



FAT-R

Fonemische Analyse Test Herziene versie

PSYCHOMETRISCHE EIGENSCHAPPEN
WHITE PAPER 2

White paper

FAT-R

Fonemische Analyse Test

Herziene versie

Psychometrische eigenschappen
White paper 2

Drs. M. Kerkmeer
Psychometrician, Pearson Assessment and Information B.V.

Drs G.I. Leemans
Product Developer, Pearson Assessment and Information B.V.

Januari 2016

PEARSON

Inhoud

1	Samenvatting	4
2	Inleiding	5
3	Normering	6
4	Betrouwbaarheid	10
5	Validering	13
6	Referenties	18

1 Samenvatting

De Fonemische Analyse Test FAT-R (De Groot, Van den Bos, & Van der Meulen, 2015) is een instrument voor fonemische analysevaardigheid en bestaat uit twee onderdelen: FoneemWeglating en FoneemVerwisseling. Daarnaast is er een Indexscore beschikbaar. De subtests en de FAT-R-Indexscore hebben zogeheten T-scores, waardoor de prestaties van kinderen adequaat te vergelijken zijn met hun leeftijdsgenoten. De FAT-R is genormeerd op $n=2134$ kinderen in de leeftijd van 7 tot en met 14 jaar uit Nederland.

De betrouwbaarheid is onder meer onderzocht via inter-itemrelaties, test-hertest- en interbeoordelaarsbetrouwbaarheid. De inter-itemrelatiebetrouwbaarheid was goed met λ 's tussen 0,75 en 0,94 voor de subtests en 0,82 en 0,95 voor de FAT-R-Indexscore. Ook de resultaten van het test-hertestonderzoek waren goed en onderbouwen de betrouwbaarheid van de FAT-R met een gemiddelde correlatie van 0,89 tussen de eerste en tweede afname. De interbeoordelaarsbetrouwbaarheid was goed met een gemiddelde kappa van 0,76 voor FoneemWeglating en 0,92 voor FoneemVerwisseling. De validiteit is onderzocht met neventests, klinische groepen en via voorspelling van criteria als opleidingstype, thuistaal en leesstoornis. De samenhang met de neventests was volgens verwachting. De FAT-R is in het algemeen een goede voorspeller voor leesstoornis.

2 Inleiding

De FAT-R (De Groot, Van den Bos, & Van der Meulen, 2014) is een instrument om bij normaal horende kinderen, in de leeftijd van 7 tot en met 14 jaar, de fonemische analysevaardigheid te onderzoeken.

Problemen met lezen en spellen zijn meestal de aanleiding om de FAT-R als test in te zetten. Een specifiek gebruiksdoel is dan ook om de (verklarende) diagnostiek van taal- en lees- en spellingstoornissen te ondersteunen, vooral bij (vermoedens van) dyslexie en comorbide stoornissen zoals de combinatie van SLI en dyslexie of ADHD en dyslexie. Ook bij het nagaan van de effecten van specifieke interventies op fonologisch gebied kan de FAT-R behulpzaam zijn.

De FAT-R is de opvolger van de FAT (Van den Bos, Lutje Spelberg & De Groot, 2010) en is ontwikkeld aan de Rijksuniversiteit Groningen. De FAT is een toonaangevend instrument om fonemische analysevaardigheden mee in kaart te brengen en met de FAT-R zijn er nieuwe normen, scoringsmethoden en betrouwbaarheid- en valideringsonderzoek beschikbaar.

Elke testafname moet verzorgd worden door een testleider. Bij het afnemen van de FAT-R heeft de testleider, naast het monitoren van de instructie, twee taken. Eerst moet bij een antwoord van het kind de timer van de computer gestopt worden en vervolgens moet de testleider beoordelen of het gegeven antwoord goed of fout is. Hierbij kan de testleider gebruik maken van de handleiding (De Groot, Van den Bos, & Van der Meulen, 2015); met name hoofdstuk 6 en 7 gaan over afname en interpretatie.

De scoring gebeurt in het scoringsprogramma van de testuitgever Pearson. Direct na de afname is het testrapport beschikbaar. De testleider kan bovendien het rapport openen en heeft hier de mogelijkheid om fout gescoorde items en overige foutieve testgegevens te herstellen.

3 Normering

Steekproeftrekking

De normen zijn in twee fases verzameld: eerst van 2005 tot 2009, deze steekproef bleek onvoldoende kwaliteit te hebben¹ en daarom is van 2011 tot 2012 bijverzameld.

Grootte

In totaal zijn 2341 afnames voor de normering gedaan. Er zijn 1860 afnames voor Foneem-Verwisseling, en 2293 afnames voor FoneemWeglatting. De groepsgrootte ligt tussen 170 en 381 per normgroep.

Representativiteit

De uiteindelijke steekproef van N=2341 is gecontroleerd op representativiteit voor de variabelen sekse, schooltype VO, regio, urbanisatiegraad, etnische herkomst en sociaal-economische status (via opleidingsniveau van de ouders). In onderstaande tabellen wordt een overzicht gegeven van de resultaten.

Tabel 1

Sekse per normgroep

Normgroep	Leeftijdsbereik	Meisje	Jongen	Totaal
7	6;6-7;5	204 (53.5%)	177 (46.5%)	381
8	7;6-8;5	204 (53.5%)	177 (46.5%)	381
9	8;6-9;5	162 (55.3%)	131 (44.7%)	293
10	9;6-10;5	110 (43.7%)	142 (56.3%)	252
11	10;6-11;5	85 (50.0%)	85 (50.0%)	170
12	11;6-12;5	143 (51.6%)	134 (48.4%)	277
13	12;6-13;5	143 (51.6%)	126 (46.8%)	269
14	13;6-14;5	152 (47.8%)	166 (52.2%)	318
FAT-R Totaal		1.203 (51.4%)	1.138 (48.6%)	2.341
Nederland		49.7%	50.3%	

In het algemeen is de steekproef representatief voor sekse. In een enkele groep (de 9-jarige en 10-jarige meisjes en jongens) is de afwijking iets hoger dan 5% maar door weging is dit gecorrigeerd.

¹ Voornamelijk vanwege de representativiteit; er was een oververtegenwoordiging van kinderen uit Oost en Noord en een ondervetegenwoordiging van West en Zuid. Dit leek niet veel effect op de scores te hebben

Tabel 2

Representativiteit schooltype voortgezet onderwijs

Normgroep	opleidingstype	N	% vmbo	streven	% VO	streven
13	vmbo-lwoo	12	10,8	10	6,1	5,5
	vmbo-bk/tl	99	89,2	90	50,3	49,5
	vmbo-totaal	111	100	100	56,4	55
	havo/vwo	86			43,6	45
	VO totaal	197			100	100
14	vmbo-lwoo	21	12,2	10	6,6	5,5
	vmbo-bk/tl	151	87,8	90	47,5	49,5
	vmbo-totaal	172	100	100	54,1	55
	havo/vwo	145			45,9	45
	VO totaal	318			100	100

De steekproefpercentages liggen dicht bij de streefcijfers. De afwijking bedraagt nergens meer dan 5%. Bovendien is door weging gecorrigeerd voor afwijkingen.

Tabel 3

Representativiteit voor regio en urbanisatiegraad

Urbanisatiegraad	Noord (%)	Oost (%)	West (%)	Zuid (%)	Totaal (%)
Zeer sterk en sterk stedelijk	15,1	33,7	69,6	25,5	37,3
streefpercentage	11,9	29,5	62,3	32,5	43,4
Matig, weinig en niet stedelijk	84,9	66,3	30,4	74,5	63,7
streefpercentage	88,1	70,5	37,7	67,5	56,6

In de regio's West en Zuid wijken de percentages meer dan 5% af van de streefpercentages. Voor de totale steekproef geldt dat er iets te veel 'matig, weinig en niet stedelijk' en te weinig '(zeer) sterk stedelijk' afnames zijn. Dit is echter gecorrigeerd door weging.

Voor de variabelen etnische herkomst en opleidingsniveau ouders (voor SES) is een ietwat andere benadering gebruikt. Bij het begin van het onderzoek waren daar namelijk geen gegevens over verzameld, dus er was te veel *missing data*. Daarom is het als volgt aangepakt. De Dienst Uitvoering Onderwijs (DUO, 2011) weet per klas of groep wat de zogeten rugzakjes van de leerlingen zijn, de gewichtenregeling. Deze wordt bepaald door de Centrale Financiën Instellingen (CFI, 2006) op de volgende indeling:

- 0.00: De ouder heeft een opleiding van drie of vier jaar mavo (c- of d-niveau), drie of vier jaar vmbo gemengde leerweg of theoretische leerweg genoten. Of de ouder heeft meer dan twee jaar havo of vwo, een mbo, hbo of universitaire opleiding gevolgd (of vergelijkbare niveaus).
- 0.30: De ouder heeft maximaal lbo/vbo, praktijkonderwijs of vmbo basis- of kader-beroepsgerichte leerweg gedaan. Of de ouder heeft maximaal twee jaar onderwijs in een andere schoolopleiding in het voortgezet onderwijs aansluitend op het basis-onderwijs gehad.
- 1.2: Een van de ouders heeft maximaal basisonderwijs (tot en met het 13e levensjaar) of (v)so-zmlk gehad (CFI, maart 2006).

Vervolgens zijn de schoolproporties vergeleken met de landelijke gemiddelden per regio per normgroep. In Tabel 4 is te zien dat de FAT-R-steekproef in het algemeen goed overeenkomt met de landelijke gegevens.

Tabel 4

Opleidingsniveau van de ouders, middels gewichtenregeling ('rugzakjes')

Normgroep		Regio												Totaal FAT-R / Landelijk		
		Noord			Oost			Zuid			West			o	o.3	1.2
		o	o.3	1.2	o	o.3	1.2	o	o.3	1.2	o	o.3	1.2			
6;6 - 7;5	%	83	10	7	87	9	4	71	13	16	onbekend			83 / 87	10 / 6	7 / 7
	n	31			51			21						103		
7;6 - 8;5	%	91	7	2	95	5	1	73	11	16	91	9	0	90 / 87	7 / 7	3 / 7
	n	212			34			23			43			312		
8;6 - 9;5	%	94	5	1	94	6	0	92	5	3	98	1	1	95 / 86	4 / 7	1 / 7
	n	144			76			21			53			294		
9;6 - 10;5	%	96	2	2	97	2	0	81	3	16	97	3	1	95 / 85	3 / 8	3 / 7
	n	139			25			23			65			252		
10;6 - 11;5	%	94	3	3	100	0	0	89	6	6	90	8	2	93 / 84	4 / 8	2 / 7
	n	50			38			22			60			170		
11;6 - 12;5	%	71	4	25	100	0	0	86	14	0	65	12	23	76 / 71	6 / 13	18 / 16
	n	101			40			19			53			213		
Totaal	%	90	5	6	95	4	1	80	11	10	88	6	5	89 / 86	6 / 7	5 / 7
	n	677			264			132			274			1.347		

Er is enige ondervertegenwoordiging van gezinnen met laag opgeleide ouders, maar in het algemeen stemt de steekproef goed overeen met de landelijke cijfers. Bovendien is door weging gecorrigeerd voor afwijkingen. Overigens geldt voor alle variabelen, dat de wegingsfactor nooit groter dan 2 is.

In Tabel 5 staan de percentages voor etnische herkomst van de steekproef. Ook hierbij is gebruik gemaakt van de gegevens van DUO/ CFI gebaseerd op onderwijstype en postcode van de respondenten.

Tabel 5

Representativiteit voor etnische herkomst per onderwijstype, gebaseerd op onderwijstype en postcodegebied

	Autochtoon (%)	Westers allochtoon (%)	Niet-westers allochtoon (%)
PO (N = 1.763)	86.1	7.4	6.5
VO (N = 564)	86.9	7.2	5.9
Nederland Totaal	86.5	7.6	5.9

Te zien is dat de steekproef representatief is voor de variabele etnische herkomst.

Standardscores

De normen zijn uitgedrukt in een T-scoreschaal, die aansluit bij andere instrumenten op het gebied van taal. Het gemiddelde van deze schaal is altijd 50 en de $SD=10$. Dit houdt in dat 68% van de kinderen een score heeft tussen 40 en 60, en dat 95% van de kinderen een score heeft tussen 30 en 70. Hiermee kunnen kinderen vergeleken worden met hun leeftijdgenoten, en kan de behandelaar zien of een kind voor die leeftijd gemiddeld presteert.

De oorspronkelijke FAT-2010 had normen gebaseerd op de zogenoemde Wechslerschaal met een gemiddelde van 10 en $SD=3$. Maar de gemiddelde ruwe scores van de FAT-R (voor FoneemWeglating tussen de 16 en 64, voor FoneemVerwisseling tussen de 58 en 304) rechtvaardigen een meer verfijnd scoresysteem. Een zogenaamde IQ-schaal met een gemiddelde van 100 en $SD=15$ zou niet erg zinvol zijn geweest, omdat op die manier de ruwe scores te veel worden opgeblazen en de verfijning een schijnnaauwkeurigheid biedt. Kortom, het systeem is verfijnder dan de Wechslerschaal en we blazen het aantal scoreklassen niet onnodig hoog op.

Conclusie normering

De normen zijn gebaseerd op een voldoende grote en representatieve steekproef van Nederlandse kinderen van 7 tot en met 14 jaar. Eventuele afwijkingen van streefcijfers zijn gecorrigeerd door weging. De gekozen normschaal sluit aan bij het mogelijke bereik van de ruwe scores.

4 Betrouwbaarheid

Inleiding

In het normeringsonderzoek van de FAT-R is ook de betrouwbaarheid onderzocht. Dit gebeurde op de volgende vier manieren:

- Inter-itemrelaties
- Paralleltestmethode
- Test-hertest
- Interbeoordelaars

In dit whitepaper zullen we de inter-itemrelaties, de test-hertestbetrouwbaarheid en de interbeoordelaarsbetrouwbaarheid behandelen. Voor een beschrijving van de paralleltestmethoden verwijzen we naar de FAT-R handleiding (paragraaf 4.3, pag. 65-68).

Interitemrelaties

De interitemrelaties zijn onderzocht op de ongewogen dataset. De gebruikte maten zijn coëfficiënt alfa en Guttman's lambda-2. Coëfficiënt alfa geldt als ondergrens van de betrouwbaarheid; lambda geeft een schatting van de betrouwbaarheid die dichterbij de echte betrouwbaarheid ligt (Evers, Lucassen, Meijer & Sijsma, 2010). In Tabel 6 staat een overzicht van de waardes per normgroep.

Tabel 6

Betrouwbaarheid (coëfficiënt alfa en Guttman's lambda-2) per normgroep

Groep	FoneemWeglating			FoneemVerwisseling			FAT-R-index		
	N	alfa	labda2	N	alfa	labda2	N	alfa	labda2
7	380	0,75	0,82	-*	-*	-*	380	0,75	0,82
8	380	0,73	0,80	290	0,82	0,89	290	0,86	0,90
9	293	0,75	0,82	288	0,82	0,90	288	0,88	0,92
10	252	0,75	0,82	251	0,83	0,91	251	0,87	0,92
11	170	0,77	0,84	170	0,81	0,89	170	0,87	0,91
12	277	0,68	0,75	277	0,83	0,91	277	0,85	0,90
13	269	0,73	0,80	269	0,83	0,91	269	0,87	0,91
14	271	0,79	0,87	318	0,86	0,94	271	0,90	0,95
Totaal	2292	0,75	0,82	1863	0,83	0,91	2196	0,87	0,92

* De subtest FoneemVerwisseling is bij de 7-jarigen niet afgenomen. De betrouwbaarheidsmaten voor FoneemWeglating van deze leeftijdsgroep gelden ook voor de FAT-R-Index.

De waardes voor alfa zijn voldoende tot goed, en de waardes voor labda zijn goed tot zeer goed.

Test-hertest

De stabiliteit van de FAT-R-scores FoneemWeglating, FoneemVerwisseling en de FAT-R-Index is onderzocht onder in totaal 92 respondenten. De test-hertestgroep was vrijwel gelijk verdeeld over geslacht en normgroep. De deelnemers waren allemaal uit de regio Noord afkomstig. In Tabel 7 staat een overzicht van de resultaten voor de FAT-R-Index.

Tabel 7

Gegevens Test-hertest

Normgroep	vrouw	man	Eerste afname		Tweede afname		r	rho
			gem	SD	gem	SD		
7	-	-	-	-	-	-	0,87	0,74
8	7	5	49,9	(8,9)	56,0	(8,5)	0,93	0,90
9	7	5	45,4	(14,1)	48,7	(14,6)	0,93	0,75
10	5	8	44,4	(10,2)	49,0	(11,7)	0,94	0,82
11	4	6	51,6	(9,5)	55,4	(10,2)	0,92	0,81
12	8	6	50,2	(4,3)	55,2	(6,8)	0,68	0,76
13	4	4	50,9	(8,0)	56,6	(4,0)	0,91	0,86
14	7	7	52,8	(8,2)	59,4	(8,3)	0,73	0,77
Totaal	42	41	49,2	(9,5)	54,3	(10,2)	0,89	0,84

In het algemeen zijn de resultaten goed. Bij de oudste groep en de 12-jarigen is de correlatie wat lager, maar de resultaten zijn nog steeds voldoende te noemen.

Interbeoordelaarsbetrouwbaarheid

Een aantal testleiders was gevraagd om een elke tiende afname met een voice-recorder op te nemen. Hierdoor kon een tweede beoordelaar de accuratesse van het antwoord dat het kind gaf beoordelen. Deze procedure leverde uiteindelijk 31 bruikbare en onafhankelijk gescoorde gepaarde waarnemingen op. De respondenten uit deze interbeoordelaarsgroep waren evenredig verdeeld over geslacht en normgroep. In Tabel 8 staan de Kappa's gerapporteerd als maten van de interbeoordelaarsbetrouwbaarheid per item. Voor de items 1, 9 en 10 van FoneemWeglating zijn geen Kappa's vermeld omdat deze items in de steekproef constant gescoord werden. Kappa als associatiemaat kan niet berekend worden als een der beide beoordelaars constante scores heeft op deze items. Deze constante scores (c) waren in alle gevallen 'goed'.

Tabel 8

Kappa's voor FAT-R (FoneemWeglating en FoneemVerwisseling) per item

	FoneemWeglating	FoneemVerwisseling
Item 1	c	1
Item 2	0,59	0,87
Item 3	0,76	1
Item 4	0,84	0,71
Item 5	0,63	0,92
Item 6	0,63	1
Item 7	0,74	0,77
Item 8	c	1
Item 9	c	1
Item 10	0,82	1
Item 11	0,83	0,82
Item 12	0,91	1
Gemiddelde	0,76	0,92

FoneemVerwisseling heeft een (zeer) goede interbeoordelaarsbetrouwbaarheid, en voor FoneemWeglating is deze voldoende tot goed.

Conclusie Betrouwbaarheid

Al met al is de FAT-R een betrouwbaar instrument voor het meten van fonemische analysevaardigheid in de doelgroep. De betrouwbaarheid van de FAT-R is onderzocht met verschillende betrouwbaarheidsmaten, met steekproeven van aanvaardbare grootte en kwaliteit. Omdat de FAT-R vooral ingezet wordt voor minder belangrijke beslissingen, zijn de gevonden waarden, vooral voor de FAT-R-Index, hoog genoeg om de betrouwbaarheid goed te noemen. Daarom wordt vooral het gebruik van de FAT-R-Index aanbevolen, omdat deze van de drie gestandaardiseerde FAT-R-scores het meest betrouwbaar is.

5 Validering

Inleiding

Er is veel valideringsonderzoek verricht met de FAT-R. De interne structuur is onderzocht met factoranalyse, correlaties tussen de ruwe vaardigheidsscores, en tussen de FAT-R-Index en de FAT-2010-Index. De convergente en divergente validiteit is onderzocht met de EMT (1999), Klepel (1994), CB&WL (2010), PI-dictee (2000), SVS (2001), DLE-test spellen (2001) en de WISC-III-nl (2005). De begripsvaliditeit is verder onderzocht onder de groepen met een leesstoornis, met een diagnose dyslexie, met ADHD, naar opleidingstype, thuistaal, en met leeftijdsgroepen. Verder is de voorspellende waarde van de FAT-R onderzocht met een aantal klinische steekproeven. Hieronder zal een selectie worden besproken.

Factorstructuur

De auteurs hebben een confirmatieve factoranalyse gedaan om het best passende model te achterhalen. De volgende drie modellen zijn vergeleken:

1. Een twee-factorenmodel met FW en FV als latente factoren;
2. Een meer spaarzaam model met één latente factor voor fonemische vaardigheid;
3. Een drie-factorenmodel, met naast FW en FV nog Aandachtsspanne vanwege te verwachten concentratieverlies aan het eind van FV. De laatste drie items van FV zijn ondergebracht in de factor Aandachtsspanne.

In Tabel 9 zijn de uitkomsten weergegeven van genoemde analyses. Het spaarzame één-factor-model geeft de data het minst goed weer, gevolgd door het twee-factorenmodel, dat marginaal onvoldoende scoort met CFI/TLI-waarden van respectievelijk .85 en .84 en een RMSEA-index van .05. De beste model-fit volgt uit het drie-factorenmodel, waarvoor de CFI- en TLI-waarden duiden op een goede fit (.95) en ook de RMSEA-index goed te noemen is (.03).

Tabel 9

Resultaten confirmatieve factoranalyse

Maat	1 factor	2 factoren	3 factoren
CFI	0,79	0,85	0,95
TLI	0,77	0,84	0,95
RMSEA	0,06	0,05	0,03

Verschillen tussen groepen

Thuistaal

Het is aannemelijk dat een andere thuistaal dan Nederlands (zoals Berber, Turks en dergelijke) kinderen enigszins belemmert bij een Nederlandstalige test van hun fonemische vaardigheden. Dit doet zich vermoedelijk meer voor bij jonge kinderen dan bij oudere kinderen met een andere thuistaal. Deze laatste groep zal meer gelegenheid hebben gehad gewend te raken aan de Nederlandse fonologie, dan de relatief jongste groep. We voorspellen daarom dat de FAT-R-Index tussen de twee thuistaalcondities differentieert en dat het effect bij jonge kinderen groter is dan bij de oudere kinderen.

Om deze voorspellingen te toetsen hebben de auteurs Thuistaal Nederlands onderscheiden van Thuistaal Anders. De groep *Jong* wordt bevat kinderen tot en met 10 jaar, en de groep *Oud* kinderen van 11-14 jaar. In Tabel 10 staan de gemiddelde FAT-R-Indexscores.

Tabel 10

Gemiddeld RAT-R-Indexscores per thuistaal, naar leeftijdsklasse

	Jong	Oud
Nederlands	47.9 (10.8) n=338	48.8 (9.5) n=441
Anders*	39.5 (10.7) n=33	48.8 (9.4) n=52

* Deze categorie omvat alle talen anders dan Nederlands die overwegend in het gezin worden gesproken, inclusief het Fries.

De scores werden in een 2 x 2-variantieanalyse geanalyseerd. De F-waarden voor de hoofd-effecten voor thuistaal en leeftijd zijn respectievelijk $F=12.7$, $p<.001$, $\eta^2=.02$; $F=18.7$, $p<.001$, $\eta^2=.02$. De F-waarde van het interactie-effect is $F=12.9$, $p<.001$, $\eta^2=.02$.

Dit resultaat komt overeen met onze verwachting. De groep *Jong* toont een duidelijk verschil in fonemische analysevaardigheid. In deze groep scoort de categorie Anders significant lager dan Nederlands, terwijl er bij de groep *Oud* geen verschil op gemiddelde tussen de thuistalen is.

Leesstoornis

Er zijn twee gangbare criteria voor een leesstoornis: een score van 2 SD onder het gemiddelde, of een score op of onder het 10^e percentiel (dit komt overeen met 1,3 SD onder het gemiddelde). Dit komt overeen met een T-score van respectievelijk 30 en 37 op de EMT en de Klepel, die nu veelal ingezet worden in dyslexieonderzoek.

Deze beide afkappunten zijn onderzocht voor de FAT-R, om na te gaan of kinderen met een leesstoornis door de FAT-R eruit gefilterd worden. De leesstoornis is vastgesteld door de score op het onderdeel Woordlezen van de CB&WL (Bos, van den & Lutje Spelberg, 2010). De aantallen volgens beide indelingen staan in Tabel 11.

Tabel 11

Aantallen kinderen met en zonder leesstoornis, volgens twee criteria

	2 SD beneden gemiddeld: WL-score \leq 30		Op of onder 10 ^e percentiel: WL-score \leq 37	
	Aantal	Percentage	Aantal	Percentage
Geen leesstoornis	1441	96.6	1332	89.3
Wel leesstoornis	50	3.4	159	10.7

De interpretatie van de resultaten is als volgt:

	Test: wel leesstoornis	Test: geen leesstoornis
Werkelijkheid: wel leesstoornis	Sensitiviteit	Vals-negatief
Werkelijkheid: geen leesstoornis	Vals-positief	Specificiteit

De eerste maat is sensitiviteit: het vermogen van een test (de FAT-R Indexscore in dit geval) om mensen die in werkelijkheid een stoornis hebben, er inderdaad uit te pikken. Als er volgens de FAT-R niets aan de hand is, terwijl mensen in werkelijkheid wel een stoornis hebben, is er sprake van een vals-negatief.

De tweede maat is specificiteit. Dit houdt in dat mensen die in werkelijkheid geen stoornis hebben, ook niet als zodanig gelabeld worden. Als mensen zonder stoornis wél in de klinische groep vallen volgens de FAT-R, is er sprake van een vals-positief, en is de specificiteit niet zo goed. Bij screeningsinstrumenten wordt meestal de sensitiviteit belangrijker gevonden dan de specificiteit, omdat je zo veel mogelijk mensen wilt detecteren die mogelijk hulp of zorg nodig hebben.

Een laatste maat is de AUC (Area Under the Curve). Dit is een verhouding van de vals-positieve waarden afgezet tegen de sensitiviteit van een test. Als een test slecht voorspelt, en niet beter is dan toeval, is de AUC 0,50. Als een test perfect voorspelt wie wel of geen stoornis heeft, is de AUC 1,00. Een AUC vanaf 0,80 geldt als een goed resultaat.

De sensitiviteit, specificiteit en AUC van de FAT-R (bij een Afkappunt FAT-R Index ≤ 37 , dus onder het 10^e percentiel) bij twee studies staan in Tabel 12. De Valideringsstudie betreft 599 respondenten uit de regio's Noord en Oost, tussen 9 en 12 jaar oud. De Normeringssteekproef betreft 1491 respondenten uit het normeringsonderzoek, van wie ook de WL-score bekend is. Dit is een composietscore, gebruikt als maat voor leesstoornis, samengesteld uit de normscores van de EMT, Klepel en de subtest Monosyll uit CB&WL.

Tabel 12

Sensitiviteit en specificiteit van de FAT-R ten opzichte van de WL-score

	Sensitiviteit	Specificiteit	AUC
Valideringsstudie afkap WL 37 (=gem. -1.3 SD)	0,46	0,91	0,82
Valideringsstudie afkap WL 30 (=gem. -2.0 SD)	0,86	0,89	0,94
Normeringssteekproef afkap WL 37 (=gem. -1.3 SD)	0,49	0,92	0,81
Normeringssteekproef afkap WL 30 (=gem. -2.0 SD)	0,68	0,90	0,89

De AUC-waardes zijn allemaal boven de 0,80 en zijn daarmee goed te noemen. De sensitiviteit is bij een criterium van 37 op WL laag, namelijk onder de 0,50; voor een screener wil je een sensitiviteit van 0,80 of hoger. Dit pleit ervoor om het strengere criterium van 30 op WL te hanteren. Bij de specificiteit maakt het niet veel uit welk criterium gehanteerd wordt: WL van 37 (onder het 10e percentiel) of WL van 30 (twee standaarddeviaties beneden het gemiddelde). Bij al deze studies was het afkappunt van de FAT-R Index gelijk aan 37.

Neventests

Continu Benoemen

Continu Benoemen werd gemeten met de CB&WL (Bos, van den & Lutje Spelberg, 2010). Er is onderscheid tussen twee vormen van Continu Benoemen: Alfanumeriek (gemeten met Cijfers- en Letters benoemen) en Non-Alfanumeriek (gemeten met Kleuren- en Plaatjes benoemen). De verwachting was dat het verband tussen de FAT-R en Alfanumeriek benoemen hoger zou

zijn dat het verband tussen de FAT-R en Non-Alfanumeriek benoemen. In Tabel 13 staan de correlaties tussen scores op de FAT-R, de scores op Continu Benoemen: Alfa-numeriek en de scores op Continu Benoemen: Non-Alfanumeriek.

Tabel 13

Het verband tussen fonemische analysevaardigheid en de Alfa-numerieke en Non-Alfanumerieke Benoemingsnelheid

Normgroep	Continu Benoemen Alfa-numeriek		Continu Benoemen Non-Alfanumeriek	
	n	r	n	r
8	216	0,17	216	0,13
9	175	0,36	175	0,26
10	128	0,26	128	0,40
12	126	0,34	126	0,23
13	131	0,49	51	0,36
14	153	0,32	153	0,19
Totaal	929	0,32	849	0,25

Te zien is dat deze verwachting in het algemeen is uitgekomen. Alleen in de groep 10-jarigen was de relatie met Non-Alfanumeriek benoemen hoger dan de relatie met Alfa-numeriek benoemen. Daarmee is deze studie een onderbouwing van de validiteit van de FAT-R.

EMT-Klepel

De EMT-Klepel² (Bos, van den; Lutje Spelberg; Scheepstra; & Vries, de, 1994 en Brus & Voeten, 1972) werd gebruikt om de relatie van de FAT-R met woordlezen te onderzoeken. Er waren twee duidelijke theoretische verwachtingen, namelijk dat er een verband tussen deze twee is, en dat het verband afneemt naarmate kinderen ouder worden. In Tabel 14 staan de correlaties van de FAT-R-Index met de scores op de EMT, de Klepel en de Composietscore EMT-Klepel.

Tabel 14

Correlaties van de FAT-R-Index met de EMT, de Klepel en de Composietscore EMT-Klepel

Normgroep	N	EMT	Klepel	EMT-Klepel composietscore
7	85	0,54	0,55	0,56
8	73	0,67	0,57	0,65
9	108	0,65	0,72	0,71
10	118	0,67	0,67	0,69
11	148	0,60	0,58	0,61
12	150	0,45	0,56	0,54
13	127	0,51	0,51	0,54
14	87	0,42	0,40	0,44
Totaal	887	0,55	0,56	0,58

2 De composietscore EMT-Klepel, het gemiddelde van de afzonderlijke standaardcores op de EMT en de Klepel, uitgedrukt op een T-schaal

Te zien valt dat er een duidelijk verband tussen de maten is, en dat dit inderdaad afneemt met de leeftijd. Daarmee is ook deze studie een onderbouwing van de validiteit van de FAT-R.

Conclusie validering

De validiteit van de FAT-R is op verschillende manieren onderzocht. De onderzoeksgroepen sluiten aan bij de doelgroep van de FAT-R. Zowel de interne als de externe structuur zijn onderzocht. De resultaten ondersteunen de validiteit van de FAT-R. In dit whitepaper is een greep gedaan uit de studies, maar in de handleiding van de FAT-R staan nog veel meer onderzoeken gepresenteerd. Al met al is de validiteit van de FAT-R goed.

6 Referenties

- Bos K.P. van den & Lutje Spelberg H.C. (2007, 2010). *Continu benoemen en woorden lezen (CB&WL). Een test voor het diagnosticeren van taal- en leesproblemen*. Gebruikershandleiding en verantwoording. Amsterdam: Boom test uitgevers.
- Bos, K.P. van den, Lutje Spelberg, H.C., Scheepstra, A.J.M. & De Vries, J.R. (1994). *De Klepel. Vorm A en B. Een test voor de leesvaardigheid van pseudowoorden. Verantwoording, handleiding, diagnostiek en behandeling*. Amsterdam: Pearson Assessment and Information.
- Brus, B. Th. & Voeten, M.J.M. (1972). *Een-minuut-test*. Amsterdam: Pearson Assessment and Information. Zesde ongewijzigde druk, 2009.
- CBS, *Kerncijfers*, <http://statline.cbs.nl/statweb/>, (2005, 2009, 2011).
- CFI, *Open data van DUO*, http://www.cfi.nl/organisatie/open_onderwijsdata/default.asp (2006).
- Dienst Uitvoering Onderwijs, *Open data van DUO*, https://duo.nl/organisatie/open_onderwijsdata/default.asp (2011).
- Evers, A., Lucassen, W., Meijer, R. & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP.
- Groot, B.J.A. de; Bos, K.P. van den & Meulen, B.F. van der (2014). *FAT-R Fonemische Analyse Test Herziene versie*. Amsterdam: Pearson Assessment and Information.
- Van den Bos, K.P., Lutje Spelberg, H.C. & De Groot, B.J.A. (2010). *FAT Fonemische Analyse Test Verantwoording en Handleiding*. Amsterdam: Pearson Assessment and Information.



Pearson Assessment and Information BV

Radarweg 60-A1, 1043 NT Amsterdam

Postbus 78, 1000 AB Amsterdam

T: +31 (0)20 581 5500

E: info-nl@pearson.com

www.pearsonclinical.nl

www.pearsonclinical.be

Twitter: [@PearsonNL](https://twitter.com/PearsonNL)