

25 **Abstract**

26 **Objectives:** The purpose of this literature review was to clarify the sufficiency and efficiency
27 needs, as well as the stratification and the assignment scoring principles of the International
28 Classification of Functioning, Disability and Health (ICF) core set, in conjunction with the Item
29 Response Theory (IRT)-derived functioning outcome measures.

30 **Methods:** A literature search has been done on the PubMed database. We included studies
31 involving the IRT-derived ICF evaluation tool. Our main outcome was the assignment and
32 stratification scoring principle of IRT-derived ICF-based functioning outcome measures.

33 **Results:** 46 relative articles met the eligibility criteria of our review. On this basis, this review
34 was undertaken to track the research progress of the ICF-based and IRT-modeling measurement
35 tools and screen the final core set. This review included a classification of IRT modeling for ICF
36 studies and a summary of existing research paradigms of IRT-derived ICF outcome measures.
37 Moreover, this review identified blind spots related to scoring assignment principles and
38 “difficulty parameter” in current ICF-based IRT studies.

39 **Conclusion:** The ICF is the worldwide terminology system for functioning categories, while
40 ICF-based evaluation in clinical practice and scientific research has elicited the underlying
41 challenges. However, the algorithm of the ICF-based IRT modeling may advance the
42 understanding of ICF clinical application and enable a new paradigm for the design of IRT-
43 derived ICF questionnaires, namely the parsimonious core set of ICF. Additionally, the Wright
44 map has the potential to facilitate an understanding of the rehabilitation process and tailor
45 rehabilitation goals.

46 **Keywords:** Item Response Theory (IRT), International Classification of Functioning, Disability
47 and Health (ICF), Wright map, Research paradigm

48 **Introduction**

49 Rehabilitation science is a discipline dedicated to comprehending and facilitating human
50 functioning. In order to guide and inform clinical decisions, it is crucial to identify potential
51 disabilities and monitor changes in functioning across the continuum of care. However,
52 inconsistent and numerous outcome measure scales pose a major challenge to multidisciplinary
53 communication. Additionally, determining the most appropriate functioning assessment scale
54 presents several challenges. Firstly, it is inevitable to encounter a trade-off between the
55 sufficiency and efficiency of assessment tools. Relying solely on disease-specific scales may fail
56 to provide a comprehensive functioning assessment for patients with multiple health issues. A
57 thorough assessment can provide valuable insights into rehabilitation management and functional
58 recovery, while efficient clinical assessment tools that can rapidly and conveniently achieve
59 outcome measures(Rolstad, Adler, and Rydén, 2011). The ideal functioning assessment should
60 consider the length and complexity of assessment tools limited to a certain extent, so as to
61 minimize the response burden on frail patients and optimize resource efficiency. Another concern
62 is that many rehabilitation assessment scales, such as manual muscle testing, Barthel index, etc.
63 use ordinal scales, that may not precisely capture the full range of functioning changes over time.

64 To tackle these barriers, the World Health Organization has introduced the International
65 Classification of Functioning, Disability and Health (ICF), which offers a standardized
66 international terminology for defining and operationalizing function management across multiple
67 domains. Accurate and timely assessment of an individual's function is crucial in making clinical
68 decisions for treatment planning and rehabilitation(Finch, Higgins, Wood-Dauphinee, and Mayo,
69 2009). To achieve this, the assessment scale must include items that have high discrimination,
70 enabling the determination of functioning performance. The Item Response Theory (IRT) is a
71 well-established psychometric paradigm that measures the interaction between examinees and
72 test items to estimate an individual's ability level (represented by the latent trait, θ). Additionally,
73 the IRT model is a statistical technique that can provide theoretical and algorithmic support for
74 the developing clinical assessment tools in healthcare practice(Stucki, Rubinelli, and Bickenbach,
75 2020), which can transform raw ICF scores into the IRT-based interval scores. Recently,
76 emerging studies have advocated the use of IRT-based Rasch models to process ICF evaluation
77 data and generate the new version of the ICF core set that contains the most discriminable and
78 minimal items.

79 This review aims to offer an overview of the research paradigm regarding the ICF-based
80 functioning evaluation in combination with the IRT technique tools. Incorporating these two
81 techniques, a comprehensive and continuous assessment of human functioning with a minimal
82 core set can be constructed for patients undergoing rehabilitation. It also explores the challenges
83 of operationalizing and analyzing the ICF core set assessment, specifically in terms of
84 stratification and assignment scoring principles. This review emphasizes the value of IRT
85 modeling in ICF studies and further categorizes IRT-derived ICF outcome measures according to
86 the scoring principles. Moreover, the review proposes appropriate assignment principles that
87 match the difficult parameter in IRT modeling and discusses the benefits of using Wright maps to
88 visualize rehabilitation goals and potential functioning barriers. Overall, this review focuses on
89 introducing a technique-solid and framework-sound questionnaire construction approach to
90 develop the parsimonious core set of ICF and the Wright map that can enhance clinical decision-
91 making in rehabilitation settings.

92

93 **Materials and Methods**

94 **Review Criteria**

95 This review focused on studies that explore the psychometric properties of the IRT-derived
96 ICF evaluation tool. The search was carried out in January 31st 2023 based on the PubMed
97 database. The keywords used during the search included the following: [(Rasch) AND
98 (International Classification of Functioning, Disability and Health OR ICF), [(Item response
99 theory OR IRT) AND (International Classification of Functioning, Disability and Health OR
100 ICF)], and [(Mokken) AND (International Classification of Functioning, Disability and Health
101 OR ICF)]. The reference lists of the resulting articles were scanned to identify further relevant
102 studies, which totally resulted in 190 additional results. Note that in this paper, we adopted Rasch
103 as an umbrella term that includes Rasch and the 1-parameter logistic (1PL) model.

104 After removing duplicate results, the search resulted in 169 pertinent articles. Given the
105 search results, we employed the following inclusion and exclusion criteria to select relevant
106 works for detailed review.

107

108 **Eligibility Criteria**

109 The inclusion and exclusion criteria for studies are as follows:

110 1) *Study design*: There was no strict study design required in this review. Reviews regarding
111 the IRT-derived ICF evaluation tool were considered, languages: “English”, conference abstracts
112 without the full text, and unpublished studies were excluded;

113 2) *Participant characteristics*: Participants irrespective of gender, age, and health conditions
114 were evaluated via the ICF categories;

115 3) *Outcome measures*: The outcome measure or self-report questionnaire include ICF items
116 in body structure, body function, activity and participation, and environmental factors.

117

118 **Results**

119 Our approach resulted in the identification of 46 publications that met all predetermined
120 criteria ([Figure 1](#)). From each article, we extracted the information on several key variables,
121 including the specific ICF core set, the original number of ICF categories, the number of IRT-
122 based ICF categories, sample size, settings, response options, collapse strategy, scoring strategy,
123 and data analysis software, which were summarized in [Table 1](#), [Table 2](#), [Table 3](#) and [Table 4](#).

124 **Sufficiency and Efficiency in Clinical Evaluation: Challenges of** 125 **Operationalizing the ICF Questionnaire**

126 The ICF provides a comprehensive taxonomy of function and disability, organized into four
127 components, namely body structure, body functions, activity and participation, and
128 environmental and personal contextual factors. However, the ICF functional category bank
129 includes nearly 1,500 categories, making it impractical to assess each category in clinical
130 practice. To address this challenge, the derivatives of ICF known as the ICF core set have been
131 developed to evaluate functioning performance for specific health conditions or healthcare
132 settings (Selb et al, 2015). There are two versions of the ICF core set: the comprehensive core set
133 and the brief core set. A comprehensive ICF core set is inclined to exhaustively describe disease-
134 specific functions, while it overloads the burden of answering lengthy and multi-item
135 questionnaires on patients and clinicians. On the other hand, a brief ICF core set sacrifices the
136 sufficiency of the questionnaire to satisfy the minimal standard of assessing the most essential
137 functional categories. Despite the advantages of both core sets, two main issues are faced with

138 the operationalization of the ICF questionnaire, namely the complex scoring system and data
139 analysis in the ICF Likert scale.

140 Firstly, the ICF category in the core set requires scoring on a 0-4 Likert scale, which might
141 not be an efficient application in real clinical scenarios. There are controversies regarding how to
142 apply the 5-level Likert scale to measure the degree of “disability” regarding each ICF category,
143 namely, “no impairment” = 0, “mild impairment” = 1, “moderate impairment” = 2, “severe
144 impairment” = 3, and “complete impairment” = 4 (Rauch, Cieza, and Stucki, 2008; Jette, 2018).
145 Practitioners have not achieved full consensus on ICF category descriptions and differentiation
146 between scale points, especially the qualifiers 2 to 4. Although researchers have suggested the
147 employment of other assessment tools to assist in determining qualifier values, it significantly
148 increases the complexity of the assessment process after expanding one questionnaire into
149 multiple questionnaires. It will greatly expand the evaluation time due to a five-point rating scale
150 for each item in either comprehensive or brief core sets. Thorough scoring options and extensive
151 categories qualify the sufficiency of clinical assessment, but it is hard to satisfy the efficiency of
152 evaluation in rapid-paced and hectic clinical work.

153 Another prominent barrier is the data analysis of the ICF Likert scale. Firstly, the summated
154 scores represent the level of dysfunction depending on the uneven interval of the 5-point
155 qualifier scale, which diminishes the inter- and intra-individual comparability of outcome
156 measures. Secondly, each ICF category is considered as equal difficulty in the daily assessment
157 rather than hierarchically categorize in terms of parameters such as item difficulty and
158 discrimination. Without considering the quantitative values for item difficulties regarding certain
159 individual abilities, simple qualifier values of each item are insufficient to represent the whole
160 picture of clinical outcomes.

161 162 **Personal Ability and Item Difficulty: Principles and Values of the IRT-** 163 **derived Medical Outcome Measures**

164 Chang et al.(1995) proposed that the IRT could be a promising tool for measuring health
165 outcomes and quantifying the degree of health, which has been applied in the medical field since
166 the early 2000s (Hays, Morales, and Reise, 2000). The IRT model differentiates between an
167 individual’s ability levels and the difficulty levels of test items, constructing an Item Response
168 Function (IRF) by taking θ as an independent variable and the probability of correct answer rate

169 (P(θ)) as a dependent variable, under the condition that test items do not interact with each other.
170 For example, Liu et al.(2002) conducted a Rasch analysis on the Stroke Impairment Assessment
171 Set (SIAS). After reducing subsets and combining similar items, the IRT-derived SIAS is not
172 only transformed into a linear scale but also can reflect person abilities corresponding to item
173 difficulties.

174 In recent years, there has been increasing interest in the use of IRT-based Rasch models for
175 processing ICF evaluation data, which can be broadly categorized into two perspectives. The
176 first approach involves handling ICF categories as questionnaire items, which has been used in
177 various studies (Table 1, Table 2, Table 3, Table 4), such as the development of a 41-item
178 activity measurement scale for post-acute care by Haley et al.(2004). Cieza et al.(2009b) also
179 constructed the Rasch modeling based on the comprehensive assessment of the functional level
180 among patients with rheumatoid arthritis (AS) and osteoarthritis (OA). By investigating 437
181 patients with OA in Germany, Italy, Hungary, Serbia, and Singapore, researchers achieved the
182 first ICF-based cross-cultural clinical functional measurement. They found that the difficulty of
183 the same ICF category varied in different countries. The Rasch analysis also explored ICF
184 categories for various conditions or diseases, including AS(Cieza et al, 2009a), OA(Cieza et al,
185 2009b; Kurtaiş et al, 2011), chronic widespread pain(Prodinger et al, 2012), low back
186 pain(Bagraith, Strong, Meredith, and McPhail, 2017; Røe et al, 2009; Røe, Bautz-Holter, and
187 Cieza, 2013), breast cancer(Yang, Shin, Shin, and Lim, 2014), children and youth with
188 disabilities(Illum and Gradel, 2015), the ICF Generic Set(Ehrmann et al, 2018), the
189 Rehabilitation Set(Gao et al, 2020), spinal cord injury (SCI)(Li et al, 2018; Jia et al, 2020),
190 stroke(Feng, Jiang, Sun, and Lin, 2022), haemophilia(Feng et al, 2022), and cerebral palsy(Jiang
191 et al, 2023).

192 The second approach involves linking questionnaire items from scales to ICF items, which
193 has been applied in several studies(Finch, Higgins, Wood-Dauphinee, and Mayo, 2009; Maritz et
194 al, 2022; Nilsson, Westergren, Carlsson, and Hagell, 2010, 2010; Prodinger, O'Connor, Stucki,
195 and Tennant, 2017; Van de Winckel et al, 2019). Finch et al.(2009) discerned 44 items from
196 other scales and constructed a Rasch model based on the ICF framework to evaluate the level of
197 physical function in patients three months post-stroke. They obtained an item difficulty list
198 consisting of 52 items, and suggested that stroke rehabilitation goals and treatment plans could
199 be developed based on person abilities and item difficulties. This approach has been shown to be

200 useful for identifying functional limitations and developing interventions that target specific
201 areas of impairment. Other studies have used similar approaches to link scales to the ICF
202 framework for various health conditions, including Parkinson’s Disease(Nilsson, Westergren,
203 Carlsson, and Hagell, 2010), stroke(Van de Winckel et al, 2019), spinal cord injury(Maritz et al,
204 2022), demonstrating the versatility and applicability of this method in the assessment of
205 functional outcomes.

206

207 **Research Progress and Deficiency: Clarification of the ICF-based IRT** 208 **Scoring Principles**

209 ICF-based IRT studies employ two dimensions of scoring principles, stratification and
210 assignment. The stratification dimension involves sparse and dense stratification principles to
211 grade each ICF category. The sparse stratification principle (SSP) can involve scoring each ICF
212 category with qualifiers 0 for ‘no impairment/limitation’, 1 for ‘impairment/limitation’ instead of
213 adopting the original 5-level Likert scale(2005), or utilizing collapse strategies, item combination,
214 and item deletion during data analysis. Some studies suggest that reducing the five-point to
215 three-point qualifiers leads to a better model fit. For example, Li et al. (2018) found that reducing
216 the five-point (01234) to the three-point (01112) qualifiers improves the overall model fit. Gao et
217 al.(2020) also recommend using a collapse strategy and converting from 01234 to 01223 or
218 01112 in the Rasch model when operating the ICF five-point qualifiers as a rating scale. On the
219 other hand, the dense stratification principle (DSP) expands qualifiers to capture more detailed
220 clinical information and reduce the ceiling and floor effect, using an 11-point numeric rating
221 scale and transforming raw scores to the Rasch-based interval score(Ehrmann et al, 2018). The
222 SSP is broadly endorsed for the qualifiers in terms of enough items, such as 65 ICF categories of
223 SCI(Li et al, 2018). The DSP is more comprehensive but requires fewer items, such as the
224 Generic 6 set(Ehrmann et al, 2018) in [Table 2](#).

225 Considering assignment dimension, most ICF-based IRT studies have typically employed
226 the “descending assignment principle or assigning by dysfunction (ABD)” to address item scores
227 based on functional ability, where lower functionality corresponds to a higher score. Some
228 examples of studies that have employed this principle include Farin et al.(2007), Cieza et
229 al.(2009a), Røe et al.(2013), Ehrmann et al.(2018), Guo et al. (2020), and Gao et al.(2020).
230 However, a few studies have taken an alternative approach, using the “ascending assignment

231 principle or assigning by functioning (ABF)”, where a higher score corresponds to higher
232 functionality(Feng et al, 2022; Feng, Jiang, Sun, and Lin, 2022; Jiang et al, 2023). Okochi et
233 al.(2013) implemented the binary response options, scoring items as “yes” for functioning
234 (assignment = 1), or “no” for dysfunction (assignment = 0). According to the final scores, they
235 classified individuals into functional stages based on their activity level. The Lucerne ICF-based
236 Multidisciplinary Observation Scale (LIMOS) also adopted the ascending assignment principle
237 by generating 45 ICF-similar items based on the degree of assistance in daily life(Van de
238 Winckel et al, 2019).

239

240 **Discussion**

241 To achieve an optimal balance between sufficiency and efficiency in clinical assessment, it
242 is imperative to develop a concise and effective core set from the larger and more comprehensive
243 ICF core set. This core set should contain minimal yet discriminative items, capturing continuous
244 functioning changes, and ensuring good construct validity and internal consistency. The
245 implementation of the IRT statistical approach has been used to derive such a measurement tool,
246 which we refer to as the ICF parsimonious core set (PCS).

247 By overcoming the limitation of assuming that each functional item is equally difficult, the
248 IRT modeling approach has proven to be an effective way to extract PCSs. It has been applied
249 since the inception of the ICF and has demonstrated its effectiveness over the last two decades.
250 The IRT-derived outcome measure has the potential to serve as a pragmatic tool that provides
251 comprehensive and holistic information that can facilitate decision-making in clinical settings.

252

253 **Future Implication: the ICF Framework and IRT Technique**

254 As the most commonly used IRT model, the Rasch analysis can optimize clinical scales due
255 to its ability to enhance the construct validity of clinical scales and reduce the indiscriminate
256 items. Based on different configurations of parameters, the IRF can be presented as an S-shaped
257 Item Characteristic Curve (ICC, [Figure 2](#)) on the coordinate axis. This statistical model
258 overcomes the limitation of the classical test model, which assumes that each item is equally
259 difficult. The IRT recognizes that item difficulty is represented by parameters and maps the
260 locations of individuals and the difficulties of test items onto the same invariant scale, allowing
261 direct comparison.

262 The IRT-based measurement of individual ability has important implications for clinical
263 assessment in rehabilitation practice. It can help identify an individual's functional level
264 compared to others with the same disease, as well as assess whether a specific function has
265 improved following intervention. Additionally, the IRT-based estimation of item difficulty can
266 aid in prioritizing rehabilitation goals by determining which functional items are most
267 appropriate for an individual's abilities and which dysfunctions should be addressed first.

268 [Supplemental Table 1](#) displays four parameters within the IRT model, where the parameter
269 “difficulty (d)” is primarily reserved in parametric IRT models. As an illustration, the Rasch
270 model only estimates the “difficulty” parameter, and uses default values of $a = 1$ (moderate
271 discrimination), $g = 0$ (no guess), and $u = 1$ (completely carefulness). The parameters θ and d can
272 be expressed numerically on the same scale, namely logit or the log-odds (the logarithm of the
273 frequency ratio of scored and non-scored items). Consequently, individual functional status can
274 be sorted by person abilities, and different items can be hierarchized according to item
275 difficulties. Moreover, person abilities and item difficulties can be compared on the same scale,
276 allowing for the identification of whether an item is relatively difficult ($\theta < d$) or relatively easy
277 ($\theta > d$) for a person with a certain ability level (θ). The IRT model can also convert an ICF
278 Likert scale with ordinal response options into an equidistant scale(Prodingner et al, 2020).

279 The parameters, g and u , offer an opportunity to identify and address issues with the quality
280 of the ICF assessment scale. For instance, if a model reports an item with high guessing or
281 carelessness, the evaluators or clinical practices in the institution should audit certain functioning
282 evaluation processes to improve the assessment quality.

283

284 **The Scoring Principle: Sparse Scoring principle and Assigning by Functioning**

285 The SSP and DSP principles are useful in different contexts, with the SSP being more
286 efficient and the DSP being more comprehensive, as shown in [Figure 3](#). From an efficiency
287 perspective, the stratification principle of qualifiers aligns with the SSP, such as screening tools
288 ([Table 3](#)). In contrast, research tools with a large number of items and DSP can capture
289 functional limitations comprehensively. To strike a balance between sufficiency and efficiency, it
290 is recommended to adopt more items in ICF core sets, while using SSP for detailed assessments
291 ([Table 4](#)). A recent study by Feng et al.(2022) extracted 118 categories of “b” and “d”
292 components from the ICF categories for stroke and used a dichotomous qualifying strategy

293 (dysfunction = 0 or functioning = 1) for IRT modeling. The final Rasch-derived ICF scale
294 consisted of 45 items. This approach enabled simplified assessment of comprehensive ability
295 levels and allowed for scoring of functional categories based on ICF qualifier scores. By utilizing
296 a binary response pattern, the approach may further enhance efficiency.

297 In terms of assignment scoring principles, the ABD principle and the ABF principle have no
298 difference in assessing an individual's functional capacity. A reduction in dysfunction or an
299 improvement in functional level is typically considered an improvement result. However,
300 differences between the two principles become evident in IRT analyses and guiding treatment
301 plans. The logit value of item difficulty is the primary contradiction of evaluation and treatment
302 in rehabilitation services. The ABD principle presents a semantic dilemma in clinical thinking as
303 it indicates how difficult it is to disable a certain function. On the other hand, the ABF principle
304 aligns with the clinical treatment thinking process of improving functional capacity, and only the
305 logit value of an individual's potential traits derived from the principle of ABF would reflect the
306 level of individual function. As assessment is in the service of treatment, researchers have
307 recommended adopting the ABF principle in future ICF studies(Feng, Jiang, Sun, and Lin, 2022;
308 Feng et al, 2022).

309

310 **The IRT-derived ICF Parsimonious Core Set and Wright Map**

311 The ICF-based IRT questionnaire, constructed with the principles of sufficiency and
312 efficiency, uses a binary response pattern to increase efficiency. ABF scoring aligns with
313 difficulty parameter in the IRT modeling, and the resulting IRT-derived PCS has high internal
314 consistency and construct validity. This ICF PCS contains the most discriminable and minimal
315 items, making it ideal for a clinical-pragmatic approach. Empirical evidence from PCS
316 measurement tools, such as the stroke simplified IRT assessment tool(Feng, Jiang, Sun, and Lin,
317 2022), IRT-derived and ICF-based haemophilic activity and participation assessment tool(Feng
318 et al, 2022), and ICF-based IRT assessment tool(Jiang et al, 2023) for children with cerebral
319 palsy, demonstrates its efficacy in assessing functional stages.

320 A novel approach to prioritizing daily activities intervention can be achieved by
321 understanding the difficulty levels of diverse functional activities. The Wright map, which plots
322 person abilities and item difficulties with the same unit (i.e., logit), orders the level of functional
323 ability of participants on the left side and the difficulty of functional categories on the right side

324 (Figure 4 © 2022 John Wiley & Sons Ltd.). The most difficult functional categories are at the
325 top of the scale, while the easiest ones are at the bottom. Subjects with low functional ability
326 struggle with even the easiest activities, while those with higher functional ability experience no
327 difficulty performing body functions or activities.

328 Recent advancements in technology and data analytics enable the use of IRT-derived PCS
329 functioning outcome measures to personalize interventions that improve functioning and quality
330 of life. The Wright map provides evidence for customized rehabilitation goals by integrating
331 person abilities and item difficulties. Item difficulties can be used to estimate rehabilitation goals,
332 such as the level of difficulty in improving a specific function from a low stage to a high stage.
333 Additionally, several functioning properties can provide more proxy goals for rehabilitation
334 therapy.

335

336 **Conclusion**

337 The use of the IRT technique in ICF studies has highlighted a number of contradictions,
338 such as the trade-offs between efficiency and adequacy, comprehensiveness and precision,
339 assessment and treatment, and disability and health. In the context of humanism in healthcare,
340 the fundamental goal of function evaluation and treatment intervention must be to optimize
341 patient outcomes. To achieve this, we can consider the following suggestions through
342 comparative analysis: (1) Streamline the stratification of qualifiers to achieve efficient evaluation,
343 while increasing the number of scale items to ensure comprehensive assessment; (2) Adhere to
344 the principle of “assigning by functioning” when scoring items to promote clinical-relevant and
345 intuitive thinking in tailoring treatment plans; (3) Adoption of the Wright map (person-item
346 threshold map) might facilitate patient-oriented care according to person abilities and item
347 difficulties.

348

349 **Acknowledgment**

350 Not applicable.

351 **Declaration of Interest Statement**

352 The authors declare that the research was conducted in the absence of any commercial or
353 financial relationships that could be construed as a potential conflict of interest.

354

355 **Ethical Approval and Consent to Participate**

356 Not applicable.

357 **Data Availability Statement**

358 Data sharing not applicable to this article as no datasets were generated or analyzed during the
359 current study.

360 **Funding**

361 This research is sponsored by Shanghai Sailing Program (23YF1433700).

362 **Author contribution**

363 CF and LF conceived the original idea. CF and LF designed and performed the review and wrote
364 the manuscript. LF offered essential idea to the review design. CF and LF edited the manuscript.
365 SGL gave comments on the review design. All authors discussed the conclusions and
366 commented on the manuscript.

367

368 **Reference**

369 Bagraith KS, Strong J, Meredith PJ, McPhail SM 2017 Rasch Analysis Supported the Construct Validity of
370 Self-Report Measures of Activity and Participation Derived from Patient Ratings of the ICF Low Back Pain
371 Core Set. *Journal of Clinical Epidemiology* 84: 161–172.

372 Chang WC, Chan C 1995 Rasch Analysis for Outcomes Measures: Some Methodological Considerations.
373 *Archives of Physical Medicine and Rehabilitation* 76: 934–939.

374 Cieza A, Hilfiker R, Boonen A, Heijde D van der, Braun J, Stucki G 2009a Towards an ICF-Based Clinical
375 Measure of Functioning in People with Ankylosing Spondylitis: A Methodological Exploration. *Disability
376 and Rehabilitation* 31: 528–537.

377 Cieza A, Hilfiker R, Chatterji S, Kostanjsek N, Ustün BT, Stucki G 2009b The International Classification of
378 Functioning, Disability, and Health Could Be Used to Measure Functioning. *Journal of Clinical
379 Epidemiology* 62: 899–911.

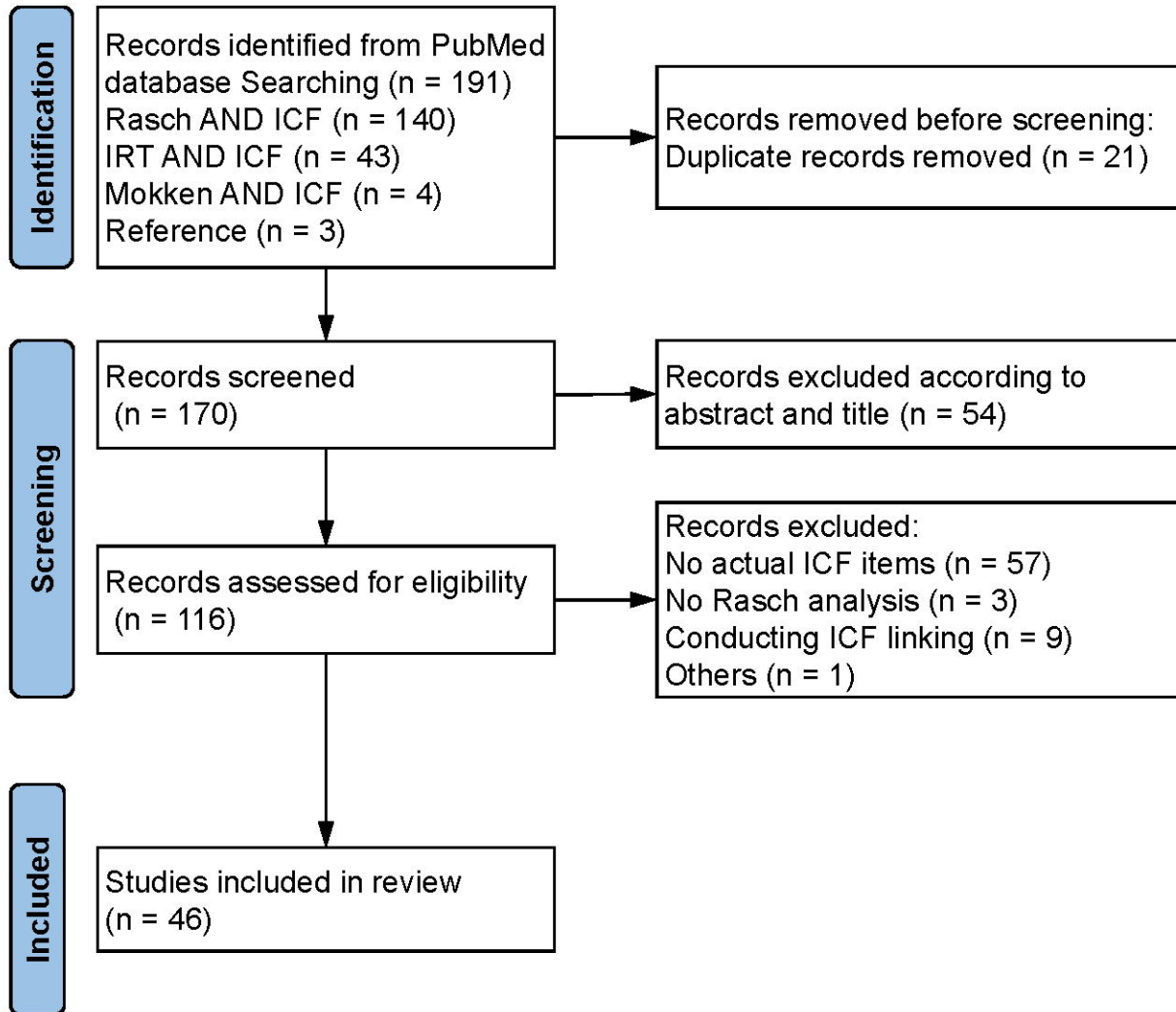
380 Ehrmann C, Prodinger B, Stucki G, Cai W, Zhang X, Liu S, Liu S, Li J, Reinhardt JD 2018 ICF Generic Set as
381 New Standard for the System Wide Assessment of Functioning in China: A Multicentre Prospective Study
382 on Metric Properties and Responsiveness Applying Item Response Theory. *BMJ Open* 8: e021696.

383 Farin E, Fleitz A, Frey C 2007 Psychometric Properties of an International Classification of Functioning,
384 Disability and Health (ICF)-Oriented, Adaptive Questionnaire for the Assessment of Mobility, Self-Care
385 and Domestic Life. *Journal of Rehabilitation Medicine* 39: 537–546.

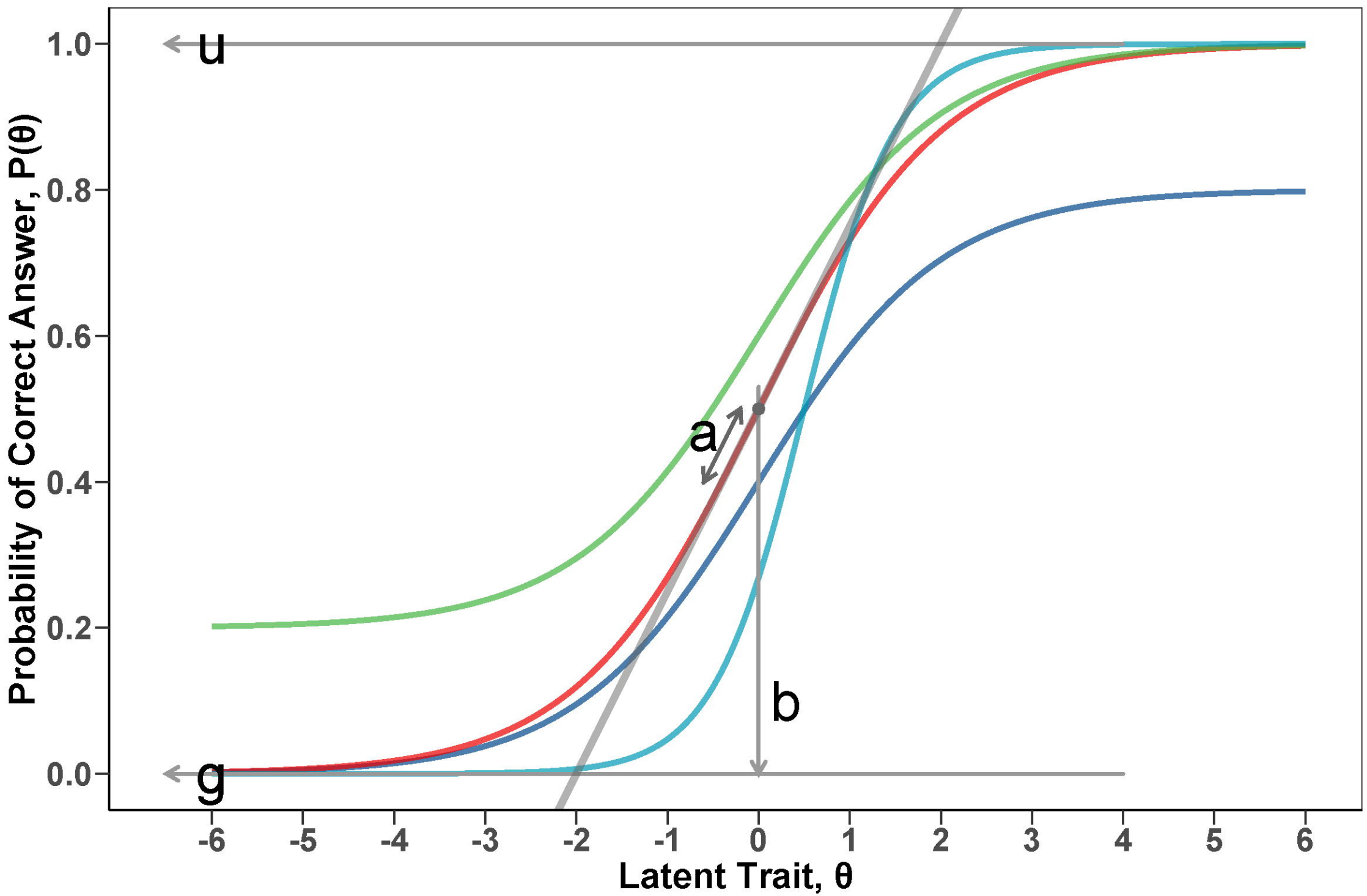
- 386 Feng C, Geng B-F, Liu S-G, Jiang Z-L, Lin F 2022 Activity and participation in haemophiliacs: Item response
387 modelling based on international classification of functioning, disability and health. *Haemophilia*.
- 388 Feng C, Jiang Z-L, Sun M-X, Lin F 2022 Simplified Post-stroke Functioning Assessment Based on ICF via
389 Dichotomous Mokken Scale Analysis and Rasch Modeling. *Frontiers in Neurology* 13: 827247.
- 390 Finch LE, Higgins J, Wood-Dauphinee SL, Mayo NE 2009 A Measure of Physical Functioning to Define
391 Stroke Recovery at 3 Months: Preliminary Results. *Archives of Physical Medicine and Rehabilitation* 90:
392 1584–1595.
- 393 Gao Y, Yan T, You L, Li K, Zhang L, Zhang M 2020 Psychometric Properties of the International
394 Classification of Functioning, Disability and Health Rehabilitation Set: A Rasch Analysis. *International*
395 *Journal of Rehabilitation Research* 44: 144–151.
- 396 Grill E, Stucki G, Boldt C, Joisten S, Swoboda W 2005 Identification of relevant ICF categories by geriatric
397 patients in an early post-acute rehabilitation facility. *Disability and Rehabilitation* 27: 467–473.
- 398 Guo C, Liu Y, Hao S, Xie L, Xiang G, Wu Y, Li S 2020 The Reliability and Validity of the “Activity and
399 Participation” Component in the Brief ICF Core Set for Chronic Obstructive Pulmonary Diseases Based on
400 Rasch Analysis. *International Journal of Chronic Obstructive Pulmonary Disease* 15: 1191–1198.
- 401 Haley SM, Coster WJ, Andres PL, Ludlow LH, Ni P, Bond TL, Sinclair SJ, Jette AM 2004 Activity outcome
402 measurement for postacute care. *Medical Care* 42: 149–161.
- 403 Hays RD, Morales LS, Reise SP 2000 Item Response Theory and Health Outcomes Measurement in the
404 21st Century. *Medical Care* 38: 1128–1142.
- 405 Illum NO, Gradel KO 2015 Assessing Children With Disabilities Using WHO International Classification of
406 Functioning, Disability and Health Child and Youth Version Activities and Participation D Codes. *Child*
407 *Neurology Open* 2: 2329048X15613529.
- 408 Jette AM 2018 The Utility of and Need for Improving the ICF. *Physical Therapy* 98: 629–630.
- 409 Jia M, Tang J, Xie S, He X, Wang Y, Liu T, Yan T, Li K 2020 Using a Mobile App-Based International
410 Classification of Functioning, Disability, and Health Set to Assess the Functioning of Spinal Cord Injury
411 Patients: Rasch Analysis. *JMIR MHealth and UHealth* 8: e20723.
- 412 Jiang Y-E, Zhang D-M, Jiang Z-L, Tao X-J, Dai M-J, Lin F 2023 ICF-Based simple scale for children with
413 cerebral palsy: Application of Mokken scale analysis and Rasch modeling. *Developmental*
414 *Neurorehabilitation* 1–18.
- 415 Kurtaiş Y, Öztuna D, Küçükdeveci AA, Kutlay Ş, Hafiz M, Tennant A 2011 Reliability, construct validity and
416 measurement potential of the ICF comprehensive core set for osteoarthritis. *BMC Musculoskeletal*
417 *Disorders* 12: 255.
- 418 Li K, Yan T, You L, Xie S, Li Y, Tang J, Wang Y, Gao Y 2018 Psychometric Properties of the International
419 Classification of Functioning, Disability and Health Set for Spinal Cord Injury Nursing Based on Rasch
420 Analysis. *Disability and Rehabilitation* 40: 338–345.

- 421 Liu M, Chino N, Tuji T, Masakado Y, Hase K, Kimura A 2002 Psychometric Properties of the Stroke
422 Impairment Assessment Set (SIAS). *Neurorehabilitation and Neural Repair* 16: 339–351.
- 423 Maritz R, Fellinghauer C, Brach M, Curt A, Gmünder HP, HOPFE M, HUND-GEORGIADIS M, JORDAN X,
424 SCHEEL-SAILER A, STUCKI G 2022 A Rasch-Based Comparison of the Functional Independence Measure
425 and Spinal Cord Independence Measure for Outcome and Quality in the Rehabilitation of Persons with
426 Spinal Cord Injury. *Journal of Rehabilitation Medicine* 54: 82.
- 427 Nilsson MH, Westergren A, Carlsson G, Hagell P 2010 Uncovering Indicators of the International
428 Classification of Functioning, Disability, and Health from the 39-Item Parkinson’s Disease Questionnaire.
429 *Parkinson’s Disease* 2010: 984673.
- 430 Okochi J, Takahashi T, Takamuku K, Escorpizo R 2013 Staging of Mobility, Transfer and Walking Functions
431 of Elderly Persons Based on the Codes of the International Classification of Functioning, Disability and
432 Health. *BMC Geriatrics* 13: 16.
- 433 Prodinge B, Küçükdeveci AA, Kutlay S, Elhan AH, Kreiner S, Tennant A 2020 Cross-Diagnostic Scale-
434 Banking Using Rasch Analysis: Developing a Common Reference Metric for Generic and Health
435 Condition-Specific Scales in People with Rheumatoid Arthritis and Stroke. *Journal of Rehabilitation
436 Medicine* 52: jrm00107.
- 437 Prodinge B, O’Connor RJ, Stucki G, Tennant A 2017 Establishing Score Equivalence of the Functional
438 Independence Measure Motor Scale and the Barthel Index, Utilising the International Classification of
439 Functioning, Disability and Health and Rasch Measurement Theory. *Journal of Rehabilitation Medicine*
440 49: 416–422.
- 441 Prodinge B, Salzberger T, Stucki G, Stamm T, Cieza A 2012 Measuring Functioning in People with
442 Fibromyalgia (FM) Based on the International Classification of Functioning, Disability and Health (ICF)--a
443 Psychometric Analysis. *Pain Practice: The Official Journal of World Institute of Pain* 12: 255–265.
- 444 Rauch A, Cieza A, Stucki G 2008 How to apply the International Classification of Functioning, Disability
445 and Health (ICF) for rehabilitation management in clinical practice. *Eur J Phys Rehabil Med* 44: 329–42.
- 446 Røe C, Bautz-Holter E, Cieza A 2013 Low Back Pain in 17 Countries, a Rasch Analysis of the ICF Core Set
447 for Low Back Pain. *International Journal of Rehabilitation Research Internationale Zeitschrift Fur
448 Rehabilitationsforschung Revue Internationale De Recherches De Readaptation* 36: 38–47.
- 449 Røe C, Sveen U, Geyh S, Cieza A, Bautz-Holter E 2009 Construct Dimensionality and Properties of the
450 Categories in the ICF Core Set for Low Back Pain. *Journal of Rehabilitation Medicine* 41: 429–437.
- 451 Rolstad S, Adler J, Rydén A 2011 Response Burden and Questionnaire Length: Is Shorter Better? A
452 Review and Meta-analysis. *Value in Health* 14: 1101–1108.
- 453 Selb M, Escorpizo R, Kostanjsek N, Stucki G, Üstün B, Cieza A 2015 A Guide on How to Develop an
454 International Classification of Functioning, Disability and Health Core Set. *European Journal of Physical
455 and Rehabilitation Medicine* 51: 105–117.
- 456 Stucki G, Rubinelli S, Bickenbach J 2020 We Need an Operationalisation, Not a Definition of Health.
457 *Disability and Rehabilitation* 42: 442–444.

- 458 Van de Winckel A, Ottiger B, Bohlhalter S, Nyffeler T, Vanbellinghen T 2019 Comprehensive ADL Outcome
459 Measurement after Stroke: Rasch Validation of the Lucerne ICF-Based Multidisciplinary Observation
460 Scale (LIMOS). Archives of Physical Medicine and Rehabilitation 100: 2314–2323.
- 461 Yang EJ, Shin E-K, Shin H-I, Lim J-Y 2014 Psychometric Properties of Scale Constructed from the
462 International Classification of Functioning, Disability and Health (ICF) Core Set for Breast Cancer Based
463 on Rasch Analysis. Supportive Care in Cancer: Official Journal of the Multinational Association of
464 Supportive Care in Cancer 22: 2839–2849.



— $a=1, b=0, g=0, u=0.8$ — $a=1, b=0, g=0, u=1$ — $a=1, b=0, g=0.2, u=1$ — $a=2, b=0.5, g=0, u=1$



ABD, e.g., 0 = functioning, 1=dysfunction

Assignment Principles

θ of individual (latent ability/location):
Disability level

ABF, e.g., 0 = dysfunction, 1 = functioning

b of item (item difficulty):
How hard is it to disable the function?

Assessment

θ of individual (latent ability/location):
Functioning level

&

Treatment

b of item (item difficulty):
How hard is it to improve the functional capacity?
To what extent is this function treatable?

Efficiency

With few qualifiers
e.g., 0 & 1, totally 2 levels
Simplest for each item
Less information
High efficiency

SSP

Screening Tools
SSP
& Less items

Clinical A&T
SSP
& More items

With more qualifiers
e.g., 0-10, totally 11 levels
Difficult for grading
Sufficient information
Low efficiency

DSP

X

Research Tools
DSP
& More Items

Low

Less items

High

More Items

Sufficiency

Abbreviations:

ABF: Assigning by Functioning

ABD: Assigning by Dysfunction

A&T: Assessment and Treatment

SSP: Sparse Stratification Principle

DSP: Dense Stratification Principle

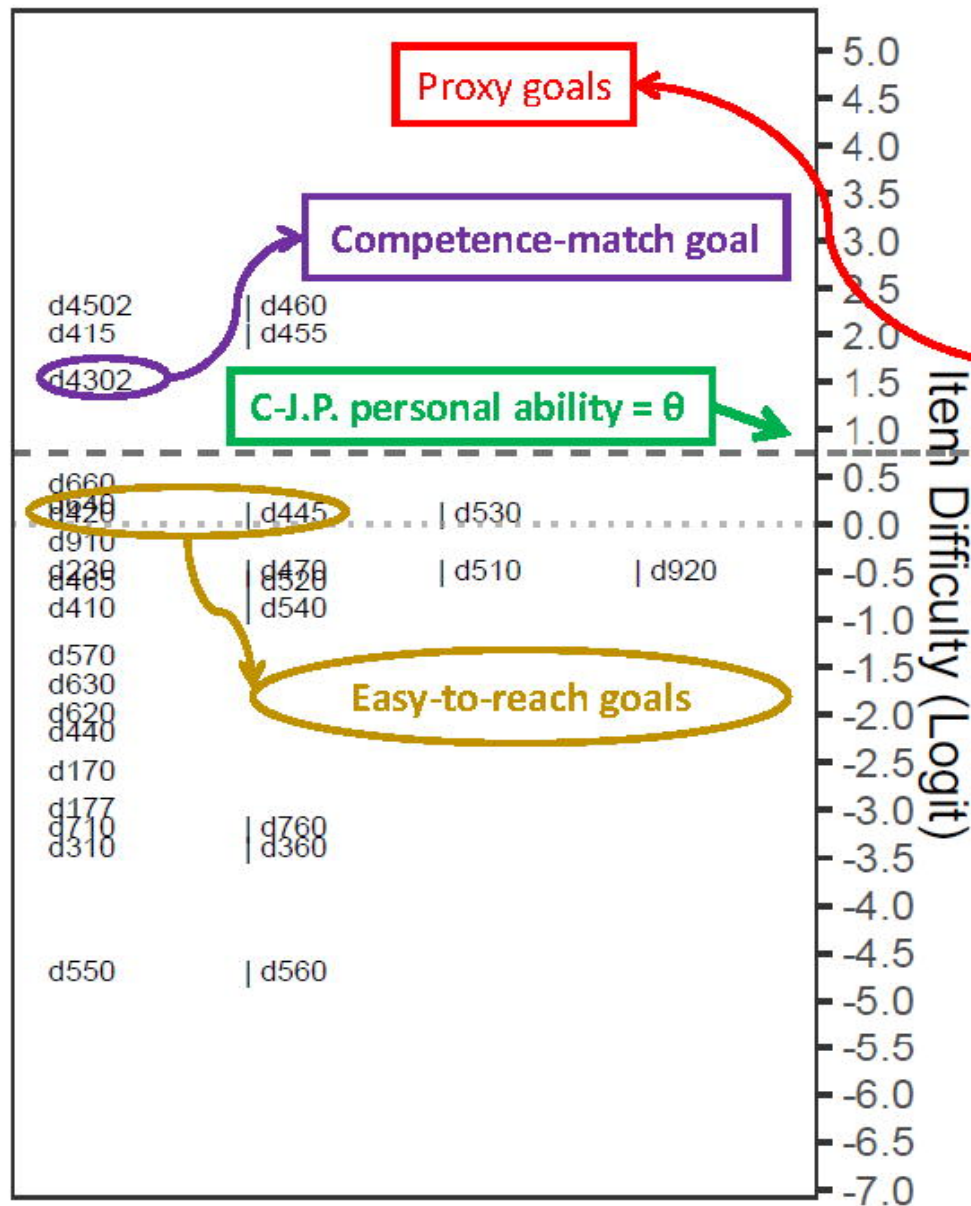
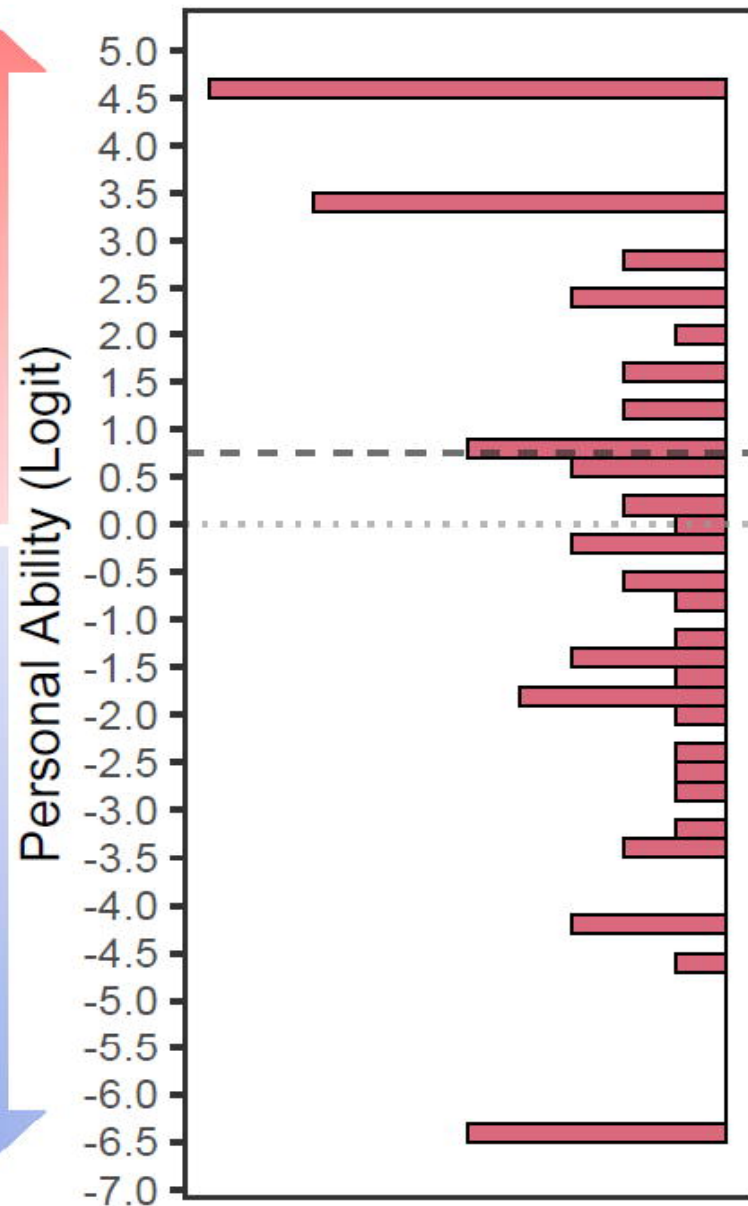
To reduce item number.
Minimum is only 1 item.
The latent ability is defined by a few items or even single item.
Most specific.
Information richness depends on stratification principles.

To expand item number.
Maximum can exceed 150 items, e.g., stroke core set with 166 items.
The latent ability is measured by an integrated set of items.
Most comprehensive.
Time consumption depends on stratification principles.

Stratification Principles

Specific Person: C-J.P.
 Dashed Line: $\theta = 0.7564$ logit.
 Dotted Line: reference line of 0 logit.

Dysfunction items



The higher, the more difficult

C-J.P. Dysfunction Lists

d4502 | d460
 d415 | d455
 d4302
 d640
 d420 | d445

The lower, the easier

- Reference:
- d4502 Walking on different surfaces
 - d460 Moving around in different locations
 - d415 Maintaining a body position
 - d455 Moving around
 - d4302 Carrying in the arms
 - d640 Doing housework
 - d420 Transferring oneself
 - d445 Hand and arm use

Figure legend

Figure 1. The Flow diagram of study selection.

Figure 2. Schematic diagram of the ICC curve.

The illustration of the effect of four parameter were respectively depicted on the shape of the ICC (a = discrimination/slope, b = difficulty/location/threshold, g = pseudo-guessing/lower asymptote, u = upper asymptote/slipping/pseudo-carelessness)

Figure 3. Exploration of efficiency and sufficiency in the research paradigm of the ICF-based evaluation tools.

Figure 4. Personal ability and item difficulty constructed in the Wright map. The left column illustrates the personal ability corresponding to the item difficulty in the right column © 2022 John Wiley & Sons Ltd.

Table 1. Dense Stratification Principle and more items based on the ICF-based IRT modeling

Author, year	ICF Items	Candidate ICF Item	Item bank	Sample Size	Setting	Response Options	Collapse Strategy	Assigning Principle (A or D)	Software
Haley et al., 2004	ICF Activity domains	58	41	477	Post-acute care	Not at all/a little/somewhat/a lot/complete, can't do	NA	A	WINSTEPS
Farin et al., 2007	"d5 self-care", "d6 domestic life", "d4 mobility", MOSES Patient Questionnaire	95	58	1019	Musculoskeletal disease, cardiac disease, neurology rehabilitation	No difficulty/mild difficulty/moderate difficulty/severe difficulty/impossible	NA	A	WINSTEPS
Farin et al., 2009	MOSES Physician Questionnaire	58 items of Moses-Patient	47	1019	Musculoskeletal disease, cardiac disease, neurology rehabilitation	No difficulty/mild difficulty/moderate difficulty/severe difficulty/impossible	NA	A	WINSTEPS
Ballert et al., 2014	ICF Core Sets for SCI	264	Not Reported	1048	Early post-acute and long-term	ICF generic qualifier scale	NA	D	R software eRm package
Tutelyan et al., 2014	HF-ICF-60 measure	60	60	500	Disability and healthy respondents	No problem (0) to extreme problem (5)	NA	D	RUMM2030
Illum et al., 2015	ICF-CY	57	39	332	Children with disabilities	ICF-CY generic qualifier scale	NA	D	Andrich Rating-Scale Model; Winsteps 3.74.0

SCI: Spinal cord injury

MOSES: Mobility, Self-Care, and Domestic Life

ICF-CY: International Classification of Functioning, Disability, and Health for Children and Youth core set

RUMM: Rasch Unidimensional Measurement Model

Assigning Principle:

A: Less assistance with higher scores (Ascending Scoring Principle);

D: Descending Scoring Principle (Higher disability with higher scores)

ICF generic qualifier scale:

0 = 0–4% No problem,

1 = 5–24% Mild problem,

2 = 25–49% Moderate problem,

3 = 50–95% Severe problem,

4 = 96–100% Complete problem

8 = Not specified (used when there is insufficient information to quantify the extent of the problem)

9 = Not applicable (used to indicate when a category does not apply to a particular person)

ICF-CY generic qualifier scale:

1 = Child's ability is as expected for his or her age.

2 = Child has difficulties, but functioning is still in the expected range for his or her age.

3 = Child needs help from another person with functions, activities, and participation.

4 = Child needs help and care; the child has only limited ability with respect to body functions, activities, and participation.

5 = Child is totally dependent on others for body functions, activities, and participation.

Table 2. Dense Stratification Principle and Less Items based on the ICF-based IRT Modeling

Author, year	ICF Items	Candidate ICF Item	Item bank	Sample Size	Setting	Response Options	Collapse Strategy	Assigning Principle (A or D)	Software
Prodinge et al., 2012	Brief ICF Core Set for fibromyalgia	19	14	256	Patients with Fibromyalgia	ICF generic qualifier scale	01123	D	RUMM 2020
Reinhardt et al., 2016	ICF Generic Set categories	7	7	761	Musculoskeletal, neurological health condition, SCI and TBI		NA	D	RUMM 2030
Illum and Gradel, 2017	ICF-CY	26	26	162	Parents of children with disability		NA	D	Winsteps 3.74.0
Illum et al., 2018	ICF-CY	26	26	162	Parents of children with disability		NA	D	Winsteps 3.74.0
Reinhardt et al., 2016	ICF Generic Set categories	7	7	761	Musculoskeletal, Neurological health condition		NA	D	RUMM 2030
Guo et al., 2020	Activity and participation categories in the brief ICF core set for COPD	4	4	103	Patients with COPD		NA	D	FACETS (Minifac) 3.67.0
Alviar et al., 2012	ICF core set for OA	32	25	316	Adults with OA who were either waiting for hip or knee arthroplasty (pre-operative), or within one year of		01123	D	RUMM 2030

					surgery (post-operative)				
Jia et al., 2020	SCI-related ICF items	31	28	112	Families of SCI patients		NA	D	RUMM 2030
Misajon et al., 2008	Self-care, Mobility, Participation, Relationships, and Psychological Well-being	23	23	210	Mobility impairment	Impact scale: “no impact” to “extreme impact”, or Distress scale: “no distress” to “extreme distress”; 6-point scale (012345)	011223	D	RUMM 2020
Pallant et al., 2009		23	22	259	Patients with OA		NA	D	RUMM 2020
Prodinger et al., 2019	Mobility, Selfcare, Domestic life	25	Not reported	2927	Stroke, RA, OA	No problem to complete problem; 5-point scale (0-4)	NA	D	RUMM 2030
Janeslätt et al., 2015	Time-Self rating ICF environmental factors	5	5	64	Adults with/without mental disability	Not at all/a little/a lot/very much (1234)	NA	A	Winsteps
Prodinger et al., 2016	SCIM-SR	17	17	1549	Traumatic or nontraumatic SCI	Total assistance/partial assistance/independent but need adaptive devices or specific equipment/ independent and do not need adaptive devices or specific equipment	NA	A	RUMM 2030

SCIM-SR: Spinal Cord Independence Measure – Self Report; TBI: traumatic brain injury ; COPD: chronic obstructive pulmonary disease; RA: rheumatoid arthritis; OA: Osteoarthritis

Table 3. Sparse Stratification Principle and Less Items based on the ICF-based IRT Modeling

Author, year	ICF Items	Candidate ICF Item	Item bank	Sample Size	Setting	Response Options	Collapse Strategy	Assigning Principle (A or D)	Software
Ostir et al. 2006	Home and Community Participation	20	20	594	Inpatient rehabilitation	Performance: “0 – Did not participate in this life situation”; “1 – Participated monthly (once every 3–4 weeks)”; “2 – Participated bi-weekly (once every 2 weeks)”; “3 – Participated weekly (1–4 days per week)”; and “4 – Participated daily/almost daily (5 or more days per week)”	3-level scoring system	A	WINSTEPS
Pallant et al., 2006	Mental functions, self-care, domestic life, mobility, interpersonal interactions & relationships, major life areas, community, social, and civic life	23	23	169	Mobility impairment	Impact scale: “no impact” to “extreme impact”, or Distress scale: “no distress” to “extreme distress”; 6-point scale (012345)	011112	D	RUMM2020
Mukaiino et al., 2020	home environment checklist	10	10	244	Individuals with disability	Complete mismatch between the individual’s ability and home environment/Partial mismatch between the individual’s ability and home environment/No mismatch between the individual’s ability and home environment/no need for a specific environment; 3-point scale (210)	NA	A	RUMM2030
Bagraith et al., 2017	LBP Core Set Self-Report Checklist	33	27	308	Adults with chronic LBP	No difficulty/some difficulty/severe difficulty, could not do; 3-point scale (012)	NA	D	Winsteps (v3.92) with JMLE

Author, year	ICF Items	Candidate ICF Item	Item bank	Sample Size	Setting	Response Options	Collapse Strategy	Assigning Principle (A or D)	Software
Gao et al., 2021	Rehabilitation Set	28	28	515	Subacute or chronic phase recovery	ICF generic qualifier scale	01223 or 01112	D	RUMM2030
Luo et al., 2022	ICF-CY with CHD 1 year after surgery	23	22	340	Children aged 3–6 years after CHD surgery		01122	A	Version 3.72.3 WINSEPS
Cieza et al., 2009	ICF Core Set of OA	38	74 country-specific and 7 common ICF categories	437	OA from Germany, Italy, Hungary, Serbia, and Singapore		01233, 01222, 01123, 01111, 01122, 01123, 01233, 01222, 01112,	D	RUMM 2020
Rauch et al., 2009	Comprehensive ICF Core Sets for RA and AS	24	Not Reported	230	Not Reported		environment: 1 and -1	D	Conquest software
Algurén et al., 2011	Brief ICF Core Set for Stroke	15	20 categories (6 country-common)	757	China, Germany, Italy, and Sweden		01122	D	RUMM2030
Yang et al., 2014	Brief ICF core set for breast cancer	43	30	158	Routinely referred after breast cancer treatment		Environmental factors: 012	D	WINSTEPS
Bostan et al., 2014	Activities and Participation ICF categories	22	Capacity scale: 16 items; Performance scale: 18 items; Environmental: 13 items	296	CWP, LBP, OA, OP and RA		01122, environmental factors: -1 0 +1	D	RUMM2030
Okochi et al., 2013	Basic mobility, walking	33	8	4424	Persons over 65 years old		Person performed the task (1) or not (0).	NA	A

LBP: low back pain; CHD: congenital heart disease

Table 4 Sparse Stratification Principle and More Items based on the ICF-based IRT Modeling

Author, year	ICF Items	Candidate ICF Item	Item bank	Sample Size	Setting	Response Options	Collapse Strategy	Assigning Principle (A or D)	Software
Uhlig et al., 2007	Comprehensive ICF Core Set for rheumatoid arthritis	78	73	25	Outpatient RA	ICF generic qualifier scale	Qualifiers: 5 to 3, environmental factors: 9 to 3	D	RUMM 2020
Cieza et al., 2009	AS related ICF item	68	64	111	Not Reported	ICF generic qualifier scale	01233	D	RUMM2020
Røe et al., 2009	Comprehensive ICF Core Set for LBP	78	59	118	Acute, subacute or chronic	ICF generic qualifier scale	01123, 00123; environmental factors: 012	D	RUMM 2020
Hilfiker et al., 2009	Comprehensive ICF Core Set for LBP	78	Not Reported	61	Not Reported	ICF generic qualifier scale or modified qualifier scale (0,1,8,9 or -1,0,1,8,9)	01222, 01111, 01122, -1-1-1-101122	D	RUMM 2020
Kurtaiş et al., 2011	Comprehensive ICF core set for OA	55	44	100	Outpatient knee and/or hip OA	ICF generic qualifier scale	01122, 01111	D	RUMM2030
Bostan et al., 2012	LBP, breast cancer, and stroke ICF items	194	Not Reported	361	Patients with LBP, breast cancer, and stroke	ICF generic qualifier scale	01122	D	RUMM2020
Røe et al., 2013	Comprehensive ICF Core Set for LBP	78	Not Reported	972	Acute, subacute or chronic	ICF generic qualifier scale	00001,01122,00012,01122,00123,0122,00011,01233,01223	D	RUMM2020
Ballert et al., 2014	Brief ICF Core Sets for Spinal Cord	64	Not Reported	1048	Early post-acute and long-term health care	No problem or not applicable (0), problem (1)	NA	D	ConQuest and R
Kiltz et al., 2015	Comprehensive ICF Core Set for AS	251	17	1752	Not Reported	Dichotomous item	NA	D	RUMM 2030
Li et al., 2016	ICF-SCIN	65	61	140	Not Reported	ICF generic qualifier scale	01112	D	RUMM 2030

Müller M, et al., 2016	PaArticular Scales	105	24 items for the Activities scale and 11 items for the Participation scale	191	Individuals in geriatric care	0 for “no problem”, 1 for “mild to moderate problem”, 2 for “severe problem” and 3 for “complete problem”	0112, 0111	D	R 2.15.2
Feng et al., 2022	118 stroke-related ICF items (59 items in b and d domains)	118	45	130	Inpatient stroke	ICF generic qualifier scale; Functioning/disability (1/0)	11000	A	R (mokken and ltm package)
Feng et al., 2022	31 haemophilic-specific activity and participation ICF items	112	31	70	Haemophiliacs		10000	A	R 4.1.0 (mokken and mirt package)
Jiang et al., 2023	34 items in b domain and 58 items in d domain	92	67	150	Children with CP		10000	A	R (mokken and ltm package)
Oude Voshaar et al., 2017	ICF domestic life chapter	72	41	1128	Outpatient populations	3-point	NA	A	Mokken R-package, Rasch model (PCM)

RA: rheumatoid arthritis; ICF-SCIN: ICF set for spinal cord injury nursing; CP: cerebral palsy