Pain Archival Analysis Report

Prepared by
Wen-Hung Chen, PhD
Dennis Revicki, PhD

On behalf of the PROMIS
SCC Analysis Team

July, 2006

PROMIS: IRT Analysis for Pain Items in IMMPACT Surveys and NU-CORE Item Bank

The SCC directed secondary analyses were planned and conducted to provide information about the psychometric characteristics of existing items in the five targeted PROMIS domains (physical function, emotional distress, fatigue, pain, social/role). The main interest was in examining whether sets of items measuring the same domain were sufficiently unidimensional for IRT analyses and estimating item parameters for the items in the selected domains, in this case pain. These analyses are intended to provide preliminary information about item characteristics and performance that might help inform and guide decisions for the PROMIS field test.

Two separate data sets included items measuring pain-related concepts of pain intensity, pain quality and pain interference: (1) the IMMPACT Survey and (2) Northwestern University CORE item bank (NU-CORE). These pain-related items were selected from the both the IMMPACT Main Survey and Module 4 (which overlaps with the NU-CORE item bank) and the NU-CORE item bank assess pain intensity, pain interference and pain quality. There were differences in the sample sizes between the three samples (see below). In addition, we examined two different methods for linking item calibration across separate samples, simultaneous and separated IRT calibration, since for the PROMIS field test, this will be an important issue.

Data

The IMMPACT Survey was a cross sectional internet-based survey of individuals with chronic pain from the American Chronic Pain Association website. The purpose of the survey was to gather data on the importance and relevance of different domains impacted by chronic pain, measure pain intensity, quality and interference, and functional and psychological well-being outcomes in people with chronic pain. The sample consisted of 959 respondents with one or more chronic pain conditions (see Table 1). The survey included a main survey section that included a total of 43 pain-related items and Module 4 which included the 61 original items from the NU-CORE item bank. A total of 959 subjects completed the main survey and 148 subjects completed Module 4; the Module 4 respondents also completed the main survey.

The NU-CORE pain item bank data was collected in a cross-sectional survey of 402 cancer patients.[1] Table 2 summarizes the demographic and clinical characteristics of this sample. The original NU-CORE pain item bank contained 61 items, but this item bank was reduced to 43 items. Items without good psychometrical characteristics were excluded.[1]

The psychometric and IRT analysis of pain items were conducted on three samples: (1) IMMPACT Main survey (N=959); (2) IMMPACT Module 4 survey (N=148); and (3) NU-CORE item bank (N=402). The 148 subjects that completed the Module 4 survey also completed the Main survey. There are 43 items in the IMMPACT Main survey, and 36 items in Module 4. There is only one item that is common between the IMMPACT main survey and the Module 4 survey; it is a Pain Intensity domain item. NU-CORE item bank has 61 items. At later stage only 43 items from NU-CORE are used for these analyses. There are 7 items that are common between the IMMPACT main survey and the NU-CORE item bank. These 7 items are all Pain Interference domain items. All Module 4 items are items included in the NU-CORE item bank. There are 12 Pain Interference items, 14 Pain Quality items, and 10 Pain Intensity items.

Basically, this data is composed of two samples of subjects (IMMPACT and NU-CORE), and two sets of items (IMMACT main survey and NU-CORE item bank). Table 3 shows the relationship between the samples and the instruments. The keys to the linking of the items from the two samples (i.e., IMMPACT main survey and NU-CORE item bank) are the 148 subjects that completed both the IMMPACT main survey and a subset of the NU-CORE item bank, and the 7 common items that were completed by everyone. The list of items, their rating scales, and their domains are shown in Appendices A, B, and C.

Analysis Outline:

First, descriptive statistics, correlations, and item response frequencies are produced for general analyses of psychometric properties. These are summarized in Tables 4 to 6 for IMMPACT main survey items. Tables 7 to 9 show the same statistics for IMMPACT Module 4 items. Tables 10 to 12 summarize these data for NU-CORE items. We will not discuss the descriptive statistics in detail. However, there are three general observations. First, response categories of some of the items have to be collapsed because few or no subject endorsed them. Second, some of the item-by-item correlations are low. This is an indication that these items may not be unidimensional. Third, based on the correlations and response frequencies, it seems that the IMMPACT sample and the NU-CORE sample are heterogeneous, that is, they represent different chronic disease populations (i.e., chronic pain and oncology).

Factor analysis was next conducted to examine whether these pain items form a single unidimensional construct. There may be two sets of separate factor analyses: one for IMMPACT main survey using all 959 subjects; one for NU-CORE item bank using NU-CORE sample of 402 subjects. Further decisions were made on which items are to be included in the IRT analysis based on the factor analysis results. Please see discussion of factor analysis results in the next section.

IRT analyses were conducted in several sets for exploratory analyses for linking the data sets. First, we performed a simultaneous calibration for both item sets with all subjects. This approach will automatically put all items, regardless of the item sets, onto the same measurement scale. This method treats the two item sets as if they were of one big item bank. The subjects from the NU-CORE item bank were treated as a separate sample from the IMMPACT subjects. This will not affect the item parameter calibration, but will result in different means for the two samples of subjects.

The second linking approach that we used was by separated item calibration followed by scale transformation. One IRT calibration was conducted for the IMMPACT main survey and IMMPACT Module 4 combined. The second IRT calibration was conducted for the NU-CORE item bank. Next, common items shared between the IMMPACT Main and Module 4 Survey, and NU-CORE item bank were used as a “bridge” to link the IMMPACT items and the NU-CORE items. This linking was accomplished by using several different scale transformation methods[2,3]. After the scale transformation, the two item sets will then be on the same measurement scale. Using these two linking approaches, we were able to demonstrate how items from two studies can be linked and the results validated for the IRT calibration and for establishing of a single item bank.

Exploratory Factor Analysis Results

There were 43 items and 601 observations included in the exploratory factor analysis for the IMMPACT main survey data. Based on a priori expectations as to the factor structure, a four-factor solution was chosen. There are seven eigen-values that are greater than one. The largest is 14.32 and the second largest is 3.88. The factor analysis indicates the multidimensional structure underlying these items. However, the un-rotated factor pattern and the one very large eigen-value also hint at one super factor. This observation is more obvious for the NU-CORE item bank items. There were 61 items and 326 observations included in the exploratory factor analysis for the NU-ORE item bank data. As with the previous exploratory factor analysis, a four-factor solution is chosen. There are also seven eigen-values that are greater than one. The largest is 33.14, the second largest is 2.86. All items show larger loadings on the first factor in the un-rotated factor pattern table.

Next, exploratory factor analyses are conducted separately for subsets of the items based on their hypothetical domains. The two hypothesized domains for the IMMPACT main survey items are: Pain Interference and Pain Quality. There are 17 items and 788 observations included in the

exploratory factor analysis for the Pain Inference domain. Three eigen-values were greater than one. The largest is 9.24. Again, the un-rotated factor pattern shows high loadings of all items on the first factor. The rotated factor pattern shows three factors. But with closer examination, it was suspected that the response scale (i.e., 7 points vs. 11 points) may be playing a role in separating the additional factors.

Twenty-four items and 660 observations are included in exploratory factor analysis for the Pain Quality domain. Five eigen-values were greater than one. The largest is 7.98. It is clear that all 24 items are highly correlated, but there are also subtle distinctions between the types of pain. It seems that there are sub-domains such as sharp pain, fear caused by pain, tingling and numbness, etc. But, the results indicate that the items fit well enough into one dimension for the purpose of IRT analysis.

There were three hypothesized domains for the IMMPACT Module 4 items: Pain Intensity, Pain Interference, and Pain Quality. These items were also included in the NU-CORE item bank. There were 10 items and 395 observations included for the exploratory factor analysis for the Pain Intensity domain. There was only one eigen-value that was greater than one. This set of items relate to different aspects of having pain. It was clear that these items are unidimensional.

There were 12 items and 393 observations included in the exploratory factor analysis for the Pain Interference domain. There was only one eigen-value that is greater than one. Based on this criterion, one can assume that these 12 items were also unidimensional.

There were 14 items and 395 observations included in the exploratory factor analysis for the Pain Quality domain. There were three eigen-values that were greater than one. This result was similar to the Pain Quality domain for the IMMPACT main survey items. In both analyses, there seems some subtle distinctions that can be made about different types of pain quality. However, in both case, the un-rotated factor patterns show high loadings of all items on the first factor. An argument can also be made to support unidimensional assumption that is sufficient for IRT analysis.

Based on the exploratory factor analysis results and after discussion among the study team, it was decided that the pool of items from both datasets can be classified into three pain domains: interference, quality, and intensity. Each item was classified into one of the domains. Confirmatory factor analyses, using the same data, were conducted for further examination of the dimensionality and the item classification of these items.

Confirmatory Factor Analysis Results

Confirmatory factor analyses were conducted to examine the factor structure within the specified domains. The first analysis was for the 43 items of IMMPACT main survey that classified into interference and quality domains. There are 17 items for the interference domain, and 24 for quality domain (Appendix A). The RMR was 0.068, the goodness-of-fit index (GFI) was 0.71, goodness-of-fit index adjusted for degree of freedom (AGFI) is 0.68, NFI equals 0.71 and non-normed fit index was 0.74. The fit indexes suggest relatively poor fit. The second confirmatory factor analysis was for the 17 interference items as one dimension. The RMR was 0.70, GFI was 0.71, the AGFI was 0.63, NFI was 0.78 and non-normed fit index was 0.76. The third confirmatory factor analysis was for the 24 quality items. The RMR was 0.076, the GFI was 0.77, and the AGFI is 0.72, NFI was 0.68 and non-normed fit index was 0.68.

Confirmatory factor analyses were conducted using the 43 NU-CORE items with the three specified domains. There were 3 domains: interference, quality, and intensity, each with 19, 14, and 10 items, respectively (Appendix B). The RMR was 0.056, GFI was 0.65, the AGFI was 0.61. NFI was 0.79, and non-normed fit index was 0.82. Next the confirmatory factor analyses were conducted for each of the three domains separately, each tested as a single factor structure. For pain interference, the RMR was 0.052, the GFI was 0.66, AGFI was 0.58, NFI was 0.83 and non-

normed index was 0.83. For pain intensity, the RMR was 0.04, GFI was 0.85, AGFI was 0.76, NFI was 0.92, and non-normed fit index was 0.91. For pain quality, the RMR was 0.07, GFI was 0.83, AGFI was 0.77, NFI was 0.80 and non-normed fit index was 0.79.

Based on the exploratory factor analysis and confirmatory factor analysis findings, there was only relatively fair support for the unidimensionality of these pain item data. The pain intensity items seemed to indicate good fit to unidimensionality criteria, while the pain interference and pain quality items appeared to have only fair consistency with unidimensionality criteria.

IRT Item Calibration

IRT item calibration refers to the process of fitting item response data to the IRT model and obtaining the item parameters for each item. One of the main goal (and challenge) of this study was to link the two sets of items, IMMPACT and NU-CORE, so that they are on the same measurement scale. In this analysis the linking was possible because some of the items in the two surveys were exactly the same. These items act as the linkage and are called the "shared common items." The linking process is called "scaling" and is applicable because IRT methodology is used. There are different ways of linking sets of items using IRT method depending on the data collection designs. There are two ways to achieve linking for this study where we have common items with two non-equivalent sample groups. One way of linking is by "simultaneous IRT calibration." The other is by "separated IRT calibration" then link the two sets of calibration by "scale transformation."

Simultaneous IRT Calibration

Simultaneous IRT calibration is the easiest and most direct way to link two sets of items when there are shared common items. This approach takes advantage of the IRT assumption that the item characteristics are invariant across occasions and samples. In this case, during the IRT calibration, the item parameters of the common items are constrained to be equal across the samples. In essence, each common item is treated as one single item administered to different groups of samples. When the simultaneous calibration is completed all items are automatically on the same measurement scale. As a rule, the mean and standard deviation of the population distribution of latent trait (θ) are also estimated during the calibration. When there are multiple groups, one of the group will have mean fixed to zero, and other groups' mean will be estimated.

The program used was MULTILOG 7.0. MULTILOG allows multiple groups of samples. However, the maximum number of response categories allowed is 10. Therefore, for the 11-categoriy items, two of the response categories were combined into one category. A decision was made to combine the two highest response categories. In this study, the higher the response score indicates worst the pain or problems. Because most of the items were multiple-response categories, the graded response model is used.

The results of our simultaneous IRT calibration are summarized in Tables 13, 14, and 15, for the pain interference, pain quality, and pain intensity domains, respectively. There were 29 pooled items and 1,364 subjects for the pain interference domain. The slope parameters were all reasonable large from 1.84 to 3.74, and all the threshold parameters were monotonically increasing. The mean θ for the NU-CORE sample was fixed at zero, and the mean θ for the IMMPACT sample was 2.22. The item characteristic curves (Figure 1) suggest that 10 response categories may be too many. Some of the middle response categories overlap with each other. Subjects with similar degrees of interference (based on the estimated θ) were equally likely to endorse any of the overlapping response categories. On the average, the IMMPACT sample reported higher degree of pain interference. This makes sense, since the NU-CORE subjects were mostly general cancer patients plus some randomly selected prostate cancer patients whereas, the IMMPACT sample were patients who had chronic pain from rheumatoid arthritis, osteoarthritis, lower back pain, fibromyalgia, diabetic neuropathy, and other neuropathic pain.

There were 38 pooled items and 1,364 subjects for the pain quality domain. The slope parameters ranged from 0.53 to 3.24. The mean θ of the IMMPACT sample was 1.90 higher than the mean for the NU-CORE sample. The item characteristic curves (Figure 2) show that for these types of pain quality items the 10 response categories do not perform very well. Most subjects either endorsed the lowest (none) response or the two highest responses (worst possible). On the other hand, the dichotomous version of these pain quality items may not be very informative. Given these results, some compromise between dichotomous and 0 to 10 response categories seems reasonable. It is suggested that 5 response categories for these items may be more appropriate.

There were 10 pooled items and 1,364 subjects for the pain intensity domain. There was only one item that was answered by all subjects. Only the subjects administered the IMMPACT Module 4 survey and the NU-CORE subjects answered the other 9 items. The slope parameters were very high, from 2.78 to 4.73, except the item "I have minor aches and pains" which has a slope of 0.90. This item was one of the unfolding model items, where the patients with no pain or with severe pain would both answer "None of the time." Basically, this item does not fit the graded response model as well as the other 9 items. Mean θ of the IMMPACT sample was 2.41 higher than the NU-CORE sample. The item characteristic curves (Figure 3) show that, again, the items with 10 response categories can be reduced to be with fewer response categories.

Separated IRT Calibration

IRT calibration can be conducted separately for each of the groups. This method results in separate sets of item parameters for each group. These two sets of item parameters will not be on the same measurement scale because the groups are non-equivalent. To link the separate sets of item parameters, "scale transformation" is performed on the common items. Scale transformation calculates the constants that transform one set of item parameters to be equal to the other set of item parameters. The constants are called the "transformation constants." This method takes advantage of the "item invariant" assumption in IRT. Basically, the same items should have the same item parameters. Because the measurement scale in IRT is unobservable, the scale is arbitrarily given a mean of 0 and standard deviation of 1. When the same items are calibrated separately for two samples and/or two instruments, because of the difference between the samples or the instruments, the same items do not usually have the same parameter estimates. These two sets of parameter estimates are separated by a linear transformation. There are different scale transformation methods[2,3]. The most commonly known are the mean/mean, mean/sigma, and test characteristic curve methods. However, there is little research focused on comparing these different methods. Appendix D provides a brief introduction of these transformation methods.

Before the scale transformation, item parameters have to first be derived separately. Tables 16 to 21 summarize the estimated item parameters resulting from the separated item calibrations. Tables 16 and 17 are for the interference domain for the IMMPACT and NU_CORE samples, respectively. Tables 18 and 19 are for the quality domain, and Tables 20 and 21 are for the intensity domain. One noticeable result relates to the calibration of the intensity domain items for the IMMPACT sample. Only one item was answered by all IMMACT subjects, and other 9 items were answered by the 148 Module 4 Survey subjects. As a result, because of the small sample size and since many of the response categories have zero counts (see Table 9, e.g., rate worse pain in the last week, rate average pain in the last week, and rate pain right now) the calibration is not stable. One can easily see (Table 20) that one of the slopes is as low as 0.13. On the other hand, some of the thresholds are as high as 16.6 or 37.0, and are not monotonically increasing (see Table 20, e.g., rate worse pain in the last week, rate average pain in the last week, and rate pain right now). Based on this result, it is determined, for the intensity domain, the linking cannot be achieved based on the separated calibration approach.

Similar to the pain intensity domain, pain quality domain has the same problem with small sample size for the common items. One noticeable item is the 'Nagging Pain' item for the pain quality

domain in the IMMPACT set. This item has a very flat estimated slope and a very large threshold parameter. In general, for the IMMPACT data set, because the common items were answered by only the 148 subjects (of Module 4 survey), their estimated item parameters are not stable. These are two examples of data files with too few observations for the common items to successfully calibration the item parameters. As a result, they are not appropriate for use as the bridge to link with other items. In this situation, simultaneous calibration should be the approach for linking; it has the advantage of augmenting the sample sizes.

A good example of the separated calibration approach is for the pain interference domain. There are 7 common items between the IMMPACT main survey and NU-CORE items with 959 and 402 subjects in each study, respectively. In addition, there are 12 common items between IMMPACT module 4 survey and NU_CORE items with 148 and 402 subjects, respectively. The process starts with estimating item parameters for the pain interference items for the IMMPACT sample. These include 959 subjects who answered the Main survey and 148 of them also answered Module 4 surveys. There are 29 interference items in this data set, 17 item from the Main survey and 12 from Module 4. The estimated item parameters are shown in Table 16. Item calibration is also completed for the NU-CORE interference items. There are 19 interference items with 402 subjects. The estimated item parameters are shown in Table 17. There are 19 common items between these two sets of items (7 from the main survey and 12 from the Module 4 survey). Different from the simultaneous calibration where the mean θ is fixed for one of the sample and estimated for the other, for the each separated calibration the mean θ is always fixed at zero. That is, the mean θ are fixed at zero for both the IMMPACT and NU-CORE samples when their parameters were being calibrated in separate runs.

After the item parameters are estimated, the next step of linking is the scale transformation. The process is to linearly transform one set of item parameter estimates to be the same as the other set. In this case, we transformed the item parameters from the IMMPACT calibration to be the same as the NU-CORE calibration. The computer program STUIRT[5] is used to obtain the transformation constants. STUIRT is one of many equating and scale transformations computer programs that are available at the web-site www.uiowa.edu/~casma. These programs can be used in conjunction, as exercises, with the book by Kolen and Brennan[3]. Appendix F is the input file used by STUIRT to calculate the transformation constants for the pain interference items shown as an example.

For the purpose of comparison, we also scale transform the pain quality items even though we determine that the parameter estimates for the IMMPACT sample is not stable. There are 38 pain quality items and 959 subjects for the IMMPACT data, 24 from the main survey and 14 from the Module 4. Their estimated item parameters are shown in Tables 16 to 21. The output of the STUIRT is the transformation constants, A and B, using the four different scaling methods. These results are shown in Table 22 for both the interference and quality domains.

As shown in Table 22, the transformation constants are similar between Haebara and Stocking-Lord approaches, and this is consistent with general findings. These constants then are used to transform the item parameters of the non-common items, and the items from the two separated calibration are now on the same scale.

Comparing the scale transformation between the pain interference and pain quality domains, it is found that the transformation constant is very close for pain interference among the four transformation methods. In other words, the result of the transformation will be very close regardless of which transformation methods is used. On the other hand, the transformation constants for the pain quality domain vary considerably. For example, the largest difference between the B constants is 0.49, comparing to 0.019 for pain interference domain. Because of the variability of the transformation constants, the linking results will differ widely depending on the transformation methods. This makes the separately approach less useful in this situation. This is the direct result of the unstable item parameter estimated owing to small sample size.

Next, we compared the simultaneous approach and the separate approach for the pain interference domain. Note, that although we are comparing the item parameters of the same items between the two approaches, the comparison is not direct. The IMMPACT and NU-CORE items are on the same scale as a result of the simultaneous calibration approach. In addition, the IMMPACT and NU-CORE items are also on the same scale as the result of the separated calibration approach. However, the measurement scales for two approaches are not on the same scale. In fact, we achieve two sets of "item bank" calibrations based of the same items. However, the scales of the two item banks are only a linear transformation apart. On the other hand, direct comparison is applicable. All it requires is another scale transformation to make the two banks on the same scale. This is only for research interest to compare the simultaneous calibration and separated calibration approaches.

We correlated the item parameters of the two approaches. We can also examine the descriptive statistics and correlations of the IRT scores of the two approaches. The correlation between the slope parameters is 0.923. The correlation between the threshold parameters range between 0.911 to 0.992, except the first threshold parameter (b1). The scatter plots of the item parameters show that they mostly form a straight line, except an outlier for b1. It turns out that the b1 parameter for the item "Pain interference with your daily activities," a 4-point scale item, is very small. It is a result of zero observation for the lowest response option. Without this outlier, the two sets of item parameters will likely be even closer to each other. These data are summarized in Figures 4 to 8. Table 23 shows the descriptive statistics and correlation of the IRT score for the pain interference domain for the two approaches. The two scales differ approximately by the factor of 0.784 which is the ratio of the standard deviations for the IRT scores of the NU-CORE sample (1.047/0.821). This is because we use NU-CORE sample as the base for both of the calibration approaches, that is, constrain NU-CORE sample to have zero mean for its population Theta score. The correlations between the IRT scores of the two approaches (see Table 23), are as high as 0.999 for the both IMMPACT sample and NU-CORE sample, and for overall. These results demonstrate that the two calibration approaches produce very similar item characteristics.

In this study we demonstrated how the items from separate surveys can be linked to be on the same measurement scale and to form a single item bank. This is possible given the common items between the two surveys and the application of IRT analysis. For the datasets that are used in this study, because the common items for the pain quality and pain intensity domains are answered by only 148 subjects, we found that the item parameter estimates are not robust if they are calibrated separately. Because of the small sample size, we expect that the linking result is better when simultaneous calibration is used than separated calibration for these two pain domains. This finding corresponds to other simulation studies where the general finding is the simultaneous calibration produced more accurate results than separated calibration when the IRT model fits the data[3]. On the other hand, for pain interference domain, because of the sufficient sample size, the linking of the two studies yield similar results from simultaneous calibration and separated calibration. Therefore, in general, the results of this study suggest that simultaneous calibration, when possible, should be preferable when linking two item surveys. However, with sufficient sample size and common items, separate calibration yield similar result with simultaneous calibration.

References

1. Lai JS, Dineen K, Reeve BB, et al. An item response theory-based pain item bank can enhance measurement precision. J Pain Symptom Manage 2005; 30: 278-288.

2. Cohen AS, Kim SH. An investigation of linking methods under the graded response model. Applied Psychological Measurement 1998; 22: 116-130.

3. Kolen MJ, Brennan, RL. Test Equating, Scaling, and Linking: Methods and Practices. New York: Springer-Verlag 2004; 208-227.

4. Thissen DA, Chen WH, Bock, D. MULTILOG 7.0. 2002. Lincolnwood, IL: SSI.

5. Kim S, Kolen MJ. STUIRT: A computer program for scale transformation under unidimensional item response theory models, version 1.0, 2004. Iowa City: Iowa Testing Programs, Univ. of Iowa.

Appendix A

Pain Items in Northwestern University CORE Item Bank

Item ID	Pain Item	Rating Scale	Core/ Peripheral/ Reject	Comments
P1	I have pain	0=none of the time; 4=all of the time	C[1]	intensity
P3	Pain interferes with my ability to concentrate	0=none of the time; 4=all of the time	C[1]	interference
P4	Pain interferes with my ability to pay attention (e.g., details in conversations	0=none of the time; 4=all of the time	C[1]	interference
P5	My pain is unbearable	0=none of the time; 4=all of the time	C[1]	intensity
P8	I suffer from my pain	0=none of the time; 4=all of the time	C[1]	intensity
P10	I have minor aches and pains	0=none of the time; 4=all of the time	C[1]	intensity
P14	Pain interferes with my ability to think clearly	0=none of the time; 4=all of the time	C[1]	interference
P15	Pain makes me depressed	0=none of the time; 4=all of the time	C[1]	affective
P16	Pain interferes with my family life	0=none of the time; 4=all of the time	C[1]	interference
P18	Please rate your pain by circling the one number that best describes your pain at its **worst** in the last week	0=no pain; 10=pain as bad as you can imagine	C[1]	intensity
P20	Please rate your pain by circling the one number that best describes your pain on the **average**	0=no pain; 10=pain as bad as you can imagine	C[1]	intensity
P21	Please rate your pain by circling the one number that tells how much pain you have **right now**	0=no pain; 10=pain as bad as you can imagine	C[1]	intensity
P22	Circle Yes or No if the adjective applies to your pain: Aching	0=no; 1=yes	C[1]	quality
P23	Circle Yes or No if the adjective applies to your pain: Throbbing	0=no; 1=yes	C[1]	quality
P24	Circle Yes or No if the adjective applies to your pain: Shooting	0=no; 1=yes	C[1]	quality
P25	Circle Yes or No if the adjective applies to your pain: Stabbing	0=no; 1=yes	C[1]	quality
P26	Circle Yes or No if the adjective applies to your pain: Gnawing	0=no; 1=yes	C[1]	quality
P27	Circle Yes or No if the adjective applies to your pain: Sharp	0=no; 1=yes	C[1]	quality
P28	Circle Yes or No if the adjective applies to your pain: Burning	0=no; 1=yes	C[1]	quality
P29	Circle Yes or No if the adjective applies to your pain: Exhausting	0=no; 1=yes	C[1]	quality

Appendix A (continued)

Item ID	Pain Item	Rating Scale	Core/ Peripheral/ Reject	Comments
P30	Circle Yes or No if the adjective applies to your pain: Tiring	0=no; 1=yes	C[1]	quality
P31	Circle Yes or No if the adjective applies to your pain: Penetrating	0=no; 1=yes	C[1]	quality
P32	Circle Yes or No if the adjective applies to your pain: Nagging	0=no; 1=yes	C[1]	quality
P33	Circle Yes or No if the adjective applies to your pain: Numb	0=no; 1=yes	C[1]	quality
P34	Circle Yes or No if the adjective applies to your pain: Miserable	0=no; 1=yes	C[1]	quality
P35	Circle Yes or No if the adjective applies to your pain: Unbearable	0=no; 1=yes	C[1]	quality
P36	Circle the one number that describes how, during the past week, pain has interfered with your: General activity	0=does not interfere; 10=completely interferes	C[2]	interference
P37	Circle the one number that describes how, during the past week, pain has interfered with your: Mood	0=does not interfere; 10=completely interferes	C[2]	interference
P38	Circle the one number that describes how, during the past week, pain has interfered with your: Walking ability	0=does not interfere; 10=completely interferes	C[2]	interference
P39	Circle the one number that describes how, during the past week, pain has interfered with your: Normal work (includes both work outside the home and housework)	0=does not interfere; 10=completely interferes	C[2]	interference
P40	Circle the one number that describes how, during the past week, pain has interfered with your: Relations with other people	0=does not interfere; 10=completely interferes	C[2]	interference
P41	Circle the one number that describes how, during the past week, pain has interfered with your: Sleep	0=does not interfere; 10=completely interferes	C[2]	interference
P42	Circle the one number that describes how, during the past week, pain has interfered with your: Enjoyment of life	0=does not interfere; 10=completely interferes	C[2]	interference
P43	Have you had pain	1=not at all; 4=very much	C[1]	intensity
P44	Did pain interfere with your daily activities	1=not at all; 4=very much	C[1]	interference
P45	I have pain	0=not at all; 4=very much	C[1]	intensity
P46	How much bodily pain have you had in the past 4 weeks	1=none; 6=very severe	C[1]	intensity
P47	During the past 4 weeks, how much did pain interfere with the following things? Your mood	1=not at all; 5=extremely	C[1]	interference

Appendix A (continued)

Item ID	Pain Item	Rating Scale	Core/ Peripheral/ Reject	Comments
P48	During the past 4 weeks, how much did pain interfere with the following things? Your ability to walk or move about	1=not at all; 5=extremely	C[1]	interference
P49	During the past 4 weeks, how much did pain interfere with the following things? Your sleep	1=not at all; 5=extremely	C[1]	interference
P50	During the past 4 weeks, how much did pain interfere with the following things? Your normal work (including both work outside the home and housework)	1=not at all; 5=extremely	C[1]	interference
P51	During the past 4 weeks, how much did pain interfere with the following things? Your recreational activities	1=not at all; 5=extremely	C[1]	interference
P52	During the past 4 weeks, how much did pain interfere with the following things? Your enjoyment of life	1=not at all; 5=extremely	C[1]	interference
P2	Pain interferes with my close personal relationships	0=none of the time; 4=all of the time	P	interference
P6	My pain is so severe, I can hardly think of anything else	0=none of the time; 4=all of the time	P	interference
P7	Pain interferes with my ability to remember things	0=none of the time; 4=all of the time	P	interference
P9	Pain interferes with my ability to do the things I usually do for fun	0=none of the time; 4=all of the time	P	interference
P19	Please rate your pain by circling the one number that best describes your pain at its **least** in the last week	0=no pain; 10=pain as bad as you can imagine	P	intensity
P53	During the past 4 weeks, how often have you had pain or discomfort?	1=once or twice; 5=every day or almost every day	P	intensity
P54	Circle the one statement that most closely indicates how you have been feeling in the last week	1=I almost never have pain; 5=I am in some degree of pain almost constantly	P	intensity
P11	I have soreness after minor physical activity	0=none of the time; 4=all of the time	R	Concept issue
P12	I have soreness after moderate physical activity	0=none of the time; 4=all of the time	R	Concept issue
P13	I have soreness after strenuous physical activity	0=none of the time; 4=all of the time	R	Concept issue
P17	Pain interferes with my sex life	0=none of the time; 4=all of the time	R	Misfit
P55	Circle the one statement that most closely indicates how you have been feeling in the last week	1=when I do have pain, it is very mild; 5=the pain I have is almost unbearable	R	Confusing response scale

Appendix A (continued)

Item ID	Pain Item	Rating Scale	Core/ Peripheral/ Reject	Comments
P56	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Family/Home responsibilities. This category refers to activities related to the home or family. It includes chores or duties performed around the house (e.g., yard work, house cleaning) and errands or favors for other family members (e.g., driving the children to school).	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy
P57	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Recreation. This category includes hobbies, sports, and other similar leisure time activities.	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy
P58	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Social activity. This category refers to activities which involve participation with friends and acquaintances other than family members. It includes parties, theater, concerts, dining out, and other social functions.	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy
P59	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Occupation. This category refers to activities that are a part of or directly related to own's job. This includes non-paying jobs as well, such as housewife or volunteer.	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy
P60	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Self-care. This category includes activities which involve personal maintenance and independent daily living (e.g., taking a shower, driving, getting dressed).	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy
P61	Circle the number on the scale which describes the level of disability you have experienced due to your pain in each area over the past week: Sexual behavior. This category refers to the frequency and quality of one's sex life.	0=no disability; 10=total disability	R	Some degree of misfit, complex instructions, cumbersome, lengthy

[1] This is a common item with IMMPACT Module 4 survey

[2] This is a common item with IMMPACT Main survey

Appendix B

Pain Items in IMMPACT Main Survey

Item ID	Pain Item	Rating Scale	Comments
S4Q1_r1	Please rate your pain by selecting the one number that best describes your average pain during the past week.	0=no pain; 10=pain as bad as you can imagine	intensity
S4Q5_r1	Select the number below that describes how, during the past week, pain had interfered with your: General activity	0=does not interfere; 10=completely interferes	interference
S4Q5_r2	Select the number below that describes how, during the past week, pain had interfered with your: Mood	0=does not interfere; 10=completely interferes	interference
S4Q5_r3	Select the number below that describes how, during the past week, pain had interfered with your: Walking ability	0=does not interfere; 10=completely interferes	interference
S4Q5_r4	Select the number below that describes how, during the past week, pain had interfered with your: Normal work (including both work outside the home and housework)	0=does not interfere; 10=completely interferes	interference
S4Q5_r5	Select the number below that describes how, during the past week, pain had interfered with your: Relations with other people	0=does not interfere; 10=completely interferes	interference
S4Q5_r6	Select the number below that describes how, during the past week, pain had interfered with your: Sleep	0=does not interfere; 10=completely interferes	interference
S4Q5_r7	Select the number below that describes how, during the past week, pain had interfered with your: Enjoyment of life	0=does not interfere; 10=completely interferes	interference
S5Q1_r1	In general, how much does your pain interfere with your day-to-day activities?	0=no interference; 7=extreme interference	interference
S5Q2_r1	Since the time your pain began, how much had your pain changed your ability to work?	0=no interference; 7=extreme interference	interference
S5Q2_r2	How much has your pain changed the amount of satisfaction or enjoyment you get from taking part in social and recreational activities?	0=no interference; 7=extreme interference	interference
S5Q2_r3	How much has your pain changed your ability to take part in recreational and other social activities?	0=no interference; 7=extreme interference	interference
S5Q2_r4	How much has your pain changed the amount of satisfaction or enjoyment you get from family related activities?	0=no interference; 7=extreme interference	interference
S5Q2_r5	How much has your pain changed your relationship with your spouse, family, or significant other?	0=no interference; 7=extreme interference	interference
S5Q2_r6	How much has your pain changed the amount of satisfaction or enjoyment you get from work?	0=no interference; 7=extreme interference	interference
S5Q2_r7	How much has your pain changed your ability to do household chores?	0=no interference; 7=extreme interference	interference

Appendix B (continued)

Item ID	Pain Item	Rating Scale	Comments
S5Q2_r8	How much has your pain changed your friendships with people other than your family?	0=no interference; 7=extreme interference	interference
S7Q5	During the past 4 weeks, how much did pain interfere with your normal work (including both work outside the home and housework)?	0=not at all; 4=extremely	interference
S6Q1_r1	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Throbbing pain	0=none; 10=worst possible	quality
S6Q1_r2	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Shooting pain	0=none; 10=worst possible	quality
S6Q1_r3	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Stabbing pain	0=none; 10=worst possible	quality
S6Q1_r4	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Sharp pain	0=none; 10=worst possible	quality
S6Q1_r5	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Cramping pain	0=none; 10=worst possible	quality
S6Q1_r6	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Gnawing pain	0=none; 10=worst possible	quality
S6Q1_r7	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Hot-burning pain	0=none; 10=worst possible	quality
S6Q1_r8	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Aching pain	0=none; 10=worst possible	quality
S6Q1_r9	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Heavy pain	0=none; 10=worst possible	quality
S6Q1_r10	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Tender	0=none; 10=worst possible	quality
S6Q1_r11	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Splitting pain	0=none; 10=worst possible	quality

Appendix B (continued)

Item ID	Pain Item	Rating Scale	Comments
S6Q1_r12	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Tiring-exhausting	0=none; 10=worst possible	quality
S6Q1_r13	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Sickening	0=none; 10=worst possible	quality
S6Q1_r14	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Fearful	0=none; 10=worst possible	quality
S6Q1_r15	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Punishing-cruel	0=none; 10=worst possible	quality
S6Q1_r16	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Dull pain	0=none; 10=worst possible	quality
S6Q1_r17	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Electric-shock pain	0=none; 10=worst possible	quality
S6Q1_r18	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Squeezing-pressure pain	0=none; 10=worst possible	quality
S6Q1_r19	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Cold-freezing pain	0=none; 10=worst possible	quality
S6Q1_r20	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Piercing	0=none; 10=worst possible	quality
S6Q1_r21	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Pain caused by light touch	0=none; 10=worst possible	quality
S6Q1_r22	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Itching	0=none; 10=worst possible	quality
S6Q1_r23	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Tingling or “pins and needles”	0=none; 10=worst possible	quality
S6Q1_r24	Please click the number that best describes the intensity of the pain and related symptoms you felt during the past week: Numbness	0=none; 10=worst possible	quality

Appendix C

Pain Items in IMMPACT Module 4 Survey

Item ID	Pain Item	Rating Scale	Comments
M4Q1-r1	I have pain	0=none of the time; 4=all of the time	intensity
M4Q1-r2	Pain interferes with my ability to concentrate	0=none of the time; 4=all of the time	interference
M4Q1-r3	Pain interferes with my ability to pay attention (e.g., details in conversations	0=none of the time; 4=all of the time	interference
M4Q1-r4	My pain is unbearable	0=none of the time; 4=all of the time	intensity
M4Q1-r5	I suffer from my pain	0=none of the time; 4=all of the time	intensity
M4Q1-r6	I have minor aches and pains	0=none of the time; 4=all of the time	intensity
M4Q1-r7	Pain interferes with my ability to think clearly	0=none of the time; 4=all of the time	interference
M4Q1-r8	Pain makes me depressed	0=none of the time; 4=all of the time	affective
M4Q1-r9	Pain interferes with my family life	0=none of the time; 4=all of the time	interference
M4Q2-r1	Please rate your pain by circling the one number that best describes your pain at its **worst** in the last week	0=no pain; 10=pain as bad as you can imagine	intensity
M4Q3-r1	Please rate your pain by circling the one number that best describes your pain on the **average**	0=no pain; 10=pain as bad as you can imagine	intensity
M4Q4-r1	Please rate your pain by circling the one number that tells how much pain you have **right now**	0=no pain; 10=pain as bad as you can imagine	intensity
M4Q5-r1	Circle Yes or No if the adjective applies to your pain: Aching	0=no; 1=yes	quality
M4Q5-r2	Circle Yes or No if the adjective applies to your pain: Throbbing	0=no; 1=yes	quality
M4Q5-r3	Circle Yes or No if the adjective applies to your pain: Shooting	0=no; 1=yes	quality
M4Q5-r4	Circle Yes or No if the adjective applies to your pain: Stabbing	0=no; 1=yes	quality
M4Q5-r5	Circle Yes or No if the adjective applies to your pain: Gnawing	0=no; 1=yes	quality
M4Q5-r6	Circle Yes or No if the adjective applies to your pain: Sharp	0=no; 1=yes	quality
M4Q5-r7	Circle Yes or No if the adjective applies to your pain: Burning	0=no; 1=yes	quality
M4Q5-r8	Circle Yes or No if the adjective applies to your pain: Exhausting	0=no; 1=yes	quality
M4Q5-r9	Circle Yes or No if the adjective applies to your pain: Tiring	0=no; 1=yes	quality

Appendix C (continued)

Item ID	Pain Item	Rating Scale	Comments
M4Q5-r10	Circle Yes or No if the adjective applies to your pain: Penetrating	0=no; 1=yes	quality
M4Q5-r11	Circle Yes or No if the adjective applies to your pain: Nagging	0=no; 1=yes	quality
M4Q5-r12	Circle Yes or No if the adjective applies to your pain: Numb	0=no; 1=yes	quality
M4Q5-r13	Circle Yes or No if the adjective applies to your pain: Miserable	0=no; 1=yes	quality
M4Q5-r15	Circle Yes or No if the adjective applies to your pain: Unbearable	0=no; 1=yes	quality
M4Q6-r1	Have you had pain	1=not at all; 4=very much	intensity
M4Q6-r2	Did pain interfere with your daily activities	1=not at all; 4=very much	interference
M4Q7-r1	I have pain	0=not at all; 4=very much	intensity
M4Q8	How much bodily pain have you had in the past 4 weeks	1=none; 6=very severe	intensity
M4Q9-r1	During the past 4 weeks, how much did pain interfere with the following things? Your mood	1=not at all; 5=extremely	interference
M4Q9-r2	During the past 4 weeks, how much did pain interfere with the following things? Your ability to walk or move about	1=not at all; 5=extremely	interference
M4Q9-r3	During the past 4 weeks, how much did pain interfere with the following things? Your sleep	1=not at all; 5=extremely	interference
M4Q9-r4	During the past 4 weeks, how much did pain interfere with the following things? Your normal work (including both work outside the home and housework)	1=not at all; 5=extremely	interference
M4Q9-r5	During the past 4 weeks, how much did pain interfere with the following things? Your recreational activities	1=not at all; 5=extremely	interference
M4Q9-r6	During the past 4 weeks, how much did pain interfere with the following things? Your enjoyment of life	1=not at all; 5=extremely	interference

Appendix D

Brief Introduction to IRT Scale Transformation Methods

Test equating is a necessary step for the development of an assessment. Sometimes it is owing to the use of multiple forms. Sometimes it is owing to a common score scale is to maintain throughout the administrations of the assessment at different time frame. The widely use of IRT technique in assessment development requires a different set of test equating tools. Prior to the test equating when using IRT, a necessary step is to transform the item parameters onto the common scale. This is usually called the Transformations of IRT scales. The basic approach of the transformation of IRT scale is to calculate of the transformation constants. The idea is that the underlying two latent traits are a liner transformation of the each other. The transformation of one metric to the other can be expressed as,

$$\theta^* = A\theta + B, \quad (1)$$

where, *A* is the slope, and *B* is the intercept, and are called the transformation constants. One set of item parameters then can be transformed to the target metric using the transformation constants as follows,

$$a_j^* = \frac{a_j}{A}, \quad (2)$$

$$b_j^* = Ab_j + B, \quad (3)$$

$$c_j^* = c_j, \text{ and} \quad (4)$$

$$d_{jk}^* = Ad_{jk}. \quad (5)$$

After this scale transformation, items parameters estimates from separated IRT calibration are now on the same scale. Once the item parameters are on the same scale, the resulting proficiency estimates are on the same scale as well.

Scale transformation is necessary when the samples are from non-equivalent groups and the item parameters are estimated separately. In this case, it is also necessary that they share a certain number of common items. The well-known scale transformation methods include the mean/sigma, mean/mean and the characteristic curves methods. For characteristic methods, there are also the Haebara and Stocking-Lord approaches.

Mean/Mean Method

The most straightforward scale transformation method is to use the ratio of the means of the slope parameters of the common items for the *A* transformation constant. The difference of the means of the threshold parameters are used to drive the *B* constant.

$$A = \frac{mean(a_j)}{mean(a_j^*)}, \text{ and} \quad (6)$$

$$B = mean(b_j^*) - A \times mean(b_j). \quad (7)$$

Mean/Sigma Method

Another simple method is to use the ratio of the standard deviations of the threshold parameters of the common items for the *A* transformation constant. Then the difference of the means of the threshold parameters are used to drive the *B* constant.

$$A = \frac{STD(b_j^*)}{STD(b_j)}, \text{ and} \quad (8)$$

$$B = mean(b_j^*) - A \times mean(b_j). \quad (9)$$

Characteristic Curve Method

Basically, with the characteristic curve approach, the transformation constants can be obtained by minimized a quadratic loss function based on the difference between the IRT true scores of the common items. Specifically, the loss function is expressed as,

$$F = \frac{1}{N}\sum_{i=1}^{N}(T_i - T_i^*)^2, \quad (10)$$

where, N is the number of examinees, and T_i and T_i^* are the true scores for examinee i based on the old and the transformed item parameters, respectively. By summing over the examinees, Equation 2 minimizes the difference between the two true score distributions. This is only suitable for horizontal equating, where the same group of examinees has taken the common anchor items in both test administrations.

When Stocking-Lord approach is referred to as minimizing the difference between the test characteristic curves, the summation could be over a set of arbitrary values of ability. In this case, the loss function can be expressed as,

$$F = \frac{1}{Q}\sum_{i=1}^{Q}(T_i - T_i^*)^2, \quad (11)$$

where Q is the number of arbitrary quadrature points.

The true scores, T_i and T_i^* are calculated as,

$$T_i = \sum_{j=1}^{n}\sum_{k=1}^{m_j} u_{jk} P_{jk}(\theta_i), \text{ and} \quad (12)$$

$$T_i^* = \sum_{j=1}^{n}\sum_{k=1}^{m_j} u_{jk} P_{jk}^*(\theta_i), \quad (13)$$

where n is the number of items, m_j is the number of score categories for item j, u_{jk} is the score assigned to category k of item j, and $P_{jk}(\theta_i)$ is the probability of obtaining u_{jk} given the examinee whose ability is θ_i.

An iterative multivariate technique was used to find the coefficient that minimizes the loss function (Davidon, 1959; Fletcher & Powell, 1963). The method requires setting the first derivative with respect to the two coefficients to equal zero, for Equation 2,

$$\frac{\partial F}{\partial A} = \frac{-2}{N}\sum_{i=1}^{N}(T_i - T_i^*)\frac{\partial T_i^*}{\partial A} = 0, \text{ and} \quad (14)$$

$$\frac{\partial F}{\partial B} = \frac{-2}{N}\sum_{i=1}^{N}(T_i - T_i^*)\frac{\partial T_i^*}{\partial B} = 0, \quad (15)$$

and Q replaces N if Equation 3 is used.

The derivative of the target true score with respect to A, by the chain rule, is,

$$\frac{\partial T_i^*}{\partial A} = \sum_{j=1}^{n}\sum_{k=1}^{m_j} u_{jk}\left(\frac{\partial P_{jk}^*(\theta_i)}{\partial a_j^*}\frac{\partial a_j^*}{\partial A} + \frac{\partial P_{jk}^*(\theta_i)}{\partial b_j^*}\frac{\partial b_j^*}{\partial A} + \frac{\partial P_{jk}^*(\theta_i)}{\partial c_j^*}\frac{\partial c_j^*}{\partial A} + \sum_{t=1}^{m_j}\frac{\partial P_{jk}^*(\theta_i)}{\partial d_{jt}^*}\frac{\partial d_{jt}^*}{\partial A}\right), \quad (16)$$

and, with respect to B,

$$\frac{\partial T_i^*}{\partial B} = \sum_{j=1}^{n}\sum_{k=1}^{m_j} u_{jk}\left(\frac{\partial P_{jk}^*(\theta_i)}{\partial a_j^*}\frac{\partial a_j^*}{\partial B} + \frac{\partial P_{jk}^*(\theta_i)}{\partial b_j^*}\frac{\partial b_j^*}{\partial B} + \frac{\partial P_{jk}^*(\theta_i)}{\partial c_j^*}\frac{\partial c_j^*}{\partial B} + \sum_{t=1}^{m_j}\frac{\partial P_{jk}^*(\theta_i)}{\partial d_{jt}^*}\frac{\partial d_{jt}^*}{\partial B}\right). \quad (17)$$

Equations 16 and 17 are the generic formula to calculate the transformation constants for most IRT models. One just need to find the derivatives for the corresponding parameters and models, and substituted into Equations 16 and 17. The new expressions are then substituted into Equations 14 and 15 to find the coefficients that minimize the loss function.

Haebara approach is different from the Stock-Lord approach on how the lost function is defined. While Stock-Lord's lost function is to minimize the difference between the sum of the total true scores, whereas Haebara's lost function is to minimize the sum of the difference of individual item true scores. Harbara;s lost function is as following,

$$F = \frac{1}{Q}\sum_{i=1}^{Q}\sum_{j}^{n}(T_{ij} - T_{ij}^*)^2, \quad (18)$$

and the individual item true scores, T_{ij} and T_{ij}* are calculated as,

$$T_{ij} = \sum_{k=1}^{m_j} u_{jk} P_{jk}(\theta_i), \text{ and} \quad (19)$$

$$T_{ij}^* = \sum_{k=1}^{m_j} u_{jk} P_{jk}^*(\theta_i), \quad (20)$$

Summary

These transformation methods usually yield different transformation constants. There is no consensus on which scale transformation method is superior than the others. The results from limited number of studies are inconclusive. It is general found that the characteristic curve methods are most robust and accurate.

Appendix F

Input file used by STUIRT to calculate the transformation constants for pain interference items

```
/*
   This is the input file for STUIRT for IRT scale transformation.
   The project is PROMIS Pain Item IRT analysis (A2-3489).
   The new form is IMMPACT data of Main Survey and Module 4 Survey.
   The old form is Northwestern U. CORE item bank.
   All Module 4 items are also items in NU-CORE item bank, thus
   they are the common items.
   Only 148 out of the 959 IMMPACT subject answered Module 4 survey.
   NU-CORE data have 402 subjects.
   This is linking for the Pain Interference domain.
   The item parameters are estimated by MULTILOG.
   April 1, 2006 by WH Chen.
*/

NE 29
1 GR 7 DW 1.7 2.98836 -3.02156 -2.57366 -1.97739 -1.23054 -0.49632 0.21264
2 GR 7 DW 1.7 2.11631 -2.37427 -2.12376 -1.79901 -1.27229 -0.77242 -0.25563
3 GR 7 DW 1.7 2.86645 -2.68414 -2.32479 -1.89037 -1.38317 -0.79122 -0.02249
4 GR 7 DW 1.7 2.84749 -2.71570 -2.33314 -1.85916 -1.43211 -0.88398 -0.12413
5 GR 7 DW 1.7 2.52722 -2.44740 -1.87430 -1.49769 -1.06770 -0.32783  0.41254
6 GR 7 DW 1.7 1.80452 -2.48125 -1.95289 -1.38384 -0.85540 -0.36967  0.45045
7 GR 7 DW 1.7 2.02706 -2.34917 -2.08984 -1.69085 -1.32576 -0.83499 -0.19606
8 GR 7 DW 1.7 2.37400 -2.87333 -2.24134 -1.76486 -1.23394 -0.56738  0.20293
9 GR 7 DW 1.7 2.17423 -1.89269 -1.36077 -1.06796 -0.61379 -0.06254  0.67464
10 GR 5 DW 1.7 2.05509 -3.31187 -2.38648 -1.10504  0.37812
11 GR 5 DW 1.7 2.25096 -2.52440 -1.53764 -0.15288  0.74246
12 GR 5 DW 1.7 2.12174 -2.03665 -1.22666  0.11149  0.86151
13 GR 5 DW 1.7 2.04713 -2.09225 -1.10881  0.00684  1.16811
14 GR 5 DW 1.7 1.31328 -2.59845 -1.29589 -0.18556  0.90540
15 GR 5 DW 1.7 1.89326 -2.41879 -1.71609 -0.44648  0.60184
16 GR 10 DW 1.7 2.61934 -2.93045 -2.77713 -2.43649 -2.03730 -1.64365 -1.18144 -0.85566 -0.40565 0.12006
17 GR 10 DW 1.7 1.90497 -3.15580 -2.82865 -2.49166 -2.10104 -1.85285 -1.40586 -1.02909 -0.57793 0.05841
18 GR 10 DW 1.7 1.58216 -2.27948 -2.00936 -1.73586 -1.53085 -1.28206 -0.90898 -0.57155 -0.15231  0.36727
19 GR 10 DW 1.7 2.87077 -2.71773 -2.34032 -2.10660 -1.90767 -1.67036 -1.32809 -1.02770 -0.64702 -0.21906
20 GR 10 DW 1.7 2.24676 -2.28908 -1.93174 -1.63926 -1.36832 -1.11432 -0.79186 -0.50611 -0.14639  0.41260
21 GR 10 DW 1.7 1.19753 -3.72739 -3.17770 -2.80484 -2.46739 -2.12149 -1.70588 -1.37068 -0.89969 -0.25206
22 GR 10 DW 1.7 2.56791 -2.77114 -2.62349 -2.27106 -2.10581 -1.84158 -1.45817 -1.15535 -0.84369 -0.41051
23 GR 4 DW 1.7 2.77223 -5.56438 -1.24629 -0.14582
24 GR 5 DW 1.7 1.81674 -2.92725 -1.34358 -0.42321  1.04458
25 GR 5 DW 1.7 1.78100 -2.08096 -1.17751 -0.54285  0.65752
26 GR 5 DW 1.7 1.25996 -2.79491 -2.00520 -0.76976  0.25222
27 GR 5 DW 1.7 2.64415 -2.92869 -1.72874 -0.66298  0.35286
28 GR 5 DW 1.7 2.23122 -3.12728 -1.70070 -0.75998  0.15384
29 GR 5 DW 1.7 2.40768 -3.06567 -1.88952 -0.78176  0.14991

OL 19
1 GR 5 DW 1.7 3.63982 0.34396 0.88675 1.51103 2.25282
2 GR 5 DW 1.7 3.13251 0.52693 1.05788 1.92941 2.70518
3 GR 5 DW 1.7 2.98397 0.76292 1.21422 1.91443 2.63564
4 GR 5 DW 1.7 2.48696 0.39015 0.99732 1.72149 2.67653
5 GR 5 DW 1.7 3.25362 0.45515 1.07295 1.72057 2.29883
6 GR 10 DW 1.7 4.71229 -0.04664 0.19882 0.48615 0.69773 0.86028 1.06479 1.20704 1.42747 1.81388
```

```
7 GR 10 DW 1.7 4.33194 -0.07623 0.21947 0.51188 0.72522 0.84861 1.11228 1.27771 1.47238 1.81881
8 GR 10 DW 1.7 3.62997 -0.10847 0.14864 0.38889 0.58051 0.72621 0.97181 1.07360 1.26223 1.51455
9 GR 10 DW 1.7 5.49629 -0.07479 0.19237 0.46127 0.60180 0.71542 0.95997 1.03533 1.18942 1.46826
10 GR 10 DW 1.7 5.05058 0.31526 0.53880 0.76275 0.93150 1.07854 1.26857 1.41896 1.60744 1.92224
11 GR 10 DW 1.7 3.40586 -0.02517 0.24941 0.53478 0.73255 0.87619 1.10579 1.28439 1.50877 1.81124
12 GR 10 DW 1.7 5.95595 0.04949 0.27580 0.49873 0.63889 0.73940 0.97543 1.06924 1.26722 1.47026
13 GR 4 DW 1.7 4.86488 0.09946 1.03807 1.56739
14 GR 5 DW 1.7 4.11799 0.02869 0.83962 1.38079 2.26996
15 GR 5 DW 1.7 3.65595 -0.06676 0.64430 1.08511 1.88336
16 GR 5 DW 1.7 2.90338 -0.04141 0.78863 1.36137 2.23367
17 GR 5 DW 1.7 4.32139 -0.06620 0.59466 1.10213 1.80628
18 GR 5 DW 1.7 4.02012 -0.09545 0.52499 0.94222 1.58472
19 GR 5 DW 1.7 4.34903 -0.01914 0.63402 1.15890 2.15230

CI 19 MA
11 1
12 2
13 3
14 4
15 5
16 6
17 7
18 8
19 9
20 10
21 11
22 12
23 13
24 14
25 15
26 16
27 17
28 18
29 19

OP
ND 33 ED -4.5 4.5 / EQ
OD 33 ED -4.5 4.5 / EQ
FS DO DO
SY BI BI
LM 1.0 0.0 / 10 3.0 0.0001 FI INTERFERENCE_LOCMIN_HIS.OUT
KO SL
BY
```