



Analyzing differential item functioning of the Nottingham Health Profile by Mixed Rasch Model

Selcen Yüksel¹, Atilla Halil Elhan², Derya Gökmen², Ayşe Adile Küçükdeveci³, Şehim Kutlay³

¹Department of Biostatistics, Ankara Yıldırım Beyazıt University Medical School, Ankara, Turkey

²Department of Biostatistics, Ankara University Medical School, Ankara, Turkey

³Department of Physical Medicine and Rehabilitation, Ankara University Medical School, Ankara, Turkey

Received: February 2018 Accepted: March 2018

ABSTRACT

Objectives: This study aims to evaluate whether items in the Turkish version of the Nottingham Health Profile (NHP) function differently according to age, sex, and duration of pain using the Mixed Rasch Model (MRM).

Patients and methods: The NHP data of patients with low back pain from a previous study was analyzed. To analyze differential item functioning (DIF) within the items of the NHP, the MRM was used. Age, sex, and duration of pain were considered factors which could cause DIF. The most powerful factor to define latent classes derived from the MRM was estimated using the Rasch tree method.

Results: The two-class mixture version of the nominal response model was identified as the best fitting model for the physical mobility, sleep, and social isolation sections. For the physical mobility dimension, some items showed DIF by age and for the social isolation dimension some items showed DIF by sex. For the sleep dimension, latent classes were unable to be explained by age, sex, and duration of pain. Items in other dimensions of the NHP were DIF-free and no items showed DIF according to age, sex, and duration of pain.

Conclusion: The results of this study may shed light on explaining the different response behavior of patients on the items of the NHP. Age and sex were found to be variables affecting item responses of NHP. Contrary to expectations, duration of pain did not cause DIF for any item. From the perspective of DIF, it can be concluded that NHP scale is a robust scale in terms of validity.

Keywords: Differential item functioning; low back pain; Mixed Rasch Model; Nottingham Health Profile.

Physical and rehabilitation medicine aims to enable persons with health conditions who experience, or are likely to experience disability, to achieve and maintain optimal functioning, and to enhance quality of life.^[1] Rehabilitation is a problem-solving and educational process that requires the use of assessments to identify problems and monitor patients during treatment.^[2] Standardized measurement tools of functioning and quality of life, including patient-reported outcome measures (PROMs), are increasingly considered relevant and important sources of information for outcome measurement in physical and rehabilitation medicine.

One of the commonly used scales to assess perceived health status of patients in the field of rheumatology and rehabilitation is the Nottingham Health Profile (NHP).^[3] The development and first validation study of the scale was conducted in the United Kingdom and, then, was used in a wide range of health problems and general population studies.^[4] The NHP is a generic patient-reported health status (or health-related quality of life) measure developed to record perceived distress of patients in physical, emotional, and social domains. It comprises 38 items (answered “Yes” or “No”) in six dimensions: physical mobility (8 questions), pain (8 questions), sleep (5 questions), emotional reactions (9 questions) social isolation (5 questions), and energy

Corresponding author: Selcen Yüksel, MD. Ankara Yıldırım Beyazıt Üniversitesi Tıp Fakültesi, Biyoistatistik Anabilim Dalı, 06760 Çubuk, Ankara, Turkey.

e-mail: selcenpehlivan@gmail.com

Cite this article as:

Yüksel S, Elhan AH, Gökmen D, Adile Küçükdeveci A, Kutlay Ş. Analyzing differential item functioning of the Nottingham Health Profile by Mixed Rasch Model. Turk J Phys Med Rehab 2018;64(x):xx-xx.

Presented at Rasch 2010- Conference Held in Copenhagen at June 13-16, 2010.

level (3 questions). The original language of the NHP is English, although adaptations into various languages, including Danish, Dutch, Finnish, French, German, Italian, Norwegian, Spanish, Swedish, and Turkish have been made.^[5] The Turkish adaptation was made in 2000 and the psychometric properties in the Turkish population were investigated in a group of patients with osteoarthritis.^[6] In the Turkish adaptation study, test-retest reliability of the subscales (dimensions) of the scale varied between 0.70 and 0.92, and internal consistency values ranged between 0.56 and 0.87, confirming its validity and reliability. Since then, NHP has been used in various studies in Turkey. In some of the articles, the NHP was used as comparator to evaluate the external construct validity of the scale which is at the main theme of content.^[7,8] The reliability and validity of the NHP were shown for various diagnostic groups.^[9-12]

The Rasch model is a mathematical measurement model developed by the Danish mathematician Georg Rasch.^[13] It can be applied in the development of a new scale, where it is possible to design the item set to fit the model expectations from the outset.^[14] Second, Rasch analysis can be used in reviewing the psychometric properties of existing ordinal scales.^[14] It is used extensively within the medical outcomes field to test the unidimensionality of scales, the ways in which their categories function, whether or not items are biased for key personal factors such as age or sex, and whether or not scales work in the same way across countries or across diagnostic groups.^[15,16]

Evaluating whether the responses given to the items in a scale differ in terms of certain personal factors such as age, sex, or clinical variable of the individual is very important to strengthen the internal construct validity of the scale. The fact that the answers given to the items in the scale do not differ in terms of personal factors is an important proof that the scale used is valid. However, responses to certain items in the data set may show bias in terms of these personal factors. This bias is referred to differential item functioning (DIF) in the Rasch literature.^[15]

One of the solutions envisaged for resolving the DIF problem is the Mixed Rasch Model (MRM) approach. The MRM is a combination of latent class analysis and Rasch model first defined by Rost.^[17] The MRM model assigns patients to latent classes by taking into account the answers they give to the items. In terms of the latent structure that the scale is expected to measure, the homogeneity of the patient responses in the latent classes and the heterogeneity

of the patient responses between the latent classes are maximum. In addition, the MRM allows item difficulties of each item to be estimated differently in latent classes, thus solving the DIF problem.^[18]

The disadvantage of the MRM is that the probability of defining latent classes with personal factors is low. For instance, if two latent class structures are provided with the MRM and sex is considered as a variable to affect the item results in the data set, the possibility of overlapping sex with two latent classes is low for most data sets. Identification of latent classes obtained from the MRM is provided by Strobl et al.^[19] with the Rasch tree method. This method recursively tests all groups that can be defined based on the available personal factors, thus preserving interpretability, but still exploring a very wide set of potential sources of DIF. In this study, the Rasch tree method was used to define latent classes derived from the MRM by personal factors.

In the literature, there is no study evaluating the item bias of the NHP in low back pain according to personal factors. In this study, using the MRM model solution for each dimension of the scale, we aimed to investigate whether the responses given to the NHP subscale items were influenced by personal factors such as age, sex, and duration of pain.

PATIENTS AND METHODS

The NHP data collected in a previous study by Elhan et al.^[20] was analyzed secondarily for this study. In that study, they included a total of 399 outpatients with low back pain who were admitted for rehabilitation between February 2007 and November 2007. A written informed consent was obtained from each patient. The study protocol was approved by the Medical Faculty of Ankara University Ethics Committee. The study was conducted in accordance with the principles of the Declaration of Helsinki.^[20] In the current study, patients with missing data according to personal factors (n=49) were excluded. Therefore, this secondary analysis included the data of 350 patients with low back pain.

Evaluating differential item functioning

The DIF of the NHP items was assessed using the MRM.^[21] A variety of statistical methods is available for detecting DIF in the Rasch model. The MRM approach^[17] can be considered as the most stringent test for the Rasch model, as it tests for item parameter differences between all possible groups of subjects, regardless of person factors.^[1] Since item responses are binary, dichotomous MRM was selected to examine

the number of latent classes. More formally, the probability that individual i answers item j correctly for dichotomous MRM is:

$$P(X_{ij} = 1 | \theta_i, \varphi_i, \beta_j) = \frac{\exp(\theta_i - \beta_j \varphi_i)}{1 + \exp(\theta_i - \beta_j \varphi_i)}$$

where θ_i is the individual i 's ability, φ_i indicates which latent class individual i belongs to, and β_j denotes the item j 's difficulty which depends on the discrete latent group variable φ .

In the MRM, information criteria are used to determine the number of latent classes the model predicts. These information criteria depend on “-2 log likelihood” statistics. If the factors such as the number of model parameters and the sample size are equal, the model with the least number of parameters in two models with equal log likelihoods is expressed as the best model. The most frequently used information criteria are Akaike and Bayesian Information Criteria. In the simulation study done by Preinerstorfer and Formann,^[22] the Bayes Information Criteria (BIC) proved a better fit for the MRM. Therefore, in this study, the BIC were used when determining latent classes with the MRM.

To provide item fit statistics of MRM solutions, the standardized Q-index of all items for both classes was given. The standardized Q-index and its probabilities indicated that the observed item response patterns for both classes did not statistically deviate from the expected patterns predicted by the model.

As mentioned above, the MRM does not provide straightforward interpretation of the latent groups with respect to manifest characteristics like age and sex. The Rasch tree method was used to define latent classes derived from the MRM.

Statistical software

The “psychotree” package in R i386.3.2.1 language^[23] and WINMIRA 2001 software^[24] were used for the Rasch tree solution and MRM respectively. Type I error rate was taken as $\alpha=0.05$ for statistical significance.

RESULTS

The NHP data of 350 patients with low back pain was analyzed. A total of 17.1% of the patients were male ($n=60$) and 82.9% were female ($n=290$). The percentage of patients up to the median age of 58 years was 69.7% ($n=244$) and over the age of 58 was 30.3% ($n=106$). A total of 45.1% of the patients ($n=158$) had pain

Table 1. Mixed Rasch Model solution for each dimension of Nottingham Health Profile

NHP dimensions	Number of latent classes	Log-likelihood statistics	Number of parameters	Bayesian information criterion	Number of optimal latent classes
Physical mobility	1	-776.60	9	1605.93	
	2	-737.41	19	1586.13	2
	3	-736.99	29	1643.86	
Pain	1	-1067.93	9	2188.57	
	2	-1052.56	19	2216.42	1
Sleep	1	-951.62	6	1938.39	
	2	-894.93	13	1866.01	2
	3	-889.23	20	1895.61	
Emotional reactions	1	-1613.66	10	3285.89	
	2	-1590.43	21	3303.88	1
Social isolation	1	-825.65	6	1686.45	
	2	-794.74	13	1665.64	2
	3	-783.35	20	1683.86	
Energy level	1	-412.07	4	847.57	
	2	-409.94	9	872.60	1

Table 2. Item fit Q-index for physical mobility dimension from 2-class Mixed Rasch Model solution

Items	Class-1		Class-2	
	Z _(Q-index)	p	Z _(Q-index)	p
NHP 10	0.52	0.30	0.40	0.35
NHP 11	-0.02	0.51	-0.17	0.57
NHP 14	0.52	0.30	1.20	0.12
NHP 17	0.01	0.50	-0.13	0.55
NHP 18	0.02	0.50	0.07	0.47
NHP 25	0.14	0.45	0.10	0.46
NHP 27	-0.01	0.51	-0.12	0.54
NHP 35	-0.61	0.73	-0.59	0.72

NHP: Nottingham Health Profile.

between 0 and 4 years, whereas 54.9% (n=192) had pain over five years or more.

As mentioned above, to solve the DIF problem with MRM, the patients are assigned to latent classes which homogeneous within itself and heterogeneous among each other. If the model predicts more than one latent class, then the result indicates that the responses of items differ according to some personal factors in these classes. If MRM suggest one latent class, it is clear that items are not affect by personal factors. It is the second step of the DIF solution to expose these latent classes in terms of personal

factors. Therefore, when analyzing data set for the DIF solution with MRM, we first found the number of latent class for each dimension; if more than one latent class existed, it was questioned that which personal factor had defined these latent classes.

Mixed Rasch Model Solution for DIF

The number of optimal latent classes for each dimension is given in Table 1.

To define these latent classes by personal factors, the item responses were assessed with respect to possible group differences related to the three

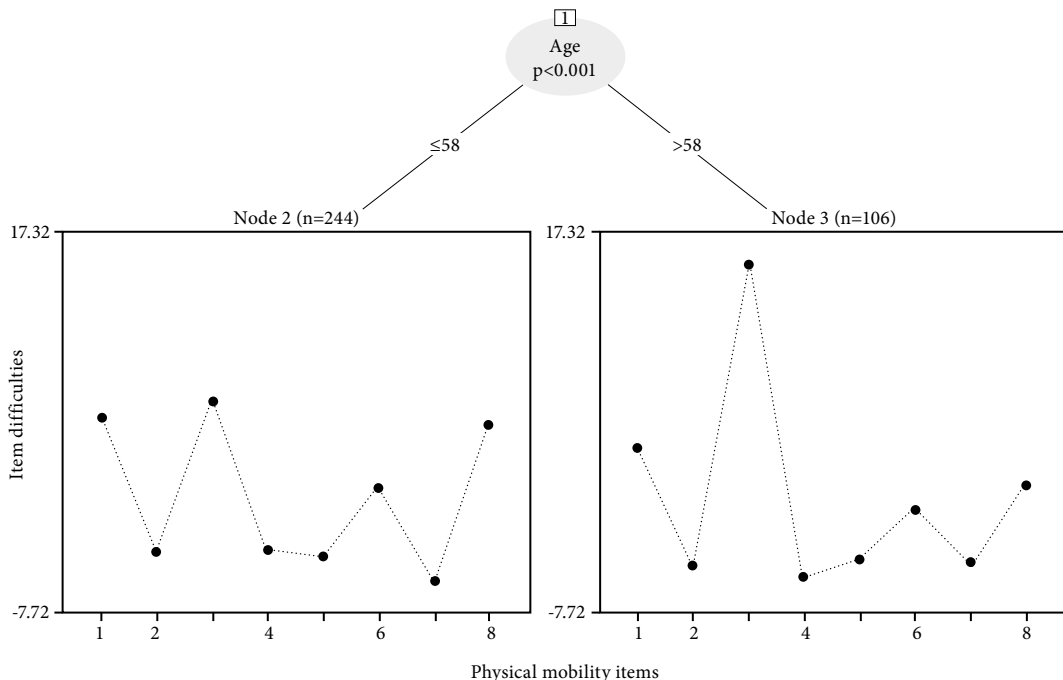


Figure 1. The Rasch tree for physical mobility exhibiting differential item functioning for age. In the terminal nodes, estimates of the item difficulty are displayed for each of the 8 items of the physical mobility dimension.

Table 3. Item difficulty estimations for physical mobility dimension of Nottingham Health Profile

	Age ≤58	Age >58
NHP 10 I can walk about only indoors	5.15	3.20
NHP 11 I find it hard to bend	-3.52	-4.41
NHP 14 I'm unable to walk at all	6.26	15.24
NHP 17 I have trouble getting up and downstairs and steps	-3.52	-5.26
NHP 18 I find it hard to reach for things	-3.89	-4.20
NHP 25 I find it hard to get dressed by myself	0.51	-0.89
NHP 27 I find it hard to stand for long	-5.64	-4.41
NHP 35 I need help to walk about outside	4.64	0.72

NHP: Nottingham Health Profile.

personal factors as age, sex, and duration of pain. In each of the terminal nodes of the tree, the item parameter estimates for the items were displayed (a high value indicates a high difficulty of the item). Each of the six dimensions of the NHP was analyzed separately.

Physical mobility

Table 1 indicates that a 2-class MRM solution is associated with the responses for the physical mobility dimension. Table 2 provides the item fit statistics of the physical mobility dimension from the 2-class MRM

Table 4. Item fit Q-index for sleep dimension from 2-class Mixed Rasch Model solution

Items	Class-1		Class-2	
	Z _(Q-index)	<i>p</i>	Z _(Q-index)	<i>p</i>
NHP 5	-0.53	0.70	-0.02	0.51
NHP 13	0.16	0.43	0.05	0.48
NHP 22	-0.29	0.61	0.01	0.50
NHP 29	0.45	0.32	0.09	0.47
NHP 33	-0.52	0.69	-0.13	0.55

NHP: Nottingham Health Profile.

Table 5. Item fit Q-index for social isolation dimension from 2-class MRM solution

Items	Class-1		Class-2	
	Z _(Q-index)	<i>p</i>	Z _(Q-index)	<i>p</i>
NHP 9	-0.04	0.52	0.10	0.46
NHP 15	0.03	0.49	-0.41	0.66
NHP 21	0.04	0.48	-0.08	0.53
NHP 30	0.18	0.43	0.55	0.29
NHP 34	-0.23	.059	-0.46	0.68

MRM: Mixed Rasch Model; NHP: Nottingham Health Profile.

Table 6. Item difficulty estimations for social isolation dimension of Nottingham Health Profile

	Male	Female
NHP 9 I feel lonely	-0.79	-1.30
NHP 15 I'm finding it hard to make contact with people	1.61	1.30
NHP 21 I feel there is nobody that I am close to	0.34	-1.11
NHP 30 I feel I am a burden to people	-1.71	-0.64
NHP 34 I'm finding it hard to get along with people	0.55	1.74

NHP: Nottingham Health Profile.

solution. The $Z_{(Q\text{-index})}$ of all 8 items for both classes revealed good item fit.

The resulting model for the physical mobility dimension, which was partitioned with respect to age, is presented in Figure 1. Following the tree from top to bottom, we found that different item parameters resulted for patients up to the age of 58 (represented in Node 2) and over the age of 58 (represented in Node 3). It is clear from Figure 1 and Table 3 that patients over the age of 58 are disadvantaged by NHP14 (they mostly say yes to that item) and that patients up to the age of 58 are disadvantaged by NHP35.

Sleep

A 2-class MRM solution was associated with the responses for the sleep dimension (Table 1). Table 4 provides the item fit statistics of the sleep dimension from the 2-class MRM solution. The $Z_{(Q\text{-index})}$ of all 5 items for both classes revealed good item fit.

For the sleep dimension, the Rasch tree showed only one single node according to age, sex, and duration of pain. However, the MRM ensured that there were two distinct groups according to the responses for the sleep dimension. No manifest characteristic was a deterministic factor for latent classes.

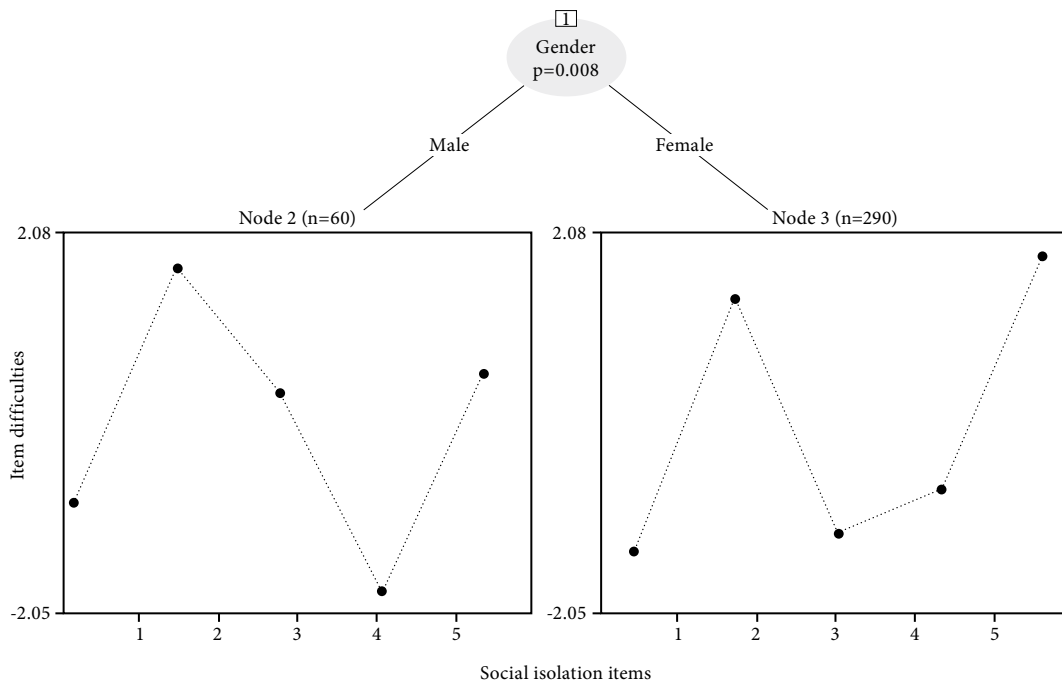


Figure 2. The Rasch tree for social isolation exhibiting differential item functioning for sex. In the terminal nodes, estimates of the item difficulty are displayed for each of the 5 items of the social isolation dimension.

Social isolation

Another 2-class MRM solution was found for the social isolation dimension (Table 1). Table 5 provides the item fit statistics of the social isolation dimension from the 2-class MRM solution. The $Z_{(Q\text{-index})}$ of all 5 items for both classes revealed good item fit.

The resulting model for the social isolation dimension, which was partitioned with respect to sex, is presented in Figure 2. One can conclude from Figure 2 and Table 6 that males were disadvantaged by NHP9 and NHP21 and that females were disadvantaged by NHP30 and NHP34.

Pain, Emotional reactions, Energy level

For the pain, emotional reactions and energy dimensions a one-class model was the best fitting model, as expected, as item fits were higher than 50% and unidimensionality was provided by the one-class solution. None of items showed DIF by age, sex, and duration of pain for these dimensions.

DISCUSSION

Differential item functioning occurs when individuals from different groups have different probabilities or likelihoods of success on an item, after they have been matched on the ability of interest. Studies of DIF use a priori known grouping information. The present study demonstrated how the MRM could be used to overcome the cause of DIF. When the response behavior can be described by a one-class model, the items function the same way for all individuals. However, if individuals have different response behavior to the items, more than one latent class may be identified. For each latent class, the MRM estimates a set of Rasch parameters for each item and individual.

In our study, DIF analysis showed that the three dimensions of the NHP (pain, energy level, and emotional reactions) were DIF-free. Although remaining three dimensions showed two latent classes, the differences between BIC values obtained for each number of latent class were low. For the physical mobility dimension, two latent classes differed with regard to age groups. For the item “*I am unable to walk at all*”, patients over the age of 58 had more distress than patients up the age of 58. This can be explained by the decrease in participation in social life with increasing age in the Turkish population. Another item that has different difficulty estimates within two classes was “*I need help to walk about outside*”, patients up the age of 58 had more distress than patients over

the age of 58. This may be due to the expectation limitation of patients over the age of 58. This result is may be artificially defined by the MRM as a result beyond clinical expectations. Therefore, although this item is considered as showing DIF by MRM, it cannot be said that the item is biased. For the sleep dimension, no personal factor was a deterministic factor to define latent classes. Two latent classes differed with regard to sex for the social isolation dimension. Males had more distress on the “*I feel lonely*” and “*I feel there is nobody that I am close to*” items than females. For the items “*I feel I am a burden to people*” and “*I’m finding it hard to get along with people*”, females had more distress than males. For the pain, emotional reactions and energy level dimensions, a one-class model was the best fitting model. Therefore, we can conclude that the items function the same way for all individuals for these dimensions. Hagell et al.^[10] found sex DIF to be more than 30% of items for emotional reaction for patients receiving care at a neurology department, but did evaluate the response behaviors of patients for each item. For the energy dimension, the results obtained from the MRM may have been influenced by the low number of items. Duration of pain did not cause any DIF for items for this data set. For the sleep dimension, no personal factor was a deterministic factor for latent classes. Thus, it is evident that more personal factors must be evaluated to expose DIF for NHP items.

Latent classes can be obtained to solve DIF problem by using MRM and these latent classes can be defined with personal factors by Rasch tree method. The clinician should be aware that items of the physical mobility dimension of the NHP can be biased by age, whereas items of social isolation dimension can have bias sex for low back pain patients. Although there were two latent classes for sleep dimension, these classes could not be explained by age, sex, and duration of pain.

In conclusion, although the three dimension showed two latent classes, it is obvious that BIC values from each number of latent classes were similar. The fact that the BIC values, which allow us to decide the optimal number of latent classes, are close to the BIC values obtained from one latent class solution, is an indicator that the DIF problem is not tragic, even in dimensions with two latent classes. Therefore, from the perspective of DIF, it can be concluded that NHP scale is a robust scale in terms of validity. However, there is a need for further studies to explain DIF for NHP with more personal factors and larger sample size.

Declaration of conflicting interests

The authors declared no conflicts of interest with respect to the authorship and/or publication of this article.

Funding

The authors received no financial support for the research and/or authorship of this article.

REFERENCES

- Gutenbrunner C, Meyer T, Melvin J, Stucki G. Towards a conceptual description of Physical and Rehabilitation Medicine. *J Rehabil Med* 2011;43:760-4.
- Küçükdeveci AA, Tennant A, Grimby G, Franchignoni F. Strategies for assessment and outcome measurement in physical and rehabilitation medicine: an educational review. *J Rehabil Med* 2011;43:661-72.
- Hunt SM, McEwen J, McKenna SP. Measuring health status: a new tool for clinicians and epidemiologists. *J R Coll Gen Pract* 1985;35:185-8.
- European Group for Quality of Life Assessment and Health Measurement. *European Guide to the Nottingham Health Profile*. Brookwood, Surrey: Brookwood Medical Publications; 1993.
- Available from: <http://www.galen-research.com/measures-database/>
- Küçükdeveci AA, McKenna SP, Kutlay S, Gürsel Y, Whalley D, Arasil T. The development and psychometric assessment of the Turkish version of the Nottingham Health Profile. *Int J Rehabil Res* 2000;23:31-8.
- Duruöz MT, Doward L, Turan Y, Cerrahoglu L, Yurtkuran M, Calis M, et al. Translation and validation of the Turkish version of the Ankylosing Spondylitis Quality of Life (ASQOL) questionnaire. *Rheumatol Int* 2013;33:2717-22.
- Kutlay S, Küçükdeveci AA, Elhan AH, Oztuna D, Koç N, Tennant A. Validation of the World Health Organization disability assessment schedule II (WHODAS-II) in patients with osteoarthritis. *Rheumatol Int* 2011;31:339-46.
- Kutlay S, Nergizoglu G, Kutlay S, Keven K, Erturk S, Ates K, et al. General or disease specific questionnaire? A comparative study in hemodialysis patients. *Ren Fail* 2003;25:95-103.
- Hagell P, Whalley D, McKenna SP, Lindvall O. Health status measurement in Parkinson's disease: validity of the PDQ-39 and Nottingham Health Profile. *Mov Disord* 2003;18:773-83.
- Busija L, Pausenberger E, Haines TP, Haymes S, Buchbinder R, Osborne RH. Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQoL). *Arthritis Care Res (Hoboken)* 2011;63 Suppl 11:S383-412.
- Post MW, Gerritsen J, van Leusen ND, Paping MA, Prevo AJ. Adapting the Nottingham Health Profile for use in people with severe physical disabilities. *Clin Rehabil* 2001;15:103-10.
- Bechger TM, Maris G. A Statistical Test for Differential Item Pair Functioning. *Psychometrika* 2015;80:317-40.
- Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis Rheum* 2007;57:1358-62.
- Tennant A, Penta M, Tesio L, Grimby G, Thonnard JL, Slade A, et al. Assessing and adjusting for cross-cultural validity of impairment and activity limitation scales through differential item functioning within the framework of the Rasch model: the PRO-ESOR project. *Med Care* 2004;42(1 Suppl):I37-48.
- Elhan AH, Kutlay S, Küçükdeveci AA, Cotuk C, Oztürk G, Tesio L, et al. Psychometric properties of the Mini-Mental State Examination in patients with acquired brain injury in Turkey. *J Rehabil Med* 2005;37:306-11.
- Rost J. Rasch models in latent classes: an integration of two approaches to item analysis. *Applied Psychological Measurement*, 1990;14:271-82.
- Yüksel S. Ölçeklerde saptanan madde işlev farklılığının Karma Rasch modelleri ile incelenmesi [Tez]. Ankara: Ankara Üniversitesi Sağlık Bilimleri Enstitüsü Biyoistatistik Anabilim Dalı; 2012.
- Strobl C, Kopf J, Zeileis A. Rasch Trees: A New Method for Detecting Differential Item Functioning in the Rasch Model. *Psychometrika* 2015;80:289-316.
- Elhan AH, Oztuna D, Kutlay S, Küçükdeveci AA, Tennant A. An initial application of computerized adaptive testing (CAT) for measuring disability in patients with low back pain. *BMC Musculoskelet Disord* 2008;9:166.
- Rasch G. *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press; 1960.
- Preinerstorfer D, Formann AK. Parameter recovery and model selection in Mixed Rasch Models. *Br J Math Stat Psychol* 2012;65:251-62.
- R: A language and environment for statistical computing. Vienna, Austria; 2008. Available from: <https://cran.r-project.org/bin/windows/base/>
- Von Davier M. WINMIRA 2001: Software and user manual. Available from: <http://208.76.80.46/~svfklumu/wmira/index.html>