



# Rasch validation of the Danish version of the shoulder pain and disability index (SPADI) in patients with rotator cuff-related disorders

Karl Bang Christensen<sup>1</sup> · Kristian Thorborg<sup>3,4</sup> · Per Hölmich<sup>3</sup> · Mikkel Bek Clausen<sup>2,3,4</sup>

Accepted: 11 November 2018 / Published online: 19 November 2018  
© Springer Nature Switzerland AG 2018

## Abstract

**Purpose** The psychometric properties of the shoulder pain and disability index (SPADI) have been extensively evaluated using classical test theory, but very few studies have applied Rasch analysis. The purpose of this study was to validate the Danish version using Rasch analysis.

**Methods** Responses to the SPADI from 229 patients (48% female, mean age 54.5) were included in the Rasch analysis. Overall fit, individual item fit, local response dependence, dimensionality, targeting, reliability, and differential item functioning (DIF) were examined.

**Results** After iterative analyses, good fit to the Rasch model was observed, with acceptable targeting and uni-dimensionality. SPADI should be reported as two separate subscales: Pain and Functional Disability. The pain subscale initially demonstrated misfit due to local dependence and DIF, but a log linear Rasch model showed good fit to the Rasch model with acceptable targeting and uni-dimensionality. A six-item version of the disability subscale exhibited adequate fit in the Danish version. The same items were also found to fit the Rasch model in the English version.

**Conclusions** The measurement properties of the Danish SPADI are similar to those of the English version. SPADI should be reported as two separate subscales. For the pain subscale, DIF with respect to age was disclosed, but the impact was small. The eight-item disability subscale did not fit the Rasch model. A six-item version of the disability subscale exhibited adequate fit in the Danish version. The same items were also found to fit the Rasch model in the English version.

**Keywords** Rasch model · Shoulder pain and disability index · Psychometrics

## Introduction

A systematic review of outcomes measures for shoulder disorders identified pain, range of motion, and function as the most commonly assessed domains [2], but pain and physical function have been identified as core outcome domains [22]. The shoulder pain and disability index (SPADI) consists of a five-item pain subscale and an eight-item disability subscale [25]. It originally used visual analogue scales, but items are now scored using an ordinal rating scale from zero to ten, the latter indicating the highest level of pain/difficulty. Different scoring strategies have been used, with some authors reporting the mean of all 13 items and others calculating an average for the five pain and eight disability items separately. The latter method can also be used for reporting the mean of the two sub scores, thus giving equal weight to the two domains. The SPADI is recommended for clinical practice and research [10, 26]. However, these recommendations are

✉ Karl Bang Christensen  
kach@sund.ku.dk

<sup>1</sup> Section of Biostatistics, Department of Public Health, University of Copenhagen, Copenhagen, Denmark

<sup>2</sup> Department of Physiotherapy and Occupational Therapy, Faculty of Health and Technology, Metropolitan University College, Copenhagen, Denmark

<sup>3</sup> Sports Orthopedic Research Center-Copenhagen (SORC-C), Department of Orthopedic Surgery, Copenhagen University Hospital, Amager-Hvidovre, Copenhagen, Denmark

<sup>4</sup> Physical Medicine and Rehabilitation Research-Copenhagen (PMR-C), Amager-Hvidovre Hospital, Copenhagen University Hospital, Copenhagen, Denmark

based on studies of the validity, reliability, and responsiveness of SPADI using classical test theory (CTT).

Rasch analysis [24, 12, 6] is a modern approach based on item response theory (IRT) [28] that is used for the development and testing of patient-reported outcome (PRO) instruments. All IRT models possess certain desirable properties, and while some consider the Rasch model to be an overly simplified statistical model, its simplicity is exactly what gives it a special status among IRT models, representing a measurement model where the sum score of item responses contains all information about the underlying latent variable that the scale is intended to measure. A recent validation study of the SPADI using Rasch analysis identified strengths and limitations not previously observed using CTT methods [15]. This study concluded that the SPADI should be treated as two separate subscales and that, while the pain subscale fits the Rasch model well, the disability subscale does not fit the Rasch model and that clinicians should exercise caution when interpreting score changes on the disability subscale and attempt to compare their scores to age- and sex-stratified data. The Danish version of SPADI has been cross-culturally adapted and validated using CCT [4]. The purpose of this study is to validate the two subscales of the Danish translation of the SPADI using Rasch analysis, evaluate differential item functioning (DIF), and study how well the two subscales are targeted to patients with rotator cuff-related disorders.

## Methods

The data for this validation study came from a consecutive cohort of patients with rotator cuff-related disorders (subacromial impingement with or without rotator cuff tear) [8]. The cohort included patients with a shoulder problem referred to orthopedic specialist assessment at a public secondary care outpatient clinic during a 3-month period (March to June 2014). Eligible patients received an information letter explaining that an extended assessment was offered immediately prior to the orthopedic specialist examination. As part of the additional assessment, the information letter contained the SPADI questionnaire and instructions to fill it in and bring it on the day of examination. Orthopedic specialists were blinded to results of the assessments, and patients were diagnosed according to the clinical judgement of the orthopedic specialist performing the examination. Study methods and results have been described elsewhere [8, 9, 29].

Overall fit to the Rasch model was assessed using the Andersen conditional likelihood ratio test [1] and individual item fit was evaluated by comparing observed and expected item-restscore associations [18]. We also evaluated item fit graphically by dividing the sample into five

score groups (often denoted ‘class intervals’ in the Rasch literature) and, for each item, plotting the item mean scores in each interval, and comparing these to 95% confidence regions for the model expectations. To test the assumption of uni-dimensionality, we compared, observed, and expected subscore correlations [14]. Differential item functioning (DIF) [13] occurs when responses systematically differ by some other factor or variable like age or gender. Local response dependence occurs when items are almost identical (redundancy) or when they share features, e.g., wording response format or are associated with some other underlying trait. We evaluate local response dependence and test for DIF with respect to gender and age using log linear Rasch model tests [16] and item screening [19]. The ability of the subscales to discriminate between respondents is evaluated using Cronbach’s coefficient alpha and the person separation index (PSI) [23]. In all analyses, we adjust *p* values using the Benjamini–Hochberg [15] procedure to control the false discovery rate.

Disordered thresholds occur when participants cannot consistently discriminate between the available response options. Jerosch-Herold et al. examined category probability curves and proposed a re-scoring (00112233445). Our sample was deemed to be too small to estimate threshold parameters with sufficient precision and we used this re-scoring for all items.

Analyses were done using DIGRAM [17, 20] diagram and version 9.4 of the SAS software package. Person-item location maps were created using a SAS macro [6].

## Results

The validation sample consisted of 229 patients (48% female) with a mean age of 54.5 (*SD* = 14.2). Information about employment status was available for 221, of whom 115 were currently working, 106 were not working (sick leave, retired, unemployed). A total of 21 patients reported being on part time (*n* = 5) or sick leave/unemployed (*n* = 16) due to the shoulder disorder. The mean SPADI original score was 55 (*SD* = 22) and the dominant side was affected in 56.5%, (122 of 216). Average pain last week (on NPRS, 0–10) was 5 (*SD* = 2) (*n* = 210), and duration of shoulder problem (*n* = 226): 0–1 months: 3 (1%), 1–3 months: 37 (16%), 3–6 months: 50 (22%), 6 or more: 136 (60%).

### Rasch analysis of SPADI pain subscale

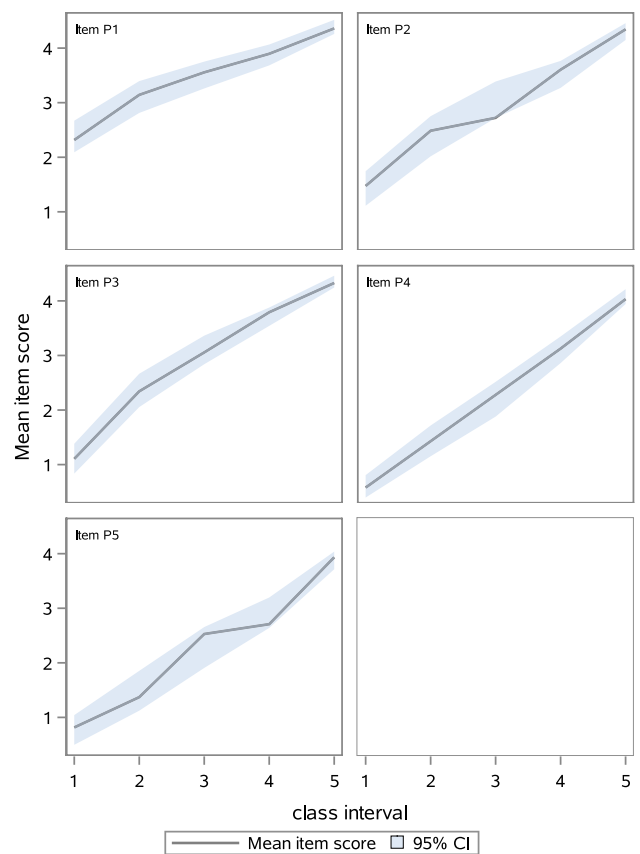
Initial analysis of the pain subscale revealed poor fit to the Rasch model (Andersen  $\chi^2 = 58.7$ , *df*=23, *p* = 0.0001). The item screening indicated local response dependence for two item pairs: P1 (‘at its worst’) and P2 (‘lying on affected side’) (*p* = 0.0001) and P3 (‘reaching for object on a high shelf’)

and P4 (‘touching the back of your neck’) ( $p < 0.0001$ ) and DIF by age for P1 ‘pain at worst’ ( $p = 0.0131$ ). Adding these yielded a log linear Rasch model with excellent overall (Andersen  $\chi^2 = 48.4$ ,  $df=56$ ,  $p = 0.7540$ ) fit. Regarding individual item fit, the item fit statistics (Table 1) and the plots of observed and expected item mean scores (Fig. 1) indicated that the data fit the log linear Rasch model.

Reliability was high with a Cronbach coefficient alpha of 0.86 and a person separation index (PSI) of 0.84 and the person-item location map (Fig. 2, left panel) shows that the subscale works well at different levels of the construct.

**Rasch analysis of SPADI disability subscale**

Initial analysis of the 8-item disability subscale revealed misfit to the Rasch model (Andersen  $\chi^2 = 61.9$ ,  $df=39$ ,  $p = 0.0112$ ) and evidence substantial misfit for item D7 ‘carry a heavy object’ (observed item-restscore association 0.52, expected item-restscore association 0.66,  $p < 0.0001$ ). Deleting the items D3 ‘putting on undershirt or jumper’ and D7 ‘carry heavy object’ from the subscale yielded a model with excellent overall fit to the Rasch model (Andersen  $\chi^2 = 36.3$ ,  $df=29$ ,  $p = 0.1647$ ) and with excellent item fit (Table 1; observed and item mean scores corresponded to Rasch model predictions, Fig. 3). There was no evidence of DIF with respect to age and gender (results not shown), but evidence of local response dependence for the items D4 ‘putting on a shirt that buttons at front’ and D5 ‘putting on trousers’ ( $p < 0.0001$ ). Adding this yielded a log linear Rasch model with excellent overall (Andersen  $\chi^2 = 51.4$ ,  $df=45$ ,  $p = 0.2366$ ) and individual item fit (Table 1, Fig. 3). Reliability was high with a Cronbach coefficient alpha of



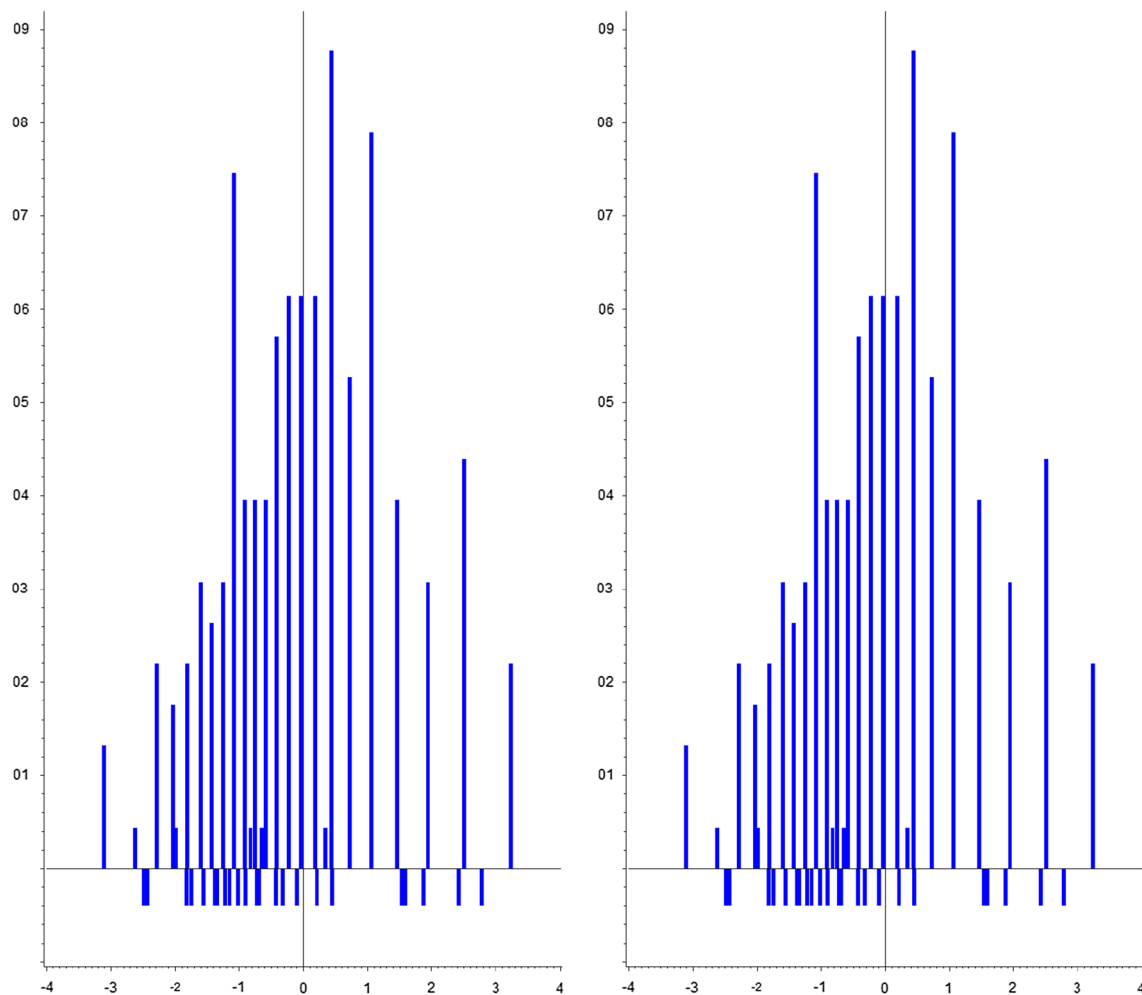
**Fig. 1** Item fit plot for the Pain subscale. Item mean scores (solid lines) and 95% confidence regions for expected mean scores (shaded areas)

**Table 1** Item fit statistics

Subscale	Item	Difficulty	Item-restscore association		P
			Observed	Expected	
P	P1 at its worst	- 1.29	0.60	0.60	0.9339
P	P2 lying on affected side	- 0.49	0.62	0.57	0.2924
P	P3 reaching for object on a high shelf	- 0.55	0.74	0.74	0.8937
P	P4 touching the back of your neck	0.13	0.70	0.71	0.8128
P	P5 pushing with involved arm	0.13	0.61	0.58	0.5674
D	D1 Washing your hair?	- 0.13	0.71	0.65	0.1266
D	D2 Washing your back?	- 1.13	0.67	0.65	0.5297
D	D3 Putting on undershirt or jumper <sup>a</sup>				
D	D4 Putting on a shirt that buttons down the front?	0.49	0.65	0.70	0.1959
D	D5 Putting on your pants?	0.69	0.69	0.70	0.7247
D	D6 Placing an object on a high shelf?	- 1.13	0.63	0.65	0.6009
D	D7 Carry heavy object <sup>a</sup>				
D	D8 Removing something from your back pocket?	- 0.24	0.69	0.66	0.2798

Observed and expected item-restscore association in the revised scales

<sup>a</sup>Not included



**Fig. 2** Person-item location maps. Distribution of person location estimates above the x-axis and item threshold distribution below the x-axis for the pain (left panel) and disability (right panel) subscales

0.89 and a person separation index (PSI) of 0.87 and the person-item location map (Fig. 2, right panel) shows that the subscale works well at different levels of the construct.

### Dimensionality

Testing the assumption of uni-dimensionality by comparing observed and expected subscale correlations [14] showed the SPADI to be two-dimensional: expected subscale correlation 0.698 (s.e.=0.0262), observed subscale correlation 0.620,  $P = 0.0029$ .

### Impact of DIF

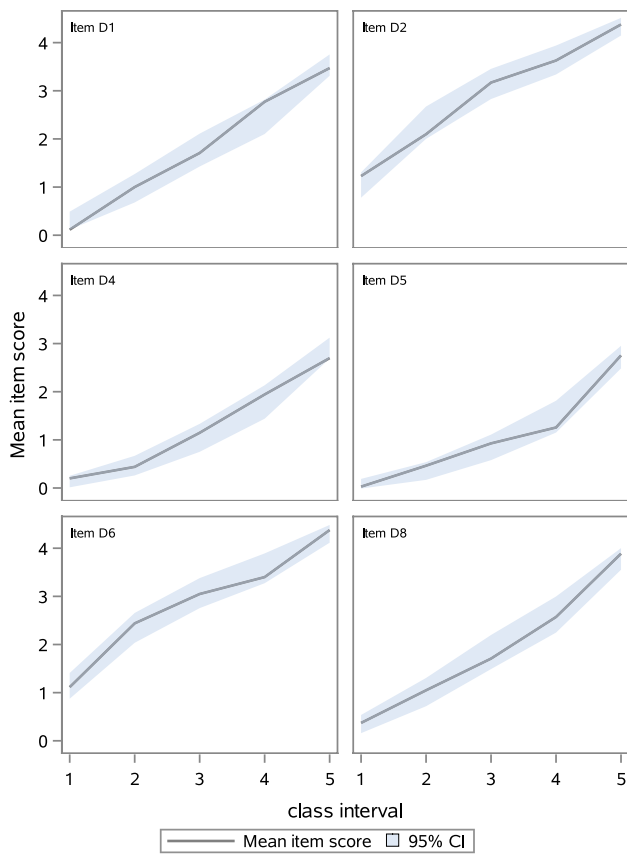
Scores derived from the SPADI should be interpreted with caution. Firstly, it should be treated as a five-item pain subscale and an eight six disability subscale. We disclosed evidence of DIF by age for P1 ('pain at worst'). In order to

evaluate the impact of this, we computed equated scores and found the difference to be smaller than 0.61 (Table 2).

### Discussion

We validated the Danish version of the SPADI and found results very similar to those Jerosch-Herold et al. found for the English version. The Danish version of the SPADI should be reported as two separate subscales. The pain subscale has some DIF, but the impact appears to be small. The disability subscale cannot be used in its current form, but a six-item version was found to fit the Rasch model adequately. Rasch Model analysis of the SPADI has identified some strengths and limitations not previously observed using CTT methods.

For the pain subscale, Jerosch-Herold et al. found DIF by age for P1 'pain at worst' and by gender for P5 'pain when pushing with involved arm,' but no evidence of local



**Fig. 3** Fit plot for the disability subscale. Item mean scores (solid lines) and 95% confidence regions for expected mean scores (shaded areas)

response dependence. We replicated the finding regarding DIF by age for P1 only. Regarding local response dependence we found evidence for two item pairs. Computation of equated scores indicated the difference to be relatively small.

For the disability subscale, we replicated the finding of Jerosch-Herold et al. that the six-item version resulting from removing the items D3 ‘putting on undershirt or jumper’ and D7 ‘carrying heavy object’ showed reasonable fit to the Rasch model, but where Jerosch-Herold et al. found DIF for D1 and D4 by gender and for D5 by age, our analysis did not disclose evidence of DIF. Again we disclosed significant evidence for local response dependence (for the item pair D4 ‘putting on a shirt that buttons at front’ and D5 ‘putting on trousers’) where Jerosch-Herold et al. did not.

Regarding the reason for misfit of the item D3 (‘Putting on undershirt or jumper’), we speculate that because in the Danish translation the two garments are quite different the item is double-barreled, and regarding misfit of D7 (‘Carry heavy object’) it is likely that because respondents are not equally strong that they do not rate ‘Carrying a heavy object of 10 pounds (4.5 kg)’ consistently. A DIF item “behaves differently for various subgroups after controlling for the overall differences between subgroups on the construct being measured” [11]. Regarding the DIF for the item P1 (‘pain at its worst’), where respondents over 60 consistently score slightly lower, we speculate that their reference for pain is shifted.

Jerosch-Herold et al. found more evidence of DIF than we did. We speculate that the difference in sample size could be the reason for this. Regarding local response dependence we found more evidence than Jerosch-Herold et al. and speculate that the reason could be that they did not follow the recommendation by Marais [21] that LD should be considered relative to the average residual correlation, cf. Christensen et al. [8].

Beyond the smaller sample size, our study population differed from the population of Jerosch-Herold et al. in other ways. Most importantly, Jerosch-Herold et al. [15] included all patients treated for shoulder pain irrespective of shoulder disorder, where this study only includes patients with rotator cuff-related disorders. Furthermore, the mean total of the original SPADI score of 55 (SD 22) in our sample, is somewhat higher than the 48 (SD 22) in the sample included by Jerosch-Herold et al. [15].

**Clinical implications**

In conclusion, score derived from the SPADI should be interpreted with caution. Firstly, it should be treated as a five-item pain subscale and a six-item disability subscale. Reporting of scores can still be done using a linear transformation to a zero to 100 scale (for the validation sample studied here this would yield SPADI pain 64 (SD = 22) and SPADI disability 44 (SD = 25)). Secondly, clinicians should attempt to compare their scores to age-stratified data, even though the impact of differential item function seems small.

**Table 2** Equated scores showing the impact of DIF

Age ≤ 60	1	2	3	4	5	6	7	8	9	10	11	12
Age > 60	1.38	2.43	3.57	4.60	5.61	6.60	7.58	8.54	9.50	10.46	11.42	12.39
Age > 60	13	14	15	16	17	18	19	20	21	22	23	24
Age ≤ 60	13.38	14.37	15.37	16.37	17.36	18.36	19.37	20.40	21.42	22.39	23.30	24.16

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethical approval** This paper is based on a secondary analysis of data. The trial protocol, the informed consent forms, and other requested documents have been reviewed and approved by the Capitol Regional Ethics Committee in Denmark (H-16016763) with respect to the scientific content and the compliance to the applicable health science regulations. All procedures performed in the study involving human participants were in accordance with the ethical standards of the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

## References

- Andersen, E. B. (1973). A goodness of fit test for the rasch model. *Psychometrika*, 38(1), 123–140.
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1), 289–300.
- Buchbinder, R., Page, M. J., Huang, H., Verhagen, A. P., Beaton, D., Kopkow, C., et al. (2017). A preliminary core domain set for clinical trials of shoulder disorders: A report from the OMERACT 2016 shoulder core outcome set special interest group. *The Journal of Rheumatology*, 44(12), 1880–1883.
- Christensen, K. B., Kreiner, S., & Mesbah, M. (Eds.). (2013). *Rasch models in health*. Hoboken: Wiley.
- Christensen, K. B., Makransky, G., & Horton, M. (2017). Critical values for Yen's Q 3: Identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*, 41(3), 178–194.
- Christensen, K. B., & Olsbjerg, M. (2013). Marginal maximum likelihood estimation in polytomous Rasch models using SAS. *Pub Inst Stat Univ*, 57, 69–84.
- Christiansen, D. H., Andersen, J. H., & Haahr, J. P. (2013). Cross-cultural adaption and measurement properties of the Danish version of the shoulder pain and disability index. *Clinical Rehabilitation*, 27(4), 355–360.
- Clausen, M., Witten, A., Holm, K., Christensen, K., Attrup, M., Hölmich, P., et al. (2017). Glenohumeral and scapulothoracic strength impairments exists in patients with subacromial impingement, but these are not reflected in the shoulder pain and disability index. *BMC Musculoskeletal Disorders*, 18(1), 302.
- Clausen, M. B., Merrild, M. B., Witten, A., Christensen, K. B., Zebis, M. K., Hölmich, P., et al. (2018). Conservative treatment for patients with subacromial impingement: Changes in clinical core outcomes and their relation to specific rehabilitation parameters. *PeerJ*, 6, e4400.
- Dawson, J., Harris, K. K., Doll, H., Fitzpatrick, R., & Carr, A. (2016). A comparison of the Oxford shoulder score and shoulder pain and disability index: Factor structure in the context of a large randomized controlled trial. *Patient Related Outcome Measures*, 7, 195–203.
- Edelen, M. O., Thissen, D., Teresi, J. A., Kleinman, M., & Ocepek-Welikson, K. (2006). Identification of differential item functioning using item response theory and the likelihood-based model comparison approach: Application to the mini-mental state examination. *Medical Care*, 44(11), S134–S142.
- Fischer, G. H., & Molenaar, I. W. (Eds.). (1995). *Rasch models*. New York: Springer.
- Holland, P. W., & Wainer, H. (1993). *Differential Item Functioning*. Hillsdale: Erlbaum.
- Horton, M., Marais, I., & Christensen, K. B. (2013). Dimensionality. *Rasch models in health* (pp. 137–158). Hoboken: Wiley.
- Jerosch-Herold, C., Chester, R., Shepstone, L., Vincent, J. I., & MacDermid, J. C. (2017). An evaluation of the structural validity of the shoulder pain and disability index (SPADI) using the Rasch model. *Quality of Life Research*, 27(2), 389–400.
- Kelderman, H. (1984). Loglinear Rasch model tests. *Psychometrika*, 49(2), 223–245.
- Kreiner, S. (2003). Introduction to DIGRAM. Research Report 10, Department of Statistics, University of Copenhagen.
- Kreiner, S. (2011). A note on item-restscore association in Rasch models. *Applied Psychological Measurement*, 35(7), 557–561.
- Kreiner, S., & Christensen, K. B. (2011). Item screening in graphical loglinear Rasch models. *Psychometrika*, 76(2), 228–256.
- Kreiner, S., & Nielsen, T. (2013). *Item analysis in DIGRAM 3.04: Part I: Guided tours*. Department of Biostatistics, University of Copenhagen.
- Marais, I. (2013). *Local dependence*. In *Rasch models in health* (pp. 111–130). Hoboken: Wiley.
- Page, M. J., Huang, H., Verhagen, A. P., Gagnier, J. J., & Buchbinder, R. (2018). Outcome reporting in randomized trials for shoulder disorders: Literature review to inform the development of a core outcome set. *Arthritis Care & Research*, 70(2), 252–259.
- Pallant, J. F., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology*, 46(1), 1–18.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish National Institute for Educational Research.
- Roach, K. E., Budiman-Mak, E., Songsiridej, N., & Lertratanakul, Y. (1991). Development of a shoulder pain and disability index. *Arthritis Care and Research: The Official Journal of the Arthritis Health Professions Association*, 4(4), 143–149.
- Roy, J.-S., MacDermid, J. C., & Woodhouse, L. J. (2009). Measuring shoulder function: A systematic review of four questionnaires. *Arthritis and Rheumatism*, 61(5), 623–632.
- SAS Institute (2013). *SAS 9.4 Language reference: Concepts*. Cary: SAS Institute Inc.
- Van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Witten, A., Clausen, M. B., Thorborg, K., Attrup, M. L., & Hölmich, P. (2018). Patients who are candidates for subacromial decompression have more pronounced range of motion deficits, but do not differ in self-reported shoulder function, strength or pain compared to non-candidates. *Knee Surgery, Sports Traumatology, Arthroscopy: Official Journal of the ESSKA*, 26(8), 2505–2511.