



Methodologische achtergronden van het testinstrumentarium

HRorganizer.com is een online HR systeem waarin selectie-instrumenten, ontwikkelingsgerichte methodes en beoordelingssystematiek in samenhang met elkaar worden aangeboden. Dit document geeft achtergronden en cijfermatige onderbouwing over de tests en zelfbeschrijvingstechnieken die standaard aanwezig zijn in HRorganizer.com

Indien er functie- of competentiegericht assessment moet plaatsvinden stelt het systeem automatisch het beste testprogramma samen* en rapporteert automatisch op de relevante competenties.

Naast deze standaard aanwezige instrumenten beschikt HRorganizer over vele instrumenten die klant-specifiek kunnen worden ingebouwd.

Notitie Juni 2010: vanaf 25 juni zijn zowel de geactualiseerde (101) competenties alsook de vernieuwde set aan instrumenten beschikbaar in HRorganizer.com. Als gevolg daarvan zullen de in dit document aanwezige normgegevens in de periode direct na juni 2010 regelmatig gewijzigd en aangevuld worden. Van dit document zullen wij in de tweede helft van 2010 nog regelmatig nieuwe versies publiceren.

* Uit het aanwezige testinstrumentarium



Inhoud

A.	Verantwoording	P. 3
B.	Introductie	P. 4
C.	Betrouwbaarheid	P. 5
D.	Validiteit	P. 10
	D1. Constructvaliditeit	P. 10
	D2. Predictieve validiteit	P. 13
E.	Procesverloop	P. 16
F.	Samenstellen testprogramma's	P. 17
	F1. Keuze van instrumenten	P. 17
	F2. Afname volgorde	P. 18
	F3. Itemkeuze	P. 19
	F3.1. Item generatie	
	F3.2. Equivalentie	
	F4. Adaptief testen	P. 21
	F4.1. Branching	
	F4.2. Flexibel	
	F4.3. De gekozen oplossing	
	F5. Keuze normgroep en normering	P. 23
G.	Conclusie	P. 25
H.	Referenties	P. 26

A. Verantwoording

In opdracht van HRorganizer B.V. is op basis van beschikbare research – zoals uitgevoerd door Cut-e GmbH – dit document samengesteld. Hierbij is omwille van de compactheid en leesbaarheid een selectie gemaakt van de belangrijkste bevindingen. Uitvoerige psychometrische documentatie (Engelstalig) is verkrijgbaar via HRorganizer B.V. (contactgegevens zie www.HRorganizer.com). Opgemerkt wordt dat dit document een 'levend' document is. Het geeft de stand van zaken weer zoals die in juni 2010 was. Er zijn voortdurend researchactiviteiten gaande, waarvan de resultaten, indien relevant, toegevoegd zullen worden aan dit document.

De persoonlijkheidsvragenlijst en de cognitieve testbatterij zijn door DNV Noorwegen gecertificeerd tegen de criteria zoals die zijn opgesteld door de EFPA (European Federation of Psychologists' Associations) voor alle beschikbare taalversies. Meer informatie is verkrijgbaar via HRorganizer.



HRorganizer.com is een systeem dat bedoeld is voor gebruikers binnen Europa, inclusief Nederland. Ten einde na te kunnen gaan in welke mate de beschikbare normgroepen uitwisselbaar zijn, is statistisch onderzoek uitgevoerd. De centrale vraag daarbij was: in welke mate zijn er verschillen tussen Europese normgroepen op de diverse instrumenten. Er zijn gegevens beschikbaar over meer dan 30 Europese landen. In het onderzoek is statistisch getoetst of er sprake is van significante verschillen tussen landen. Er zijn normen beschikbaar per land én voor Europa als geheel. De normgegevens voor Nederland zijn verzameld van 2006-2010.

Dit document is oorspronkelijk samengesteld door Dr. Wouter Schoonman, Psy Tech industrial psychology. Binnen Psy Tech vervult hij de rollen van onafhankelijk bedrijfspsycholoog, methodoloog en consultant HRM. Daarnaast is hij part time verbonden aan Saxion Hogescholen als lector Assessment.

Maarten Hack, directeur HRorganizer B.V.

Wageningen, juni 2010



B. Introductie

HRorganizer.com heeft testtechnologie en testconcepten van Cut-e GmbH integraal opgenomen. In de verdere bespreking zullen we voor het leesgemak hiernaar verwijzen met de term 'instrumenten'. Het zijn instrumenten gericht op het meten van individuele verschillen op het gebied van *Drijfveren (motivatie), Gedragsvoorkeuren en Cognitieve capaciteiten*. In dit hoofdstuk worden algemene kenmerken van deze instrumenten besproken. Gedetailleerde informatie per instrument is beschikbaar in diverse aparte documenten (Preuss, 2002-2010).

In onderstaand overzicht worden de beschikbare instrumenten in HRorganizer.com weergegeven:

Naam	Meetpretentie	# items	Afnametijd	Adaptief	Itemgeneratie
Shapes Executive	Gedragsvoorkeuren	192	≈ 20	ja	n.v.t.
Shapes	Gedragsvoorkeuren	144	≈ 15	ja	n.v.t.
Shapes admin	Gedragsvoorkeuren	90	≈ 10	ja	n.v.t.
Views	Drijfveren	84	≈ 10	ja	n.v.t.
Scales numerical consumer	Numeriek redeneren	37	12	nee	nee
Scales numerical admin	Numeriek redeneren	36	12	nee	nee
Scales verbal consumer	Verbaal redeneren	49	12	nee	nee
Scales verbal admin	Verbaal redeneren	42	12	nee	nee
Scales clues	Informatieverwerking	50	15	nee	nee
Scales cls	Inductief logisch redeneren	12	12	nee	ja
Scales fx	Deductief logisch redeneren	12	15	nee	ja
Scales stm	Korte termijn geheugen	10	5	nee	ja
Scales eq1 (a)	Rekenvaardigheid	variabel ¹	5	ja	ja
Scales spr	Ruimtelijk inzicht	12	10	nee	ja
Scales mt	Multi tasking	variabel ¹	5	ja	ja
Scales lct	Vermogen om te leren	6	5	nee	ja
Scales ix (a)	Inductief logisch redeneren (laag en midden niveau)	variabel ¹	5	ja	ja
Scales lst (a)	Deductief logisch redeneren (laag en midden niveau)	variabel ¹	6	ja	ja

¹ Afhankelijk van de prestaties van de kandidaat



Tabel 1. Overzicht beschikbare instrumenten

Voor verschillende instrumenten zijn verschillende versies beschikbaar, afhankelijk van het gewenste niveau. Zie de betreffende paragraaf over normen (blz 23 e.v.).

C. Betrouwbaarheid

In onderstaande tabel de betrouwbaarheid van de cognitieve capaciteiten tests:

Naam	Score	Betrouwbaarheid ²	N
Scales numerical consumer	Performance	.84	213
	Precisie	.89	
Scales numerical admin	Performance	.86	162
	Precisie	.76	
Scales verbal consumer	Performance	.87	204
	Precisie	.86	
Scales verbal admin	Performance	.84	162
	Precisie	.75	
Scales clues	Performance	.79	285
Scales cls	Performance	.86	271
	Accuratesse	.81	
Scales fx	Performance	.84	258
	Accuratesse	.89	
Scales stm	Performance	.93	317
	Reactiestijl	.88	
	Accuratesse	.91	
Scales mt	Performance	.91	1163
Scales eq1	Performance	.89	1538

² Bij speedtests is de standaard Cronbach's alpha niet (zonder meer) toepasbaar. De split-half betrouwbaarheid geeft een betere indicatie en staat in deze tabel.

	Snelheid	.87	
Naam	Score	Betrouwbaarheid	N
Scales lct	Performance	.88	11064
Scales ix			
Scales lst			
Scales spr	Performance	.81	211
	Accuratesse	.67	

Tabel 2. Overzicht betrouwbaarheid capaciteitentests

Uit de tabel blijkt de hoge interne consistentie. De COTAN (Evers et al, 2000) houdt bijvoorbeeld een ondergrens aan van 0.8, wat erg streng is. Alle instrumenten/scores voldoen hieraan (0.79 is afgerond 0.8). Zie ook de paragraaf over Equivalentie, waar de test-hertest betrouwbaarheid aan de orde komt. In de volgende tabellen gaat het over de betrouwbaarheid van de vier vragenlijsten die in HRorganizer.com zijn opgenomen (Shapes executive, Shapes, Shapes basic en Views).

Dimensie	Alpha	M	SD
leidinggevend (directing)	.75	19.7	4.3
overtuigend (persuasive)	.69	18.8	4.0
zelfverzekerd (socially confident)	.73	19.9	6.8
gericht op relaties (sociable)	.70	17.6	4.2
gericht op samenwerking (agreeable)	.65	18.5	4.5
gericht op gedrag (behavioral)	.59	20.3	4.4
verstandig (prudent)	.71	19.8	3.2
resultaatgericht (focused on results)	.73	20.4	3.6
planmatig (planful)	.53	18.3	4.2
nauwgezet (conscientious)	.69	17.0	5.7
analytisch (analytical)	.72	19.3	3.5
conceptueel (conceptual)	.71	18.0	3.0

Dimensie	Alpha	M	SD
vindingrijk (imaginative)	.70	16.0	4.4
veranderingsgezind (open for change)	.67	19.1	3.9
zelfsturend (autonomous)	.59	21.6	4.0
prestatiegericht (achieving)	.72	22.6	5.8
competitief (competitive)	.60	19.2	3.6
gedreven (energetic)	.64	20.9	4.8

Tabel 3. Betrouwbaarheden Shapes

In tabel 3 wordt een overzicht gegeven van de betrouwbaarheden (Cronbach's alpha) van de 18 schalen van Shapes. Daarnaast de gemiddelden en standaarddeviaties, alsmede de gecorrigeerde gemiddelde itemtotaalcorrelatie. De steekproef bestaat uit 405 personen.

De test-hertest betrouwbaarheid van de schalen in Shapes wordt weergegeven in onderstaande tabel (N = 97, 28% vrouwelijk, 27-59 jaar, voornamelijk academische opleiding, werkzaam in verschillende branches, werkervaring 2-15 jaar):

Dimensie	Hertest-betr.
leidinggevend (directing)	.81
overtuigend (persuasive)	.74
zelfverzekerd (socially confident)	.87
gericht op relaties (sociable)	.79
gericht op samenwerking (agreeable)	.75
gericht op gedrag (behavioral)	.72
verstandig (prudent)	.84
resultaatgericht (focused on results)	.82
planmatig (planful)	.69
nauwgezet (conscientious)	.76
analytisch (analytical)	.73
conceptueel (conceptual)	.83
vindingrijk (imaginative)	.77

veranderingsgezind (open for change)	.71
Dimensie	Hertest-betr.
zelfsturend (autonomous)	.70
prestatiegericht (achieving)	.81
competitief (competitive)	.83
gedreven (energetic)	.76

Tabel 4. Test-hertest betrouwbaarheid Shapes

De test-hertest betrouwbaarheid is hoog wat wijst op de stabiliteit van de meting in de tijd.

Een variant op Shapes is Shapes basic. Bij een steekproef met deze vragenlijst van $N = 1104$, 42% vrouwelijk, leeftijd 14-24, schoolverlaters werden de volgende betrouwbaarheden geconstateerd:

Dimensie	M	SD	alpha
uitdaging	16.9	3.7	.79
zingeving	17.4	3.4	.63
nauwgezetheid	18.9	3.8	.78
vindingrijkheid	13.7	4.2	.80
omzichtigheid	15.5	3.1	.75
werkpret	16.5	4.4	.72
eendracht	18.2	3.9	.77
gezelligheid	14.4	4.0	.76
samenwerkingsgerichtheid	18.0	3.2	.72
zelfstandigheid	15.0	3.5	.70
flexibiliteit	16.2	3.2	.75
waardering	10.0	4.8	.74
zelfredzaamheid	17.4	3.2	.77
vasthoudendheid	16.0	3.2	.74
ijver	16.7	3.0	.76
consistentie	4.0	1.0	

Tabel 5. Betrouwbaarheid Shapes basic

Ook hier zien we goede interne consistenties. De betrouwbaarheid is ook bij deze vragenlijst goed. Hierna volgt nog een overzicht van de interne betrouwbaarheid van een vierde vragenlijst (Views).

Dimensie	Alpha	M	SD
professionele uitdaging	.71	17.23	3.72
prestatiewaardering	.81	14.76	3.59
financiële beloning	.79	11.84	5.24
geborgenheid	.61	14.21	3.69
werkpret	.64	14.83	4.15
zingeving	.74	18.21	3.77
eendracht	.73	11.87	4.54
eerlijkheid	.63	17.26	3.12
samenwerking	.78	16.28	3.35
integriteit	.67	17.44	2.95
rechtvaardigheid	.73	17.41	2.89
hiërarchie	.84	14.46	4.12
structuur	.72	15.54	3.66
tempo van verandering	.64	13.25	3.64
ontwikkelingsmogelijkheden	.64	17.17	4.16
afwezigheid van stress	.78	12.26	3.89
beïnvloedingsmogelijkheden	.62	17.85	3.78
fysieke werkomgeving	.71	10.88	3.54

Tabel 6. Betrouwbaarheden Views

In de tabel staan de 18 betrouwbaarheden vermeld. De steekproef is N = 477.

De betrouwbaarheden van de vragenlijsten zijn lager dan die van de capaciteitentests. Dit is vaak het geval en verdedigbaar door te wijzen op het effect dat Akkerman (1984) beschreven heeft³. Wanneer een hoge betrouwbaarheid wordt nagestreefd, bestaat het risico dat het gemeten concept erg smal wordt. Bovendien wordt het aantal items erg hoog en qua inhoud gelijken ze zeer sterk op elkaar. In de

³ In zijn geval demonstreerde hij dat het mogelijk is een alpha van .98 te behalen met slechts 11 – min of meer identieke – items.



factoranalytische literatuur staan zulke nauwe schalen bekend als 'bloated specific' (Kline, 1998).
Ontraden wordt bij instrumentconstructie zulke schalen op te nemen.



D. Validiteit

Validiteit wil zeggen of een instrument aan zijn doel beantwoordt. Er zijn twee perspectieven over validiteit: het interne perspectief en het externe perspectief. Binnen beide perspectieven zijn twee manieren van kijken:

Intern	Extern
Convergent	Gelijktijdig
Divergent	Predictief

Tabel 7. Soorten validiteit

Vanuit het intern perspectief kan worden gekeken of het instrument daadwerkelijk meet wat de bedoeling is. Een algemeen woord is constructvaliditeit. Met Convergente validiteit wordt bedoeld of er verband bestaat tussen twee tests die hetzelfde pretenderen te meten. Divergent is precies het omgekeerde: een test met een bepaalde meetpretentie mag niet samenhangen met een test die heel iets anders meet. De tests moeten divergeren.

Bij het externe perspectief wordt gekeken naar samenhang tussen testcores en iets in de buitenwereld, bijvoorbeeld arbeidsprestatie. De testgegevens en de beoordeling van arbeidsprestatie kunnen gelijktijdig (Gelijktijdig) worden verzameld of na verloop van tijd (Predictief).

Er is veel validiteitonderzoek gedaan naar de instrumenten binnen HRorganizer.com. Een volledig overzicht is op te vragen bij HRorganizer BV. Hieronder enkele voorbeelden van de verschillende soorten onderzoek.

D1. Constructvaliditeit

Rond de opgenomen vragenlijsten in HRorganizer.com is validiteitonderzoek uitgevoerd. Hieronder een voorbeeld van onderzoek naar de constructvaliditeit van Shapes. Het eerste betreft een factor analyse naar de onderliggende structuur van de 18 dimensies. De gevonden oplossing komt overeen met het Great Eight competentie model (Kurz & Bartram, 2002) wat op zijn beurt weer voortbouwt op het Big5 persoonlijkheidsmodel. De oplossing is als volgt (voor details zie Preuss, 2002-2009):

Dimensie	F1	F2	F3	F4	F5	F6	F7	F8
leidinggevend	.69	-.20	-.26	-.09	-.16	.08	-.01	-.27
overtuigend	.79	-.17	-.04	.05	.14	.13	.02	.00
zelfverzekerd	.66	.05	-.02	.05	.21	-.21	-.05	-.05
gericht op relaties	.36	.35	.59	.16	-.07	-.15	-.20	.16
gericht op samenwerking	-.13	.67	.16	-.20	-.16	-.28	-.20	.02
gericht op gedrag	-.06	.05	.16	-.79	-.01	-.11	-.19	.13
verstandig	.02	-.20	-.06	-.54	.25	.39	.32	.26
resultaatgericht	.00	.18	-.14	.55	.12	.34	.14	.35
planmatig	-.07	-.07	.01	.03	-.14	-.20	.82	.12
nauwgezet	-.34	.38	-.05	.08	-.16	.22	.53	.19
analytisch	-.11	-.16	-.10	-.16	-.04	-.01	.13	.83
conceptueel	-.04	-.37	.04	-.10	.12	-.07	.01	.34
vindingrijk	.25	-.46	.14	-.02	.64	.18	.04	.02
veranderingsgezind	.05	-.10	-.09	-.03	.84	-.06	-.16	-.03
zelfsturend	-.02	-.03	-.04	.01	-.01	.84	-.19	-.02
prestatiegericht	.15	-.14	-.11	.83	-.06	.01	-.03	.03
competitief	.41	-.30	-.15	.47	-.12	.14	-.26	.24
gedreven	-.08	.27	-.37	.52	.15	-.20	-.13	.02

Tabel 8. Constructvaliditeit Shapes (1): Factoroplossing in relatie tot Great Eight

De acht factoren zijn gemakkelijk interpreteerbaar volgens het Great Eight model:

- | | |
|-----------------------------|-----------------------------|
| 1. Need for Power & Control | 5. Openness |
| 2. Agreeableness | 6. Neuroticism |
| 3. Extraversie | 7. Conscientiousness |
| 4. Need for Achievement | 8. Analysing & Interpreting |



Een tweede onderzoek naar de constructvaliditeit van Shapes maakt gebruik van de OPQ32i. Hieronder de correlaties (geen decimaalpunt) tussen de 18 x 32 dimensies.

Shapes OPQ32i	Directing	Persuasive	Socially confident	Sociable	Agreeable	Behavioural	Prudent	Focused on results	Playful	Conscientious	Analytical	Conceptual	Imaginative	Open for change	Autonomous	Achieving	Competitive	energetic
Persuasive	44**	45**	28*	30*	-33*	-04	11	13	-04	-26*	01	23	33*	23	07	10	33*	13
Controlling	53**	36**	30*	24	-15	-21	14	12	-14	-25*	13	13	17	24	12	21	16	05
Outspoken	33*	-04	24	07	43**	-08	-12	13	01	14	10	17	33*	03	29*	07	17	-10
Independent Minded	12	03	24	11	14	03	13	24	11	-24	05	20	24	10	40**	05	05	14
Outgoing	18	27*	30*	54**	18	13	-20	-23	09	05	-24	-10	11	14	-11	10	19	13
Affiliative	-20	-43**	09	34*	44**	22	-24	-19	-08	20	03	-22	-33*	-22	-29	-02	-18	07
Socially Confident	34*	44**	52**	33*	13	-10	10	10	03	-16	-13	11	32*	14	-13	11	17	12
Modest	-22	-29*	-12	-24	34*	11	11	12	14	17	22	14	-15	13	13	-11	-26*	07
Democratic	03	20	18	31*	40**	11	13	-19	14	23	05	-14	-11	03	-23	07	-25*	05
Caring	-10	11	16	26*	32*	12	-02	-10	09	13	09	03	13	03	13	-09	-18	-09
Data Rational	13	12	-13	14	-14	13	26*	20	22	13	44**	17	09	13	13	03	13	10
Evaluative	07	-10	14	-02	-24	-13	42**	20	20	11	40**	26*	11	-06	21	10	-06	-31*
Behavioural	-18	-11	05	03	21	43**	22	-24	-21	09	05	13	-04	05	05	-06	-20	-26*
Traditional	-31*	07	11	-11	23	03	-10	-06	14	24	11	-13	-32*	-34*	12	03	-04	18
Conceptual	12	-10	13	-19	-22	05	27*	05	10	-22	19	53**	16	23	10	05	05	-20
Innovative	26*	23	16	-09	-27*	-11	34*	07	01	-20	16	44**	36**	30*	09	07	23	-02
Variety Seeking	13	18	11	19	-04	01	11	12	-12	-20	05	14	32*	43**	09	05	05	-24
Adaptable	-06	33*	09	13	34*	14	07	-12	-16	11	-13	14	-24	19	-23	24	01	07
Forward Thinking	10	03	-08	-17	14	07	17	13	26*	24	23	28*	03	05	22	07	05	11
Detail Conscious	-04	-16	11	-10	11	-11	05	10	-04	09	21	-13	-19	13	07	09	05	14
Conscientious	-34*	-16	11	07	16	10	05	15	-14	43**	07	10	12	01	13	24	-18	01
Rule Following	-23	-31*	11	07	31*	-14	24	13	03	43**	14	-08	14	-10	-17	14	07	19
Relaxed	-06	11	09	07	-17	01	-22	-16	03	-25*	14	10	-24	01	-21	-17	13	27*
Worrying	07	19	-13	-21	16	09	11	13	07	20	05	14	14	-28*	23	14	23	-17
Tough Minded	24	09	11	-09	-09	12	07	14	11	-22	03	10	-11	03	19	09	24	27*
Optimistic	09	13	22	03	-14	-14	-14	-08	13	-09	12	-14	22	20	11	09	03	11
Trusting	-23	22	13	21	22	-14	-06	-16	01	-08	-02	03	14	17	-16	12	03	05
Emotionally Controlled	19	-18	07	-21	17	14	23	24	13	19	07	-09	10	-13	13	12	09	21
Energetic	14	12	-09	11	-13	10	12	25*	14	-11	09	03	-10	05	-11	24	21	24
Competitive	36**	36**	21	-15	-16	09	-04	12	23	-24	-04	14	22	07	07	23	42**	10
Achieving	34*	20	03	-20	-04	-20	13	33*	12	-11	19	-04	03	22	18	43**	22	05
Decisive	05	12	10	-24	-19	07	03	27*	07	17	09	12	20	20	30*	22	05	18

Tabel 9. Constructvaliditeit Shapes (2): Correlaties tussen Shapes en OPQ32i



Uit beide onderzoeken blijkt de constructvaliditeit van Shapes. De correlatiepatronen zijn volgens verwachting en de data clusteren na een factor analyse tot een acht factoren model met bekende validiteit.

D.2 Predictieve validiteit

Een voorbeeld van een onderzoek naar de predictieve validiteit van instrumenten uit de Scales serie volgt hieronder. In 2003-2004 werd een steekproef van managers en projectleiders samengesteld. De volgende gegevens zijn bekend:

- Zelfbeschrijving op basis van competenties (N = 194)
- Beoordeling door managers op basis van competenties (N = 152)
- Scores op Scales numeriek en verbaal.

Er werden verbanden verwacht tussen de Performance scores op beide Scales instrumenten en de volgende competenties:

- Analyse en oordeel
- Professionele expertise

Tevens werden correlaties verwacht tussen de Accuratesse scores en twee andere competenties:

- Uitvoering (Execution)
- Systematische aanpak

(voor definities zie elders).

De resultaten staan in onderstaande tabellen:

	Performance verbaal ⁴	Performance numeriek
Analyse en oordeel (zelf ⁵)	.38 **	.43 **
Analyse en oordeel (manager ⁶)	.36 **	.39 **
Professionele expertise (zelf)	ns	.27 *
Professionele expertise (manager)	ns	.22 *

Tabel 10. Predictieve validiteit Scales verbaal en numeriek (1)

⁴ * = significant op $p < 0.05$

** = significant op $p < 0.01$

⁵ Zelfbeoordeling

⁶ Beoordeling door manager

	Accuratesse verbaal ⁷	Accuratesse numeriek
Uitvoering (zelf)	.21 *	ns
Uitvoering (manager)	ns	.23 *
Systematische aanpak (zelf)	.32 **	.36 **
Systematische aanpak (manager)	.29 *	.31 **

Tabel 11. Predictieve validiteit Scales verbaal en numeriek (2)

Conform verwachting zijn de correlaties met de zelfbeoordelingen hoger. Verder worden er op een aantal competenties significante verbanden gevonden met relevante competenties. Dit is een duidelijke indicatie van de predictieve validiteit van beide instrumenten uit de Scales serie. Meta-analytisch onderzoek (waarbij data uit meerdere studies worden gecombineerd) wijst uit dat cognitieve capaciteitentests tot de beste voorspellers behoren (Schmidt & Hunter, 1998). Ook dit validiteitonderzoek is daarmee in overeenstemming.

Ook naar de vragenlijsten is validiteitonderzoek in predictieve zin gedaan. Een steekproef had de volgende kenmerken:

- Shapes vragenlijst ingevuld door zelf + Zelfbeoordeling op basis van competenties (N = 179)
- Beoordeling door managers op basis van competenties (N = 113)
- Beoordeling door ondergeschikten op basis van competenties (N = 265)
- Beoordeling door collega's op basis van competenties (N = 204)
- Beoordeling door anderen op basis van competenties (N = 215)

In onderstaande tabel staan de correlaties tussen Shapes en de beoordelingen door de vijf andere groepen (Z = Zelf, M = Manager, O = Ondergeschikten, C = Collega's en A = Anderen). Hierbij is een vertaalslag gemaakt van de Shapes dimensies naar een competentiemodel op basis van een zogenaamde 'mapping'. De beoordelingen van het werkgedrag zijn gebaseerd op de competenties uit het cut-e model.

⁷ * = significant op $p < 0.05$
 ** = significant op $p < 0.01$



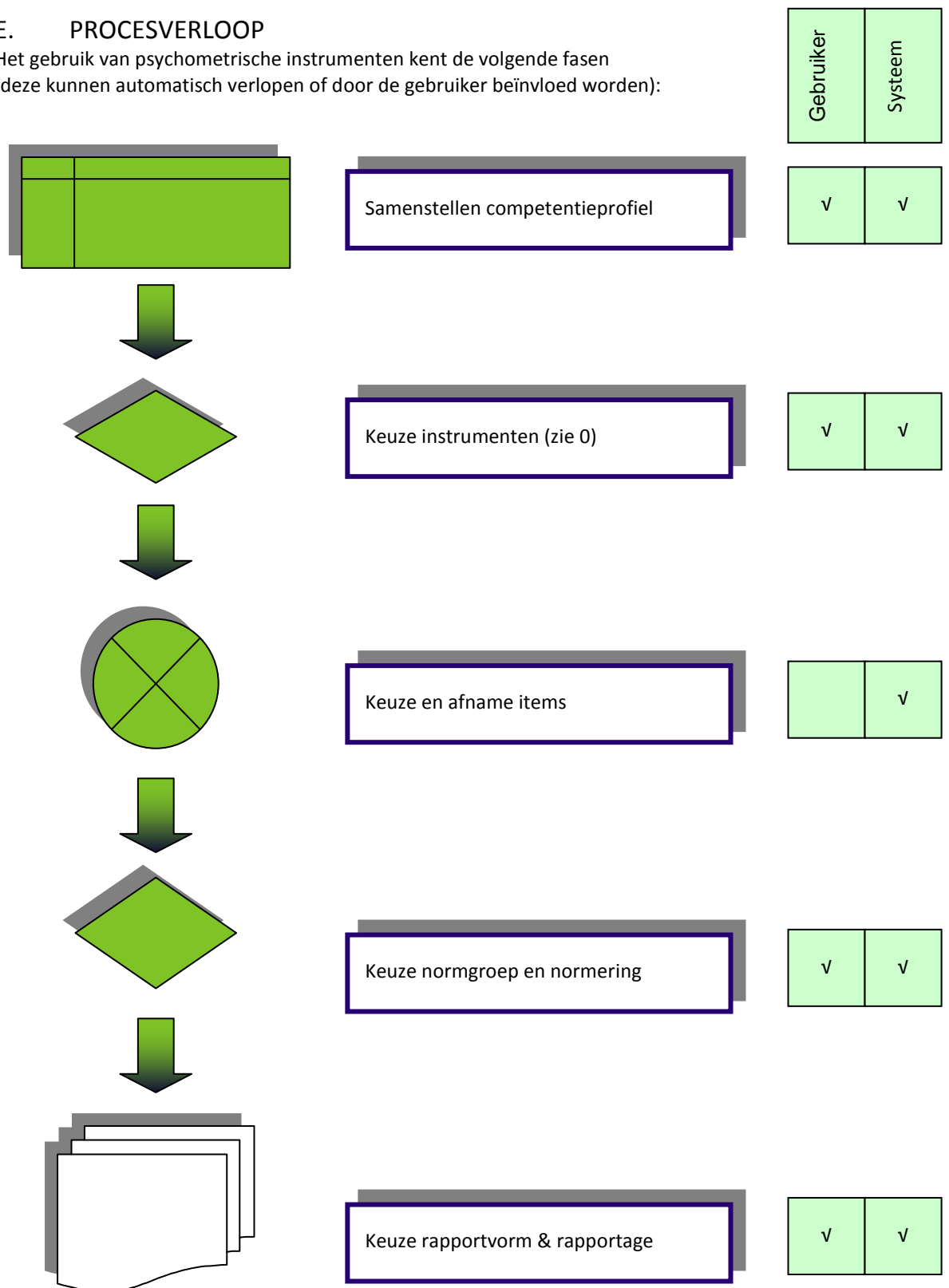
Competency	Z	M	O	C	A
Vision & Strategy	.71**	.37**	.23*	.21*	.18*
Initiative & Responsibility	.76**	.39**	.28*	.23*	.15*
Business Development	.63**	.34**	.21*	ns	ns
Bottom line focus	.74**	.41**	.31**	.17*	.19*
Influence	.69**	.37**	.29**	.26*	.21*
Networking	.73**	.26*	.17*	.25*	ns
People management	.72**	.31**	.27*	.21*	.16*
People development	.69**	.29**	.18*	ns	.18*
Organizational awareness	.61**	.33**	ns	.22*	ns
Execution	.69**	.39**	.19*	.25*	.21*
Systematic approach	.59**	.28*	ns	.18*	.16*
Steadiness	.63**	.38**	.23*	.24*	ns
Analysis & Judgement	.60**	.29**	.19*	.26**	.20*
Professional expertise	.63**	.21*	ns	.17*	ns
Innovation	.73**	.31**	.28**	.20*	.22*
Effective communication	.75**	.35**	.26**	.25*	.25*
Constructive teamwork	.71**	.36**	.22*	.16*	ns
Self-development	.67**	.22*	.17*	ns	.17*

Tabel 8. Predictieve validiteit Shapes

Ten eerste bestaan er sterke correlaties tussen Shapes en de Zelfbeschrijving van gedrag (de Z-kolom). Dit is een stevige indicatie dat Shapes inderdaad gedrag voorspelt. Ook de correlaties tussen Shapes en de oordelen van zowel Managers, Ondergeschikten en Collega's zijn grotendeels significant. De laagste correlaties bestaan met de oordelen van Anderen. Een verklaring is dat deze heterogene groep het minst in staat is geweest het actuele werkgedrag van de betrokkenen te beoordelen. Al me al bestaan er duidelijke verbanden tussen Shapes en (de beoordeling van) werkgedrag.

E. PROCESVERLOOP

Het gebruik van psychometrische instrumenten kent de volgende fasen (deze kunnen automatisch verlopen of door de gebruiker beïnvloed worden):



Figuur 1. Procesverloop inzet instrumenten



F. SAMENSTELLEN TESTPROGRAMMA

Het samenstellen van het competentieprofiel c.q. de testbatterij kan binnen HRorganizer.com op drie manieren plaatsvinden, namelijk via:

- Het maken van een functieprofiel
- Het kiezen van competenties
- Het kiezen van instrumenten

Wanneer gekozen wordt voor een van de twee eerste opties, dan biedt het systeem hulp bij het samenstellen van de testbatterij. Dit gebeurt als volgt:

F.1 Keuze van instrumenten

Op basis van expertkennis is binnen het systeem een matrix beschikbaar waarin de relatie tussen scores van tests, dimensies van vragenlijsten en de competenties is vastgelegd.

Op basis van het competentieprofiel en de gewenste werk- en denkniveau wordt een voorstel voor een verzameling tests gedaan. Hierbij worden de volgende uitgangspunten gehanteerd:

- Indien een functietypering als basis wordt gebruikt, krijgen de belangrijkste competenties een grotere invloed op de keuze van de instrumenten.
- Verder wordt er rekening gehouden met de ontwikkelbaarheid van competenties: moeilijk ontwikkelbare competenties hebben een grotere invloed dan eenvoudig ontwikkelbare competenties.
- Instrumenten worden geselecteerd op hun mate van bijdrage aan de meting van de te meten competenties (dit laatste noemen we de *relevantie*).
- Voor elke afzonderlijke competentie is bepaald in welke mate cognitie versus persoonlijkheid/motivatie/waarden een rol spelen. Ook deze verhouding is meegenomen in de berekening van de relevantie.



Het programma optimaliseert het vanuit de matrix aangereikte testprogramma als volgt:

- Van instrumenten die hetzelfde meten wordt de meest relevante gekozen. Hierbij wordt onder andere rekening gehouden met het vereiste werk- en denkniveau.
- Instrumenten die minder dan 20% aan de meeting bijdragen worden uit het testprogramma verwijderd.
- Een instrument wordt niet verwijderd als er een competentie is die uitsluitend door dit instrument wordt gemeten.
- Een instrument wordt ook niet verwijderd indien dit instrument een test is die als enige bepaalde cognitieve componenten in één of meerdere competenties meet.

Zodra het systeem het testprogramma wat betreft samenstelling heeft bepaald, wordt de afname volgorde berekend. De samenstelling van het testprogramma kan handmatig worden gewijzigd.

F.2 Afname volgorde

Bij het bepalen van de afname volgorde wordt rekening gehouden met de intensiteit van de instrumenten. De intensiteit bestaat uit drie componenten:

- Kwantitatieve belasting (breedte) - Power
- Tijdsdruk - Speed
- Kwalitatieve belasting (diepte) - Quality

Het systeem plaatst de instrumenten in een volgorde waarbij de intensiteit vanaf het begin wordt opgebouwd en eventueel naar het einde toe wordt afgebouwd.



F.3 Itemkeuze

Er zijn twee manieren om items te kiezen tijdens de afname. De traditionele manier is een lineaire afname met vaste items. Elk kandidaat krijgt dezelfde items in dezelfde volgorde voorgelegd. Dit heeft twee nadelen:

- De antwoorden op de items kunnen bekend worden of er kan worden afgekeken
- De testlengte kan hinderlijk worden

Bij het cut-e instrumentarium worden twee andere vormen bij de keuze van items gebruikt. Bij cognitieve capaciteitentests wordt gebruik gemaakt van item generatie. Bij de vragenlijsten (shapes, views) wordt gebruik gemaakt van een adaptief algoritme, dat wil zeggen er wordt – op basis van gegeven antwoorden - afgeweken van de standaard lineaire volgorde. In beide gevallen krijgen kandidaten een ‘test op maat’.

F.3.1 Item generatie

Dit wil zeggen dat op het moment van afname een item wordt gemaakt op grond van een aantal regels (algoritmen). Deze regels bepalen welke typen items kunnen voorkomen en zorgen er voor dat de items vergelijkbaar zijn. De regels hebben bovendien tot functie de moeilijkheidsgraad van de voor te leggen items te standaardiseren. Elke kandidaat krijgt dus een unieke test voorgelegd. Binnen de instrumenten van HRorganizer.com worden verschillende soorten regels gebruikt, dit afhankelijk van het soort test. Voor alle Scales instrumenten bestaan technische documenten (Preuss, 2002 – 2010) waarin precies beschreven wordt hoe de items gegenereerd worden. Voor het dagelijks gebruik is dit echter minder relevant. Wel relevant is de vraag naar equivalentie.

F.3.2. Equivalentie

Equivalentie betekent de mate waarin twee instrumenten met dezelfde meetpretentie in psychometrisch opzicht gelijk aan elkaar zijn. Een bekend voorbeeld is een papieren versie en een gecomputeriseerde versie van eenzelfde instrument (Schoonman, 1992, 1995). Bij de instrumenten binnen HRorganizer.com die gericht zijn op cognitieve capaciteiten en waarbij gebruik gemaakt wordt van item generatie is sprake van duizenden 'parallel tests'. De vraag naar equivalentie is dan ook aan de orde en is het best te beantwoorden door te kijken naar de test-hertest betrouwbaarheid.

Wanneer mensen op verschillende tijdstippen twee keer dezelfde test afleggen kan de test-hertest betrouwbaarheid (ook wel stabiliteit genoemd) berekend worden. In het geval van de hier bedoelde instrumenten is dat gebeurd waarbij de deelnemers – als gevolg van item generatie – twee keer een 'parallel' test hebben gemaakt waarbij dus verschillende items zijn voorgelegd (het interval tussen de twee afnames was een week). In onderstaande tabel staan de hertest-betrouwbaarheden – in termen van correlatie - voor een aantal instrumenten:

Instrument / score	N	M	SD	Hertest-betr.
verbaal				
Performance	84	13.9	4.9	.81
Snelheid	84	27.6	6.9	.53
Precision	84	75.3	8.0	.84
numeriek				
Performance	84	17.1	5.0	.83
Snelheid	84	27.4	6.5	.65
Precision	84	88.3	6.2	.79
stm				
Performance	116	59.6	6.3	.82
Reactiestijl	116	39.7	3.1	.86
Accuratesse	116	87.6	4.8	.81
cls				
Performance	91	63.4	8.7	.84
Snelheid	91	9.6	1.8	.73
Accuratesse	91	79.6	10.8	.91
eql				
Performance	197	12.3	2.7	.83
Snelheid	197	11.9	1.5	.69
Accuratesse	197	82.7	7.4	.82
mt				
Performance	186	108.4	17.9	.81
Vangen	186	56.5	1.8	.83
Rekenen	186	19.2	3.0	.74
Checken	186	21.7	1.7	.72
Instabiliteit	186	29.3	13.5	.71

Tabel 9. Test-hertestbetrouwbaarheden Scales instrumenten

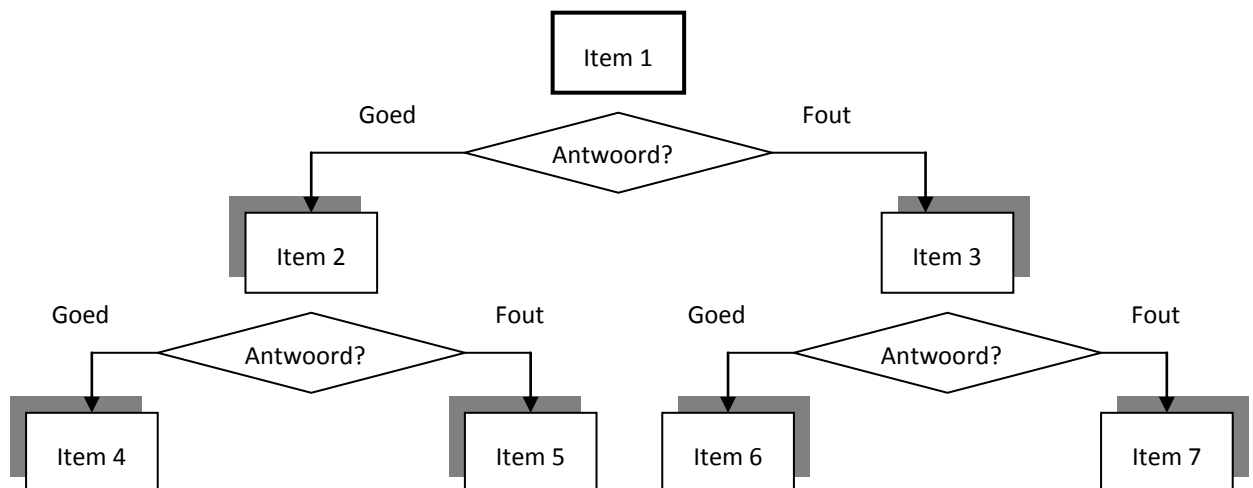
Uit de tabellen blijkt een hoge test-hertestbetrouwbaarheid (rond 0.8). Voor de scores op Snelheid geldt in een aantal gevallen een iets lagere betrouwbaarheid. De hoge correlaties wijzen op equivalentie van testversies waarbij verschillende items zijn gegenereerd.

F.4 Adaptief testen

Er zijn twee hoofdvormen van adaptief testen Fixed en Flexibel.

F.4.1. Branching

Bij Fixed adaptief testen (branching) is sprake van een zogenaamde itemboom. Hierin is vastgelegd welke volgorde bij de keuze van items moet worden aangehouden gegeven de antwoorden van de kandidaat. Bijvoorbeeld:



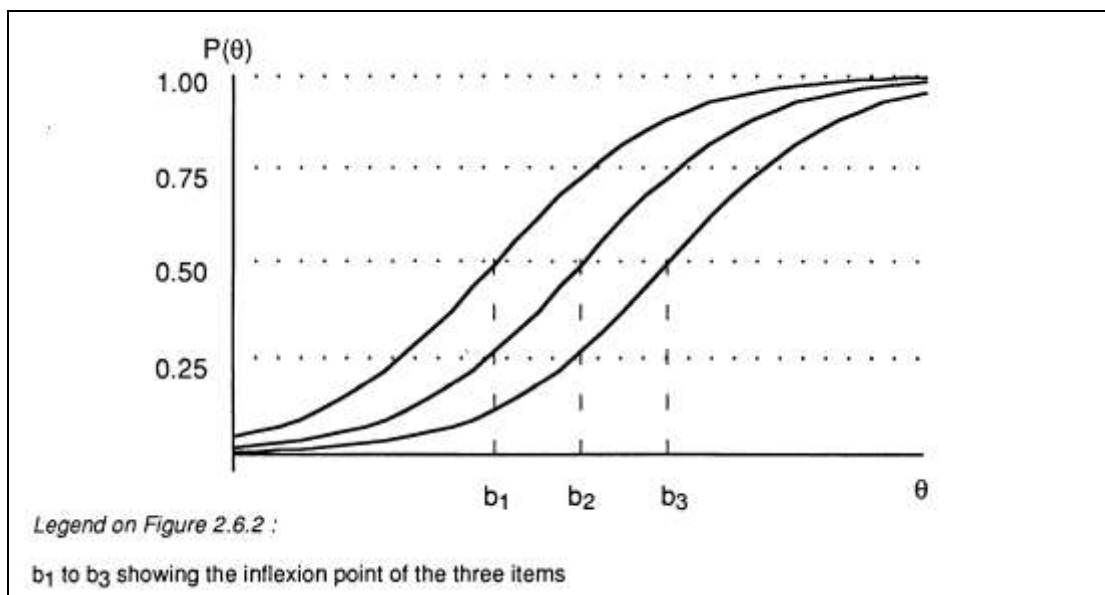
Figuur 2. Branching als item selectie

Het nadeel van deze manier van testen is dat er erg veel items nodig zijn ($2^x - 1$, waarbij x het aantal af te nemen items is). Bij een test van 10 items is dit dus al $2^{10} - 1 = 1023$ items. Een samenhangend nadeel is dat de meeste items maar weinig gebruikt zullen worden. Bij een test van drie items is dit $3/7 = 42\%$, bij een test van tien items nog geen 1%.

F.4.2. Flexibel

Bij Flexibel adaptief testen kan gebruik gemaakt worden van Item Response Theory (IRT). Deze opvolger van de Klassieke Test Theorie is wetenschappelijk superieur omdat veel nauwkeuriger wordt omschreven wat er gebeurt wanneer een persoon een item "ontmoet".

Volgens deze theorie ontstaat er een stochastisch proces waarbij de kans dat de persoon het goede antwoord geeft, afhangt van zijn vaardigheid en een of meer kenmerken van het item (bijvoorbeeld moeilijkheidsgraad, gokkans). In onderstaande afbeelding staan drie items. De grafieken (de zogenaamde Item Characteristic Curves) geven de kans aan dat de persoon het goede antwoord geeft, gegeven het niveau van de vaardigheid en het kenmerk(en) van het item. In onderstaande grafiek worden drie items weergegeven met van ieder item het kansverloop op een correct (of positief) antwoord (ontleend aan Schoonman, 1989). Op de Y-as staat de kans op een correct (of positief) antwoord. Op de X-as staat het vaardigheidsniveau.



Figuur 3. Item karakteristieke curven bij adaptief testen gebaseerd op IRT

Er kleven vier grote nadelen aan adaptief testen op basis van IRT. Ten eerste zijn de psychometrische eisen die aan de items gesteld worden hoog. Er moet sprake zijn van een goede 'modelpassing', dat wil zeggen dat de items zich in empirische zin 'gedragen' zoals het model voorschrijft. Ten tweede zijn grote aantallen kandidaten nodig om te kunnen komen tot stabiele parameterschattingen. Het derde nadeel is dat een deel van de gewonnen tijd (er worden in de regel minder items afgenomen dan bij een lineaire test) weer verloren gaat doordat de kandidaat steeds met relatief moeilijke items geconfronteerd wordt (immers de test past zich aan bij het niveau). Schoonman spreekt hier over de 'productiviteitsparadox'. Het laatste nadeel is dat kandidaten tijdens de afname steeds 'op hun tenen lopen'. De voldoening die ontstaat wanneer je als kandidaat zeker weet dat je een aantal items juist beantwoord hebt, verdwijnt bij adaptief testen. Dit is deels op te lossen door iets gemakkelijkere items (bijvoorbeeld met 60% in plaats van 50% kans op het goede antwoord) uit te kiezen, maar hierdoor gaat weer een deel van de winst verloren.



F.4.3. De gekozen oplossing

Bij de vragenlijsten binnen HRorganizer.com is op pragmatische gronden gekozen voor het idee van Adaptive Allocation of Content. Dit is een algoritme voor item keuze. Er zijn twee principes die hierbij een rol spelen:

- a. Partial Credit Scoring
- b. Weging van itemantwoorden

Items worden in groepjes (een 'blok') aangeboden (bij shapes in groepjes van drie) en de deelnemer kan aan elk item nul tot zes punten toekennen, afhankelijk van de mate waarin hij vindt dat het item zijn gedrag beschrijft. Dit is Partial Credit. Na zes blokken zijn alle 18 dimensies middels één item aan de beurt geweest. Deze $6 \times 3 = 18$ items wordt een sector genoemd (het aantal sectoren is gelijk aan het aantal items per schaal). Nadat de items uit de eerste sector zijn afgenomen, wordt de tweede sector samengesteld door de scores op de 18 dimensies te sorteren van hoog naar laag. Deze volgorde wordt aangehouden bij het uitkiezen en afnemen van de items uit de tweede sector. Items uit schalen die hoge scores hebben, komen bij elkaar in een blok te staan. Vanaf de tweede sector worden de gegeven antwoorden (nul tot en met zes) vermenigvuldigd met het blokgewicht (het totaal van alle scores in het blok).

Het gevolg van deze werkwijze is dat er snel maximale variatie ontstaat: de scores op de diverse schalen groeien snel uit elkaar. Het adaptieve karakter zit in het feit dat de kandidaat gedwongen wordt moeilijke keuzen te maken, immers de populariteit van de drie alternatieven per blok is ongeveer gelijk. Een bijkomend voordeel is dat door toepassing van deze methode de scores minder gevoelig zijn voor sociale wenselijkheid (zie ook Preuss, 2002).

F.5 Keuze normgroep en normering

Na afloop van een afname van een instrument worden een of meer ruwe scores berekend, bijvoorbeeld aantal goed en fout beantwoorde items, aantal overgeslagen items, de gemiddelde responstijd, enzovoort. Deze ruwe scores zijn op zich betekenisloos (wat zegt het als iemand op een test 13 goede antwoorden heeft). De ruwe scores krijgen betekenis na normering, dat wil zeggen het vergelijken met een groep mensen die de test eerder gedaan heeft. De keuze van de normgroep is een belangrijke en moeilijke beslissing. Binnen HRorganizer.com wordt de gebruiker hierbij ondersteund en er zijn diverse normgroepen beschikbaar. In onderstaande tabel een overzicht van de beschikbare normgroepen. Deze tabel is nu te gebruiken om te bekijken voor welke niveaus de instrumenten geschikt zijn.

Opmerking 1: plus (+) staat voor werkervaring, dit betekent niet dat de mensen uit deze groep gemiddeld ook hoger scoren op de tests.

Opmerking 2: over vrijwel de gehele linie is het zo dat jongere net afgestudeerde mensen hoger scoren op tests dan oudere mensen (met veel werkervaring).



Instrument	werknemer, VMBO opleiding (eqf 1)	zelfstandig werknemer, MBO opleiding (eqf 2)	beginnend beroepsbeoefenaar met een MBO of HBO opleiding (eqf 3)	zelfstandig beroepsbeoefenaar met een MBO of HBO opleiding (eqf 4)	superviserend beroepsbeoefenaar met een MBO of HBO opleiding (eqf 5)	junior professional of manager met een HBO of WO opleiding (eqf 6)	zelfstandig professional of manager met een HBO of WO opleiding (eqf 7)	senior professional of manager met een HBO of WO opleiding (eqf 8)
Shapes executive						senior management (202)	senior management (202)	senior management (202)
shapes		autonomous laborer (193)	junior employee (1264)	advanced employee (1375)	senior employee (1856)	junior expert/manager (1876)	advanced expert/manager (1164)	senior expert/manager (1527)
Shapes basic	starting laborer/semi-skilled (214)	autonomous laborer/semi-skilled (196)	junior employee (212)	advanced employee (162)	senior employee (158)			
views		autonomous laborer (216)	junior employee (1823)	advanced employee (721)	senior employee (808)	junior expert/manager (1245)	advanced expert/manager (1076)	senior expert/manager (1234)
scales numerical (consumer)				advanced employee (1129)	senior employee (1053)	junior expert/manager (2574)	advanced expert/manager (763)	senior expert/manager (774)
scales verbal (consumer)				advanced employee (532)	senior employee (476)	junior expert/manager (3274)	advanced expert/manager (786)	senior expert/manager (791)
scales cls				advanced employee (621)	senior employee (813)	junior expert/manager (3697)	advanced expert/manager (1849)	senior expert/manager (2561)
scales fx						junior expert/manager (836)	advanced expert/manager (276)	senior expert/manager (249)
scales clues			junior employee (871)	advanced employee (1389)		junior expert/manager (2364)		
scales spr			employee (347)	employee (347)	employee (347)	employee (347)	employee (347)	
scales eq1 (a)	starting laborer (469)	autonomous laborer (537)	junior employee (832)	advanced employee (346)	senior employee (287)	junior expert/manager (216)	advanced expert/manager (256)	
scales mt	starting laborer (286)	autonomous laborer (253)	junior employee (1242)	advanced employee (513)	senior employee (542)	junior expert/manager (1667)	advanced expert/manager (984)	
scales stm	starting laborer (263)	autonomous laborer (183)	junior employee (486)	advanced employee (161)	senior employee (173)	junior expert/manager (422)	advanced expert/manager (256)	
scales lct	general adult population (11064)	general adult population (11064)	general adult population (11064)	general adult population (11064)	general adult population (11064)	general adult population (11064)		
scales ix (a)	general adult population (432)	general adult population (432)	general adult population (432)	general adult population (432)	general adult population (432)			
scales lst (a)	general adult population (172)	general adult population (172)	general adult population (172)	general adult population (172)	general adult population (172)	general adult population (172)	general adult population (172)	
scales numerical admin	starting laborer/semi-skilled (186)	autonomous laborer/semi-skilled (163)	junior employee (179)	service selection (162)				
scales verbal admin	starting laborer/semi-skilled (241)	autonomous laborer/semi-skilled (274)	junior employee (548)	advanced employee (221)	senior employee (217)			

Tabel 10. Normgroepen in relatie tot werkniveau; European Qualification Framework (normgrootte tussen haakjes).



G. Conclusie

In dit overzicht is het instrumentarium binnen HRorganizer.com nader beschreven. De nadruk heeft hierbij gelegen op de volgende punten:

- Betrouwbaarheid (zowel interne consistentie als hertest)
- Validiteit (zowel construct als predictief)
- Item generatie
- Adaptief testen
- Normering

Gesteld kan worden dat de instrumenten goed onderzocht en erg innovatief zijn. Daarnaast is er sprake van hoge face-validity, grote normgroepen en een brede dekking. Ook technologisch zijn deze producten state-of-the-art (web based, flash player, centrale databases, enzovoort).



H. Referenties

Akkerman, A.E. (1984). De Leidse VrolijkheidsStemmings-Questionnaire, de LVSQ. Een alternatief voor therapie-evaluatie? *De Psycholoog*, 19, 4, 167-169

Evers, A., Van Vliet-Mulder, J.C. & Groot, C.J. (2000). *Documentatie van tests en testresearch in Nederland. Deel 1 en 2*. Assen: Van Gorcum

Kline, P. (1998). *The New Psychometrics. Science, Psychology and Measurement*. London: Routledge

Kurz, R. & Bartram, D. (2002). *Competency and individual performance: Modelling the world of work*. In I. Robertson, M. Callinan & D. Bartram (Eds.). *Organisational effectiveness: The role of psychology*. Chichester: Wiley

Preuss, A. (2000-2005). *Inside-serie. Serie technische documenten over alle cut-e instrumenten*. Hamburg: cut-e

Preuss, A. (2002). *ADALLOC. Adaptive Allocation of Consent*. Hamburg: cut-e

SHL (1999). *OPQ32 Manual and User's Guide*. Surrey, UK: SHL Group plc

Schmidt, F.L. & Hunter, J.H. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 24, 262-274

Schoonman, W. (1989). *An applied study on computerized adaptive testing*. Lisse: Swets & Zeilinger

Schoonman, W. (1992). Equivalentie tot elke prijs? *Psychologie & Computers*, 9, 3-7

Schoonman, W. (1995). Equivalentie: prettig maar irrelevant. *Psychologie & Computers*, 12, 56-58