

INHOUD

Woord vooraf

P.F. Sanders en T.J.H.M. Eggen

1 Inleiding	1
1.1 Testindelingen	1
1.2 Toetsconstructie	3
1.3 Het valideren van meetinstrumenten	9
1.4 Psychometrie in de praktijk	12

N.H. Veldhuijzen en F.G.M. Kleintjes

2 Dataverzameling	17
2.1 Van waarnemingen tot data	18
2.2 Schaalniveaus	18
2.3 Meten per fiat	21
2.4 Procedures voor dataverzameling	21
2.5 Betrouwbaarheid en validiteit	22
2.6 Steekproeven	24
2.2.1 <i>Representativiteit van steekproeven</i>	25
2.6.2 <i>Nauwkeurigheid</i>	25
2.6.3 <i>Aselecte steekproeven</i>	26
2.6.4 <i>Gestratificeerde steekproeven</i>	27
2.6.5 <i>Getrapte steekproeven</i>	27
2.6.6 <i>Intraklassecorrelatie</i>	28
2.7 Proefopzetten	30
2.8 Stimuli	31
2.9 Meetmodellen	31

N.H. Veldhuijzen, P. Goldebeld en P.F. Sanders

3 Klassieke testtheorie en generaliseerbaarheidstheorie	33
3.1 Ware score	34
3.2 De centrale formule van de klassieke testtheorie	35
3.3 Betrouwbaarheid	36
3.4 Standaardmeetfout	37
3.5 Schattingen van de ware score	38
3.6 Het schatten van de betrouwbaarheid en de standaardmeetfout	40
3.6.1 <i>Parallele metingen</i>	40
3.6.2 <i>Test-hertestmethode</i>	42
3.6.3 <i>Toetsverlenging</i>	42
3.6.4 <i>Coëfficiënt alpha</i>	44
3.7 Toets- en itemanalyse	46
3.7.1 <i>Toets- en itemindices bij toetsen met meerkeuzevragen</i>	46
3.7.2 <i>Itemindices bij toetsen met meerkeuzevragen</i>	47
3.7.3 <i>Toetsindices bij toetsen met meerkeuzevragen</i>	51
3.7.4 <i>Toets- en itemindices bij toetsen met open vragen</i>	52
3.7.5 <i>Itemindices bij toetsen met open vragen</i>	53
3.7.6 <i>Toetsindices bij toetsen met open vragen</i>	54
3.8 Betrouwbaarheid en standaardmeetfout	54
3.8.1 <i>Coëfficiënt alpha en de KR-20</i>	54
3.8.2 <i>Verschilscores</i>	55
3.9 Nauwkeurigheid van toets- en itemindices	56
3.9.1 <i>Standaardfout van een p-waarde</i>	57
3.9.2 <i>Standaardfout van een gemiddelde toetsscore en van een p'-waarde</i>	58
3.9.3 <i>Standaardfout van een r_{it}-waarde</i>	58
3.9.4 <i>Standaardfout van coëfficiënt alpha</i>	59
3.10 Normen voor toets- en itemindices	59
3.10.1 <i>Normen voor p- en p'-waarden</i>	60
3.10.2 <i>Normen voor r_{it}-waarden</i>	60
3.10.3 <i>Normen voor de betrouwbaarheid</i>	61
3.11 Generaliseerbaarheidstheorie	62
3.12 Design met een facet	64
3.12.1 <i>Generaliseerbaarheidsstudie</i>	66
3.12.2 <i>Decisiestudie</i>	70
3.13 Design met twee facetten	73

3.13.1	<i>Generaliseerbaarheidsstudie</i>	74
3.13.2	<i>Decisiestudie</i>	79
3.14	Andere aspecten van de generaliseerbaarheidstheorie	81

N.D. Verhelst

4	Itemresponstheorie	83
4.1	Begrippen en algemene theorie	86
4.1.1	<i>Het Raschmodel</i>	89
4.1.2	<i>Lokale stochastische onafhankelijkheid</i>	91
4.2	Het schatten van de parameters in het Raschmodel	93
4.2.1	<i>Grootste-aannemelijkheidsschatters: een voorbeeld</i>	93
4.2.2	<i>JML-schatting in het Raschmodel</i>	99
4.2.3	<i>CML-schatting in het Raschmodel</i>	103
4.2.4	<i>MML-schatting in het Raschmodel</i>	111
4.2.5	<i>Een voorbeeld</i>	114
4.3	Het toetsen van het Raschmodel	118
4.3.1	<i>De veronderstellingen van het Raschmodel</i>	120
4.3.2	<i>Relaties tussen het Raschmodel en het multinomiale model</i>	124
4.3.3	<i>Likelihood-ratio-toetsen</i>	126
4.3.4	<i>Wald-toetsen</i>	131
4.3.5	<i>Veralgemeende Pearson χ^2-toetsen</i>	136
4.3.6	<i>Een voorbeeld</i>	158
4.4	Het Raschmodel en onvolledige designs	161
4.5	Het schatten van de persoonsparameters	167
4.5.1	<i>Drie methoden om de persoonsparameter θ te schatten</i>	169
4.5.2	<i>Een voorbeeld</i>	175
4.5.3	<i>Passingsindices voor individuele antwoordpatronen</i>	176

C.A.W. Glas en N.D. Verhelst

5	Een overzicht van itemresponsmodellen	179
5.1	Het lineair-logistische testmodel	180
5.1.1	<i>Parameterschatting in het LLTM</i>	181
5.1.2	<i>Het toetsen van het LLTM</i>	184
5.1.3	<i>Een toepassing van het LLTM</i>	187
5.2	Indelingsprincipes van IRT-modellen	190

5.3	Unidimensionale modellen voor dichotome items	195
5.3.1	<i>Het twee- en drieparameter logistisch model</i>	196
5.3.2	<i>Het éénparameter logistisch model (OPLM)</i>	205
5.3.3	<i>Modellen zonder de assumptie van lokale stochastische onafhankelijkheid</i>	208
5.4	Unidimensionale modellen voor polytome items	211
5.4.1	<i>Het partial credit model (PCM)</i>	213
5.4.2	<i>Generalisaties van het partial credit model</i>	218
5.5	Multidimensionale IRT-modellen	226
5.5.1	<i>Een OPLM met een multivariate vaardigheidsverdeling</i>	229
5.5.2	<i>Het multidimensionale model van Rasch</i>	231
5.5.3	<i>Compensatorische IRT-modellen</i>	233
5.5.4	<i>Conjunctieve IRT-modellen</i>	236
5.6	Nabeschouwing	237

T.J.H.M. Eggen

6	Itemresponstheorie en onvolledige gegevens	239
6.1	De relatie tussen onvolledige gegevens en IRT	240
6.1.1	<i>Efficiëntie van de schattingen</i>	241
6.1.2	<i>Calibratie in onvolledige designs en linken</i>	243
6.2	De datamatrices van structureel onvolledige designs	247
6.3	De stochastische structuur van structureel onvolledige designs	251
6.3.1	<i>Gerandomiseerd onvolledig design</i>	252
6.3.2	<i>Meerfasen onvolledig design</i>	253
6.3.3	<i>Groepsgericht onvolledig design</i>	255
6.4	Algemene voorwaarden voor calibratie in onvolledige designs	256
6.5	Voorwaarden voor calibratie in stochastische designs	259
6.5.1	<i>MML in stochastische designs</i>	264
6.5.2	<i>CML in stochastische designs</i>	272
6.6	Schatten van persoonsparameters in stochastische designs	280
6.6.1	<i>ML- en WML-vaardigheidsschatting in stochastische designs</i>	280
6.6.2	<i>EAP vaardigheidsschatting in stochastische onvolledige designs</i>	282

N.D. Verhelst en F.G.M. Kleintjes

7 Toepassingen van itemresponstheorie	285
7.1 De PPON-rekenpeiling	286
7.2 De Cito leesbaarheidsindex voor het basisonderwijs	295
7.3 De diagnostische verborgen-figurentest	304

R.J.H. Engelen en T.J.H.M. Eggen

8 Equivaleren	309
8.1 Overzicht equivaleren	310
8.1.1 <i>Psychometrische voorwaarden voor equivaleren</i>	312
8.1.2 <i>Designs voor equivaleren</i>	315
8.2 Equivaleren in de klassieke testtheorie	320
8.2.1 <i>Basismethoden voor equivaleren</i>	321
8.2.2 <i>Equivaleren met behulp van het ankertoetsdesign</i>	328
8.3 Equivaleren met itemresponstheorie	332
8.3.1 <i>Calibratie</i>	334
8.3.2 <i>Verschillende vormen van equivalering in de itemresponstheorie</i>	336
8.3.3 <i>Equivaleren met behulp van een itembank</i>	341
8.3.4 <i>Quasi-multidimensionaal IRT-equivaleren</i>	344
8.4 De kwaliteit van de equivalente methoden vergeleken	346

C.A.W. Glas en M.J. Ouborg

9 Vraagonzuiverheid	349
9.1 Definitie van onzuiverheid	350
9.2 Methoden voor het bepalen van vraagonzuiverheid	353
9.2.1 <i>De Mantel-Haenszel-procedure</i>	354
9.2.2 <i>Procedure met IRT-modellen</i>	356
9.2.3 <i>De relatie tussen de Mantel-Haenszel-procedure en de IRT-procedure</i>	363
9.2.4 <i>Een voorbeeld van het bepalen van vraagonzuiverheid met behulp van het OPLM</i>	364
9.3 Conclusie	370

F.H. Kamphuis en R.J.H.Engelen

10	Het meten van veranderingen	371
10.1	Individuele groei	372
10.1.1	<i>Longitudinale data en modellering</i>	372
10.1.2	<i>Het vaststellen van de individuele groei bij zuigelingen</i>	373
10.1.3	<i>Problemen bij het volgen van individuele leerlingen</i>	375
10.2	Klassieke testtheorie en groeiscoringen	378
10.2.1	<i>Artificiële longitudinale data</i>	378
10.2.2	<i>Statische benadering</i>	379
10.2.3	<i>Dynamische benadering</i>	384
10.2.4	<i>Evaluatie statische en dynamische benadering</i>	390
10.2.5	<i>Schattingen van structurele parameters</i>	394
10.3	Itemresponstheorie en groeiscoringen	396
10.3.1	<i>Schaal Vorderingen en Spellingvaardigheid</i>	396
10.3.2	<i>Het schatten van de latente vaardigheid</i>	398
10.4	Epiloog	406

T.J.J.M. Theunissen, P.F. Sanders en A.J. Verschoor

11	Het samenstellen van toetsen	409
11.1	Mathematisch programmeren	410
11.2	Het samenstellen van toetsen in de itemresponstheorie	416
11.2.1	<i>Lineaire programmeringsproblemen</i>	417
11.2.2	<i>Praktijkvoorbeelden</i>	420
11.2.3	<i>Specificeren van restricties en relaties</i>	428
11.3	Het samenstellen van toetsen in de klassieke testtheorie	431
11.4	Het samenstellen van toetsen in de generaliseerbaarheidstheorie	438

A.P.J.M. Heuvelmans en P.F. Sanders

12 Beoordelaarsovereenstemming	443
12.1 Definitie van beoordelaarsovereenstemming	444
12.2 Beoordelaarsovereenstemming bij data van nominaal niveau	444
12.3 Beoordelaarsovereenstemming bij data van ordinaal niveau	451
12.4 Beoordelaarsovereenstemming bij data van intervalniveau	457
12.5 Lage beoordelaarsovereenstemming: oorzaken en remedies	466
12.6 Tot besluit	469

H.H.F.M. Verstralen

13 Schalen, normen en cijfers	471
13.1 Het niveau van de schaal	472
13.2 Normschalen	474
13.2.1 <i>Cumulatieve verdelingen</i>	475
13.2.2 <i>Genormeerde lineaire transformaties</i>	478
13.2.3 <i>Genormaliseerde schalen</i>	479
13.2.4 <i>Ontwikkelingsschalen</i>	483
13.2.5 <i>De nauwkeurigheid van normschalen</i>	485
13.3 Beheersingsschalen	486
13.4 Het rapporteren van meetnauwkeurigheid	487
13.5 De cesuur voldoende/onvoldoende en andere normen voor cijfergeving	492
13.5.1 <i>Traditionele methoden van cesuurbepaling</i>	492
13.5.2 <i>Cesuurbepaling en overige cijfers binnen itemresponstheorie</i>	503
13.6 Conclusie	509

Literatuur	511
-------------------	-----

Personenregister	529
-------------------------	-----

Zakenregister	533
----------------------	-----