International Journal of Nursing Studies 87 (2018) 49–59

Contents lists available at ScienceDirect

International Journal of Nursing Studies

journal homepage: www.elsevier.com/locate/ijns

Scaling properties of pain intensity ratings in paediatric populations using the Faces Pain Scale-revised: Secondary analyses of published data based on the item response theory

Alexander Avian[a,b,c,*,1], Brigitte Messerer[d], Andreas Frey[c,e], Winfried Meissner[b], Annelie Weinberg[f], William Ravekes[g], Andrea Berghold[a]

[a] Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2, 8036 Graz, Austria
[b] Department of Anesthesiology and Intensive Care, Jena University Hospital, Jena, Lobeda, Germany
[c] Department of Research Methods in Education, Friedrich Schiller University Jena, Am Planetarium 4, 07743 Jena, Germany
[d] Division of Anesthesiology for Cardiovascular Surgery and Intensive Care Medicine, Medical University of Graz, Auenbruggerplatz 29, 8036 Graz, Austria
[e] Centre for Educational Measurement (CEMO) at the University of Oslo, Postboks 1161 Blindern, 0318 Oslo, Norway
[f] Department of Orthopedics and Orthopedic Surgery, Medical University of Graz, Auenbruggerplatz 5, 8036 Graz, Austria
[g] Division of Pediatric Cardiology, Johns Hopkins University School of Medicine, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Pain measurement
Pediatrics
Scale propertiers
Surveys and questionnaires

ABSTRACT

Background: The Faces Pain Scale-revised (FPS-r) has been developed as an interval scale. For other pain measurement instruments, several studies found evidence for and against an interval level of measurement.
Objectives: The primary aim of the current study was to evaluate the scale properties of the FPS-r using an item response theory approach.
Design: Secondary analysis of published data.
Setting: Three studies; Study 1 and study 2: One university hospital; Study 3: international pain registry.
Participants: Study 1: n = 246, female: 41%, age: 11–18 years, 3 pain items; Study 2: n = 240, female: 43%, age: 11–18 years, 9 pain items; Study 3: n = 2266, female: 41%, age: 4–18 years, 3 pain items.
Methods: The rating scale model (interval scale), the graded response model (no interval scale, ordered response categories) and the partial credit model (no interval scale) were used to scale the data.
Results: In all three studies, the rating scale model was outperformed by the graded response model or the partial credit model in terms of model fit. Overlapping response categories were found in items associated with less pain. Response category widths were wider for categories associated with low pain intensity and smaller for categories associated with high pain intensities. Smallest response categories were 1%–67% smaller compared to the widest response category of the same item.
Conclusion: According to these findings, the interval scale properties of the FPS-r may be questioned. Item response theory methods may help to solve the problem of missing linearity in pain intensity ratings using FPS-r.

What is already known about the topic?

- Faces Pain Scale-revised (FPS-r) has been developed as a linear interval scale.
- The interval scale properties of the FPS-r are questioned.

What this paper adds

- Responses to the FPS-r cannot be assumed interval scaled.
- When reporting responses to the FPS-r, nonparametric (e.g. median, interquartile range) parameters or the number of patients above/

* Corresponding author at: Institute for Medical Informatics, Statistics and Documentation, Medical University of Graz, Auenbruggerplatz 2, 8036 Graz, Austria.
E-mail addresses: alexander.avian@medunigraz.at, Alexander.avian@uni-jena.de (A. Avian), brigitte.messerer@medunigraz.at (B. Messerer), andreas.frey@uni-jena.de (A. Frey), Winfried.meissner@med.uni-jena.de (W. Meissner), Annelie.Weinberg@t-online.de (A. Weinberg), wravekes@jhmi.edu (W. Ravekes), andrea.berghold@medunigraz.at (A. Berghold).
[1] www.medunigraz.at.

https://doi.org/10.1016/j.ijnurstu.2018.07.009
Received 1 December 2017; Received in revised form 10 July 2018; Accepted 13 July 2018
0020-7489/

below a certain pain level should be used.
- Parametric parameters (e.g. mean, standard deviation) for reporting FPS-r responses should not be used.

1. Introduction

An adequate assessment tool is crucial for effective pain management in hospitals. Pain should be evaluated regularly for monitoring reasons, to follow the course of pain intensity in patients and to evaluate the effectiveness of pain therapy. Furthermore, pain is assessed for research purposes in experimental and clinical settings. Several pain assessment instruments have been established that differ in the number of items used and the target population. For children older than 4 years, the Faces Pain Scale-revised (FPS-r) is often recommended (von Baeyer, 2009).

The FPS-r has been developed as a linear interval scale. Two faces have been labeled “no pain” and “very much pain” and further four faces have been chosen in between these two faces to represent equal intervals between each of these six faces. Within the development of the FPS-r 101 different faces were presented on the computer to participants. Participants had to choose four faces that corresponded to predefined pain intensity levels on a scale with fixed endpoints (no pain to very much pain) (Hicks et al., 2001). Consequently, in many publications FPS-r values are treated as being located on an interval scale and therefore statistics relying on that scale level such as mean scores are calculated and parametric analyses are applied (e.g. Birnie et al., 2016; Brown et al., 2016; Ferreira-Valente et al., 2011; Park et al., 2015; Sánchez-Rodríguez et al., 2012). Some publications are using nonparametric analyses due to distributional concerns (e.g. Crevatin et al., 2016; McLaughlin et al., 2016). Only in a few publications, the interval scale properties of FPS-r are questioned and therefore nonparametric analyses applied (de Azevedo et al., 2014; Hirunwiwatkul et al., 2009). Some publications even describe the FPS-r as a measurement on an ordinal scale but report parametric statistics (Ho et al., 2015). In several studies analyses of other pain measurement tools found evidence for and against an interval scale of measurements (e.g. Oliveira et al., 2014; Shields et al., 2003a,b). According to von Baeyer (2009), the interval scale property of a pain measurement tool has to be questioned even if it was explicitly designed to measure on a linear interval scale like the FPS-r. This is especially true for pain assessment in younger children (von Baeyer, 2009).

The primary aim of the current study was to evaluate the scale properties of the FPS-r and therefore assess whether the assumption that they have the properties of an interval scale holds. In three different samples responses were analyzed (e.g. response category widths, overlapping of response categories) using three different item response theory models (c.f. Box 1) for polytomous responses (Ostini and Nering, 2006). Different pain items and different age groups of patients were analyzed to examine whether the scale properties of the FPS-r are dependent upon the sets of items used and the patient’s age.

2. Methods

2.1. Sample and design

This is a secondary analysis based on three previously reported studies by our group. The rationale and design for these three studies have been reported in detail elsewhere (Avian et al., 2016, 2017). Briefly, Study 1 (Avian et al., 2016) and study 2 (Avian et al., 2017) were prospective studies (Study 1:between July 2010 and March 2012; study 2: between October 2013 and May 2014) that included patients between eleven and 18 years who underwent surgery at the Department of Pediatric and Adolescent Surgery, Medical University of Graz (Austria). Patients had to be able to speak German. Intensive care patients and patients with cognitive impairment were not included. Patients were asked by an independent researcher not involved in patient care to rate their pain-intensity.

In study 1, patients rated their pain at rest, during movement and their worst pain. This study aimed to evaluate possible order effects in children and adolescents and the possible influence of sex on order effects. Therefore, three pain items (pain at rest, during movement and their worst pain) were presented in six different orders.

Study 2 varied from the first study in that patients rated their worst pain after surgery and the pain while carrying out eight different activities. Six of these eight activities were included in this manuscript: (1) eating, (2) drinking, (3) turning over in the bed, (4) getting up from bed, (5) coughing, and (6) lying in bed. Two activities were excluded to get a unidimensional model. Study 2 aimed to analyze inconsistencies and the test-retest reliability in worst pain ratings in children and adolescents. Inconsistencies were defined as lower worst pain ratings compared to pain intensity ratings for activity pain items. In study 2, pain assessments were performed twice (t1, t2), separated by one to two hours [median time between assessments: 75 min, interquartile range (IQR): 70–85; Range: 60–120 min]. In our current analysis we only included the first of the two pain assessment ratings collected in study 2.

In study 3 (Avian et al., 2017) data from an international pediatric acute pain registry (Quality Improvement in Postoperative Pain Treatment in children; QUIPSi) were included. Within the QUIPSI registry, patient data from German, Austrian and Swiss pediatric patients are collected (http://www.quips-projekt.de/). This registry includes (1) outcome measurements (pain intensity measurements, pain-related interference e.g. pain when coughing, side effects e.g. vomiting), and (2) relevant process parameters (e.g. kind of surgery, medication). Children at the age of 4 to 18 years can be included in this registry. These children were admitted for pediatric surgery in participating hospitals. These hospitals were collecting these patient data for quality improvement reasons. Within the QUIPSI registry, it is possible to compare the hospital’s outcomes with all other hospital outcomes on the hospital level or e.g. on a surgery level. Of the 5970 included patients, only those answering the questionnaire alone without any help (n = 2266, 46% female, age: 13.3 ± 2.7 years) were analyzed.

For all pain assessments, the FPS-r was used (Hicks et al., 2001). The FPS-r shows acceptable reliability in children rating their actual pain (r = .77) (Tsze et al., 2013) and moderate to high correlations with other pain assessment tools (r = .66–.87) (Perrott et al., 2004; Tsze et al., 2013).

2.2. Ethical considerations

All three studies comply with all institutional guidelines related to patient confidentiality and research ethics including institutional review board approval (Study 1 and 2: Medical University Graz Ethics Committee, IRB00002556; Study 3: University Ethics committee of Jena University Hospital, Thuringia, Germany, IRB00004153).

2.3. Data analysis

The data sets of the three studies were analyzed separately. The R-package mirt (version 1.25) (Chalmers, 2012, 2017) was used for data analysis. The software R (Version 3.4.1, 2017-06-30; R Foundation for statistical Computing) was used for all analyses. Missing data were not imputed. Response categories were collapsed if less than 10 responses within a category were observed. To analyze the scale properties of the polytomous pain ratings, three different item response theory models were compared: the rating scale model (Andrich, 1978), the graded response model (Samejima, 1969) and the partial credit model (Masters, 1982).

While the rating scale model assumes an interval scale, the graded response model only assumes ordinal scale and the partial credit model