NATIONAAL COHORTONDERZOEK ONDERWIJS van het NRO

Technische Toelichting 2de meting Masterplan Basisvaardigheden

Technische Toelichting

December 2024

Versie 1.0

Van Vugt, L., Haelermans, C., Abbink, H., Baumann, S., Havermans, W., Ronda, S. & Smeets, C.

Inleiding

Deze technische toelichting hoort bij de factsheets (2de meting) die het NCO in 2024 heeft uitgebracht over de ontwikkeling van basisvaardigheden in taal en rekenen van leerlingen in het primair onderwijs (PO). In deze factsheets kijken we naar de landelijke cijfers voor de Cito-vaardigheidsscores van begrijpend lezen en spelling ten behoeve van de basisvaardigheid *taalvaardigheid*, en van rekenen-wiskunde ten behoeve van de basisvaardigheid *rekenvaardigheid*. In de factsheets bekijken we de beheersing van de taal- en rekenvaardigheid tot en met schooljaar 2023/2024. We kijken voornamelijk naar de Eind-toets (E-toets) en presenteren zowel de ontwikkeling van de absolute gemiddelde vaardigheidsscores als percentages leerlingen met een bepaald referentieniveau.

In factsheet 1 en 2 tonen we trendanalyses voor taalvaardigheid en rekenvaardigheid voor zes cohorten van groep 3 tot en met groep 7, voor de schooljaren 2018/2019 tot en met 2023/2024. We laten het startniveau in groep 3 zien en vertalen vaardigheidsscores naar referentieniveaus voor groep 6 en 7. In factsheet 3 en 4 splitsen we de analyses van factsheet 1 en 2 uit naar leerling- en schoolkenmerken. In factsheet 3 worden de analyses uitgesplitst naar opleidingsniveau ouders, gezinssamenstelling of geslacht. Factsheet 4 toont een uitsplitsing naar schoolweging, stedelijkheid of schoolgrootte.

Deze technische toelichting dient als toelichting bij de gepresenteerde statistieken en is opgebouwd uit vier onderdelen:

1. Het proces van dataverzameling en bewerking van ruwe LVS-data
2. De additionele databewerking specifiek voor de factsheetdata en de relevante variabelen
3. Representativiteit van de LVS-steekproef
4. Een korte toelichting op de bijgevoegde factsheet statistieken (in Excel)

Privacy

De leerling- en schoolgegevens uit het NCO en de daarop gebaseerde schoolrapportages zijn vertrouwelijk. Ze worden niet gedeeld met en zijn niet toegankelijk voor het Ministerie van Onderwijs, Cultuur en Wetenschap, de Inspectie van het Onderwijs of anderen. Alle informatie wordt door het Centraal Bureau voor de Statistiek (CBS) en onderzoekers van het Nationaal Cohortonderzoek Onderwijs (NCO) in een beveiligde omgeving opgeslagen en bewerkt.

1. Dataverzameling en bewerking ruwe data

1.1 Dataverzameling

Het doel van deze dataverzameling is om meer inzicht te krijgen in de ontwikkeling van vaardigheidsscores van leerlingen in Nederland, zowel op schoolniveau (via school-specifieke rapportages) als op nationaal niveau (via factsheets). Daartoe worden alle schoolbesturen in het PO door het NCO benaderd met het verzoek om gegevens uit het LeerlingVolgSysteem (LVS) beschikbaar te stellen. In het schooljaar 2023/2024 waren zo'n 4.225 scholen aangemeld voor het NCO-LVS project.

De procedure voor het aanleveren van LVS-data door scholen aan het NCO verloopt als volgt: Indien een school na het ontvangen van een informatiepakket aangeeft LVS-data met het NCO te willen delen, moeten besturen eerst een contract tekenen met het Nationaal Regieorgaan Onderwijzonderzoek (NRO). De besturen zijn immers juridisch verantwoordelijk voor, en 'eigenaar' van, de data. Omdat LVS-data zogenaamde 'niet-bijzondere persoonsgegevens' zijn, is het niet nodig dat ouders expliciet toestemming geven voor het delen van de data. Wel moeten de deelnemende scholen alle ouders informeren over het voornemen om de toetsgegevens te leveren aan CBS met het doel deze te koppelen aan de NCO-dataset.[1] Ouders worden in de gelegenheid gesteld om daar bezwaar tegen te maken gedurende een bepaalde periode. De scholen registreren deze bezwaren in het LeerlingAdministratieSysteem (LAS)) van hun softwareleveranciers (Cito LOVS, ParnasSys en ESIS) via een ingebouwde 'bezwaarknop'. Om de feitelijke upload naar het CBS vervolgens te laten plaatsvinden, moeten de scholen op een 'verzendknop' klikken om aan te geven dat men alle procedures doorlopen heeft en de softwareleveranciers tot levering over kunnen gaan. De softwareleveranciers leveren vervolgens de gegevens via een beveiligd kanaal aan het CBS. Het CBS 'verrint' deze gegevens (proces van pseudonimisering) en stelt de data beschikbaar aan het NCO-team. Op deze manier kunnen onderzoekers (in dit geval de onderzoekers van het NCO) niet achterhalen om welke leerlingen en scholen het gaat, maar wél inzichten genereren met behulp van deze data. Het NCO-team bewerkt de data in een beveiligde digitale 'remote access' omgeving, en maakt de factsheets op basis van deze bewerkte data.

[1] Voor meer informatie over de NCO-dataset, zie: Haelermans, C., Huijgen, T., Jacobs, M., Levels, M., van der Velden, R., van Vugt, L., van Wetten, S., (2020). Using Data to Advance Educational Research, Policy, and Practice: Design, Content, and Research Potential of the Netherlands Cohort Study on Education. *European Sociological Review 36*(4), p. 643–662, https://doi.org/10.1093/esr/jcaa027

1.2 Data aanlevering

De data is verzameld via negen leveringen, de exporten van de data vonden plaats op: 30 november 2020, 18 januari 2021, 1 april 2021, 1 augustus 2021, 28 maart 2022, 1 augustus 2022, 13 maart 2023, 1 augustus 2023 en 1 augustus 2024. De gegevens van de leerlingen waarvoor de ouder(s) en/of het bevoegd gezag geen bezwaar hebben gemaakt voor het leveren van de data, zijn via de softwareleveranciers Cito LOVS, ParnasSys en ESIS verstrekt aan het Centraal Bureau voor de Statistiek (CBS). CBS pseudonimiseert vervolgens de identificerende informatie van de leerlingen en van de school. Op deze manier kunnen onderzoekers (in dit geval de onderzoekers van het NCO) niet achterhalen om welke leerling en om welke school het gaat.

In de drie losse datasets van Cito LOVS, ParnasSys en ESIS zitten gegevens die de school in hun softwaresysteem over de leerling heeft geregistreerd zoals:

- Brinnummer (CBS pseudonimiseert deze)
- Vestigingsnummer
- Postcode school
- LeerlingID (CBS pseudonimiseert deze)
- Inschrijfdatum op school
- Eventuele uitschrijfdatum van school
- Jaargroep
- Klasnaam
- Vaardigheidsscore
- OSOtoetscode
- Jaargroep van afname toets
- Afnamedatum toets
- Geslacht
- Geboortemaand en jaar

De volgende groepen zijn opgevraagd:

- Schooljaar 2013/2014: leerlingen groep 3
- Schooljaar 2014/2015: leerlingen groep 3 t/m 4
- Schooljaar 2015/2016: leerlingen groep 3 t/m 5

- Schooljaar 2016/2017: leerlingen groep 3 t/m 6
- Schooljaar 2017/2018: leerlingen groep 3 t/m 7
- Schooljaar 2018/2019: leerlingen groep 3 t/m 8 – *vanaf hier volledig cohort*
- Schooljaar 2019/2020: leerlingen groep 3 t/m 8
- Schooljaar 2020/2021: leerlingen groep 3 t/m 8
- Schooljaar 2021/2022: leerlingen groep 3 t/m 8
- Schooljaar 2022/2023: leerlingen groep 3 t/m 8
- Schooljaar 2023/2024: leerlingen groep 3 t/m 8

In **Tabel 1.1** ziet u een overzicht van de ontvangen data, uitgesplitst naar aantal scholen, aantal unieke leerlingen en aantal toetsrecords per softwareleverancier. De enige selectie die hiervoor heeft plaatsgevonden is dat er enkel gegevens van afnames in groep 3 t/m 8 meegenomen worden. In totaal hebben we data ontvangen van 3.975 scholen. Tabel 1.1 bevat ook historische data.

Tabel 1.1 Inhoud ruwe data

Softwareleverancier	Aantal unieke scholen	Aantal unieke leerlingen	Aantal toetsrecords[2]
Cito[3]	343	73.470	3.184.881
ParnasSys	3.280	942.879	62.356.048
ESIS	766	258.387	18.410.123
Totaal[4, 5]	**3.975**	**1.174.768**	**83.951.052**

1.3 Opschoning van ruwe data

Het NCO heeft de data opgeschoond zodat er zo weinig mogelijk gegevens met administratieve fouten en dubbele leerlingen (bijvoorbeeld door een schoolwissel) in de data voorkomen. Hieronder wordt uitgelegd welke stappen en beslissingen er zijn genomen om tot een opgeschoonde dataset te komen.

[2] Data levering december 2020 + januari 2021 + april 2021 + augustus 2021 + maart 2022 + augustus 2022 + maart 2023 + augustus 2023 + augustus 2024
[3] Van Cito LOVS hebben we geen nieuwe dataleveringen ontvangen.
[4] Leerlingen en scholen kunnen in de verschillende data van de softwareleveranciers voorkomen bijvoorbeeld doordat een leerling van school wisselt die een andere softwareleverancier heeft of omdat de school is overstapt van softwareleverancier. Het totaal is daarom geen exacte optelling van de aantallen van Cito LOVS, ParnasSys en ESIS.
[5] Deze data is inclusief data van andere toetsaanbieders (Boom, Diataal en Bureau ICE (IEP)) en bevat ook SBO-scholen.

Stappen en selecties:

1. Niet alle leerlingen zijn door CBS gekoppeld. Enkel leerlingen die in de Basisregistratie Persoonsgegevens (BRP) (Gemeentelijke Basis Administratie (GBA)-bestand CBS) staan ingeschreven zijn gekoppeld en dus behouden in het bestand.
2. Enkel scholen waarvan het contract is ondertekend houden we in de data. We koppelen de data aan onze eigen administratielijst waarbij enkel de volledig aangemelde scholen worden meegenomen. Dit betekent dat scholen die onverhoopt toch in de data zijn gekomen, terwijl zij bijvoorbeeld niet het contract hebben ondertekend of zich afgemeld hebben, verwijderd zijn uit de data.
3. Een deel van de dubbele toetsrecords lijkt te zijn ontstaan door schoolwisselaars, waarbij de nieuwe school ook de gegevens van de oude school heeft overgenomen. Deze leerlingen hebben op twee verschillende scholen exact dezelfde toetsen gemaakt met dezelfde toetsresultaten. Om hiervoor te corrigeren wordt gekeken naar de in- en uitschrijfdatum van de leerling. Wanneer de toets is afgenomen in de periode tussen de inschrijfdatum en eventuele uitschrijfdatum, is het aannemelijk dat de toets op die desbetreffende school is afgenomen. Op deze manier zijn enkel de toetsrecords overgebleven die op de school ten tijde van de afnamedatum hebben plaatsgevonden.
4. Toetsrecords waar de vaardigheidsscore, OSOtoetscode en afnamedatum van ontbreekt worden verwijderd.
5. Identieke dubbele leerlingen zijn verwijderd. Hierbij is gekeken naar identieke waardes op: brinnummer, vestigingsnummer, geslacht, leerlingID, inschrijfdatum, uitschrijfdatum, jaargroep, klasnaam, vaardigheidsscore, OSOtoetscode, afnamedatum, bronbestand, postcode school en geboortedatum.
6. Van toetsrecords waarbij enkel de jaargroep niet identiek is, maar de overige variabelen wel, wordt de hoogste jaargroep behouden.
7. Onder één rinpersoon kunnen meerdere leerlingID's bestaan en onder één leerlingID blijken soms verschillende personen te zijn gekoppeld. Er is bekeken in hoeverre dit kwam door een eventuele schoolwisseling of door een incorrecte combinatie van rinpersoon en leerlingID. Indien dat laatste het geval bleek hebben wij via de NCO-data gekeken of achterhaald kon worden welke combinatie de juiste was (op basis van geboortedatum). De toetsrecords waar rinpersoon en leerlingID met de geboortedatum overeenkwamen zijn behouden in de data. De overige toetsrecords binnen de dubbelingen zijn verwijderd.

8. Indien een behaalde vaardigheidsscore niet binnen de juiste minimale en maximale score valt die hoort bij de betreffende OSOtoetscode, dan wordt deze op missing gezet.
9. Enkel toetsrecords waarvan de jaargroep van afname bekend is, zijn behouden.
10. Toetsrecords van alle M- en E-toetsen die betrekking hebben op de drie domeinen begrijpend lezen, spelling en rekenen-wiskunde zijn meegenomen.
11. Begrijpend lezen en rekenen-wiskunde generatie 2 toetsen worden omgezet naar generatie 3 toetsen door middel van een formule die is geleverd door Cito.
12. Spelling generatie 2 vaardigheidsscores worden verwijderd aangezien deze niet omgezet kunnen worden naar generatie 3 vaardigheidsscores.
13. Enkel de opgevraagde cohorten zoals genoemd op pagina 6 worden behouden.
14. De verwijzing van een M-toets en E-toets is gebaseerd op de afnamemaand. Toetsen die tussen september en maart zijn afgenomen noemen we M-toets. Toetsen die tussen april en augustus zijn afgenomen noemen we E-toetsen. Bij minder dan 2% van de toetsen blijkt dat de toets niet is afgenomen in het juiste tijdsframe.
15. Sommige leerlingen blijken meerdere keren in een schooljaar getoetst. In eerste instantie wordt de toets meegenomen die op het juiste afnamemoment is afgenomen. Dat wil zeggen: een M-toets in de maanden september t/m maart of een E-toets in de maanden april t/m augustus. Indien dit geen uitsluitsel geeft over welke toets de juiste is wordt de laatst afgenomen toets behouden zodat de data uiteindelijk van iedere leerling per schooljaar maximaal één M-toets en één E-toets bevat.

Na deze stappen zijn de toetsrecords gekoppeld aan de desbetreffende leerling waardoor de data op iedere rij één leerling bevat met daarbij de desbetreffende toetsresultaten over de hele bassischoolloopbaan.

De variabele die (per toets) in het bestand blijven staan zijn:

- Brinnummer + vestigingsnummer school
- Jaargroep ten tijde van de toets
- Afnamedatum toets
- Vaardigheidsscore
- Soort toets (bijvoorbeeld “M7 Digi-toets”)

De variabelen zijn geconstrueerd voor ieder schooljaar van 2013/2014 tot en met 2023/2024 en voor ieder domein: begrijpend lezen, spelling (niet-werkwoorden) en rekenen-wiskunde.

Er zijn nog wel een paar dingen waar rekening mee gehouden moet worden met betrekking tot de Cito-toetsen:

- De generatie 3 toets voor spelling bestaat pas sinds schooljaar 2014/2015, dus voor schooljaar 2013/2014 is geen informatie over spelling bekend.
- De M-toets in begrijpend lezen wordt niet afgenomen in jaargroep 3. Vandaar dat deze gegevens niet aanwezig zijn.

2. Databewerking en variabelen factsheetdata

2.1 Databewerking factsheetdata

Na het bewerken en opschonen van de ruwe LVS-data, worden de data nog verder bewerkt om tot een structuur te komen die geschikt is voor het produceren van statistieken ten behoeve van de factsheets Masterplan Basisvaardigheden.

Stappen en selecties:

1. Ten eerste wordt een correctie gemaakt voor leerlingen die in augustus, september en oktober 2020 een CITO-toets hebben gemaakt. Omdat het hier waarschijnlijk gaat om door COVID-19 vertraagde E2019 toetsen, worden deze scores teruggezet naar die periode. Er wordt een indicator aangemaakt om deze scores nadien te kunnen identificeren.
2. Vervolgens worden de data zo bewerkt dat elke observatie overeenkomt met een combinatie van een unieke leerling en een uniek afnamemoment (zoals '2017E' of '2020M'). Dit beperkt het aantal extra variabelen specifiek voor elk afnamemoment dat nodig is per observatie, en vereenvoudigt de vergelijking van vaardigheidsscores tussen verschillende afnamemomenten. Observaties zonder vaardigheidsscore op ten minste één van de drie domeinen worden uit de dataset verwijderd.
3. De LVS-data kennen een toets-specifieke schoolcode (BRIN), wat een koppeling met schoolkenmerken compliceert. Indien de BRINs behorend bij de gemaakte toetsen op één afnamemoment niet overeenkomen, wordt de meest-voorkomende BRIN als uitgangspunt genomen.
4. Toetsscores van verschillende toetsaanbieders kunnen niet met elkaar worden vergeleken. Leerlingen met toetsrecords van toetsen met een andere toetsaanbieder dan Cito (Boom, Bureau ICE (IEP) en Diataal) worden verwijderd uit de data. Ditzelfde geldt voor alle leerlingen op scholen waar op enig toetsmoment op ten minste één domein meer dan 75% van de afgenomen toetsen van een andere toetsaanbieder dan CITO afkomstig is. Dit laatste gebeurt om de impact van het selectie-effect van scholen die overstappen van toetsleveranciers op de uitkomsten te beperken.
5. Enkel vaardigheidsscores van leerlingen uit het reguliere basisonderwijs kunnen op een juiste manier worden meegenomen in onze analyses. Voor zover vaardigheidsscores behorend bij het speciaal basisonderwijs niet al zijn verwijderd uit de ruwe data, wordt dit alsnog gedaan.

6. Vaardigheidsscores die negatief zijn of gelijk zijn aan 0 worden uit de dataset verwijderd omdat dit geen realistische waarden zijn. Observaties die ten gevolge van bovenstaande aanpassingen op dit punt geen vaardigheidsscore op ten minste één van de drie domeinen meer hebben, worden uit de dataset verwijderd.
7. Direct voorafgaand aan het berekenen van de statistieken die ten grondslag liggen aan de factsheets, worden de vaardigheidsscores die behoren tot het eerste en honderdste percentiel van de vaardigheidsscoredistributie per domein, schooljaar, afnamemoment en jaargroep in de data geïdentificeerd. Deze waarden worden vervolgens op 'missing' gezet. Dit dient ertoe het effect van extreme uitschieters op de berekende statistieken te beperken.

2.2 Variabelen factsheetdata

2.2.1 Doelvariabelen

Vaardigheidsscore: de meeste figuren in de factsheets presenteren gemiddelde vaardigheidsscores op de domeinen rekenen-wiskunde, begrijpend lezen en spelling. Omdat vaardigheidsscores van verschillende toetsleveranciers vooralsnog niet eenvoudig met elkaar kunnen worden vergeleken, baseren we ons enkel op Cito-vaardigheidsscores (welke het meest voorkomen). In vrijwel alle gevallen tonen we vaardigheidsscores ten tijde van E-toetsen, gemiddeld per schooljaar en groep.

Referentieniveaus: voor een aantal figuren in de factsheets zijn de vaardigheidsscores vertaald naar referentieniveaus.[6] De fundamentele niveaus (F-niveaus) beschrijven wat zo veel mogelijk leerlingen op een bepaald moment in de schoolloopbaan moeten beheersen op het gebied van taal en rekenen. Het streefniveau (S-niveau) is voor leerlingen die meer aankunnen. De gemiddelden bij deze doelvariabele vertegenwoordigen hiermee het aandeel leerlingen in een bepaalde groep dat het beschreven referentieniveau heeft behaald. Niveau 1F is het minimale fundamentele niveau voor het einde van het primair onderwijs. Niveau 1S correspondeert met het bijbehorende streefniveau voor rekenen, terwijl bij begrijpend lezen voor dit streefniveau de standaard 2F wordt gehanteerd.

[6] Zie Expertgroep doorlopende leerlijnen taal en rekenen. (2008). Over de drempels met taal en rekenen : hoofdrapport van de Expertgroep doorlopende leerlijnen taal en rekenen. Enschede: SLO. https://www.slo.nl/@4230/drempels-taal/

2.2.2 Leerlingkenmerken

Bij leerlingkenmerken gebruiken we telkens de waarde van de kenmerken zoals ze waren in het schooljaar waarvoor een statistiek wordt weergegeven.

Jaargroep: Dit betreft de groep waar de leerling in zat op het moment dat de toets werd afgenomen, zoals geregistreerd door DUO.

Opleidingsniveau ouders: Indien minstens één van de juridische ouders hoogopgeleid was (minimaal een hbo-opleiding afgerond), dan valt de leerling onder de categorie 'hoog opgeleide ouders' (2). Indien de hoogst behaalde opleiding van minimaal één van de ouders mbo 2-4, havo of vwo was, dan valt de leerling onder de categorie 'gemiddeld opgeleide ouders' (1) en indien beide ouders laag opgeleid zijn (maximaal vmbo-gt, havo/vwo onderbouw) dan valt de leerling onder de categorie 'lage opgeleide ouders' (0).

Gezinssamenstelling/ouderlijke structuur: Er is een opsplitsing gemaakt tussen leerlingen die in tweeoudergezinnen wonen en leerlingen die in eenoudergezinnen wonen. Onder tweeoudergezinnen verstaan wij leerlingen die wonen met beide juridische ouders of één van de juridische ouders met een partner (0). Onder eenoudergezinnen vallen leerlingen die staan ingeschreven in een huishouden met één juridische ouder, zonder partner (1). Merk op dat het hier gaat om de geregistreerde inwoners bij de Basisregistratie Persoonsgegevens (BRP) (Gemeentelijke Basis Administratie (GBA)-bestanden CBS) waardoor het kan voorkomen dat als bijvoorbeeld de partner van de moeder (nog) niet staat ingeschreven in dit huishouden, de leerling onterecht als wonend in een eenoudergezin wordt beschouwd. Leerlingen die zonder ouders wonen, bijvoorbeeld omdat ze begeleid wonen, zijn buiten beschouwing gelaten.

Geslacht: Geslacht zoals vastgelegd in de BRP. In geval van geslachtwijziging is alleen het laatst geldende geslacht vastgelegd. Personen waarvan het geslacht onbekend is worden als vrouw bestempeld in lijn de CBS werkwijze. meisje (1) of jongen (0).

Een overzicht van de beschrijvende statistieken vindt u in **Tabel 2.1**.

Tabel 2.1 Beschrijvende statistieken leerlingkenmerken

Beschrijving	Aantal observaties	Percentage
Opleiding ouders		
Laag opgeleid	168.451	9%
Gemiddeld opgeleid	574.811	30%
Hoog opgeleid	947.481	50%
Onbekend	197.839	10%
Ouderlijke structuur		
Eenoudergezinnen	311.619	17%
Tweeoudergezinnen	1.550.503	82%
Zonder ouders	11.375	1%
Onbekend	15.085	1%
Geslacht		
Jongens	665.007	50%
Meisjes	662.292	50%
Onbekend	11.470	1%

2.2.3 Leerlingenpopulatie- en schoolkenmerken

Net als bij de leerlingkenmerken gebruiken we telkens de waarde van de kenmerken zoals ze waren in het schooljaar waarvoor een statistiek wordt weergegeven.

Schoolweging: De schoolweging wordt berekend door het CBS aan de hand van vijf omgevingskenmerken: het opleidingsniveau van de ouders, het gemiddeld opleidingsniveau van alle moeders op school, het land van herkomst van de ouders, de verblijfsduur van de moeder in Nederland en of de ouders in de schuldsanering zitten.[7] De schoolweging heeft een schaal tussen 20 en 40, en is een maat voor de complexiteit van de leerlingenpopulatie op een school. Een hoge weging betekent een hoge mate van complexiteit. We hebben zeven categorieën onderverdeeld; drie waarden van de schoolweging in één categorie, waarbij de hoogste vier en laagste vier waarden samen genomen zijn in (elk) een eigen categorie.

Stedelijkheid: Dit is gebaseerd op de plaats waar de school staat en het aantal adressen per km^2. We hebben dit onderverdeeld in vijf categorieën: 0) zeer sterk stedelijk (>= 2500 adressen/km^2), 1) sterk stedelijk (1500 – 2500 adressen/km^2), 2) matig stedelijk (1000 – 1500 adressen/km^2), 3) weinig stedelijk (500 – 1000 adressen/km^2) en 4) niet stedelijk (<500 adressen/km^2).

Schoolgrootte: Dit is gebaseerd op het aantal leerlingen op een school en onderverdeeld in drie categorieën: 0) kleine scholen (maximaal 140 leerlingen), 1) gemiddelde scholen (tussen de 141 en 220

[7] Zie https://www.onderwijsinspectie.nl/trends-en-ontwikkelingen/onderwijsdata/schoolweging-po.

leerlingen) en 2) grote scholen (minimaal 221 leerlingen). De categorieën zijn, op basis van NCO-data, in drie ongeveer gelijke groepen verdeeld.

Een overzicht van beschrijvende statistieken van de leerlingenpopulatie- en schoolkenmerken vindt u in **Tabel 2.2**.

Tabel 2.2 Beschrijvende statistieken leerlingenpopulatie- en schoolkenmerken

Beschrijving	Aantal observaties	Percentage
Schoolweging		
Schoolweging <23	94.441	5%
Schoolweging 23-25	248.538	13%
Schoolweging 26-28	546.712	29%
Schoolweging 29-31	533.581	28%
Schoolweging 32-34	223.666	12%
Schoolweging 35-37	161.781	9%
Schoolweging 38+	49.810	3%
Onbekend	30.053	3%
Stedelijkheid		
Zeer sterk	461.279	24%
Sterk	593.792	31%
Matig	287.100	15%
Weinig	394.455	21%
Niet	140.486	7%
Onbekend	11.470	1%
Schoolgrootte		
Tot en met 140 leerlingen	430.681	23%
141-220 leerlingen	563.436	30%
221 leerlingen of meer	882.995	47%
Onbekend	11.470	1%

3. Representativiteit

Voor alle leerling- en schoolkenmerken hebben we een representativiteitscheck uitgevoerd om te kijken of de LVS-steekproef representatief is vergeleken met de NCO-data. De NCO-data bevat alle leerlingen in het Nederlands bekostigd onderwijs in het PO. Door te vergelijken met de NCO-data kunnen we zien welk type leerlingen en/of scholen in de LVS-steekproef over- of ondervertegenwoordigd zijn.

We zien dat de LVS-steekproef op sommige kenmerken afwijkt van de algehele populatie. De verschillen zijn echter relatief klein. Omdat de NCO-data gemiddeld bijna een miljoen leerlingen per jaar bevat, zijn ook kleine verschillen significant. We kijken daarom niet naar significantie, maar in hoeverre het procentuele verschil afwijkt van 0. In **Tabel 3.1** staan de uitkomsten van de representativiteitscheck van de leerlingkenmerken. Te zien is dat de verschillen tussen de NCO-populatie en de LVS-steekproef erg klein zijn. In **Tabel 3.2** staan de uitkomsten van de representativiteitscheck van de schoolkenmerken. Betreft schoolkenmerken werkt de LVS-steekproef sterker af van de NCO-populatie. De steekproef bevat bijvoorbeeld meer leerlingen van scholen met een schoolweging hoger dan 35, en juist minder leerlingen van scholen met een schoolweging lager dan 32. Ook zijn leerlingen op scholen in matig-stedelijk gebied ondervertegenwoordigd. Ten slotte zijn ook grote scholen meer aanwezig in de LVS-steekproef dan in de NCO-populatie.

Tabel 3.1 Representativiteitscheck voor leerlingkenmerken

	Nationaal Cohortonderzoek		LVS-steekproef				
Variabelen	**Percentage**	**Non-missing N**	**Percentage**	**Non-missing N**	**Verschil**	**T-statistic**	**P-waarde**
Opleidingsniveau ouders							
Laag	10,08%	3.746.726	9,96%	1.690.743	0,12%	4,2	0,00
Gemiddeld	34,60%	3.746.726	34,00%	1.690.743	0,61%	13,8	0,00
Hoog	55,32%	3.746.726	56,04%	1.690.743	-0,72%	-15,7	0,00
Ouderlijke structuur							
Tweeoudergezinnen	81,87%	4.225.504	82,76%	1.873.497	-0,89%	-26,5	0,00
Eenoudergezinnen	17,25%	4.225.504	16,63%	1.873.497	0,62%	18,8	0,00
Geslacht							
Jongens	50,91%	4.253.743	49,98%	1.877.112	0,94%	21,4	0,00
Meisjes	49,09%	4.253.743	50,02%	1.877.112	-0,94%	-21,4	0,00

Tabel 3.2 Representativiteitscheck voor leerlingenpopulatie- en schoolkenmerken

	Nationaal Cohortonderzoek		LVS-steekproef				
Variabelen	**Percentage**	**Non-missing N**	**Percentage**	**Non-missing N**	**Verschil**	**T-statistic**	**P-waarde**
Schoolweging							
Schoolweging <23	6,54%	4.032.043	5,05%	1.882.020	1,49%	70,8	0,00
Schoolweging 23-25	12,82%	4.032.043	13,27%	1.882.020	-0,44%	-14,9	0,00
Schoolweging 26-28	27,96%	4.032.043	29,49%	1.882.020	-1,53%	-38,3	0,00
Schoolweging 29-31	31,13%	4.032.043	28,70%	1.882.020	2,43%	59,8	0,00
Schoolweging 32-34	12,76%	4.032.043	12,12%	1.882.020	0,64%	21,7	0,00
Schoolweging 35-37	7,02%	4.032.043	8,70%	1.882.020	-1,68%	-72,0	0,00
Schoolweging 38+	1,77%	4.032.043	2,67%	1.882.020	-0,90%	-72,1	0,00
Stedelijkheid							
Zeer sterk	23,9%	4.253.802	24,6%	1.877.112	-0,68%	-18.1	0,00
Sterk	30,8%	4.253.802	31,6%	1.877.112	-0,87%	-21.4	0,00
Matig	17,3%	4.253.802	15,3%	1.877.112	1,99%	61.0	0,00
Weinig	21,3%	4.253.802	21,0%	1.877.112	0,31%	8.5	0,00
Niet	6,7%	4.253.802	7,5%	1.877.112	-0,75%	-33,7	0,00
Schoolgrootte							
< 140 leerlingen	24,78%	4.253.802	22,94%	1.877.112	1,83%	48,8	0,00
141 tot 220 leerlingen	31,60%	4.253.802	30,02%	1.877.112	1,58%	39,0	0,00
> 221	43,62%	4.253.802	47,04%	1.877.112	-3,42%	-78,5	0,00

4. Toelichting op statistieken

De statistieken die ten grondslag liggen aan de figuren die worden gepresenteerd in de factsheets Masterplan Basisvaardigheden, zijn te vinden in de Excel-bijlage bij deze Technische Toelichting. De naam van de sheet verwijst telkens naar achtereenvolgend (1) het type figuur/statistiek dat is opgenomen in de sheet, (2) het domein waarover statistieken worden gepresenteerd, en (3) het eventueel achtergrondkenmerk waarnaar de statistieken zijn uitgesplitst.

De Excel-sheets tonen in de eerste kolommen een beschrijving van de groep waarvoor statistieken worden gepresenteerd met in de kolommen daarna achtereenvolgend het aantal observaties (unieke leerling*schooljaar combinaties), de gemiddelde waarde van de doelvariabele[8], en de bijbehorende standaarddeviatie (SD). Daarnaast worden de ondergrens en bovengrens van een 95% betrouwbaarheidsinterval rond het gemiddelde weergegeven. Bij sheets waarin wordt uitgesplitst naar achtergrondkenmerken wordt onderaan de sheet de definitie van het betreffende achtergrondkenmerk weergegeven.

4.1 Type figuren/statistieken

Trend

Deze trendstatistieken worden gebruikt om de gemiddelde vaardigheidsscore per groep te laten zien voor verschillende schooljaren (zoals factsheet 1 figuur 1) óf voor verschillende niveaus van een achtergrondkenmerk in één schooljaar (zoals factsheet 3 figuur 1). De bijbehorende Excel-sheetnamen beginnen met ‘Trend’.

Cohort

Deze statistieken worden gebruikt voor het weergeven van de gemiddelde vaardigheidsscore-ontwikkeling voor verschillende cohorten (zoals factsheet 1 figuur 2) óf voor verschillende niveaus van een achtergrondkenmerk binnen één cohort (zoals factsheet 3 figuur 1). Als uitgangspunt voor een cohort gebruiken we het schooljaar waarop een leerling in groep 7 zit, waarbij we per schooljaar terug redeneren in welke groep diezelfde leerling in het eerdere schooljaar had moeten zitten bij reguliere doorstroom.

[8] Zie 2.2.1 voor een overzicht van de doelvariabelen.

De gehanteerde cohortdefinitie is dynamisch in de zin dat zittenblijvers en leerlingen die een groep overslaan van cohort veranderen. De bijbehorende Excel-sheetsnamen beginnen met ‘Cohort’.

Startniveau

Deze statistieken worden gebruikt voor het tonen van de gemiddelde vaardigheidsscore ten tijde van de eerste LVS-toets in groep 3 per schooljaar (zoals factsheet 1 figuur 3). Merk op dat het hier in tegenstelling tot alle andere figuren bij de domeinen rekenen-wiskunde en spelling om een M-toets gaat. Omdat in groep 3 geen M-toets begrijpend lezen wordt afgenomen, betreft het hier wél een E-toets, en komen de weergegeven statistieken normaalgesproken ook overeen met die voor groep 3 in de sheet met cross-sectionele gegevens. De bijbehorende Excel-sheetnamen beginnen met ‘Start’.

Referentieniveau

Deze statistieken worden gebruikt om te laten zien welk percentage leerlingen in groep 6 en 7 de referentieniveaus voor het eind van het primair onderwijs (1F en 1S/2F) al heeft behaald voor verschillende schooljaren (zoals factsheet 1 figuur 4) óf voor verschillende niveaus van een achtergrondkenmerk in één schooljaar (zoals factsheet 3 figuur 4). De bijbehorende Excel-sheetnamen beginnen met ‘Ref’.

4.2 Exceldocument ‘Statistieken bij Figuren Factsheets Masterplan Basisvaardigheden’

Excel-sheetnamen met de toevoeging ‘alg’ wijzen erop dat statistieken in de betreffende sheet enkel zijn uitgesplitst naar jaargroepen en schooljaren. Excel-sheetnamen met een andere toevoeging zijn daarnaast binnen jaargroepen en schooljaren verder uitgesplitst naar een specifiek achtergrondkenmerk. Zie hoofdstukken 2.2.2 en 2.2.3 voor een nadere beschrijving van de beschikbare achtergrondkenmerken.

4.3 Exeldocument ‘Aanvullende Statistieken bij Factsheets Masterplan Basisvaardigheden’

Deze Excel bevat de statistieken uitgesplitst naar specifieke achtergrondkenmerken voor de domeinen begrijpend lezen en spelling.