



Bayley-III-NL

PSYCHOMETRISCHE EIGENSCHAPPEN
WHITE PAPER 2

White paper

Bayley—III-NL

Psychometrische eigenschappen
White paper 2

Margreet Kerkmeer
Janke Zijlstra
Joëlle Dek

Januari 2015

PEARSON

Inhoud

1	Samenvatting	4
2	Inleiding	5
3	Bayley-III-NL Normering	6
	3.1 Steekproef	6
	3.2 Normscores	7
4	Betrouwbaarheid	9
	4.1 Inleiding	9
	4.2 Betrouwbaarheid: interitemrelaties	9
	4.3 Betrouwbaarheid: test-hertest stabiliteit	10
	4.4 Conclusies betrouwbaarheid	10
5	Validiteit	11
	5.1 Inleiding	11
	5.2 Verschillen tussen groepen	11
	5.3 Neventests	13
	5.4 Conclusies validiteit	14
6	De Vragenlijsten	15
	6.1 Inleiding	15
	6.2 Normen	15
	6.3 Betrouwbaarheid	16
	6.4 Validiteit	17
	6.5 Conclusie	18
7	Referenties	19
	Bijlage A	21
	Indeling normgroepen naar leeftijd	

1 Samenvatting

De Bayley-III-NL (Baar, Steenis, Verhoeven & Hessen, 2014) is de lang verwachte opvolger van de BSID-II-NL (Meulen, van der et al. 2002). Het is een instrument om de ontwikkeling bij heel jonge kinderen vast te stellen op het gebied van taal, motoriek en cognitie. De Bayley-III (Bayley, 2006) bestaat in plaats van drie uit vijf schalen: Cognitie, Taalbegrip, Taalproductie, Grove Motoriek en Fijne Motoriek. Bovendien kan de Bayley-III-NL extra informatie leveren over vaardigheden en sociaal-emotionele ontwikkeling door de toevoeging van twee vragenlijsten: de Adaptief Gedrag Schaal¹, en de Sociaal-Emotionele Ontwikkeling Schaal².

De Bayley-III-NL is genormeerd op 1945 kinderen tussen 2 weken en 3,5 jaar oud. Deze kinderen vormen een goede afspiegeling van de Nederlandse bevolking van die leeftijd. In de normeringssteekproef zijn ook 10% kinderen met ontwikkelingsproblemen opgenomen. Bijvoorbeeld te vroeg geboren kinderen en kinderen met Down syndroom.

De betrouwbaarheid wordt onderbouwd met coëfficiënt alfa, Guttman's lambda-2, test-hertest-onderzoek en interbeoordelaarsbetrouwbaarheid. De resultaten waren goed.

De validiteit wordt onderbouwd met intercorrelaties, item-restcorrelaties, factorstructuur, convergente en divergente validiteit, verschillen tussen relevante groepen, en voorspellende waarde. De resultaten wijzen erop dat de Bayley-III-NL een valide instrument is.



¹ gebaseerd op de Adaptive Behavior Assessment System – Second Edition; Harrison & Oakland, 2003

² gebaseerd op de Greenspan Social-Emotional Growth Chart; Greenspan, 2004

2 Inleiding

Het standaardisatieonderzoek is een van de laatste fases in de ontwikkeling van een test. Tijdens deze fase wordt de data verzameld waarmee de normen bepaald worden, daarnaast wordt onderzoek gedaan naar de betrouwbaarheid en de validiteit van de test. Betrouwbaarheid en validiteit zijn belangrijke psychometrische eigenschappen van een test. De betrouwbaarheid van een test geeft aan of de test elke keer hetzelfde meet en de validiteit van een test geeft aan of de test meet wat hij moet meten.

In deze whitepaper worden enkele onderzoeken en resultaten besproken van het standaardisatieonderzoek dat gedaan is voor de Bayley-III-NL en op basis waarvan de normen berekend zijn. Daarnaast worden de methode en de resultaten van het validiteits- en betrouwbaarheidsonderzoek besproken. Dit zodat (aankomende) gebruikers van de Bayley-III-NL bekend zijn met de psychometrische eigenschappen van de Bayley-III-NL en zij bij het interpreteren van de resultaten van cliënten hier rekening mee kunnen houden.

3 Bayley-III-NL Normering

3.1 Steekproef

Het standaardisatieonderzoek voor de Bayley-III-NL vond plaats tussen april 2011 en september 2013. Dit onderzoek is uitgevoerd onder Nederlandse kinderen met en zonder ontwikkelingsachterstand. De kinderen waren verdeeld in 17 leeftijdsgroepen om recht te doen aan de snelle ontwikkeling die zij doormaken. De groepen worden aangeduid met letter A tot en met Q (zie Bijlage A voor de indeling van de leeftijden). Het onderzoek werd uitgevoerd met in totaal 71 testleiders en de gemiddelde afnameduur was 31-86 minuten, afhankelijk van de leeftijdsgroep.

Om een goede afspiegeling van de doelgroep te verkrijgen, bestaat net als in het Amerikaanse onderzoek 10% van de steekproef uit kinderen met een verhoogd risico op ontwikkelingsproblemen of met een specifieke klinische diagnose:

- Prematuur (geboren voor 37 weken zwangerschap);
- Syndromen, zoals het syndroom van Down;
- Beperkingen in taal en/of spraak of motoriek;
- Een algehele ontwikkelingsachterstand zonder aanwijsbare reden.

Het CBS (Centraal Bureau voor de Statistiek) verstrekt gegevens over de samenstelling van de Nederlandse bevolking. Voor een aantal demografische variabelen met betrekking tot kinderen tussen 2 weken en 3,5 jaar oud (geslacht, opleidingsniveau van de moeder, regio, etnische achtergrond) zijn streefpercentages opgesteld op basis van gegevens van het CBS, om ervoor te zorgen dat de data die verzameld werd, overeen kwam met de populatie (kinderen tussen 2 weken en 3,5 jaar). Na afloop is extra gecontroleerd of de steekproef representatief was, en is zo nodig gewogen om verschillen te minimaliseren. De wegingsfactor was nooit groter dan 2, de maximaal aanvaarde wegingsfactor volgens de COTAN (Evers et al., 2010).

In Tabel 1 Kenmerken gewogen steekproef worden de kenmerken van de steekproef gepresenteerd nadat gewogen is voor geslacht, opleidingsniveau van de moeder, herkomst van de ouders en regio en het wel of niet behoren tot de klinische groep.

Tabel 1

Kenmerken gewogen steekproef

Groep	N	Geslacht (%)		Herkomst ouders (%)		Opleiding moeders (%)			Regio (%)				Gezonde kinderen	
		M	V	Nederlands	Niet-Nederlands	Laag	Midden	Hoog	Noord	Oost	Zuid	West	N	%
Totaal	1945	51,2	48,8	74,9	25,1	15,6	39,0	45,4	9,6	20,5	22,5	47,4	1753	90,1
Streef%		51,0	49,0	75,0	25,0	16,0	39,0	45,0	10,0	21,0	22,0	47,0		90,0
A	74	58,1	41,9	78,1	21,9	13,7	39,7	46,6	12,3	19,2	27,4	41,1	65	87,8
B	73	54,8	45,2	76,7	23,3	16,2	31,1	52,7	9,5	20,3	21,6	48,6	69	94,5
C	70	50,0	50,0	75,7	24,3	11,4	44,3	44,3	12,9	18,6	21,4	47,1	69	97,2
D	76	44,7	55,3	75,0	25,0	15,8	35,5	48,7	10,7	20,0	20,0	49,3	75	98,7
E	73	54,8	45,2	79,7	20,3	17,6	40,5	41,9	9,5	21,6	23,0	45,9	66	91,7
F	74	47,3	52,7	77,0	23,0	18,9	35,1	45,9	9,3	22,7	22,7	45,3	72	97,3
G	156	52,6	47,4	75,5	24,5	14,1	39,7	46,2	9,6	19,9	23,7	46,8	149	95,5
H	122	50,0	50,0	73,0	27,0	16,4	37,7	45,9	9,8	21,3	22,1	46,7	114	94,2
I	109	48,6	51,4	73,6	26,4	14,5	42,7	42,7	8,3	22,9	23,9	45,0	100	90,9
J	148	52,7	47,3	75,7	24,3	15,0	40,8	44,2	10,9	21,1	21,1	46,9	129	87,2
K	149	51,0	49,0	74,5	25,5	15,4	41,6	43,0	7,4	18,2	23,6	50,7	131	87,9
L	109	51,4	48,6	76,1	23,9	15,6	39,4	45,0	9,3	19,4	20,4	50,9	99	90,8
M	109	50,5	49,5	71,6	28,4	16,5	39,4	44,0	10,1	20,2	21,1	48,6	100	91,7
N	103	52,4	47,6	75,7	24,3	17,5	36,9	45,6	8,8	19,6	19,6	52,0	88	85,4
O	181	52,5	47,5	74,6	25,4	14,9	39,2	45,9	8,9	20,6	23,3	47,2	151	83,4
P	211	49,8	50,2	71,0	29,0	16,2	36,7	47,1	9,5	21,0	23,8	45,7	179	84,8
Q	108	49,1	50,9	77,8	22,2	16,5	39,4	44,0	9,3	22,2	22,2	46,3	97	89,8

Noot Laag opleidingsniveau verwijst naar speciaal onderwijs, basisschool, vmbo, mbo-1;
Midden opleidingsniveau verwijst naar mbo-2-4, havo, vwo;
Hoog opleidingsniveau verwijst naar hbo, universiteit bachelor, master of hoger niveau.
Noord zijn de provincies Groningen, Friesland, Drenthe;
Oost de provincies Gelderland, Overijssel, Flevoland;
Zuid de provincies Limburg en Noord-Brabant;
West de provincies Noord-Holland, Zuid-Holland, Utrecht en Zeeland.

Te zien valt dat de steekproef in het algemeen erg goed voldoet aan de streefpercentages en vrijwel nergens meer dan 5% afwijkt, het internationaal aanvaarde afwijkingsverschil van de steekproef van de populatie

3.2 Normscores**3.2.1 Scoring**

Bij de Bayley-III-NL is het mogelijk om online door middel van een scoringsprogramma de scores te berekenen. Hierbij wordt gebruik gemaakt van dagnormen. Tevens is er een handleiding beschikbaar om de scores handmatig te berekenen. Hierbij is gebruik gemaakt van maandnormen. Bij de onlinescoring zijn de normen dus preciezer en is er een kleinere meetfout. Ook bestaat hier niet het risico van fouten bij het berekenen van de scores of opzoekfouten. Uit het scoringsprogramma komt een rapport met alle genormeerde scores, vergelijkingscores, profielen en groeigrafieken.

3.2.2 Bepalen geschaalde subtest scores

Voor iedere subtest van de Bayley-III-NL zijn dagnormen beschikbaar. Er is gebruik gemaakt van een zogenaamde Box Cox transformatie om ruwe scores om te zetten in geschaalde scores (Box & Cox, 1964). Deze transformatie gaat niet alleen uit van het gemiddelde van de resultaten per leeftijdsgroep, maar ook van de scheefheid en gepiektheid van de verdeling van de resultaten per leeftijdsgroep. Hierdoor sluiten de normen nog beter aan bij de verdeling in de populatie. Elke geschaalde score kan liggen tussen 1 en 19, met een vast gemiddelde van 10 en een SD van 3.

3.2.3 Bepalen indexscores

Voor het bepalen van de indexscores zijn de geschaalde subtestsscores per indexscore bij elkaar opgeteld. Deze sommen zijn getransformeerd naar een (Wechsler)schaal met een gemiddelde van 100 en een SD van 15.

3.2.4 Ontwikkelingsleeftijdsequivalenten.

Ontwikkelingsleeftijdsequivalenten (OLE), of kortweg ontwikkelingsleeftijden, geven aan bij welke leeftijd een bepaalde ruwe score gemiddeld is (dus een standaardscore van 10 oplevert). Om OLE's te bepalen is per subtest voor elke leeftijdsgroep nagegaan bij welke totale ruwe score een geschaalde subtestscore van 10 werd behaald. Zo kun je zeggen dat een kind van 12 maanden een OLE van 14 maanden heeft voor Grove Motoriek (dan loopt het kind als het ware voor), of van 9 maanden voor Taalbegrip (dan loopt het kind als het ware achter). Er zitten veel haken en ogen aan deze methode, die worden besproken in de handleiding in paragraaf 8.5.6 en bijvoorbeeld in Evers & Resing (2007).

4 Betrouwbaarheid

4.1 Inleiding

Om de betrouwbaarheid van de Bayley-III-NL te onderzoeken is voor de subtests en de schalen een schatting van de betrouwbaarheid gemaakt door de interitemrelaties te bekijken (ook wel interne consistentie genoemd). Deze maat, Guttman's λ , wordt besproken in paragraaf 4.2 van deze whitepaper. De λ 's zijn ook gebruikt om de standaardmeetfout en betrouwbaarheidsintervallen te berekenen. Bovendien zijn de test-hertestbetrouwbaarheid en de interbeoordelaarsbetrouwbaarheid bepaald. De test-hertestbetrouwbaarheid wordt besproken in paragraaf 4.3 van deze whitepaper, en informatie over de interbeoordelaarsbetrouwbaarheid is te vinden in de handleiding van de Bayley-III-NL (Baar e.a., 2014).

4.2 Betrouwbaarheid: interitemrelaties

Bij het berekenen van de interitemrelaties is gebruik gemaakt van Guttman's λ^2 (Guttman, 1945). Coëfficiënt alfa is de meest bekende coëfficiënt voor interne consistentie. Alfa wordt echter gezien als ondergrens van de betrouwbaarheid (Novick & Lewis, 1967) en de schatting ervan komt vrijwel altijd lager uit dan de werkelijke testbetrouwbaarheid (Evers e.a., 2010). De waarden van andere schatters liggen dicht bij de werkelijke betrouwbaarheid (Sijtsma, 2009), waaronder λ^2 . Voor het schatten van de betrouwbaarheid is daarom deze coëfficiënt berekend. Ook voor deze maat geldt dat de samenhang van de antwoorden op de vragen groter is, naarmate de betrouwbaarheidsmaat dicht bij 1 ligt (hoogst mogelijke waarde voor betrouwbaarheid). In Tabel 2 staan de waarden van λ^2 vermeld voor alle kinderen uit de normeringssteekproef.

Tabel 2

Waarden van λ^2 per normgroep, voor Subtests en Indexscores

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	Gem. Rxx
Cognitie; index	0.73	0.78	0.76	0.80	0.89	0.83	0.77	0.76	0.80	0.84	0.90	0.87	0.89	0.94	0.86	0.96	0.87	0.87
Taal; index	0.66	0.69	0.62	0.64	0.75	0.75	0.77	0.71	0.84	0.91	0.94	0.94	0.96	0.95	0.95	0.98	0.96	0.90
Motoriek; index	0.75	0.81	0.80	0.77	0.87	0.87	0.87	0.87	0.91	0.90	0.94	0.85	0.85	0.90	0.87	0.97	0.91	0.90
Taalbegrip	0.55	0.60	0.55	0.53	0.61	0.65	0.65	0.58	0.74	0.84	0.89	0.90	0.91	0.89	0.90	0.96	0.91	0.83
Taalproductie	0.61	0.65	0.63	0.57	0.68	0.70	0.70	0.64	0.79	0.87	0.91	0.93	0.94	0.95	0.94	0.98	0.95	0.88
Fijne Motoriek	0.65	0.65	0.69	0.76	0.85	0.81	0.76	0.68	0.64	0.75	0.86	0.68	0.82	0.84	0.83	0.95	0.88	0.81
Grove Motoriek	0.74	0.81	0.82	0.67	0.78	0.87	0.88	0.91	0.92	0.90	0.94	0.83	0.81	0.87	0.84	0.96	0.85	0.88
N	79	73	69	73	72	72	153	120	107	145	150	107	108	103	179	233	110	1953

Te zien valt dat in de jongere leeftijdsgroepen A tot en met H (2 weken tot en met 10 maanden oud) de betrouwbaarheid varieert van 0.53 tot 0.91. Waardes tot 0.70 zijn te laag, met name bij de taalschalen komt dit voor, maar de waarden van 0.70 en hoger zijn acceptabel tot goed. Een verklaring hiervoor kan zijn dat deze kinderen niet zo stabiel presteren door specifieke omstandigheden, zoals moeheid, honger en geringe vertrouwdheid met de omgeving.

Bovendien is bij de jongste groepen vaak sprake van een bodemeffect. Als er nauwelijks variatie in de scores is, is het een statistisch artefact dat betrouwbaarheden erg laag kunnen uitvallen. Voor de oudere leeftijdsgroepen I tot en met Q (11 maanden tot en met 42 maanden en 15 dagen) zijn de betrouwbaarheidscoëfficiënten ten minste voldoende voor alle subtests, behalve voor Fijne Motoriek voor leeftijdsgroepen I (0.64) en L (0.68). Dit betekent dat de Bayley-III-NL al met al een voldoende tot goede betrouwbaarheid heeft.

4.3 Betrouwbaarheid: test-hertest stabiliteit

Het onderzoek naar de test-herteststabiliteit van de Bayley-III-NL is uitgevoerd onder 183 kinderen. Hiermee onderzoek je of de respondent dezelfde scores behaalt, als je de test bij dezelfde respondent op verschillende momenten afneemt. Met andere woorden of je elke keer hetzelfde meet. De kinderen die deelnamen aan dit onderzoek varieerden in leeftijd van 16 dagen tot 42 maanden en 15 dagen oud. In Tabel 3 staan de gemiddelde scores op de eerste en tweede afname, de correlaties tussen de twee afnames en de effectgrootte van deze correlatie. Deze gegevens zijn ook per leeftijdsband berekend, deze zijn te vinden in de technische handleiding van de Bayley-III-NL (Baar e.a., 2014).

Tabel 3

Gemiddelden, correlaties en de effectgrootte van de gehele test-herteststeekproef

Totaal, N = 183	Eerste afname		Tweede afname		r_{12}	Corrected r^a	Effect-grootte
	Gem.	SD	Gem.	SD			
Cognitie; index	103,88	14,92	105,43	14,87	0,72**	0,72	0,03
Taal; index	101,32	14,27	103,51	14,44	0,74**	0,76	0,04
Motoriek; index	102,08	13,75	103,22	14,19	0,74**	0,78	0,02
Cognitie	10,67	3,15	11,00	3,15	0,72**	0,69	0,03
Taalbegrip	10,52	3,17	10,72	3,03	0,63**	0,59	0,02
Taalproductie	9,89	3,08	10,50	3,30	0,73**	0,71	0,05
Fijne Motoriek	10,43	3,10	10,89	3,00	0,68**	0,66	0,04
Grove Motoriek	10,01	2,92	9,92	3,07	0,75**	0,76	-0,01

Uit de gevonden correlaties en effectgroottes blijkt dat de kinderen op beide afnames gelijkwaardige scores halen, met dat verschil dat de scores op de tweede afname iets hoger liggen. Dit kan veroorzaakt zijn doordat het kind zich inmiddels al verder ontwikkeld heeft of door een leereffect van de test; omdat het kind vorige keer de opgaven ook al gedaan heeft, kan het nu wel de benodigde taak uitvoeren. De effectgroottes zijn zeer klein, wat betekent dat de scores behoorlijk stabiel zijn tussen beide afnames. De test blijkt dus elke keer hetzelfde te meten.

4.4 Conclusies betrouwbaarheid

De betrouwbaarheid van de Bayley-III-NL is op diverse manieren onderzocht. Zowel de inter-itemrelaties als de test-hertest- en interbeoordelaarsbetrouwbaarheid gaven, zeker voor de kinderen vanaf 10 maanden, goede resultaten. Hierdoor kunnen we concluderen dat de Bayley-III-NL een betrouwbaar instrument is.

5 Validiteit

5.1 Inleiding

Om de validiteit van de Bayley-III-NL te onderbouwen, zijn verschillende onderzoeken uitgevoerd. Zo zijn er diverse groepen onderzocht op verschillen tussen de geschaalde subtestscores van de Bayley-III-NL. Bij deze groepen werd op basis van de literatuur soms wel en soms geen verschil verwacht in scores. Er is gekeken naar Sekse, Etnische herkomst, Opleidingsniveau moeder en Regio. In dit whitepaper worden alleen de resultaten met betrekking tot opleidingsniveau moeder gepresenteerd, omdat daarvan vanuit de literatuur een duidelijke verwachting was dat kinderen van moeders van verschillende opleidingsniveau van elkaar zouden verschillen. De overige onderzoeken zijn te vinden in de Technische handleiding (Baar e.a, 2014). Daarnaast zijn de volgende Klinische groepen bekeken: kinderen met het syndroom van Down, kinderen met een ontwikkelingsachterstand en te vroeg geboren kinderen. Van deze groepen werd verwacht dat zij van afwijkende scores zouden behalen, gezien hun stoornis of beperking. In dit whitepaper wordt alleen het onderzoek met de kinderen met het syndroom van Down behandeld; de overige informatie is in de handleiding te vinden. Ter onderbouwing van de convergente en divergente validiteit zijn neventest-onderzoeken uitgevoerd met de BSID-II-NL, WPPSI-III-NL, Lexilijst en Schlichting Test voor Taalbegrip en Taalproductie. In dit whitepaper worden de resultaten met de WPPSI-III-nl besproken. Verder is ter ondersteuning van de validiteit de interne structuur onderzocht met Confirmatieve Factoranalyse, item-restcorrelaties en intercorrelaties. Als laatste is de criteriumvaliditeit onderzocht, in hoeverre zijn de resultaten op de test voorspellend voor toekomstige ontwikkeling van het kind. Hiervoor is gekeken in hoeverre de test voorspellend is voor het hebben van een stoornis; met behulp van de sensitiviteit en specificiteit van de test. Deze laatstgenoemde onderzoeken worden niet in dit whitepaper besproken.

5.2 Verschillen tussen groepen

Opleidingsniveau

In Tabel 4 staan de gemiddelden en SD's voor kinderen van moeders die laag, midden of hoog opgeleid zijn per groep. Hierbij geldt de CBS-definitie van hoogst afgeronde opleiding, dus niet opleidingen die iemand eventueel op dit moment volgt. De verwachting is dat naarmate de moeder hoger is opgeleid, de kinderen gemiddeld iets hoger zullen scoren op de Bayley-III-NL. Dit omdat intelligentie voor een deel genetisch bepaald is en hoog opgeleide moeders daarnaast bijvoorbeeld gemiddeld genomen iets meer voorlezen en spelen met hun kind, wat een gunstige invloed heeft op hun ontwikkeling. Dit blijkt ook uit bijvoorbeeld de resultaten op intelligentietests (WPPSI-III-NL, Wechsler, 2009).

Tabel 4

Gemiddelden en SD's voor de groepen Opleidingsniveau moeder: Laag, Midden, Hoog.

	Laag (N = 223)		Midden (N = 627)		Hoog (N = 863)	
	Gem.	SD	Gem.	SD	Gem.	SD
Cognitie; index	97,34	12,97	100,66	13,31	102,55	13,48
Taal; index	96,69	13,00	100,68	13,25	103,08	13,44
Motoriek; index	100,02	12,51	101,72	13,72	101,95	13,41
Cognitie	9,28	2,76	9,99	2,82	10,39	2,86
Taalbegrip	9,35	2,73	10,04	2,82	10,50	2,95
Taalproductie	9,35	2,90	10,12	2,90	10,56	2,83
Fijne Motoriek	9,58	2,79	10,11	3,01	10,29	3,04
Grove Motoriek	10,13	2,72	10,19	2,92	10,10	2,96

Laag: lager onderwijs, vmbo, MBO-1

Midden: havo, vwo, MBO2-4

Hoog: HBO, WO

Te zien valt dat de scores licht toenemen met opleidingsniveau van de moeder. De verschillen tussen de groepen zijn significant op .05 niveau en sommigen op .01 niveau, met uitzondering van Motoriek en de schaal Grove Motoriek waarop de kinderen niet verschillen. Dit bevestigt de verwachting dat kinderen van hoger opgeleide moeders hoger scoren op de Cognitie en Taalschalen.

Klinische groepen

De geschaalde scores van een groep van 27 kinderen met het syndroom van Down zijn vergeleken met de geschaalde scores van een gematchte steekproef van 27 zich normaal ontwikkelende kinderen. De kinderen zijn gematcht op leeftijd, sekse, herkomst ouders, opleiding van de moeder en regio. In Tabel 5 staan de gemiddelde scores van de twee groepen met de minimum- en maximumscore, de significantie van eventuele verschillen en de effect-grootte daarvan. Verwacht werd dat kinderen met het syndroom van Down over de hele linie lager zouden scoren dan zich normaal ontwikkelende kinderen.

Tabel 5

Resultaten van de kinderen met het syndroom van Down vergeleken met de resultaten van de kinderen uit de gematchte groep

	Controle				Syndroom van Down				Sig.	Effect grootte
	Min	Max	Gem.	SD	Min	Max	Gem.	SD		
Cognitie; index	87	139	107,41	11,66	58	72	64,89	4,61	0,000	0,86
Taal; index	87	130	107,70	10,21	54	73	59,56	5,40	0,000	0,90
Motoriek; index	72	129	100,81	14,40	49	69	56,63	4,92	0,000	0,81
Cognitie	7	18	11,41	2,45	1	4	2,41	0,97	0,000	0,86
Taalbegrip	8	16	11,26	2,01	2	6	3,00	1,24	0,000	0,86
Taalproductie	7	15	11,52	2,26	1	4	1,93	0,96	0,000	0,89
Fijne Motoriek	4	18	9,74	3,60	1	5	2,30	0,91	0,000	0,68
Grove Motoriek	4	13	10,26	2,51	1	4	2,26	1,02	0,000	0,82

De resultaten zijn in de verwachte richting: op alle subtests en indexscores scoren de kinderen met het syndroom van Down significant lager dan de controlegroep. In de laatste kolom staan de effectgroottes. De effectgrootte is een statische maat voor het verschil tussen twee groepen. Het getal ligt rond de nul als er geen effect is en rond de 0,50 bij een matig effect, en de maat is hetzelfde voor alle soorten schalen of meeteenheden. De effectgroottes zijn bij de kinderen met het syndroom van Down alle boven de 0,80 wat wijst op een groot effect. Dit ondersteunt de validiteit van de Bayley-III-NL.

5.3 Neventests

Het onderzoeken van de relaties tussen de Bayley-III-NL en andere tests geeft informatie over de convergente en divergente validiteit (of soortgenootvaliditeit) van de Bayley-III-NL. Hiermee onderzoek je of de resultaten van jouw test positief samenhangen met een test met dezelfde meetpretentie, en negatief samenhangen met een test met een andere meetpretentie.

Bij 56 kinderen van 30 tot 42 maanden oud is zowel een WPPSI-III-NL als een Bayley-III-NL afgenomen, met een tussentijd van 1 tot 48 dagen. Bij 23 kinderen is eerst de Bayley-III-NL afgenomen en daarna de WPPSI-III-NL, en bij 33 kinderen was dat andersom. Verwacht wordt dat er een duidelijke samenhang (convergente validiteit) naar voren komt tussen de Cognitie- en Taalschaal van de Bayley-III-NL en de Totale en Verbale Indexen, evenals de Algemene Taal Index-score van de kinderen op de WPPSI-III-NL. Tussen de Motoriekschaal van de Bayley-III-NL en de WPPSI-scores wordt geen samenhang verwacht (divergente validiteit), omdat de performale taken van de WPPSI-III-NL ontworpen zijn om de cognitieve vaardigheden te meten en de motorische taken van de Bayley-III-NL sec de motorische vaardigheden meten.

Tabel 6

Correlaties tussen Bayley-III-NL en WPPSI-III-NL

	VIQ	PIQ	TIQ	ATI
Cognitie; index	0.45**	0.10	0.38**	0.27*
Taal; index	0.53**	0.03	0.40**	0.50**
Motoriek; index	0.24	0.18	0.27*	0.23
Cognitie	0.45**	0.10	0.38**	0.28*
Taalbegrip	0.52**	0.06	0.40**	0.44**
Taalproductie	0.44**	0.01	0.31*	0.45**
Fijne Motoriek	0.24	0.22	0.28*	0.22
Grove Motoriek	0.15	0.06	0.14	0.15

* significant op .05-niveau

** significant op .01-niveau

In Tabel 6 komt inderdaad naar voren dat er matig sterke (0.30- 0.50) tot sterke (>0.50) relaties worden gevonden tussen verwante subtests, zoals verwacht. Zo is de correlatie tussen het Verbaal IQ van de WPPSI-III-NL en Taalbegrip en Taalproductie van de Bayley-III-NL respectievelijk 0.52 en 0.44 en met de Taalschaal 0.53. Ook met de Algemene Taal Index van de WPPSI-III-NL en de taalsubtests zijn de gevonden correlaties in de verwachte richting; 0.44 met Taalbegrip en 0.45 met Taalproductie en met de totale Taalschaal van 0.50. Met de Performale intelligentiescore van de WPPSI-III-NL is geen statistisch significante samenhang. Zoals verwacht hangen de Motoriekschaal en de resultaten op de WPPSI-III-NL niet significant samen. De correlatie tussen de Motoriek-index en het Totaal IQ van de WPPSI-III-NL is wel significant, maar een matige correlatie. Deze bevindingen bevestigen de verwachtingen en vormen dus onderbouwing voor de validiteit van de Bayley-III-NL dat deze deels hetzelfde meet als de WPPSI-III-NL schalen.

5.4 Conclusies validiteit

De validiteit van de Bayley-III-NL is op diverse manieren onderzocht. De verschillen tussen relevante groepen en de neventestonderzoeken onderbouwen dat de Bayley-III-NL meet wat hij beoogt te meten. De onderzoeken naar het onderscheidingsvermogen tonen aan dat de Bayley-III-NL voorspellende waarde heeft voor relevante kenmerken in de doelgroep. Al met al kunnen we vaststellen dat de Bayley-III-NL een valide instrument is.

6 De Vragenlijsten

6.1 Inleiding

Naast de Bayley-III-NL die wordt afgenomen door een testleider, zijn er ook twee vragenlijsten over de sociale, emotionele en adaptief gedragsontwikkeling van het kind. Deze worden ingevuld door de ouder/verzorger.

De Adaptive Behavior Assessment System – Second Edition (ABAS-II; Harrison & Oakland, 2003) vormt de basis van de Adaptief Gedrag Schaal. De items zijn gegroepeerd in tien domeinen. Deze domeinen worden gecombineerd tot drie samengestelde scores en een totale schaal de Adaptief Gedrag Indexscore; AGI. In figuur 1 wordt aangegeven welke vaardigheidsdomeinen met de Adaptief Gedrag Schaal in kaart worden gebracht en tot welke drie vaardigheden deze samengenomen kunnen worden.

Vaardigheidsdomeinen:

Communicatie
Omgang buitenshuis
Voorschoolse vaardigheden
Huishoudelijke vaardigheden
Gezondheid en veiligheid
Spel
Zelfredzaamheid
Zelfregulatie
Sociaal contact
Motoriek

Samengestelde scores:

Conceptuele vaardigheden CON

Communicatie
Voorschoolse activiteiten
Zelfregulatie

Sociale vaardigheden SOC

Spel
Sociaal contact

Praktische vaardigheden PRV

Gedrag buitenshuis
Huishoudelijke activiteiten
Gezondheid en veiligheid
Zelfredzaamheid

De Sociaal-Emotionele Ontwikkeling Schaal beoordeelt de sociale en emotionele ontwikkeling van jonge kinderen en loopt belangrijke mijlpalen in de sociaal-emotionele ontwikkeling na die horen bij een bepaalde leeftijd. De Greenspan Social-Emotional Growth Chart (Greenspan, 2004) vormt de basis van de Bayley-III Sociaal-Emotionele Ontwikkeling Schaal. De vragenlijst bestaat uit 35 vragen. Op verschillende punten in de vragenlijst wordt aangegeven wanneer de ouders mogen stoppen, gezien de leeftijd van hun kind.

6.2 Normen

In totaal hebben 1845 moeders de Sociaal Emotionele Ontwikkeling Schaal ingevuld en 1846 de Adaptief Gedrag Schaal. Daarnaast hebben 321 vaders de Sociaal Emotionele Ontwikkeling Schaal ingevuld en 293 de Adaptief Gedrag Schaal. De streefcijfers zijn bepaald volgens CBS-gegevens en de steekproeven waren na weging representatief voor de populatie:

Tabel 7

Beschrijving steekproeven Sociaal Emotionele Ontwikkeling Schaal en Adaptief Gedrag Schaal inclusief streefpercentages

		Sociaal Emotioneel	Adaptief Gedrag	streef%
aantal	voor weging	1845	1846	
	na weging	1787	1691	
seks kind	man	50,7%	51,1%	51,0%
	vrouw	49,3%	48,9%	49,0%
herkomst	Nederlands	75,6%	76,8%	75,0%
	niet-Nederlands	24,4%	23,2%	25,0%
opleiding moeder	laag	15,8%	14,4%	16,0%
	midden	38,9%	38,7%	39,0%
	hoog	45,3%	46,9%	45,0%
regio	noord	9,2%	9,0%	10,0%
	oost	21,1%	21,3%	21,0%
	zuid	23,2%	22,5%	22,0%
	west	46,5%	47,2%	47,0%
gezonde kinderen	N	1614	1530	
	%	90,3%	90,5%	90,0%

In Tabel 8 is te zien dat de afwijking van de streefcijfers nergens meer dan 5% bedraagt. Ook voor de vragenlijsten zijn dagnormen berekend op basis van continue normering. De Sociaal Emotionele Ontwikkeling Schaal en de domeinen van de Adaptief Gedrag Schaal hebben een geschaald gemiddelde van 10 (SD=3). De samengestelde scores op de vaardigheden van de Adaptief Gedrag Schaal hebben een gemiddelde van 100 (SD=15), waarbij is opgelegd dat de scores tussen 40 en 160 moeten liggen.

6.3 Betrouwbaarheid

Net als bij het de kernschalen van de Bayley-III-NL, wordt bij de vragenlijsten Guttman's labda-2 (Guttman, 1945) gepresenteerd. Ook is de interbeoordelaarsbetrouwbaarheid tussen vaders en moeders onderzocht.

6.3.1 Labda's Controlegroep

De labda-coëfficiënten zijn voor de Sociaal Emotionele Ontwikkeling Schaal gemiddeld 0,92 en liggen tussen 0,81 (0-3 maanden) en 0,96 (31-42 maanden). Voor de Adaptief Gedrag Vragenlijst zijn de waarden per leeftijdsgroep op de vaardigheidsdomeinen voldoende tot goed, op één te lage waarde na (0,47 voor het vaardigheidsdomein Gezondheid en Veiligheid in 4-7 mnd), met een gemiddelde van 0,83. Voor de totale groep zijn de betrouwbaarheden goed tussen 0,76 en 0,89. De samengestelde scores (CON, SOC en PRV) hebben een gemiddelde labda van 0,89 met een range van 0,71 tot 0,97. Deze zijn dus alle goed.

6.3.2 Interbeoordelaarsbetrouwbaarheid

Van 321 kinderen hebben zowel de moeder als de vader de Vragenlijsten ingevuld. Uit de resultaten van de gepaarde t-test en de gepaarde correlatie blijkt voor de Sociaal Emotionele Ontwikkeling Schaal dat vaders en moeders overeenstemmen in hun beoordeling, gezien de

correlatie van 0.59. Maar de kinderen krijgen van de vaders wel systematisch een lagere score dan van de moeders, $t(320) = 5.427$, $p = .000$, met een matig effect van 0.07. Bij de Adaptief Gedrag Schaal lagen de correlaties tussen 0.47 (Zelfregulatie) en 0.67 (Voorschoolse competenties). Ook hier krijgen kinderen over het algemeen lagere cijfers van hun vader dan van hun moeder. Wel moet men zich er van bewust zijn dat de normering is uitgevoerd op de scores gegeven door de moeders. Dus indien een vragenlijst ingevuld wordt door vader, zou het kunnen dat deze een iets andere score zou geven, waar bij de interpretatie rekening mee gehouden moet worden.

6.3.3 Conclusie

Al met al is de betrouwbaarheid van de Sociaal Emotionele Ontwikkeling Schaal en de Adaptief Gedrag Schaal voldoende tot goed te noemen.

6.4. Validiteit

Ten eerste zijn de intercorrelaties onderzocht tussen de subtest- en domeinscores. De Sociaal Emotionele Ontwikkeling Schaal laat een zwakke tot matige samenhang zien met de vaardigheidsdomeinen en samengestelde scores van de Adaptief Gedrag Vragenlijst, variërend van 0.23 met Gezondheid en veiligheid tot 0.38 met de samengestelde score voor Sociaal contact. Tussen de samengestelde scores van de Adaptief Gedrag Vragenlijst onderling blijkt een matige tot sterke samenhang te bestaan, variërend van 0.36 tussen Zelfredzaamheid en Communicatie tot 0.65 tussen Gezondheid en veiligheid en Huishoudelijke vaardigheden. Dit bevestigt de verwachtingen.

Verder zijn de item-restcorrelaties bekeken ter ondersteuning van de constructvaliditeit; meten de items binnen een schaal wel hetzelfde. Deze item-restcorrelaties liggen gemiddeld tussen 0.58 en 0.86, wat goed is.

De scores van kinderen met verschillende achtergrondkenmerken zijn vergeleken. Hierbij is een variantieanalyse uitgevoerd voor een van de genoemde kenmerken, waarbij is gecontroleerd voor de overige kenmerken. Het doel hiervan is om te kijken of verschillen tussen kinderen te verklaren zijn door dat specifieke achtergrondkenmerk of ook voorkomen uit andere verschillen tussen de kinderen. Voor sekse blijkt dat meisjes over alle leeftijdsgroepen heen wat hogere scores halen dan jongens (soms statistisch significant). Wat betreft etniciteit blijkt dat Nederlandse kinderen wat hogere scores behalen dan niet-Nederlandse kinderen op de Sociaal-Emotionele Schaal; voor de Adaptief Gedrag Schaal is dit juist andersom. Voor opleidingsniveau van de moeders blijken de kinderen hogere scores te krijgen naarmate de moeders hoger zijn opgeleid. Voor regio komen geen verschillen naar voren. Overigens zijn de effectgroottes in het algemeen aan de lage kant (η^2 tussen 0.002 en 0.048).

Tot slot zijn klinische groepen vergeleken met zich normaal ontwikkelende kinderen. Bij de 26 kinderen met het syndroom van Down wordt duidelijk dat zij ook in hun sociaal-emotionele ontwikkeling en hun adaptieve gedrag een duidelijke achterstand laten zien in vergelijking met 23 kinderen uit een controlegroep ($F(10,34) = 44.70$, $p = .534$, $\eta^2 = .93$). Dit bleek te gelden voor de Sociaal-Emotionele Ontwikkeling en voor alle vaardigheids domeinen van de Adaptief Gedrag Vragenlijst met een sterke effectgrootte (bij alle domeinen groter dan 0.43). De resultaten van de andere groepen staan beschreven in de handleiding.

6.5 Conclusie

De Vragenlijsten kunnen een waardevolle aanvulling zijn op de gegevens van de Bayley-III-NL, omdat ze naast de Cognitie, Taal en Motoriek een beeld geven van de Sociaal Emotionele Ontwikkeling en het Adaptief Gedrag, zoals de ouders dit zien. Dit zorgt voor een completer beeld van de ontwikkeling van het kind en ook voor meer betrokkenheid van de ouders bij het diagnostisch proces. De vragenlijsten zijn in een grote, representatieve steekproef genormeerd en zijn betrouwbaar en valide.

7 Referenties

- Baar, A.L., Steenis, L.J.P., Verhoeven, M., & Hessen, D.J. (2014). *Bayley-III- NL, Technische Handleiding*. Amsterdam: Pearson Assessment and Information B.V.
- Bayley, N. (2006). *Bayley Scales of Infant and Toddler Development - Third Edition*. San Antonio, NCS Pearson.
- Box, George E. P.; Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B* 26 (2): 211–252. JSTOR 2984418. MR 192611.
- Centraal Bureau voor Statistiek [CBS] – Statline (2011). Retrieved 2010 from [http:// www.cbs.nl](http://www.cbs.nl)
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2010). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP
- Evers, A. & Resing, W. (2007) Het drijfzand van didactische leeftijdsequivalenten. *De Psycholoog*, jaargang 42 (9) pag. 466-472. Te vinden op bijv. http://www.omdatelkkindtelinzuidoost.nl/media/files/rekenspecialisten/plg_rs_100318/gc_2007_12_16_eversresing_het_drijfzand_van_dle.pdf
- Greenspan, S. I. (2004). *Greenspan social-emotional growth chart: A screening questionnaire for infants and young children*. San Antonio, TX: Harcourt Assessment.
- Guttman, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255-282.
- Harrison, P. L., & Oakland, T. (2003). *Adaptive behavior assessment system – Second edition*. San Antonio, TX: The Psychological Corporation.
- Meulen, B.F. van der, Ruiter, S.A.J. , Spelberg, H.C. Lutje, & Smrkovsky M. (2002), *Bayley Scales of Infant Development Second Edition – Nederlandse Versie*. Amsterdam: Pearson Assessment and Information B.V.
- Novick, M.R. & Lewis, C. (1967). Coefficient alpha and the reliability of composite measurements. *Psychometrika*, 32, 1-13.
- Schlichting, J.E.P.T., & Lutje Spelberg, H.C. (2007). *Lexilijst Begrip, handleiding*. Amsterdam: Pearson Assessment and Information B.V.
- Schlichting, J.E.P.T., & Lutje Spelberg, H.C. (2010a). *Schlichting Test voor Taalbegrip, Handleiding*. Houten: Bohn Staf leu van Loghum.
- Schlichting, J.E.P.T., & Lutje Spelberg, H.C. (2010b). *Schlichting Test voor Taalproductie-II, Handleiding*. Houten: Bohn Stafleu van Loghum.

- Sijtsma, K. (2009). Over misverstanden rond Cronbachs alfa en de wenselijkheid van alternatieven. *De Psycholoog*, 44, 561-567.
- Visser, L. (2014). The Bayley-III-NL Special Needs Addition. Proefschrift, Rijksuniversiteit Groningen.
- Wechsler, D. (2009). Weschler preschool and primary scale of intelligence - Third Edition: Nederlandstalige bewerking. Amsterdam: Pearson Assessment and Information B.V.

Bijlage A:

Indeling normgroepen naar leeftijd

Instap	Leeftijd
A	16 dagen - 1 maand 15 dagen
B	1 maand 16 dagen - 2 maanden 15 dagen
C	2 maanden 16 dagen - 3 maanden 15 dagen
D	3 maanden 16 dagen - 4 maanden 15 dagen
E	4 maanden 16 dagen - 5 maanden 15 dagen
F	5 maanden 16 dagen - 6 maanden 15 dagen
G	6 maanden 16 dagen - 8 maanden 30 dagen
H	9 maanden en 0 dagen - 10 maanden en 30 dagen
I	11 maanden en 0 dagen - 13 maanden en 15 dagen
J	13 maanden 16 dagen - 69 maanden 15 dagen
K	16 maanden 16 dagen - 19 maanden 15 dagen
L	19 maanden 16 dagen - 22 maanden 15 dagen
M	22 maanden 16 dagen - 25 maanden 15 dagen
N	25 maanden 16 dagen - 28 maanden 15 dagen
O	28 maanden 16 dagen - 32 maanden 30 dagen
P	33 maanden 0 dagen - 38 maanden 30 dagen
Q	9 maanden en 0 dagen - 42 maanden en 15 dagen



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)