Assessing the invariance of a culturally competent multi-lingual unmet needs survey for immigrant and Australian-born cancer patients: a Rasch analysis

McGrane, J. A.[1], Butow, P. N.[2], Sze, M.[2], Eisenbruch, M.[3], Goldstein, D.[4], King, M. T.[2], on behalf of the Psycho-Oncology Co-operative Research Group (PoCoG) CALD team*.

1. Pearson Psychometric Laboratory, Faculty of Education, The University of Western Australia, Perth, WA, Australia
2. Psycho-Oncology Co-operative Research Group (PoCoG),The University of Sydney, Sydney, NSW, Australia
3. School of Psychology and Psychiatry, Monash University, Melbourne, VIC, Australia
4. Prince of Wales Clinical School, Faculty of Medicine, University of New South Wales, Sydney, NSW, Australia

* The PoCoG CALD team also includes: Ms Lynley Aldridge, Dr Melanie Bell, Dr Priya Duggal, Prof Afaf Girgis, A/Prof Michael Jefford, A/Prof Penny Schofield, A/Prof Bettina Meiser, Prof Elizabeth Lobb and Prof Patsy Yates.

Short title: Invariance of a culturally competent multi-lingual unmet needs survey

Corresponding author: Dr Joshua McGrane, Pearson Psychometric Laboratory, Faculty of Education, M428, The University of Western Australia, 35 Stirling Hwy, Crawley WA, 6009 AUSTRALIA

Email: joshua.mcgrane@uwa.edu.au; ph.: 61 8 6488 1456; fax: 61 8 6488 1052

Abstract

Purpose: To assess the invariance of a culturally competent multi-lingual unmet needs survey.

Methods: A cross-sectional study was conducted among immigrants of Arabic, Chinese and Greek-speaking backgrounds, and Anglo-Australian-born controls, recruited through Cancer Registries (*n = 591*) and oncology clinics (*n = 900*). The survey included four subscales, with newly developed items addressing unmet need in culturally competent health information and patient support (CCHIPS), and items adapted from existing questionnaires addressing physical and daily living (PDL), sexuality (SEX) and survivorship (SURV) unmet need. The measure was translated into Arabic, Chinese and Greek. Rasch analysis was carried out on the four domains.

Results: While many items were mistargeted to less prevalent areas of unmet need, causing substantial floor effects in person estimates, reliability indices were all acceptable. The CCHIPS domain showed differential item functioning (DIF) for cultural background and language and the PDL domain showed DIF for treatment phase and gender. The results for SEX and SURV domains were limited by floor effects and missing responses. All domains showed adequate fit to the model after DIF was resolved and a small number of items were deleted.

Conclusions: The study highlights the intricacies in designing a culturally competent survey that can be applied to culturally and linguistically diverse groups across different treatment contexts. Overall, the results demonstrate that this survey is somewhat invariant with respect to these factors. Future refinements are suggested to enhance the survey's cultural competence and general validity.

Identification of the unmet supportive care needs of cancer patients and survivors informs appropriate resource allocation and optimisation of service provision relative to actual needs [1-4]. To date, such assessment has focused on non-immigrant cancer patients. Immigrant groups diagnosed with cancer are worse off than comparable non-immigrant groups in terms of survival [5], psychological morbidity and quality of life [6]. These disparities may arise from a lack of familiarity with the healthcare system, lack of culturally appropriate information, communication difficulties for culturally and linguistically diverse (CALD) patients, and cultural attitudes and beliefs toward cancer [7-9]. Thus there is an imperative to develop and validate a questionnaire to assess unmet needs in immigrant cancer patients and survivors [2, 10, 11].

Existing unmet needs questionnaires address themes that may be common to immigrant and non-immigrant cancer survivors, including needs related to the health system and information, physical and daily living, patient care and support, sexuality, and survivorship [1, 4, 7]. However, immigrants may have unique needs as a result of the issues outlined above [7]. Furthermore, the interpretation of items and scores must remain invariant across diverse languages and cultural groups to enable valid comparisons.

Rasch analysis provides a robust means of assessing measurement invariance across subgroups of respondents, including CALD groups, and has been used to assess cross-cultural validity of health outcomes assessments [12, 13]. In broad terms, Rasch analysis enables the rigorous investigation of various assumptions of traditional scaling approaches, including the fit of a set of item responses to a unidimensional model, targeting of a multi-item scale to a particular population, functioning of response categories/scales, and differential item functioning (DIF). Pallant and Tennant [14] provide an introduction to the Rasch model and Andrich [15] provides a technical overview.

DIF analysis is particularly informative for cross-cultural validation of health outcomes measures [16, 17]. DIF occurs when people of the same level of a latent construct have a substantially different response probability for a specific item as a result of a group factor(s). For example, in a multi-item survey designed to assess patient information needs, non-fluent English speakers may find it systematically easier to endorse a high level of unmet need for interpreter services than fluent English speakers, even if they have the same or similar overall level of unmet need as estimated across all the information need items. Significant cultural/linguistic DIF has been identified by Rasch analysis in the cross-cultural validation of a number of commonly used health outcomes measures [12, 13, 18-20]. This highlights the importance of rigorously testing for DIF, as undiagnosed DIF may lead to biased item and person estimates between different cultural/language groups. This may then confound inferences made using a questionnaire, including cross-cultural comparisons [21].

Moreover, unlike scale-level analyses that have traditionally been used for cross-cultural validation, e.g., Factor Analysis, Rasch analysis provides much more fine-grained diagnostic information, as it is sensitive to item-level biases [22, 23].

Aims

This aim of the analyses reported in this paper was to assess the invariance of a multi-lingual survey of the unmet health-care needs of immigrant cancer patients. We used Rasch analysis to examine invariance across four factors: Immigrant vs. Australian-born; questionnaire language; cohort (current treatment vs. survivor); and, gender.

Methods

Participants

Participants were first-generation immigrants to Australia from Arabic, Greek or Chinese speaking backgrounds, plus Australian-born controls. Eligible patients were aged 18 years or over and diagnosed with a new, primary or secondary, histologically confirmed cancer comprising one of the top 12 most common cancers by incidence (all stages). Immigrant participants were born in a country where Chinese, Greek or Arabic is spoken. Two cohorts were recruited from two key phases of cancer: 1) a treatment cohort was recruited through 16 Oncology clinics in New South Wales, Northern Territory and Victoria between 2010 and 2012, diagnosed and treated with cancer within 12 months of recruitment; 2) a survivorship cohort was recruited through State Cancer Registries in New South Wales, Queensland and Victoria between 2009 and 2011, diagnosed between 2004 and 2007. Detailed recruitment procedures are described elsewhere [24]. Participant characteristics are given in Table 1.

Questionnaire construction

To inform the construction of a culturally competent survey of the health-care needs of immigrant cancer patients, we conducted a qualitative study [7], and reviewed original research on the unmet needs of immigrants with cancer. We then assessed available unmet needs surveys, and selected the Supportive Care Needs Survey (SCNS) and the Cancer Survivors Unmet Needs survey (CaSUN), as the most suitable basis for developing a culturally competent assessment that covered the needs expressed by our CALD groups. A 28-item immigrant version and a 24-item Australian-born version of the questionnaire were constructed covering four domains (see Table 2 for content of both versions).

The physical and daily living (PDL) (7 items) and sexuality (SEX) (3 items) domain items were adapted from the SCNS [1]. The 4 survivorship (SURV) domain items were adapted from the CaSUN, and were only included for the survivor cohort. Fourteen novel items were developed based upon our qualitative study [7]; these formed a new domain that we called the "culturally competent health system, information and patient support" domain (CCHIPS). This included 4 items specific to linguistically diverse immigrants which addressed their need for interpreter services and health professionals that speak their native language. These items were shown to have good Cronbach's alpha coefficients and an Exploratory Factor Analysis supported a single-factor structure [24]. The questionnaire was translated into Arabic, Chinese and Greek using forward-backward translation procedures [25]. Translated Questionnaires were field tested and revised accordingly [26].

The item response format was based on the SCNS, with expanded descriptors of the response categories to improve comprehension. The 5-point rating scale had two levels of description, presence of need and degree of unmet need, for each response option (see Table 2).

Procedure

For the survivor cohort, eligible participants were identified by the State Cancer Registries, who then approached the participants' doctors and then the participants themselves requesting their participation. Once the registry received consent, the details of eligible participants were released to the research team and the participants were then invited by phone/mail to participate in the study [25]. For the treatment cohort, potential participants were identified by doctors in participating Oncology clinics. Immigrant participants had the option to complete in their native-language or in English (giving 7 culture-by-language groups in our dataset: 3 immigrant by native-language, 3 immigrant by English, 1 control). All participants were provided with a language appropriate study package which were completed at home and mailed to the researchers. The study package included a comprehensive survey, including this unmet needs survey, as well as other psychosocial outcomes assessments and additional questions about the respondents demographics, disease and treatment status.

Data Analysis

The Rasch analysis was performed on the subscales using RUMM2030 [27]. Specifically, the partial-credit model was applied, which was justified by a statistically significant likelihood ratio test ($p < .001$) between the partial-credit and rating-scale parameterisations for each subscale, with Weighted Maximum Likelihood estimation. For the CCHIPS and PDL domains, this analysis was carried out on a reduced sample size (N = 623), whereby the six larger culture-by- language groups were randomly sampled down to the size of the smallest culture-by-language group (Arabic immigrants who responded in Arabic, $n = 89$). This reduction was performed to obtain item estimates that were relatively unbiased toward any culture-by-language group(s) in the sample, as if bias was present for an item, the Australian-born group (or Chinese for the immigrant-specific items) would dominate the estimate as a result of the effect of their substantially larger sample size (see Table 1) on the estimation procedure [15]. Furthermore, both the chi-square tests of model fit and the Analysis of Variance (ANOVA) used to identify items with DIF are known to be highly sample-size dependent, and so this sample-size reduction helped reduce that concern whilst maximising the information of the considerably smaller Arabic and Greek groups [14, 15]. The overall reduced sample size exceeds sample-size recommendations for the robust estimation of polytomous items and the culture-by-language subsample sizes are appropriate for robust exploratory estimation [28, 29]. There were no substantial differences in the characteristics of the random subsample and the full sample of the Chinese, Greek and Australian-born groups presented in Table 1. This sample-reduction procedure was not followed for two domains: SEX (33% of respondents had all missing values and 61% of the remaining had the minimum score); SURV (35% had all missing responses and 42% of remaining had the minimum score) domains due to their already diminished sample sizes. Given these high levels of missing and extreme responses, results for the SEX and SURV domains are interpreted with caution.

Items exhibited good fit to the Rasch model if they met a number of criteria, including: good targeting supported by graphical inspection of the person-item threshold distribution and a Person Separation Index (PSI) greater than .7, which supports the use of the subscale for group comparisons [14]; properly ordered categories reflected by the ordering of the Rasch threshold estimates; low absolute values between ±2.5 for the standardised item fit residuals, which were supported by graphical inspection of the Item Characteristic Curves (ICCs); values less than +2.5 for at least 95% of respondents for the standardised person fit residuals, indicating that the overwhelming majority of response patterns were sufficiently consistent with the modelled item ordering; non-statistically significant chi-square tests of item fit after Bonferroni adjustment for multiple comparisons; absence of local and trait dependence (multidimensionality) in the person-by-item residuals, with residual correlations less than .3 and less than 5% of the sample having significantly different individual person estimates on item subsets identified by the Principal Component Analysis (PCA) of the residuals [14]. Items that showed significant dependence were combined into a testlet, i.e., they were added together to create a single, polytomous item, and refitted to the model with all other items to observe the effect on the PSI, as a reduction in reliability using this procedure is further evidence of substantial dependence between items [30].

Finally, fit was assessed with respect to DIF following the iterative procedure recommended by Andrich and Hagquist [31] to distinguish between real and artificial DIF. Artificial DIF is not attributable to a substantive source(s), but rather is an artefact of model estimation when real DIF is present. If an item with real DIF is resolved, real DIF in other items will remain apparent, but artificial DIF will diminish. Therefore, firstly, the largest, real DIF effect was graphically and statistically identified. This DIF was resolved by splitting the item across the differential groups, i.e., it became a unique item for each group. Once resolved, the procedure was repeated until no more real DIF was detected. Once all DIF was resolved, the overall Chi-square values were inspected as a final check of fit.

Results

CCHIPS domain

The PSI was .83, providing evidence of good targeting and sufficient power for other tests of fit. However, the person-item threshold distribution (Figure 1) indicated that the items collectively were somewhat mistargeted toward persons with higher levels of unmet need, with a person location mean of -1.62 logits (*SD* = 1.70). All items except CCHIPS-2 had disordered thresholds. For these 13 items, the third category was never the most probable response for any level of the latent construct (Figure 2), and was collapsed with the fourth category, which resolved the disordering and increased the PSI to .84.

The standardised item residuals revealed some misfitting items (*M* = -.14, *SD* = 2.25), particularly CCHIPS-3 (res. = 5.16). However, the Item Characteristic Curve (ICC) showed that only 3 of the 10 class intervals substantially deviated from the model expectations (Figure 3). Therefore, all items were retained for further tests of fit. The standardised person fit residuals also indicated misfit (*M* = -.74, *SD* = 2.10), but only 3% exceeded +2.5.

Significant positive residual correlations were found between three of the immigrant-specific items (CCHIPS-11/12/13), and for the family-focused items (CCHIPS-9/10). These dependent items were combined into testlets and the presence of local dependence was supported by the reduction of the PSI to .82 and the thresholds of the testlets were disordered [32]. The PCA of residuals showed that a number of language/information specific items (CCHIPS-5, 11/12/13 & 14) loaded significantly positively on the first residual component versus a number of negatively loading, non-language specific items (CCHIPS-1, 3, 8, 9 & 10). However, only 4.9% (95% CI: 3.4-6.8%) of paired t-tests between estimates from the subsets were statistically significant.

The greatest, significant uniform-DIF was found for CCHIPS-14 between immigrant non-English responders and immigrant English responders (Figure 4). After resolving this item, the next largest DIF was observed for CCHIPS-5 between immigrant non-English responders and both immigrant and Australian-born English responders. Significant uniform-DIF was then observed for the CCHIPS-11/12/13 testlet between Immigrant non-English responders and immigrant English responders. After resolving this item, significant uniform-DIF remained for CCHIPS-8 between Chinese language responders and all other culture-by-language groups. The same pattern of uniform-DIF was then observed for CCHIPS-6. Further uniform-DIF was observed for the resolved, English-language CCHIPS-5 between Australian-born responders and immigrant English responders. Significant uniform-DIF was then found for CCHIPS-7 between Australian-born responders and immigrant responders. Finally, CCHIPS-3 showed uniform DIF between English-language responders and non-English responders. No uniform DIF was observed for the gender or stage of treatment person factors and no significant, non-uniform DIF was observed (see Table 3 for DIF statistics and specific group estimate differences).

Fit was then re-examined for the new item and person estimates. The overall chi-square test of fit showed misfit to the model, but when the four most misfitting items were removed, adequate fit to the model was achieved ($\chi^2(60) = 86.08$, $p = .02$) and the PSI only marginally reduced to .78. The final location estimates and fit statistics are presented in Table 3.

PDL domain

While the PSI was good at .80, the person-item threshold distribution (Figure 6) showed a marked floor effect (23% of respondents at the minimum score). The items were somewhat mistargeted toward persons with higher levels of unmet need, with person location mean of -1.59 logits (*SD* = 1.79).

PDL-1 and PDL-5 had disordered thresholds. Collapsing the third and fourth categories resolved this and PSI increased to .81. PDL-4 and PDL-5 had residuals that exceeded the ±2.5 threshold (-4.20 and 3.95 respectively), but the ICCs showed that the observed responses only moderately deviated from the model expectations at the very top and bottom of the person distribution (Figure 7). Given the ambiguous findings regarding their fit, and as we aimed to maintain as many items as possible to maximise content coverage and to avoid capitalising on chance when re-testing the model after the post-hoc deletion of items, these items were maintained for further examinations of model fit. The standardised person fit residuals also indicated some misfit (*M* = -.57, *SD* = 1.43), but less than 1% exceeded +2.5.

No residual correlations exceeded the acceptable threshold, but there was evidence of minor violations of trait dependence. Specifically, the more functional items (PDL-5, 6 & 7) loaded positively on the first residual factor versus the more physical items (PDL-1, 2, 3 & 4) which loaded negatively. However, only 3.4% (95% CI: 2.2-5.1%) of paired t-tests between the subsets' estimates were statistically significant.

The greatest, significant uniform-DIF was found for PDL-3 between the treatment and the survivorship cohorts (Figure 8). After resolving this item, the next largest uniform-DIF was observed for PDL-4 between the treatment and the survivor cohorts. Finally, uniform-DIF was observed for PDL-6 between females and males. No uniform-DIF was observed for the cultural group or questionnaire language person factors and no non-uniform-DIF was observed (see Table 3 for DIF statistics and specific group estimate differences).

After resolving DIF, the overall chi-square test of fit showed initial misfit to the model, but this became non-significant with the deletion of PDL-4 Treatment cohort resolved item ($\chi^2(72) = 93.54$, $p = .05$) and the PSI only marginally reduced to .78. The final item location estimates and fit statistics are presented in Table 3.

SEX domain

While the PSI of .69 indicated reasonable separation, the person location mean of -3.12 (*SD* = 2.32) indicated that these items were quite mistargeted toward persons with higher levels of unmet need (Figure 9). None of the items showed disordered thresholds. Inspection of the standardised item fit residuals revealed no misfit (*M* = -.65, *SD* = 2.11). No standardised person fit residuals exceeded the +2.5 threshold (*M* = -1.78, *SD* = 2.03).

Inspection of the item residual correlations revealed large, negative correlations between SEX-3 and the other two items, but these values must be interpreted with caution given the small number of items, and for similar reasons, the PCA of residuals was not interpreted. No significant DIF was detected for any of the person factors. The overall chi-square test of fit revealed good fit to the model, supporting the invariance of the ordering of these items along the different levels of the latent construct, $\chi^2(9) = 15.56$, $p = .07$. The final location estimates and fit statistics are presented in Table 4.

SURV domain

These items were quite mistargeted toward higher levels of survivorship unmet need, with a PSI of .52 and mean person estimate of -2.87 (*SD* = 2.01) (Figure 10). The low PSI undermined the power of other tests of fit. All items had disordered thresholds, similar to the CCHIPS and PDL domains. Once the third and fourth categories were collapsed, the thresholds were properly ordered and the PSI increased to .61. The standardised item fit residuals did not reveal any significant misfit. No standardised person fit residuals exceeded the acceptable threshold.

Inspection of the item residual correlations revealed large, negative correlations between SURV-3, but again this was confounded by the small number of items. No DIF was detected for any of the person factors. The overall chi-square test of fit revealed marginal fit to the model, $\chi^2(12) = 25.96$, $p = .01$. The final location estimates and fit statistics are presented in Table 4.

Discussion

Rasch analysis revealed a number of problems, including mistargeting, disordered response categories, item-dependencies, and a range of DIF across various person factors. Once these anomalies were accounted for, all domains showed reasonable fit to the model with minimal deletion of misfitting items, providing some support for the psychometric validity of the four domains.

The new culturally-centred items in the CCHIPS domain showed significant variation in terms of questionnaire language and cultural grouping. As might be expected, the people that responded in non-English generally found these items easier to indicate as unmet needs, which is consistent with previous findings on the patient reported outcomes of bi-lingual cancer patients, whereby item bias was more associated with linguistic rather than cultural grouping [33]. From a measurement perspective, including items with DIF can either exaggerate or diminish mean differences between groups, which in this case resulted in inflated difficulty estimates for the non-English responders. This may be avoided by giving the language-specific items to only non-English responders and using the Rasch person estimates which can take account of systematic missing data, rather than total scores [34]. Alternatively, some questions could be revised to have a common, easily translated phrasing across all groups. For example, CCHIPS-5 could be changed to, "Written information in your *preferred* language" and given to both immigrant and non-immigrant responders alike. A similar strategy could be applied to CCHIPS-7.

CCHIPS-3 ("'alternative' medicines *or medicines from your culture*") was included as a result of earlier consultation with Chinese-background patients [7], thus this item could be included to obtain information from this group, but not included in the overall domain score when making cross-cultural comparisons. Such a strategy could be used generally when researchers want to include group specific items to increase content validity for particular groups. Moreover, Chinese-language responders found the practical support (CCHIPS-8) and religious representative (CCHIPS-6) items significantly more difficult to endorse as unmet needs, which is consistent with the cultural differences in emphases on types of unmet healthcare need identified in the qualitative development of this domain [7]. It may be that the Chinese-language responders actually had greater support of this kind from their communities than did the other groups. However, given this DIF was language specific, an alternative explanation is that the Chinese translation had a somewhat different interpretation to the other language versions, despite rigorous forward and backward translation, as translation issues have commonly been found to be the source of DIF in patient-reported outcome assessments [35, 36]. Future qualitative research should investigate source(s) of bias in these items for Chinese language responders.

The results for the PDL domain items were far more positive in terms of cultural and/or linguistic invariance, but the DIF observed for the nausea (PDL-3) and illness (PDL-4) items for stage of treatment was somewhat unsurprising since nausea as a specific unmet healthcare need in the cancer context is intertwined with treatment, e.g., chemotherapy [37]. The magnitude of the difference in the resolved item estimates for the nausea/illness items suggests they should not be included in a PDL domain total score when comparing survivor and treatment cohorts. The housework item (PDL-6) showed DIF by gender, which is sensible given that Australian females tend to play a greater role in this daily task [38], but the magnitude of the difference between the resolved items was relatively small.

Additional insights for future refinements of this survey, and for the development of other cross culturally valid assessments, include the need for improved item targeting and the alleviation of substantial floor effects, particularly for the sexuality and survivorship domains. Such floor effects are not uncommon in unmet needs surveys and may confound inferences and comparisons [4, 39]. Future applications should attempt to include items which are more prevalent areas of unmet need. Moreover, the response scale was found to be problematic, particularly for the CCHIPS and SURV domains, as respondents did not consistently discriminate between the “needed a little more help” and “needed some more help” response categories. Combining these in future versions of the survey may enhance the validity of the response scale by simplifying the response structure for participants.

Item dependencies were observed for the cultural items concerned with spoken language and family support. From a psychometric perspective, this level of redundancy could be addressed in future versions by removal of dependent items, or by presenting out of succession to encourage more independent engagement with them. However, this would not overcome the obvious logical dependence between, in particular, the spoken language items, and may compromise the content validity if all three items matter to patients.

The prevalence of missing values and floor effects for SEX and SUR domains only permit tentative conclusions, but these initial findings support the invariance of these items across the range of person factors for both Immigrant and Australian-born responders. The marginal fit of the survivorship domain items to the model may be attributable to the fact that they were selected from an existing instrument without considering thematic content. For the sexuality domain, the prevalence of missing data and floor effect may be attributable to people's, and particularly immigrants' unwillingness to share such sensitive information [25, 40].

Consideration of the differences in person estimates obtained before and after resolving the various cases of DIF goes beyond the scope of the current manuscript. In practical terms, the fine-grained, item-level analysis employed in the present study, and the particular cases of DIF that have been observed, may make minimal difference to the group-level comparisons and inferences that psycho-oncology researchers are interested in [41]. In a future paper, we will examine this issue using the unresolved and resolved person estimates from the present analysis.

In summary, the disparities in physical and psychosocial outcomes for immigrant cancer patients versus non-immigrants highlight the need for assessments that are both culturally appropriate and psychometrically sound [6]. Our findings are consistent with others that have highlighted the intricacies in designing a multi-lingual, culturally competent survey that can be applied in a culturally diverse setting across different treatment contexts [16, 35]. Overall, our results demonstrate that this unmet needs survey is somewhat invariant with respect to these factors. Rasch analysis identified a number of substantive anomalies in the survey, which could be addressed in future versions of it to enhance the validity of substantive comparisons made with it across immigrant and non-immigrant groups.

References

1. Bonevski, B., Sanson-Fisher, R., Girgis, A., Burton, L., Cook, P., et al. (2000). Evaluation of an instrument to assess the needs of patients with cancer. Supportive Care Review Group. *Cancer, 88*(1), 217-25.
2. Au, A., Lam, W. W., Kwong, A., Suen, D., Tsang, J., et al. (2011). Validation of the Chinese version of the short-form Supportive Care Needs Survey Questionnaire (SCNS-SF34-C). *Psycho-Oncology, 20*(12), 1292-300.
3. Boyes, A., Girgis, A., and Lecathelinais, C. (2009). Brief assessment of adult cancer patients' perceived needs: development and validation of the 34-item Supportive Care Needs Survey (SCNS-SF34). *Journal of Evaluation in Clinical Practice, 15*(4), 602-6.
4. Hodgkinson, K., Butow, P., Hunt, G. E., Pendlebury, S., Hobbs, K. M., et al. (2007). The development and evaluation of a measure to assess cancer survivors' unmet supportive care

needs: the CaSUN (Cancer Survivors' Unmet Needs measure). *Psycho-Oncology, 16*(9), 796-804.
5. Chu, K. C., Miller, B. A., and Springfield, S. A. (2007). Measures of racial/ethnic health disparities in cancer mortality rates and the influence of socioeconomic status. *Journal of the National Medical Association, 99*(10), 1092-1104.
6. Colagiuri, B., King, M. T., Butow, P. N., McGrane, J. A., Luckett, T., et al. (2012). A comparison of the FACT-G and the Supportive Care Needs Survey (SCNS) in women with ovarian cancer: unidimensionality of constructs. *Quality of Life Research, 21*(5), 887-897.
7. Butow, P. N., Sze, M., Duggal-Beri, P., Mikhail, M., Eisenbruch, M., et al. (2011). From inside the bubble: migrants' perceptions of communication with the cancer team. *Supportive Care in Cancer, 19*(2), 281-290.
8. Goldstein, D., Thewes, B., and Butow, P. (2002). Communicating in a multicultural society. II: Greek community attitudes towards cancer in Australia. *Internal Medicine Journal, 32*(7), 289-96.
9. Moore, R. and Butow, P. (2005). Culture and oncology: Impact of context effects. In D. Speigel (Ed.), *Cancer, Communication and Culture*. New York: Kluwer Academic/ Plenum Publishers.
10. Ali, N. S., Khalil, H. Z., and Yousef, W. (1993). A comparison of American and Egyptian cancer patients' attitudes and unmet needs. *Cancer Nursing, 16*(3), 193-203.
11. Guthrie, B. (2005). Bridging Health Care Disparities: Addressing Unmet Needs of Women of Color. *Journal of Obstetric, Gynecologic, & Neonatal Nursing, 34*(3), 385-385.
12. Custers, J. W., Hoijtink, H., van der Net, J., and Helders, P. J. (2000). Cultural differences in functional status measurement: analyses of person fit according to the Rasch model. *Quality of Life Research, 9*(5), 571-578.
13. Tennant, A., Penta, M., Tesio, L., Grimby, G., Thonnard, J.-L., et al. (2004). Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Medical Care, 42*(1), I-37.
14. Pallant, J. F. and Tennant, A. (2007). An introduction to the Rasch measurement model: an example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*, 1-18.
15. Andrich, D. (1988). *Rasch models for measurement*. Beverly Hills: Sage Publications, Inc.
16. Pagano, I. S. and Gotay, C. C. (2005). Ethnic differential item functioning in the assessment of quality of life in cancer patients. *Health and Quality of Life Outcomes, 3*(1), 60.
17. Teresi, J. A. and Fleishman, J. A. (2007). Differential item functioning and health assessment. *Quality of Life Research, 16*(1), 33-42.
18. Tesio, L. and Tennant, A. (2005). Psychometric properties of the Mini-Mental State Examination in patients with acquired brain injury in Turkey. *Journal of Rehabilitation Medicine, 37*, 306-311.
19. Geyh, S., Fellinghauer, B., Kirchberger, I., and Post, M. (2010). Cross-cultural validity of four quality of life scales in persons with spinal cord injury. *Health and Quality of Life Outcomes, 8*, 94.
20. Choi, Y., Mericle, A., and Harachi, T. W. (2006). Using Rasch analysis to test the cross-cultural item equivalence of the Harvard Trauma Questionnaire and the Hopkins Symptom Checklist across Vietnamese and Cambodian immigrant mothers. *Journal of Applied Measurement, 7*(1), 16.
21. Baranik, L. E., Meade, A. W., Lakey, C. E., Lance, C. E., Hu, C., et al. (2008). Examining the Differential Item Functioning of the Rosenberg Self-Esteem Scale Across Eight Countries. *Journal of Applied Social Psychology, 38*(7), 1867-1904.
22. Zumbo, B. D. (2003). Does item-level DIF manifest itself in scale-level analyses? Implications for translating language tests. *Language Testing, 20*(2), 136-147.

23. Salzberger, T., Sinkovics, R. R., and Schlegelmilch, B. B. (1999). Data Equivalence in Cross-Cultural Research: A Comparison of Classical Test Theory and Latent Trait Theory Based Approaches. *Australasian Marketing Journal, 7*(2), 23-38.
24. Butow, P. N., Bell, M., Aldridge, L. J., Sze, M., Eisenbruch, M., et al. (2013). Unmet needs in immigrant cancer survivors: a cross-sectional population based study. *Supportive Care in Cancer, 21*, 2509-2520.
25. Bell, M. L., Butow, P. N., and Goldstein, D. (2013). Informatively missing quality of life and unmet needs sex data for immigrant and Anglo-Australian cancer patients and survivors. *Quality of Life Research*, 1-4.
26. Schuman, H. (1966). The random probe: a technique for evaluating the validity of closed questions. *American Sociological Review*, 218-222.
27. Andrich, D., Sheridan, B., and Luo, G. (2010). Rasch models for measurement: RUMM2030. RUMM Laboratory, Perth, Western Australia.
28. Chen, W.-H., Lenderking, W., Jin, Y., Wyrwich, K., Gelhorn, H., et al. (2014). Is Rasch model analysis applicable in small sample size pilot studies for assessing item characteristics? An example using PROMIS pain behavior item bank data. *Quality of Life Research, 23*(2), 485-493.
29. Linacre, M. (1994). Sample Size and Item Calibration Stability. *Rasch Measurement Transactions, 7*, 328.
30. Marais, I. and Andrich, D. (2008). Effects of varying magnitude and patterns of response dependence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(2), 105-124.
31. Andrich, D. and Hagquist, C. (2012). Real and Artificial Differential Item Functioning. *Journal of Educational and Behavioral Statistics, 37*(3), 387-416.
32. Marais, I. and Andrich, D. (2008). Formalising dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(3), 200-215.
33. Scott, N. W., Fayers, P. M., Aaronson, N. K., Bottomley, A., Graeff, A., et al. (2007). The use of differential item functioning analyses to identify cultural differences in responses to the EORTC QLQ-C30. *Quality of Life Research, 16*(1), 115-129.
34. Andrich, D. (2005). The Rasch model explained. In S. Alagumalai, D. Durtis, and N. Hungi (Eds.), *Applied Rasch Measurement: A book of exemplars* (pp. 308-328). Dordrecht, Netherlands: Springer-Kluwer.
35. McKenna, S. P., Wilburn, J., Thorsen, H., and Brodersen, J. (2012). Adapting Patient-Reported Outcome Measures for Use in New Languages and Cultures. In K.B. Christensen, S. Kreiner, and M. Mesbah (Eds.), *Rasch Models in Health* (pp. 303-315). London UK: Wiley.
36. Petersen, M. A., Groenvold, M., Bjorner, J. B., Aaronson, N., Conroy, T., et al. (2003). Use of differential item functioning analysis to assess the equivalence of translations of a questionnaire. *Quality of Life Research, 12*(4), 373-85.
37. Gershon, R., Cella, D., Dineen, K., Rosenbloom, S., Peterman, A., et al. (2003). Item response theory and health-related quality of life in cancer. *Expert Review of Pharmacoeconomics & Outcomes Research, 3*(6), 783-791.
38. ABS, *Australian Social Trends (4102.0)*, A.B.o. Statistics, Editor 2010, AGPS: Canberra.
39. Hobart, J. C. (2003). Rating scales for neurologists. *Journal of Neurology Neurosurgery and Psychiatry, 74*, 22-26.
40. Flynn, K. E., Jeffery, D. D., Keefe, F. J., Porter, L. S., Shelby, R. A., et al. (2011). Sexual functioning along the cancer continuum: focus group results from the Patient-Reported Outcomes Measurement Information System (PROMIS®). *Psycho-Oncology, 20*(4), 378-386.
41. Borsboom, D. (2006). When does measurement invariance matter? Commentary. *Medical Care, 44*(11), S176-S181.

Figure Captions

Figure 1. Person-item threshold distribution for the CCHIPS domain, showing the distribution of people (above x-axis) relative to the distribution of item thresholds (below x-axis).

Figure 2. Category probability curve for the CCHIPS-4 item (Information about your cancer and its treatment) illustrating that the third response category “needed a little more help” (scored as 2 in the model) was never the most probable response across any level of the latent construct.

Figure 3. Item-Characteristic Curve for the CCHIPS-3 item (Information about “alternative” medicines *or medicines from your culture*) showing the observed proportions (dots) for the different class intervals along the latent construct relative to the model expected values.

Figure 4. Item-Characteristic Curve for the CCHIPS-14 item (*professional interpreter to come with you at doctor visits*), demonstrating the DIF between the English (ENG - circles) and non-English (NONENG - crosses) responding groups.

Figure 5. Item-Characteristic Curve for the non-Australian born resolved CCHIPS-7 item (Support services *in your language)*.

Figure 6. Person-item threshold distribution for the PDL domain.

Figure 7. Item-Characteristic Curve for the PDL-4 item (Feeling unwell).

Figure 8. Item-Characteristic Curve for the PDL-3 (Nausea/vomiting) item demonstrating the DIF between the survivor (S - circles) and current treatment (CT - crosses) cohorts.

Figure 9. Person-item threshold distribution for SEX domain.

Figure 10. Person-item threshold distribution for the SURV domain.

Tables

Table 1.
Participant characteristics by immigrant group and Australian-born controls.

	First-generation immigrants				
	Arabic	Chinese	Greek	Australian-born controls	All
	N=199	N=389	N=255	N=648	N=1491
Males (%)	87 (44)	164 (58)	142 (56)	301 (47)	694 (47)
Mean Age in Years (SD)	60.6 (11.9)	58.9 (12.6)	67.5 (9.5)	63.3 (11.3)	62.5 (11.8)
Responded in native language (%)	89 (45)	264 (68)	134 (53)	N/A	N/A
Mean Years in Australia (SD)	27.0 (11.9)	20.2 (11.1)	45.9 (8.6)	N/A	N/A
Mean Years in Australia for non-English responders (SD)	19.2 (13.7)	17.4 (8.9)	45.3 (8.2)	N/A	N/A
Cohort (%)					
- Survivor	55 (28)	141 (36)	77 (30)	318 (49)	591 (40)
- Treatment	144 (72)	248 (64)	178 (70)	330 (51)	900 (60)
Education (%)					
- Did not complete High School	97 (49)	110 (28)	180 (71)	315 (49)	702 (47)
- High School/Tech College	55 (28)	135 (35)	55 (22)	212 (33)	457 (31)
- University	45 (23)	144 (37)	15 (6)	117 (18)	321 (22)
Cancer type (%)					
- Breast	75 (38)	128 (33)	52 (20)	188 (29)	443 (31)
- Colorectal	35 (18)	60 (15)	27 (11)	120 (19)	242 (17)
- Prostate	26 (13)	39 (10)	36 (14)	98 (15)	199 (14)
- Lung	16 (8)	57 (15)	27 (11)	69 (11)	169 (12)
- Leukaemia, lymphomas	5 (3)	16 (4)	15 (6)	40 (6)	76 (5)
- Head and neck	10 (5)	28 (7)	11 (4)	15 (2)	64 (4)
- Bladder, kidney	6 (3)	7 (2)	10 (4)	33 (5)	56 (3)
- Hepatobiliary	4 (2)	7 (2)	19 (8)	25 (4)	55 (3)
- Other/Unknown	22 (11)	47 (12)	58 (22)	60 (9)	187 (13)
Degree of spread at diagnosis (%)					
- Localized	128 (64)	252 (65)	170 (67)	428 (66)	978 (66)
- Regional/Distant	66 (33)	128 (33)	76 (30)	212 (33)	482 (32)
- Unknown/Missing	5 (3)	9 (2)	9 (3)	8 (1)	31 (2)
Mean months since 1st diagnosis (SD)					
- Survivor	95 (16)	98 (28)	95 (22)	86 (22)	91 (24)
-Treatment	61 (32)	64 (35)	62 (32)	60 (44)	62 (38)

Table 2.

Australian-born and Immigrant versions of the unmet needs survey across the four domains responded to on a 5-point rating scale: first category = “no need; not relevant to me”; second = “no need - satisfied with help received”; third = “some need - needed a little more help”; fourth = “some need - needed some more help”; fifth = “some need - needed a lot more help”.

Domain	Australian-born version, *Immigrant version – additional phrases/items in italics*
Culturally Competent Health System, Information & Patient Support (CCHIPS)	1. Someone to guide you if you have a medical problem
	2. Someone to help you to ask questions
	3. Information about ‘alternative’ medicines *or medicines from your culture*
	4. Information about your cancer and its treatment
	5. Written information *in your language*
	6. A religious representative to talk with
	7. Support services *in your language*
	8. Practical support
	9. Someone to help you to talk with your family about your illness
	10. Someone to help the members of your family if they need support
	11. A general practitioner who speaks your language
	12. A specialist who speaks your language
	13. Other health professionals who speak your language
	14. A professional interpreter to come with you at doctor visits
Physical and Daily Living (PDL)[a]	1. Pain
	2. Lack of energy/tiredness
	3. Nausea/vomiting
	4. Feeling unwell
	5. Not sleeping well
	6. Work around the home
	7. Not being able to do the things you used to
Sexuality (SEX)[a]	1. Changes in sexual feelings
	2. Changes in your sexual relationships
	3. To be given information about sexual relationships
Survivorship (SURV)[b]	1. Moving on with your life
	2. Exploring your spiritual beliefs
	3. Managing your concerns about the cancer coming back
	4. Developing new relationships after your cancer

[a] These items were adapted from the Supportive Care Needs Survey (SCNS)
[b] These items were adapted from the Cancer Survivors Unmet Needs survey (CaSUN). These items were only given to the survivor cohort.

Table 3.

CCHIPS domain resolved items & testlet location estimates (δ), standardised item fit residuals (SFR), chi-square tests of fit (χ^2), and tests of DIF (DIF) broken down by Cultural, Language, Treatment and Gender groups for significant DIF items (items in italics were deleted to improve fit and estimates are based on the pre-deletion analysis). Sample sizes, excluding extreme values, were 527 for the CCHIPS Domain and 474 for the PDL Domain.

Item/Testlet	Grp.[a]	Lang.[b]	δ (*SE*)	SFR	χ^2 (*p*)	DIF
CCHIPS Domain					(*df* = 4)	
CCHIPS-1			-0.29 (.07)	-0.62	14.88 (.005)	
CCHIPS-2			*0.22 (.06)*	*-1.86*	*25.77 (.000)**	
CCHIPS-3		E	-0.25 (.08)	1.49	2.22 (.70)	*F*(1) = 30.83, *p* < .004
CCHIPS-3	Ar C G	NE	0.48 (.10)	0.14	4.55 (.34)	
CCHIPS-4			-0.75 (.07)	0.56	13.43 (.01)	
CCHIPS-5	Ar C G	NE	-0.79 (.09)	1.07	3.21 (.52)	*F*(1) = 29.60, *p* < .004
CCHIPS-5	Au	E	-0.85 (.20)	-0.95	3.58 (.47)	*F*(3) = 5.86, *p* < .004
CCHIPS-5	Ar C G	E	-0.09 (.10)	0.13	3.88 (.42)	
CCHIPS-6	C	C	1.38 (.17)	1.02	8.49 (.08)	*F*(3) = 10.75, *p* < .004
CCHIPS-6		E A G	0.53 (.08)	0.71	5.17 (.27)	
CCHIPS-7	Au	E	-0.71 (.18)	-0.88	4.70 (.32)	*F*(3) = 6.27, *p* < .004
CCHIPS-7	*Ar C G*		*0.17 (.07)*	*-2.78*	*26.98 (.000)**	
CCHIPS-8	C	C	0.99 (.14)	0.49	2.48 (.65)	*F*(3) = 9.81, *p* < .004
CCHIPS-8		E A G	-0.11 (.07)	-0.74	6.21 (.18)	
CCHIPS-9/10			0.12 (.04)	-1.82	4.08 (.40)	
CCHIPS-11/12/13		*E*	*-0.09 (.05)*	*2.17*	*17.73 (.001)**	*F*(1) = 30.83, *p* < .004
CCHIPS-11/12/13	*Ar C G*	*NE*	*-0.39 (.05)*	*1.96*	*15.87 (.003)**	
CCHIPS-14		E	0.72 (.11)	-0.24	4.81 (.31)	*F*(1) = 61.22, *p* < .004
CCHIPS-14	Ar C G	NE	-0.37 (.08)	1.88	4.42 (.35)	
PDL Domain					(*df* = 8)	
PDL-1			-0.08 (.07)	1.07	10.86 (.21)	
PDL-2			-0.48 (.06)	-1.03	14.57 (.07)	
PDL-3	S		1.26 (.16)	0.46	9.58 (.30)	*F*(1) = 18.55, p < .007
PDL-3	CT		0.41 (.07)	1.19	13.00 (.11)	
PDL-4	S		0.30 (.13)	-1.15	5.34 (.72)	*F*(1) = 24.13, *p* < .007
PDL-4	*CT*		*-0.16 (.07)*	*-3.94*	*25.57 (.001)**	
PDL-5			-0.36 (.06)	2.96	8.78 (.36)	
PDL-6	F		-0.37 (.07)	-2.43	8.32 (.40)	*F*(1) = 8.21, *p* < .007
PDL-6	M		-0.09 (.08)	-1.05	12.28 (.14)	
PDL-7			-0.59 (.05)	-1.67	10.79 (.21)	

[a] Ar = Arabic-speaking born; C = Chinese-speaking born; G = Greek-speaking born; Au = Australian born; S = Survivor cohort; CT = Current treatment cohort; F = Female; M = Male.
[b] A = Arabic; C = Chinese; E = English; G = Greek; NE = All non-English languages.

* Significant at Bonferroni adjusted <.004 for CCHIPS domain and <.007 for PDL domain critical levels.

Table 4.
SEX and SURV domain item location estimates (δ), standardised item fit residuals (SFR) and chi-square tests of fit (χ^2). Sample sizes, excluding extreme values, were 367 for the SEX Domain and 223 for the SURV Domain.

Item	δ (*SE*)	SFR	χ^2 (*df* = 3)	*p*
SEX Domain				
SEX-1	-0.22 (.08)	-1.55	6.32	.12
SEX-2	0.09 (.08)	-2.17	6.92	.15
SEX-3	0.13 (.08)	1.75	4.52	.40
SURV Domain				
SURV-1	-0.39 (.12)	-0.49	8.70	.03
SURV-2	1.74 (.13)	-0.72	9.48	.02
SURV-3	-1.59 (.11)	1.03	3.03	.39
SURV-4	0.25 (.12)	-0.45	4.75	.19

Figures

Figure 1.

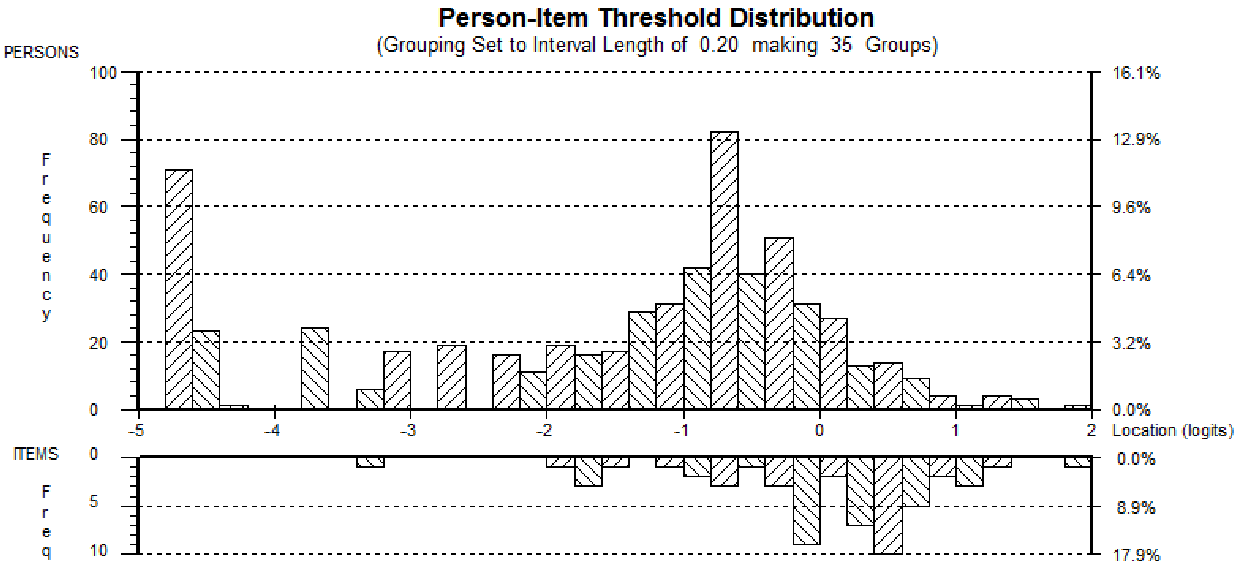


Figure 2.

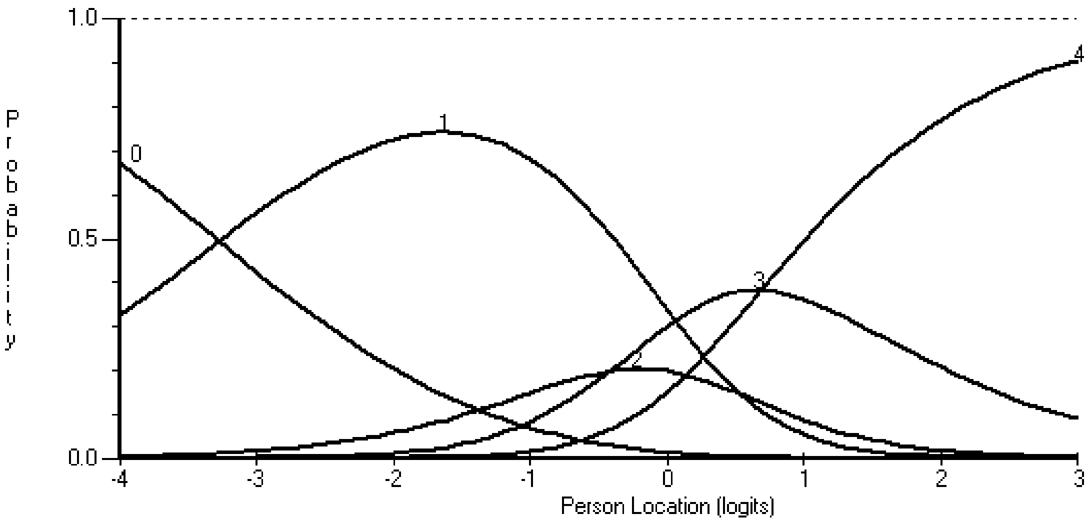


Figure 3.

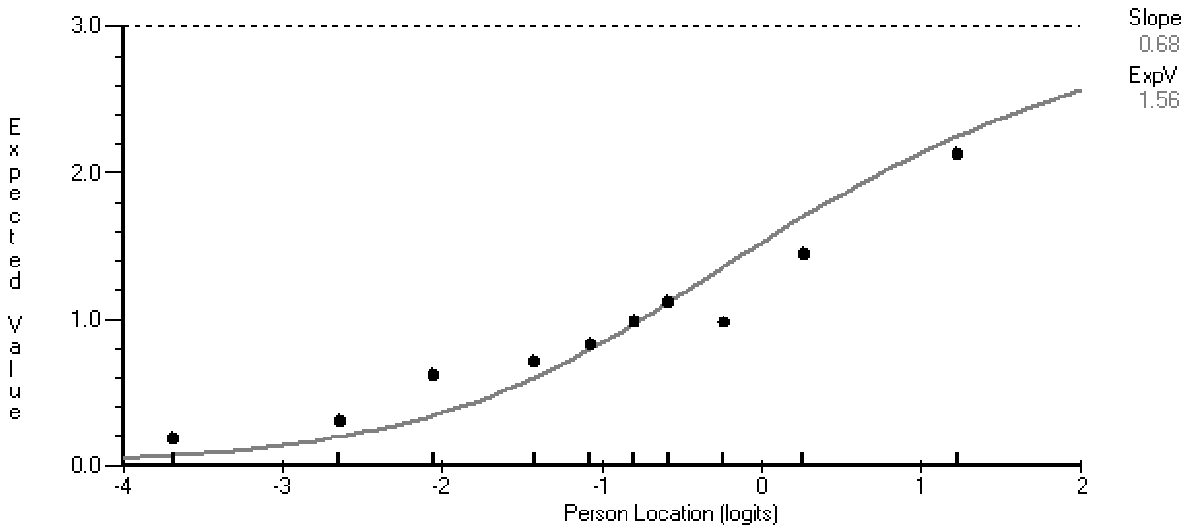


Figure 4.

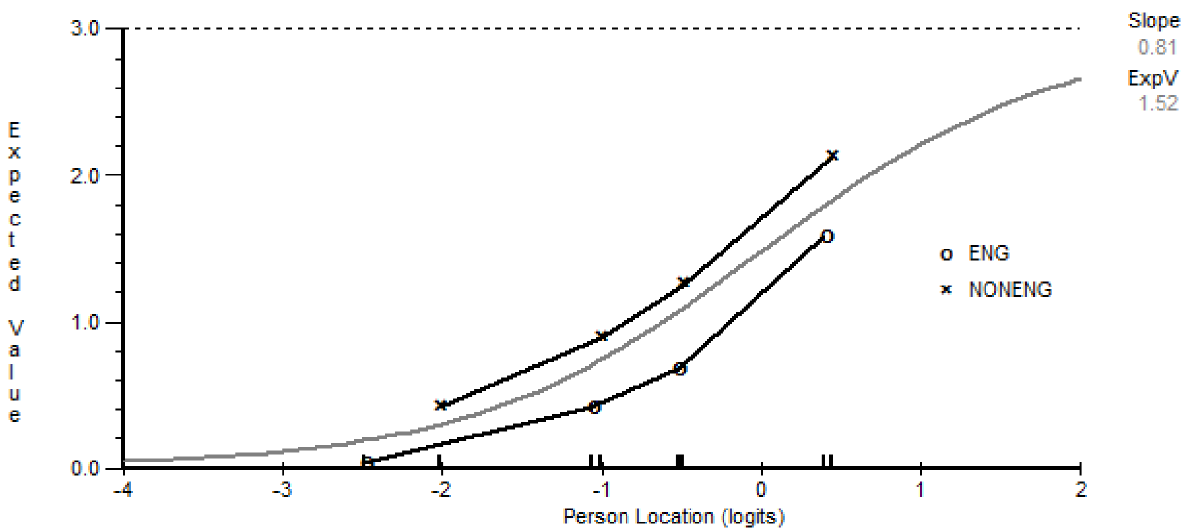


Figure 5.

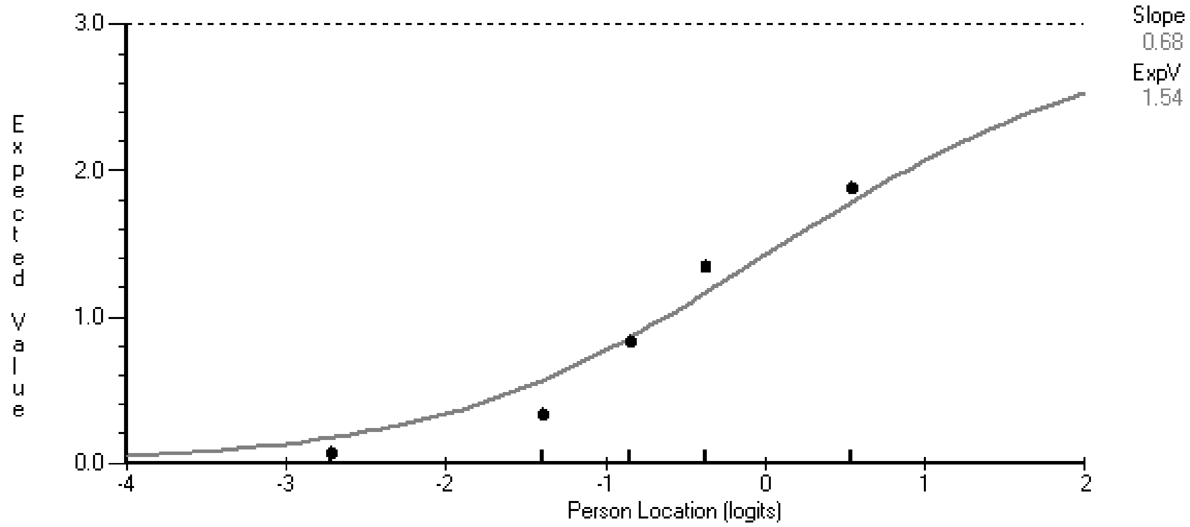


Figure 6.

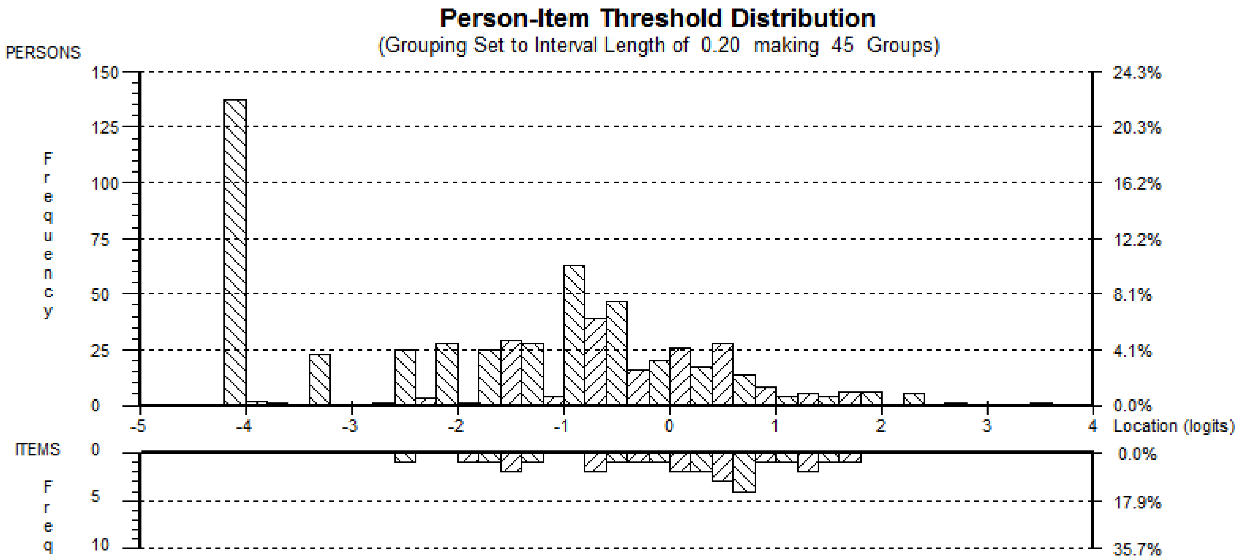


Figure 7.

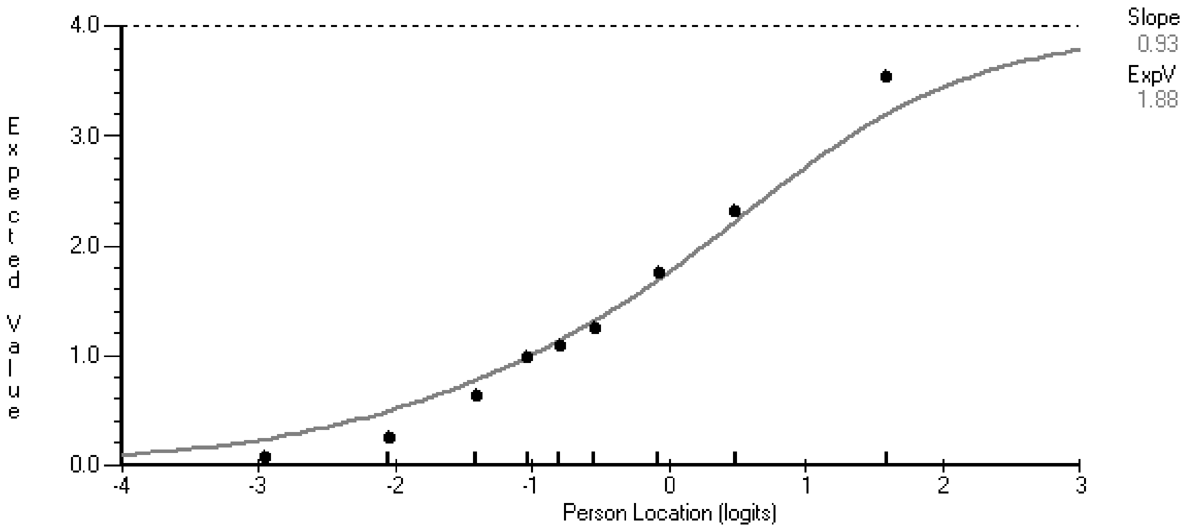


Figure 8.

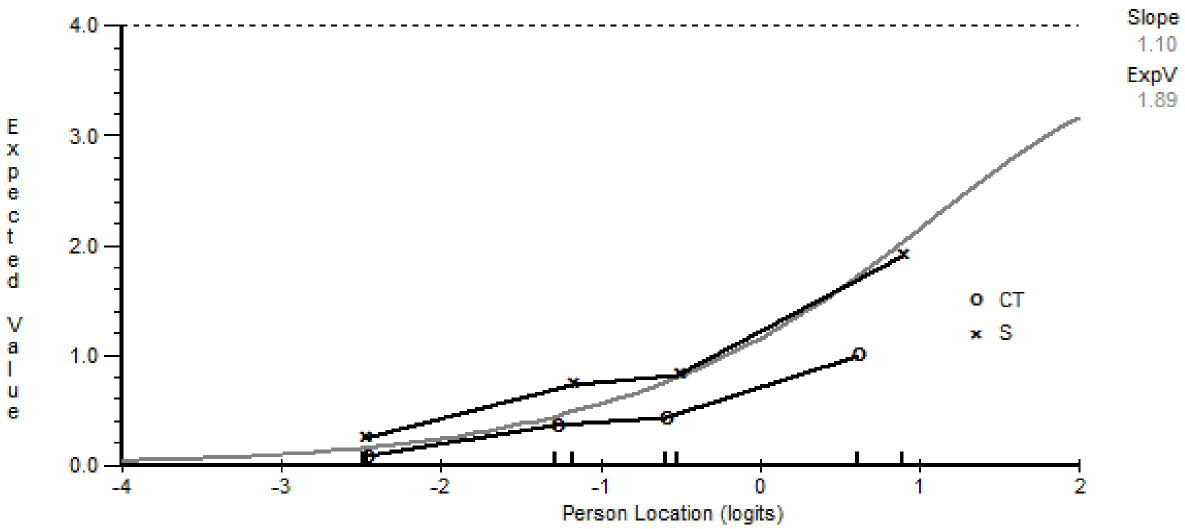
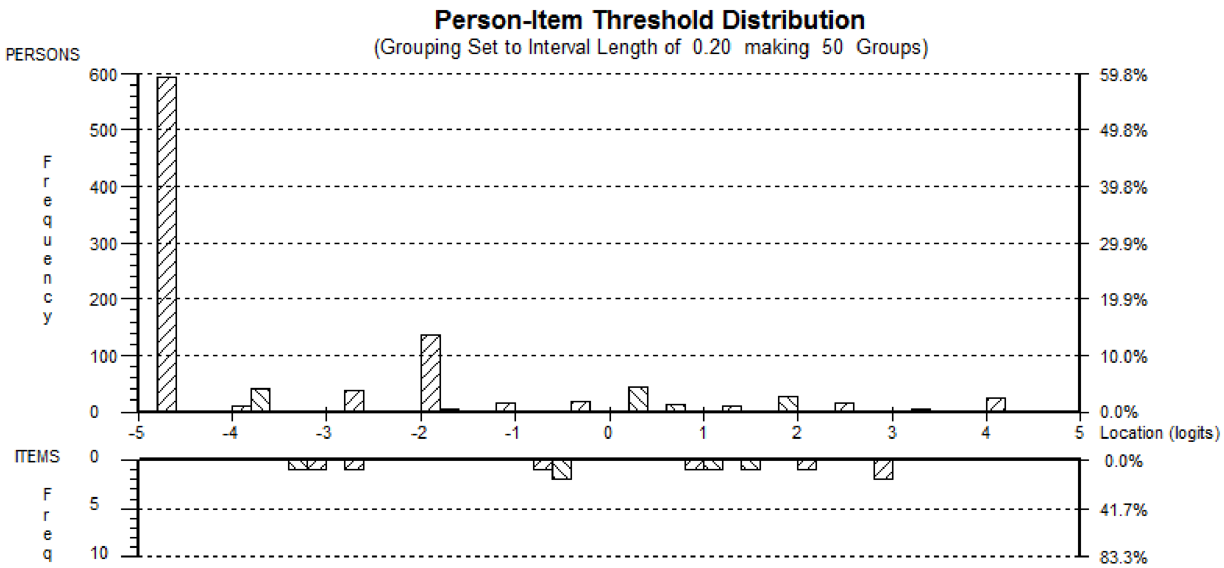
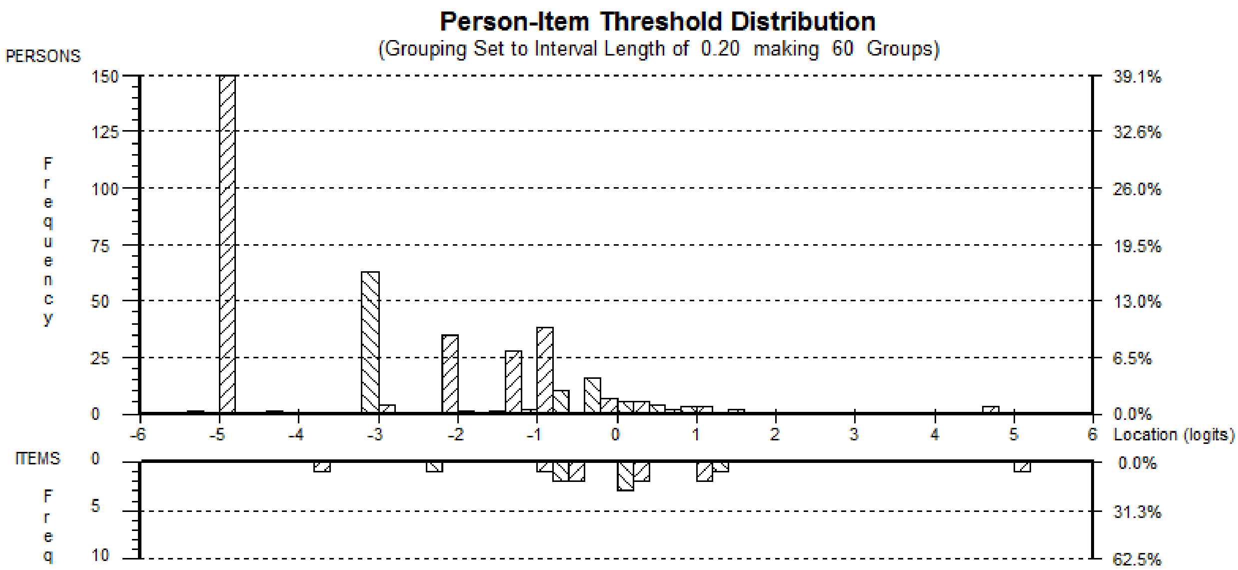


Figure 9.



Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.20 making 50 Groups)
PERSONS
Frequency
600
500
400
300
200
100
0
59.8%
49.8%
39.8%
29.9%
19.9%
10.0%
0.0%
-5
-4
-3
-2
-1
0
1
2
3
4
5
Location (logits)
ITEMS
Freq
0
5
10
0.0%
41.7%
83.3%

Figure 10.



Person-Item Threshold Distribution
(Grouping Set to Interval Length of 0.20 making 60 Groups)
PERSONS
Frequency
150
125
100
75
50
25
0
39.1%
32.6%
26.0%
19.5%
13.0%
6.5%
0.0%
-6
-5
-4
-3
-2
-1
0
1
2
3
4
5
6
Location (logits)
ITEMS
Freq
0
5
10
0.0%
31.3%
62.5%