p = 0.172$, $b_6 = 1.790$). This aligns with the KFF study which found this was the most difficult item with only 33% of U.S. adults able to correctly identify the correct definition (Norton et al., 2014). Interestingly, Nobles and colleagues (2018) found that 74% of college students in their sample correctly identified the meaning of health insurance formulary; however, it is likely that this is attributed to their item's poor response distractors:

> "A ______ is a list of prescription drugs covered by a prescription drug plan or another insurance plan offering prescription drug benefits. This is also called a drug list."
>
> Response options: Network; Referral; Formulary; Summary of benefits and coverage (SBC)

These inappropriate response distractors may lead respondents to arrive to the correct answer through process of elimination, and thus contribute to an overestimation of student health insurance knowledge and have negative implications for health education specialists who are

Author's Accepted Manuscript

using the instrument to identify topics to cover in health education programming; that is, if students are scoring highly on knowledge of a health insurance formulary, then formularies may not be covered in as much detail in practice because a need is not being demonstrated.

A limitation of the current study is that the KFF 10-item quiz was not developed according to test construction best practices (e.g., response validity studies, etc.; Crocker & Algina, 1986) and was not explicitly developed for the proposed use of health promotion program planning. However, the evidence of validity based on content and internal structure suggests that the quiz elicits information about the underlying, unidimensional construct of health insurance knowledge. The results from our analyses add to the evidence of validity for using scores from this quiz to measure health insurance knowledge. Further, the evidence of strong correlation between raw scores and latent scores supports the use of raw scores in-practice; this is important because conducting an IRT study is impractical and not feasible in most health education practice settings, and health education specialists are unlikely to have the analytic resources (e.g., training, access to software, sample size, etc.) necessary to estimate scores in an IRT framework. Additional research is needed to identify meaningful cut-off scores to better inform practice-based identification of students with inadequate knowledge, and validity studies – ensuring adequate content coverage and external relations with relevant variables – are needed: this can include Delphi studies with experts in health services and insurance, examining the relation of quiz scores to health literacy scores and healthcare utilization, and sub-group analyses. An additional limitation is the lack of a representative sample of college students in the U.S., as this study only includes students from a single university; this may inhibit the estimation of measurement precision for the population of college students in the U.S. However, the sample does include a broad range of scores that may provide accurate samples of the IRT model and

Author's Accepted Manuscript

parameters. Overall, our findings support the use of scores elicited by the KFF's health insurance knowledge quiz to identify health insurance information needs among college students.

Relevance to Nursing Practice, Education, or Research

The KFF's 10-item health insurance knowledge quiz can be integrated into practice in community and collegiate health care settings as a tool for identifying college students' health insurance knowledge-levels and assessing the need for intervening to ensure college students make the most cost-efficient decisions for their healthcare. This would provide an opportunity to better serve college student patients by providing them tailored, usable information regarding health insurance and potentially referring them to financial support staff or health education specialists for more information. Further, these results have implications for health educators developing programs and materials to improve the health insurance knowledge among college student populations. IRT difficulty parameters provide insight into the performance of our university's students with specific domains of health insurance knowledge. Thus, health education specialists should target areas of lower knowledge (i.e., higher difficulty) when developing programs and materials. For our sample, this would include more information on health insurance formularies and calculating co-insurance and out-of-pocket costs.

Conclusion

This study applied classical and modern measurement theory to evaluate the psychometric properties of the KFF's 10-item health insurance knowledge quiz. Results suggest that the quiz elicits reliable responses from participants, and those responses represent a unidimensional construct of health insurance knowledge. The 10-item quiz is user-friendly, publicly available, and reviewed positively by college health education specialists and healthcare

Author's Accepted Manuscript

administrators. This study supports the use of the KFF’s quiz to measure health insurance knowledge among college students.

Author's Accepted Manuscript

References

Allen, E. M., Call, K. T., Beebe, T. J., McAlpine, D. D., & Johnson, P. J. (2017). Barriers to care and healthcare utilization among the publicly insured. *Medical Care*, *55*(3), 207–214. https://doi.org/10.1097/MLR.0000000000000644

Bandura, A. (2004). Health promotion by social cognitive means. *Health Education & Behavior*, *31*(2), 143–164.

Berkman, N. D., Sheridan, S. L., Donahue, K. E., Halpern, D. J., & Crotty, K. (2011). Low health literacy and health outcomes: An updated systematic review. *Annals of Internal Medicine*, *155*(2), 97–107.

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*(6). https://doi.org/10.18637/jss.v048.i06

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.

Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Belmont, CA: Wadsworth Publishing Company.

Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, *38*(1), 67–86. https://doi.org/10.1111/j.2044-8317.1985.tb00817.x

Emmereé, J., James, T. G., Varnes, J. R., & Arceneaux, D. (2015). *Healthy Gators Student Survey Report, 2013*. Gainesville, FL: Healthy Gators Coalition.

Glanz, K., Rimer, B. K., & Viswanath, K. (2015). *Health behavior: Theory, research, and practice*. San Francisco, CA: John Wiley & Sons.

Author's Accepted Manuscript

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

James, T. G., Emmereé, J., & Brantley, J. (2015). *Healthy Gators Student E-cigarette Survey, Spring 2015.* Gainesville, FL: Healthy Gators Coalition.

James, T. G., & Ryan, S. J. (2018). HIV knowledge mediates the relationship between HIV testing history and stigma in college students. *Journal of American College Health*, *66*(07), 561–569. https://doi.org/10.1080/07448481.2018.1432623

James, T. G., Sullivan, M. K., Dumeny, L., Lindsey, K., Cheong, J., & Nicolette, G. (in press). Health insurance literacy and health service utilization among college students. *Journal of American College Health*.

Muthén, L. K., & Muthén, B. O. (2011). *Mplus user's guide* (6th ed.). Los Angeles: Muthén & Muthén.

Nobles, A. L., Curtis, B. A., Ngo, D. A., Vardell, E., & Holstege, C. P. (2018). Health insurance literacy: A mixed methods study of college students. *Journal of American College Health*, *0*(ja), 1–37. https://doi.org/10.1080/07448481.2018.1486844

Norton, M., Hamel, L., & Brodie, M. (2014). *Assessing Americans' familiarity with health insurance terms and concepts*. Retrieved from Kaiser Family Foundation website: https://www.kff.org/health-reform/poll-finding/assessing-americans-familiarity-with-health-insurance-terms-and-concepts/

Paez, K. A., Mallery, C. J., Noel, H., Pugliese, C., McSorley, V. E., Lucado, J. L., & Ganachari, D. (2014). Development of the Health Insurance Literacy Measure (HILM):

Author's Accepted Manuscript

Conceptualizing and measuring consumer ability to choose and use private health insurance. *Journal of Health Communication*, *19*(sup2), 225–239.

Quincy, L. (2012). Measuring health insurance literacy: A call to action. *Yonkers: Consumers Union*.

Rizopoulos, D. (2007). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, *17*(5). https://doi.org/10.18637/jss.v017.i05

Vujicic, M., Buchmueller, T., & Klein, R. (2016). Dental care presents the highest level of financial barriers compared to other types of health care services. *Health Affairs*, *35*(12), 2176–2182. https://doi.org/10.1377/hlthaff.2016.0800

Willse, J. T. (2018). CTT: Classical Test Theory functions (Version 2.3.3) [R]. Retrieved from https://cran.r-project.org/web/packages/CTT/CTT.pdf

Yang, L. (2016). *Young adults' attitudes and perceptions on health insurance and their health insurance literacy levels* (Minnesota State University Mankato). Retrieved from https://cornerstone.lib.mnsu.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1616&context=etds

Author's Accepted Manuscript

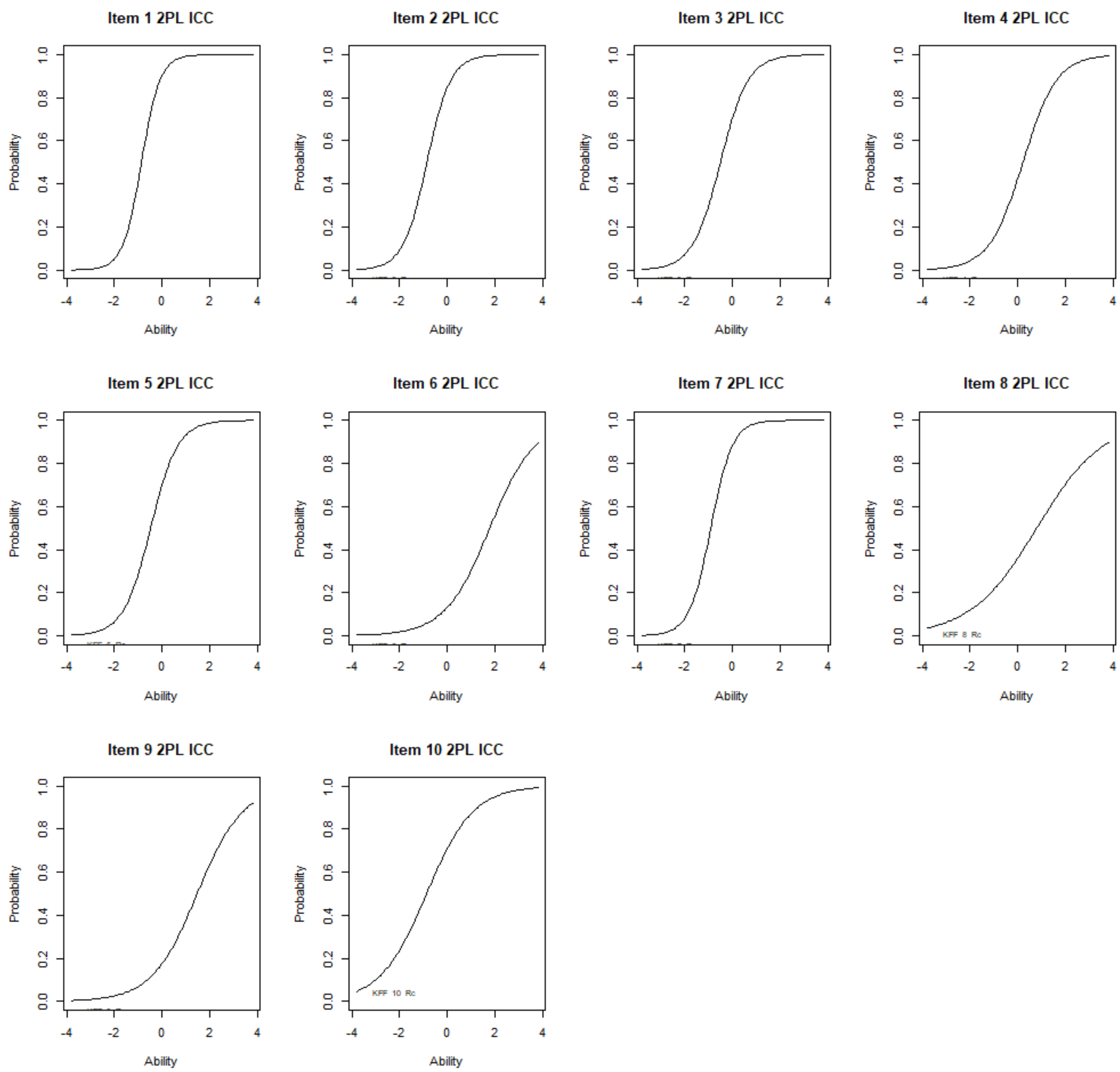

Figure 1. Individual item characteristic curves of KFF health insurance knowledge quiz items estimated under the 2PL model.

Auth

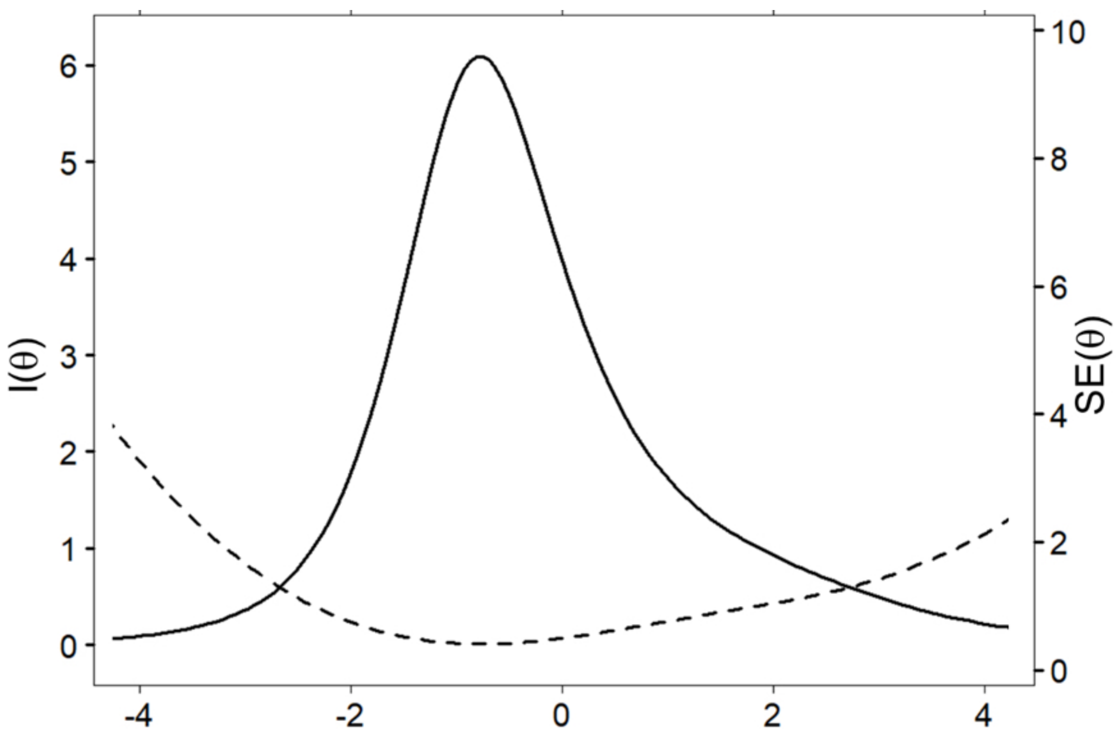

Figure 2. Test information function (solid) with standard error of measurement (dotted) for the range of theta scores (X-axis) elicited from the KFF health insurance knowledge quiz estimated under the 2PL model.

Author's Accepted Manuscript

Table 1

Health Insurance Quiz Items

#	Item	Response Options
1	Which of the following is the best definition of the term "health insurance premium"?	A: The best type of health insurance you can buy. B*: The amount health insurance companies charge each month for coverage. C: A bonus you get at the end of the year if you stay covered. D: I don't know.
2	Is a health insurance premium something you must pay every month, regardless of whether you use health care services, or do you only have to pay your health insurance premium during months when you use health care services?	A*: Must pay every month, regardless of whether you use services. B: Only have to pay in months when you use health care services. C: I don't know.
3	Which of the following is best definition of the term "annual health insurance deductible"?	A: The amount that is deducted from your paycheck each year to pay for your policy. B: The amount of health expenses you can subtract from your income on your yearly tax return. C*: The amount of covered health care expenses you must pay yourself each year before your insurance will begin to pay.
4	Suppose that under your health insurance policy, hospital expenses are subject to a $1,000 deductible and $250 per day copay. You get sick and are hospitalized for 4 days, and the bill comes to $6,000. How much of that hospital bill will you have to pay yourself?	A: $0 B: $1,000 C*: $2,000 D: $4,000 E: $6,000 F: I don't know.

Author Manuscript

5	Which of the following best describes the "annual out-of-pocket limit" under a health insurance policy?	A*: The most you will have to pay in deductibles, copays, and coinsurance for covered care received in network for the year. B: The most your insurance policy will pay for covered services in a year. C: The most you will have to pay for premiums in a year. D: I don't know.
6	Which of the following best describes a "health insurance formulary"?	A: The form you send to your insurance company when you need to have a medical bill paid. B: The name for permission you must get from your insurance company before surgery will be covered. C*: The list of prescription drugs your health plan will cover. D: I don't know.
7	Which of the following best describes a health plan "provider network"?	A*: The hospitals and doctors that contract with your health plan to provide services for an agreed-upon rate or fee schedule. B: The computer system doctors and hospitals use to submit bills to insurance companies. C: A website where consumers can find information about the best doctors. D: I don't know.
8	True or false: If you receive inpatient care at a hospital that participates in your health plan's provider network, all the doctors who care for you while you're in the hospital will also be in network.	A: True B*: False C: I don't know.
9	Suppose your health plan covers lab tests in full if you go to an in-network lab, but only pays 60% of allowed charges if you go out of network. You forget to check and get your blood test at a lab that turns out to be out of network. The lab	A: \$0 B: \$40 C: \$80 D*: \$88

Author's Accepted Manuscript

	bills you $100 for the blood test. Your health insurance allows only a $20 charge for that test. How much would you have to pay out of pocket for that lab test?	E: $100 F: I don't know.
10	True or false: If your health insurance plan refuses to pay for a service that you think is covered and your doctor says you need, you can appeal the denial and possibly get the insurance company to pay the claim.	A*: True B: False C: I don't know.

Note. * indicates correct response option.

Author's Accepted Manuscript

Table 2

Descriptive Statistics, Item Parameter Estimates, and Item Fit Statistics of the 10-item Health Insurance Knowledge Quiz Under the Two-Parameter Logistic (2PL) Model

	Classical Test Theory			Item Response Theory: 2PL Model					
Item	Difficulty (*p*)	Corrected item-total correlation	α if deleted	Difficulty (*b*)	SE (*b*)	Discrimination (*a*)	SE (*a*)	Yen's Q1	*p*, Q1
1	0.763	0.525	0.716	-0.865	0.041	2.578	0.186	81.271	0.347
2	0.742	0.485	0.721	-0.856	0.045	2.006	0.132	64.461	0.673
3	0.640	0.486	0.720	-0.507	0.041	1.717	0.106	107.159	0.020
4	0.444	0.429	0.729	0.219	0.042	1.414	0.090	113.011	0.347
5	0.634	0.503	0.717	-0.450	0.040	1.761	0.109	100.248	0.069
6	0.172	0.279	0.748	1.790	0.124	1.058	0.090	89.183	0.089
7	0.764	0.515	0.717	-0.913	0.044	2.230	0.151	67.652	0.624
8	0.374	0.269	0.753	0.801	0.088	0.717	0.062	35.622	0.584
9	0.217	0.293	0.747	1.480	0.101	1.049	0.084	58.872	0.673
10	0.675	0.357	0.740	-0.865	0.068	1.027	0.073	80.204	0.228
Overall	M = 0.543	M = 0.414	0.752						

Note. *p*-values of Yen's Q1 estimate were simulated in the R package *ltm* to account for the large sample size.

Author's Accepted Manuscript