The Development of a Short Measure of Physical Function for Knee OA

KOOS-Physical Function Short-form (KOOS-PS) – An OARSI/OMERACT Initiative

Anthony V. Perruccio[1], L. Stefan Lohmander[2], Mayilee Canizares[3], Alan Tennant[4], Gillian A. Hawker[5], Philip G. Conaghan[6], Ewa M. Roos[7], Joanne M. Jordan[8], Jean-Francis Maillefert[9], Maxime Dougados[10], Aileen M. Davis[11]

1. Division of Health Care and Outcomes Research and Arthritis Community Research and Evaluation Unit, Toronto Western Research Institute; Department of Public Health Sciences, University of Toronto, Toronto, Canada
2. Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden
3. Division of Health Care and Outcomes Research and Arthritis Community Research and Evaluation Unit, Toronto Western Research Institute, Toronto, Canada
4. Academic Unit of Musculoskeletal and Rehabilitation Medicine, University of Leeds, Leeds, UK
5. Division of Rheumatology, Department of Medicine, Women's College Hospital, Departments of Medicine and Health Policy, Management and Evaluation, University of Toronto, Toronto, Canada
6. Academic Unit of Musculoskeletal Disease, University of Leeds, Leeds, UK
7. Department of Orthopaedics, Clinical Sciences Lund, Lund University, Lund, Sweden; Institute of Sports Science and Clinical Biomechanics, University of Southern Denmark, Odense, Denmark
8. Departments of Medicine and Orthopaedics, Thurston Arthritis Research Center, University of North Carolina, Chapel Hill, NC
9. Dijon University Hospital and INSERM U887, University of Burgundy, Dijon, F-21079, France
10. Department of Rheumatology, Faculty of Medicine, Cochin Hospital Paris, France
11. Division of Health Care and Outcomes Research and Arthritis Community Research and Evaluation Unit, Toronto Western Research Institute; Departments of Physical Therapy, Rehabilitation Science, Health Policy, Management and Evaluation and Institute of Medical Science, University of Toronto, Toronto, Canada

Corresponding Author: Anthony V Perruccio
Division of Health Care and Outcomes Research
and Arthritis Community Research and Evaluation Unit
Toronto Western Research Institute
399 Bathurst Street - MP10-322
Toronto, Ontario M5T 2S8
Tel: 416 603-6269
Fax: 416 603-6288
Email: perrucci@uhnres.utoronto.ca

Running headline: KOOS-Physical Function Short-form (KOOS-PS)

Abstract

Objective: To develop a short measure of physical function for knee osteoarthritis (OA) using multi-national data from individuals with varying degrees of severity of knee OA.

Methods: Rasch analysis, based on the partial credit model, was conducted on KOOS/WOMAC data from individuals with knee OA, ranging from community to pre-total knee replacement samples from 5 countries. Fit of the data to the Rasch model was evaluated by overall model fit and item-level fit statistics (chi-square, size of residual, F-test). Invariance across age, gender and country was evaluated. Unidimensionality was evaluated by factor analysis of residuals. The derived short measure was further tested for fit through re-analyses in individual sub-samples. A nomogram converting raw summed scores to Rasch-derived interval scores was developed.

Results: Thirteen data sets were included (n=2145), with an age range of 26 to 95 years, and a male/female ratio of 1:1.4. The final model included seven of the original 22 items. From easiest to most difficult, the items (logit) were: rising from bed (1.366), putting on socks/stockings (1.109), rising from sitting (0.537), bending to the floor (0.433), twisting/pivoting on injured knee (-0.861), kneeling (-1.292) and squatting (-1.292). Sub-sample analyses confirmed findings.

Conclusion: Based on the use of accepted Rasch-based measurement methods and the compliment of countries, languages and OA-severity represented in this study, our seven item short measure of physical function for knee OA is likely generalizable and widely applicable. This measure has potential for use as the function component in an OA severity scoring system.

INTRODUCTION

This paper is a companion to the publication by Davis et al[1]. The distinction is the latter's focus on osteoarthritis (OA) of the hip, while the present focus is on knee OA. The conceptualization, objective and methods of the studies are shared and the reader is directed to Davis' et al publication for a detailed description of the background and methodology of the study. A brief synopsis is provided here.

OA, particularly of the hip and knee, is a major cause of disability globally [2-5]. It has a high impact due to the prevalence and worsening of pain and physical functioning over time [6-10], with concomitant reductions in independence and quality of life [11-13].

To date, interventions for hip and knee OA-associated pain and loss of function have been predominantly focused on end stage OA, with total joint replacement (TJR) regarded as the most effective treatment for severe hip and knee OA [14;15]. However, due to its slow progression over time, studies intended to examine and understand the natural course of hip and knee OA, which may lead to TJR, have been limited. This lack of knowledge has made it difficult to assess the progression of OA severity and to test interventions that might alter the course of disease, pain and disability in these individuals. The use of 'time to TJR' as an endpoint is problematic as there is known variability in the decision to perform and undergo surgery [16-21]. Understanding states of hip and knee OA is critical for a number of reasons, including improved definition of eligibility criteria for clinical trials, defining criteria for TJR and for evaluating outcomes from non-surgical intervention studies.

As described by Gossec et al [22], and recognizing the need for an adequate measure of OA severity, an OARSI and OMERACT international working group was established to evaluate issues related to OA severity and to construct a composite measure capable of defining severity states in OA of the hip and knee. It was decided that three domains should be included in this composite measure of severity: pain, functional status and structural damage.

Our focus was the functional status component of this composite measure. The objective was to develop physical function states that represent the progression of physical disability from early to late disease for individuals with OA of the knee. The aim was that the set of items identified as being adequate to classify or categorize severity based on functional status should represent a unidimensional construct, be free of age, gender and cultural biases, represent difficulties across the range of OA severity and be parsimonious.

The Western Ontario McMaster Universities' Osteoarthritis Index (WOMAC) [23-25] and the Knee injury and Osteoarthritis Outcome Score (KOOS) [26;27] are the most common measures of functional status that have been used for knee OA. Seventeen items make up the WOMAC physical function subscale, selected on the basis of their relative significance to people with OA of the hip and knee [23-25]. In addressing activities of daily living (ADL) alone, it has been put forward that these 17 items, failing to capture physically more demanding activities, are limited in range of difficulty [28;29]. The KOOS [26] was in part developed to address this gap. It subsumes the 17 physical function items of the WOMAC 3.0 (ADL component), and further includes 5 higher-level items, sport and recreational activities (Sport/Rec component), increasing the complement of functional status items to 22. The WOMAC 3.0 and KOOS items are similarly scaled, scored 0 to 4 with response options for rating the amount of difficulty on an activity ranging from 'None' to 'Extreme'. Raw item scores are summed to compute total scores, ranging from 0 to 68 for the WOMAC 3.0 physical function scale, and 0 to 88 for the two KOOS subscales. For the KOOS subscales, the raw subscale scores are then calculated as a percentage score. The raw responses of the 22 items for the two KOOS subscales were used for these analyses.

A series of methodological studies have supported the reliability and validity of the WOMAC and KOOS. The KOOS ADL and Sport/Rec construct validity have been determined in comparison with the SF-36 physical function subscale [30] with correlations ranging from 0.5 to 0.7 in U.S.- and Swedish-based studies [26;31], and content and face validity by a panel of patients and orthopaedic surgeons [26]. The internal-consistency reliability of the ADL and Sport/Rec components were found to be 0.9 and 0.8, respectively, in the Swedish study [31]. The reliability, validity, and responsiveness of the physical function subscale of the WOMAC have been well studied and demonstrated in a number of studies across a range of patient groups and interventions and many of its psychometric properties are known [23;24;32].

The Rasch one-parameter model, based on item response theory, has been used to construct, evaluate and shorten the structure of measures [33-35]. It is a probabilistic method that makes possible the examination of unidimensionality and ordering of items on a measurement continuum. This model is used to place items and persons on a common measurement scale. Using the Rasch model, we analyzed raw WOMAC 3.0 and KOOS 2.0 data from individuals with knee OA accrued to North American and European studies, ranging from community cohorts to individuals awaiting total knee replacement surgery (TKR), with the aim of developing a short measure of OA-related physical function across the OA spectrum.

METHODS

A description of the sub-samples comprising our larger sample is presented in Table 1. Country contributions included Sweden (5 samples), Canada (4 samples), France (2 samples), Estonia (1 sample), and the Netherlands (1 sample). Individual study accrual methods have been described elsewhere (cited in Table 1). For those pre-TKR, patients were booked for their surgery and completed the questionnaires either as a part of routine care or in relation to a specific research study. The sample for Study 5 is made up of individuals from a knee OA clinical trial group, recruited from outpatient clinics. Ages for the combined sample ranged from 26 to 95 years and the ratio of males to females was 1:1.4; total n=2145. All data were based on the WOMAC Likert-type version 3.0 or KOOS Likert-type version 2.0 questionnaires. This secondary analysis was approved by the institutional ethics review board.

[Table 1]

Analysis

The logistic function of the Rasch model specifies that the probability of an individual endorsing a particular item is dependent on the individual's ability and the difficulty of the item. The Rasch model allows us to estimate person abilities, based on physical functioning, and item difficulties along a shared measurement scale. A linear, interval level scale is achieved with Rasch measurement. An extension of this model is the partial credit model [33;36;37], appropriate for multiple response option data and where no assumption is made as to the equivalence of the difficulty of moving through item categories between items. All analyses were carried out using the partial credit model and RUMM 2020 software [38].

Our iterative analyses commenced with an assessment of items in the total sample. The final items retained from the total sample were subsequently evaluated in four sub-samples, independently, to ensure that results were consistent. The sub-samples were characterised by relative OA severity and included only those for which an n ≥100 was available, to ensure sufficient sample size for testing within class intervals, and for which final selected items were available; the samples were the community sample (study no. 1, 2, and 3), OA cohort sample (study no. 5), OA biomarker sample (study no. 6) and the pre-TKR sample (study no. 10, 11, 12, and 13). The criteria for interval level data included fit of the data to the model, demonstration of appropriate response category ordering, lack of item bias or differential item functioning (DIF) and unidimensionality. For all study analyses, statistical significance was based on a critical value of 0.05 with a Bonferroni correction factor for multiple testing; p-values <0.002 were considered statistically significant.

The first step of analyses involved assessing summary statistics of measures of fit. Under ideal circumstances, item and person score fit residuals have means of 0 and standard deviations of 1. The internal consistency and reliability of the model was evaluated by a person separation index that is equivalent to Cronbach's alpha [39]. Values of approximately 0.80 and greater are acceptable [40]. Finally, overall fit of the data to the Rasch model was assessed by a χ^2 test, where a non-significant χ^2 is interpreted as evidence of good fit.

Item response categories were examined to determine if they produced sequentially ordered thresholds. Response categories were collapsed if misordered thresholds were found.

Rasch analysis allows for a determination of the undimensionality of a set of items based on goodness of fit statistics for each item. Item fit represents the consistency between observed and expected data. The data were considered to fit the Rasch model when individual item χ^2 probabilities were not significant, item residuals were small (absolute value < 2.5) and F-test statistics were not significant. Items displaying misfit were first evaluated for DIF.

Rasch modelling makes possible an evaluation of variation of item characteristics across different samples [33;41], and therefore the construction of an invariant construct of functional status (i.e. free of DIF) that can be used to compare abilities and discriminate between levels of OA severity based on physical functioning. We evaluated DIF by age, gender and country. Age was dichotomized into <65 years of age and ≥65 years. Items displaying DIF were sequentially removed and not retained in subsequent iterative analyses. Throughout the analyses, qualifying items were removed one at a time and model and item fit re-evaluated.

Finally, principal component analysis of the residuals was carried out to ensure that remaining items conformed to a unidimensional construct. If unidimensionality holds, no factor structure should be found in the residuals. Person score estimates were compared based on subsets of items from this factor analysis. Scores were generated from independent sets of items, items with positive factor loadings of 0.30 and higher and items with negative factor loadings of

-0.30 and lower. T-tests were used to compare the estimates and the percentage of tests outside ± 1.96 (95% confidence interval) was calculated [42;43].

To make the results of the present study applicable and practical in both research and clinical settings, we include a formula which can be used to convert a raw summed score of the final items to the equivalent Rasch-based person score. The formula was developed from the fitting of a cubic model, regressing Rasch-based person scores on the raw summated scores. Only individuals with complete data on the final items (n=1154) were included in this analysis. A table is provided whereby one can quickly determine the person score based on the raw summed scores.

RESULTS

Table 2 presents summary measures of model fit from the first iteration of analyses for the total sample. With all 22 items included, the large standard deviations for the item fit residuals and the highly significant overall model χ^2 indicated poor fit of the data to the Rasch model. Except for item *twisting/pivoting on injured* knee, all items displayed misfit to the model, identified by a significant χ^2 or F-statistic probability (using a Bonferroni corrected p-value of 0.002) or a large residual (>2.5) (data not shown). *Jumping* was the only item that showed disordered thresholds and was therefore rescored.

The process of eliminating items, one at a time, and re-evaluating model fit at each iteration began with an evaluation of DIF for those items displaying misfit. As a consequence of finding significant DIF, eliminated items included *sitting, lying in bed, getting in/out of bath/shower,* and *heavy household chores* which showed DIF by both age and gender, and *running, getting on/off toilet, ascending* and *descending stairs* and *standing* which displayed DIF by age. Each of these items also displayed DIF by country. Finally, although not displaying DIF by age and gender, six further items showed DIF by country and were sequentially eliminated. Along with the summary measures of model fit for the initial model, Table 2 also includes the same measures for three models evaluated in the process of arriving at the final, well-fitting model. These are included for illustration purposes to show the sequential improving model fit from initial to final model analysis.

[Table 2]

Following the iterative process, seven items remained, each showing good fit to the model. The summary measures for this final model are shown in Table 2 and item level statistics are presented in Table 3. The overall model χ^2 indicated good fit to the Rasch model, with a χ^2 of 73.34 and p=0.1751. An example of the absence of DIF for three of these items is displayed in Figure 1. Panel A, Figure 1, shows the item characteristic curves for item *rise from sitting* by age, Panel B item *twisting/pivoting on injured knee* by gender, and Panel C item *rising from bed* by country. The overlap of curves is indicative of no DIF. Principal component analysis of the residuals did not reveal any detectable patterns or systematic information, suggesting that the retained items formed a unidimensional structure. Rotated factor loadings showed that no more than one item loaded on each of the components, indicating that none was highly correlated with the others. Each item had a factor loading >0.90 on their respective component and loadings <0.22 on all other components. Further confirming the unidimensional structure, the t-test comparing person scores, estimated using the two sets of items distinguished by positive and negative factor loadings from the residual analyses, indicated that only 3.6% of observations lay outside the 95% confidence interval (Table 3).

[Table 3]

[Figure 1]

The internal consistency and reliability of the final model was very good with a person separation index of 0.904 (Tables 2 and3). The seven items covered a wide range of difficulties with mean item locations (logit values) ranging from –1.3 for item *squatting* (the most difficult item) to 1.4 for item *rising from bed* (the easiest item) (Table 3); threshold values ranged from –2.75 to 4.32. The Rasch-based person scores ranged from –4.37 to 5.59. The distribution of item thresholds and person scores of the person-item threshold distribution in Figure 2 are on the same scale and show that a substantial number of persons are closely targeted to the items and vice versa. As well, the means of items and persons is similar at 0 and 0.205, respectively (Table 2 and Figure 2). As expected, the distribution of person scores (or abilities) along the underlying physical functioning measure captured by the seven items corresponded with the relative severity of OA. Thus, the distribution of abilities for the community sample, for example, was nearest the most difficult items, and the distribution for the pre-TKR sample (individuals with more difficulty) nearest the easier items. Figure 3 displays the distribution of Rasch-based person scores, scaled 0 to 100, by sample type.

[Figure 2]

[Figure 3]

Finally, data for the seven items were tested for fit to the Rasch model in four sub-samples. With some slight variations between samples, not unexpected due to sample severity differences, individual sub-sample results were consistent with total sample results, displaying a well-fitting model (Table 4). No item misfit was detected in sub-sample analyses.

[Table 4]

A scatter plot of the raw summed scores for the seven items and the Rasch-based person scores (or abilities) is shown in Figure 4. Included as well in Figure 4 is the fitted curve from a cubic model estimation regressing the Rasch-based person scores on the raw summed score. Model summary and coefficient estimates are presented in Table 5, as are descriptives for the raw summed scores, observed and predicted Rasch-based person scores and residuals. Finally,

Figure 5 displays the scatter plot of the observed and predicted person scores. Results presented in Table 5 and Figure 5 are consistent, and are indicative of the appropriateness of the cubic model for prediction purposes throughout the range of values.

[Figure 4]

[Table 5]

[Figure 5]

Results from the fitting of the cubic model (Table 5) allow for the estimation of a Rasch-based person score (RPS) based on the raw summed score from the seven items (rSS), where

$$RPS = -4.214126559441 + (0.5698144707377)*(rSS) + (-0.0336880193327)*(rSS)^2 + (0.0009162754582646)*(rSS)^3$$

Table 6 provides Rasch-based, model predicted person score estimates for values of raw summed scores. Predicted estimates are shown both in their original scale and rescored on a 0 to 100 scale.

[Table 6]

DISCUSSION

The goal of this study was to develop a short measure of physical functioning for OA of the knee. Using data from individuals presenting with varying degrees of OA, from community samples to TKR candidates, and spanning a number of North American and European countries, we report that the Rasch model supports a short measure, the 7-item KOOS-PS (Appendix 1). The items display consistency, with an underlying unidimensional construct, are free of differential item functioning based on age, gender and country, display overall reliability, and cover a range of physical functioning difficulty. As a result, a Rasch-based, interval level, single KOOS-PS score can be computed for individuals.

By virtue of the prolonged disease evolution of OA over time, the preponderance of measures available in the literature and in use clinically for knee OA have been specific to end stage, severe OA. This makes it difficult to assess the effectiveness of potential disease and symptom course-altering interventions. The work presented here, based on functional status, is a component step in the development of a useful measure to fill this gap. This short measure has the potential to be used to characterize OA states and establish criteria for surgery [22], in combination with pain and structural damage components, for use in research and intervention studies, and by clinicians.

In recognition of the need to join together component, complimentary measures into a composite index, it was determined that the original 22 items were too long for defining physical function states for our purposes. As well, the number of items of the WOMAC physical function subscale in combination with its limited range in difficulty has raised issues of redundancy within the scale [29;44]. Also, the demand by regulators and the burden on study participants and clinicians, issues of feasibility and compliance, must be considered and balanced in the development of a measure. Increasing the parsimony of a measure must not compromise its validity and responsiveness. We tested the fit of the seven items retained from the total sample in individual sub-samples characterised by relative OA severity and report very reasonable fit in each of these samples, with minor variations between them and results from the total sample. Further work and expert opinion are needed for validation studies and to define severity cut points to establish, for instance, common eligibility criteria for surgical intervention.

While additional studies are needed to evaluate potential item variations across additional cultures and OA groups, the compliment of countries, languages and OA-severity represented in this study give sufficient reason to believe that our short measure of physical functioning, based on accepted Rasch-based measurement methods, is generalizable, widely applicable, and feasible for use in both research and clinical settings. We have largely addressed the applicability requirements of the 'OMERACT filter' [45] in this study. However, as next steps, validation studies to define a cutpoint in the KOOS-PS, given its proposed use in a composite measure, and studies assessing the responsiveness of the KOOS-PS over time (i.e. measuring change) are needed.

Acknowledgement of funding sources

New Emerging Team Grant in Early OA sponsored by Canadian Arthritis Network and Canadian Institutes of Health Research

Pfizer

Novartis

Negma

Astra Zeneca

AV Perruccio is funded by a CIHR Canada Graduate Scholarship

There was no involvement of the funding sources in the study design, data collection, analysis and interpretation of the data, in the writing of the manuscript or in the decision to publish the work.

Reference List

(1) Davis AM, Perruccio AV, Canizares M, Tennant A, Hawker GA, Conaghan PG et al. The development of a short measure of physical function for hip OA. An OARSI/OMERACT initiative. Osteoarthritis Cartilage 2007; submitted.

(2) Kaplan W, Laing R. Priority Medicines for Europe and the World. 2004. WHO:Department of Essential Drugs and Medicines Policy.

(3) Brooks PM. Impact of osteoarthritis on individuals and society: how much disability? Social consequences and health economic implications. Curr Opin Rheumatol 2002; 14(5):573-577.

(4) Brooks PM. The burden of musculoskeletal disease--a global perspective. Clin Rheumatol 2006; 25(6):778-781.

(5) Woolf AD, Pfleger B. Burden of major musculoskeletal conditions. Bull World Health Organ 2003; 81(9):646-656.

(6) More Hip and Knee Replacements Performed on Younger Canadians. 2003. Canadian Institute for Health Information.

(7) Badley EM, Rasooly I, Webster GK. Relative importance of musculoskeletal disorders as a cause of chronic health problems, disability, and health care utilization: findings from the 1990 Ontario Health Survey. J Rheumatol 1994; 21(3):505-514.

(8) Badley EM, Webster GK, Rasooly I. The impact of musculoskeletal disorders in the population: are they just aches and pains? Findings from the 1990 Ontario Health Survey. J Rheumatol 1995; 22(4):733-739.

(9) Perruccio AV, Power JD, Badley EM. Revisiting arthritis prevalence projections--it's more than just the aging of the population. J Rheumatol 2006; 33(9):1856-1862.

(10) Power JD, Perruccio AV, Desmeules M, Lagace C, Badley EM. Ambulatory physician care for musculoskeletal disorders in Canada. J Rheumatol 2006; 33(1):133-139.

(11) Buckwalter JA, Saltzman C, Brown T. The impact of osteoarthritis: implications for research. Clin Orthop Relat Res 2004;(427 Suppl):S6-15.

(12) Elders MJ. The increasing impact of arthritis on public health. J Rheumatol Suppl 2000; 60:6-8.

(13) Salaffi F, Carotti M, Stancati A, Grassi W. Health-related quality of life in older adults with symptomatic hip and knee osteoarthritis: a comparison with matched healthy controls. Aging Clin Exp Res 2005; 17(4):255-263.

(14) Zhang W, Doherty M, Arden N, Bannwarth B, Bijlsma J, Gunther KP et al. EULAR evidence based recommendations for the management of hip osteoarthritis: report of a

task force of the EULAR Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). Ann Rheum Dis 2005; 64(5):669-681.

(15) Jordan KM, Arden NK, Doherty M, Bannwarth B, Bijlsma JW, Dieppe P et al. EULAR Recommendations 2003: an evidence based approach to the management of knee osteoarthritis: Report of a Task Force of the Standing Committee for International Clinical Studies Including Therapeutic Trials (ESCISIT). Ann Rheum Dis 2003; 62(12):1145-1155.

(16) Ballantyne PJ, Gignac MA, Hawker GA. A patient-centered perspective on surgery avoidance for hip or knee arthritis: lessons for the future. Arthritis Rheum 2007; 57(1):27-34.

(17) Hawker GA, Wright JG, Coyte PC, Williams JI, Harvey B, Glazier R et al. Determining the need for hip and knee arthroplasty: the role of clinical severity and patients' preferences. Med Care 2001; 39(3):206-216.

(18) Hawker GA. Who, when, and why total joint replacement surgery? The patient's perspective. Curr Opin Rheumatol 2006; 18(5):526-530.

(19) Hudak PL, Clark JP, Hawker GA, Coyte PC, Mahomed NN, Kreder HJ et al. "You're perfect for the procedure! Why don't you want it?" Elderly arthritis patients' unwillingness to consider total joint arthroplasty surgery: a qualitative study. Med Decis Making 2002; 22(3):272-278.

(20) Maillefert JF, Hawker GA, Gossec L, Mahomed NN, Lohmander S, Dieppe PA et al. Concomitant therapy: an outcome variable for musculoskeletal disorders? Part 2: total joint replacement in osteoarthritis trials. J Rheumatol 2005; 32(12):2449-2451.

(21) Woolhead GM, Donovan JL, Chard JA, Dieppe PA. Who should have priority for a knee joint replacement? Rheumatology (Oxford) 2002; 41(4):390-394.

(22) Gossec L, Hawker G, Davis AM, Maillefert JF, Lohmander LS, Altman R et al. OMERACT/OARSI initiative to define states of severity and indication for joint replacement in hip and knee osteoarthritis. J Rheumatol 2007; 34(6):1432-1435.

(23) Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988; 15(12):1833-1840.

(24) Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of the WOMAC: a health status instrument for measuring clinically-important patient relevant outcomes following total hip or knee arthroplasty in osteoarthritis. J Orthop Rheumatol 1988; 1:95-108.

(25) Bellamy N. WOMAC osteoarthritis index. User Guide IV. 2000. 2000. London.

(26) Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)--development of a self-administered outcome measure. J Orthop Sports Phys Ther 1998; 28(2):88-96.

(27) Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. Health Qual Life Outcomes 2003; 1(1):17.

(28) Davis AM, Badley EM, Beaton DE, Kopec J, Wright JG, Young NL et al. Rasch analysis of the Western Ontario McMaster (WOMAC) Osteoarthritis Index: results from community and arthroplasty samples. J Clin Epidemiol 2003; 56(11):1076-1083.

(29) 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 1999; 12(5):331-335.

(30) Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care 1992; 30(6):473-483.

(31) Roos EM, Roos HP, Ekdahl C, Lohmander LS. Knee injury and Osteoarthritis Outcome Score (KOOS)--validation of a Swedish version. Scand J Med Sci Sports 1998; 8(6):439-448.

(32) McConnell S, Kolopack P, Davis AM. The Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC): a review of its utility and measurement properties. Arthritis Rheum 2001; 45(5):453-461.

(33) Bond TG, Fox CM. Applying the Rasch Model: Fundamental Measurement in the Human Sciences. Mahwah, NJ: Lawrence Erlbaum Associates, Inc., 2001.

(34) Hambleton RK, Swaminathan H, Rogers HJ. Fundamentals of Item Response Theory. Newbury Park, CA: Sage Publications, 1991.

(35) Cole JC, Rabin AS, Smith TL, Kaufman AS. Development and validation of a Rasch-derived CES-D short form. Psychol Assess 2004; 16(4):360-372.

(36) Wright BD, Masters GN. Rating Scale Analysis. Chicago, IL: MESA Press, 1982.

(37) Masters GN. A Rasch Model for Partial Credit Scoring. Psychometrika 1982; 47(2):149-174.

(38) RUMM 2020 Version 4.0 for Windows. Perth: RUMM Laboratory Pty Ltd., 2004.

(39) Cronbach LJ. Coefficient Alpha and the Internal Structure of Tests. Psychometrika 1951; 16(3):297-334.

(40) Nunnally JC, Bernstein IH. Psychometric Theory. 3 ed. The Mcgraw-Hill Companies, 1994.

(41) Andrich D. Rasch Models for Measurement. Newbury Park, CA: Sage Publications, 1988.

(42) Pallant JF, Tennant A. Unidimensionality Matters. Rasch Measurement Transactions 2006; 20(1):1048-1051.

(43) Smith EV, Jr. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. J Appl Meas 2002; 3(2):205-231.

(44) Sun Y, Sturmer T, Gunther KP, Brenner H. Reliability and validity of clinical outcome measurements of osteoarthritis of the hip and knee--a review of the literature. Clin Rheumatol 1997; 16(2):185-198.

(45) Boers M, Brooks P, Strand CV, Tugwell P. The OMERACT filter for Outcome Measures in Rheumatology. J Rheumatol 1998; 25(2):198-199.

(46) Thorstensson CA, Petersson IF, Jacobsson LT, Boegard TL, Roos EM. Reduced functional performance in the lower extremity predicted radiographic knee osteoarthritis five years later. Ann Rheum Dis 2004; 63(4):402-407.

(47) Ornetti P, Parratte S, Gossec L, Tavernier C, Argenson JN, Roos EM et al. Cross-cultural adaptation and validation of the French version of the Knee injury and Osteoarthritis Outcome Score (KOOS) in knee osteoarthritis patients. Osteoarthritis Cartilage 2007.

(48) Tamm A, Lintrop M, Veske K, Hansen U, Tamm A. Prevalence of Patello- and Tibiofemoral Osteoarthritis in a Population in Southern–Estonia Aged 34-55: Elva Study. J Rheumatol 2007;(in press).

(49) Pham T, Maillefert JF, Hudry C, Kieffert P, Bourgeois P, Lechevalier D et al. Laterally elevated wedged insoles in the treatment of medial knee osteoarthritis. A two-year prospective randomized controlled study. Osteoarthritis Cartilage 2004; 12(1):46-55.

(50) Dahl A, Toksvig-Larsen S, Roos EM. A 2-year prospective study of patient-relevant outcomes in patients operated on for knee osteoarthritis with tibial osteotomy. BMC Musculoskelet Disord 2005; 6:18.

(51) Groot IBd, Favejee MM, Reijman M, Verhaar JAN, Terwee JB. The Dutch version of the Knee Injury and Osteoarthritis Outcome Score: a validation study. Health Qual Life Outcomes 2007;(in press).

Table 1. Study sample descriptions.

Study	Country	Type of Sample	N	Mean age (sd, range)	Sex* M:F	Measure
1[46]	Sweden	Community	120	56.7 (5.8, 47-66)	66:54	KOOS
2[17]	Canada	Community	360	76.5 (7.0, 58-95)	77:283	KOOS
3	Sweden	Community	118	58.8 (1.1, 37-85)	89:29	KOOS
4[47]	France	Clinical	67	-	-	KOOS
5	Sweden	OA cohort	181	62.4 (8.8, 42-81)	84:97	KOOS
6[48]	Estonia	OA biomarker study	161	45.4 (6.1, 32-55)	61:100	KOOS
7[49]	France	Medial wedge	166	65.7 (10.8, 38-91)	43:123	WOMAC
8[50]	Sweden	Pre-Osteotomy	58	54.3 (6.9, 36-69)	30:28	KOOS
9[51]	Netherlands	Post-ACL Post-Osteotomy Pre-TKR OA cohort	36 63 47 54	59.0 (15.7, 27-89)	104:94	KOOS
10	Canada	Pre-TKR	140	68.7 (9.8, 38-89)	45:95	WOMAC
11[27]	Sweden	Pre-TKR	105	71.3 (8.5, 43.86)	39:66	KOOS
12	Canada	Pre-TKR	223	68.4 (9.5, 26-89)	135:87	WOMAC
13	Canada	Pre-TKR	246	65.0 (10.5, 30-89)	91:154	KOOS

TKR=total knee replacement
*M:F ratio does not equal the sample size in some cases due to missing data.

Table 2. Summary measures of model fit: initial, selected mid-point and final models.

		Items				Persons				Item-trait Int.*		
		Location		Fit Residual		Location		Fit Residual				
Model	**n(no. items)**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	χ^2	χ^2 **prob**	**PSI****
Initial	2063 (22)	0.000	1.041	-0.581	4.734	-0.195	1.961	-0.448	1.567	1044	0.0000	0.972
5	2059 (19)	0.000	1.263	-0.610	4.037	0.001	2.117	-0.429	1.374	800.1	0.0000	0.970
10	2056 (14)	0.000	1.299	-0.533	2.221	0.263	2.045	-0.409	1.223	332.9	0.0000	0.956
15	2049 (10)	0.000	1.462	-0.740	1.234	0.440	1.965	-0.409	1.087	166.2	0.0000	0.931
Final	2037 (7)	0.000	1.129	-0.578	1.416	0.205	1.955	-0.474	1.124	73.34	0.1751	0.904

* Item-trait Int: Item-trait interaction; **PSI: Person Separation Index

Table 3. Total sample - final model with 7 items.

Item	Location	SE	Residuals	X^2	p-value	F-stat	p-value
Squatting	-1.292	0.038	0.304	5.07	0.8286	0.474	0.8928
Kneeling	-1.292	0.037	-0.585	12.14	0.2056	1.200	0.2910
Twisting/pivoting on injured knee	-0.861	0.036	-1.256	17.75	0.0382	2.549	0.0067
Bending to floor	0.433	0.030	2.058	9.24	0.4154	0.954	0.4771
Rising from sitting	0.537	0.032	-0.874	7.83	0.5519	1.098	0.3607
Putting on socks/stockings	1.109	0.031	-1.323	7.75	0.5598	1.073	0.3793
Rising from bed	1.366	0.032	-2.369	13.58	0.1381	2.072	0.0289
Mean	0.000		-0.578	**Chi-square**		73.34	
SD	1.129		1.416	**p-value**		0.1751	
Separation Index	0.904						
T-test for unidimensionality	Proportion outside 95% CIs				3.6%		

Table 4. Summary measures of model fit from sub-sample analyses; final model with 7 items.

		Items				Persons				Item-trait Int.*		
		Location		Fit Residual		Location		Fit Residual				
Sub-sample	**n**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	**Mean**	**SD**	χ^2	χ^2 **prob**	**PSI****
OA Biomarker	151	0.000	1.010	-0.061	0.954	-2.567	1.977	-0.340	1.025	43.30	0.9726	0.891
Community	530	0.000	1.893	-0.424	1.076	-0.528	2.757	-0.383	0.927	77.13	0.1086	0.949
OA cohort	181	0.000	1.465	0.636	0.849	0.901	1.526	-0.298	1.127	58.17	0.6488	0.818
Pre-TKR†	697	0.000	1.094	0.155	1.387	1.112	1.671	-0.439	1.044	64.15	0.4010	0.828

*Item-trait Int: Item-trait interaction; **PSI: Person Separation Index; †Pre-total knee replacement

Table 5. Cubic model summary, coefficients and variable descriptives; Rasch-based person scores regressed on raw summed scores, final 7 items (n=1154).

	R	R^2	Adjusted R^2	Std. Error
	0.999	0.998	0.998	0.084

	Coefficients			
	B	Std. Error	t	p-value
(Raw sum score)	0.5698	0.003	207.056	0.000
$(\text{Raw sum score})^2$	-0.0337	0.000	207.056	0.000
$(\text{Raw sum score})^3$	0.0009	0.000	133.690	0.000
(Constant)	-4.2141	0.008	-560.477	0.000

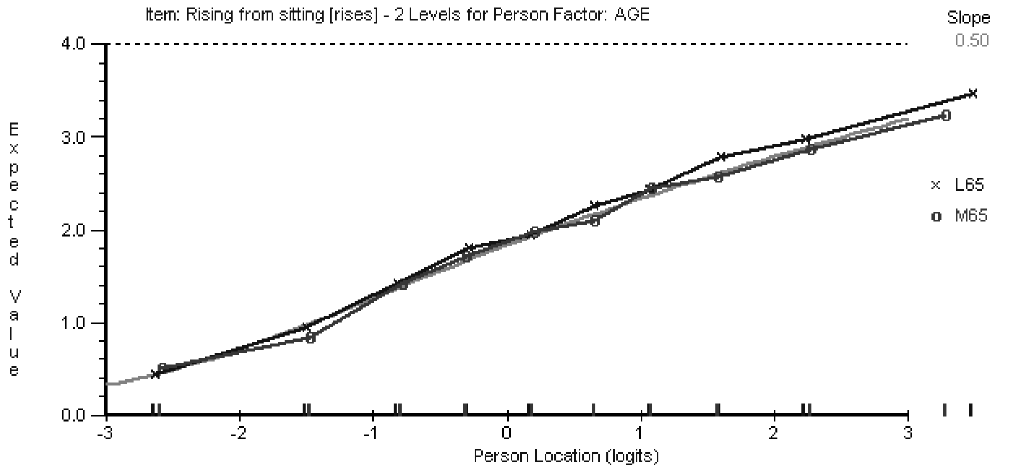
	Minimum	Maximum	Mean	Std. Dev.
Raw sum score	0.000	28.000	13.227	7.618
Observed Rasch person scores	-4.370	5.588	-0.397	1.979
Predicted person scores	-4.214	5.443	-0.397	1.977
Residuals	-0.156	0.223	0.000	0.084

Table 6. Conversion table – raw summed score from 7 items to predicted Rasch-based person score.

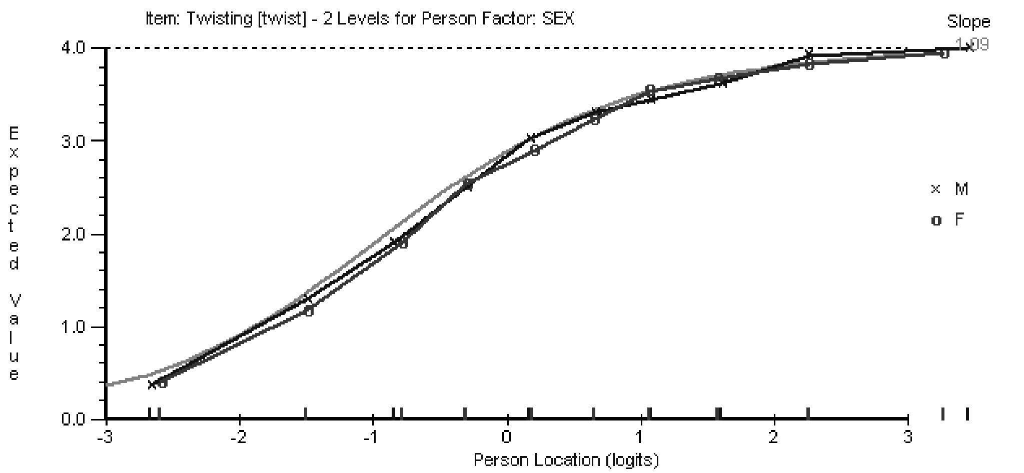
Raw summed score	Model predicted person score	Model predicted person score (0 to 100 scale)
0	-4.21	0.00
1	-3.68	5.56
2	-3.20	10.48
3	-2.78	14.82
4	-2.42	18.63
5	-2.09	21.97
6	-1.81	24.89
7	-1.56	27.46
8	-1.34	29.73
9	-1.15	31.76
10	-0.97	33.61
11	-0.80	35.32
12	-0.64	36.97
13	-0.49	38.60
14	-0.33	40.27
15	-0.15	42.04
16	0.03	43.97
17	0.24	46.11
18	0.47	48.52
19	0.74	51.25
20	1.04	54.38
21	1.38	57.94
22	1.77	62.00
23	2.22	66.61
24	2.72	71.84
25	3.29	77.73
26	3.93	84.35
27	4.65	91.76
28	5.44	100.00

Figure 1. Item characteristic curves for item *rise from sitting* by age (A), item *twisting/pivoting on injured knee* by gender (B) and item *rising from bed* by country (C).

(A)



(B)



(C)

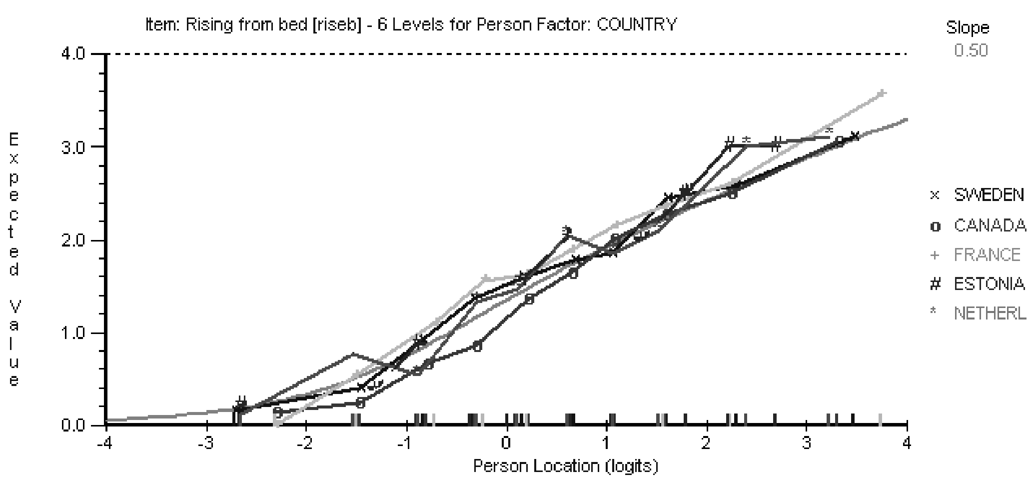


Figure 2. Person-Item threshold distribution, total sample – final model with 7 items.

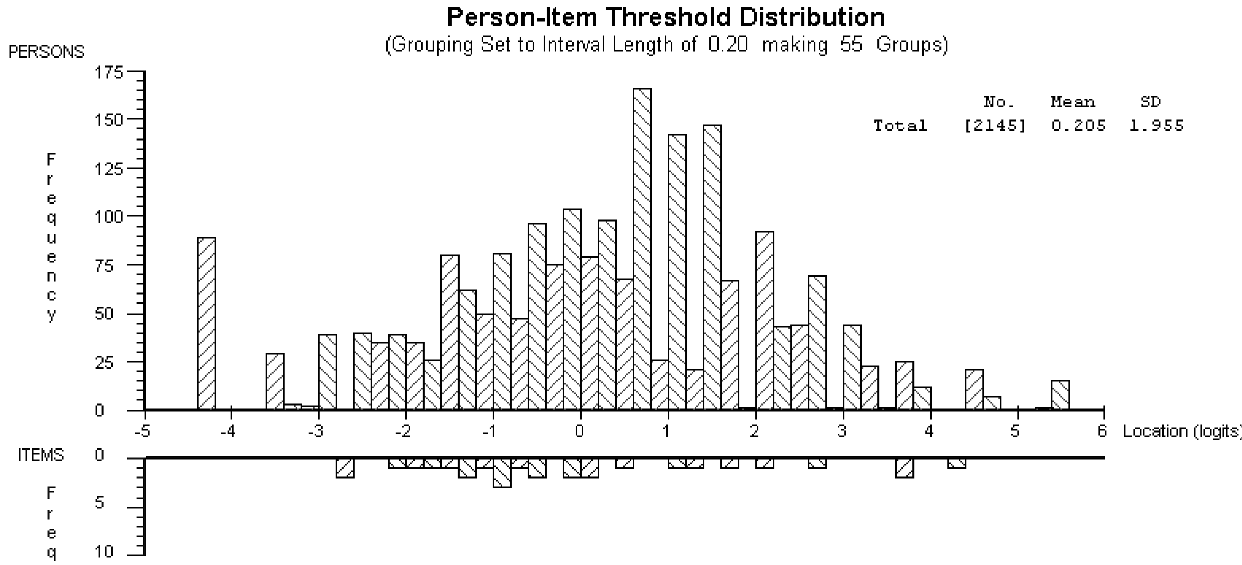


Figure 3. Distribution of Rasch-based person scores (scaled 0 to100) by sample type.

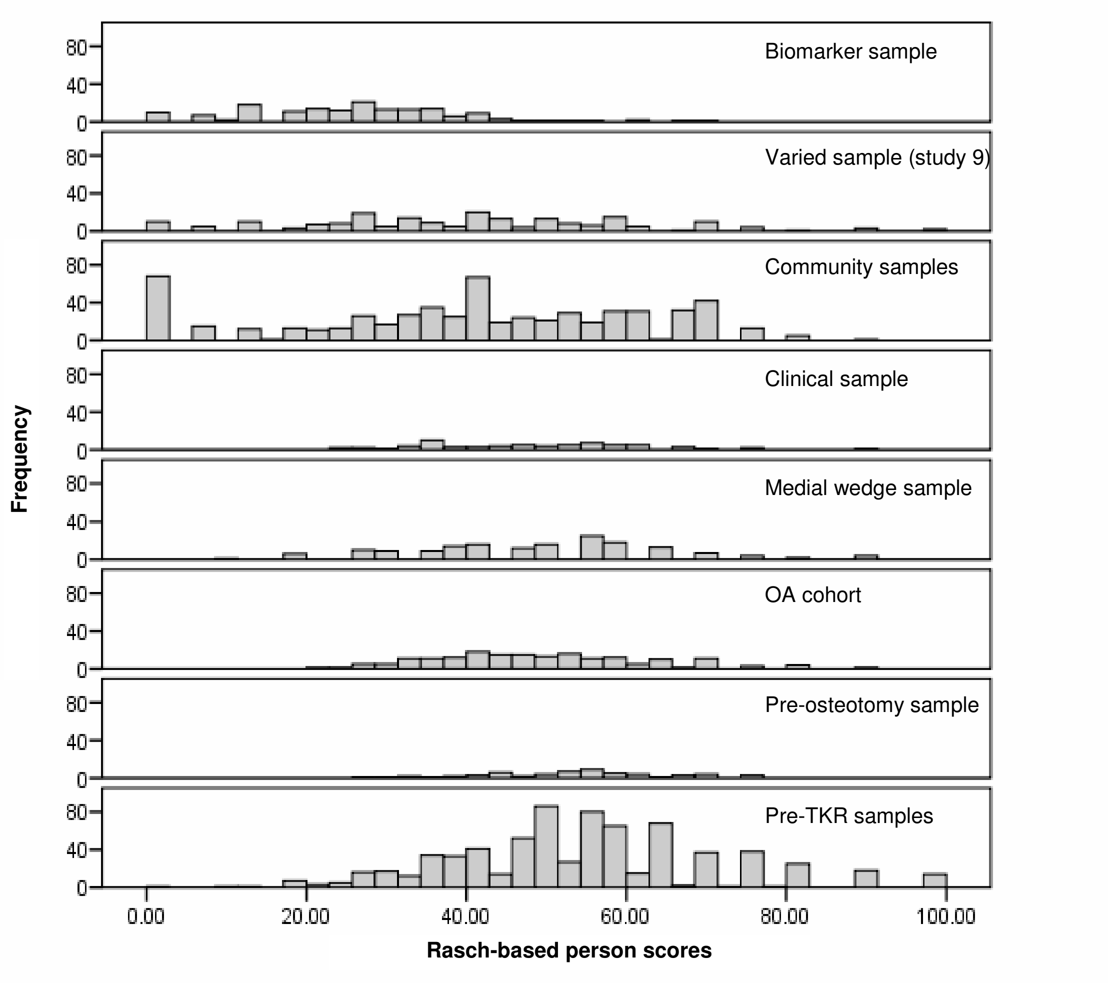


Figure 4. Scatter plot of raw summed scores and Rasch-based person scores for final 7 items and predicted curve from cubic model (n=1154).

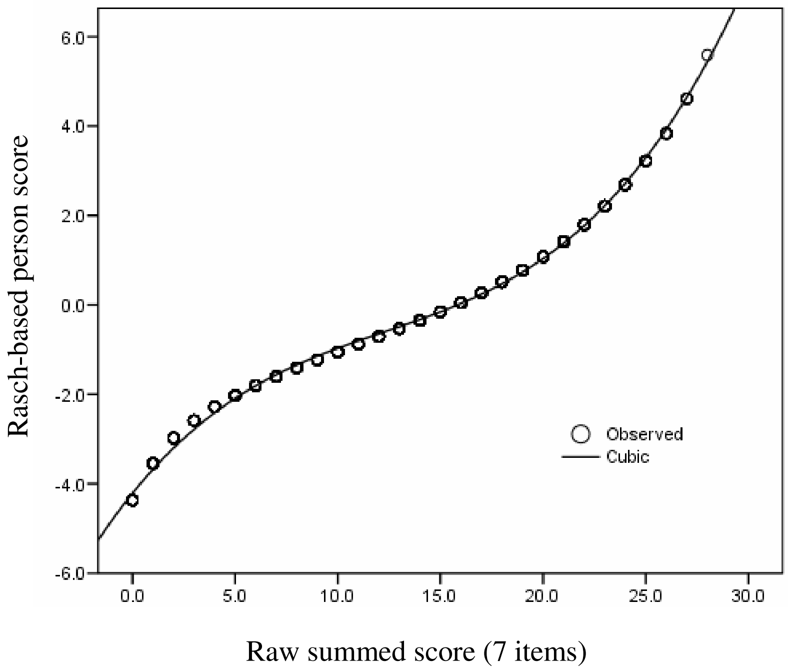
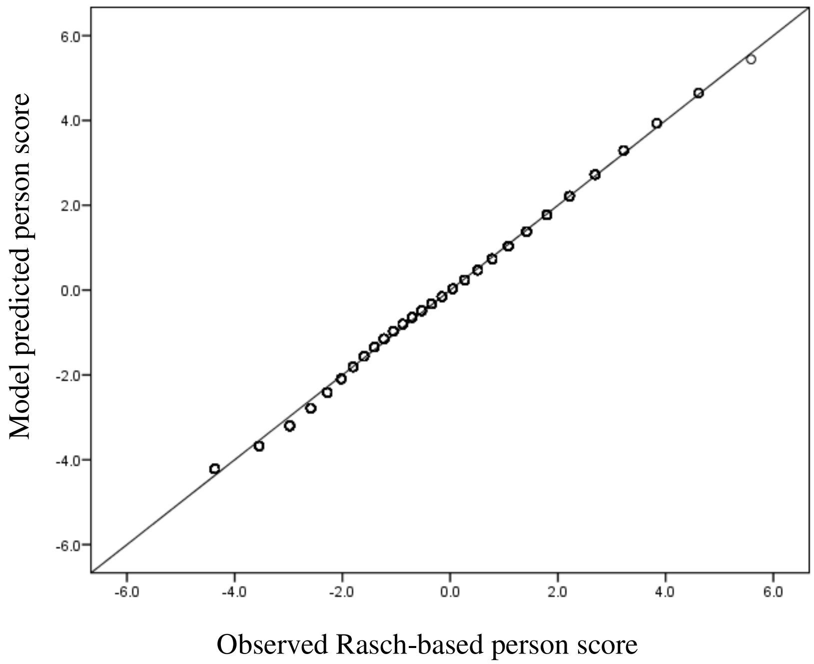


Figure 5. Scatter plot of observed and predicted Rasch-based person scores.



Appendix 1

KOOS-PHYSICAL FUNCTION SHORTFORM (KOOS-PS)

This survey asks for your view about your knee. This information will help us keep track of how well you are able to perform different activities. Answer every question by ticking the appropriate box, only one box for each question. If you are unsure about how to answer a question, please give the best answer you can so that you answer all the questions.

The following questions concern your level of function in performing usual daily activities and higher level activities. For each of the following activities, please indicate the degree of difficulty you have experienced in the **last week** due to your knee problem.

1. Rising from bed

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

2. Putting on sock/stockings

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

3. Rising from sitting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

4. Bending to the floor

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

5. Twisting/pivoting on your injured knee

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

6. Kneeling

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐

7. Squatting

None	Mild	Moderate	Severe	Extreme
☐	☐	☐	☐	☐