

The use and quality of reporting of Rasch analysis in nursing research: A methodological scoping review



Minna Stolt^{a,*}, Anders Kottorp^b, Riitta Suhonen^{a,c}

^a Department of Nursing Science, University of Turku, Finland

^b Faculty of Health and Society, Malmö University, Malmö, Sweden

^c Turku University Hospital and City of Turku, Welfare Division, Turku, Finland

ARTICLE INFO

Article history:

Received 14 December 2021

Received in revised form 28 March 2022

Accepted 31 March 2022

Keywords:

Nursing

Psychometrics

Rasch analysis

Reproducibility of results

Scoping review

ABSTRACT

Background: Rasch analysis is widely used in the life sciences. Rasch analysis is a mathematical and probabilistic model based on the assumption that the probability of passing a single item is governed by a person's ability and the difficulty of the item. However, its use in nursing science remains unclear.

Aim: To (i) describe the use of Rasch analysis in nursing research and (ii) determine the quality of reporting in nursing studies using Rasch models.

Methods: A methodological scoping review of literature was conducted. The systematic electronic literature search was initially conducted on 1 February 2020 and updated on 16 April 2021 from PubMed/Medline and CINAHL databases. The search was limited to covering the timeframe from the earliest literature available until 31 December 2020. The search terms used were Rasch, IRT, item response theory, and nursing. The search was limited to the English language and title/abstract level. The analysis included quantification and content analysis.

Results: In total, 388 hits were identified. Following a two-phase retrieval process, 88 articles were included in the final analysis. Rasch analysis was used to test the psychometric properties of the newly developed instrument, and validate or test a short version of the existing instrument. The reporting of Rasch analysis demonstrated large variability in quality. Rating scale functioning, internal scale validity using goodness-of-fit statistics, and unidimensionality were the most frequently reported outcomes.

Conclusion: The use of Rasch analysis in nursing science was found to be unsystematic. Rasch analysis could provide new possibilities for investigating measurement properties. However, robust, comprehensive, and precise reporting of the methodological choices and results of Rasch analysis is needed. Furthermore, the use of Rasch analysis in nursing science is encouraged.

What is already known

- Rasch analysis is a mathematical and probabilistic model based on the assumption that the probability of passing a single item is related to a person's ability and the difficulty of the item.
- Rasch analysis is widely used in the life sciences. However, its use and quality of reporting in nursing science have yet to be explored.

What this paper adds

- Rasch analysis is rarely used in nursing science, although its use is increasing.
- This review has identified deficiencies in the reporting of nursing validation studies using Rasch analysis.
- Minimum standards for the reporting of Rasch analysis in nursing research are proposed.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

* Corresponding author at: Department of Nursing Science, FI-20014 University of Turku, Finland.

E-mail addresses: minna.stolt@utu.fi (M. Stolt), anders.kottorp@mau.se (A. Kottorp), riisuh@utu.fi (R. Suhonen).

1. Introduction

In nursing, many instruments have been developed as outcome measures for patients (such as Patient-Reported Outcome Measures, Patient Reported Experience Measures), significant others, nurses, and managers to measure certain phenomena. The instruments focus, for example, on symptoms, competence, attitudes, and healthcare quality. The fundamental methodological requirement for constructing new instruments or using existing instruments is adequate psychometric properties. Psychometric properties are intrinsic characteristics of an instrument that are tested using statistical analysis (Ginty, 2013). Psychometric properties provide information about an instrument's appropriateness, meaningfulness, usefulness, and more broadly, its validity and reliability (Ginty, 2013). Evidence of psychometric properties supports the validity and reliability of the results obtained using a certain instrument (American Educational Research Association, 2014).

Psychometric properties can be evaluated using the Classical Test Theory or the Item Response Theory. Both approaches are suitable and often used complementarily. However, these two approaches differ in their evaluation levels. Classical Test Theory focuses on test-level evaluation, whereas Item Response Theory focuses on item-level evaluation (Tractenberg, 2010). Classical Test Theory, also known as 'True Score Theory', analyses test results based on test scores (Magno, 2009). Examples of the Classical Test Theory are methodological approaches such as item analysis and factor analysis (Bortolotti et al., 2013). Item Response Theory, also called 'Modern Test Theory', is a newer approach for testing the psychometrics of the instrument and examining the pattern of responses that the respondent makes to the set of items (Kline, 2005). In Item Response Theory, there are several models, such as the one-parameter logistic model, two-parameter logistic model, and three-parameter logistic model (Kline, 2005). Of these, the one-parameter logistic model, also known as the Rasch model (Kline, 2005), is perhaps the most common.

The Rasch model is an analytical model used to evaluate the measurement properties of rating scales using probability estimates (Rasch, 1980). The Rasch model is a logistic latent trait Item Response Theory model that focuses on the quality of the outcome measures. The analysis is based on the assumption that the probability of a person passing a single item in a test is related to assertions about a person's ability and the difficulty of the item toward which the empirical data are tested (Bond and Fox, 2015). Ultimately, the purpose of Rasch analysis is to provide more precise and accurate measurements of both persons and items that can support various aspects of validity and precision (American Educational Research Association, 2014).

Reporting of Rasch analysis depends on the focus of the report. Some central requirements for Rasch analysis should always be reported (Smith et al., 2003; Tennant and Conaghan, 2007; Salzberger, 2013). However, their universal use in reporting is selective, focusing primarily on reporting item goodness-of-fit (Leung et al., 2014). The central parts of Rasch analysis are the selection of the Rasch model and software, analysis of item hierarchy, rating scale functioning, unidimensionality, internal scale validity, person-separation reliability, and different item functioning.

The selection of the **Rasch model** depended on the response scale of the instrument. The Rasch model can be either dichotomous or polytomous (including the rating scale and the partial credit model). The dichotomous Rasch model is used with instruments of two response options, and a polytomous model is used in instruments with three or more response options (Lerdal et al., 2016). Rasch analysis is usable in instrument development and testing of the psychometric properties of existing instruments (Boone, 2016). Particularly in the instrument development process, Rasch analysis has advantages in terms of item reduction (Tennant et al., 2004). To conduct Rasch analysis, there are several statistical **software packages**, and Winsteps, RUMM, and Conquest are the most used in Rasch analysis (Leung et al., 2014).

Rasch analysis provides psychometric evidence of the function of a single item and persons. It orders the items and persons on the same continuum, where the most likely achieved items are located at the bottom of the continuum (Lerdal et al., 2016). The **hierarchy of the items** along the continuum determines the order of item challenge locations (also called calibrations, Prieto et al., 2003) in relation to the distribution of persons. Hierarchy can be evaluated using an item map (Wright map, Boone et al., 2014). Based on this information, test item targeting can be evaluated by comparing the mean item measures provided for the items and persons. Ideally, the mean item measures are at the same level, indicating that the items are neither too easy nor too difficult for the respondents (Boone, 2016). In addition, the flooring and ceiling effects are easily evaluated from the map. The flooring effect means there are no items for persons with low ability, whereas in the ceiling effect, items for persons with high ability are lacking (Mitchell and Jolley, 2010).

Rating scale functioning evaluates how each category of the instrument is working. Based on category functioning, the shape of the distribution and number of responses in each category can be evaluated. Regular distributions, such as uniform, normal, bimodal, and slightly skewed distributions, are expected. Categories with low frequencies are challenging because they lack the power to provide stable category threshold values (Bond and Fox, 2015). At least 10 responses in each category are considered minimum (Linacre, 1999).

Unidimensionality is the basic assumption of Rasch analysis, meaning that the test measures a single construct (Kline, 2005). Unidimensionality can be evaluated using a principal component analysis of the residuals. The usual criteria are that the first component should explain at least 50% of the variance and the second component should explain less than 5% (or eigenvalue less than 2.0, Linacre, 2011).

Internal scale validity and person response validity are typically reported using goodness-of-fit statistics. The goodness-of-fit is expressed using infit statistics and can be assessed at the item and person levels (Lerdal et al., 2016; Bond and Fox, 2015; Wright and Masters, 1982). It is crucial to identify items that do not fit the Rasch model. Including items with poor fit hampers the quality of measurement and decreases instrument precision. These items should be either removed, revised, or rewritten, and tested again. Goodness-of-fit statistics indicate how well each item fits the underlying test construct (Bond and Fox, 2015). In Rasch, goodness-of-fit statistics are reported using the mean square (MnSq) and standardised *z* statistics. The acceptable criterion is an item mean square infit of 0.6–1.4 (Wright and Linacre, 1994), but it is also recommended to be adjusted due to sample size (Smith et al., 2008). Higher mean square values mean larger variance in item response patterns than the Rasch model was expecting (Bond and Fox, 2015).

Related to person fit, Rasch model analysis produces evidence of **person-separation reliability**. It describes the separation of persons by their pattern of scores and better separation further reflects a more precise measurement (Wright, 1996; Bond and Fox, 2015). The person separation index should be evaluated against the measurement purpose and number of response options. However, the higher the value, the better the separation. Usually, the minimum requirement for the index is 2 or higher, which indicates that the instrument can separate persons from at least two strata, for example, low and high ability (Fisher, 1992).

Differential item functioning (DIF) can be used to evaluate how items work in different groups. For example, the data can be divided based on gender or age and by comparing the differential item functioning estimates in different groups. Evaluations focus on item functioning, in which items can function in either similar or different ways between different groups. If the functions differ between groups, this is a sign of differential item functioning (Bond and Fox, 2015).

Based on the previous reviews, Rasch analysis have been used in testing scales measuring mobility (Belvedere and de Morton, 2010),

musculoskeletal disease (Leung et al., 2014) and upper extremity outcomes (Hong and Bonilha, 2017). Rasch analysis has been widely used in psychology, education, and medicine (Bortolotti et al., 2013). It can be used in the development, evaluation, and administration of standardised measurements (Bortolotti et al., 2013).

However, a comprehensive review of the use of Rasch analysis in nursing research is lacking. This would clarify and identify the ways in which Rasch analysis could advance nursing research methodologically and increase the understanding of the instruments used to measure constructs of interest in nursing science. This methodological scoping review aimed to (i) describe the use of Rasch analysis in nursing research and (ii) determine the quality of reporting in studies using Rasch models in nursing. The ultimate goal was to emphasise the possibility of using Rasch analysis in nursing science as a choice in the evaluation of psychometric properties of the instruments.

2. Methods

2.1. Protocol and registration

This is a methodological scoping review. The study protocol was not published.

2.2. Eligibility criteria

Eligibility criteria were applied to each record, and the articles were included if they were: 1) conducted in a nursing context (studies focusing on nursing care in different care facilities or nursing education with nursing professionals, nursing students, or patients as informants), 2) empirical articles reported in English, and 3) Rasch analysis method was used to assess the instrument's measurement properties. Articles were excluded if they 1) used Rasch analysis as a method to evaluate values other than the measurement properties of the instrument (to compare outcome values between different groups).

2.3. Information sources and the search

A systematic electronic literature search was conducted on 1 February 2020 and updated on 16 April 2021 from PubMed/Medline and CINAHL databases (Fig. 1) by one researcher (MS). These databases are comprehensive in the field of nursing science (Subirana et al., 2005). The search was limited to covering the timeframe from the earliest (CINAHL since 1981, PubMed/Medline 1946) to 31 December 2020. The search terms (keywords) used were Rasch, IRT, item response theory, and nursing. To ensure adequate coverage of the search, it was conducted in two phases using two independent search sentences for both databases: first sentence was (Rasch [Title/Abstract]) AND (nursing[Title/Abstract]), second sentence, (IRT[Title/Abstract] OR "item response theory"[Title/Abstract]) AND (nursing[Title/Abstract]). The search was limited to the English language and title/abstract level.

2.4. Selection of sources of evidence

The selection process for sources of evidence was performed in two phases. The first two authors independently (MS, RS) evaluated the hits received from a two-phased search at the title and abstract levels and second, at the full-text level. After both phases, the researchers discussed the results and reached a consensus. In case of disagreement, a third author (AK) participated in the decision-making process.

2.5. Data charting process and data items

The data were first gathered in an Excel spreadsheet, including the following descriptive information: author, year, country, aim, setting,

sample population, and size. Following this, information on the measurement properties was tabulated, including scale rating, scale functioning, internal scale validity, person-response validity, person-separation validity, and internal consistency (Bonsaksen et al., 2013; Lerdal et al., 2014, 2016; Gay et al., 2016).

2.6. Synthesis of results

Using the tabulated information of the measurement properties, the data were analysed using quantification and content analyses.

3. Results

3.1. Selection of sources of evidence

In total, 388 hits were identified. After screening abstracts and titles, 218 were excluded because they were not of a nursing context, leaving 170 hits for full-text inspection. After full-text reading, 82 articles were excluded because they were in the field of medicine or occupational therapy ($n = 47$). Rasch analysis was used, for example, to generate interval measures to evaluate the effects of interventions instead of instrument testing as the focus ($n = 35$). In total, 88 articles were included in the final analysis.

3.2. Characteristics of sources of evidence

The size of the study samples varied (Table 1). The subjects of these studies were patients ($n = 37$), their next of kin ($n = 1$), nurses ($n = 38$), and nursing students ($n = 17$). The sample size varied from 43 to 13,113 (mean, 737). The contents of the studies/instruments focused predominantly on clinical care and practice from the perspectives of the competence of professionals or students ($n = 33$), well-being of patients ($n = 25$), physical activity of patients ($n = 11$), work well-being of professionals ($n = 10$), and ethics in health care ($n = 9$).

3.3. Use of Rasch analysis in nursing

The first article to report the use of Rasch analysis in nursing science was published in 1996 (Fig. 2). Until 2006, there was approximately one article per year, but after 2005, the use of Rasch analysis increased. In the last five years, the proportion of Rasch analysis papers has been approximately nine articles per year.

Rasch analysis was used to test the psychometric properties of a newly developed instrument ($n = 22$), to validate an already existing instrument with a different sample or language ($n = 63$), or to test a short version of the previously validated instrument ($n = 3$).

3.4. Reporting of the Rasch analysis

The reporting of Rasch analysis was of variable quality (Table 2). Winsteps (45%) was the most frequently used Rasch software, followed by RUMM and its different versions (30%). The mathematical derivative of the Rasch model (dichotomous or multi/partial credit) was stated in half of the articles (50%). The polytomous item-response Rasch model was predominantly used (34%). Rating scale functioning, internal scale validity using goodness-of-fit (infit) statistics, and unidimensionality were the most often reported (75%, 81%, and 67%, respectively). Person-response validity using person goodness-of-fit values was reported in 42% of articles. An item map showing item distribution along the continuum was illustrated in 39% of the articles. Differential item functioning, testing how responses differ based on age or gender, was used in half (50%) of the studies.

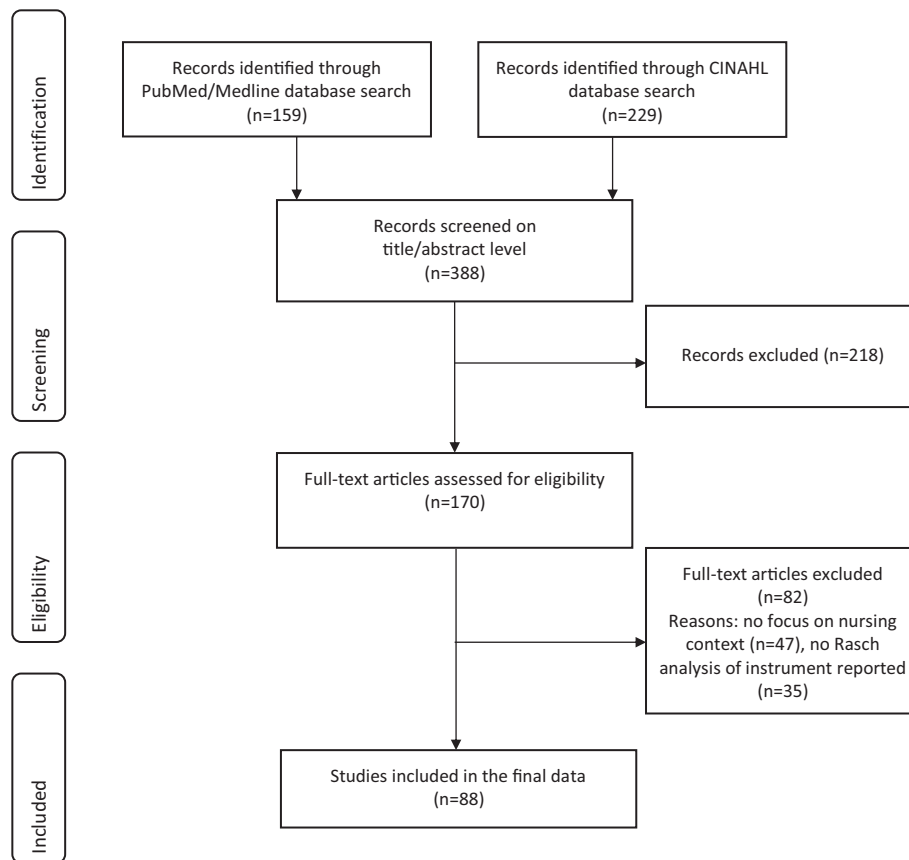


Fig. 1. Flow chart of the literature search process.

3.5. Synthesis of results

Since 1996, the use of Rasch analysis in nursing science has progressed steadily. Between 2010 and 2019, approximately six papers per year published Rasch analysis results. The specific reporting of Rasch analysis outcomes varied. Most studies reported the specific Rasch software used; however, only half reported the name of the Rasch model. Outcomes related to rating scale functioning, internal scale validity (including goodness-of-fit values), unidimensionality, and person-separation reliability were the most often reported values. Based on the analysis of the reporting of Rasch analysis in nursing research, a minimum reporting standard is suggested (Table 3). Uniform reporting increases transparency and quality of reporting and promotes the accurate use of scientific evidence of the instrument's psychometric properties.

4. Discussion

4.1. Summary of evidence

Rasch analysis has been increasingly used in nursing practice. This review demonstrated a smooth increase in the use of Rasch analysis over time. The quality of reporting seems to have improved over the years, potentially due to an increase in different reporting guidelines such as COSMIN (Gagnier et al., 2021). Classical Test Theory has been the most used approach in the assessment of the psychometric properties of instruments, providing evidence of scale-level properties. Instead, Rasch analysis can provide more detailed information and evidence regarding item- and person-level characteristics. It may yield information about sensitivity and provide instrument developers with more specific details to support

the development refinement process of an instrument. Where Classical Test Theory approaches are largely deductive and try to identify what items best cover the underlying construct, the Item Response Theory and Rasch analysis aim to approach the included items in a test and their individual contributions to capture the unidimensional concept within the item content in relation to empirical data gathering.

The focus in the reporting of Rasch results has been unsystematic and technically reported. Most of the articles reported item fit (how the actual item responses matched the expected responses), but other Rasch characteristics were reported to a more limited extent. This result is in line with a study by Leung et al. (2014), where reporting focused on fit statistics rather than on Rasch model assumptions. The use of selective reporting in instrument development and validation studies leaves the analysis of the validity and reliability of the instrument in question to a very superficial level. Outcomes in relation to unidimensionality, the basic assumption of Rasch analysis, were reported in only two-thirds of the articles. However, a more systematic reporting of the properties of the Rasch procedure is highly important. Therefore, a minimum reporting standard based on the results of this review was provided to the researcher to systematise the quality of reporting.

The studies reported different statistical software or programs used for Rasch analysis. Each software has its own benefits; however, each program produces different fit statistics, which may hamper the understandability and comparison of results. This also raises the problem of using incorrect method-based terminology for reporting. To advance the use of Rasch analysis, for example, in nursing science, clear guidelines and systematic approaches for constructing the analysis are needed. Many instruments have been criticised for being subjective,

Table 1
Descriptive characteristics of the studies (n = 88) included.

Author, year, country	Concept	Instrument	n of items	Dimensions	Sample, size
Pollak et al., 1996 USA	Functional independence	Functional Independence Measure (FIM)	18 items; 1 (= dependent)-7 (= independent) ordinal scale	Motor and cognitive	49, community, sheltered care, nursing facility
Smith Jr. et al., 2002 USA	Physical self-efficacy	The Perceived Physical Ability (PPA) subscale from the Physical Self Efficacy Scale	10 items (in PPA): 6-point rating scale from 1 = strongly agree to 6 = strongly disagree	Not reported	206 members of a national nursing association
Dimitrov and Shelestak, 2003 USA	Nursing students' performance	NLN Diagnostic Readiness Test	172 items; dichotomous 1 = mastered 0 = non-mastered	5: Client Needs, Nursing process, Categories of Human Functioning, Health alterations, Clinical Nursing	646 student RN
Finlayson et al., 2005 USA	Activities of Daily Living	Activities of Daily Living (ADL) and Instrumental Activities of Daily Living (IADL)	24 items	Not reported	607 older people
Slade et al., 2006 UK	Dependency	The Leeds Elderly Assessment Dependency Screening tool (LEADS)	17 items (result)	Resulting Modified Barthel Index; Abbreviated mental test; Northwich park Dependency Scale	258 older people
Gilworth et al., 2007 UK	Work disability	Nurse-Work Instability Scale Nurse-WIS	43 items; 5 levels 0-4 (high)	Not reported	296 nursing staff
Pearce et al., 2007 USA	Subjective fall concerns	The Activities-Specific Fall Caution (AFC) scale	13 items; 7 response options = > 4 category	1	57 staff, 234 residents nursing home
Resnick et al., 2007a USA	Restorative care activities	Restorative care Behaviour Checklist (RCBC)	10 observable activities; 3 level, n/a, yes, no	Not reported	386 nursing assistants
Resnick et al., 2007b USA	Ability to consent	Evaluation to Sign Consent (ESC)	5 items	1	346 residents nursing home
Velozo et al., 2007 USA	Functional Independence Measure and the Minimum Data Set	Functional Independence Measure (FIM) and the Minimum Data Set (MDS)	FIM: 18 items MDS: 284 items	Not reported	236 patients
Bouffoulox et al., 2008 Belgium	Satisfaction of activities and participation in the actual environment	SATIS-STROKE	Result: 36 items; 4-level scale: 0 = Very Dissatisfied to 3 = Very Satisfied.	1	101 stroke patients
Resnick et al., 2008 USA	Confidence in performing restorative care activities	Nursing Assistant Self-Efficacy for Restorative care Scale (NASERC)	NASERC; 10 items, 0 = no confidence, 10 very confident	NASERC 2: Functional Skill sub-scale Challenges sub-scale	384 nursing assistants
Wang et al., 2008 USA	Perceived benefits of restorative care Physical functioning and cognition scales	Nursing Assistant Outcome Expectations for Restorative Care Scale (NAOERC) Minimum Data Set	NAOERC; 9 items, 1 = strongly disagree, 5 strongly agree 5-point rating scale; 0 = independent, 4 = total dependence	NAOERC 1 Physical function and Cognition	654 veterans
Hagquist et al., 2009 Sweden	Self-Efficacy	Nursing Self-Efficacy scale	9 items: 0-11 category responses	1	1379 nursing students
Iacono et al., 2009 Australia	Interaction with Disabled Persons	Interaction with Disabled Persons scale (IDP)	20 items: six response categories ranging from strongly agree to strongly disagree 57 items	6	373 healthcare students
Lamoureux et al., 2009 Australia	Quality of life	Nursing Home Vision-Targeted Health-Related Quality-of-Life questionnaire (NHVQoL)		9 subscales: 1) general vision, 2) reading, 3) ocular symptoms, 4) mobility, 5) psychological distress, 6) activities of daily living, 7) social activities and hobbies, 8) adaptation and coping, 9) and social interaction Dimensions: 1) Emotional, 2) Emotional/ Cognition, 3) Self-determination, 4) Social relationships and School, 5) Physical function and 6) Experiences of care	76 care facility residents
Nisell et al., 2009 Sweden	Imperforate anus	Imperforate Anus Psychosocial Questionnaire (IAPSQ)	45 items: Likert scales	Dimensions: 1) Emotional, 2) Emotional/ Cognition, 3) Self-determination, 4) Social relationships and School, 5) Physical function and 6) Experiences of care five domains: 1) self-esteem, 2) positive affect, 3) negative affect, 4) feelings of belonging, and 6) sense of aesthetics	87 children
Adler and Resnick, 2010 USA	Quality of life	Dementia Quality of Life (DQOL)	29 items: 5-point scale from 1 = never to 5 = very often	1) physical function, 2) psychosocial function, 3) pain or discomfort 2 dimensions: 1) exhaustion, 2) disengagement	486 nursing home residents
Franchignoni et al., 2010 Italy	Oral health	Geriatric Oral Health Assessment Index	12 items: a five-point scale from 0 = never to 4 = always		85 long-term residents
Gustavsson et al., 2010 Sweden	Burnout	Oldenburg Burnout Inventory (OLBI)	16 items: 4-point response scale ranging from 'Does not apply at all' to 'Applies completely'		933 nursing professionals
Hula et al., 2010 USA	Aphasia	Western Aphasia Battery (WAB)	NR	Not reported	101 individuals with aphasia
Farin et al., 2011 Germany	Communication preferences of patients with chronic illness	Communication preferences of chronically ill patients (KOPRA questionnaire)	32 items (result) Six response options: (1) not so important-(5)extremely important, and (6) "unsuitable/bad"	4 subscales: a) Patient participation and patient orientation, b) Effective and open communication, c) Emotionally supportive communication, d) Communication about personal circumstances	472 patients
Winstanley and White, 2011 Australia	Clinical supervision	Manchester Clinical Supervision Scale (MCSS)	36 items: 5-point scale from strongly disagree to strongly agree	6	385 nursing staff

(continued on next page)

Table 1 (continued)

Author, year, country	Concept	Instrument	n of items	Dimensions	Sample, size
Flannery et al., 2012 USA	Job attitude	Job Attitude Scale (JAS)	17 items: 5-point Likert scale; a higher score indicates a higher degree of job satisfaction	5 domains: 1) pay, 2) interaction/organisational factors, 3) task requirements, 4) job status, 5) autonomy	508 nursing assistants
Panella et al., 2012 Italy	Risk screening	Blaylock Risk Assessment Screening Score (BRASS)	10 items: delivers a summative score ranging from 0 to 40	Not reported	104 patients
González-de Paz et al., 2012 Spain	Nurses' ethical sensitivity	Ethical Sensitivity Scale	35 items: 5-point Likert scale with a score ranging from 0 = almost never to 4 = always.	Not reported	143 nurses
Forkmann et al., 2013 Germany	Depression	Patient Health questionnaire (PHQ-9)	9 items: response categories range from 0 = not at all to 3 = nearly every day	Not reported	1631 older people
Gerrard, 2013 USA	Functional status	Katz Index of Independence in Activities of Daily Living (Katz ADL)	6 items, patients give functional grade from A to G based on their abilities to perform the tasks. (A = independence with all ADLs, G = dependence in all ADLs)	6 activities of daily living (ADL): 1) bathing, 2) dressing, 3) toileting, 4) transferring, 5) continence, and 6) feeding	13,113 patients
Guedes Ede et al., 2013 Brazil	Power as an intentional participation in changes	Power as Knowing Participation in Change Tool (PKPCT)	48 items: ranging from 1 to 7 points.	4 operational indicators of power: 1) awareness, 2) choices, 3) freedom to act intentionally, 4) involvement in creating change	952 nursing assistants 627 baccalaureate nurses
La Porta et al., 2013 Italy	Coma recovery	Coma Recovery Scale-Revised (CRS-R)	29 items	6 subscales addressing 1) auditory, 2) visual, 3) motor, 4) oromotor/verbal, 5) communication, 6) arousal functions.	129 patients (258 observations)
McMullen and Resnick, 2013 USA	Self-esteem	Rosenberg Self-Esteem Scale (RSES)	10 items: 4-point Likert scale from strongly agree to strongly disagree	2 dimensions: 1) positive self-esteem, 2) negative self-esteem	508 nursing assistants
Müller, 2013 Switzerland	Nurse competence	Nurse Competence Scale	73 items: visual analogue scale (0–100 mm) 0 = low level to 100 = very high level of competence) Also 4-point scale from 0 = not to 3 = very often, to state how often these 73 competencies were used.	7-factor model: 1) helping role, 2) teaching-coaching, 3) diagnostic functions, 4) managing situations, 5) therapeutic interventions, 6) ensuring quality, 7) work role competencies. Result: 7-factor model confirmed	679 nurses
Nicholson et al., 2013 Australia	Competence of an instrument nurse	Performance Based Scoring Rubric	(1) Behavioural descriptors 1–4 (4 = highest level of performance) (2) Fours broad band performance levels: Beginning, Developing, Consolidating and Effective Practitioner	Includes two parts: 1) Analytical Observation Form, 2) Holistic Competence Rubrics	95 instrument nurses (were observed and assessed by 32 nurse educators and preceptors)
Suhonen et al., 2013 Finland Greece Sweden UK	Individualised care	Individualised Care Scale (patient)	34 items: 5-point Likert scale ranging from 1 = fully disagree to 5 = fully agree	Dimensions: 1) Support of Individuality ICS-A, 2) Perceptions of Individuality ICS-B Both dimensions (A&B) include three subscales: 1) Clinical Situation: Clin A and Clin B; 2) Personal Life Situation: Pers A and B; 3) Decisional Control: Dec A and Dec B	1093 orthopaedic and trauma patients
Guttersrud et al., 2014 Norway	Critical nutrition literacy	Critical nutrition literacy (CNL) instrument	19 items: 5-point Likert scale ranging from 1 = disagree strongly to 5 = agree strongly	Two scales: 1) the 'engagement in dietary habits, 2) the 'taking a critical stance toward nutrition claims and their sources	473 students
Holloway et al., 2014 UK	Confidence in responding to alcohol use in patients	The Clinical Confidence Questionnaire	40 items: 4-point Likert scale using from 4 = Very easy to do - 1 too difficult to do.	Not reported	43 registered nurses
Ma et al., 2014 Taiwan	Workplace bullying	Negative Acts Questionnaire-Revised (NAQ-R)	22 items: 5 response alternatives 1 = never, 2 = occasionally, 3 = monthly, 4 = weekly, 5 = daily	5 categories: 1) physical aspects, 2) psychological aspects, 3) interpersonal relations at work, 4) willingness to work, 5) quality of work	300 nurses
Strout and Howard, 2014 USA	Wellness	The Wellness Assessment Tool (WEL)	22 items	5 dimensions: 1) social wellness, 2) intellectual wellness, 3) physical wellness, 4) emotional wellness, 5) spiritual wellness	5604 community-dwelling adults
Blackman and Giles, 2015 Australia	Evidence-Based Practice	Evidence-Based Practice Tool	27 items: 4-point Likert scale ranging from 1 = very difficult to do to 4 = very simple to do	Not reported	471 nursing students
Bourke et al., 2015 USA	Clinical Stress	Nursing Students' Clinical Stress Scale	59 items: 5-point Likert type scale ranging from 1 = Very much to 5 = None	Not reported	110 students
Classen et al., 2015 USA Canada	Fitness to drive	Fitness-to-Drive Screening Measure (FTDS)	54 items: 5-point adjectival scale 1 = cannot do - 5 = not difficult	Not reported	200 older drivers 200 caregivers
Snowden et al., 2015 UK	Emotional Intelligence	Trait Emotional Intelligence Questionnaire Short form (TEIQue-S)	30 items: seven possible responses ranging from 1 = Completely Disagree to 7 = Completely Agree	Four factor structure: 1) Well-being, 2) Self-control, 3) Emotionality, 4) Sociability	938 undergraduate nursing and computing students
Spurlock Jr and Wonder, 2015 USA	Evidence-based practice knowledge	Evidence-Based Practice Knowledge Assessment in Nursing (EKAN)	Result: 20 items: 5-point Likert-type scale	Not reported	200 undergraduate nursing students

Table 1 (continued)

Author, year, country	Concept	Instrument	n of items	Dimensions	Sample, size
Ahmad et al., 2016 Malaysia England	Work satisfaction	Index of Work Satisfaction (IWS)	44 items	Six components: a) pay, b) autonomy, c) task requirements, d) professional status, e) interaction, f) organisational policies	556 nurses
Chan et al., 2016 HongKong	Core competencies for infection control nurses	ICN Advanced Practice Core Competency Scale	83 items: 5-point rating scale: 1 = not very important to 5 = very important	Not reported	117 infection control nurse
Jeon et al., 2016 Australia	Depression in dementia	The Cornell Scale for Depression in Dementia (CSDD-19)	19 items: 2-point scale (0 = absent, 1 = mild or intermittent, 2 = severe) Result: Short version includes 4 items (CSDD-4)	Not reported	556 nursing home residents
Müller et al., 2016 Germany	Activities and participation in older individuals with joint contractures	PaArticular Scales	Original scale from "no problem" to "severe problem".	Result: Two independent scales. 1) International Classification of Functioning, Disability and Health component Activities, 2) International Classification of Functioning, Disability and Health component Participation	191 older people
Stolt et al., 2016 Cyprus, Greece, Finland, Sweden	Trust in nurses	Trust in Nurses Scale (TNS)	4 items: 5-point Likert-type (1 = never, 5 = always)	1	599 cancer patients
Blackman and Giles, 2017 Australia	Evidence-based practice	Evidence-Based Practice Survey	27 items: 4-point Likert scale (4 = very simple to do, 1 = very difficult to do)	Not reported	375 graduating nursing students
Carretta et al., 2017 Italy	Patient satisfaction	Customer Satisfaction Audit	22 items: (5-point scale 1 = unimportant, 5 = important)	Not reported	3320 discharged patients
Finbråten et al., 2017 Norway	Health literacy	European Health Literacy Survey Questionnaire (HLS-EU-Q47)	47 items: four-point rating scale 1 = very easy - 4) very difficult	a) accessing, b) understanding, c) applying, d) health information	999 members from the Diabetes Association
Galik et al., 2017 USA	Resistiveness to care	Resistance to Care Scale	Observation over 5 min period: 1) frequency of the resistiveness to care behaviour 2) duration of the behaviour 3) intensity of the behaviour (1 = mild, 2 = moderate, 3 = extreme)	Not reported	261 nursing home residents
Jung et al., 2017 South Korea	(1) Nursing assistant self-efficacy for restorative care (2) Nursing assistant outcome expectations for restorative care	1) Nursing Assistant Self-Efficacy for Restorative Care Scale (2) Nursing Assistant Outcome Expectations for Restorative Care Scale	(1) 10 items: 10-point Likert (2) 9 items: 5-point Likert	(1) Two subscales: a) NASERC for functional skills, b) NASERC for challenges associated with restorative care	697 direct care workers
Kleinknecht-Dolf et al., 2017 Switzerland	Moral distress	Moral Distress Scale	10 items: 5-point scale 0 = never, 4 = very often	Not reported	2153 nurses and 1965 nurses
Ma et al., 2017 Taiwan	Online bullying	Negative Acts Questionnaire-Revised (NAQ-R)	22 items: 5 response options 1 = never, 2 = occasionally, 3 = monthly, 4 = weekly, 5 = daily	Not reported	963 nurses
Nguyen et al., 2017 Australia	Dementia specific health	dementia-specific health state classification system based on the QOL-AD instrument (quality of life in Alzheimer's disease)	15 items: four-point scale: 1 = poor, 4 = excellent	Not reported	284 residents in long-term care facilities
Rojas Sánchez et al., 2017 Colombia	Cardiac output	Decreased cardiac output (DCO)	21 items: Dichotomous scale rating "absent"/"present"	Not reported	200 patients
Sanabria-Arenas et al., 2017 Colombia	Patient satisfaction	the Scale for Evaluation of Haemodialysis Patient's Satisfaction with Service provided	44 items: Likert scale 1 = very unsatisfied, 5 = very satisfied	1) overall satisfaction, 2) personnel at the unit, medications and supplies, facilities and processes, phone contact	370 haemodialysis patients
Álvarez-García et al., 2018 Spain	Environmental health	Children's Environmental Health Knowledge Questionnaire (ChEHK-Q)	26 items: three response options True, False, Don't know	Not reported	308 nursing students
Heritage et al., 2018 Australia	Quality of life	the Children's Environmental Health Skills Questionnaire (ChEHS-Q)	12 items: Likert scale values from 1 (strongly disagree) to 5 (strongly agree)	compassion satisfaction, secondary traumatic stress, and burnout	Registered nurses or nurse officers n = 1615
Li et al., 2018 China	Functioning	Professional Quality of Life (ProQOL)	21 items	three components: body functions, body structures, activities	140 spinal injury patients
Lundberg et al., 2018 Sweden	Research utilisation	Research Utilisation Questionnaire	29 items: 5-point Likert scale 1 = strongly agree, 5 = strongly disagree	attitudes toward research, availability and support for implementation of research findings, use of research findings	163 nurses and nurse aides in nursing home

(continued on next page)

Table 1 (continued)

Author, year, country	Concept	Instrument	n of items	Dimensions	Sample, size
Milliken et al., 2018	Ethical awareness	Ethical Awareness Scale	18 items: three response options	Not reported	nurses n = 116
Parra-Anguita et al., 2018 Spain	Dementia knowledge	Dementia Knowledge Assessment Tool 2 (DKAT2-Sp)	21 items: three answer choice (Yes, No, I don't know)	Not reported	nursing staff n = 361, nursing students n = 297
Resnick et al., 2018 USA	Quality of life	the Quality of Life in Late-Stage Dementia (QUALID) Scale	29 items: responses range from 1 = never to 5 = very often	5 domains: self-esteem, positive affect and humour, negative affect, feelings of belonging, and sense of aesthetics	137 older people
Volz-Sidiropoulou et al., 2018 Germany	Disease related impact	The ICIQ-Cog consists of two scales: a) ICIQ-Cog-P measures disease-specific bother b) ICIQ-Cog-C measures cognitive impairment	a) 12-items: 4-point Likert type scale (rarely; occasionally; frequently; very frequently) b) 4 items: 4-point Likert scale (strongly disagree; somewhat disagree; somewhat agree; strongly agree)	Not reported	Proxy ratings were collected for these 60 residents
Happell et al., 2019a Australia, Ireland, Norway, Finland, the Netherlands	Stigma of mental illness	Opening Minds Scale	20 statements: to be rated on a range of (dis)agreement: 1-strongly agree, 2-agree, 3-neither agree nor disagree, 4-disagree, 5-strongly disagree.	social distance, attitudes of health care providers and disclosure/help-seeking.	Undergraduate nursing student n = 423
Happell et al., 2019b Australia, Ireland, Norway, Finland, the Netherlands	Attitudes to people labelled with a diagnosis of mental illness	Mental Health Nurse Education Survey	24 items: 7-point scale ranging from 'Strongly Disagree' to 'Strongly Agree'	7 domains: Preparedness for Mental Health Field, Knowledge of Mental Illness, Negative Stereotypes, Anxiety Surrounding Mental Illness, Future Career in Mental Health Nursing, Valuable Contributions, and Course Effectiveness	University nursing students in Australia and Western Europe n = 424
Jacob et al., 2019 Australia	Critical thinking	Not reported	five unfolding scenarios covering a variety of common clinical situations which incorporated 18 items: score 3 always has ethical implications, score 2 may have ethical implications, score 1 never has ethical implications	25 multiple choice questions relating to these scenarios	128 nursing students
Milliken et al., 2019	Ethical awareness	Ethical Awareness Scale	10 indicators of the NOC Swallowing status: 1 (worst result) to 5 (best health status)	Not reported	n = 240 nurses
Oliveira-Kumakura et al., 2019	Swallowing status	The Nursing Outcomes Classification (NOC)	23 items: three options: "Yes", "No", "I don't know"	Not reported	227 post-stroke patients
Parra-Anguita et al., 2019	Alzheimer's disease knowledge	UJA Alzheimer's Care Scale	30 affirmations: each judged relation with their musculoskeletal pain	Not reported	262 nursing workers
Petersen et al., 2019	Work instability	the Brazilian Nurse-Work Instability Scale	29 items that assess type and frequency of missed care and 17 items that identify reasons for missed care	B part 29 items: 1 to 5 (from "never missed" to "always missed"). C part 17 items: to 4 ("not a reason", "minor reason", "moderate reason" and "significant reason")	331 nurses
Riklikienė et al., 2019 Lithuania	Missed nursing care	Missed Nursing Care in Infection Prevention and Control Survey A part: background information, B part: type and frequency of missed nursing care, C part: indication why this care might be missed	40 items: 2 parallel parts for patients and nurses. 4-point scale 4 (fully agree) to 1 (fully disagree); the option 0 no opinion is also possible.	Both parts 7 quality categories: (A) characteristics of actors, (B) nursing actions, (C) preconditions for care, (D) environment, (E) proceeding of the process, (F) patient management strategies, and (G) collaboration with family members/ significant others.	surgical patients (n = 476) and nurses (n = 167)
Wilberforce et al., 2019 UK	Person-centred care	Person-Centred Climate Questionnaire - Staff version (PCQ-S)	14 items: 6-point Likert scale (1 = No, I disagree completely, to 6 = Yes, I agree completely).	three subscales: spanning safety, homeliness and community.	4831 nurses
Areskoug-Josefsson and Rolander, 2020 Sweden and Denmark	Sexual health	the Students' Attitudes toward Sexual Health (SA-SH)	22 items: 5-point Likert scale; disagree, partly disagree, partly agree, agree, strongly agree	four domains: present feelings of comfortableness, future working environment, fear of negative influence on future patient relations, and educational needs	699 students from nursing, physiotherapy, occupational therapy, orthopaedic engineering, and social work.
Browall et al., 2020 Sweden	Attitudes toward care of dying patients	The Frommelt Attitudes Toward Care of the Dying (FATCOD)	30 items: 5-point Likert scale (1 disagree, 2 [no category name], 3 means neither/nor, 4 [no category name], 5 means agree).	positive attitudes toward caring for dying patients (17 items), perceptions of patient and family-centered care (8 items)	39 with no formal education, 125 enrolled students, 477 undergraduate students, 71 registered nurses, 287 specialist nursing students
Dickens et al., 2020 Australia	Violence prevention climate	violence prevention climate scale (VPC-14)	14 items: 5-point Likert scale (Strongly Agree to Strongly Disagree).	patient actions' (things patients do that prevent violence) and 'staff actions' (the things that staff do)	213 responses from nursing staff and patients
Hübsch et al., 2020 Switzerland	Performed and missed nursing care	the MISSCARE questionnaire	24 listed nursing activities: 5-point response scale how frequently ('always missed' to 'never missed') they or their unit's staff missed any of listed nursing activities	Not reported	1030 nurses and midwives

Table 1 (continued)

Author, year, country	Concept	Instrument	n of items	Dimensions	Sample, size
Kalánková et al., 2020 Slovak Republic	Rationing of care	the Perceived Implicit Rationing of Nursing Care (PIRNCA)	31 items: 5-point frequency scale (0 – 'not needed', 1 – 'never', 2 – 'rarely', 3 – 'sometimes', 4 – 'often').	including activities related to assistance with physical care, implementation of a prescribed treatment plan, emotional support and education, patient surveillance, coordination of care and discharge planning and documentation of care	895 registered nurses
López-Franco et al., 2020 Spain	Pressure injury prevention	Pressure Injury Prevention Knowledge (PIPK) questionnaire	35 items: True, False, I don't know	Not reported	438 nursing professionals
Nick et al., 2020 USA	Evidence-based practice	Evidence-based Practice Knowledge Assessment in Nursing (EKAN) instrument	20 items: three answer options	research appraisal, quality improvement, the EBP process, and other topics pertinent to the EBP movement.	103 baccalaureate nursing students
Resnick et al., 2020 USA	Person-centered care	the Knowledge of Person-Centered Behavioural Approaches for behavioural and psychological symptoms associated with dementia	10 items: correct, incorrect, missing	how to manage a resident with a specific behaviour (6 items), ways to prevent BPSD in residents with dementia (3 items), assessment of underlying capability of the resident to guide the development of appropriate person-centered care plans (1 item)	1071 nurses
Samur et al., 2020 Turkey	Standard precautions	the Compliance with Standard Precautions Scale	20 items: a four-point scale (never, seldom, sometimes and always)	Not reported	411 nurses
Stolt et al., 2020 Finland	Person-centered care climate	Person-Centered care Climate Questionnaire-Patient	17 items: 7-point Likert scale (from 1 = No, I disagree completely to 7 = Yes, I agree completely).	climate of safety (10 items), climate of everydayness (4 items) and climate of hospitality (3 items)	111 older people with heart failure
Westergren et al., 2020 Sweden	Participation in the care of older people	Next of Kin Participation in Care (NoK-PiC).	37 items: five possible responses were: disagree (0), somewhat disagree (1), neither agree nor disagree (2), somewhat agree (3), and strongly agree (4).	seven subject areas: trusting the staff (8 items), being present (3 items), conversations and information (5 items), relationship with the staff (s6 items), completing a task (5 items), being respected for one's knowledge (2 items), and being acknowledged as part of the care team (8 items)	364 next of kin of older people in nursing homes
Westergren and Melgaard, 2020 Denmark	Eating	The Minimal Eating Observation Form – II (MEOF-II)	9 items: the number of eating difficulties 0–9 in the total scale, and 0–3 in the subscales.	three subscales: food intake (3 items), swallowing/mouth (3 items), energy/appetite (3 items)	302 acute geriatric patients

and interpretations of the results obtained may be biased due to self-reporting (Althubaiti, 2016). Rasch analysis may help researchers add systematic, critical sensitivity, and specificity of the measurement while being able to analyse how the items capture different individuals' responses. This adds to our understanding of the objectiveness of measurement (American Educational Research Association, 2014). Most importantly, this method also enables us to understand how the measure captures the topic under measurement. Unlike classical test theory, Rasch models test the relationship between the performance on an individual item and the respondent's overall ability to measure the construct of the survey designs in a probabilistic manner.

One of the advantages of performing Rasch analyses on instruments is that it provides researchers with evidence in relation to response processes both at the item level and at the person level. As Classical Test Theory usually focuses on item functioning, Rasch analysis provides researchers with information on “both sides of the coin” (item and person responses). A test may show a high degree of item goodness-of-fit but still demonstrate relatively large proportions of persons not meeting the criteria of person goodness-of-fit. If such differences in responses among persons are systematic, this can also require an analysis of DIF, followed by additional actions to minimise such problems in test use. Therefore, the findings from this review call for further action when

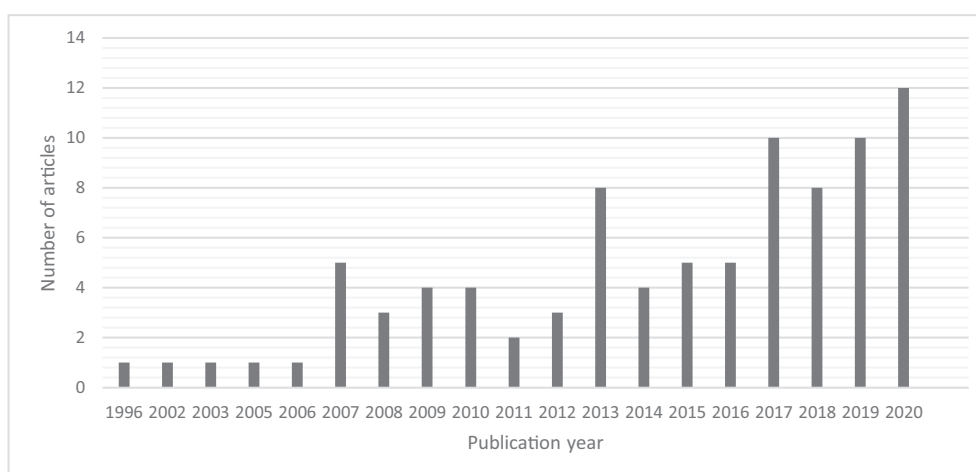


Fig. 2. Number of published articles reporting Rasch analysis results in nursing science between 1996 and 2020.

Table 2
Measurement properties reported from Rasch analysis (n = 88).

Reference	Program name			Model		Rating scale function		Goodness-of-fit	Unidimensionality	Person-response validity	Person-separation reliability	Differential item functioning	Item map			
	Winsteps	RUMM	ConQuest	R (eRM)	Other	Not reported	Dichotomous							Multi/Partial credit	Not reported	
total f. (%)	40 (45%)	26 (30%)	3 (3%)	4 (5%)	10 (11%)	6 (7%)	14 (16%)	30 (34%)	44 (50%)	66 (75%)	71 (81%)	59 (67%)	37 (42%)	61 (69%)	44 (50%)	34 (39%)
Pollak et al., 1996	x								x							
Smith Jr. et al., 2002	x								x							
Dimitrov and Shelestak, 2003					x				x							
Finlayson et al., 2005	x								x							
Slade et al., 2006		x							x							
Gilworth et al., 2007	x						x		x							
Pearce et al., 2007	x								x							
Resnick et al., 2007a	x								x							
Resnick et al., 2007b					x				x							
Velozo et al., 2007	x								x							
Bouffieux et al., 2008		x							x							
Resnick et al., 2008					x				x							
Wang et al., 2008	x						x		x							
Hagquist et al., 2009	x						x		x							
Iacono et al., 2009	x								x							
Lamoureux et al., 2009	x								x							
Nisell et al., 2009	x								x							
Adler and Resnick, 2010									x							
Franchignoni et al., 2010	x								x							
Gustavsson et al., 2010		x							x							
Hula et al., 2010	x								x							
Farin et al., 2011									x							
Winstanley and White, 2011		x							x							
Flannery et al., 2012	x								x							
Panella et al., 2012	x								x							
González-de Paz et al., 2012	x								x							
Forkmann et al., 2013									x							
Gerrard, 2013									x							
Guedes Ede et al., 2013	x								x							
La Porta et al., 2013		x							x							
McMullen and Resnick, 2013	x								x							
Müller, 2013									x							
Nicholson et al., 2013		x							x							
Suhonen et al., 2013	x								x							
Gutterstrud et al., 2014		x							x							

Table 2 (continued)

Reference	Program name		Model		Rating scale function		Goodness-of-fit	Unidimensionality	Person-response validity	Person-separation reliability	Differential item functioning	Item map
	Winsteps	RUMM	ConQuest R (eRM)	Other	Not reported	Dichotomous						
Areskog-Josefsson and Rolander, 2020	x					x			x			
Browall et al., 2020		x				x	x		x			
Dickens et al., 2020				x		x	x		x			x
Hübsch et al., 2020			x			x	x		x			
Kalánková et al., 2020	x					x	x		x			
López-Franco et al., 2020				x			x					
Nick et al., 2020												
Resnick et al., 2020		x				x	x		x			x
Samur et al., 2020				x			x					
Stolt et al., 2020		x				x	x		x			x
Westergren and Meigaard, 2020						x	x		x			
Westergren et al., 2020	x						x					

Table 3

Minimum reporting standards of Rasch analysis in nursing research (criteria base on Wright and Linacre, 1994, Linacre, 2004, Tennant et al., 2004, Tennant and Conaghan, 2007, Lerdal et al., 2014, Gagnier et al., 2021).

Measurement properties	Description of the information to be reported
Rasch software used	The name of the software used for Rasch analysis
Sample size	Number of participants and justification for sample size
Model	Dichotomous or multi/partial credit
Rating scale functioning	Does the rating scale function consistently across items?
Internal scale validity (GFS)	How well do the actual item responses match the expected responses from the Rasch model?
Internal scale validity, unidimensionality	Is the scale unidimensional?
Person-response validity (GFS)	How well do the actual individual responses match the expected responses from the Rasch model?
Person-separation reliability	Can the scale distinguish at least 2 distinct groups of people in the sample?
Differential Item Functioning	Are item difficulty calibrations stable in relation to demographic and clinical variables?
Person-Item map	Distribution of the persons and items along the continuum

applying Rasch analysis in nursing science, as only 42% and 50% of the studies reported person goodness-of-fit and differential item functioning, respectively. To demonstrate evidence in relation to response processes and fairness in testing (American Educational Research Association, 2014), such Rasch outcomes should be reported to a much greater extent in nursing.

Rasch analysis has many advantages worth highlighting. Compared to Classical Test Theory, Rasch analysis can be conducted with smaller samples ($n = 30-50$, Cappelleri et al., 2014). However, the sample size must be estimated considering the type and number of items and response categories, as well as how well the test items are targeted to the sample (Linacre, 1994; Linacre, 2002). However, with relatively small sample sizes, Rasch analysis allows a thorough investigation of items, persons, and their difficulty or ability with relatively precise confidence (Linacre, 1994). With this investigation, it is possible to understand item-level strengths and weaknesses and to test how values change if some items are removed from the analysis.

The use of Rasch analysis in nursing science has increased rapidly since 2006. This growth may relate to advancing methodological competence among researchers and the desire to seek modern statistical analyses to supplement classical testing. Classical Test Theory has a strong status in instrumentation in nursing science. Although the levels of measurement using categorical data are common in health and social sciences and have also been criticised (e.g. Prieto et al., 2003), Classical Theory testing has been largely used. Therefore, new methodological approaches and programs should be used as they may provide alternative, differently focused, and detailed information about measures and measurements, thereby advancing the details and levels of analysis (Waltz et al., 2016).

4.2. Limitations

This review had some limitations which must be considered. First, only two databases (PubMed/Medline and CINAHL) were used to locate the studies. The use of other databases would have increased the number of potential studies. However, Medline and CINAHL are comprehensive and relevant, particularly in nursing (Subirana et al., 2005). As this was a methodological review, the studies found were used to illustrate and advance our understanding of the use of the Rasch model, and not to cover all possible studies. Second, the search terms were wide, simple, and free. Rasch analysis or item response theory is not categorised in the Medical Subject Headings (Leung et al., 2014), so the use of free search terms is obvious. The search was limited to the title and abstract levels, and because of this, there might have been some articles left

outside the review if the search terms were not mentioned in the title or abstract. However, the search yielded a large number of hits, indicating the appropriateness of the search. The retrieval process of the studies was conducted by two researchers (MS, RS) who independently increased the robustness of the study selection. All possible discrepancies were discussed by the research team, and a consensus was achieved. The review focused only on studies using Rasch analysis in the nursing context. Despite extensive inclusion criteria and systematic search process, there might be studies conducted by nursing professionals in the nursing context which were not identified in this review. Third, the review focused on describing the use and quality of reporting in Rasch analysis. To achieve a more in-depth description of reporting of Rasch analysis, the quality of the statistical analysis could have been beneficial. However, the use of different statistical software may challenge the uniform analysis and reporting of Rasch statistics. Fourth, the quality appraisal of the studies was not performed. The focus of this review was to analyse the use and reporting quality of Rasch analysis in nursing research, including all possible studies with different methodologies. Therefore, studies were not excluded based on methodological quality. This might have limited the validity of the review; however, it provides an overall description of the implementation of Rasch analysis in nursing research.

5. Conclusions

The use of Rasch analysis in the field of nursing has increased in recent years. Despite its advantages in the analysis of the psychometric properties of instruments, its use in nursing is still relatively rare. As instruments developed specifically from the point of view of nursing science have increased over the years, they have matured and are used in cross-cultural studies. This analytical method provides strength for the validation of the developed instruments. However, in future, more emphasis should be placed on systematic reporting of the results of Rasch analysis to obtain quality publications. Regardless of the limited use of all the possibilities Rasch analysis can provide and the missing information in earlier reporting, while Rasch analysis has been widely implemented in research, the use of Rasch analysis in nursing science is highly recommended, and researchers are encouraged to apply it.

Funding

None.

Declaration of Competing Interest

None.

References

- Adler, E., Resnick, B., 2010. Reliability and validity of the dementia quality of life measure in nursing home residents. *West. J. Nurs. Res.* 32, 686–704. <https://doi.org/10.1177/0193945909360780>.
- Ahmad, N., Oranye, N.O., Danilov, A., 2016. Rasch analysis of Stamps's index of work satisfaction in nursing population. *Nurs. Open* 4, 32–40. <https://doi.org/10.1002/nop2.61>.
- Althubaiti, A., 2016. Information bias in health research: definition, pitfalls, and adjustment methods. *J. Multidiscip. Healthc.* 9, 211–217. <https://doi.org/10.2147/JMDH.S104807>.
- Álvarez-García, C., Álvarez-Nieto, C., Pancorbo-Hidalgo, P.L., Sanz-Martos, S., López-Medina, I.M., 2018. Student nurses' knowledge and skills of children's environmental health: instrument development and psychometric analysis using item response theory. *Nurse Educ. Today* 69, 113–119. <https://doi.org/10.1016/j.nedt.2018.07.008>.
- American Educational Research Association, 2014. *The Standards for Educational and Psychological Testing*. American Educational Research Association, Washington D.C.
- Areskoug-Josefsson, K., Rolander, B., 2020. Value of performing a Rasch analysis on a reliable and valid instrument-case study of the SA-SH. *J. Nurs. Meas.* 28, 322–342. <https://doi.org/10.1891/JNM-D-18-00106>.
- Belvedere, S.L., de Morton, N.A., 2010. Application of Rasch analysis in health care is increasing and is applied for variable reasons in mobility instruments. *J. Clin. Epidemiol.* 63, 1287–1297. <https://doi.org/10.1016/j.jclinepi.2010.02.012>.
- Blackman, I.R., Giles, T., 2015. Psychometric evaluation of a self-report evidence-based practice tool using Rasch analysis. *Worldviews Evid.-Based Nurs.* 12, 253–264. <https://doi.org/10.1111/wvn.12105>.
- Blackman, I.R., Giles, T.M., 2017. Can nursing students practice what is preached? Factors impacting graduating Nurses' abilities and achievement to apply evidence-based practices. *Worldviews Evid.-Based Nurs.* 14, 108–117. <https://doi.org/10.1111/wvn.12205>.
- Bond, T.G., Fox, C.M., 2015. *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. 3rd ed. Routledge, New York, NY.
- Bonsaksen, T., Kottorp, A., Gay, C., Fagermoen, M.S., Lerdal, A., 2013. Rasch analysis of the general self-efficacy scale in a sample of persons with morbid obesity. *Health Qual. Life Outcomes* 11, 202. <https://doi.org/10.1186/1477-7525-11-202>.
- Boone, W.J., 2016. Rasch analysis for instrument development: why, when, and how? *CBE life Sci. Educ.* 15, rm4. <https://doi.org/10.1187/cbe.16-04-0148>.
- Boone, W.J., Staver, J.R., Yale, M.S., 2014. *Wright maps: First steps. Rasch Analysis in the Human Sciences*. Springer, Dordrecht.
- Bortolotti, S.L.V., Tezza, R., de Andrade, D.F., Borna, A.C., de Sousa, Jr, A. F., 2013. Relevance and advantages of using the item response theory. *Qual. Quant.* 47, 2341–2360. <https://doi.org/10.1007/s11135-012-9684-5>.
- Bouffouix, E., Arnould, C., Thonnard, J.L., 2008. SATIS-stroke: A satisfaction measure of activities and participation in the actual environment experienced by patients with chronic stroke. *J. Rehabil. Med.* 40, 836–843. <https://doi.org/10.2340/16501977-0272>.
- Bourke, M., Wallace, L., Greskamp, M., Tormoehlen, L., 2015. Improving objective measurement in nursing research: Rasch model analysis and diagnostics of the nursing Students' clinical stress scale. *J. Nurs. Meas.* 23, E1–15. <https://doi.org/10.1891/1061-3749.23.1.E1>.
- Browall, M., Pakpour, A.H., Melin-Johansson, C., Lundh Hagelin, C., Österlind, J., Henoch, I., 2020. Development and psychometric evaluation of a new short version of the Swedish Frommelt attitudes toward Care of the Dying Scale. *Cancer Nurs.* <https://doi.org/10.1097/NCC.0000000000000794>.
- Cappelleri, J.C., Jason Lundy, J., Hays, R.D., 2014. Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. *Clin. Ther.* 36, 648–662. <https://doi.org/10.1016/j.clinthera.2014.04.006>.
- Carretta, E., Bond, T.G., Cappiello, G., Fantini, M.P., 2017. Looking through the Patients' eyes: measuring patient satisfaction in a public hospital. *J. Patient Experience* 4, 121–128. <https://doi.org/10.1177/23743735170706614>.
- Chan, W.F., Bond, T.G., Adamson, B., Chow, M., 2016. Identifying Core competencies of infection control nurse specialists in Hong Kong. *Clin. Nurse Spec.* 30, E1–E9. <https://doi.org/10.1097/NUR.0000000000000174>.
- Classen, S., Velozo, C.A., Winter, S.M., Bédard, M., Wang, Y., 2015. Psychometrics of the fitness-to-drive screening measure. *OTJR: Occup. Particip. Health* 35, 42–52.
- Dickens, G.L., Tabvuma, T., Hadfield, K., Hallett, N., 2020. Violence prevention climate in general adult inpatient mental health units: validation study of the VPC-14. *Int. J. Ment. Health Nurs.* 29, 1101–1111. <https://doi.org/10.1111/inm.12750>.
- Dimitrov, D.M., Shelestak, D., 2003. Psychometric analysis of performance on categories of client needs and nursing process with the NLN diagnostic readiness test. *J. Nurs. Meas.* 11, 207–223.
- Farin, E., Gramm, L., Kosiol, D., 2011. Development of a questionnaire to assess communication preferences of patients with chronic illness. *Patient Educ. Couns.* 82, 81–88.
- Finbråten, H.S., Pettersen, K.S., Wilde-Larsson, B., Nordström, G., Trollvik, A., Guttersrud, Ø., 2017. Validating the European health literacy survey questionnaire in people with type 2 diabetes: latent trait analyses applying multidimensional Rasch modeling and confirmatory factor analysis. *J. Adv. Nurs.* 73, 2730–2744. <https://doi.org/10.1111/jan.13342>.
- Finlayson, M., Mallinson, T., Barbosa, V.M., 2005. Activities of daily living (ADL) and instrumental activities of daily living (IADL) items were stable over time in a longitudinal study on aging. *J. Clin. Epidemiol.* 58, 338–349. <https://doi.org/10.1016/j.jclinepi.2004.10.008>.
- Fisher, W., 1992. Reliability, separation, strata statistics. *Rasch Meas. Trans.* 6, 238.
- Flannery, K., Resnick, B., Galik, E., Lipscomb, J., McPhaul, K., 2012. Reliability and validity assessment of the job attitude scale. *Geriatr. Nurs.* 33, 465–472. <https://doi.org/10.1016/j.gerinurse.2012.05.002>.
- Forkmann, T., Gauggel, S., Spangenberg, L., Brähler, E., Glaesmer, H., 2013. Dimensional assessment of depressive severity in the elderly general population: psychometric evaluation of the PHQ-9 using Rasch analysis. *J. Affect. Disord.* 148, 323–330. <https://doi.org/10.1016/j.jad.2012.12.019>.
- Franchignoni, M., Giordano, A., Levrini, L., Ferriero, G., Franchignoni, F., 2010. Rasch analysis of the geriatric Oral health assessment index. *Eur. J. Oral Sci.* 118, 278–283. <https://doi.org/10.1111/j.1600-0722.2010.00735.x>.
- Gagnier, J.J., Lai, J., Mokkink, L.B., Terwee, C.B., 2021. COSMIN reporting guideline for studies on measurement properties of patient reported outcome measures. *Qual. Life Res.* 30, 2197–2218.
- Galik, E., Resnick, B., Vigne, E., Holmes, S.D., Nalls, V., 2017. Reliability and validity of the Resistiveness to care scale among cognitively impaired older adults. *J. Am. Med. Dir. Assoc.* 18, 59–64. <https://doi.org/10.1016/j.jamda.2016.08.008>.
- Gay, C.L., Kottorp, A., Lerdal, A., Lee, K.A., 2016. Psychometric limitations of the Center for Epidemiologic Studies-Depression Scale for assessing depressive symptoms among adults with HIV/AIDS: A Rasch analysis. *Depress. Res. Treat.*, 2824595 <https://doi.org/10.1155/2016/2824595>.
- Gerrard, P., 2013. The hierarchy of the activities of daily living in the Katz index in residents of skilled nursing facilities. *J. Geriatr. Phys. Ther.* 36, 87–91. <https://doi.org/10.1519/JPT.0b013e318268da23>.
- Gilworth, G., Bhakta, B., Eyres, S., Carey, A., Anne Chamberlain, M., Tennant, A., 2007. Keeping nurses working: development and psychometric testing of the nurse-work instability scale (nurse-WIS). *J. Adv. Nurs.* 57, 543–551.

- Ginty, A.T., 2013. Psychometric properties. In: Gellman, M.D., Turner, J.R. (Eds.), *Encyclopedia of Behavioral Medicine*. Springer, New York, NY https://doi.org/10.1007/978-1-4419-1005-9_480.
- González-de Paz, L., Kostov, B., Sisó-Almirall, A., Zabalegui-Yárnoz, A., 2012. A Rasch analysis of nurses' ethical sensitivity to the norms of the code of conduct. *J. Clin. Nurs.* 21, 2747–2760. <https://doi.org/10.1111/j.1365-2702.2012.04137.x>.
- Guedes Ede, S., Orozco-Vargas, L.C., Turrini, R.N., de Sousa, R.M., dos Santos, M.A., da Cruz Dde, A., 2013. Rasch analysis of the power as knowing participation in change tool—the Brazilian version. *Rev. Latino-Am. Enfermagem* 21, 163–171.
- Gustavsson, J.P., Hallsten, L., Rudman, A., 2010. Early career burnout among nurses: modelling a hypothesized process using an item response approach. *Int. J. Nurs. Stud.* 47, 864–875. <https://doi.org/10.1016/j.ijnurstu.2009.12.007>.
- Guttersrud, O., Dalane, J.Ø., Pettersen, S., 2014. Improving measurement in nutrition literacy research using Rasch modelling: examining construct validity of stage-specific 'critical nutrition literacy' scales. *Public Health Nutr.* 17, 877–883. <https://doi.org/10.1017/S1368980013000530>.
- Hagquist, C., Bruce, M., Gustavsson, J.P., 2009. Using the Rasch model in nursing research: an introduction and illustrative example. *Int. J. Nurs. Stud.* 46, 380–393. <https://doi.org/10.1016/j.ijnurstu.2008.10.007>.
- Happell, B., Platania-Phung, C., Scholz, B., Bocking, J., Horgan, A., Manning, F., Doody, R., Hals, E., Granerud, A., Jan van der Vaart, K., Allon, J., Lahti, M., Pulli, J., Vatula, A., Ellilä, H., Griffin, M., Russell, S., MacGabhann, L., Björnsson, E., Biering, P., 2019a. Assessment of the opening minds scale for use with nursing students. *Perspect. Psychiatr. Care* 55, 661–666. <https://doi.org/10.1111/ppc.12393>.
- Happell, B., Platania-Phung, C., Scholz, B., Bocking, J., Horgan, A., Manning, F., Doody, R., Hals, E., Granerud, A., Lahti, M., Pulli, J., Vatula, A., Ellilä, H., van der Vaart, K.J., Allon, J., Griffin, M., Russell, S., MacGabhann, L., Björnsson, E., Biering, P., 2019b. Nursing student attitudes to people labelled with 'mental illness' and consumer participation: A survey-based analysis of findings and psychometric properties. *Nurse Educ. Today* 76, 89–95. <https://doi.org/10.1016/j.nedt.2019.02.003>.
- Heritage, B., Rees, C.S., Hegney, D.G., 2018. The ProQOL-21: A revised version of the professional quality of life (ProQOL) scale based on Rasch analysis. *PLoS One* 13, e0193478. <https://doi.org/10.1371/journal.pone.0193478>.
- Holloway, A., Blackman, I., Flynn, F., 2014. A study examining the appropriateness of a self-rated alcohol-related clinical confidence tool as a method of measurement among registered hospital nurses using Rasch analysis. *J. Adv. Nurs.* 70, 937–945. <https://doi.org/10.1111/jan.12264>.
- Hong, I., Bonilha, H.S., 2017. Psychometric properties of upper extremity outcome measures validated by Rasch analysis: a systematic review. *Int. J. Rehabil. Res.* 40, 1–10. <https://doi.org/10.1097/MRR.0000000000000202>.
- Hübsch, C., Müller, M., Spirig, R., Kleinknecht-Dolf, M., 2020. Performed and missed nursing care in Swiss acute care hospitals: conceptual considerations and psychometric evaluation of the German MISSCARE questionnaire. *J. Nurs. Manag.* 28, 2048–2060. <https://doi.org/10.1111/jonm.13026>.
- Hula, W., Donovan, N.J., Kendall, D.L., Gonzalez-Rothi, L.J., 2010. Item response theory analysis of the Western aphasia battery. *Aphasiology* 24, 1326–1341.
- Iacono, T., Tracy, J., Keating, J., Brown, T., 2009. The interaction with disabled persons scale: revisiting its internal consistency and factor structure, and examining item-level properties. *Res. Dev. Disabil.* 30, 1490–1501. <https://doi.org/10.1016/j.ridd.2009.07.010>.
- Jacob, E.A., Duffield, C., Jacob, A.M., 2019. Validation of data using RASCH analysis in a tool measuring changes in critical thinking in nursing students. *Nurse Educ. Today* 76, 196–199.
- Jeon, Y.H., Liu, Z., Li, Z., Low, L.F., Chenoweth, L., O'Connor, D., Beattie, E., Davison, T.E., Brodaty, H., 2016. Development and validation of a short version of the Cornell scale for depression in dementia for screening residents in nursing homes. *Am. J. Geriatr. Psychiatr.* 24, 1007–1016. <https://doi.org/10.1016/j.jagp.2016.05.012>.
- Jung, D., Byun, J., Lee, M., Kim, H., 2017. Psychometric testing of Korean versions of self-efficacy and outcome expectations for restorative care activities scales. *Geriatr. Nurs.* 38, 207–212.
- Kalánková, D., Suhonen, R., Stolt, M., Kurucová, R., Katajisto, J., Žiaková, K., Gurková, E., 2020. Psychometric testing of perceived implicit rationing of nursing care (PIRNCA). *J. Adv. Nurs.* 76, 1469–1482. <https://doi.org/10.1111/jan.14351>.
- Kleinknecht-Dolf, M., Spichiger, E., Müller, M., Bartholomeyczik, S., Spirig, R., 2017. Advancement of the German version of the moral distress scale for acute care nurses—A mixed methods study. *Nurs. Open* 4, 251–266. <https://doi.org/10.1002/nop.291>.
- Kline, T.J., 2005. *Modern test theory: Assumptions, equations, limitations, and item analyses*. In: Kline, T.J. (Ed.), *Psychological Testing: A Practical Approach to Design and Evaluation*. SAGE Publications Ltd, Thousand Oaks, CA.
- La Porta, F., Caselli, S., Ianes, A.B., Cameli, O., Lino, M., Piperno, R., Sighinolfi, A., Lombardi, F., Tennant, A., 2013. Can we scientifically and reliably measure the level of consciousness in vegetative and minimally conscious states? Rasch analysis of the coma recovery scale-revised. *Arch. Phys. Med. Rehabil.* 94, 527–535. <https://doi.org/10.1016/j.apmr.2012.09.035>.
- Lamoureux, E.L., Fenwick, E., Moore, K., Klaic, M., Borschmann, K., Hill, K., 2009. Impact of the severity of distance and near-vision impairment on depression and vision-specific quality of life in older people living in residential care. *Invest. Ophthalmol. Vis. Sci.* 50, 4103–4109. <https://doi.org/10.1167/iov.08-3294>.
- Lerdal, A., Kottorp, A., Gay, C.L., Grov, E.K., Lee, K.A., 2014. Rasch analysis of the Beck depression inventory-II in stroke survivors: a cross-sectional study. *J. Affect. Disord.* 158, 48–52. <https://doi.org/10.1016/j.jad.2014.01.013>.
- Lerdal, A., Kottorp, A., Gay, C., Aouizerat, B.E., Lee, K.A., Miasowski, C., 2016. A Rasch analysis of assessments of morning and evening fatigue in oncology patients using the Lee fatigue scale. *J. Pain Symptom Manag.* 51, 1002–1012. <https://doi.org/10.1016/j.jpainsymman.2015.12.331>.
- Leung, Y.-Y., P, M.-E., Conaghan, P., Tennant, A., 2014. A systematic literature review on the application of Rasch analysis in musculoskeletal disease – A special interest group report of OMERACT 11. *J. Rheumatol.* 41, 159–164.
- Li, K., Yan, T., You, L., Xie, S., Li, Y., Tang, J., Wang, Y., Gao, Y., 2018. Psychometric properties of the international classification of functioning, disability and health set for spinal cord injury nursing based on Rasch analysis. *Disabil. Rehabil.* 40, 338–345. <https://doi.org/10.1080/09638288.2016.1250169>.
- Linacre, J.M., 1994. Sample size and item calibration stability. *Rasch Meas. Trans.* 7, 328.
- Linacre, J.M., 1999. Understanding Rasch measurement: estimation methods for Rasch measures. *J. Outcome Meas.* 3, 381–405.
- Linacre, J.M., 2002. Optimizing rating scale category effectiveness. *J. Appl. Meas.* 3, 85–106.
- Linacre, J.M., 2004. Rasch model estimation: further topics. *J. Appl. Meas.* 5, 95–110.
- Linacre, J.M., 2011. Rasch measures and Unidimensionality. *Rasch Meas. Trans.* 24, 1310.
- López-Franco, M.D., Parra-Anguila, L., Comino-Sanz, I.M., Pancorbo-Hidalgo, P.L., 2020. Development and psychometric properties of the pressure injury prevention knowledge questionnaire in Spanish nurses. *Int. J. Environ. Res. Public Health* 17, 3063. <https://doi.org/10.3390/ijerph17093063>.
- Lundberg, V., Boström, A.M., Malinowsky, C., 2018. A psychometric evaluation of the Swedish version of the research utilization questionnaire using a Rasch measurement model. *Scand. J. Caring Sci.* 32, 586–593. <https://doi.org/10.1111/scs.12482>.
- Ma, S.C., Chien, T.W., Wang, H.H., Li, Y.C., Yui, M.S., 2014. Applying computerized adaptive testing to the negative acts questionnaire-revised: Rasch analysis of workplace bullying. *J. Med. Internet Res.* 17, e50. <https://doi.org/10.2196/jmir.2819>.
- Ma, S.C., Wang, H.H., Chien, T.W., 2017. A new technique to measure online bullying: online computerized adaptive testing. *Ann. General Psychiatry* 16, 26. <https://doi.org/10.1186/s12991-017-0149-z>.
- Magno, C., 2009. Demonstrating the difference between classical test theory and item response theory using derived test data. *Int. J. Educ. Psychol. Assess.* 1, 1–11.
- McMullen, T., Resnick, B., 2013. Self-esteem among nursing assistants: reliability and validity of the Rosenberg self-esteem scale. *J. Nurs. Meas.* 21, 335–344.
- Milliken, A., Ludlow, L., DeSanto-Madeya, S., Grace, P., 2018. The development and psychometric validation of the ethical awareness scale. *J. Adv. Nurs.* 74, 2005–2016.
- Milliken, A., Ludlow, L., Grace, P., 2019. Ethical awareness scale: replication testing, invariance analysis, and implications. *AJOB Empir. Bioeth.* 10, 231–240. <https://doi.org/10.1080/23294515.2019.1666176>.
- Mitchell, M.L., Jolley, J.M., 2010. *Research Design Explained*. 7th ed. Wadsworth, Belmont, CA.
- Müller, M., 2013. Nursing competence: psychometric evaluation using Rasch modelling. *J. Adv. Nurs.* 69, 1410–1417. <https://doi.org/10.1111/jan.12009>.
- Müller, M., Oberhauser, C., Fischer, U., Bartoszek, G., Saal, S., Strobl, R., Meyer, G., Grill, E., 2016. The PaArticular scales - A new outcome measure to quantify the impact of joint contractures on activities and participation in individuals in geriatric care: development and Rasch analysis. *Int. J. Nurs. Stud.* 59, 107–117. <https://doi.org/10.1016/j.ijnurstu.2016.04.002>.
- Nguyen, K.H., Mulhern, B., Kularatna, S., Byrnes, J., Moyle, W., Comans, T., 2017. Developing a dementia-specific health state classification system for a new preference-based instrument AD-5D. *Health Qual. Life Outcomes* 15, 21. <https://doi.org/10.1186/s12955-017-0585-0>.
- Nicholson, P., Griffin, P., Gillis, S., Wu, M., Dunning, T., 2013. Measuring nursing competencies in the operating theatre: instrument development and psychometric analysis using item response theory. *Nurse Educ. Today* 33, 1088–1093. <https://doi.org/10.1016/j.nedt.2012.04.008>.
- Nick, J.M., Roberts, L., Adrien, F.M., Wonder, A.H., 2020. Establishing A global EBP tool to measure evidence-based practice knowledge: translating and testing the French version of EKAN. *J. Nurs. Meas.* 28, 370–381. <https://doi.org/10.1891/JNM-D-19-00003>.
- Nisell, M., Brodin, U., Christensson, K., Rydelius, P.A., 2009. The imperforate anus psychosocial questionnaire (IAPSQ): its construction and psychometric properties. *Child Adolesc. Psychiatry Ment. Health* 14, 15. <https://doi.org/10.1186/1753-2000-3-15>.
- Oliveira-Kumakura, A.R.S., Alonso, J.B., de Carvalho, E.C., 2019. Psychometric assessment of the nursing outcome swallowing status: Rasch model approach. *Int. J. Nurs. Knowl.* 30, 197–202.
- Panella, L., La Porta, F., Caselli, S., Marchisio, S., Tennant, A., 2012. Predicting the need for institutional care shortly after admission to rehabilitation: Rasch analysis and predictive validity of the BRASS index. *Eur. J. Phys. Rehab. Med.* 48, 443–454.
- Parra-Anguila, L., Moreno-Cámara, S., López-Franco, M.D., Pancorbo-Hidalgo, P.L., 2018. Validation of the Spanish version of the dementia knowledge assessment tool 2. *J. Alzheimers Dis.* 65, 1175–1183. <https://doi.org/10.3233/JAD-180290>.
- Parra-Anguila, L., Sánchez-García, I., Del Pino-Casado, R., Pancorbo-Hidalgo, P.L., 2019. Measuring knowledge of Alzheimer's: development and psychometric testing of the UJA Alzheimer's care scale. *BMC Geriatr.* 19, 63. <https://doi.org/10.1186/s12877-019-1086-2>.
- Pearce, N.J., Myers, A.M., Blanchard, R.A., 2007. Assessing subjective fall concerns in residential living seniors: development of the activities-specific fall caution scale. *Arch. Phys. Med. Rehabil.* 88, 724–731.
- Petersen, R.S., Tennant, A., Nakagawa, T.H., Marziale, M., 2019. Translation, adaptation and validation of the nurse-work instability scale to Brazilian Portuguese. *Rev. Latino-Am. de Enfermagem* 27, e3170. <https://doi.org/10.1590/1518-8345.2943.3170>.
- Pollak, N., Rheault, W., Stoecker, J.L., 1996. Reliability and validity of the FIM for persons aged 80 years and above from a multilevel continuing care retirement community. *Arch. Phys. Med. Rehabil.* 77, 1056–1061.
- Prieto, L., Alonso, J., Lamarca, R., 2003. Classical test theory versus Rasch analysis for quality of life questionnaire reduction. *Health Qual. Life Outcomes* 1, 27. <https://doi.org/10.1186/1477-7525-1-27>.
- Rasch, G., 1980. *Probabilistic Model for some Intelligence and Achievement Tests*. University of Chicago Press, Chicago, IL.

- Resnick, B., Rogers, V., Galik, E., Gruber-Baldini, A.L., 2007a. Measuring restorative care provided by nursing assistants: reliability and validity of the restorative care behavior checklist. *Nurs. Res.* 56, 387–398.
- Resnick, B., Gruber-Baldini, A.L., Pretzer-Aboff, I., Galik, E., Buie, V.C., Russ, K., Zimmerman, S., 2007b. Reliability and validity of the evaluation to sign consent measure. *Gerontologist* 47, 69–77.
- Resnick, B., Galik, E., Pretzer-Aboff, I., Rogers, V., Gruber-Baldini, A.L., 2008. Testing the reliability and validity of self-efficacy and outcome expectations of restorative care performed by nursing assistants. *J. Nurs. Care Qual.* 23, 162–169. <https://doi.org/10.1097/01.NCQ.0000313766.09891.43>.
- Resnick, B., Galik, E., Kolanowski, A., Van Hartsma, K., Boltz, M., Ellis, J., Behrens, L., Flanagan, N.M., 2018. Reliability and validity testing of the quality of life in late-stage dementia scale. *Am. J. Alzheimers Dis. Other Dement.* 33, 277–283. <https://doi.org/10.1177/1533317518765133>.
- Resnick, B., Kolanowski, A., Van Hartsma, K., Galik, E., Boltz, M., Ellis, J., Behrens, L., Eshraghi, K., Viviano, N., Madrigal, C., 2020. Reliability and validity of the knowledge of person-centered behavioral approaches for BPSD test. *J. Nurs. Meas.* 28, 472–488. <https://doi.org/10.1891/JNM-D-19-00016>.
- Riklikienė, O., Blackman, I., Bendinskaite, I., Henderson, J., Willis, E., 2019. Measuring the validity and reliability of the Lithuanian missed nursing care in infection prevention and control scales using Rasch analysis. *J. Nurs. Manag.* 28, 2025–2035. <https://doi.org/10.1111/jonm.12918>.
- Rojas Sánchez, L.Z., Hernández Vargas, J.A., Trujillo Cáceres, S.J., Roa Díaz, Z.M., Jurado Arenales, A.M., Toloza Pérez, Y.G., 2017. Usefulness of the diagnosis “decreased cardiac output (00029)” in patients with chronic heart failure. *Int. J. Nurs. Knowl.* 28, 192–198. <https://doi.org/10.1111/2047-3095.12148>.
- Salzberger, T., 2013. Reporting a Rasch analysis. In: Christensen, K.B., Kreiner, S., Mesbah, M. (Eds.), *Rasch Models in Health*. STLE, London UK.
- Samur, M., Seren Intepeler, S., Lam, S.C., 2020. Adaptation and validation of the compliance with standard precautions scale amongst nurses in Turkey. *Int. J. Nurs. Pract.* 26, e12839. <https://doi.org/10.1111/ijn.12839>.
- Sanabria-Arenas, M., Marín, J.T., Certuche-Quintana, M.C., Sánchez-Pedraza, R., 2017. Validation of an instrument for measuring satisfaction of patients undergoing hemodialysis. *BMC Health Serv. Res.* 17, 321. <https://doi.org/10.1186/s12913-017-2251-y>.
- Slade, A., Fear, J., Tennant, A., 2006. Identifying patients at risk of nursing home admission: the Leeds elderly assessment dependency screening tool (LEADS). *BMC Health Serv. Res.* 6, 31. <https://doi.org/10.1186/1472-6963-6-31>.
- Smith Jr., E.V., Conrad, K.M., Chang, K., Piazza, J., 2002. An introduction to Rasch measurement for scale development and person assessment. *J. Nurs. Meas.* 10, 189–206.
- Smith, R.M., Linacre, J.M., Smith Jr., E.V., 2003. Guidelines for manuscripts. *J. Appl. Meas.* 4, 198–204. <https://doi.org/10.1891/jnum.10.3.189.52562>.
- Smith, A.B., Rush, R., Fallowfield, L.J., Velikova, G., Sharpe, M., 2008. Rasch fit statistics and sample size considerations for polytomous data. *BMC Med. Res. Methodol.* 8, 33. <https://doi.org/10.1186/1471-2288-8-33>.
- Snowden, A., Watson, R., Stenhouse, R., Hale, C., 2015. Emotional intelligence and nurse recruitment: Rasch and confirmatory factor analysis of the trait emotional intelligence questionnaire short form. *J. Adv. Nurs.* 71, 2936–2949. <https://doi.org/10.1111/jan.12746>.
- Spurlock Jr., D., Wonder, A.H., 2015. Validity and reliability evidence for a new measure: the evidence-based practice knowledge assessment in nursing. *J. Nurs. Educ.* 54, 605–613. <https://doi.org/10.3928/01484834-20151016-01>.
- Stolt, M., Charalambous, A., Radwin, L., Adam, C., Katajisto, J., Lemonidou, C., Patiraki, E., Sjövall, K., Suhonen, R., 2016. Measuring trust in nurses – psychometric properties of the Trust in Nurses Scale in four countries. *Eur. J. Oncol. Nurs.* 25, 46–54. <https://doi.org/10.1016/j.ejon.2016.09.006>.
- Stolt, M., Katajisto, J., Kottorp, A., Leino-Kilpi, H., 2019. Measuring quality of care: A Rasch validity analysis of the good nursing care scale. *J. Nurs. Care Qual.* 34, E1–E6. <https://doi.org/10.1097/NCQ.0000000000000391>.
- Stolt, M., Koskenvuori, J., Edvardsson, D., Katajisto, J., Suhonen, R., 2020. Validation of the Finnish person-centered care climate questionnaire–patient and testing the relationship with individualised care. *Int. J. Older People Nurs.* 16, e12356. <https://doi.org/10.1111/opn.12356>.
- Strout, K., Howard, E.P., 2014. Application of the Rasch model to measure five dimensions of wellness in community-dwelling older adults. *J. Nurs. Meas.* 22, 268–290. <https://doi.org/10.1891/1061-3749.22.2.26>.
- Subirana, M., Sola, I., Garcia, J.M., Gich, I., Urrutia, G., 2005. A nursing qualitative systematic review required MEDLINE and CINAHL for study identification. *J. Clin. Epidemiol.* 58, 20–25. <https://doi.org/10.1016/j.jclinepi.2004.06.001>.
- Suhonen, R., Schmidt, L.A., Katajisto, J., Berg, A., Idvall, E., Kalafati, M., Land, L., Lemonidou, C., Välimäki, M., Leino-Kilpi, H., 2013. Cross-cultural validity of the individualised care scale – a Rasch model analysis. *J. Clin. Nurs.* 22, 648–660. <https://doi.org/10.1111/j.1365-2702.2012.04238.x>.
- Tennant, A., Conaghan, P.G., 2007. 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.* 57, 1358–1362. <https://doi.org/10.1002/art.23108>.
- Tennant, A., McKenna, S.P., Hagell, P., 2004. Application of Rasch analysis in the development and application of quality of life instruments. *Value Health* 7, S22–S26. <https://doi.org/10.1111/j.1524-4733.2004.7s106.x>.
- Tractenberg, R.E., 2010. Classical and modern measurement theories, patient reports, and clinical outcomes. *Contemp. Clin. Trials* 31, 1–3. [https://doi.org/10.1016/S1551-7144\(09\)00212-2](https://doi.org/10.1016/S1551-7144(09)00212-2).
- Velozo, C.A., Byers, K.L., Wang, Y.C., Joseph, B.R., 2007. Translating measures across the continuum of care: using Rasch analysis to create a crosswalk between the functional Independence measure and the minimum data set. *J. Rehabil. Res. Dev.* 44, 467–478. <https://doi.org/10.1682/jrrd.2006.06.0068>.
- Volz-Sidiropoulou, E., Rings, T., Wagg, A.S., Leistner, N., Gauggel, S., Kirschner-Hermanns, R., 2018. Development and initial psychometric properties of the ‘ICIQ-Cog’: a new assessment tool to measure the disease-related impact and care effort associated with incontinence in cognitively impaired adults. *BJ. Int.* 122, 309–316. <https://doi.org/10.1111/bju.14186>.
- Waltz, C.F., Strickland, O.L., Lenz, E.R., 2016. *Measurement ing and Health Research*. 4th ed. Springer Publishing Company, New York, NY.
- Wang, Y.C., Byers, K.L., Velozo, C.A., 2008. Rasch analysis of minimum data set mandated in skilled nursing facilities. *J. Rehabil. Res. Dev.* 45, 1385–1399.
- Westergren, A., Melgaard, D., 2020. The minimal eating observation form-II Danish version: psychometric and metrological aspects. *J. Nurs. Meas.* <https://doi.org/10.1891/JNM-D-18-00084>.
- Westergren, A., Behm, L., Lindhardt, T., Persson, M., Ahlström, G., 2020. Measuring next of kin’s experience of participation in the care of older people in nursing homes. *PLoS One* 15, e0228379. <https://doi.org/10.1371/journal.pone.0228379>.
- Wilberforce, M., Sköldunger, A., Edvardsson, D., 2019. A Rasch analysis of the person-Centred climate questionnaire – staff version. *BMC Health Serv. Res.* 19, 996. <https://doi.org/10.1186/s12913-019-4803-9>.
- Winstanley, J., White, E., 2011. The MCSS-26: revision of the Manchester clinical supervision scale using the Rasch measurement model. *J. Nurs. Meas.* 19, 160–178.
- Wright, B.D., 1996. Reliability and separation. *Rasch measurement*. *Transactions* 9, 472.
- Wright, B.D., Linacre, J.M., 1994. Reasonable mean-square fit values. *Rasch Meas. Trans.* 8, 370.
- Wright, B.D., Masters, G.N., 1982. *Rating Scale Analysis: Rasch Measurement*. Mesa Press, Chicago.