

# Improving Subjective Scaling of Pain Using Rasch Analysis

Konrad Pesudovs\* and Bruce A. Noble<sup>†</sup>

**Abstract:** Pain management outcomes assessment depends on valid measurement of pain. However, the validity of single-item scales, such as numeric or faces scales, with the assignment of ordinal numerical values to response scale categories, is questionable. The universal assumption that equal distances between response choices represent equal distances on the dimension being measured is essentially erroneous. Herein we demonstrate that Rasch analysis can be used to expose and repair scale inequity and reengineer scale structure. Thirty-one subjects with severe ocular surface disease repeatedly completed a 7-category faces pain scale. Rasch analysis demonstrated that response category 5 was underutilized, leading to disordering of the response scale. Collapsing category 5 into either category 4 or 6 produced an ordered 6-category faces scale that could be recalibrated with average Rasch person measures to create linear measurement on a continuous latent variable. The value of further alterations to the scale was explored, and the implication for scale redesign discussed. Rasch analysis could be applied to any subjective pain measure post-hoc to create linear measurement or applied during instrument development to optimize design.

**Perspective:** Single-item scales like the faces scale or a 1-10 numerical rating scale are commonly used for the subjective assessment of pain. However, scores applied to response categories are arbitrary, so do not represent equidistant steps in the underlying latent variable (pain). Scale inequities are easily demonstrable and repairable with Rasch analysis.

© 2005 by the American Pain Society

**Key words:** Calibration, data interpretation (statistical), outcome assessment (health care), pain, pain measurement, Rasch analysis.

Outcomes research into the management of pain encounters the difficulty of how to measure pain and therefore the effectiveness of treatments. The difficulties of patient-centered measurement are not unique to pain measurement, but the lack of valid objective measures makes subjective scoring crucial. A common approach is the inclusion of a single-item, categorical rating scale such as 0-10 numeric rating scales,<sup>18</sup> 4-point verbal rating scales,<sup>25</sup> or faces scales.<sup>7</sup> The issue addressed herein is the validity of assigning numerical values to response categories. These scales assume equal distances between response choices represent equal distances in the dimension being measured. This assumption, although universally made, is almost certainly erroneous in many cases.<sup>16,21</sup> Rasch analysis provides a method for testing scale assumptions and modifying scale structure to become a truly linear scale. Rasch analysis has been applied to subjective measurement across

medicine, including health-related quality of life<sup>22</sup> and pain.<sup>2,23,26</sup> However, most studies of pain management outcome that incorporate a patient-centred measure fail to choose a Rasch scaled measure. The aim of this study is to illustrate the use of Rasch analysis to convert ordinal data, from a single-item scale, into measures on a linear latent variable, so this procedure could be applied to similar scales used in pain measurement. Additionally, the role of Rasch analysis in rating scale engineering is explored, to illustrate the process as others could apply it.

## Materials and Methods

Thirty-one adult subjects with severe ocular surface disease, recalcitrant to conventional treatment, enrolled in a randomized controlled trial of the benefit of autologous serum eyedrops for ocular surface disease.<sup>19</sup> Ocular surface disease is defined as a chronic state including destruction of the ocular surface resulting in scarring, keratinization, loss of goblet cells, conjunctivalization of the cornea, and so on, essentially including a loss of tear film function that may arise from a number of disease processes such as Sjögren syndrome, cicatricial pemphigoid, Stevens-Johnson syndrome, and graft versus host disease. This is a group suitable for the study of pain in its chronic concept, because constant pain and discomfort are a hallmark of the condition. Informed consent was obtained from all subjects during interview. The study complied with the principals of the World Medical Asso-

Received January 20, 2005; Revised April 8, 2005; Accepted April 13, 2005.

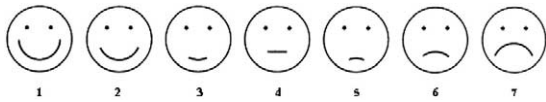
From the \*Department of Ophthalmology, Flinders Medical Centre and Flinders University, Bedford Park, Australia, and †Department of Ophthalmology, Clarendon Wing, The General Infirmary at Leeds, Belmont Grove, Leeds, Great Britain.

Konrad Pesudovs is supported by National Health and Medical Research Council of Australia (NHMRC) Sir Neil Hamilton Fairley Fellowship 0061, Canberra, Australia.

Address reprint requests to Konrad Pesudovs, PhD, Department of Ophthalmology, Flinders Medical Centre and Flinders University, Bedford Park, SA 5045, Australia. E-mail: Konrad.Pesudovs@flinders.edu.au  
1526-5900/\$30.00

© 2005 by the American Pain Society

doi:10.1016/j.jpain.2005.04.001



**Figure 1.** The faces scale used in this study. Please note that this is not offered as a model for clinical or research pain measurement, it was used here for research purposes. Better pain scales exist and should be chosen depending on the intended application.

ciation Declaration of Helsinki and was approved by the Leeds Regional Ethical Committee.

The subjects self-assessed pain daily during the course of the trial. Testing was done unsupervised, in the home environment with results recorded in the study logbook. The data extracted for this analysis were taken from 14 consecutive days during which the subjects' treatment and disease process were established to be stable on all study measures. This dataset produced 434 scores for analysis. Subjective assessment of pain utilized a faces scale. At least 14 versions of faces scales have been described for the assessment of pain.<sup>1,5</sup> These are chiefly used in children,<sup>9,10</sup> but are also applied to adults.<sup>6-8,11</sup> It is beyond the scope of this study to examine the properties of so many different scales,<sup>5</sup> because the purpose was simply to illustrate the applications of Rasch analysis to pain measurement. Therefore, a simple 7-category faces scale (Fig 1), also used for disciplines beyond pain research,<sup>17</sup> and similar in face design to that previously used in pain research,<sup>4</sup> has been used in this study. The written instructions were: "The first face shows somebody who has no pain or discomfort, the next face is somebody with a little bit of pain or discomfort. Each face along shows a little more pain or discomfort up to the last face which shows the worst possible pain. Which face comes closest to expressing the pain or discomfort of your eyes today?"

In this study, Rasch analysis was performed using Facets version 3.43 (<http://www.winsteps.com/minifac.htm>),<sup>14</sup> which calculates Wright and Masters' version of Rasch model estimates using joint maximum likelihood estimation.<sup>15,28</sup> The Rasch model does not assume values for response categories (eg, 1, 2, 3) but does assume that all categories are on the same underlying latent variable. We used a fixed-effects, 2-facet Rasch analysis approach:<sup>12</sup> at each of 14 time-points, 31 subjects rated their pain (14 observations of the same fixed effect for each subject). Because the scale is intended to have a uniform rating scale, only 1 rating scale structure is modeled. The Rasch model gives the probability of selecting a particular response category in terms of the interaction between "re-

sponse severity" and subject measure (in this case, pain) through an iterative logistic process.<sup>28</sup> The resulting response scale calibrations and person measures are expressed in log-odd units (natural logarithm of an odds ratio), or logits, positioned along a hierarchical scale with logits of greater magnitude representing increasing pain.<sup>29</sup> By definition, this scale is linear. The output can be used to score patient's responses by a direct or an indirect approach.

First, the software produces patient measures that can be directly used as the outcome measure. Alternatively, the software produces average person measures for the categories of the instrument used. These average person measures can be used to calculate patients' responses for the current or subsequent datasets. Both approaches create a linear measure, but only the direct approach is truly continuous and allows for between-person variation. However, in many settings, investigators may prefer the convenience of preestablished average person measures to acquiring new software and learning the analysis technique. Therefore we chose to illustrate this latter approach.

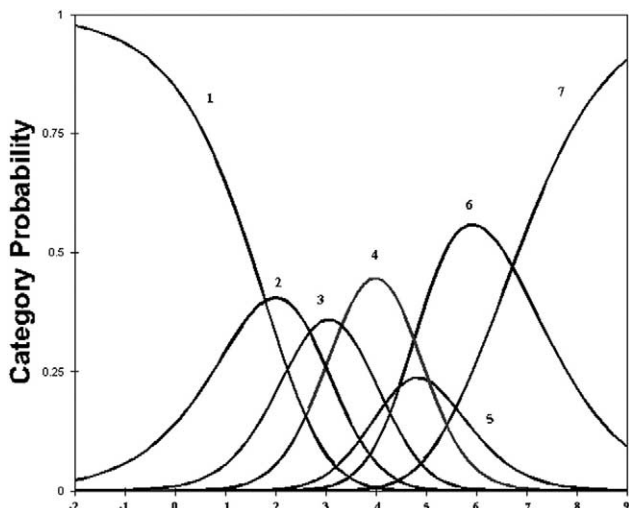
In utilizing Rasch analysis, we implicitly assume as a goal reengineering the assessment so that its measurement quality is optimized.<sup>13</sup> If response scale categories were not appropriately utilized across the whole scale (eg, low frequency, overlap with adjacent categories), the effect of merging categories was investigated.<sup>13</sup> This process could be continued until the measurement properties of the instrument were satisfactory.

## Results

Rasch analysis yields valid model statistics (separation, 7.49; reliability, 0.98, root mean square measurement error, 0.25) and person fit statistics (mean square  $\pm$  SD infit,  $1.00 \pm 1.00$ ; outfit,  $1.00 \pm 1.00$ ), although the high standard deviations were due to 5 of the 31 subjects poorly fitting the model (fit statistic  $>2.0$ ). The person measures generated could be used directly as the Rasch scaled pain measure. The average person measures for each category, as calculated by Rasch analysis, are shown in Table 1. The intervals are roughly 1 unit apart except that 4 and 5 are not effectively different (3.98 vs. 4.11), and 2 and 3 are only 0.64 apart. Ordinarily, these values could be assigned to the faces, to create an instrument with linear measurement properties. This would eliminate the overestimation of the differences in response for the central categories that occurs with traditional 1-7 scoring. However, in this case, the underutilization of

**Table 1.** The Percentage Utilization and the Average Rasch Person Measure for the Intervals on the Faces Scale

CATEGORY	1	2	3	4	5	6	7
% Utilization	16	18	19	21	8	11	6
Average Rasch Person Measure	1.02	2.40	3.04	3.98	4.11	5.40	6.82



**Figure 2.** Rasch model category probability curves for the faces scale representing the likelihood that a subject with a particular pain severity will select a category. The x-axis represents pain. For any given point along this scale, the category most likely to be chosen by a subject is shown by the category curve with the highest probability. At no point is category 5 the most likely to be selected.

face 5 (Fig 2) is causing disordering of the response scale. Therefore, we need to reengineer the scale.

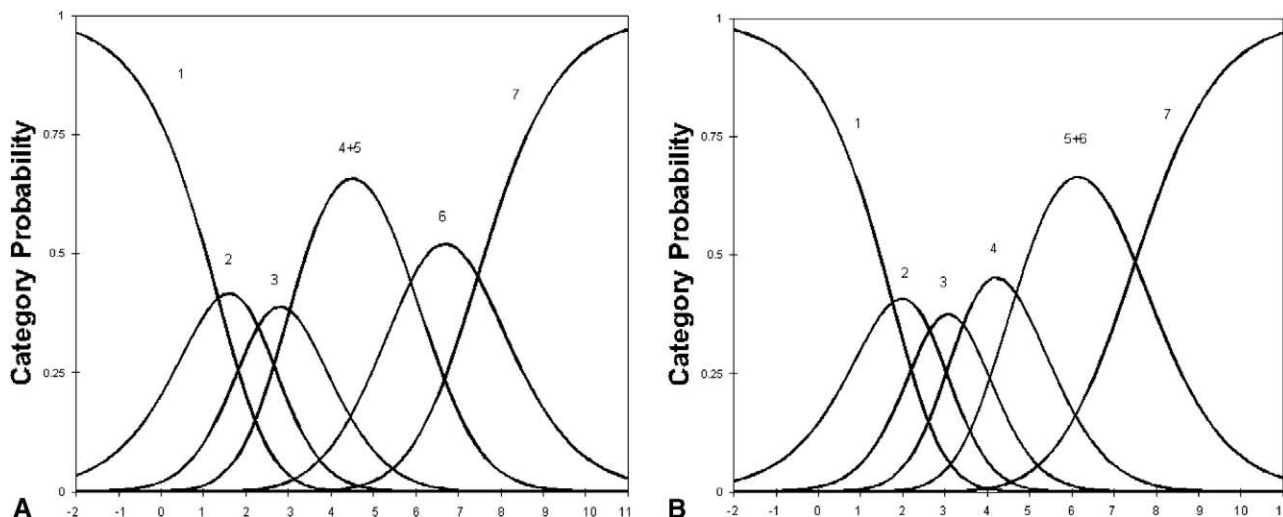
The probability that subjects with a given level of pain will select a response category is shown in Figure 2. Along the y-axis is probability and along the x-axis is the Rasch pain scale, with 7 distribution curves, 1 for each category, shown. Each probability curve illustrates the range of pain over which each category is likely to be chosen, including the range over which it is the most likely category to be chosen (the range over which a curve has the highest probability among all the curves). An optimally functioning scale should have each category most likely to be selected for an equal width of the scale (x-axis). Figure 2 illustrates the inequality of the faces scale categories with category 5 never being the most likely category selected. This suggests that category 5 could be combined with either of the adjacent categories. In this example, it is not entirely clear whether category 5 should be collapsed into category 4 or category 6, so both options were explored. Table 2 gives the percentage utilization and the average person measures for each category of the two 6-category alternatives and

their category probability curves are shown in Figure 3. An argument could be made in favor of either combination as each category has some utilization, each has a range for which it is the most likely to be chosen (albeit a wide range for the combined categories), and the average person measures are reasonably well spaced. Combining categories 4 and 5 gives a more even spacing between average person measures, but combining categories 5 and 6 evens up the percentage utilization. Therefore, we could choose either version and stop at this stage. However, this would give a lopsided faces scale with more categories below neutral than above.

One of the overarching concepts for category collapse must be that the resulting scale is “sensible.”<sup>27</sup> Therefore, it might be desirable to examine the possibility of collapsing categories 2 and 3 for symmetry. This could also be justified because categories 2 and 3 are relatively close together in terms of average patient measure and occupy a limited range of the scale for maximum probability of selection. Table 3 gives the percentage utilization and the average person measures for each category of the two 5-category alternatives and their category probability curves are shown in Figure 4. Again, an argument could be made in favor of either version. Combining categories 2 and 3 and 4 and 5 evens up the category probability curves (Fig 4A), although category 6 is less represented than categories 2 and 4. However, category 6 is not underutilized; it covers a reasonable length of the scale. Instead, the underutilization of category 6 illustrates that the population mean is in the lower half of the scale, as seen in other studies,<sup>6</sup> rather than at the center. This is also the case in the version with categories 2 and 3 and 5 and 6 combined. This version has the advantage of showing excellent symmetry in maximum probable category usage (Fig 4B), although category 4 has a minimal range over which it is the most likely category to be chosen. This version is also more sensible in terms of symmetry of a redrawn 5-category faces scale in which the mouth length of redrawn categories replacing 2 and 3 and 5 and 6 would match. Considering this issue, one could argue that if categories 4 and 5 were combined, leaving 6 on its own, category 3 should be added to 4 and 5, leaving 2 on its own. However, from the percentage utilization values in Table 2, we can calculate that 49% of the scores would be in the combined categories 3 to 5; and inspection of the curves in Figure 3A suggests that the resultant category 2 would have a very

**Table 2. The Percentage Utilization and the Average Rasch Person Measure for the Intervals on the Faces Scale for the 2 Alternatives for Collapsing Category 5 into an Adjacent Category**

CATEGORY	1	2	3	4 + 5	6	7
% Utilization	16	18	19	30	11	6
Average Rasch Person Measure	0.57	1.98	2.79	4.22	5.98	7.50
CATEGORY	1	2	3	4	5 + 6	7
% Utilization	16	18	19	21	20	6
Average Rasch Person Measure	0.98	2.37	3.06	4.21	5.18	7.47



**Figure 3.** Rasch model category probability curves for the faces scale shortened to 6 categories by (A) combining categories 4 and 5; (B) combining categories 5 and 6.

limited range of maximum probable usage. Therefore this version is not presented.

### Discussion

Testing with Rasch analysis invalidates the assumption that assigning values 1 to 7 to the categories of the faces scale will produce linear measurement. The apparent equivalence of categories 4 and 5 is a significant problem. However, this is caused by scale disordering due to the underutilization of category 5. Ordinarily, recoding the scale using the average person measures in Table 1 would convert this uneven ordinal data into a linear measure, thus facilitating simple statistical analyses. However, in this case, the scale needs to be reengineered to overcome the disordering and underutilization of category 5. The underutilization of category 5 commends its collapse into an adjacent category. Underutilized categories are a common problem in scales with many categories. Visual analog scales of pain (a 10-cm line) are effectively an 101 category scale if measured by the millimeter; however, Rasch analysis has shown that these should be collapsed into 10 categories at most and even then underutilization suggested further combination of

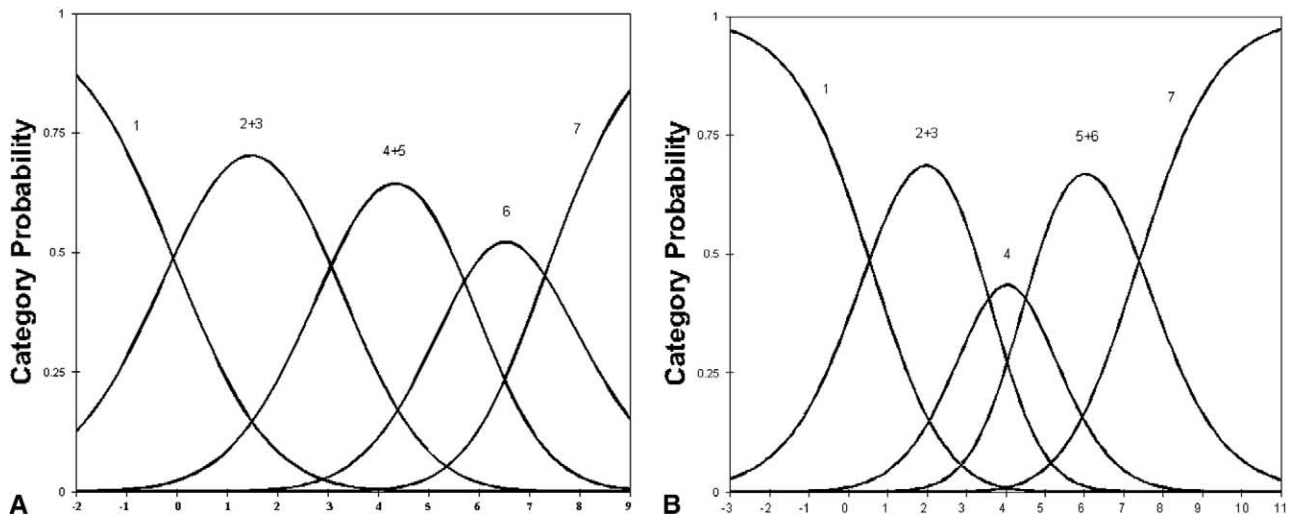
categories was possible.<sup>26</sup> The benefit of a reduction in the number of response scale categories is evident in the improvement from Figure 2 to Figure 3, with all categories being utilized and category widths becoming more even. However, in both versions the combined category occupies a large width of scale and creates an asymmetric scale with more categories below neutral than above.

Although this may be technically acceptable, and even logical given that the population mean is in the lower half of the scale, as seen in other studies,<sup>6</sup> an alternative approach would be to also collapse categories below neutral for the sake of scale symmetry. Both versions of the 5-category faces scale give fairly even category widths, but in each case the category not combined has a smaller range of maximal probable usage than the combined categories. Combining categories 2 and 3 and 5 and 6 gives good symmetry of maximal probable usage and also of a theoretical redrawn faces scale based on these combinations. The downside to reducing the number of categories from 6 to 5 is the loss of information that may occur. For this reason, further category collapse would not be contemplated. Importantly, all 4 reduced versions function better than the 7-category faces scale,

**Table 3.** The Percentage Utilization and the Average Rasch Person Measure for the Alternative Reduced 5-Interval Faces Scales Where Categories 2 and 3 Are Collapsed and Category 5 Is Collapsed into Either Adjacent Category

CATEGORY	1	2 + 3	4 + 5	6	7
% Utilization	16	37	29	11	6
Average Rasch Person Measure	-0.61	1.76	3.98	5.83	7.35
CATEGORY	1	2 + 3	4	5 + 6	7
% Utilization	16	37	21	20	6
Average Rasch Person Measure	0.02	2.25	4.00	5.05	7.36





**Figure 4.** (A) Rasch model category probability curves for the faces scale shortened to 5 categories by combining categories 2 and 3, and 4 and 5. In this model, all categories have a range over which they are most likely to be selected by a subject and show a fairly even category width. (B) Rasch model category probability curves for the faces scale shortened to 5 categories by combining categories 2 and 3, and 5 and 6. This model gives excellent symmetry with the neutral face limited in its range of maximum probable usage.

but none of these versions represents a definitive solution. Any of these would be a reasonable choice for analysis of these data.

Using this method of collapsing response scales, a 7-category instrument could be implemented but analyzed as a 6- or 5-category instrument.<sup>30</sup> This has the advantage of preventing the need for validation of a new instrument. Simply recoding the scale using the person measures in Table 2 or 3 converts this uneven ordinal data into a linear measure, thus facilitating simple statistical analyses. The conversion scores shown in Tables 2 and 3 should only be considered valid for this population. Rasch item calibrations and person measures will vary according to the disease and other characteristics of the population tested. Therefore, Rasch analysis ideally should be applied separately to each dataset, thereby producing patient measures in each case without the need to produce coefficients to recode each category. However, a categorical scale could be revalidated on a representative population and a new scoring system anchored to the categories. This would provide a convenient way for investigators to have access to a linear measure without the need to perform Rasch analysis themselves.

One could argue that these results are predictable from the appearance of the faces scale (Fig 1). Faces 3 and 5 have a short smile or frown length that is very similar to that of face 4. This similarity in mouth appearance leads to similar average person measures and underutilization of categories 3 and 5. This problem may have been accentuated by our subjects having visual impairment. Therefore, a numerical or verbal rating scale may have been a better choice for this population. Perhaps an improved faces scale could be drawn with 5 categories but the mouth lengths of the new categories 2 and 4 halfway between the mouth lengths of categories

2 and 3 and 5 and 6 in the original version. The finding that none of our reengineered versions were ideal suggests that redrawing the scale in this way would be appropriate. Certainly, our data suggest that combining categories 2 and 3, and 5 and 6, would be sensible. Hunter et al, in an ordering experiment, also found category overlap for faces near the middle of a 7-face scale and also suggested either reduction in the number of faces or modification to them.<sup>10</sup>

A shorter scale would have the advantage of being easier for young children to manage.<sup>24</sup> However, this redrawn 5-category faces scale would require validation. The subtleties underpinning these results illustrate that our analysis of this faces scale is not generalizable to other faces scales. Potentially, existing 5-category faces scales may be ideal. Other more sophisticated versions of the faces scale, with modifications specific for pain scoring—such as no-smiling face anchor for use with children because they may confuse affect and pain assessment<sup>4</sup>—and other important design subtleties<sup>5</sup> are available.<sup>8,9</sup> The confusion of affect and pain should be considered as a limitation of the choice of faces scale used in this study. Considerable effort has gone into improvement and validation of faces scales for pain,<sup>3</sup> including attempts to equalize separation between steps by psychophysical methods.<sup>9</sup> Rasch analysis should be applied to these iterations of facial pain scales to assess the scale structure and to determine the success of interval equalization. Arguments such as the need for scales to conform to a readily conceptualized numeric range—eg, 0 to 10<sup>9</sup>—should be reprioritized behind the pursuit of truly linear measurement.

Rescaling with Rasch analysis carries the important practical benefit of improving measurement precision.<sup>20</sup> In the research setting, this may mean fewer subjects are required for outcomes studies. This can also be demon-

strated using the data presented in this paper. These data were from a randomized controlled trial of autologous serum for ocular surface disease.<sup>19</sup> If we consider the group who were randomized to have conventional treatment followed by serum treatment, their raw faces scale scores were (mean  $\pm$  SD) for conventional treatment  $3.58 \pm 1.40$  and for serum  $2.36 \pm 1.24$ . After Rasch analysis the scores for conventional treatment were  $3.60 \pm 1.05$  and for serum  $2.09 \pm 1.19$ . From these data we can calculate an effect size for the raw scale of 0.92, and for the Rasch scale of 1.35. This difference represents a small but significant improvement in the sensitivity to detect a difference between groups.

Unless the rating scales that form the basis of data collection function effectively, any conclusions based on those data will be insecure. In this age of evidence-based medicine, patient-centered outcomes are as important for outcomes assessment as objective measures even if our scientific training makes us feel more comfortable with a numerical objective outcome. Rasch analysis of subjective scoring of pain can give us more confidence in patient-centered measurement by eliminating nonlinearities in a scale and by providing an insight into rating

scale function and thereby guidance for modifying its structure. The Rasch analysis has potential to improve the scoring of subjective assessment of pain and should be considered for all pain management outcomes research. Rasch analysis can be performed on any ordinal subjective rating data—such as faces scales, visual analog scales (effectively 0-101 categories), verbal rating scales or numerical rating scales—using the appropriate software. This will result in patient measures of pain that fall on a continuous linear scale. If a scale is validated using Rasch analysis and shown to perform well, then average person measures can be produced for each category to make the scale linear. These can then be used in subsequent studies to give truly linear measurement. The practical benefits of linear measurement are increased precision, which reduces the number of subjects needed for outcome studies, and increased reliability on a case-by-case basis so investigators can feel more confident in using the scores as an outcome measure. While the faces scale chosen for this investigation, the condition giving rise to pain, and the reengineered faces scales may not be applicable to any other study or population, the principles and the methodology applied herein certainly are.

## References

1. Bieri D, Reeve RA, Champion GD, Addicoat L, Ziegler JB: The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation, and preliminary investigation for ratio scale properties. *Pain* 41:139-150, 1990
2. Boeckstyns ME: Development and construct validity of a knee pain questionnaire. *Pain* 31:47-52, 1987
3. Breau LM, McGrath PJ, Craig KD, Santor D, Cassidy KL, Reid GJ: Facial expression of children receiving immunizations: a principal components analysis of the child facial coding system. *Clin J Pain* 17:178-186, 2001
4. Chambers CT, Craig KD: An intrusive impact of anchors in children's faces pain scales. *Pain* 78:27-37, 1998
5. Chambers CT, Giesbrecht K, Craig KD, Bennett SM, Huntsman E: A comparison of faces scales for the measurement of pediatric pain: children's and parents' ratings. *Pain* 83:25-35, 1999
6. Chibnall JT, Tait RC: Pain assessment in cognitively impaired and unimpaired older adults: a comparison of four scales. *Pain* 92:173-186, 2001
7. Closs SJ, Barr B, Briggs M, Cash K, Seers K: A comparison of five pain assessment scales for nursing home residents with varying degrees of cognitive impairment. *J Pain Symptom Manage* 27:196-205, 2004
8. Herr KA, Mobily PR, Kahout FJ, Wagenaar D: Evaluation of the Faces Pain Scale for use with the elderly. *Clin J Pain* 14:29-35, 1998
9. Hicks CL, von Baeyer CL, Spafford PA, van Korlaar I, Goodenough B: The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. *Pain* 93:173-183, 2001
10. Hunter M, McDowell L, Hennessy R, Cassey J: An evaluation of the Faces Pain Scale with young children. *J Pain Symptom Manage* 20:122-129, 2000
11. Kamel HK, Phlavan M, Malekgoudarzi B, Gogel P, Morley JE: Utilizing pain assessment scales increases the frequency of diagnosing pain among elderly nursing home residents. *J Pain Symptom Manage* 21:450-455, 2001
12. Linacre JM: Stress after Three Mile Island. *Rasch Measurement Transactions* 5:188, 1992
13. Linacre JM: Investigating rating scale category utility. *J Outcome Meas* 3:103-122, 1999
14. Linacre JM: A user's guide to facets Rasch measurement computer program. Chicago, MESA, 1999
15. Linacre JM, Wright BD: Understanding Rasch measurement: construction of measures from many-facet data. *J Appl Meas* 3:486-512, 2002
16. Massof RW, Fletcher DC: Evaluation of the NEI visual functioning questionnaire as an interval measure of visual ability in low vision. *Vision Res* 41:397-413, 2001
17. McDowell I, Newell C: Measuring health: A guide to rating scales and questionnaires. New York, Oxford University Press, 1987, pp 194-198
18. Mercadante S, Villari P, Ferrera P, Bianchi M, Casuccio A: Safety and effectiveness of intravenous morphine for episodic (breakthrough) pain using a fixed ratio with the oral daily morphine dose. *J Pain Symptom Manage* 27:352-359, 2004
19. Noble BA, Loh RSK, MacLennan S, Pesudovs K, Reynolds A, Bridges LR, Burr J, Stewart O, Quereshi S: Comparison of autologous serum eye drops with conventional therapy in a randomised controlled crossover trial for ocular surface disease. *Br J Ophthalmol* 88:647-652, 2004
20. Norquist JM, Fitzpatrick R, Dawson J, Jenkinson C: Comparing alternative Rasch-based methods vs raw scores in measuring change in health. *Med Care* 42:125-36, 2004

21. Pesudovs K, Garamendi E, Elliott DB: The Quality of Life Impact of Refractive Correction (QIRC) questionnaire: development and validation. *Optom Vis Sci* 81:769-777, 2004
22. Raczek AE, Ware JE, Bjorner JB, Gandek B, Haley SM, Aaronson NK, Apolone G, Bech P, Brazier JE, Bullinger M, Sullivan M: Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. *International Quality of Life Assessment. J Clin Epidemiol* 51:1203-1214, 1998
23. Ryser L, Wright BD, Aeschlimann A, Mariacher-Gehler S, Stucki G: A new look at the Western Ontario and McMaster Universities Osteoarthritis Index using Rasch analysis. *Arthritis Care Res* 12:331-335, 1999
24. Shih AR, von Baeyer CL: Preschool childrens' seriation of pain faces and happy faces in the Affective Facial Scale. *Psychol Rep* 74:659-665, 1994
25. Simpson KH, Ward J: A randomized, double-blind, cross-over study of the use of transcutaneous spinal electroanalgesia in patients with pain from chronic critical limb ischemia. *J Pain Symptom Manage* 28:511-516, 2004
26. Thomee R, Grimby G, Wright BD, Linacre JM: Rasch analysis of Visual Analog Scale measurements before and after treatment of patellofemoral pain syndrome in women. *Scand J Rehabil Med* 27:145-151, 1995
27. Wright BD: Comparing Rasch measurement and factor analysis. *Structural Equation Modeling* 3:3-24, 1996
28. Wright BD, Masters GN: *Rating Scale Analysis*. Chicago, IL, MESA Press, 1982
29. Wright BD, Stone MH: *Best Test Design*. Chicago, IL, MESA Press, 1979
30. Zhu W, Updyke WF, Lewandowski C: Post-hoc Rasch analysis of optimal categorization of an ordered response scale. *J Outcome Meas* 1:286-304, 1997