Published in final edited form as:

Support Care Cancer. 2014 August ; 22(8): 2229–2240. doi:10.1007/s00520-014-2202-7.

Item Response Theory Analysis of the Patient Satisfaction with Cancer-Related Care Measure: A Psychometric Investigation in A Multicultural Sample of 1,296 Participants

Pascal Jean-Pierre, PhD, MPH^{1,2}, Ying Cheng, PhD¹, Electra Paskett, PhD, MPH³, Can Shao, BS¹, Kevin Fiscella, MD, MPH⁴, Paul Winters, MS⁴, and Patient Navigation Research Program

¹University of Notre Dame, Department of Psychology, Notre Dame, IN

²Harper Cancer Research Institute, South Bend, IN

³Ohio State University Research Foundation Comprehensive Cancer Center, Columbus, OH

⁴University of Rochester Medical Center, Family Medicine Research, Rochester, NY

⁵Patient Navigation Research Program group

Abstract

BACKGROUND—We developed and validated a Patient Satisfaction with Cancer-Related Care (PSCC) measure using classical test theory methods. The present study applied item response theory (IRT) analysis to determine item-level psychometric properties, facilitate development of short forms, and inform future applications for the PSCC.

METHODS—We applied unidimensional IRT models to PSCC data from 1,296 participants (73% female; 18 to 86 years). An unconstrained graded response model (GRM) and a Rasch Model were fitted to estimate indices for model comparison using likelihood ratio (LR) test and information criteria. We computed item and latent trait parameter estimates, category and operating characteristic curves, and tested information curves for the better fitting model.

RESULTS—The GRM fitted the data better than the Rasch Model (LR=828, df=17, p<0.001). The log-Likelihood (−17390.38 vs. −17804.26) was larger, and the AIC and BIC were smaller for the GRM compared to the Rasch Model (AIC=34960.77 vs. 35754.73; BIC=35425.80 vs. 36131.92). Item parameter estimates (IPEs) showed substantial variation in items' discriminating power (0.94 to 2.18). Standard errors of the IPEs were small (threshold parameters mostly around 0.1; discrimination parameters: 0.1 to 0.2), confirming the precision of the IPEs.

CONCLUSION—The GRM provides precise IPEs that will enable comparable scores from different subsets of items, and facilitate optimal selections of items to estimate patients' latent

Send Correspondence requests to: **Dr. Pascal Jean-Pierre, PhD, MPH** University of Notre Dame Department of Psychology Neurocognitive Translational Research Lab Cancer Control & Survivorship Program 109 Haggard Hall, Notre Dame, IN 46556 Phone: (574) 631-2589 PJeanPierre@nd.edu.

Conflict Of Interest: The authors have no conflict of interest to declare. The authors have full control of all primary data and agree to allow the journal to review their data if requested in accordance with the data-sharing plan of the Patient Navigation Research Program (PNRP) Steering Committee.

satisfaction level. Given the large calibration sample, the IPEs can be used in settings with limited resources (e.g., smaller samples) to estimate patients' satisfaction.

Keywords

Psychometrics; Measure development; Psychometric validation; Item Response Theory; Cancer Disparities; Race-ethnicity

INTRODUCTION

Patient satisfaction with cancer-related care is an important dimension of the quality of cancer screening or treatment [1-3]. As a consequence, patient satisfaction is frequently used to systematically quantify the extent to which patients' health care experience matches the level and quality of care they expect [4-5]. Previous studies have reported significant associations between patient satisfaction and health status, treatment adherence, quality of life, as well as the relationship with, and quality of communication between patients and their health care providers [6-16].

Patient satisfaction was one of the four core outcomes evaluated in the multicenter collaborative Patient Navigation Research Program (PNRP), which was undertaken to reduce cancer disparities for socially disadvantaged and medically underserved individuals from racial-ethnic minority and lower socioeconomic groups in the United States [17-18]. The PNRP, funded by the National Cancer Institute and the American Cancer Society, evaluated the benefits of linking patients with trained patient navigators (both lay and professional) who helped patients identify and overcome cost-related, informational, attitudinal, and logistical barriers to accessing and utilizing timely cancer-related care (e.g., follow-up of abnormal cancer screening, definitive cancer diagnosis, or initiation and completion of cancer treatment) [19]. Using classical test theory methods, the PNRP also developed and psychometrically validated the Patient Satisfaction with Cancer-Related Care (PSCC) [19].

The present study involves an innovative approach to integrate the perspectives and concerns of medically underserved cancer patients into their clinical care process. We undertook item response theory (IRT) analysis to psychometrically characterize individual items of the PSCC. Item parameter estimates (e.g., difficulty and discrimination/slope parameters based on item characteristic curve) from the present analysis can be used to improve assessment of the satisfaction levels of cancer patients in future studies. For instance, scores from different subsets of items of the PSCC (e.g., different versions of short forms) can be directly compared based on IRT item parameter estimates. Additionally, IRT parameter estimates from this study will be useful in clinical and research settings with limited resources (e.g., small sample size) where item level estimates on the PSCC cannot be readily obtained. Finally, IRT parameter estimates will provide information about the psychometric properties and difficulty of specific test items to facilitate tailoring of subset of items to individual patients using a computer adaptive testing (CAT) system.

Some of the advantages of using the IRT in lieu of the CTT are as follow: 1) The IRT analysis will provide a latent satisfaction score that will explicitly accounts for measurement

error for each patient, whereas the CTT score is a raw sum score that does not account for measurement error; and 2) the IRT analysis will facilitate the assessment of conditional measurement precision. For example, the IRT analysis can help us understand whether the PSCC is best suited to assess patients with high satisfaction level or those low satisfaction level. In contrast, the CTT analysis treats reliability or measurement precision as a constant. Using IRT analysis, which focuses on item-level analytics (in contrast to CTT, which generally focuses on the whole scale) will allow us to modify or revise each item of the PSCC based on individual item psychometric properties. Additionally, the item-level information will enable CAT that is not obtainable using classical test theory methods. IRT-based CAT system will help reduce item burden without sacrificing measurement precision, and will provide an accessible, repeatable and reliable tool for assessing and monitoring patient satisfaction across the cancer care continuum.

In summary, the present study provides extensive psychometric validation of the PSCC using IRT analysis to complement previous psychometric validation of this measure using classical test theory methods. This approach will inform and facilitate the adoption of the PSCC in cancer outcome research and evidence-based clinical practices.

METHODS

Participants

The present study includes data from a multicultural sample of 1,296 participants who completed the PSCC as part of their participation in the PNRP. Participants read and provided signed informed consent to indicate their voluntary participation in the PNRP. The sample included 180 Whites, 277 Blacks/African Americans, 238 Asians, 3 American Indians or Alaska Natives, 210 Hispanics or Latinos, and 16 other races/multiple races. These participants presented with either an abnormal cancer screening test result or a definitive cancer diagnosis (i.e., breast, cervical, colorectal, or prostate cancers). The sample ranged from 18 to 86 years of age and included mostly medically underserved women from racial-ethnic minority populations and lower socioeconomic and education backgrounds. See Jean-Pierre et al., 2011 for additional sample characteristics [17]. The large multicultural sample of the present study will facilitate generalizability of the findings (See Table 1).

Procedures

Medical staff at participating clinics and hospitals referred participants to the PNRP. A research assistant met with each participant to determine eligibility for enrollment in the study. Eligible participants were randomly assigned to standard care or patient navigation by a trained lay patient navigator from their community. The PSCC was read aloud to minimize possible effects of respondents' low literacy on response outcome.

Measures

Demographic and Clinical Characteristics—We assessed demographic and clinical characteristics, including age, sex, race-ethnicity, primary language, income, education, marital status, whether the patient received care related to evaluation of cancer screening

abnormalities or treatment of cancer, and cancer types (i.e., breast, cervical, colorectal, or prostate), using questionnaires and/or data abstracted from medical records.

Patient Satisfaction with Cancer-related Care (PSCC)—A multidisciplinary team of clinical researchers developed the PSCC measure. Members of the team had expertise in measurement development and psychometric validation, biostatistics, cancer research, and experience in clinical care and treatment of medically underserved racial-ethnic minorities and the poor. The development of the PSCC was described in a previous manuscript [17]. Briefly, the measurement development team discussed relevant domains of patient satisfaction with cancer care that will be relevant to the PNRP, reviewed existing satisfaction measures, selected, modified and developed new PSCC items. Additionally, the team discussed the relevance of each item of the PSCC with experts and a few patients during the item pool development and pre-pilot phases of the study.

PSCC items are rated on 5-point Likert scales: “1 = Strongly Agree”, “2 = Agree”, “3 = Neutral”, “4 = Disagree”, and “5 = Strongly Disagree”. A lower PSCC total score indicates higher satisfaction level. Agreement or disagreement with the PSCC items connotes satisfaction or dissatisfaction, respectively. The PSCC items could have been worded to ask directly about how satisfied a patient was, for example, that his/her “health concerns were understood”. However, the PSCC development team intended to stimulate more objective and factual responses to the scale items by phrasing them as we did to facilitate a coherent set that explained the construct of satisfaction with cancer care. At the time of the development of the PSCC, the team understood that it was possible that a person could have agreed that doctors communicate well but still have concerns with his/her specific doctor-patient communication because of personal factors that are beyond doctors’ control (e.g., a patient feeling that he or she did not communicate well by failing to ask certain questions or making inquiries while communicating with the doctor).

The psychometric validation of the PSCC using classical test theory and principal components analysis (PCA) revealed a one-dimensional structure. The PCA revealed that the scale accounted for 62% of the variance in patient satisfaction. Parallel analysis suggested one underlying dimensional as well, with only one eigenvalue larger than the 95th percentile of eigenvalues obtained from randomly generated parallel samples. The largest drop also occurred between the first and the second eigenvalues. A confirmatory factor analysis (CFA) with a single factor model fitted to the data revealed all large factor loadings, ranging from 0.6 - 0.8. The one-factor CFA showed adequate model fit (RMSEA = 0.078, 95%CI of RMSEA [0.073, 0.082], and CFI = 0.996). Altogether, these aforementioned results strongly supported a unidimensional IRT model. In addition, local dependency is considered a form of violation of unidimensionality [20]. Therefore with adequate evidence of unidimensionality we did not pursue further with local dependency tests. The latent structure of the PSCC was replicated across different samples and time points. Reliability analysis revealed a high internal consistency for the PSCC, with Cronbach coefficient alphas ranging from 0.95 to 0.96. Face validity as well as discriminant and convergent characteristics of the PSCC were also demonstrated [17].

Data Analysis

We applied one-dimensional IRT models to PSCC data from the 1,296 PNRP participants. Two IRT models were fitted to the data: an unconstrained graded response model (GRM) and a constrained GRM (i.e., Rasch Model) in which all discrimination parameters across items were fixed to be equal [22]. The response scale for the PSCC goes from $k = 1$ to $(m + 1)$, where $m = 4$ for the 5-point Likert scale of this measure. The GRM posits that the probability of scoring *at or above category k* on item j at a given level of latent trait θ is as follows:

$$P_{jk}^*(\theta) = P(X_j \geq k | \theta) = \frac{\exp(\alpha_j(\theta - \beta_{j(k-1)}))}{1 + \exp(\alpha_j(\theta - \beta_{j(k-1)}))}, \quad (1)$$

where $k = 1, 2, \dots, m + 1$; X_j is the participant's response to item j ; α_j is the discrimination parameter, and the β_{jk} 's are the threshold parameters. In fact, β_{j0} is never estimated and by definition is $-\infty$. By the same token, $\beta_{j(m+1)} = \infty$. This implies that $P_{j1}^*(\theta) = 1$ and $P_{j(m+1)}^*(\theta) = 0$. For a 5-point Likert scale, $\beta_{j1} - \beta_{j4}$ will be estimated. For a polytomous item (i.e., more than two response categories), the number of threshold parameters to be estimated is the number of response categories minus 1. The discrimination parameter generally indicates how much an item can discriminate between patients with different latent satisfaction levels. The threshold parameter of β_{jk} represents the "hurdle" (i.e., the point where the latent trait level leads to an equal probability of endorsing either of two adjacent response categories) on the latent satisfaction continuum between score category k and $(k + 1)$. The probability of scoring exactly k , denoted by $P_{jk}(a)$, therefore is $P_{jk}^*(\theta) - P_{j(k+1)}^*(\theta)$. For instance, for the 5-point Likert scale of the PSCC,

$$P_{j2}(\theta) = P_{j2}^*(\theta) - P_{j3}^*(\theta) = \frac{\exp(\alpha_j(\theta - \beta_{j1}))}{1 + \exp(\alpha_j(\theta - \beta_{j1}))} - \frac{\exp(\alpha_j(\theta - \beta_{j2}))}{1 + \exp(\alpha_j(\theta - \beta_{j2}))}.$$

The GRM is a widely used IRT model for Likert-scale data when dealing with unidimensional measures [21]. The Rasch Model, however, is a more parsimonious model for which fewer parameters need to be estimated [22]. The Rasch model could be a good alternative when it offers comparable fit to the data. Therefore, we obtained model fit indices, including log-likelihood, Akaike's information coefficient (AIC), and Bayesian information coefficient (BIC) for both the GRM and the Rasch model. Then, we computed a likelihood ratio (LR) test to compare the two IRT models (viz., GRM and Rasch model). Finally, we obtained item parameter estimates and latent trait parameter estimates (i.e., patient satisfaction), category characteristic curves, operating characteristic curves, and test information curves for the better fitting of the two models. We used SAS 9.3 and IBM SPSS statistical software for the PCA and CFA analyses. We used IRTpro for the Item Response Theory (IRT) analysis.

RESULTS

The lower end of the PSCC (i.e., Strongly Disagree”) was endorsed less. For example, the percentage of respondents endorsing each response category of item 9 was: 3.52% (Strongly Disagree), 10.24% (Disagree), 4.07% (Neutral), 47.54% (Agree) and 34.64% Strongly Agree. Albeit not unsubstantial, this pattern of endorsement is typical of satisfaction measures. The “Strongly Disagree” and “Disagree” categories account for about 15% of the scale response options endorsement.

The GRM fitted the data significantly better ($LR = 828$, $df = 17$, $p < 0.001$) than the Rasch model, and had smaller AIC and BIC. Inspection of the model fit indices shows that, despite the parsimonious attributes of the Rasch model (log likelihood = -17804.26 , AIC = 35754.73 , BIC = 36131.92), the fit of the GRM (log likelihood = -17390.38 , AIC = 34960.77 , BIC = 35425.80) to the PSCC data was substantially better and more appropriate. Additionally, the enhanced fit of the GRM fully compensated for the additional parameters it inflicted. Thus, for our subsequent analyses we focus solely on the GRM. It should be noted that in the original PSCC, the response scale goes from “1 = Strongly Agree”, “2 = Agree”, “3 = Neutral”, “4 = Disagree”, and “5 = Strongly Disagree”. Items of the PSCC were appropriately reverse-coded as needed during the initial validation of the measure. In the present analysis, the first response category in the response curves refers to the “strongly disagree” category (See Figure 1).

Table 2 provides GRM estimates of the item discrimination (α_i) and threshold parameters ($\beta_{i1} - \beta_{i4}$), as well as estimates of their standard errors (SE). Item discrimination parameter estimates range from 0.94 to 2.18. Additionally, the good dispersion of these estimates suggests that the GRM would better fit the PSCC data than the Rasch model, which assumes that all the items have equal discriminating power. Because the GRM fits better, simply adding up scores on the 18 items will not produce the optimal estimate of the latent-level satisfaction of a patient. Instead, a composite of weighted item scores should be used, where item weights depend item discrimination. Equivalently, we could use the maximum likelihood estimate of θ [23].

For the PSCC, which utilizes a 5-point Likert scale, there are four threshold parameters for each item. The first threshold parameter indexes the “difficulty” (i.e., satisfaction level) needed to reach the second response category. The second threshold parameter indexes the satisfaction level needed to reach the third response category from the second, and so on for the third and fourth threshold parameters. Not surprisingly, our analyses revealed that threshold parameter estimates increased from the first threshold through the fourth. Standard errors of the item parameter estimates were generally small (mostly around 0.1 for the threshold parameters, and from 0.1 to 0.2 for the discrimination parameters), suggesting the item parameter estimates we obtained were quite precise.

Figure 1 shows the category response curves for each item. Each curve in each plot represents the probability of endorsing a certain response category. For example, the black curve shows how the probability of endorsing the first response category (i.e., “strongly disagree”) changes as a function of variations in satisfaction level. This curve is higher at the

lower end of the latent trait (see the horizontal axis), indicating that patients who are very dissatisfied with the cancer-related care they received tend to endorse the lowest category on this item. As the latent trait increases, the probability of endorsing the first category decreases while the probability of endorsing the other response categories increases. When the satisfaction level is very high, patients tend to endorse the highest response category (i.e., “strongly agree”).

With the exception of item 18 (i.e., “My regular doctor was informed about the results of the test I got”), the third response category generally has a low endorsement across the latent trait continuum (Table 3). For example, with item 8 “Making an appointment was easy”, the category characteristic curve (CCC) of the third category (the green curve) is almost flat, indicating that across the latent trait continuum, the probability of endorsing the third response category (i.e., “neutral”) is quite low. From descriptive statistics it is straightforward to find out that the overall endorsement of the category “neutral” is low for an item, but it is only with IRT analysis that we can determine if the endorsement is low across the latent trait continuum. This suggests that the response scale can be safely reduced to four categories if desired. In item parameter estimates, this is reflected by threshold parameters that are close to each other. For example, Table 2 shows that β_{i2} and β_{i3} are generally close to each other, while β_{i1} and β_{i4} are relatively far from their adjacent thresholds.

Figure 2 shows the operating characteristic curves (OCC) for each item. The OCCs can be derived from the CCCs, or vice versa. Additionally, both OCC and CCC can be derived from the item parameter estimates in Table 2. Each OCC curve represents the probability of responding in a certain category or higher and how this probability changes as a function of the latent trait level. Note that there are only four curves for each item in the OCC plot but five CCCs per item. This is because the probability of scoring in the lowest category or higher is always 1 and consequently that OCC curve is omitted. The OCC curves, from left to right, represent the probability of responding to category 2 or higher, 3 or higher, and 4 or higher. We can see that the middle two curves almost overlap each other for many items (e.g., item 8 “Making an appointment was easy” and item 9 “I knew what the next step in my care would be”), with item 18 as an exception. This is another way of showing that the third response category adds very little information at any level of patient satisfaction.

Additionally, we plotted item information curves and found that most of the items have bimodal information that generally peaks between $\theta = -2.0$ and -1.5 , and between 0 and 0.5 (Figure 3). Items 8, 9 and 18 have markedly lower information across the latent trait continuum than the other items, suggesting that if a short form of PSSC were to be developed, these items can be dropped without decreasing the scale reliability, given that the PSSC is unidimensional. On the other hand, item 2 should be retained because one of its information peaks occurs around θ between -1.0 and -0.25 , where most other items fail to yield large information. Retaining item 2 therefore allows us to better measure patients whose satisfaction level is within that range. Using CTT analysis we only obtain a general measure of information integrated over the entire range of θ and will lose such local distinction [23-24]. Most items are informative within the -2.5 to 1.0 range, which is beneficial because most people fall within this range on the latent trait (Figure 3). Further,

people above 1.0 are quite satisfied with the care they received. It is those patients who are dissatisfied with the care they received that warrant concern. Hence, items informative in the range of -2.5 to 1.0 are especially valuable. Aggregating the item information curves produces a test information curve that demonstrates a similar pattern, suggesting that the PSCC is appropriate for patients whose satisfaction level is in the -2.5 to 1.0 range (Figure 3).

DISCUSSION

We used unidimensional IRT analysis to evaluate the psychometric properties of the PSCC at the item level. IRT techniques are increasingly applied in measurement standardization (e.g., attitude measurements, behavioral ratings, and clinical testing issues) [25-28]. Additionally, IRT analysis provides a convenient means for selecting an optimal subset of test items for individual examinees and for equating scores across different subsets of items for different examinees (or across time for the same examinee in a repeated measurement condition). Further, if only a subset of items of the PSCC (e.g., a PSCC short form) is used on another population, the corresponding item parameter estimates can be consulted and the satisfaction level of people from that population can be computed accordingly. Overall, the results confirm previous classical test theory analyses by showing that the PSCC has good psychometric properties and internal validity. The GRM model fits the PSCC data well, and the item parameter estimates obtained with this large PSCC data set are relatively precise. The findings suggest possible ways to improve the PSCC measure. For example, the “neutral” response category can be removed, and items 8, 9 and 18 can be revised if needed. Also, if an overall short form of the PSCC was needed, these items may be omitted without compromising measurement precision. Nonetheless, content representation would need to be considered.

Additionally, the results of this study provide a precise characterization of the psychometric properties of each item of the PSCC through item parameter estimates and graphic representations (i.e., category characteristic curves and item information curves). The item parameter estimates obtained from the present study with a large multicultural calibration sample of 1,296 participants are accurate, generalizable and can be readily adopted by other researchers in future studies that use the PSCC to estimate latent patients’ satisfaction level with cancer care.

The present IRT analysis of the PSCC offers many advantages over the previously conducted psychometric validation using classical test theory (CTT) methods. Specifically, the present analysis provides item-level information that complement data from the CTT. The findings of the IRT analysis may help shorten the PSCC, obtain various parallel short forms, and evaluate the performance of these shortened measures across different times and/or examinees. The present study is also instrumental for the development and examination of a future CAT system to help tailor the PSCC to patients’ satisfaction level, and could also allow patients’ latent satisfaction scores to be made immediately available to their clinical care providers. This strategy, if successfully implemented, could help improve patient-centered care by enhancing patient-clinician communication, increasing clinical care providers’ understanding of patients’ health care expectations, and facilitating the

integration of patients' perspectives and their evaluation of the cancer care they received into their treatment process.

Acknowledgments

Dr. Pascal Jean-Pierre wishes to acknowledge the support of the University of Notre Dame, Walther Cancer Foundation, and Indiana University CTSI. This publication is made possible in part by the National Institutes of Health, National Center for Advancing Translational Sciences, Clinical and Translational Sciences Award Grant Number KL2 TR000163 (A. Shekhar, PI). Dr. Jean-Pierre also wishes to express his appreciation and gratitude to Drs. Scott E. Maxwell, David A. Smith, and Thomas V. Merluzzi for reviewing and providing feedback on earlier drafts of this manuscript.

Funding Sources: The PNRP was supported by grants from the National Cancer Institute (NCI) (3U01CA116924-03S1, U01 CA116924-01, 1R25CA 10261801A1, U01CA116892, U01CA 117281, U01CA116903, U01CA116937, U01CA116885, U01CA116875, and U01 CA116925) and the American Cancer Society (ACS) (SIRSG-05-253- 01).

The Patient Navigation Research Program group included Peter C. Raich, MD, Denver Health & Hospital Authority, Denver, CO; Richard Roetzheim, MD, MSPH, H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL; Steve T. Rosen, MD, Northwestern University, Robert H. Lurie Comprehensive Cancer Center, Chicago, IL; Donald J. Dudley, MD, University of Texas Health Science Center, San Antonio, TX; Karen M. Freund, MD, MPH, Tufts Medical Center, Boston, MA; Victoria Warren-Mears, PhD, Northwest Portland Area Indian Health Board, Portland, OR; and Steven R. Patierno, PhD, Duke University; Durham, NC.

REFERENCES

1. Bredart A, Razavi D, Roberson C, et al. Assessment of quality of care in an oncology institute using information on patients' satisfaction. *Oncology*. 2001; 61:120–128. [PubMed: 11528250]
2. Sanda MG, Dunn RL, Michalski J, et al. Quality of life and satisfaction with outcome among prostate-cancer survivors. *N Engl J Med*. 2008; 358:1250–1261. [PubMed: 18354103]
3. Sitzia J, Wood N. Patient satisfaction: a review of issues and concepts. *Soc Sci Med*. 1997; 45:1829–1843. [PubMed: 9447632]
4. Institute of Medicine. *Crossing the Quality Chasm: A New Health System for the 21st Century*. National Academy Press; Washington, DC: 2001.
5. Pascoe GC. Patient satisfaction in primary health care: a literature review and analysis. *Eval Program Plann*. 1983; 6:185–210. [PubMed: 10299618]
6. Cleary PD, McNeil BJ. Patient satisfaction as an indicator of quality of care. *Inquiry*. 1988; 25:25–36. [PubMed: 2966123]
7. Lenderking WR. Brief reflections on treatment satisfaction. *Value Health*. 2005; 8:S2–S5. [PubMed: 16336485]
8. Saila T, Matilla E, Kaila M, et al. Measuring patient assessments of the quality of outpatient care: a systematic review. *J Eval Clin Pract*. 2008; 14:148–154. [PubMed: 18211659]
9. Ware JE Jr, Snyder MK, Wright WR, et al. Defining and measuring patient satisfaction with medical care. *Eval Program Plann*. 1983; 6:247–263. [PubMed: 10267253]
10. Williams B. Patient satisfaction: a valid concept? *Soc Sci Med*. 1994; 38:509–516. [PubMed: 8184314]
11. Attkisson CC, Zwick R. The client satisfaction questionnaire. Psychometric properties and correlations with service utilization and psychotherapy outcome. *Eval Program Plann*. 1982; 5:233–237. [PubMed: 10259963]
12. Chow A, Meyer E, Darzi A, Athanassion T. Patient-reported outcome measures: the importance of patient satisfaction in surgery. *Surgery*. 2009; 146:435–443. [PubMed: 19715800]
13. Gold DT, McClung B. Approaches to patient education: emphasizing the long-term value of compliance and persistence. *Am J Med*. 2006; 119:S32–S37. [PubMed: 16563940]
14. Tomlinson JS. Patient satisfaction: an increasingly important measure of quality. *Ann Surg Oncol*. 2006; 13:764–765. [PubMed: 16614874]

15. Borras JM, Sanchez-Hernandez A, Navarro M, et al. Compliance, satisfaction, and quality of life of patients with colo- rectal cancer receiving home chemotherapy or outpatient treatment: a randomised controlled trial. *BMJ*. 2001; 322:826. [PubMed: 11290635]
16. Lipscomb J, Reeve BB, Clauser SB, et al. Patient-reported outcomes assessment in cancer trials: taking stock, moving forward. *J Clin Oncol*. 2007; 25:5133–5140. [PubMed: 17991933]
17. Jean-Pierre P, Fiscella K, Freund KM, Clark J, Darnell J, Holden A, Post D, Patierno SR, Winters PC. Patient Navigation Research Program Group. Structural and Reliability Analysis of a Patient Satisfaction with Cancer-Related Care Measure: A Multi-Site Patient Navigation Research Program Study. *Cancer*. 2011; 117(4):854–861. [PubMed: 20922802]
18. Jean-Pierre P, Fiscella K, Winters PC, Post D, Wells KJ, McKoy JM, Battaglia T, Simon M, Kilbourn K, Taylor E, Roetzheim RG, Patierno SR, Dudley DJ, Raich PC, Freund KM. Psychometric Development and Reliability Analysis of a Patient Satisfaction with Interpersonal Relationship with Navigator Measure: A Multi-Site Patient Navigation Research Program Study. *Psycho-Oncology*. Jun 17.2011 Doi: 1002/PON.2002. [Epub ahead of print]. PMID: 21681995.
19. Jean-Pierre P, Fiscella K, Winters PC, Paskett E, Wells K, Battaglia T, the Patient Navigation Research Program group. Cross-cultural Validation of the Patient Satisfaction with Interpersonal Relationship with Navigator Measure. *Psycho-Oncology*. Jul 1.2011 Doi: 10.1002/pon.2018. [Epub ahead of print]. PMID 21726018.
20. Baghaei P. A comparison of three polychotomous Rasch models for super-item analysis. *Psychological Test and Assessment Modeling*. 2010; 52:313–322.
21. Samejima, F. The graded response model. In: van der Linden, WJ.; Hambleton, R., editors. *Handbook of modern item response theory*. Springer; New York, NY: 1996. p. 85-100.
22. Bond, TG.; Fox, CM. *Applying the Rasch Model – Fundamental Measurement in the human sciences*. 2ed. Lawrence Erlbaum Associates, Publishers; Mahwah, NJ: 2007.
23. Embretson, SE.; Reise, SP. *Item Response Theory for Psychologists*. Lawrence Erlbaum; Mahwah, NJ: 2000.
24. Cheng Y, Yuan K-H, Liu C. Comparison of reliability measures under factor analysis and item response theory. *Educational and Psychological Measurement*. 2012; 72:52–67.
25. Engelhard, G.; Wilson, M., editors. *Objective measurement III: Theory into practice*. Ablex; Norwood, NJ: 1996.
26. Santor DA, Ramsay JO, Zuroff DC. Nonparametric item analyses of the Beck Depression Inventory: Evaluating gender item bias and response option weights. *Psych Assess*. 1994; 6:255–270.
27. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M, PROMIS Cooperative Group. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Med Care*. May; 2007 45(5 Suppl 1):S3–S11. [PubMed: 17443116]
28. Reeve BB. Item response theory modeling in health outcomes measurement. *Expert Rev Pharmacoecon Outcomes Res*. Apr; 2003 3(2):131–45. doi: 10.1586/14737167.3.2.131. [PubMed: 19807361]

Figure 1.
Item Response Category Characteristic Curves (CCC) for the 18 Items

Figure 2.
Item Operational Characteristic Curves (OCC) for the 18 Items

Figure 3. Item Information Curves (Left Panel), Test Information Curve (Upper Right Panel) and Distribution of the Latent Satisfaction Level (Bottom Right Panel).

Table 1**Demographic and Clinical Characteristics of 1,296 Participants**

Characteristic	No.	Mean (SD)
Age (18-86y)	1,296	50.6(14.0)
	No.	%
Cancer site		
Breast	875	67.6
Cervix	96	11.8
Colorectal	107	10.5
Prostate	112	9.5
Multiple concurrent cancer sites	4	0.6
Sex		
Female	950	84.2
Male	178	15.8
Race/ethnicity		
White	180	13.9
Black/African American	277	21.4
Asian	238	18.4
American Indian/Alaska Native	3	.2
Hispanic or Latino	210	16.2
Other	16	1.2
Primary language		
English	911	80.9
Spanish	191	17.0
Other	24	2.1
Birth country		
United States	801	76.1
Other	252	23.9
Marital status		
Single/never married	338	30.3
Married/living as married	473	42.4
Divorced/separated	236	21.2
Widowed	68	6.1
Education		
8th grade or less	118	11.4
Some high school	138	13.3
High school diploma/GED	254	24.5
Some college/Vocational school	236	22.7
Associate degree	75	7.2
College graduate	133	12.8
Graduate or professional degree	84	8.1
Household income		

Characteristic	No.	Mean (SD)
Less than \$10,000	286	30.2
\$10,000 to \$19,999	200	21.1
\$20,000 to \$29,999	124	13.1
\$30,000 to \$39,999	87	9.2
\$40,000 to \$49,999	46	4.9
\$50,000 or more	205	21.6
Employment status		
No current employment	588	46.7
Part-time employment	151	12.0
Full-time employment	321	25.5
Health insurance coverage		
Yes	909	82.6
No	191	15.2

Table 2

Item Parameter Estimates from the GRM and Their Standard Errors

Item	β_1 (SE)	β_2 (SE)	β_3 (SE)	β_4 (SE)	α (SE)
1	-2.441 (.118)	-1.893 (.158)	-1.593 (.146)	-0.138 (.110)	1.582 (.144)
2	-2.606 (.137)	-2.187 (.195)	-1.975 (.186)	-0.388 (.138)	1.634 (.163)
3	-2.584 (.132)	-1.873 (.172)	-1.564 (.160)	-0.101 (.128)	1.702 (.157)
4	-2.571 (.127)	-1.692 (.153)	-1.460 (.145)	-0.044 (.115)	1.456 (.126)
5	-2.384 (.112)	-1.613 (.137)	-1.326 (.127)	0.086 (.097)	1.467 (.127)
6	-2.186 (.096)	-1.449 (.126)	-1.276 (.120)	0.087 (.088)	1.705 (.152)
7	-2.108 (.089)	-1.569 (.130)	-1.354 (.122)	0.067 (.087)	1.995 (.178)
8	-2.710 (.146)	-1.781 (.151)	-1.647 (.145)	0.301 (.097)	0.939 (.087)
9	-2.410 (.114)	-1.455 (.125)	-1.260 (.118)	0.409 (.088)	1.178 (.098)
10	-2.426 (.115)	-1.617 (.138)	-1.280 (.128)	0.452 (.097)	1.465 (.120)
11	-2.050 (.085)	-1.425 (.123)	-1.252 (.118)	0.268 (.079)	2.181 (.192)
12	-2.337 (.109)	-1.511 (.130)	-1.300 (.123)	0.258 (.092)	1.550 (.129)
13	-2.230 (.096)	-1.628 (.133)	-1.422 (.127)	0.266 (.087)	1.748 (.146)
14	-2.161 (.092)	-1.604 (.138)	-1.429 (.132)	0.056 (.098)	2.175 (.188)
15	-2.089 (.087)	-1.536 (.128)	-1.298 (.120)	0.093 (.086)	2.022 (.178)
16	-2.494 (.125)	-1.729 (.153)	-1.400 (.142)	0.144 (.110)	1.507 (.133)
17	-2.299 (.108)	-1.653 (.145)	-1.315 (.135)	0.056 (.106)	1.946 (.170)
18	-2.441 (.182)	-1.893 (.197)	-1.593 (.164)	-0.138 (.141)	1.582 (.092)

Table 3**Patient Satisfaction with Cancer Care (PSCC) Scale Items**

-
1. I felt that my health concerns were understood.
 2. I felt that I was treated with courtesy and respect.
 3. I felt included in decisions about my health.
 4. I was told how to take care of myself.
 5. I felt encouraged to talk about my personal health concerns.
 6. I felt I had enough time with my doctor.
 7. My questions were answered to my satisfaction.
 8. Making an appointment was easy.
 9. I knew what the next step in my care would be.
 10. I feel confident in how I deal with the health care system.
 11. I was able to get the advice I needed about my health issues.
 12. I knew who to contact when I had a question.
 13. I received all the services I needed.
 14. I am satisfied with the care I received.
 15. The doctors seemed to communicate well about my care.
 16. I received high-quality care from my regular doctor.
 17. I received high-quality care from my specialists.
 18. My regular doctor was informed about the results of the tests I got.
-