Index

acceptability 59–60
agreement 97, see measurement error and reliability
absolute agreement 105
chance agreement, see expected agreement
expected agreement 116–117
level of agreement 206–207
observed agreement 116
analysis of variance (ANOVA) 104
random factors 104–105
fixed factors 105
anchor 246, 248, 257
clinician-based 256
patient-based 256
Apgar score 15
application
in clinical practice 45, 144, 160, 164, 169, 245
in research 45, 144, 245
area under the curve, see receiver operating characteristic (ROC) curve
Bland and Altman method 113, 163, 166, 243
Bland and Altman plot 113–114, 163, 167
interpretation of 122–123
limits of agreement 113–114, 122–123, 167, 217, 242, 260
blood pressure 122, 128, 131–137, 239
causal indicator 14
ceiling effect, see floor and ceiling effects
change 241
amount of change 257
change beyond measurement error 123
clinically important change 204
clinically relevant change 123, 245
global rating of change 214, 256
important change 204, 257
magnitude of change scores 216
minimal detectable change 243,
see smallest detectable change
minimal important change (MIC) 123, 217–218, 245, 299–300
anchor-based method 246–247, 258
distribution-based method 246–247, 258
for deterioration 258
for improvement 258
mean change method 246
ROC method 246, 248–249, 254
visual anchor-based MIC distribution 247, 249, 252–253, 255
minimal real change 243, see smallest detectable change
real change 123, see smallest detectable change
smallest detectable change (SDC) 123, 217, 242–245, 258–261, 299–300
statistical significance of 216, 245
true change 204, 243, 266
classical test theory (CTT) 13, 18–20, 68, 72, 80, 100
basic formula 19, 137, 143, 186
clinical knowledge 193
clinician’s perspective 245, 256
clinimetrics 2
coefficient of variation 115
communality, see factor analysis
components 75, see factors
comprehensibility 58, 60, 184
comprehensiveness 43, 155–157, 299
computer adaptive testing (CAT) 26, 38, 141
conceptual framework 9, 13–18, 42, 44, 187, see framework
conceptual model 7–13, 33, 79, 151, 156, 173, 205
concordance 97, see reliability
consistency 97, see reliability
construct 13, 14, 33, 35, 45, 151, 156, 205, 279, 282
complex construct 15, 151–152
definition of 31, 33, 279
development of 193
dimensionality of 169–185
multidimensional construct 12, 52
non-observable 12–13, 17
observable 12–13, 17
unidimensional construct 12, 20
unobservable construct 17, 19–20, 55, 65, 150
construct validity, see validity
COOP-WONCA scales 46, 179–181
correlation
correlation matrix 73, 75
direction and magnitude of 174, 180, 211–212
item-total correlation 81–82
inter-item correlations 72, 80–81, 84
of change scores 211
P values of 212
polychoric correlations 80
sample size 191
statistical significance of 181
tetrachoric correlations 80
correlation coefficient 163
intra-class correlation coefficient (ICC) 103–110, 127, 163, 300
confidence interval (CI) 127
for single measurements 103–107, 143
for averaged measurements 107–110, 143
for agreement (ICC_agreement) 106–107, 110, 119
for consistency (ICC_consistency) 105–107, 110
interpretation of 120
sample size calculation 127
Pearson’s correlation coefficient 80, 98, 110, 163
Spearman’s correlation coefficient 163
COSMIN boxes 286–287
checklist 284–287, 289, 293, 300, 303, 309
four-step procedure 285–286
manual 285–286
scoring system 291, 303, 309
study 3
taxonomy 3, 4, 97, 286
terminology 3
website 286
Index

Cronbach’s alpha 81–84, 112, 137–139, 243, 298, 300
’alpha if item deleted’ 82–83
interpretation of 83–84
as reliability parameter 137–139
cultural adaptation 181–183
cultural differences 185, 191
data structure 65
Decision studies (D studies) 137
design of measurements 136
dependability 97, see reliability
difference 241
direction and magnitude of 174, 177, 180–181
statistical significant 212
systematic difference 104, 113–114, 121
true difference 132
differential item functioning (DIF) 182, 185–186, 188–189
non-uniform 186–189
uniform 186–189
dimension 66, 74
number of dimensions 72
underlying dimensions 72
dimensionality 65, 71, 80
examination using CTT 72
examination using IRT 80
multidimensionality 80
of constructs 169–185
optimizing 77
discrimination
between patients 89, 91, 101
discrimination parameter, see parameter
discriminative
discriminative ability 69, 139, 210
discriminative function 232
discriminative power 69,

see discriminative ability
discriminative purpose 34, 44–45, 123, 156
distribution
of change scores 248
of items scores 68, 85, 91
of marginals 121
of population scores 70, 90–91
of scores 228–235
using CTT methods 230–231
using IRT methods 231–235
effect indicator 13
effect size 215–216, 218–219, 247
eigenvalue, see factor analysis
equivalence
of factor loadings 186
of intercepts 186
of items 189
of scores 182, 185
diagonal error term 18–19, 100, 114
random error 104, 107, 110, 114, 145
systematic error 107, 110, 111
evaluative purpose 34, 44–45, 156, 159, 202–203
expert 38
expert committee 184
expert panel 156–157
language expert 184
external criterion 55, 246, 247
factors 73, 186
number of factors 74
interpretation of 77
factor analysis 51, 65–66, 71, 81, 169–185, 266–267, 298, 300
common factor analysis 72
communality 73–74, 77
confirmatory factor analysis (CFA) 72, 79, 169–188, 266
<table>
<thead>
<tr>
<th>Term</th>
<th>Page References</th>
</tr>
</thead>
<tbody>
<tr>
<td>eigenvalue</td>
<td>74–78</td>
</tr>
<tr>
<td>exploratory factor analysis (EFA)</td>
<td>72, 169–185, 186</td>
</tr>
<tr>
<td>factor loadings</td>
<td>73–74, 77–78, 186</td>
</tr>
<tr>
<td>factor model</td>
<td>171</td>
</tr>
<tr>
<td>fit index</td>
<td>155, 170</td>
</tr>
<tr>
<td>hypothesized structure</td>
<td>72, 170–192</td>
</tr>
<tr>
<td>item loadings</td>
<td>73</td>
</tr>
<tr>
<td>multiple group factor analysis</td>
<td>186, 191</td>
</tr>
<tr>
<td>principal components analysis (PCA)</td>
<td>72, 76</td>
</tr>
<tr>
<td>rotation</td>
<td>77</td>
</tr>
<tr>
<td>sample size</td>
<td>65, 80, 191</td>
</tr>
<tr>
<td>scree plot</td>
<td>75, 77</td>
</tr>
<tr>
<td>structure</td>
<td>65, 72, 74, 186</td>
</tr>
<tr>
<td>feasibility</td>
<td>59, 60, 292</td>
</tr>
<tr>
<td>Feinstein</td>
<td>2, 18</td>
</tr>
<tr>
<td>field testing</td>
<td>31, 92</td>
</tr>
<tr>
<td>floor and ceiling effects</td>
<td>91, 216, 232–235</td>
</tr>
<tr>
<td>focus group</td>
<td>38, 70</td>
</tr>
<tr>
<td>Food and Drug Administration (FDA)</td>
<td>152, 296</td>
</tr>
<tr>
<td>formative element</td>
<td>15, see formative model</td>
</tr>
<tr>
<td>formative and reflective models</td>
<td>13–18, 43–45, 83–84</td>
</tr>
<tr>
<td>formative model</td>
<td>14, 17–18, 42, 50, 55, 65–66, 70</td>
</tr>
<tr>
<td>form of administration</td>
<td>152</td>
</tr>
<tr>
<td>four-dimensional symptom questionnaire (4DSQ)</td>
<td>19, 50</td>
</tr>
<tr>
<td>framework</td>
<td></td>
</tr>
<tr>
<td>conceptual framework</td>
<td>9, 13–17, 18, 42, 44</td>
</tr>
<tr>
<td>content framework</td>
<td>156</td>
</tr>
<tr>
<td>frequency distribution</td>
<td>85–86</td>
</tr>
<tr>
<td>Generalizability and Decision studies (G and D studies)</td>
<td>131–137</td>
</tr>
<tr>
<td>Generalizability studies (G studies)</td>
<td>131–135</td>
</tr>
<tr>
<td>Generalizability coefficient (G coefficient)</td>
<td>131–132, 135–136, 138, 145</td>
</tr>
<tr>
<td>for agreement (G agreement)</td>
<td>133, 137</td>
</tr>
<tr>
<td>for consistency (G consistency)</td>
<td>133–135</td>
</tr>
<tr>
<td>global rating scale</td>
<td>207, 241, 250</td>
</tr>
<tr>
<td>continuous</td>
<td>208</td>
</tr>
<tr>
<td>dichotomous</td>
<td>208</td>
</tr>
<tr>
<td>ordinal</td>
<td>208</td>
</tr>
<tr>
<td>Guttmann scale</td>
<td>21, 238</td>
</tr>
<tr>
<td>Guyatt’s responsiveness ratio</td>
<td>217–218, 259–260</td>
</tr>
<tr>
<td>health-related quality of life (HRQL)</td>
<td>7, 9, 10, 39, 57</td>
</tr>
<tr>
<td>hierarchical order</td>
<td>21, 43, 238</td>
</tr>
<tr>
<td>hypotheses</td>
<td></td>
</tr>
<tr>
<td>a priori defined hypotheses</td>
<td>72, 211, 218</td>
</tr>
<tr>
<td>confirmation and rejection of</td>
<td>175, 205, 214</td>
</tr>
<tr>
<td>formulate hypotheses</td>
<td>151, 173, 205</td>
</tr>
<tr>
<td>number of</td>
<td>205</td>
</tr>
<tr>
<td>specific hypotheses</td>
<td>151, 153, 169–172, 174, 185, 211, 214, 219</td>
</tr>
<tr>
<td>testing of</td>
<td>151, 169–182, 202, 211, 299, 300, see also construct validity</td>
</tr>
<tr>
<td>index</td>
<td>49, 51–53, 70</td>
</tr>
<tr>
<td>information curve</td>
<td>140</td>
</tr>
<tr>
<td>amount of information</td>
<td>140–141</td>
</tr>
<tr>
<td>level of information</td>
<td>140</td>
</tr>
<tr>
<td>internal consistency</td>
<td>80–84, 97, 137, 139, 298, 300</td>
</tr>
<tr>
<td>internal reliability</td>
<td>137, see internal consistency</td>
</tr>
<tr>
<td>International Classification of Functioning (ICF)</td>
<td>46, 158</td>
</tr>
<tr>
<td>Page</td>
<td>Section</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>332</td>
<td>interpretability 91, 227–268</td>
</tr>
<tr>
<td></td>
<td>concept of 228</td>
</tr>
<tr>
<td></td>
<td>of change scores 217–218</td>
</tr>
<tr>
<td></td>
<td>interpretation of change scores 241–268</td>
</tr>
<tr>
<td></td>
<td>of measurement properties 229</td>
</tr>
<tr>
<td></td>
<td>of single scores 235–241</td>
</tr>
<tr>
<td></td>
<td>interview 38, 70</td>
</tr>
<tr>
<td></td>
<td>cognitive interview 265</td>
</tr>
<tr>
<td></td>
<td>probing method 58</td>
</tr>
<tr>
<td></td>
<td>three-step test interview 59, 265</td>
</tr>
<tr>
<td></td>
<td>intraclass correlation coefficient, see correlation coefficient</td>
</tr>
<tr>
<td></td>
<td>item</td>
</tr>
<tr>
<td></td>
<td>item bank 38</td>
</tr>
<tr>
<td></td>
<td>item characteristics 80</td>
</tr>
<tr>
<td></td>
<td>item characteristic curve (ICC) 22–25, 38, 85, 87–88, 139, 188</td>
</tr>
<tr>
<td></td>
<td>item correlation, 72, 80–81, 84, see correlations inter-item</td>
</tr>
<tr>
<td></td>
<td>item difficulty 20, 23, 84–86, 90, 140, 189, 231</td>
</tr>
<tr>
<td></td>
<td>item discrimination 84, 140, 189</td>
</tr>
<tr>
<td></td>
<td>item functioning characteristic 84</td>
</tr>
<tr>
<td></td>
<td>item level 68, 182</td>
</tr>
<tr>
<td></td>
<td>item loadings, 73, see factor analysis</td>
</tr>
<tr>
<td></td>
<td>item location 20, 90, 231–232, 237–239</td>
</tr>
<tr>
<td></td>
<td>item reduction 65–66, 72, 77, 81, 83–84, 88</td>
</tr>
<tr>
<td></td>
<td>item redundancy 83</td>
</tr>
<tr>
<td></td>
<td>item response 186</td>
</tr>
<tr>
<td></td>
<td>item response curve 68, see item characteristic curve</td>
</tr>
<tr>
<td></td>
<td>item scores 46, 51, 66</td>
</tr>
<tr>
<td></td>
<td>item-total correlation, 81–82, see correlation</td>
</tr>
<tr>
<td></td>
<td>item variance 69</td>
</tr>
</tbody>
</table>
level of ability 22, see also trait level
level of measurement 162–163, 208
  continuous 208
  dichotomous 208
  interval level 47
  nominal level 46
  ordinal level 46, 208
  ratio level 48
Likert
  items 46
  scales 66
limits of agreement, see Bland and Altman method
logistic regression analysis 185–188, 191
  interaction term 186
  ordinal logistic regression analysis 187–188
  regression coefficient 186
measurement
  clinician-based 10
  methods of, 7, 31, 35
  objective 11, 97
  patient-based 10, 12
  standardization of 144
  subjective 11–12, 97
  parameter, see parameter of measurement error
  reduction of 145
measurement instrument
  application of 156
  appropriateness of 275
  characteristics of 10, 291–292, 303
  content analysis 158
  content of 275–311
  development 32, 72, 89, 91, 157
  format 292
missing scores or values 66–68, 170–192, 220, 290
Multidimensional Fatigue Index (MFI) 52
multidimensional inventory 80
multifactor inventory 80
multiple indicator multiple cause (MIMIC) model 55
multiple measurements, see repeated measurements
multitrait-multimethod (MTMM) 181
Neck Disability Index (NDI) 90–91, 230, 232–234, 303
noise 204, see measurement error
Objectives measurements 11
paired t-test 98–99, 113
parameter
difficulty parameter 22–25, 85–88, 90–91, 141
discrimination parameter 22, 24–25, 81, 85–88
of measurement error 101–102, 113, 120, 122–124, 247
for continuous variables 111
interpretation of 122–123
IRT analysis 139–141
of reliability 101–102, 120, 123, 229
confidence interval (CI) 126, 128
Cronbach’s alpha 137–139
for categorical variables 119
for continuous variables 103–110
interpretation of 120–122, 126
IRT analysis 139–141
summary index 140
threshold parameter 23, see difficulty parameter
patient
ability 20–21, 23, 139, 141, 237–238
groups of patients 142, 244
individual patients 142, 244
perspective 245, 256
preferences 56
patient-reported outcome (PRO) 11, 17, 30, 157, 161, 163, 207, 245
performance test 96, 125, 144, 292
physical functioning 35, 141, 158, 185, 189, 190
pilot-testing 31, 57–60, 65, 184
of non-PRO instruments 58, 60
of PRO instruments 58
population, see also sample characteristics of 278–293, 301, 305
stable population 244
target population 34, 45, 58, 65, 152, 156–157, 161, 168, 174, 184–185, 279–280, 282, 303
precision 97, see reliability
prediction 34
prediction models 35
predictive purpose 35, 156
predictive value 162, 164
preference analysis 57
principal components analysis (PCA), see factor analysis 72, 76
PRISMA statement 302
profile 51–52
PROMIS 38, 141
proxy respondents 59
psychometrics 2
publication bias 281
rating
global rating of change, see change and scale
of importance 70
recall bias 256
receiver operating characteristic (ROC) curve 163, 165–166, 168, 208, 253–254
area under the curve 165, 168–169, 208
reflective model 13, 15, 17–20, 42, 50, 55, 66, 70
regression analysis 55
regression equation 73
standardized regression coefficients 73
relevance 58, 155–157, 299
reliability 72, 96–145, 162, 194, 195, 242, 298, 300
analysis 108
improvement of 144–145
inter-rater reliability 96–97
intra-rater reliability 96–97
of mean values 108, 137, 145
parameter, see parameter of reliability
test–retest reliability 96–97, 243
reliability study
crossed design 130
design of 124–131
nested design 130
sample size 126–128
time interval 125, 290
reliable change index (RCI) 243, 260
repeatability 97, see reliability
repeated measurements 96, 100, 111, 115, 128, 131, 135–136, 143–144
reproducibility 97, see reliability
response option 45, 66, 68
dichotomous response 46, 85
response shift 261–268
adjustment for 267
assessment of 264–267
bias 267
conceptual model of 263–264
individualized measures 266
interpretation of 267–268
qualitative methods 264–267
quantitative methods 265–266
recalibration 262–263, 266
reconceptualization, 262, 266
reprioritization 262, 266
then-test 265, 267
theoretical model, see conceptual model
responsiveness 91, 196, 202–221, 235, 299
criterion approach 202, 206, 211–215
definition of 203, 215
in a clinical study 220
inappropriate measures of 215–218
responsiveness study 205, 210, 211
Roland–Morris Disability Questionnaire (RDQ) 50, 85–89, 122, 227
rotation 77, see factor analysis
orthogonal rotation 77
Varimax rotation 77–78
sample
sample-dependent 138
heterogeneous sample 102, 112, 119–120, 138, 229
homogeneous sample 101, 112, 119–120, 138, 229
sample size 191, 220, 290
scale 49, 52, 70
average scores 50
deterministic scale 21
global rating scale 207, 241, 250–251
Guttman scale 21
interval scale 89
IRT scales 50
numerical rating scale 241–242
ordinal scale 68, 163
probabilistic scale 21
scale (cont.)

scale level 66, 182
scale scores 50–51
subscale 80
types of 47
unidimensional scale 80, 83, 298
visual analogue scale (VAS) 49

score
adjusted score 191
average score 195
change score 252, see change and interpretation
clustering of 70
impact score 71
importance score 71
in formative models 51–57
in reflective models 50, 52
interpretation of 36
IRT based estimation of 231, 238
latent score 189
norm score 236
observed score 19, 98, 100, 189
population scores 70
scale score 51
single score 195
sum-scores 52, 56, 89, 239
true score 19, 20, 98, 100, 186, 195
unweighted score 51
weighted score 51, 53
scoring options 31, 46
scree plot 75, 77, see factor analysis
SEIQOL-DW 54, 266
sensibility 18
sensitivity 162–164, 246, 252, 254
sensitivity to change 216,
see responsiveness
Short-Form 36 (SF-36) 152, 177, 179–181, 187, 231–232, 236
situation-dependent 151–152, 161–162, 174
smallest detectable change, see change
Spearman 2
specificity 162–164, 246, 252, 254
stability 97, see reliability
standard deviation (SD)
half a SD (0.5 SD) 247, 259
of baseline scores 215–216
of change scores (SD\textsubscript{\text{change}}) 216, 217, 244, 260
of single measurement 111
of difference (SD\textsubscript{\text{difference}}) 111, 114, 243–244
standard error (SE) in IRT 139–142
standard error of measurement (SEM) 101, 111–113, 243, 247, 259
SEM value 111
interpretation of 122
SEM for agreement (SEM\textsubscript{agreement}) 111, 114
SEM for consistency (SEM\textsubscript{consistency}) 111, 114–115
standardized response mean (SRM) 215
structural equation modelling 56–57, 181
structural reliability 137, see internal consistency
subjective measurements 11
systematic review of measurement properties 30, 275–311
best evidence synthesis 297,
see qualitative analysis
conclusion of 300–302, 306
consistency of results 296–298, 301
data extraction 291, 301
data synthesis 286, 296–300, 306
(dis)similarity of settings 293, 296
(dis)similarity of studies 286, 297
(dis)similarity of study populations 293, 296, 301, 305
eligibility criteria 282–283
evidence 296
  conflicting evidence 308
  consistent evidence 308
  indirect evidence 282
  levels of evidence 298
  flow chart 284–285, 302
  generalizability of results 293, 296
  homogeneity of studies 296–298
  inclusion and exclusion criteria 282, 301
  qualitative analysis 296–298, 308
  quantitative analysis 296, see statistical pooling
  reporting of 302–309
  research question 276, 301
  search strategy 281, 302
  building of 279
  database 279, 283
  documentation of 283
  language restriction 281
  methodological search filter 280
  reference checking 281
  search terms 279–280
  time-limit 281
  update the search 283
  similarity of studies 286
  statistical pooling 297
  types of 276

then-test 265, 267, see response shift
  theories, see theories of measurement
theta, see trait level
  think aloud method 58, 265
  thought test 14
  time interval 125, 205, 206, 244
  trait level 68, 89–91
  transition question 256
  translation 169–182, 185, 190
  back 183
  forward 182–183

units of measurement 114, 119, 122
  utility analysis 56–57
validity or validation 150–196
  along a clinical study 192
  concept of 151–154
  concurrent validity 159–160, 163–167
  construct validity 72, 150, 169–191, 194, 207, 299, 300
  content validity 150, 154–158, 194, 216, 299
  continuous process 151, 153
  convergent validity 173, 176–178, 181
  criterion validity 150, 159–169, 191, 194
  cross-cultural validity 152, 169–181
  discriminant validity 173, 176, 177, 181
  discriminative validity 173, 175–179, 181
  face validity 154–155, 194
  hypothesis testing, see hypotheses
  known group validity, see discriminative validity
  longitudinal validity 196, see responsiveness
  of change scores 203–204, 216–218
  of single scores 151, 153, 203–204
  predictive validity 159, 160, 163, 165, 167–169
  sample size 191, 220
  structural validity 169–181, 194
  types of 150, 154, 194
VARCOMP analysis 106
  variability 97, see reliability
  variables
  categorical 48
  continuous 48
  discrete 48
Index

variance

components of 104, 106, 129–130, 132–133, 135–137

cumulative percentage of explained variance 75

due to systematic differences 104–106

error variance 100, 104–106, 111, 132, 134, 135

explained variance of factors 73

explained variance of items 73, see communalities

observed variance 100

of patients 104, 132, 134

percentage of explained variance 74–76, 78

remaining variance 74

residual variance 104–105, 135

shared variance 72

total variance 74, 78, 100, 132, 134

ture variance 100, 132

variation

between patients 101, 107

day to day variation 96

restriction of 144

sources of variation 96, 128–129, 136, 145

weighting

individual weighting 54

in formative models 57

preference weighting 56

weights 51

empirical weights 54, 57

in IRT model 51

in Rasch model 51

in two-parameter IRT model 51

judgemental weights 54, 57

method of 57

using CTT 51

WOMAC 185, 189–190, 276

Wilson and Cleary model 7–13, 15