Leervertraging door schoolsluitingen in het voortgezet onderwijs - Deel 2

Effecten na de eerste en tweede lockdownperiode

Stichting Cito

Onderzoek, Kennis & Innovatie

Auteurs
dr. Wobbe Zijlstra
dr. Silvia Rietdijk
drs. Servaas Frissen
dr. Remco Feskens

Copyright © 2021 Stichting Cito Instituut voor Toetsontwikkeling Arnhem. Dit onderzoek is met medewerking van Cito BV tot stand gekomen.

Alle rechten voorbehouden. Niets uit dit werk mag zonder voorafgaande schriftelijke toestemming van Stichting Cito Instituut voor Toetsontwikkeling worden openbaar gemaakt en/of verveelvoudigd door middel van druk, fotokopie, scanning, computersoftware of andere elektronische verveelvoudiging of openbaarmaking, microfilm, geluidskopie, film- of videokopie of op welke wijze dan ook.

Managementsamenvatting

In deze studie hebben we de effecten van de twee periodes van schoolsluitingen in 2020 en 2021 op de leerresultaten van de leerlingen in de onderbouw van het voortgezet onderwijs in kaart gebracht. Hierbij hebben we gekeken naar de ontwikkeling van de vaardigheid van leerlingen op zes onderdelen, en de leerlingprestaties in het schooljaar 2020/2021 vergeleken met die in de laatste "voor-corona" schooljaren. De zes onderdelen zijn Nederlands leesvaardigheid (nlv), Nederlands woordenschat (nws), Engels leesvaardigheid (elv), Engels woordenschat (ews), Rekenen (rek) en Wiskunde (wis). Dit onderzoek is het tweede deel van een rapportage rondom leerachterstanden in het voortgezet onderwijs. Het eerste deel van deze rapportage heeft gebruik gemaakt van afnamegegevens verzameld in het schooljaar 2019/2020. Dit tweede deel van de rapportage geeft een meer compleet beeld over de corona-gerelateerde vertragingen doordat we in dit rapport gebruik maken van gegevens verzameld na beide periodes van schoolsluitingen.

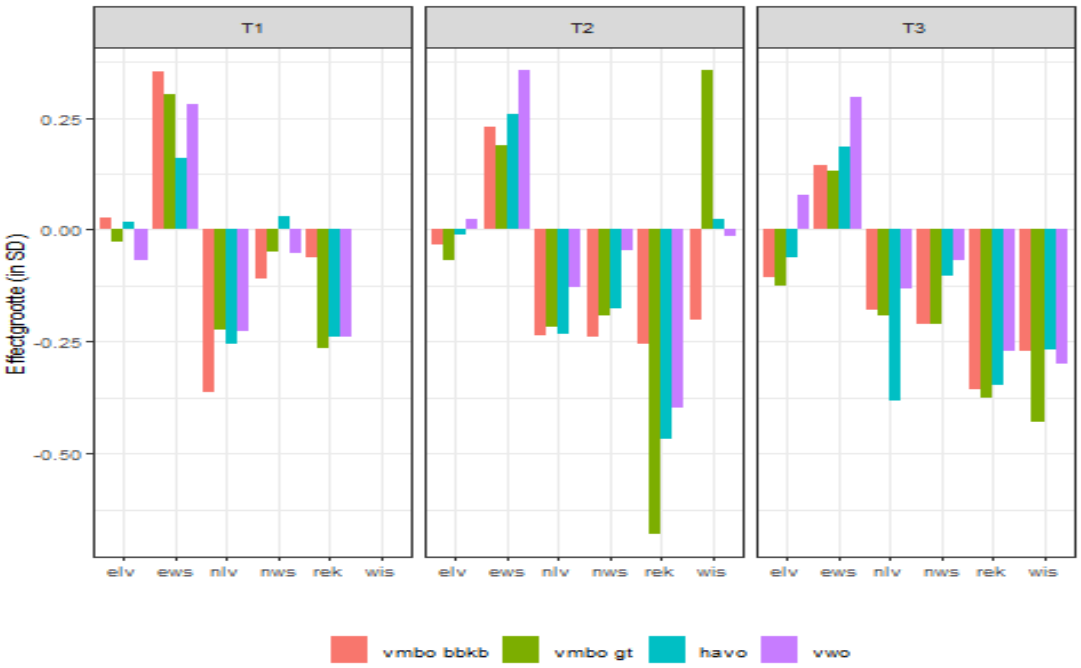
Leerlingprestaties

De belangrijkste vraag in deze studie was hoe de gemiddelde leerlingprestaties op de zes onderdelen in 2020/2021 verschillen van die in 2018/2019. Een daling van de gemiddelde scores kan namelijk wijzen op een negatief effect van de schoolsluitingen. Om zorgvuldig conclusies te kunnen trekken, vergeleken we gematchte leerlinggroepen en namen we ook de trends in eerdere schooljaren (vanaf 2016/2017) in ogenschouw. Onze bevindingen:

- De leerlingprestaties zijn vooral achteruit gegaan op Nederlands leesvaardigheid en Rekenen. Dit komt ook tot uitdrukking in de verdeling over referentieniveaus: relatief meer leerlingen presteren in 2020/2021 onder of op het 1F-niveau, en minder leerlingen op het 2F- of 3F-niveau.
- De Engelse woordenschat van leerlingen wordt al jaren steeds beter en in 2020/2021 is deze alleen maar nog sneller verbeterd.

- Op het onderdeel Engels leesvaardigheid is er weinig veranderd in 2020/2021.
- Voor Wiskunde geldt dat er in het tweede leerjaar - buiten een stijging van de resultaten binnen vmbo gt die we moeilijk kunnen verklaren - de resultaten stabiel zijn gebleven. In het derde leerjaar zien we echter ook bij Wiskunde een daling.
- In het vmbo en de havo zien we meer tekenen van leerachterstand dan in het vwo.

In Figuur 1 wordt een overzicht gegeven van de belangrijkste resultaten van deze studie. In deze figuur zijn de gemiddelde prestaties na de periode van thuisonderwijs in het voorjaar van 2020/2021 vergeleken met de gemiddelde prestaties van vergelijkbare leerlingen in 2018/2019. De verschillen tussen de gemiddelde prestaties zijn uitgedrukt als effectgroottes (in standaarddeviaties). Een negatieve effectgrootte betekent dat de gemiddelde prestatie in 2020/2021 lager was, en duidt dus op leerachterstanden.



Figuur 1: Overzicht effectgroottes

Nederlands leesvaardigheid en Rekenen zijn dus de vakken die het meest te maken hebben gehad met een daling in de leerresultaten. Om de effectgroottes van deze vakken in perspectief te zetten, kunnen we gebruik maken van de het percentage leerlingen dat onder of op niveau 1F tussen 2019 en 2021 heeft gepresteerd. Voor Nederlands leesvaardigheid geldt dat er in 2020/2021 in alle drie de leerjaren aanzienlijk meer vmbo leerlingen zijn die onder het 1F niveau presteren, en aanzienlijk minder die op 2F niveau presteren, dan in de eerdere schooljaren. Ook bij de havo leerlingen zien we dat de verdeling minder gunstig is geworden in 2020/2021: de percentages voor <1F en 1F zijn hoger, en die voor 2F en 3F lager. Bij het vwo zijn de verschillen in het tweede en derde leerjaar klein, maar in het eerste leerjaar vwo is er wel een opvallende stijging (met samen zo'n 10%) van de categorieën <1F en 1F. Ook voor Rekenen geldt dat in de meeste schooltypen en leerjaren aanzienlijk meer leerlingen bleven steken op <1F of 1F niveau en minder leerlingen het 2F of 3F niveau behaalden.

Verschillen tussen jongens en meisjes

In het onderzoek keken we specifiek naar verschillen tussen jongens en meisjes. Presteerden meisjes of jongens in coronatijden anders dan daarvoor, met groeiende verschillen tot gevolg? In 2020/2021 bleken sekseverschillen behoorlijk stabiel. Alleen voor Nederlands woordenschat groeiden de verschillen, omdat meisjes op sommige schooltypen en in sommige leerjaren lager gingen scoren. Daarentegen werden de verschillen voor Engels woordenschat kleiner. Meisjes scoorden daar in coronatijd juist hoger.

Verschillen tussen scholen op basis van hun achterstandsscore

De achterstandsscores van vo-scholen (CBS, 2021) gebruikten we om de scholen in onze data te verdelen in vier groepen: van laagste naar hoogste. De vraag was of we in coronatijd andere trends zagen op scholen met verschillende achterstandsscores. Dit bleek niet zo te zijn; de trend over de tijd was niet anders voor de vier groepen van scholen. Hierbij moet wel worden opgemerkt dat de achterstandsscore sowieso alleen maar in het vmbo bbkb enige samenhang leek te vertonen met de gemiddelde leerlingprestaties. Hierbij moet worden opgemerkt dat deze analyses naar het verband tussen leerachterstanden en achterstandsscores verricht zijn *binnen* onderwijstypes. Dat wil niet zeggen dat er geen verschillen in achterstandsscores *tussen* onderwijstypes bestaan. De eerdere conclusie dat in vmbo en havo meer tekenen van leerachterstand zijn gevonden dan in het vwo (waar de achterstandsscores gemiddeld significant lager zijn), duidt mogelijk toch op enig verband tussen de mate van leerachterstand en de achterstandsscore bezien over alle niveaus in het vo, zoals ook gevonden in het primair onderwijs[1].

Spreiding in de leerlingprestaties

De verschillen *tussen* scholen in hun leerlingprestaties op de zes onderdelen zijn in 2020/2021 vaak groter dan voorheen. Dit geldt met name voor scholen in havo en vwo. De spreiding in leerlingprestaties *binnen* scholen is soms ook wel toegenomen, maar dan in veel minder sterke mate. Samen wijst dit erop dat het verplichte afstandsonderwijs de verschillen tussen scholen meer heeft (uit-)vergroot dan de verschillen tussen individuele leerlingen.

Samenvatting

Samenvattend kunnen we uit deze studie concluderen dat de schoolsluitingen en andere invloeden van corona op het onderwijsleerproces weldegelijk effect hebben gehad op de leerprestaties van vo-leerlingen. Dit effect was vooral negatief voor Nederlands leesvaardigheid en Rekenen, maar juist positief voor Engels woordenschat. Globaal gezien vonden we in het vmbo en havo vaker (dat wil zeggen, bij meer domeinen of leerjaren) indicaties van corona-gerelateerde leerachterstanden dan in het vwo. Verder zagen we bij havo en vwo vaker en sterkere toenames in de verschillen tussen scholen dan in het vmbo.

[1] `https://www.nationaalcohortonderzoek.nl/`

Inhoudsopgave

1 Introductie 9

2 Data 11

2.1 Het Cito volgsysteem voortgezet onderwijs 11

2.2 Afnamegegevens in het cvvo 12

2.3 Externe data 14

2.4 Referentie- en ERK niveaus 14

3 Methoden 15

3.1 Analyse van de trends 16

3.2 Representativiteit 18

3.2.1 Visuele evaluatie van representativiteit 18

3.2.2 R-indicator 25

3.3 Het maken van vergelijkbare groepen 27

3.4 Spreiding in de toetsresultaten 27

4 Resultaten 29

4.1 Trendresultaten 29

4.1.1 Nederlands leesvaardigheid 29

4.1.2 Nederlands woordenschat 32

4.1.3 Engels leesvaardigheid 34

4.1.4 Engels woordenschat 36

4.1.5 Rekenen 37

4.1.6 Wiskunde 39
4.2 Trends voor subpopulaties 41
4.2.1 Geslacht 41
4.2.2 Achterstandsscore van de school 42
4.3 Spreiding in de toetsscores 43
4.3.1 Nederlands leesvaardigheid 43
4.3.2 Nederlands woordenschat 44
4.3.3 Engels leesvaardigheid 45
4.3.4 Engels woordenschat 45
4.3.5 Rekenen 46
4.3.6 Wiskunde 47

5 Conclusie en Discussie 49

6 Referenties 53

A Appendix 55
A.1 Effectgroottes 2018/2019 tussen aansluitende leerjaren of schooltypen 56
A.2 Figuren trendanalyse naar geslacht 59
A.3 Figuren trendanalyse naar achterstandsscore van de school 65
A.4 Sensitiviteitsanalyses 71

1. Introductie

In de afgelopen zestien maanden heeft de coronapandemie grote gevolgen gehad voor het onderwijs, van het basisonderwijs tot aan het hoger onderwijs. Gedurende twee periodes moesten de scholen sluiten en overgaan op online afstandsonderwijs voorzover mogelijk. Voor het voortgezet onderwijs betrof dit eerst de periode van 15 maart tot 2 juni 2020, en in het schooljaar erna de periode van 16 december 2020 tot 1 maart 2021[1]. In de tussenliggende periode en de eerste tijd hierna was er echter ook nog steeds sprake van beperkende maatregelen. Hierdoor konden niet alle leerlingen tegelijk op school aanwezig zijn, en moest er nog veel onderwijs op afstand plaatsvinden. Ook kwam het voor dat scholen tijdelijk volledig moesten sluiten vanwege corona-uitbraken[2]. In de maanden maart tot en met mei van 2021 gold de regel dat leerlingen minimaal één dag per week op school les moesten krijgen[3]. Vanaf 31 mei mochten de middelbare scholen weer helemaal opengegaan, en vanaf 7 juni zijn zij verplicht om al hun leerlingen weer volledig fysiek les te geven op school[4]. Daarbij gelden er wel nog steeds regels en richtlijnen om de verspreiding van corona tegen te gaan, zoals het gebruik van zelftesten en mondkapjes en het houden van anderhalve meter afstand voorzover dat praktisch haalbaar is.

Er zijn al een aantal onderzoeken uitgevoerd naar de effecten van de eerste lockdown-periode op de leerprestaties van leerlingen in het voortgezet onderwijs. Emons, Frissen en Straat (2021) gebruikten hiervoor cross-sectionele data uit het Cito volgsysteem voortgezet onderwijs (cvvo), en Zijlstra, Frissen, Koops, van Onna & Feskens (2021) hebben later ook longitudinale methoden toegepast op de afnames van deze volgtoetsen tot en met het

[1] https://www.vo-raad.nl/nieuws/online-onderwijs-blijft-tot-1-maart-de-norm

[2] https://www.rijksoverheid.nl/actueel/nieuws/2020/12/03/anderhalf-procent-van-de-scholen-tijdelijk-gesloten

[3] https://www.rijksoverheid.nl/actueel/nieuws/2021/02/23/middelbare-scholen-en-het-middelbaar-beroepsonderwijs-gaan-weer-open

[4] https://www.rijksoverheid.nl/actueel/nieuws/2021/05/22/middelbare-scholen-volledig-open

schooljaar 2019/2020 (inclusief uitgestelde afnames in het najaar van 2020). Daarbij is gekeken naar gemiddelde prestaties en spreiding op het niveau van cohorten, maar ook naar de groeiresultaten van leerlingen. Inmiddels zijn er nieuwe afnames geweest van de cvvo toetsen in het schooljaar 2020/2021. Dit maakt het mogelijk om nu ook te kijken naar de toetsresultaten na de tweede periode van schoolsluitingen en afstandsonderwijs. Het huidige onderzoek heeft dan ook als doel om in kaart te brengen hoe de leerlingen in het voortgezet onderwijs er nu voor staan op de onderdelen die worden getoetst in het cvvo. De aanpak is ten dele vergelijkbaar met die van het vorige onderzoek (Zijlstra, Frissen, Koops, van Onna, & Feskens, 2021). We zullen de volgende deelvragen addresseren:

1. Zijn er trends in de leerresultaten van de leerlingen in de onderbouw van het voortgezet onderwijs die verband houden met de lockdownperiodes?
2. Zijn er verschillen in de trendresultaten tussen verschillende groepen leerlingen?
3. Wat is het effect van de lockdownperiodes op de spreiding van de leeruitkomsten?

Merk op dat de onderzoeksvragen betrekking hebben op het cumulatieve effect van de coronacrisis en de twee periodes van thuisonderwijs, en niet op een geïsoleerd effect van de tweede lockdownperiode. De leerlingen hebben immers allemaal de gevolgen ondervonden van beide periodes van schoolsluitingen, of zij nu in de eerste lockdown al in het voortgezet onderwijs zaten, of nog in groep 8 van het basisonderwijs. Daardoor is het niet mogelijk om een zuiver onderscheid te maken tussen de effecten van de eerste en tweede periode van thuisonderwijs op de leerresultaten. Wel kunnen we de huidige prestaties van de leerlingen vergelijken met de prestaties van leerlingen voordat de coronacrisis begon. Dit doen we voor de zes verschillende onderdelen die in het cvvo worden afgenomen, en we gebruiken matching technieken om de leerresultaten van dit schooljaar zo zuiver mogelijk te vergelijken met die uit eerdere jaren.

Het vervolg van dit rapport is als volgt opgebouwd: In Hoofdstuk 2 beschrijven we de data die in deze studie gebruikt wordt. In Hoofdstuk 3 worden de methoden toegelicht, evenals de representativiteitsanalyse van de data. Vervolgens bespreken we de antwoorden op de onderzoeksvragen in Hoofdstuk 4, en in Hoofdstuk 5 sluiten we af met de conclusies en discussie van de resultaten.

2. Data

In dit hoofdstuk geven we eerst een beschrijving van het Cito volgsysteem voortgezet onderwijs. Vervolgens gaan we in op de afnamegegevens die we tot onze beschikking hebben. In Sectie 2.3 bespreken we de externe data die we hebben gebruikt om verschillen in kaart te brengen tussen de afnamegegevens en de landelijke gegevens. We eindigen dit hoofdstuk met een Sectie over de referentie- en ERK-niveaus (Sectie 2.4).

2.1 Het Cito volgsysteem voortgezet onderwijs

Het Cito volgsysteem voortgezet onderwijs (cvvo) is een leerlingvolgsysteem voor de onderbouw van het voortgezet onderwijs dat ontwikkeld is door Cito BV[1]. Het cvvo wordt op vier verschillende afnamemomenten afgenomen: Aan het begin van het eerste jaar van het voortgezet onderwijs (Toets 0); aan het einde van het eerste jaar (Toets 1); de derde afname vindt plaats aan het einde van het tweede jaar (Toets 2); Toets 3 tenslotte wordt afgenomen aan het einde van het derde jaar. Als leerlingen alle toetsen hebben gemaakt kunnen hun leerprestaties dus gevolgd worden van het begin van het voortgezet onderwijs tot en met het einde van de onderbouw. Binnen het cvvo worden verschillende vaardigheden gemeten. De belangrijkste vaardigheden of onderdelen zijn Nederlands leesvaardigheid (nlv), Nederlands woordenschat (nws), Engels leesvaardigheid (elv), Engels woordenschat (ews), Rekenen (rek) en Wiskunde (wis). Scholen kunnen zelf beslissen welke onderdelen zij willen afnemen op welk meetmoment.

Het cvvo kenmerkt zich door het gebruik van verschillende toetsversies. Hier is voor gekozen omdat de niveauverschillen binnen het voortgezet onderwijs te groot zijn om alle leerlingen dezelfde opgaven voor te leggen. Om deze redenen zijn er verschillende toetsversies ontwikkeld specifiek gericht op het vmbo, havo en vwo. In verband met

[1] Meer informatie over het cvvo is te vinden op de volgende website: `https://www.cito.nl/onderwijs/voortgezet-onderwijs/volgsysteem-vo`.

toetsverversing worden er periodiek nieuwe generaties van toetsversies uitgebracht. De analyses in dit onderzoek betreffen toetsversies van de derde en vierde generatie[2]. Er zijn altijd aparte toetsversies voor vmbo gt, havo en vwo. Verder zijn er bij sommige toetsen aparte versies voor vmbo bb en vmbo kb, maar bij andere toetsen is er één versie voor vmbo bb/kb; dit verschilt tussen toetsen en generaties[3]. Het is aan de scholen zelf om te kiezen welke toetsversies zij voorleggen aan welke leerlingen. In het algemeen is het echter zo dat het schooltype waarop de leerling zich bevindt, leidend is voor de keuze.

Binnen het cvvo zijn er dus vier meetmomenten waarop de vaardigheid van de leerlingen in kaart gebracht wordt. Emons, Frissen & Straat (2021) zijn al uitgebreid op de effecten van het thuisonderwijs op de resultaten van Toets 0 ingegaan. Later rapporteerden Zijlstra, Frissen, Koops, van Onna, & Feskens (2021) over de Toetsen 1, 2 en 3 van het schooljaar 2019/2020. In het huidige onderzoek focussen we opnieuw op de Toetsen 1, 2 en 3, maar nu kijken we naar afnames van het schooljaar 2020/2021[4] en vergelijken de gegevens met die uit de schooljaren vóór corona. De resultaten behaald op de verschillende toetsversies en toetsmomenten zijn direct met elkaar vergelijkbaar, omdat de toetsen *geëquivaleerd* zijn. Dit is gedaan met behulp van een statistisch model afkomstig uit de itemresponstheorie (IRT)[5]. Een IRT model houdt rekening met de moeilijkheid van een toets bij het berekenen van de vaardigheidsscore. Dit betekent dat twee leerlingen die evenveel opgaven goed maken op verschillende toetsen - zoals bijvoorbeeld een havo en een vwo toets - toch verschillende vaardigheidsscores kunnen krijgen. De leerling die de vwo toets heeft gemaakt, zal een hogere vaardigheidsscore krijgen dan de leerling die de havo toets heeft gemaakt. Dit weerspiegelt in dit geval het verschil in moeilijkheid van de toetsversies. De vaardigheidsscores zijn dus over toetsversies en over afnamemomenten direct met elkaar vergelijkbaar. Dit geldt echter alleen *binnen* een onderdeel, en niet *tussen* verschillende onderdelen (zoals leesvaardigheid en woordenschat).

2.2 Afnamegegevens in het cvvo

Figuur 2.1 toont de aantallen leerlingen per schooltype en toets (Toets 1, Toets 2 of Toets 3) en per schooljaar die zijn meegenomen in dit onderzoek. Het schooltype is afgeleid van de gemaakte toetsversie, waarbij vmbo bb en vmbo kb zijn samengevoegd. We gebruiken de afnamegegevens uit vier schooljaren: 2016/2017, 2017/2018, 2018/2019 en 2020/2021. Het betreft per schooljaar in principe de afnames van Toets 1, Toets 2 en Toets 3, maar bij het schooljaar 2016/2017 is Toets 3 van alle onderdelen weggelaten uit het databestand en de analyses, omdat dat als enige nog een generatie 2 cvvo toets was. Verder hebben we geen scores op Wiskunde Toets 1 in 2020/2021, omdat er in dit schooljaar met de vierde generatie toets werd gewerkt, waarin Rekenen en Wiskunde samen één onderdeel vormen. De meeste items in de vierde generatie Toets 1 Rekenen-Wiskunde waren echter Rekenitems,

[2]Bij Toets 3 wordt nog gewerkt met de derde generatie; bij Toets 1 is de vierde generatie in gebruik genomen vanaf schooljaar 2019/2020; en bij Toets 2 vanaf schooljaar 2020/2021.

[3]In de analyses werken we met vier niveaugroepen: vmbo bbkb, vmbo gt, havo en vwo. Afzonderlijke toetsversies vmbo bb en vmbo kb zijn waar nodig dus samengenomen, omwille van de vergelijkbaarheid.

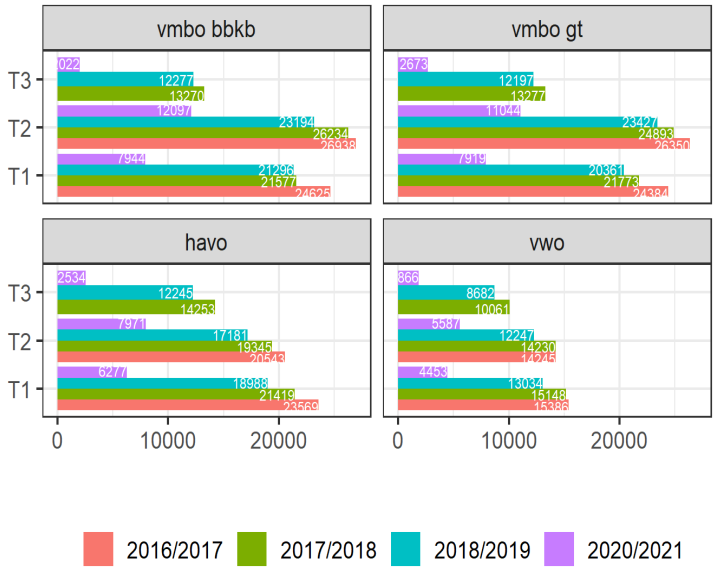
[4]De afnames liepen nog door ten tijde van de datatrekking voor dit onderzoek. We gebruiken hier de afnames die plaatsvonden voor begin juni van 2021.

[5]Meer informatie over itemresponstheorie is onder andere te vinden in Embretson & Reitse (1990) en Hambleton, Swaminathan & Rogers (1991).

dus dat toetsresultaat is alleen vertaald naar Rekenscores en niet naar Wiskundescores.

Overigens moeten we hier vermelden dat alle gegevens van het schooljaar 2020/2021 alleen afnames betreffen van voor begin juni, toen de datatrekking voor dit onderzoek plaatsvond. Dit zal (deels) verklaren waarom het aantal leerlingen in 2020/2021 veel kleiner is. Verder is het goed om te realiseren dat Figuur 2.1 alle leerlingen toont die van tenminste één onderdeel de toets hebben gemaakt. Leerlingen maken niet altijd alle onderdelen, dus in de analyses per vaardigheid gaat het om iets lagere aantallen afnames.

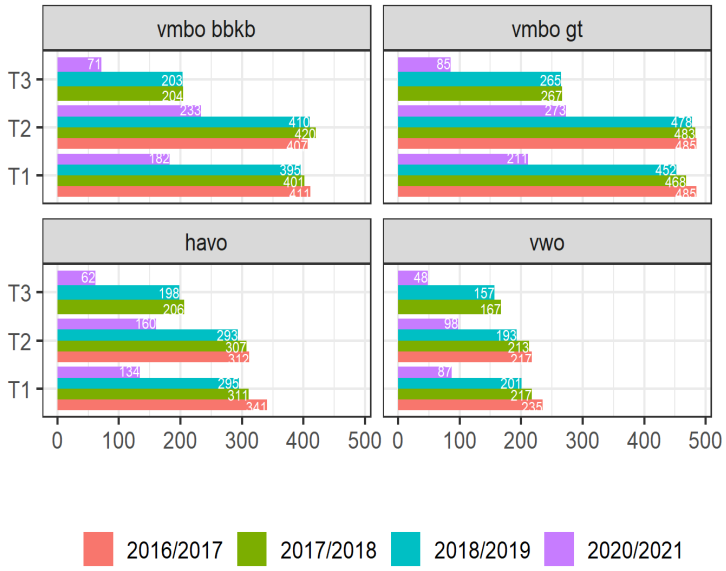
Het schooljaar 2019/2020 is niet meegenomen in dit onderzoek, omdat de coronacrisis en het afstandsonderwijs middenin dat schooljaar begonnen. Bovendien gaat onze interesse nu uit naar de vergelijking tussen het huidige schooljaar en de "voor-corona" schooljaren, omdat we willen inschatten wat het effect is geweest van de twee lockdownperiodes. In het eerdere onderzoek van Zijlstra, Frissen, Koops, van Onna, & Feskens (2021) zijn de afnames uit 2019/2020 wel geanalyseerd, waarbij speciaal rekening is gehouden met de afwijkende omstandigheden in dat schooljaar (zoals vertraagde afnames)[6].



Figuur 2.1: Aantal deelnemende leerlingen

Het aantal scholen in de onderzoeksdata is te vinden in Figuur 2.2. Ook hier valt op dat het aantal scholen in 2020/2021 veel lager is, wat echter niet vreemd is omdat de datatrekking plaatsvond voordat alle scholen de toetsen hadden kunnen afnemen. Het feit dat een kleiner aantal leerlingen en scholen in de data van 2020/2021 is opgenomen zou ertoe kunnen leiden dat deze groep niet vergelijkbaar is met de groepen van de eerdere jaren. Daarom bespreken we in Sectie 3.2 de representativiteit van de onderzoeksgroep.

[6]Dit betreft de eerste deelrapportage van dit onderzoek.



Figuur 2.2: Aantal deelnemende scholen

2.3 Externe data

Bij de afnames van het cvvo worden weinig achtergrondkenmerken van de leerlingen of scholen verzameld. Maar we beschikken wel over het brinvestigingsnummer van de scholen, wat we kunnen gebruiken om de gegevens te koppelen aan openbaar beschikbare gegevens op schoolniveau. Dit stelt ons in staat om de representativiteit van de onderzoeksgroepen te evalueren (zie Sectie 3.2). We maken daarbij gebruik van gegevens over de gemiddelde cijfers op de Centrale Examens, de achterstandsscore, de schoolgrootte, stedelijkheid, provincie, het schooltype en de geografische spreiding. De achterstandsscore is een maat die door het CBS is vastgesteld voor alle scholen in het voortgezet onderwijs[7], en deze zullen we in Sectie 4.2.2 ook gebruiken voor een inhoudelijke analyse van de leerresultaten in subgroepen.

2.4 Referentie- en ERK niveaus

Naast analyses op de toetsscores zelf, kijken we in dit onderzoek ook naar de leerprestaties in termen van de behaalde *referentieniveaus* taal en rekenen[8]. Deze niveaus zijn al gekoppeld aan bepaalde vaardigheidsscores in het cvvo, dus op basis van de toetsscores van de leerlingen kunnen we nagaan welk referentieniveau zij hebben behaald. Hetzelfde geldt voor de *ERK-niveaus* Engels leesvaardigheid[9]: uit de toetsscores op de cvvo toetsen volgt een indicatie voor het beheersen van de ERK-niveaus A2 en B1.

[7] https://www.cbs.nl/nl-nl/maatwerk/2021/26/achterstandsscores-nationaal-programma-onderwijs

[8] `http://www.toetsspecials.nl/html/referentiesets/default.shtm`

[9] `https://www.coe.int/en/web/common-european-framework-reference-languages/level-descriptions`

3. Methoden

In dit hoofdstuk bespreken we de methoden die we hebben gebruikt voor het beantwoorden van de onderzoeksvragen. De twee belangrijkste methodologische uitdagingen in dit onderzoek zijn beide gerelateerd aan de totstandkoming van de beschikbare onderzoeksgegevens. De beschikbare onderzoeksgegevens zijn de afnamegegevens van scholen die zelf kiezen om deel te nemen aan het cvvo. Dit betekent dat de gegevens niet verzameld zijn met behulp van een probabilistische steekproef en we niet zonder meer er van uit kunnen gaan dat de resultaten generaliseerbaar zijn naar alle scholen en leerlingen in Nederland. De beoordeling van de *representativiteit* van de onderzoeksgegevens is dan ook de eerste methodologische uitdaging in dit onderzoek. De tweede methodologische uitdaging is te vinden in de *vergelijkbaarheid* van groepen over jaren. De samenstelling van de groep leerlingen die toetsen uit het cvvo afneemt, is over het algemeen behoorlijk stabiel, maar voor het jaar 2020/2021 is het aantal beschikbare gegevens minder dan eerdere jaren. Dit heeft enerzijds te maken met een lichte daling in het aantal afnames en anderzijds dat op het moment van het schrijven van dit rapport nog niet alle afnamegegevens beschikbaar zijn. Voordat we beschrijven hoe we de representativiteit van de onderzoeksgegevens hebben beoordeeld en hoe we met behulp van *matching* tot zo vergelijkbaar mogelijk (jaar tot jaar) vergelijkingsgroepen zijn gekomen, beschrijven we in in Sectie 3.1 eerst de algemene aanpak van het vergelijken van de verschillende jaargroepen leerlingen in dit onderzoek. In Sectie 3.2 wordt de representativiteit van de onderzoeksgroep geëvalueerd. Sectie 3.3 zet de matchingsprocedure uiteen die we hebben gebruikt om de analyse te kunnen uitvoeren op (meer) vergelijkbaar gemaakte jaargroepen. In Sectie 3.4, ten slotte, beschrijven we hoe we de spreiding in de toetsresultaten in kaart hebben gebracht. Voor alle analyses is gebruik gemaakt van het softwarepakket *R* (R Core Team, 2017).

3.1 Analyse van de trends

Om te bepalen wat het effect is geweest van de schoolsluitingen vergelijken we de toetsprestaties in verschillende schooljaren, namelijk de "voor-corona" jaren 2016/2017, 2017/2018, en 2018/2019, en het "na-corona" jaar 2020/2021. Zoals al vermeld in Sectie 2.2 nemen we de toetsafnames van het schooljaar 2019/2020 niet mee in de analyses. Doordat we van elk van deze schooljaren drie verschillende toetsen analyseren - Toets 1 t/m Toets 3 uit het cvvo - betreffen de analyses verschillende cohorten leerlingen bij iedere afname. De leerlingen die bijvoorbeeld in 2020/2021 Toets 1 hebben gemaakt zijn brugklassers, die dus in hun laatste maanden in het basisonderwijs en een groot deel van hun eerste jaar in het voortgezet onderwijs te maken hebben gekregen met afstandsonderwijs. Deze leerlingen kunnen we ook aanduiden als het cohort 2020/2021 omdat ze in dat schooljaar zijn gestart in het voortgezet onderwijs. De leerlingen die in 2020/2021 Toets 2 of Toets 3 hebben gemaakt zaten al in het tweede of derde jaar van het voortgezet onderwijs, en kunnen we aanduiden als de cohorten 2019/2020 en 2018/2019, respectievelijk. De onderzoeksdata omvat al met al:

- Zes cohorten (2015/2016 tot en met 2020/2021)
- Vier afnamejaren (2016/2017 tot en met 2018/2019, en 2020/2021)

In de analyse kijken we in de eerste plaats naar de gemiddelde prestaties van de afnamegroep per onderdeel en per jaar. We gaan na of de groep in het schooljaar 2020/2021 anders presteert dan de groep in het laatste "voor-corona" schooljaar 2018/2019 deed. Als dat zo is, kan men vermoeden dat dit komt door het effect van de schoolsluitingen. Maar om het verschil in perspectief te plaatsen, kijken we waar mogelijk ook naar de verschillen tussen de schooljaren 2018/2019 en 2016/2017. Dit kan een indicatie geven van een eventuele trend in de gemiddelde toetsprestaties die al eerder aan de gang was, of van normale fluctuaties die kunnen optreden in twee jaar. Zo kunnen we (eventuele) gevonden verschillen beter duiden, in plaats van ze zonder meer toe te schrijven aan de coronacrisis.

In Sectie 4.1 beschrijven we de gemiddelde toetsprestaties van de leerlingen per afnamejaar, onderdeel en toets. We maken gebruik van effectgroottes en percentielverschuivingen om de relevantie van de gemiddelde verschillen tussen jaargroepen te kunnen inschatten. Effectgroottes zijn informatiever dan significantietoetsing, omdat die laatste sterk wordt beïnvloed door het aantal leerlingen in de dataset, zodat bij een grote dataset ook (erg) kleine effecten statistisch significant kunnen worden. We gebruiken *Cohen's D* als maat voor de effectgrootte (Cohen, 1998)[1]. We bepalen de effectgrootte voor het verschil tussen de schooljaren 2020/2021 en 2018/2019 omdat we zo het beste het effect van de schoolsluitingen kunnen evalueren. Ter vergelijking presenteren we waar mogelijk ook de effectgrootte van het verschil tussen de schooljaren 2018/2019 en 2016/2017[2]; ook een

[1] Ook bij Cohen's D als maat voor de effectgrootte past een kanttekening: deze hangt af van de standaarddeviatie in de betreffende populatie(s), en hier moet men rekening mee houden bij het vergelijken van effectgroottes voor verschillende onderzoekspopulaties. In het primair onderwijs is de standaarddeviatie bijvoorbeeld aanzienlijk groter dan binnen de afzonderlijke schooltypen van het voortgezet onderwijs, omdat de schooltypen homogenere groepen leerlingen betreffen.

[2] Voor Toets 3 was dit niet mogelijk omdat we geen data van Toets 3 uit 2016/2017 hebben meegenomen in het onderzoek (zie Sectie 2.2).

periode van twee jaar, maar zonder schoolsluitingen. In het algemeen kunnen we (+-) 0.06 als grenswaarde hanteren voordat we een effect als relevant beschouwen[3]. Bij de duiding van de effecten kan men daarnaast als vergelijkingsmateriaal nog gebruik maken van de effectgroottes tussen aansluitende leerjaren (T1-T2 en T2-T3) of schooltypen (vmbo bbkb-vmbo gt, vmbo gt-havo, en havo-vwo). Deze hebben we berekend voor het schooljaar 2018/2019 en in Tabel 3.1 staan de resultaten beknopt samengevat. In de Appendix staan de tabellen met per onderdeel de effectgrootte voor elk afzonderlijk contrast tussen aansluitende leerjaren of schooltypen in het schooljaar 2018/2019.

	Gemiddelde Cohen's *d* in 2018/2019	
Onderdeel	**tussen leerjaren (van een schooltype)**	**tussen schooltypen (in een leerjaar)**
Nederlands leesvaardigheid	0.39	0.79
Nederlands woordenschat	0.55	0.72
Engels leesvaardigheid	0.44	0.66
Engels woordenschat	0.54	0.45
Rekenen	0.54	0.94
Wiskunde	0.56	0.98

Tabel 3.1: Gemiddelde effectgroottes tussen aansluitende leerjaren (T1-T2 of T2-T3) of schooltypen (bv. vmbo gt-havo) in 2018/2019

Als aanvulling op de effectgrootte berekenen we ook het (afgeronde) percentiel in de scoreverdeling van het schooljaar 2018/2019 dat correspondeert met de gemiddelde score in 2020/2021, en het percentiel dat correspondeert met de gemiddelde score in 2018/2019 zelf. Zo kunnen we vervolgens het gemiddeldenverschil tussen 2020/2021 en 2018/2019 uitdrukken als percentielverschil. In wezen verschilt deze aanpak weinig van het berekenen van de effectgrootte, maar een percentielverschuiving is soms iets makkelijker om te duiden qua praktische relevantie.

In eerste instantie kunnen we kijken naar de totale groepen per jaar, zonder rekening te houden met mogelijke verschillen in de groepssamenstelling van de leerlingen die de toetsen in een bepaald schooljaar maakten. Maar voor het berekenen van de effectgroottes en percentielverschillen gebruiken we gematchte groepen (zie Sectie 3.3). In de trendfiguren in de resultatensectie worden de totale en de gematchte groepen steeds beide afgebeeld.

Naast de globale trend op de verschillende vaardigheden is het ook interessant om na te gaan of de trend voor verschillende subpopulaties van leerlingen anders is. We beschikken in de cvvo data over weinig leerlingkenmerken, maar in Sectie 4.2 kijken we wel naar de trends voor subgroepen op basis van twee variabelen: geslacht, en de achterstandsscore van de school (zie Sectie 2.3), die we in vier categorieën indelen (van laagste naar hoogste).

[3]Dit komt overeen met 1% van de totale vaardigheidsschaal.

3.2 Representativiteit

Niet alle leerlingen en niet alle scholen in Nederland maken gebruik van het cvvo. Daarnaast kunnen scholen zelf kiezen of zij gebruik willen maken van het cvvo. Om inzicht te krijgen in hoeverre het ontbreken van deze gegevens en mogelijke zelfselectie-effecten de resultaten en conclusies vertekenen, proberen we de representativiteit van de afnamegegevens in kaart te brengen. Representativiteit heeft betrekking op de vraag in welke mate de verdeling van achtergrondkenmerken in de populatie overeenkomt met de verdeling van deze kenmerken in de responsgroep. Dit zullen we voor scholen eerst beoordelen aan de hand van een visuele vergelijking van de verdeling in de populatie en steekproef van een aantal relevante achtergrondkenmerken (Paragraaf 3.2.1). Vervolgens zullen we de representativiteit van de onderzoeksgegevens ook kwantificeren met behulp van de zogenaamde *R-indicator* (Schouten, Cobben & Bethlehem, 2009) in Paragraaf 3.2.2. We beoordelen de representativiteit van de steekproef op basis van de volgende kenmerken:

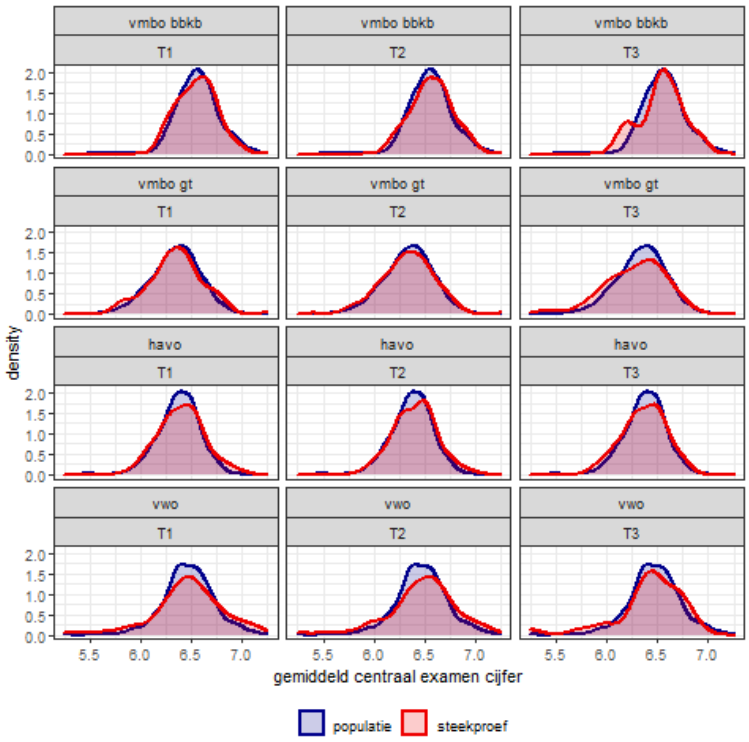
- Gemiddeld Centraal Examencijfers
- Achterstandsscores
- Stedelijkheid
- Schoolgrootte
- Provincie
- Schooltype
- Geografische spreiding

In dit onderzoek maken we niet gebruik van één steekproef, maar van meerdere vergelijkingsgroepen. De belangrijkste vergelijkingen vinden plaats tussen afnamejaren en binnen afnamejaren tussen leerjaren (Toets 1, Toets 2 en Toets 3) en niveaus (vmbo bbkb, vmbo gt, havo, vwo). Het aantal afnames zijn in 2020/2021 het laagst (zie Figuur 2.1) en de afnamegegevens van 2020/2021 zijn gematched aan de andere afnamejaren (zie Sectie 3.3), vandaar dat we de bespreking van de representativiteitsanalyses beperken tot dit afnamejaar.

3.2.1 Visuele evaluatie van representativiteit

Centraal Examen cijfers

In Figuur 3.1 is de verdeling van de gemiddelde Centrale Examencijfers te vinden. Deze zijn uitgesplitst naar niveau. De blauwe lijn geeft de verdeling van de gemiddelde school Centrale Examencijfers weer voor alle scholen in Nederland (de populatie). De rode lijn geeft de verdeling van deze cijfers weer voor de scholen die in dit onderzoek te vinden zijn. Juist de verdeling van de gemiddelde cijfers op de Centrale Examens is zeer indicatief voor de beoordeling van de representativiteit van de onderzoeksgegevens, omdat we kunnen verwachten dat er een sterke relatie bestaat tussen de eindexamencijfers en de gemiddelde schoolprestaties op de toetsen in de onderbouw van het voortgezet onderwijs. Een sterke overlap tussen de verdeling van de gemiddelde Centraal Examencijfers in de populatie en steekproef zou met andere woorden aangeven dat in de cvvo afnamegegevens er geen systematische onder- of oververtegenwoordiging is van zwakke of sterk presterende scholen.



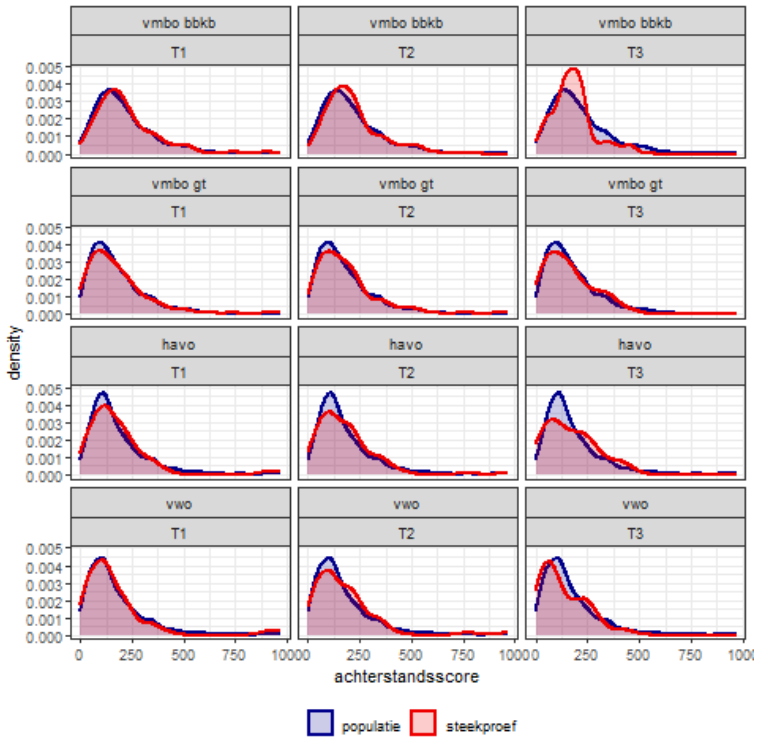
Figuur 3.1: Centraal Examen resultaten in de populatie en de responsgroep

De scholen in de responsgroep zijn wat betreft gemiddeld Centraal Examencijfer niet helemaal, maar toch behoorlijk sterk vergelijkbaar met scholen in de gehele populatie, zoals in Figuur 3.1 te zien is. Het verschilt echter wel per opleidingsniveau of de scholen in de responsgroep (minimaal) lagere of juist hogere gemiddelde eindexamenresultaten laten zien dan alle scholen in de populatie. De schooltypes vmbo gt en havo in de responsgroep behalen hogere gemiddelde Centraal Examencijfers in vergelijking met alle scholen in de populatie. Voor de niveaus vmbo bbkb en vwo is het beeld omgekeerd. Daar scoren de scholen in de populatie net iets hoger op de Centrale Examens in vergelijking met de scholen in de responsgroep. Desondanks kan op basis van de vrijwel overlappende verdelingen geconcludeerd worden dat de onderzoeksgegevens met betrekking tot de gemiddelde Centraal Examencijfers zeer representatief zijn.

Achterstandsscores

Achterstandsscores zijn een geaggregeerde school maat voor de verwachte schoolprestatie van de leerlingen op de school gezien de omgevingskenmerken en ongeacht de intelligentie van de leerlingen[4]. Een hogere achterstandsscore betekent dat er meer leerlingen op de school zitten die gegeven hun achterstand in aanmerking komen voor additionele financiering. In deze analyses vergelijken we de achterstandsscores zonder drempel van alle scholen in de populatie en de scholen in de steekproef.

[4] https://www.cbs.nl/nl-nl/maatwerk/2021/26/achterstandsscores-nationaal-programma-onderwijs



Figuur 3.2: Achterstandsscore in de populatie en de responsgroep

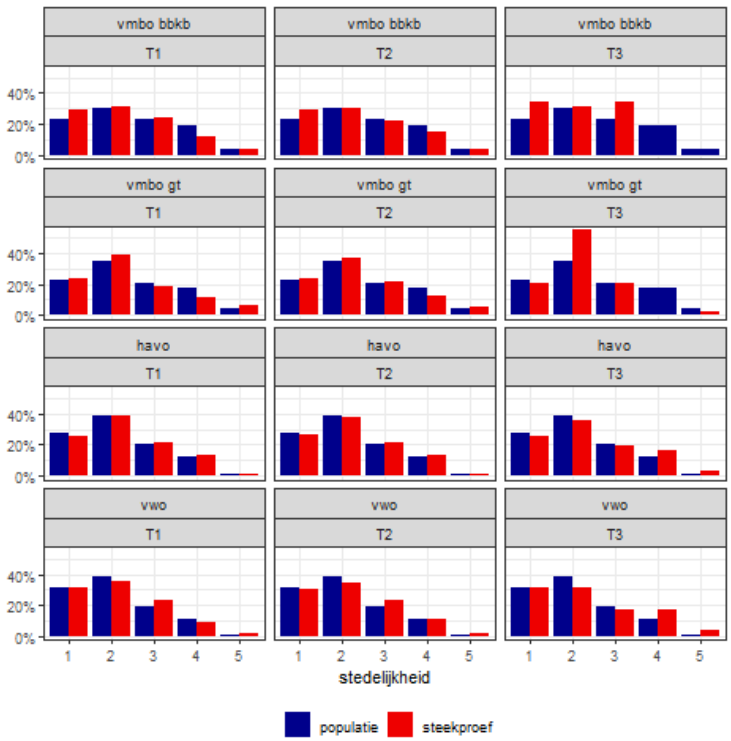
Figuur 3.2 laat zien dat de verdelingen van de populatie en de verschillende vergelijkingsgroepen behoorlijk overlappend zijn. Dat geldt zeker voor de schoolgegevens van Toets 1 en Toets 2. Voor Toets 3 wijken de verdelingen tussen de cvvo gegevens en de populatie iets meer af, maar ook hier geldt dat er geen systematische onder- of oververtegenwoording te zien is van scholen die een hogere of juist lagere achterstandsscore hebben.

Stedelijkheid

Stedelijkheid is een derde kenmerk dat we beoordelen op representativiteit. Op grond van de omgevingsadressendichtheid[5] is aan iedere buurt, wijk of gemeente een stedelijkheidsklasse toegekend (`https://www.cbs.nl/nl-nl/cijfers/detail/84583NED?q=stedelijkheid#MateVanStedelijkheid_115`). De volgende klassenindeling is gehanteerd:

1. zeer sterk stedelijk >= 2500 adressen per km^2
2. sterk stedelijk 1500 - 2500 adressen per km^2
3. matig stedelijk 1000 - 1500 adressen per km^2
4. weinig stedelijk 500 - 1000 adressen per km^2
5. niet stedelijk < 500 adressen per km^2

De gegevens zijn afkomstig van CBS Statline (`https://www.cbs.nl/nl-nl/dossier/nederland-regionaal/geografische-data/gegevens-per-postcode`). Figuur 3.3 geeft het aandeel scholen weer in elk van deze vijf categorieën in de populatie (in blauw) en in de steekproef (in rood). De verdelingen in de steekproeven komen redelijk overeen met die in de populatie. Dat geldt zeker voor Toets 1 en Toets 2. De afwijkingen zijn wat groter bij de Toets 3 gegevens. Hier is bijvoorbeeld het aandeel cvvo scholen in categorie 2 (sterk stedelijk) in de vmbo gt groep een stuk groter dan in de populatie.

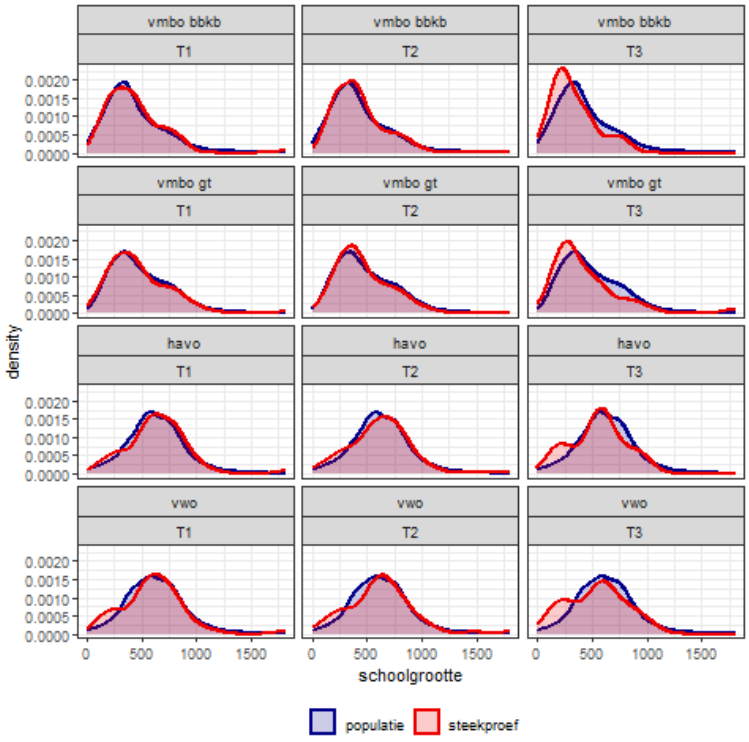


Figuur 3.3: Stedelijkheid in de populatie en de responsgroep

[5]De omgevingsadressendichtheid van een buurt, wijk of gemeente is het gemiddeld aantal adressen per vierkante kilometer binnen een cirkel met een straal van één kilometer

Schoolgrootte

De schoolgrootte van een school is bepaald door het aantal leerlingen in het derde leerjaar op een school te sommeren. Deze gegevens zijn afkomstig van DUO[6]. In Figuur 3.4 evalueren we of er in de cvvo geen onder- of juist oververtegenwoordiging is van bijvoorbeeld kleine of grote scholen.



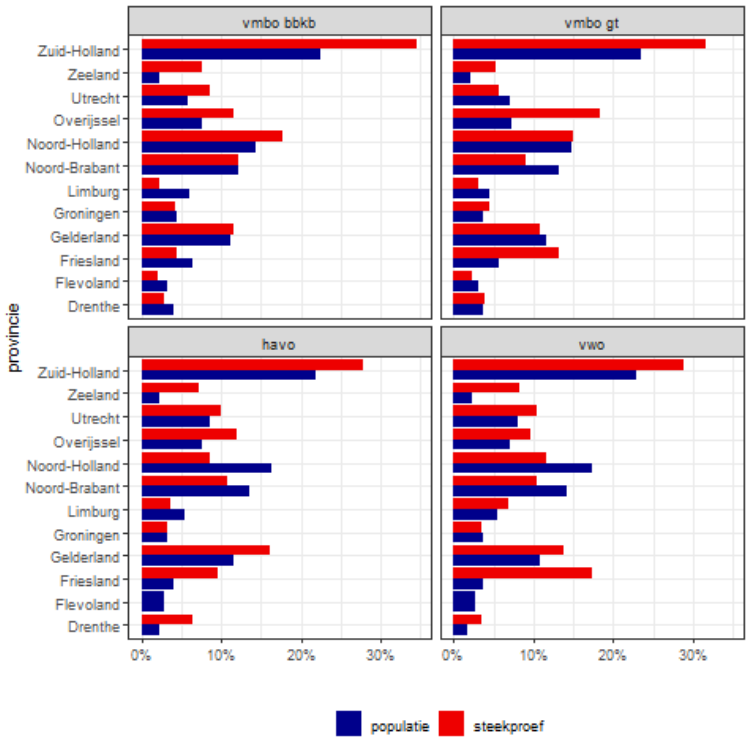
Figuur 3.4: Schoolgrootte in de populatie en de responsgroep

In Figuur 3.4 is te zien dat cvvo met betrekking tot schoolgrootte behoorlijk representatief te noemen is. De gegevens die beschikbaar zijn voor Toets 3 geven enige aanleiding om te veronderstellen dat kleinere scholen licht oververtegenwoordigd zijn in de cvvo gegevens, maar al met al zijn de gegevens ook hier wel opvallend representatief te noemen. Een laatste punt dat opvalt in dit figuur is dat de gemiddelde schoolgrootte van vmbo scholen een stuk kleiner is dan deze van havo en vwo scholen. Dit is niet relevant voor de evaluatie van de representativiteit, maar kan wel behulpzaam bij de latere evaluatie van eventuele verschillen tussen corona effecten tussen schooltypen in Sectie 4.3.

Provincie

De representativiteit met betrekking tot provincie is vanwege de overzichtelijkheid alleen per niveau in Figuur 3.5 afgebeeld. De meeste scholen zijn zowel in de populatie als in de cvvo gegevens te vinden in Zuid-Holland. Een aantal provincies zijn oververtegenwoordigd (zoals Gelderland, Zeeland en Zuid-Holland). Omgekeerd zijn scholen in Flevoland, Noord-Brabant en Noord-Holland enigszins ondervertegenwoordigd.

[6] `https://duo.nl/open_onderwijsdata/databestanden/vo/leerlingen/leerlingen-vo-1.jsp`

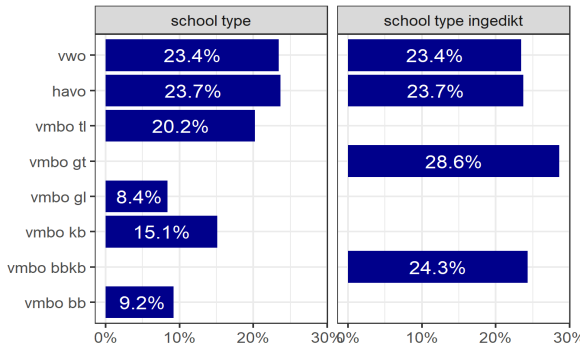


Figuur 3.5: Stedelijkheid in de populatie en de responsgroep

Schooltype

In Nederland waren er in het schooljaar 2019/2020 648 hoofdvestigingen in het voortgezet onderwijs[7]. Het aantal scholen met een eigen *brinvestiging-identificatie* was 1615.

In Figuur 3.6 is het percentage leerlingen per schooltype in het derde leerjaar gerapporteerd[8]. In het linker paneel van deze figuur zijn de percentages leerlingen per schooltype te vinden die gangbaar gebruikt worden. In het rechter paneel zijn schooltypes gecombineerd tot de indeling van schooltypes die we in dit onderzoek gebruiken.



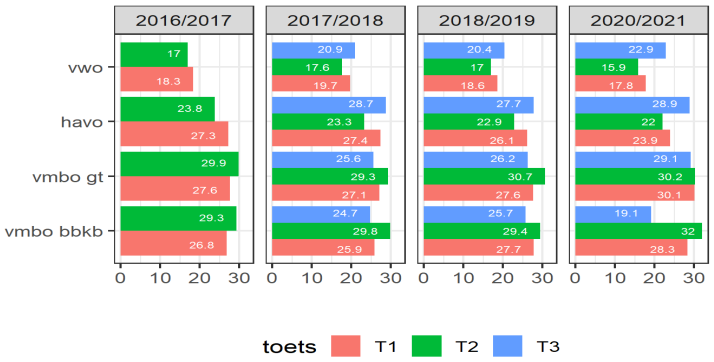
Figuur 3.6: Aantal leerlingen per schooltype in de populatie

[7]Dit is vastgesteld op basis van *brin-identificatie*, zie ook `https://duo.nl/open_onderwijsdata/databestanden/vo/adressen/adressen-vo-2.jsp`.

[8]Deze gegevens zijn afkomstig van `https://opendata.cbs.nl/statline/#/CBS/nl/dataset/80040ned/table?dl=50918`.

In Figuur 3.6 is te zien dat iets meer dan de helft van de leerlingen in het derde leerjaar in één van de leerwegen van het vmbo te vinden is. De leerwegen vmbo bb en vmbo kb - die we in dit onderzoek samen hebben gevoegd - omvat ruim 24% van de leerlingen. De meeste leerlingen (ruim 26%) zitten op het vmbo gt. Ongeveer 23% van de leerlingen is te vinden op respectievelijk de havo en het vwo.

In Figuur 3.7 staan de percentages leerlingen per afnamejaar per toets gerapporteerd.

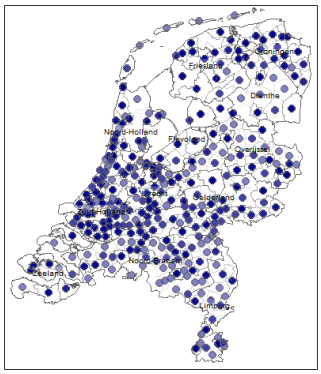


Figuur 3.7: Aantal leerlingen per schooltype in de responsgroep

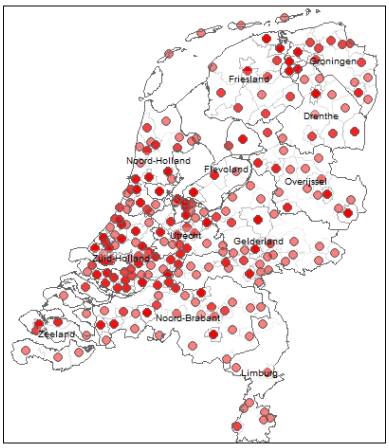
Wanneer we de verdeling van leerlingen over schooltypes van de populatie (zie Figuur 3.6) vergelijken met de gegevens die we in dit onderzoek tot onze beschikking hebben (Figuur 3.7), dan valt op dat er een behoorlijke ondervertegenwoordiging is van het aantal vwo leerlingen. Deze ondervertegenwoordiging van het aantal vwo leerlingen is in de afgelopen drie jaar ook nog groter geworden. Ruim 23% van de leerlingen zit in het derde leerjaar op het vwo in de gehele populatie. Het aandeel leerlingen dat de *toetsversie* vwo heeft gemaakt, varieert van 13.9% (Toets 2 in 2019/2020*) tot 19.8% (Toets 3 in 2017/2018). De ondervertegenwoordiging van het aantal vwo leerlingen leidt uiteraard tot een oververtegenwoordiging van het aandeel vmbo en havo leerlingen. We zullen dus voorzichtig moeten zijn met uitspraken over de ontwikkeling van een geheel cohort. Zonder rekening te houden met de afwijkende en veranderende samenstelling van de groep leerlingen waar we toetsgegevens over beschikken, kunnen snel verkeerde conclusies getrokken worden.

Geografische spreiding

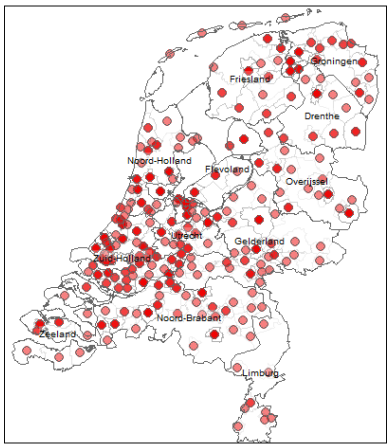
In Figuur 3.8 zijn alle voortgezet onderwijs scholen in Nederland afgebeeld. De scholen in de responsgroep zijn te vinden in Figuur 3.9. Er is geen systematisch effect zichtbaar dat in bepaalde regio's scholen minder in de responsgroep vertegenwoordigend zijn. We concluderen dan ook dat de representativiteit met betrekking tot geografische spreiding in orde is.



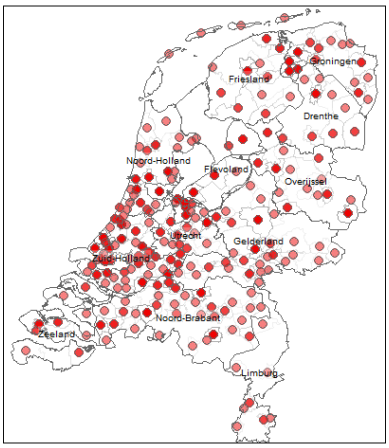
Figuur 3.8: Voortgezet onderwijsscholen in Nederland



(a) 2016/2017



(b) 2017/2018



(c) 2018/2019

(d) 2020/2021

Figuur 3.9: Voortgezet onderwijsscholen in de respons groep

3.2.2 R-indicator

De *R-indicator* of “Representativity Indicator” is een maat voor de representativiteit van enquête-onderzoek (Schouten, Cobben & Bethlehem, 2009). De R-indicator geeft een kwantificering van de mate van gelijkenis tussen de responsgroep en de populatie. Om

de R-indicator te kunnen berekenen is dan ook informatie nodig voor zowel de in dit geval de deelnemende scholen als de de niet-deelnemende scholen. De maat is ontwikkeld als alternatief voor het responscijfer om de kwaliteit van een steekproefonderzoek te beoordelen. Waarbij het responscijfer het aandeel respondenten ten opzichte van de gehele steekproef uitdrukt, probeert de R-indicator de potentiële *vertekening* door non-respons in kaart te brengen. Vertekening in de onderzoeksresultaten ontstaat zodra er conclusie voor de gehele populatie worden getrokken op basis van de beschikbare onderzoeksgegevens en de niet-beschikbare gegevens systematisch afwijken met betrekking tot de uitkomsten van het onderzoek. Binnen dit onderzoek is in strikte zin geen sprake van een responscijfer, omdat er niet gewerkt wordt vanuit een steekproef[9]. We kunnen echter wel de R-indicator gebruiken om te inventariseren of de beschikbare onderzoeksgegevens niet systematisch afwijken van de populatiekenmerken. De R-indicator neemt waarden aan tussen 0 en 1, waarbij een waarde van 1 betekent dat de steekproef representatief is ten opzichte van variabelen die zijn opgenomen in het model. In de praktijk zal een steekproef niet of nauwelijks geheel representatief zijn ten opzichte van alle beschikbare achtergrondvariabelen.

Voor het berekenen van de R-indicator hebben we gebruikt gemaakt van code zoals deze door het RISQ project ontwikkeld zijn[10]. De variabelen die we opgenomen hebben in het model zijn de gemiddelde Centraal examencijfers, achterstandsscores, schoolgrootte en stedelijkheid. Om de interpretatie van de uitkomsten te vergemakkelijken hebben we alle variabelen opgedeeld in kwartielen. In Tabel 3.2 zijn de R-indicatoren voor de twaalf steekproeven te vinden.

Schooltype	**Toets**	**R-indicator**
vmbo bbkb	Toets 1	0.889
vmbo bbkb	Toets 2	0.823
vmbo bbkb	Toets 3	0.900
vmbo gt	Toets 1	0.928
vmbo gt	Toets 2	0.894
vmbo gt	Toets 3	0.915
havo	Toets 1	0.930
havo	Toets 2	0.917
havo	Toets 3	0.941
vwo	Toets 1	0.939
vwo	Toets 2	0.929
vwo	Toets 3	0.932

Tabel 3.2: R-Indicator waarden

Zoals in Tabel 3.2 te zien is, zijn de waarden van de R-Indicatoren zijn voor alle toetsen vrij hoog. Dat betekent dat voor alle steekproeven geldt dat deze behoorlijk vergelijkbaar zijn met de gehele populatie *op basis van vier opgenomen kenmerken in het model*.

[9] Dit onderzoek zou dan ook beschouwd kunnen worden als een *convenience sample*: Scholen kunnen zelf kiezen of ze al dan niet willen deelnemen aan dit onderzoek of niet.

[10] `https://www.cmi.manchester.ac.uk/research/projects/representative-indicators-for-survey-quality/`

3.3 Het maken van vergelijkbare groepen

We kunnen niet zomaar veronderstellen dat de groepen leerlingen die de toetsen hebben gemaakt in verschillende schooljaren allemaal vergelijkbaar zijn. Er kan bijvoorbeeld sprake zijn van selectie-effecten op schoolniveau omdat niet elk jaar dezelfde scholen de toets hebben afgenomen. Het is goed denkbaar dat het afstandsonderwijs in het schooljaar 2020/2021 en de onzekerheid over hoe lang dit zou gaan duren invloed heeft gehad op het besluit om de cvvo toetsen wel of niet af te nemen. Daarnaast is het mogelijk dat er op leerlingniveau selectie-effecten zijn, bijvoorbeeld wanneer een school ervoor kiest om de toetsen vooral af te nemen bij de leerlingen van wie niet duidelijk is of ze het lesmateriaal voldoende beheersen. Deze mogelijke selectiviteit in de samenstelling van de groepen kan de vergelijking tussen de verschillende schooljaren vertekenen en maakt het dus moeilijker om geobserveerde verschillen in gemiddelden toe te schrijven aan het effect van thuisonderwijs. Om deze zogeheten populatie-effecten zo veel mogelijk uit de vergelijking te filteren, maken we gebruik van *matching* op een vergelijkbare manier als Zijlstra, Frissen, Koops, van Onna & Feskens (2021) hebben gedaan. Matching is een statistische procedure, bedoeld om het effect van een gebeurtenis (hier de schoolsluitingen) zuiverder te kunnen schatten door rekening te houden met achtergrondvariabelen die van invloed kunnen zijn op de meting (Rosenbaum & Rubin, 1983).

In een matchingsprocedure is er sprake van een zogenaamde *targetgroep* en een *donorgroep*. In dit geval bestaat de targetgroep uit de leerlingen die in het schooljaar 2020/2021 de toetsen hebben gemaakt. De donorgroep bestaat uit de leerlingen uit de eerdere afnames (dus de eerdere schooljaren). In de procedure die wij hebben toegepast, wordt elke leerling uit de targetgroep gekoppeld aan een leerling uit de donorgroep met dezelfde kenmerken op achtergrondvariabelen. Zo ontstaan twee groepen met exact dezelfde omvang en zo vergelijkbaar mogelijke achtergrondkenmerken.

De variabelen waarop we hebben gematcht zijn leeftijd, niveau, geslacht, en de afnamemaand van de toets. Binnen deze stratificatie wordt elke leerling uit de targetgroep gekoppeld aan een leerling uit de donorgroep met een zo gelijk mogelijke vaardigheidsscore op een eerdere toets. Indien het aantal leerlingen met dezelfde stratificatie in de donorgroep groter is dan in de targetgroep, blijft er uiteraard een aantal leerlingen over in de donorgroep die vervolgens afvallen. Uiteindelijk hebben we dus de groepen leerlingen op bovengenoemde kenmerken gematched zodat we hun (gemiddelde) toetsprestaties zo zuiver mogelijk kunnen vergelijken.

3.4 Spreiding in de toetsresultaten

Naast de gemiddelde toetsscores kunnen ook spreidingsmaten waardevolle informatie geven over de effecten van de schoolsluitingen op de leeruitkomsten. Met behulp van multilevel analyses kunnen we nagaan of de verschillen tussen scholen in de gemiddelde leerlingprestaties groter zijn geworden (mogelijk) door het afstandsonderwijs. In een multilevel analyse is het namelijk mogelijk de spreiding in de toetsresultaten uit te splitsen naar de spreiding op groepsniveau en de spreiding op individueel niveau (Hox, Moerbeek & Van der Schoot, 2018). In ons geval betekent dit dat we de variantiecomponent op het schoolniveau onderscheiden van de variantiecomponent op het leerlingniveau. Op het moment dat de variantie op schoolniveau toeneemt, nemen de verschillen tussen

scholen toe. Dit kan een indicatie zijn dat scholen verschillend goed konden omgaan met het aanbieden van afstandsonderwijs. Op het moment dat de variantie binnen scholen toeneemt, nemen de verschillen tussen leerlingen binnen scholen toe. Dat zou kunnen aanduiden dat de periode van thuisonderwijs vooral verschillend heeft uitgepakt voor individuele leerlingen, en niet zozeer voor hele scholen. Om de vergelijkingen over de jaren zo valide als mogelijk te maken, zijn in deze analyse alleen die scholen opgenomen die in de betreffende schooljaren altijd de cvvo toetsen hebben afgenomen. Scholen waarbij in 2020/2021 minder dan 75% heeft deelgenomen van het aantal leerlingen dat normalerwijze de toetsen maakte, zijn ook uitgesloten van de analyses. De analyses zijn uitgevoerd met behulp van het R-pakket *lme4* (Bates, Mächler, Bolker & Walke, 2015).

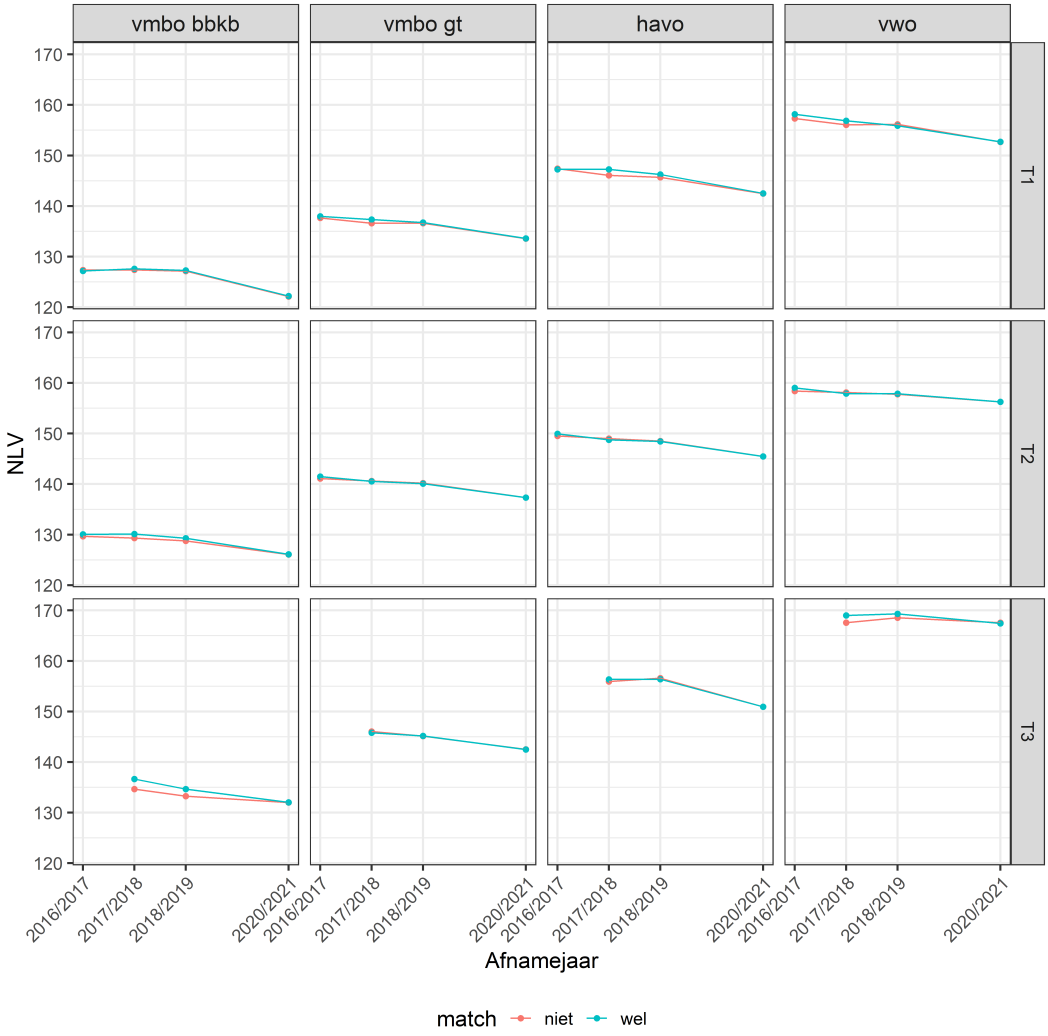
4. Resultaten

In dit hoofdstuk presenteren we de resultaten per toetsonderdeel. In Sectie 4.1 beschrijven we de trend in de gemiddelde vaardigheid over de afgelopen jaren aan de hand van figuren, effectgroottes en percentielverschillen. In de figuren vergelijken we steeds de resultaten op basis van *alle* leerlinggegevens waarover we beschikken met de resultaten op basis van gematchte groepen (zie Sectie 3.3). Bij de onderdelen Nederlands leesvaardigheid en Rekenen gaan we ook na welk percentage leerlingen een bepaald referentieniveau heeft behaald, en bij Engels leesvaardigheid kijken we naar de behaalde ERK-niveaus. Vervolgens gaan we in Sectie 4.2 na of er verschillen in de trends zijn tussen subgroepen op basis van geslacht of de achterstandsscore van de school. In Sectie 4.3 ten slotte bespreken we de spreiding in de toetsscores op basis van de multilevel analyses (zie Sectie 3.4).

4.1 Trendresultaten

4.1.1 Nederlands leesvaardigheid

In Figuur 4.1 zijn de gemiddelde vaardigheidsscores op Nederlands leesvaardigheid per afnamejaar weergegeven voor de totale (oranje lijn) en de gematchte (turquoise lijn) leerlinggroepen. De twee lijnen in de Figuur zijn vaak nauwelijks te onderscheiden, wat illustreert dat de matching op achtergrondkenmerken meestal weinig effect had op de resultaten. Dit kan zijn omdat de totale groepen al behoorlijk vergelijkbaar waren qua achtergrondkenmerken, of omdat deze kenmerken maar beperkt verband hielden met de toetsprestaties. Wat opvalt in Figuur 4.1 is dat de gemiddelde vaardigheidsscores voor Nederlands leesvaardigheid in 2020/2021 meestal lager waren. In de drie jaren van 2016/2017 tot en met 2018/2019 zagen ze er wat stabieler uit, hoewel er bij meerdere van de schooltypen en toetsen (Toets 1 t/m Toets 3) al eerder een licht dalende trend leek te zijn.



Figuur 4.1: Gemiddelde vaardigheidsscores Nederlands leesvaardigheid

In Tabel 4.1 zijn de effectgroottes (uitgedrukt in Cohen's D) gepresenteerd van het verschil in de gemiddelde vaardigheid tussen de schooljaren 2020/2021 en 2018/2019. Ter vergelijking is (waar mogelijk) ook de effectgrootte gegeven van het verschil tussen de schooljaren 2018-2019 en 2016-2017[1]. Deze effectgroottes zijn gebaseerd op de gematchte groepen en betreffen dus de trends die in Figuur 4.1 zijn afgebeeld met de turquoise lijnen.

De effectgroottes in Tabel 4.1 bevestigen de indruk dat de gemiddelde Nederlandse leesvaardigheid in het schooljaar 2020/2021 flink gedaald is ten opzichte van twee jaar eerder, en dat deze daling bij de meeste schooltypen en leerjaren groter is dan in de twee jaar daarvoor. De daling is dus niet zomaar te duiden als vervolg van een trend die al begonnen was. Het lijkt meer op een nieuwe ontwikkeling en daarom ligt het voor de hand om te denken dat de coronacrisis er een rol in heeft gespeeld. Alleen bij het vwo is de daling wat minder opvallend ten opzichte van de daling in de twee jaar ervoor. Bij de andere schooltypen is het effect daarentegen vrij uitgesproken. De effectgroottes zijn vaak -0.20 of

[1]Bij Toets 3 ontbreekt de tweede effectgrootte omdat de afnames van Toets 3 uit het schooljaar 2016/2017 niet zijn meegenomen in dit onderzoek; zie Sectie 2.2.

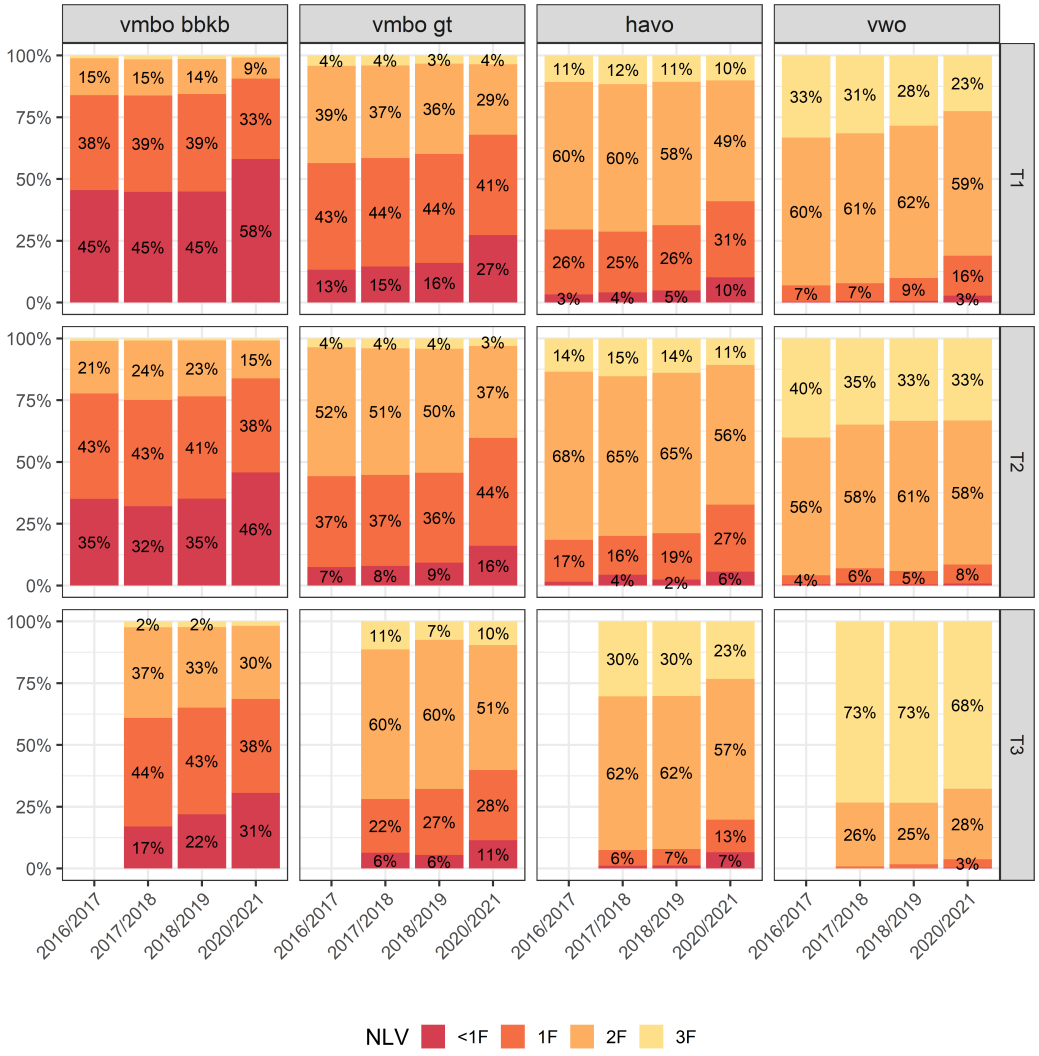
hoger, en dat is ook fors vergeleken bij de effectgrootte tussen opeenvolgende leerjaren uit 2018/2019. Gemiddeld was die laatste voor Nederlands leesvaardigheid namelijk 0.39 (zie Tabel 3.1), en specifiek tussen leerjaar (Toets) 1 en 2 was de effectgrootte van het verschil gemiddeld maar 0.19 (zie Tabel A.1 in de Appendix).

NLV		**Cohen's *d***		**Percentielverschil**
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 1	-0.36	0.01	-18
vmbo gt	Toets 1	-0.22	-0.09	-13
havo	Toets 1	-0.26	-0.07	-10
vwo	Toets 1	-0.23	-0.17	-12
vmbo bbkb	Toets 2	-0.24	-0.06	-12
vmbo gt	Toets 2	-0.22	-0.11	-11
havo	Toets 2	-0.23	-0.13	-10
vwo	Toets 2	-0.13	-0.10	-8
vmbo bbkb	Toets 3	-0.18		- 9
vmbo gt	Toets 3	-0.19		-7
havo	Toets 3	-0.38		-19
vwo	Toets 3	-0.13		-7

Tabel 4.1: Effecten op Nederlands leesvaardigheid (o.b.v. gematchte groepen)

In de laatste kolom van Tabel 4.1 zijn de effecten uitgedrukt als percentielverschil. Hiertoe zijn de gemiddelde vaardigheidsscores van de schooljaren 2020/2021 en 2018/2019 beide uitgedrukt als percentiel binnen de scoreverdeling van 2018/2019, en daarna is het verschil berekend. Over het algemeen heeft het gemiddelde van 2018/2019 een percentiel van rond 50 (de mediaan); het is vaak niet exact 50 doordat de verdelingen enigszins scheef zijn. De interpretatie van de verschuiving in het percentiel is als volgt. Bij vmbo bbkb Toets 1 is de gemiddelde score uit 2020/2021 18 percentielpunten lager dan de gemiddelde score uit 2018/2019. Een leerling die in het schooljaar 2020/2021 perfect gemiddeld scoort, zou in het leerjaar 2018/2019 dus 18% lager in de rangordening vallen dan een leerling die *toen* perfect gemiddeld scoorde. Of andersom gezegd: het gemiddelde vaardigheidsniveau in 2020/2021 is zodanig lager, dat een extra 18% van de leerlingen uit 2018/2019 dit niveau overtrof. Als we de scheefheid in de verdelingen even negeren en, ter illustratie, aannemen dat in beide leerjaren het gemiddelde exact samenviel met het 50e percentiel van de verdeling, dan zou volgen dat wat in 2020/2021 percentiel 50 heet, percentiel (50-18=) 32 zou zijn geweest in de verdeling van 2018/2019; en dat dus (50+18=) 68% van de leerlingen van toen boven het gemiddelde van 2020/2021 scoorde.

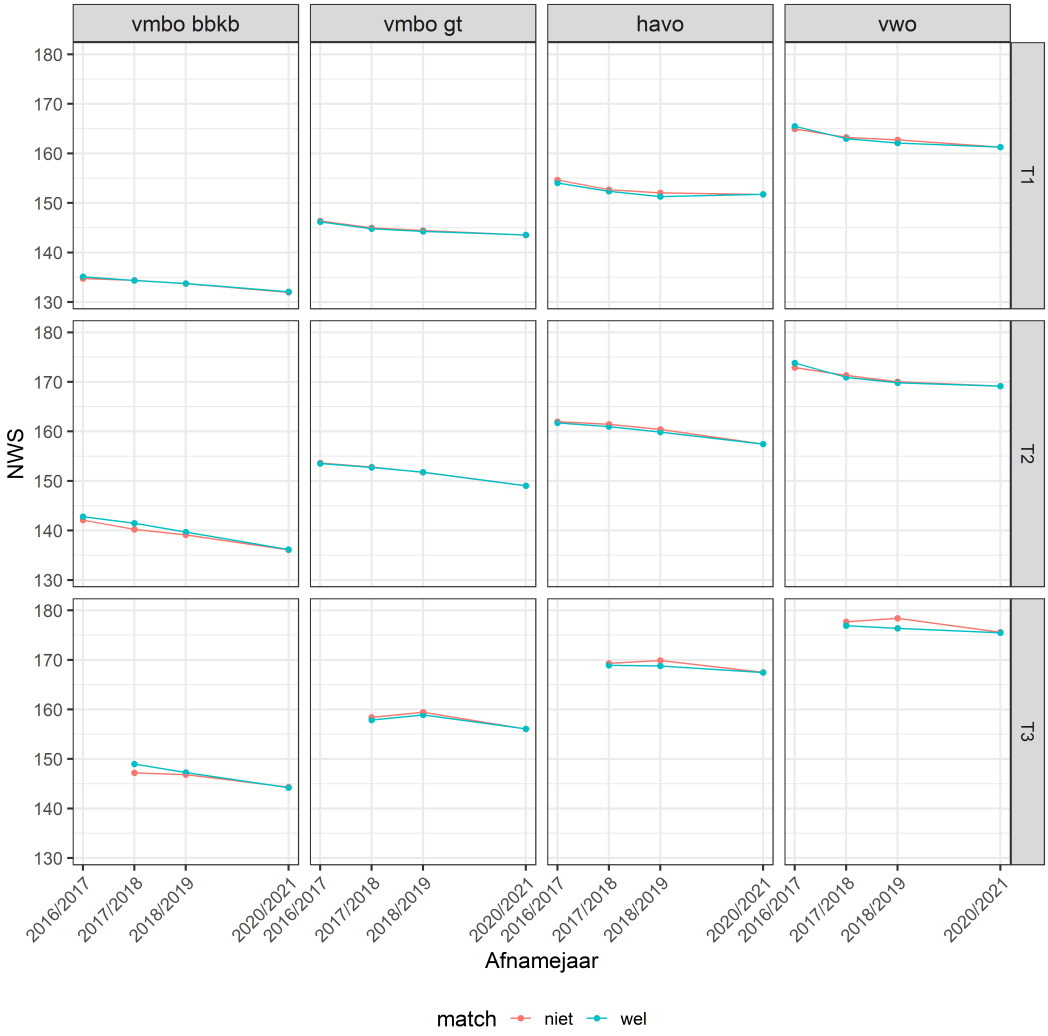
In Figuur 4.2 is te zien welke percentages leerlingen op Nederlands leesvaardigheid de referentieniveaus 1F, 2F en 3F behaalden (in schooljaar 2020/2021 en in de gematchte groepen voor de eerdere schooljaren). Hier valt op dat er in 2020/2021 in alle drie de leerjaren aanzienlijk meer vmbo leerlingen zijn die onder het 1F niveau presteren, en aanzienlijk minder die op 2F niveau presteren, dan in de eerdere schooljaren. Ook bij de havo leerlingen zien we dat de verdeling minder gunstig is geworden in 2020/2021: de percentages voor <1F en 1F zijn hoger, en die voor 2F en 3F lager. Bij het vwo zijn de verschillen in het tweede en derde leerjaar klein, maar in het eerste leerjaar vwo is er wel een opvallende stijging (met samen zo'n 10%) van de categorieën <1F en 1F.



Figuur 4.2: Verdeling over referentieniveaus Nederlands leesvaardigheid

4.1.2 Nederlands woordenschat

Figuur 4.3 toont de gemiddelde vaardigheidsscores op Nederlands woordenschat, en de effectgroottes staan in Tabel 4.2. Ook hier zien we in het vmbo een aantal keer dalingen in het niveau in 2020/2021 vergeleken met dat van 2018/2019. Het meest uitgesproken is deze daling te zien in leerjaar 3 voor zowel vmbo bbkb als vmbo gt. Daar zijn de effectgroottes (beide) -0.21, en de percentielverschillen -9 en -13 respectievelijk. Dit is best groot vergeleken bij de gemiddelde effectgrootte van leerjaar (0.55; zie Tabel 3.1 en A.3 in de Appendix). Soms was er in de twee jaar daarvoor al een vergelijkbare daling aan de gang (zoals bij vmbo bbkb Toets 2). Bij de havo zien we alleen een daling in Toets 2 en deze wijkt niet af van de dalende trend in de twee jaar ervoor. Bij het vwo tenslotte lijken de gemiddelde vaardigheidsscores in 2020/2021 nauwelijks veranderd te zijn ten opzichte van 2018/2019. Al met al lijken de resultaten op Nederlands woordenschat dus wat stabieler te zijn over de jaren dan de resultaten op Nederlands leesvaardigheid.



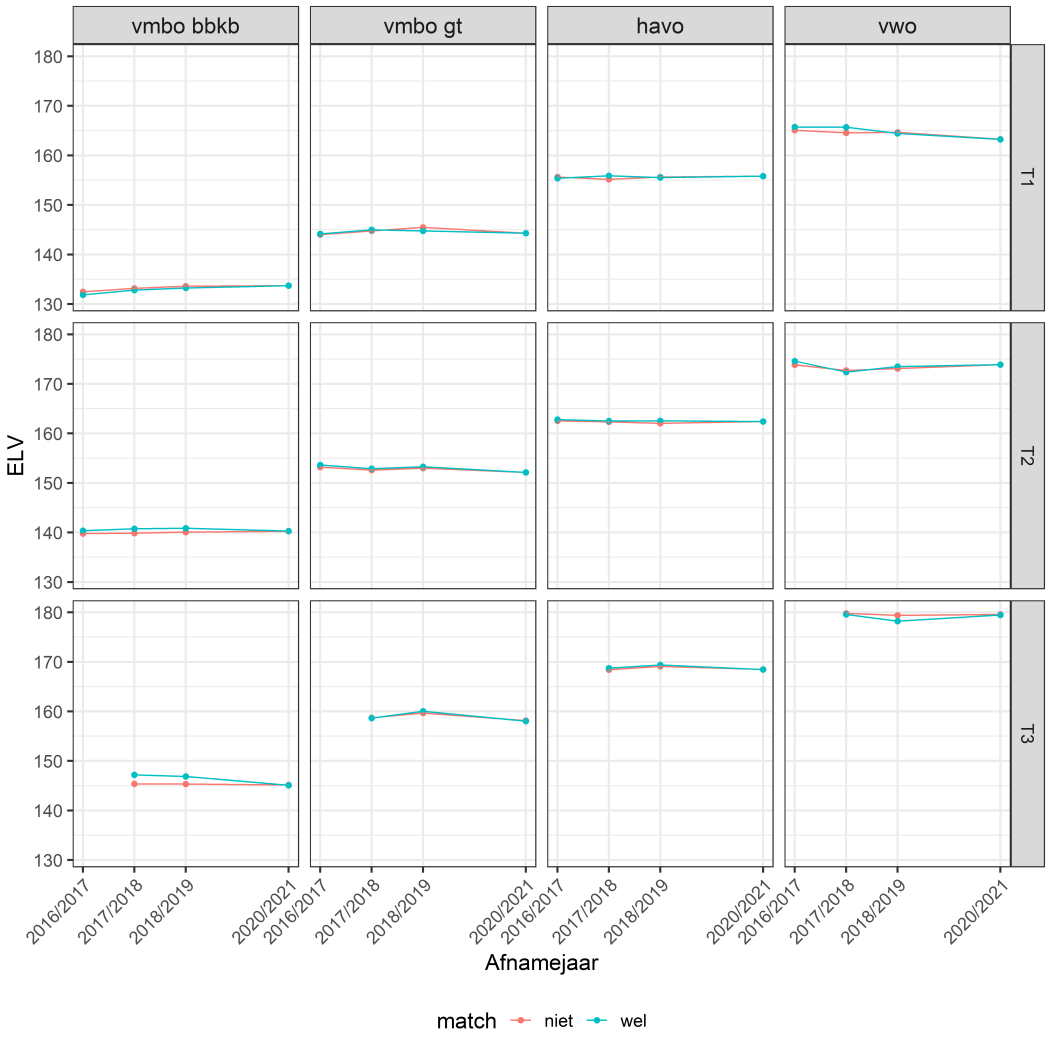
Figuur 4.3: Gemiddelde vaardigheidsscores Nederlands woordenschat

NWS		Cohen's *d*		Percentielverschil
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 1	-0.11	-0.09	-2
vmbo gt	Toets 1	-0.05	-0.14	-5
havo	Toets 1	0.03	-0.19	0
vwo	Toets 1	-0.05	-0.22	-3
vmbo bbkb	Toets 2	-0.24	-0.21	-10
vmbo gt	Toets 2	-0.19	-0.13	-11
havo	Toets 2	-0.17	-0.15	-6
vwo	Toets 2	-0.05	-0.29	0
vmbo bbkb	Toets 3	-0.21		-9
vmbo gt	Toets 3	-0.21		-13
havo	Toets 3	-0.10		-6
vwo	Toets 3	-0.07		-2

Tabel 4.2: Effecten op Nederlands woordenschat (o.b.v. gematchte groepen)

4.1.3 Engels leesvaardigheid

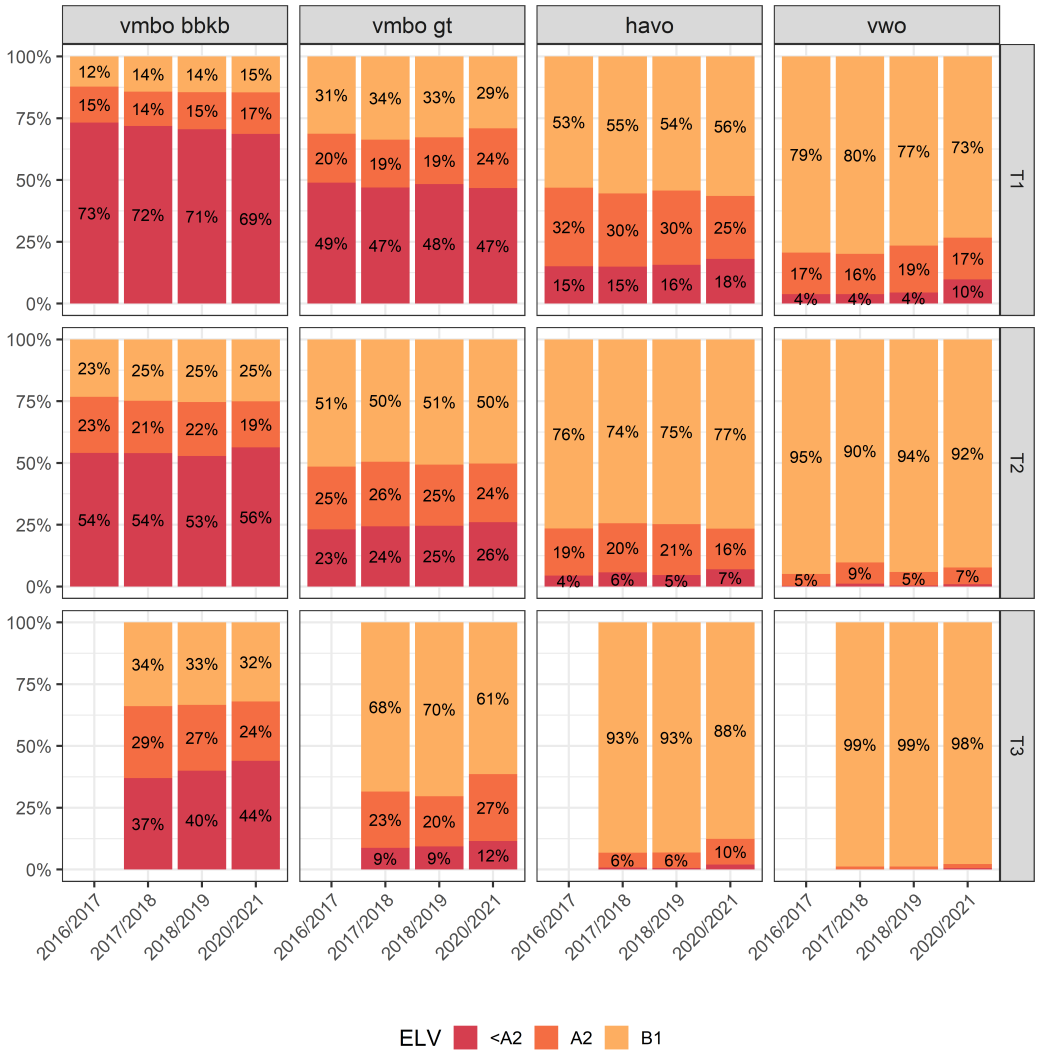
Bij het onderdeel Engels leesvaardigheid is het beeld behoorlijk anders dan bij Nederlands. Zoals te zien in Figuur 4.4 zijn de gemiddelde vaardigheidsscores nauwelijks veranderd in de afgelopen jaren en ook in 2020/2021 stabiel gebleven. Alleen bij Toets 3 in het vmbo zien we een noemenswaardige daling in de gemiddelde prestaties, met effectgroottes van -0.11 en -0.13 voor bbkb en gt respectievelijk (zie Tabel 4.3). Omdat er geen sprake was van een eerdere dalende trend, is het goed mogelijk dat deze daling een gevolg is van de coronacrisis en het thuisonderwijs. In percentielverschillen uitgedrukt zijn de effecten echter redelijk bescheiden, met -4 en -6 respectievelijk. Bij de overige leerjaren en schoolniveaus lijkt er helemaal geen effect te zijn geweest van de schoolsluitingen. Figuur 4.5 toont de percentages leerlingen die op Engels leesvaardigheid de ERK-niveaus A2 en B1 behaalden, en ook hierin valt bovenal op dat de resultaten de laatste jaren erg stabiel zijn.



Figuur 4.4: Gemiddelde vaardigheidsscores Engels leesvaardigheid

ELV		Cohen's *d*		Percentielverschil
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 1	0.03	0.08	0
vmbo gt	Toets 1	-0.03	0.03	-5
havo	Toets 1	0.02	0.01	0
vwo	Toets 1	-0.07	-0.08	-4
vmbo bbkb	Toets 2	-0.03	0.03	-1
vmbo gt	Toets 2	-0.07	-0.02	-3
havo	Toets 2	-0.01	-0.02	0
vwo	Toets 2	0.02	-0.07	0
vmbo bbkb	Toets 3	-0.11		-4
vmbo gt	Toets 3	-0.13		-6
havo	Toets 3	-0.06		-2
vwo	Toets 3	0.08		3

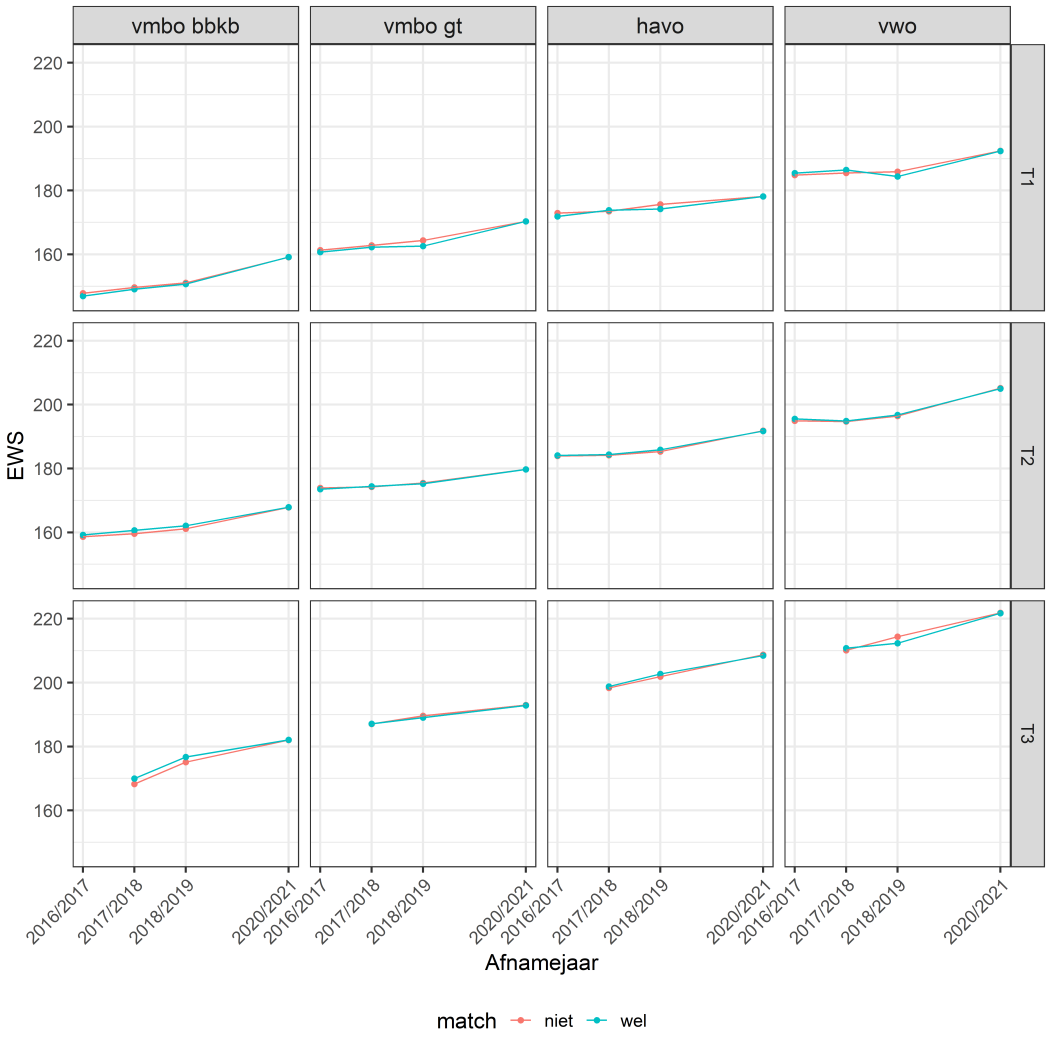
Tabel 4.3: Effecten op Engels leesvaardigheid (o.b.v. gematchte groepen)



Figuur 4.5: Verdeling over ERK-niveaus Engels leesvaardigheid

4.1.4 Engels woordenschat

De Engelse woordenschat van leerlingen in de onderbouw van het voortgezet onderwijs lijkt al jaren een stijgende lijn te vertonen, in alle leerjaren en onderwijsniveaus. Zoals Figuur 4.6 laat zien is deze stijging gewoon gecontinueerd tussen de schooljaren 2018/2019 en 2020/2021. Sterker nog, in deze periode lijkt de gemiddelde vaardigheid alleen maar meer gestegen te zijn, zoals te zien is aan de effectgroottes in Tabel 4.4, die variëren van 0.13 tot maar liefst 0.36 en meestal flink groter zijn dan de "benchmark" effectgroottes voor de vergelijking van 2018/2019 met 2016/2017. Wellicht heeft de periode van thuisonderwijs dus een positief effect gehad op de Engelse woordenschat van de leerlingen. Om de effectgroottes verder in perspectief te plaatsen, merken we op dat de gemiddelde effectgrootte tussen aansluitende leerjaren in 2018/2019 op Engelse woordenschat 0.54 was, en die tussen schooltypen 0.45 (zie Tabel 3.1). Een stijging in de vaardigheid met een effectgrootte rond 0.20 of zelfs 0.30 lijkt dan in verhouding best groot.



Figuur 4.6: Gemiddelde vaardigheidsscores Engels woordenschat

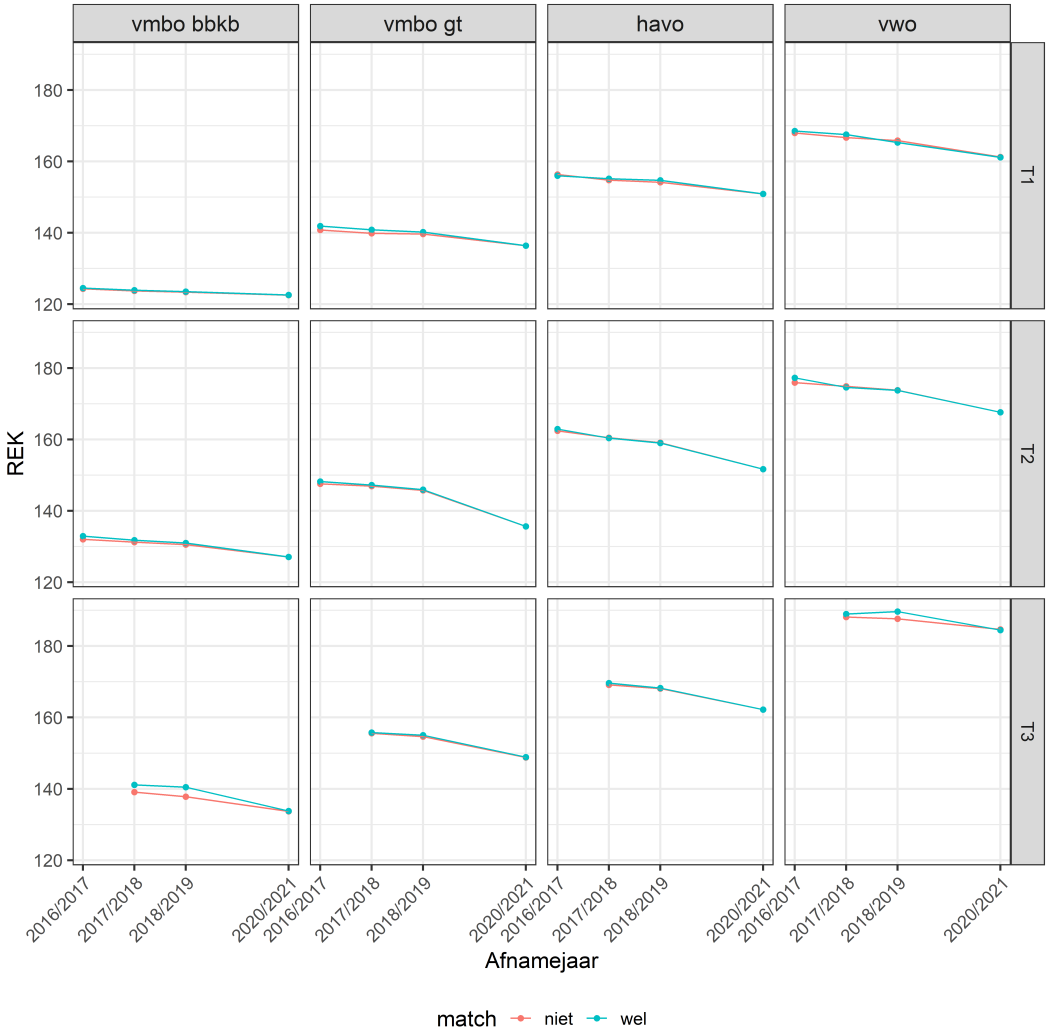
EWS		Cohen's *d*		Percentielverschil
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 1	0.35	0.16	13
vmbo gt	Toets 1	0.30	0.08	9
havo	Toets 1	0.16	0.10	5
vwo	Toets 1	0.28	-0.04	11
vmbo bbkb	Toets 2	0.23	0.12	8
vmbo gt	Toets 2	0.19	0.07	6
havo	Toets 2	0.26	0.08	9
vwo	Toets 2	0.36	0.05	13
vmbo bbkb	Toets 3	0.14		4
vmbo gt	Toets 3	0.13		3
havo	Toets 3	0.18		7
vwo	Toets 3	0.30		11

Tabel 4.4: Effecten op Engels woordenschat (o.b.v. gematchte groepen)

4.1.5 Rekenen

De gemiddelde vaardigheidsscores voor Rekenen zijn afgebeeld in Figuur 4.7. In de eerdere schooljaren leek er al sprake te zijn van licht negatieve trends, hoewel de verschillen tussen jaren vaak klein waren. Het schooljaar 2020/2021 springt er bij veel schooltypen en leerjaren behoorlijk uit omdat de gemiddelde Rekenvaardigheid is gedaald. Zoals in Tabel 4.5 te zien is, zijn de effectgroottes en de percentielverschillen voor de vergelijking tussen de schooljaren 2020/2021 en 2018/2019 vaak fors. Ter vergelijking: in 2018/2019 was de gemiddelde effectgrootte van schooltype op Rekenen 0.97 (zie Tabel 3.1); de gemiddelde effectgrootte tussen leerjaar 1 en 2 was toen 0.41; en die tussen leerjaar 2 en 3 0.67 (zie Tabel A.9). Bij de meeste schooltypen en leerjaren gaat het dus om opvallende dalingen in de vaardigheid; de enige groep waarbij de resultaten stabiel zijn gebleven is het eerste leerjaar van het vmbo bbkb. En verder geldt alleen nog voor vwo Toets 1 dat de daling van een vergelijkbare omvang was als de daling tussen 2016/2017 en 2018/2019, en dus een normaal vervolg van de eerdere trend zou kunnen zijn. Bij alle andere schooltypen en leerjaren is de daling zo groot dat deze afwijkt van de eerdere trend, en ligt het dus voor de hand om de oorzaak te zoeken in effecten van de coronacrisis, waaronder de schoolsluitingen, in 2020 en 2021.

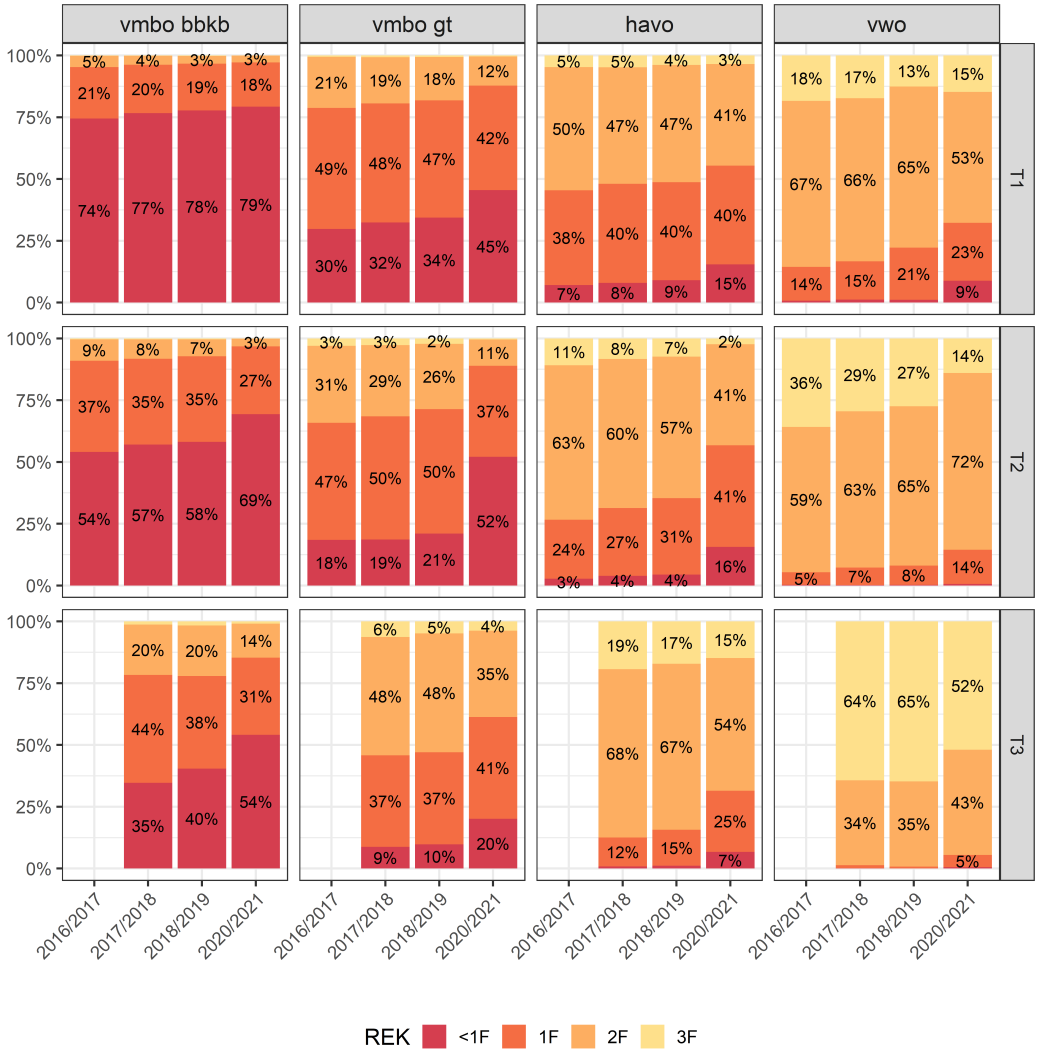
Ook in Figuur 4.8 is duidelijk te zien dat de verdeling van de vaardigheid in het schooljaar 2020/2021 omlaag is verschoven vergeleken met voorgaande jaren, waardoor er nu in de meeste schooltypen en leerjaren aanzienlijk meer leerlingen bleven steken op <1F of 1F niveau en minder leerlingen het 2F of 3F niveau behaalden.



Figuur 4.7: Gemiddelde vaardigheidsscores Rekenen

REK		Cohen's *d*		Percentielverschil
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 1	-0.06	-0.06	-1
vmbo gt	Toets 1	-0.27	-0.12	-10
havo	Toets 1	-0.24	-0.08	-12
vwo	Toets 1	-0.24	-0.20	-12
vmbo bbkb	Toets 2	-0.25	-0.12	-9
vmbo gt	Toets 2	-0.68	-0.14	-28
havo	Toets 2	-0.47	-0.25	-20
vwo	Toets 2	-0.40	-0.22	-19
vmbo bbkb	Toets 3	-0.36		-15
vmbo gt	Toets 3	-0.38		-19
havo	Toets 3	-0.35		-18
vwo	Toets 3	-0.27		-12

Tabel 4.5: Effecten op Rekenen (o.b.v. gematchte groepen)



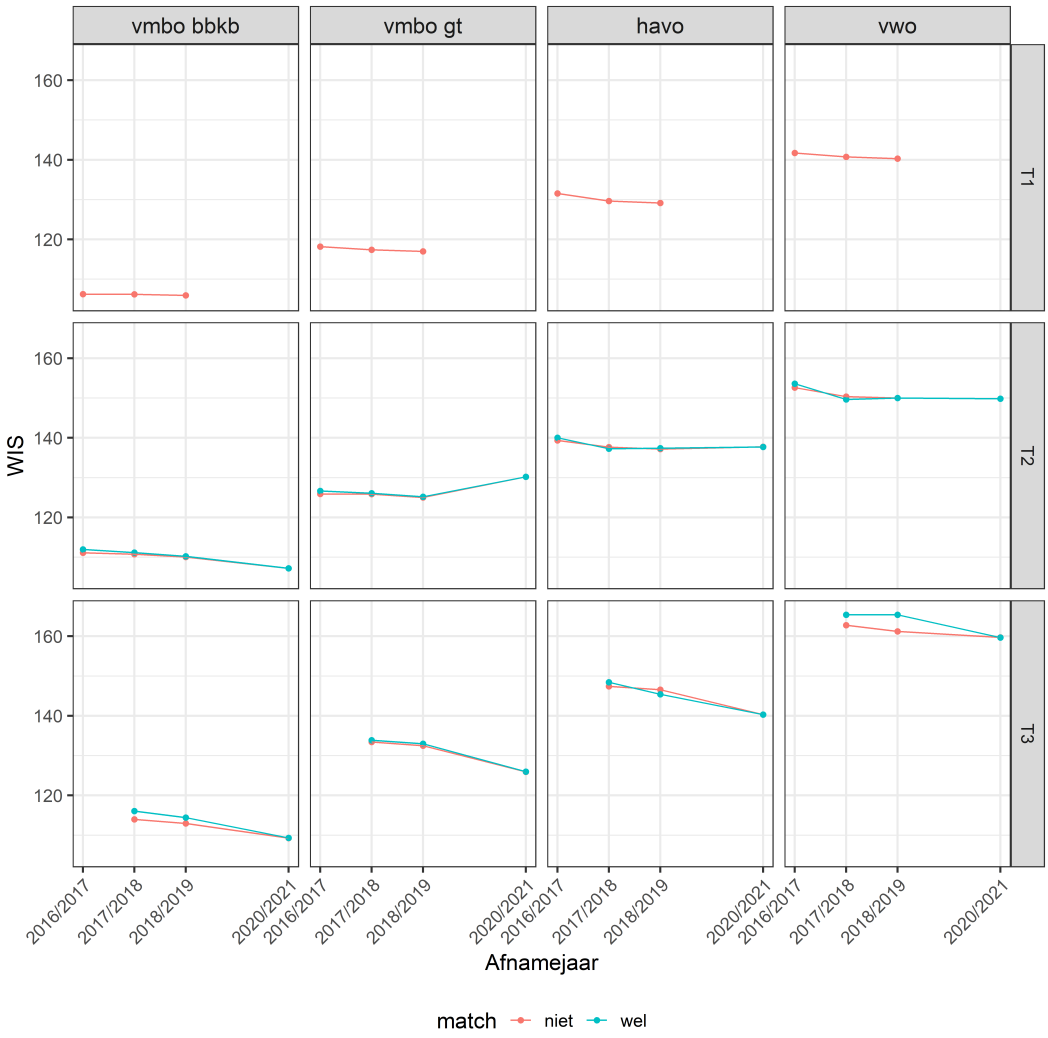
Figuur 4.8: Verdeling over referentieniveaus Rekenen

4.1.6 Wiskunde

Van het onderdeel Wiskunde hebben we voor het schooljaar 2020/2021 alleen scores op Toets 2 en Toets 3 (zie Sectie 2.2). De scores op Toets 1 in eerdere schooljaren zijn wel meegenomen in Figuur 4.9 en daarin was er nauwelijks sprake van fluctuaties. Voor Toets 2 lijken de resultaten ook redelijk stabiel, tot en met het schooljaar 2020/2021. Alleen bij vmbo bbkb zien we dat de gemiddelde vaardigheid al een aantal jaar wat daalt. Bij vmbo gt lijkt er in 2020/2021 een toename in de vaardigheid te zijn, die we echter moeilijk kunnen plaatsen. Bij havo en vwo is de gemiddelde vaardigheid bijna exact gelijk gebleven. Voor Toets 3 beschikken we over data vanaf schooljaar 2017/2018 en daar zien we bij alle schooltypen dat de prestaties in 2020/2021 gemiddeld minder goed zijn dan twee jaar geleden[2]. In Tabel 4.6 is te zien dat de effectgroottes van deze verschuiving op Toets 3 redelijk groot zijn, met een range van -0.27 tot -0.43. Ter vergelijking: de effectgrootte

[2]Als de resultaten voor de ongematchte en gematchte groepen verschillen, zoals hier bij Toets 3 in het vwo, dan interpreteren we de resultaten voor de gematchte groepen omdat we deze het meest betrouwbaar achten.

van het verschil tussen leerjaar 2 en 3 op Wiskunde in schooljaar 2018/2019 varieerde van 0.26 in het vmbo bbkb tot 0.93 in het vwo (zie Tabel A.11 in de Appendix), en de effectgrootte tussen schooltypen was toen gemiddeld 0.98 (zie Tabel 3.1). Het is lastig om met zekerheid te zeggen in hoeverre de daling in de score op Toets 3 onderdeel uit kan maken van een bestaande trend, of door de coronacrisis kan komen. Bij vmbo bbkb en havo Toets 3 was er in ieder geval ook al een daling in de vaardigheid op Wiskunde tussen 2017/2018 en 2018/2019 en zit er feitelijk geen knik in de trendlijn; bij vmbo gt en vwo Toets 3 zien we wel een duidelijke trendbreuk in 2020/2021.



Figuur 4.9: Gemiddelde vaardigheidsscores Wiskunde

WIS		Cohen's *d*		Percentielverschil
Schooltype	**Toets**	**2021-2019**	**2019-2017**	**2021-2019**
vmbo bbkb	Toets 2	-0.20	-0.12	-9
vmbo gt	Toets 2	0.36	-0.11	15
havo	Toets 2	0.02	-0.19	0
vwo	Toets 2	-0.01	-0.24	0
vmbo bbkb	Toets 3	-0.27		-12
vmbo gt	Toets 3	-0.43		-19
havo	Toets 3	-0.27		-13
vwo	Toets 3	-0.30		-13

Tabel 4.6: Effecten op Wiskunde (o.b.v. gematchte groepen)

4.2 Trends voor subpopulaties

4.2.1 Geslacht

Om na te gaan of er verschillen zijn in de trendresultaten voor jongens en meisjes hebben we de gemiddelde vaardigheid per onderdeel, afnamejaar, schooltype en toets (Toets 1, Toets 2 of Toets 3) voor jongens en meisjes geplot. De Figuren staan in de Appendix, en in deze sectie bespreken we per onderdeel wat er opviel.

Voor het onderdeel Nederlands leesvaardigheid ziet de trend van de afgelopen jaren er vergelijkbaar uit voor jongens en meisjes. Meisjes behalen gemiddeld vaak iets hogere toetsscores, met name in het vmbo en de havo. Het verschil lijkt soms ietsje kleiner te zijn geworden in 2020/2021, bijvoorbeeld bij Toets 2 in vmbo gt en havo. Maar al met al zien we geen duidelijke trendbreuk of verschuivingen in de relatieve prestaties van jongens en meisjes. Bij Nederlands woordenschat presteren jongens gemiddeld juist vaak iets beter dan meisjes. Dit verschil lijkt in 2020/2021 nog ietsje te zijn toegenomen bij een aantal schooltypen en leerjaren, zoals vmbo gt en vwo Toets 1, vmbo gt Toets 2, en vmbo gt en havo Toets 3, doordat de gemiddelde vaardigheid van de meisjes wat (meer) is gedaald.

Kijken we naar Engels leesvaardigheid, dan is er geen consistent patroon te ontwaren in de verschillen tussen jongens en meisjes over leerjaren en schooltypen heen. In sommige leerjaren en schooltypes lijken jongens gemiddeld wat hoger te scoren dan meisjes, in andere leerjaren of schooltypes zijn hun prestaties zeer vergelijkbaar. In het vmbo gt ziet het ernaar uit dat het gemiddelde verschil in vaardigheid tussen jongens en meisjes in 2020/2021 nagenoeg verdwenen is, maar dit leek ook in eerdere schooljaren al kleiner te worden doordat de twee groepen naar elkaar toe verschoven.

Bij de trend in de Engelse woordenschat van jongens en meisjes zien we duidelijk dat jongens gemiddeld op deze vaardigheid doorgaans flink hogere scores behaalden dan meisjes, in alle leerjaren en schooltypen. Bij de eerste twee leerjaren, Toets 1 en Toets 2, lijkt het verschil in 2020/2021 wel iets kleiner te zijn geworden. Bij vmbo gt en havo Toets 1, en bij vwo Toets 2, is er nauwelijks nog verschil in 2020/2021, omdat de gemiddelde vaardigheid van de meisjes wat sterker is gestegen dan die van de jongens.

Voor Rekenen ziet de trend over de tijd voor jongens en meisjes er zeer vergelijkbaar uit. Jongens behalen in alle schooltypen en leerjaren gemiddeld hogere vaardigheidsscores op dit onderdeel dan meisjes. Verder lijken de trendlijnen voor jongens en meisjes bijna parallel te lopen bij alle schooltypen en leerjaren. Wat opvalt is dat de daling in de gemiddelde

Rekenvaardigheid, vooral tussen 2018/2019 en 2020/2021, in sommige leerjaren en schooltypen behoorlijk groot is, ook relatief aan het verschil tussen jongens en meisjes.

Anders dan bij Rekenen zien we bij Wiskunde wel wat verschillende ontwikkelingen tussen jongens en meisjes, ook afhankelijk van het schooltype en de toets (Toets 2 of Toets 3). Gemiddeld genomen is de vaardigheid van jongens altijd hoger dan die van meisjes, en meestal was dat verschil op Toets 3 groter dan op Toets 2. Op Toets 2 lijkt het erop dat het verschil tussen jongens en meisjes in 2020/2021 iets groter is geworden, met name bij vmbo bbkb omdat de vaardigheid van de meisjes daar flink is gedaald, en bij vmbo gt juist omdat de vaardigheid van de jongens veel sterker is gestegen. Bij Toets 3 zien we niet zulke duidelijke verschillen in hoe de trendlijnen lopen tussen 2018/2019 en 2020/2021; de groepen zijn in die twee jaar vergelijkbaar gedaald in hun wiskundevaardigheid.

4.2.2 Achterstandsscore van de school

We hebben ook de trends vergeleken voor leerlingen van scholen met verschillende achterstandsscores (zie Sectie 2.3). Per vaardigheid hebben we de achterstandsscores van de scholen ingedeeld in vier categorieën op zo'n manier dat er evenveel leerlingen in elke categorie vallen; de gehanteerde grenswaarden van de categorieën zijn daarom niet *exact* gelijk voor verschillende vaardigheden. In de Appendix staan de trendplots waarin voor elk van de vier leerlinggroepen een andere kleur lijn is gebruikt.

Als we kijken naar de trendplots voor Nederlands leesvaardigheid dan vallen een aantal zaken op. Ten eerste zijn de vier categorieën van achterstandsscholen niet duidelijk onderscheidbaar qua gemiddelde leerlingsscores; het is niet zo dat de ordening van de gemiddelde vaardigheid netjes correspondeert met de ordening naar achterstandsscore. Alleen bij Toets 2 en Toets 3 van het vmbo bbkb lijkt dit enigszins het geval te zijn. Ten tweede lijkt de ordening van de achterstandsscores ook niet duidelijk samen te hangen met de trend(verschillen) over de jaren heen. Voorzover we soms een verschil kunnen zien in een bepaald schooljaar, bij een bepaald schooltype en toets, lijkt het geen onderdeel te zijn van een patroon. We zien hier bijvoorbeeld dus ook géén indicatie dat scholen met een hogere achterstandsscore meer moeite zouden hebben gehad met het afstandsonderwijs waardoor de Nederlandse leesvaardigheid van hun leerlingen (meer) zou zijn gedaald.

Ook bij Nederlands woordenschat lijkt de trend over de tijd weinig samen te hangen met de achterstandsscore van een school. Wel zien we bij dit onderdeel dat de ordening naar achterstandsscores in het vmbo bbkb redelijk correspondeert met de ordening van de gemiddelde vaardigheid. Bij het vmbo gt en de havo lijkt de groep met de hoogste achterstandsscore ook wel vaak de laagste vaardigheidsscore te behalen op Nederlands woordenschat, maar zijn de andere drie categorieën niet consistent geordend zoals men zou verwachten. De ontwikkelingen over de tijd lijken weinig of geen samenhang te vertonen met de achterstandsscore.

Bij de onderdelen Engels leesvaardigheid en Engels woordenschat zien we in beide gevallen dat het nauwelijks mogelijk is om een patroon te ontwaren in de trends van de verschillende groepen op basis van de achterstandsscore. Er is geen consistent verschil te bespeuren in de gemiddelde vaardigheid tussen de vier categorieën van scholen, en de trends over de tijd zijn ook vaak hetzelfde voor de vier groepen. Voorzover we hierin verschillen zien, lijken deze vooral grillig te zijn en niet interpreteerbaar. Hetzelfde geldt voor de onderdelen Rekenen en Wiskunde, alhoewel we daar wel kunnen opmerken dat de

ordening naar de achterstandsscore in het vmbo bbkb, met name op Toets 3, meestal redelijk correspondeert met de ordening van de gemiddelde vaardigheid. We zien echter geen interpreteerbare verschillen in de *trends* voor de vier categorieën van achterstandsscores.

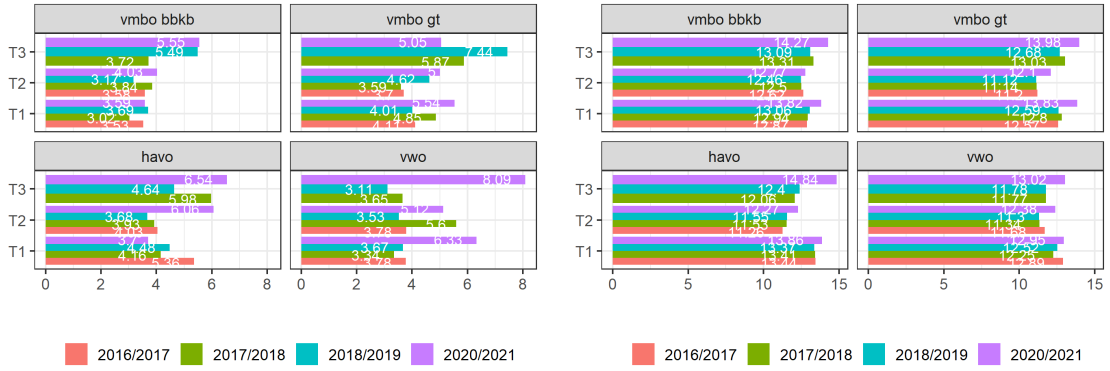
4.3 Spreiding in de toetsscores

In deze sectie bespreken we verschillen in spreiding tussen en binnen scholen. Op de gegevens van de afnamejaren 2016/2017, 2017/2018, 2018/2019 en 2020/2021 zijn afzonderlijke multilevel modellen geschat per onderdeel, toets en schoolniveau (zie Sectie 3.4). Dit leverde schattingen op van de spreiding binnen scholen, en tussen scholen. Per onderdeel vergelijken we de resultaten van 2020/2021 met die van de voorgaande jaren. Om vertekening in de resultaten door verschillende schooleigenschappen te voorkomen, is de dataset voor de multilevel analyse beperkt tot die scholen, die in alle vier van de geanalyseerde schooljaren afnames hadden. Verschillen tussen afnamejaren bij één toetsmoment (Toets 1, 2 en 3), niveau en onderdeel kunnen dus niet toegeschreven worden aan verschillen in de eigenschappen van de deelnemende scholen; zo wordt het beter mogelijk om zulke verschillen te interpreteren als effect van de schoolsluitingen.

4.3.1 Nederlands leesvaardigheid

In Figuur 4.10 staan links de standaarddeviaties van de schoolgemiddelden, en rechts die van de leerlingscores binnen scholen voor het onderdeel Nederlands leesvaardigheid. De spreiding tussen de scholen wisselde in de jaren 2016/2017, 2017/2018 en 2018/2019 al enigszins. In het schooljaar 2020/2021 is de spreiding tussen scholen vaak groter dan voorheen, al geldt dit niet voor alle leerjaren en schooltypen. Een toename in de spreiding tussen scholen is vooral te zien bij Toets 2 en Toets 3 in de havo en bij alle leerjaren in het vwo. In het derde leerjaar van het vwo is de spreiding tussen scholen het sterkste toegenomen: deze is ruim meer dan verdubbeld ten opzichte van 2018/2019.

De spreiding binnen de scholen, dus op leerlingniveau, is in 2020/2021 ook vaak iets groter dan in de voorgaande jaren, terwijl die toen redelijk stabiel was. Bij het vmbo bbkb en gt is deze spreiding in alle leerjaren iets toegenomen, terwijl de spreiding op schoolniveau daar niet altijd toegenomen was. Bij de havo en het vwo is de toename in de spreiding op leerlingniveau bescheiden vergeleken bij de toenames in de spreiding tussen scholen die we daar (overal behalve bij havo Toets 1) zien.



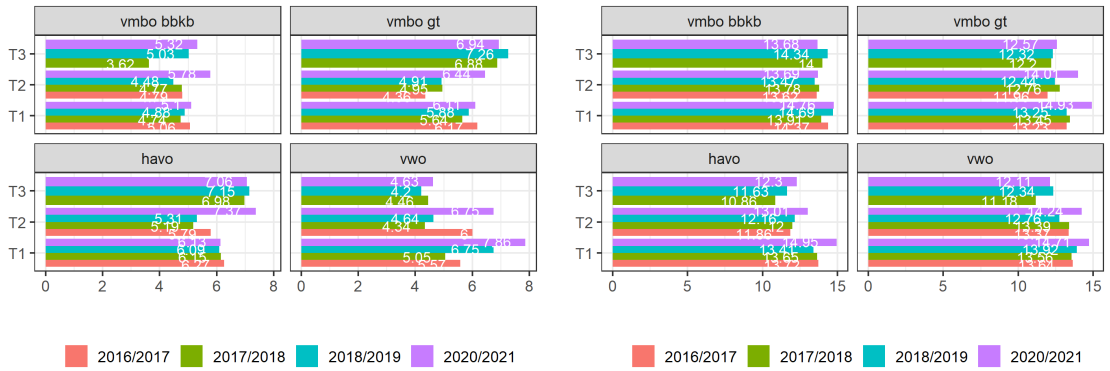
(a) Standaarddeviatie tussen scholen (b) Standaarddeviatie binnen scholen

Figuur 4.10: Spreiding leeruitkomsten Nederlands leesvaardigheid

4.3.2 Nederlands woordenschat

De spreidingsresultaten voor het onderdeel Nederlands woordenschat staan in Figuur 4.11. De spreiding tussen de scholen varieerde al enigszins in de voorgaande schooljaren. Bij het vmbo bbkb en gt en bij de havo lijkt deze spreiding in 2020/2021 alleen op Toets 2 iets te zijn toegenomen. Bij het vwo is zowel op Toets 2 als Toets 1 een toename te zien, maar de spreiding op Toets 1 was in 2018/2019 ook al flink gestegen ten opzichte van 2017/2018, dus dit lijkt op het vervolg van een al eerder ingezette trend.

In het vmbo bbkb zien we geen noemenswaardige veranderingen in de spreiding binnen scholen. Bij het vmbo gt is deze spreiding in 2020/2021 voor Toets 1 en Toets 2 wel toegenomen ten opzichte van de eerdere jaren, terwijl die toen steeds behoorlijk stabiel was. Ook bij de havo en het vwo zijn de verschillen tussen leerlingen binnen scholen iets groter geworden op Toets 1 en Toets 2. Op Toets 3 lijkt er bij alle schooltypen weinig veranderd te zijn in deze spreiding in 2020/2021.



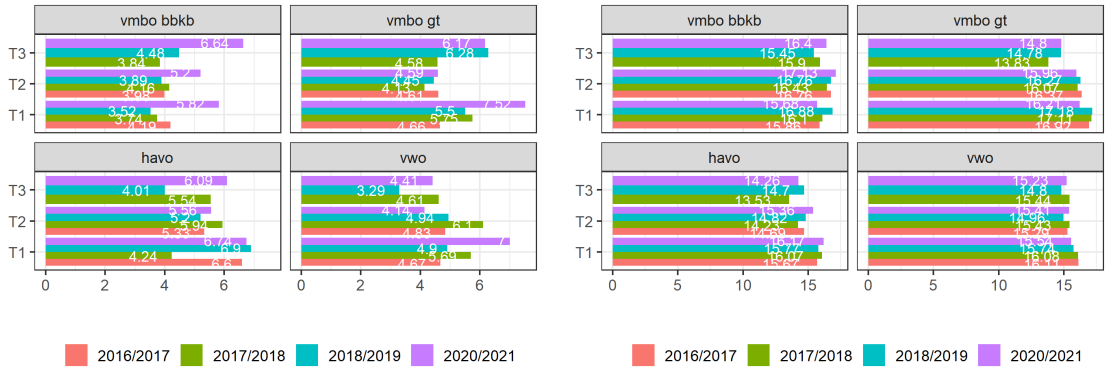
(a) Standaarddeviatie tussen scholen (b) Standaarddeviatie binnen scholen

Figuur 4.11: Spreiding leeruitkomsten Nederlands woordenschat

4.3.3 Engels leesvaardigheid

In Figuur 4.12 staan de spreidingsresultaten bij het onderdeel Engels leesvaardigheid. De spreiding tussen scholen was vrij grillig in de afgelopen schooljaren; tussen 2016/2017 en 2018/2019 zien we zowel forse toenames als afnames, afhankelijk van het leerjaar en schooltype. In 2020/2021 zien we vaak dat de spreiding flink is toegenomen, en dan hoger ligt dan in de eerdere drie schooljaren is voorgekomen. In het vmbo bbkb is deze spreiding toegenomen in alle leerjaren, maar in het vmbo gt alleen op Toets 1. Bij de havo is de spreiding alleen op Toets 3 toegenomen, maar wel fors; bij het vwo is deze toegenomen bij Toets 3 en (nog sterker) bij Toets 1.

De spreiding binnen scholen ziet er voor alle schooltypen en leerjaren erg stabiel uit over de jaren heen, en lijkt ook in 2020/2021 nauwelijks te zijn veranderd. De verschillen tussen individuele leerlingen in hun Engels leesvaardigheid zijn dus niet toegenomen in 2020/2021, maar de verschillen tussen scholen vaak wel. Dit kan erop wijzen dat scholen nogal verschilden in de mate waarin zij erin slaagden om tijdens de periodes van afstandsonderwijs te werken aan de Engelse leesvaardigheid van hun leerlingen.



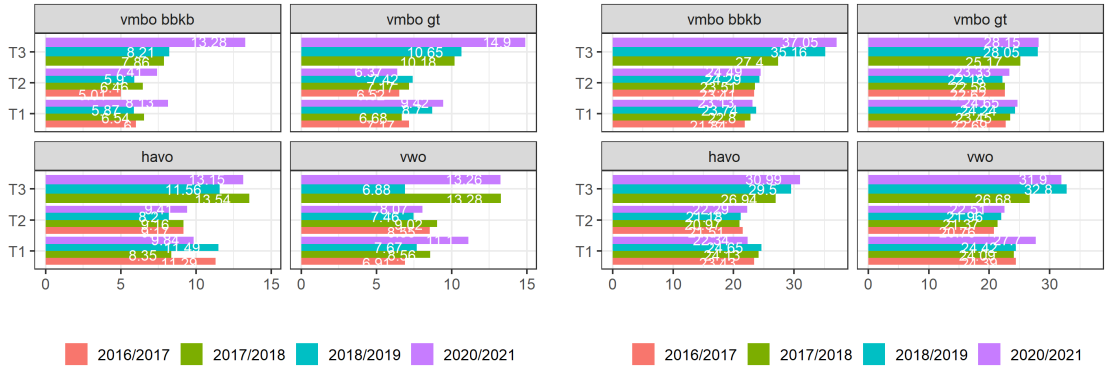
(a) Standaarddeviatie tussen scholen (b) Standaarddeviatie binnen scholen

Figuur 4.12: Spreiding leeruitkomsten Engels leesvaardigheid

4.3.4 Engels woordenschat

De spreidingsresultaten op het onderdeel Engels woordenschat zijn weergegeven in Figuur 4.13 en zien er weer anders uit dan die op Engels leesvaardigheid. De spreiding tussen scholen is opvallend sterk toegenomen op Toets 3 in 2020/2021 voor het vmbo bbkb, vmbo gt, en het vwo. Verder zien we ook een forse stijging bij Toets 1 in het vmbo bbkb en in het vwo. Maar voor de andere combinaties van schooltype/leerjaar is er geen opvallende verandering in deze spreiding.

De spreiding binnen scholen lijkt meestal weinig veranderd te zijn ten opzichte van 2018/2019. Waar die wel iets groter is geworden, lijkt dit onderdeel te zijn van een trend omdat de spreiding daar in de eerdere schooljaren ook al toenam. Er is dus geen indicatie dat de verschillen tussen leerlingen binnen scholen in Engels woordenschat opvallend zijn veranderd in 2020/2021.



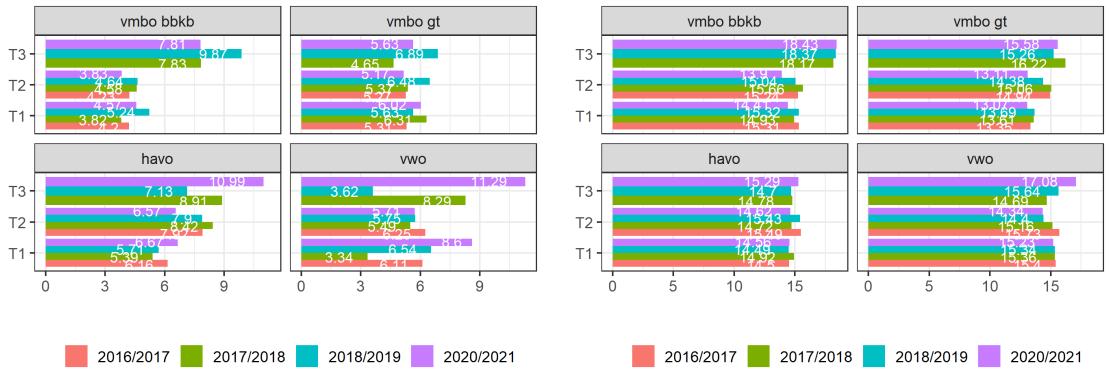
(a) Standaarddeviatie tussen scholen (b) Standaarddeviatie binnen scholen

Figuur 4.13: Spreiding leeruitkomsten Engels woordenschat

4.3.5 Rekenen

In Figuur 4.14 staan de spreidingsresultaten voor het onderdeel Rekenen. Wat direct opvalt is dat de spreiding tussen scholen in 2020/2021 heel sterk is toegenomen in het derde leerjaar van havo en vwo. Deze spreiding was daar weliswaar al wat grillig in de voorgaande schooljaren, maar is in 2020/2021 naar een recordhoogte gestegen. Ook op Toets 1 in het vwo is de spreiding toegenomen. Deze toenames kunnen wijzen op een schooleffect in hoe goed het afstandsonderwijs in Rekenen gelukt is. In het vmbo zien we geen noemenswaardige veranderingen in de spreiding tussen scholen in 2020/2021.

De spreiding binnen scholen op Rekenvaardigheid lijkt voor de meeste schooltypen en leerjaren weinig te zijn veranderd. Op Toets 1 en Toets 2 in het vmbo bbkb en gt is deze in 2020/2021 wat kleiner geworden, maar het verschil is klein. In het vwo is de spreiding op Toets 3 iets groter, maar dit lijkt een vervolg te zijn van een eerdere trend. Er is dus geen aanwijzing voor een effect van de schoolsluitingen op de spreiding tussen leerlingen binnen scholen.



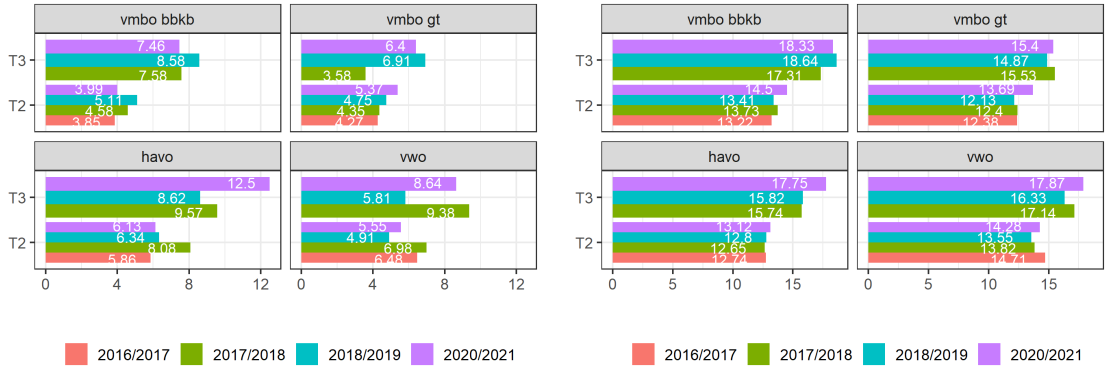
(a) Standaarddeviatie tussen scholen (b) Standaarddeviatie binnen scholen

Figuur 4.14: Spreiding leeruitkomsten Rekenen

4.3.6 Wiskunde

Figuur 4.15 toont de spreidingsresultaten voor het onderdeel Wiskunde. Net als bij Rekenen zien we een forse stijging in 2020/2021 in de spreiding tussen scholen op Toets 3 in havo en vwo. Bij vwo Toets 3 was deze spreiding echter nog hoger in 2017/2018, en kan het dus ook goed gaan om normale fluctuaties die niet met de schoolsluitingen te maken hebben. We zien verder geen relevante verschuivingen in de spreiding tussen scholen in het vmbo.

De spreiding binnen scholen is in 2020/2021 iets toegenomen in de havo en het vwo. In het vmbo is dit alleen het geval in het tweede leerjaar. Net als bij de andere onderdelen geldt wel dat als we toenames zien in de spreiding binnen scholen, deze meestal subtieler zijn dan de toenames in de spreiding tussen scholen. Dit wijst in de richting dat na twee jaren met (deels) afstandsonderwijs de verschillen tussen scholen meer zijn (uit-)vergroot dan de verschillen tussen individuele leerlingen.



(a) Standaarddeviatie tussen scholen

(b) Standaarddeviatie binnen scholen

Figuur 4.15: Spreiding leeruitkomsten Wiskunde

5. Conclusie en Discussie

In deze studie hebben we onderzoek gedaan naar mogelijke effecten van de coronacrisis en de lockdownperiodes in 2020 en 2021 op de leerresultaten in de onderbouw van het voortgezet onderwijs. Hiervoor hebben we gebruik gemaakt van toetsgegevens afkomstig uit het Cito volgsysteem voortgezet onderwijs (cvvo). Specifieker hebben we toetsresultaten gebruikt die verzameld zijn in de eerste drie leerjaren van het voortgezet onderwijs in de periode van schooljaar 2016/2017 tot en met mei van 2021, en hebben we de leerlingprestaties in het schooljaar 2020/2021 vergeleken met die in de "voor-corona" schooljaren. De onderdelen uit het cvvo zijn Nederlands leesvaardigheid, Nederlands woordenschat, Engels leesvaardigheid, Engels woordenschat, Rekenen en Wiskunde.

De resultaten van deze studie geven een wisselend beeld. Op de onderdelen Nederlands leesvaardigheid en Rekenen zien we dat de leerlingen in 2020/2021 over het algemeen slechter presteren dan in de jaren 2016/2017 tot en met 2018/2019. Bij Nederlands leesvaardigheid zien we dit het duidelijkste in het vmbo en de havo, terwijl bij Rekenen over alle schooltypen heen de prestaties flink gedaald zijn. Deze dalingen bij Nederlands leesvaardigheid en Rekenen komen ook tot uitdrukking in de verdeling over referentieniveaus: relatief meer leerlingen presteren in 2020/2021 onder of op het 1F-niveau, en minder leerlingen op het 2F- of 3F-niveau. Het aandeel leerlingen dat onder of op het 1F-niveau leesvaardigheid presteert, is onder sommige groepen leerlingen behoorlijk gestegen. Dat geldt met name voor leerlingen aan het einde van leerjaar 1. Daar is het percentage leerlingen dat onder of op dit niveau presteert gestegen met 7, 8, 10 en 10% voor respectievelijk leerlingen in het vmbo bbkb, vmbo gt, havo en vwo. In met name het vmbo en havo zijn er ook aanzienlijke effecten zichtbaar in het tweede en derde leerjaar. Ook voor Rekenen geldt dat het percentage leerlingen dat in 2020/2021 op een lager referentieniveau presteert, gestegen is ten opzichte van eerdere jaren. Op deze twee onderdelen lijkt het er dus op dat de effecten van de coronacrisis, zoals het afstandsonderwijs, tot leerachterstanden hebben geleid. Op de andere onderdelen zien we echter geen dergelijke daling van de gemiddelde

vaardigheid in 2020/2021. Bij Nederlands woordenschat en Engels leesvaardigheid lijken de gemiddelde leerlingprestaties weinig veranderd te zijn, of hooguit in lijn met trends die al voor de coronacrisis begonnen waren. En op het onderdeel Wiskunde is geen consistent patroon te ontwaren van veranderingen in 2020/2021. Bijzonder is dat op het onderdeel Engels woordenschat de gemiddelde scores al jarenlang over de hele linie (in alle schooltypen en leerjaren van de onderbouw) stijgen, en dat deze stijging in 2020/2021 alleen maar sneller heeft doorgezet. Het lijkt dus wel alsof de schoolsluitingen een positief effect hebben gehad op de Engelse woordenschat van de leerlingen.

Naast de gemiddelde vaardigheid is ook de scorespreiding onderzocht. De spreiding in de scores tussen scholen en binnen scholen geeft een ander perspectief op mogelijke effecten van de lockdownperiodes. Een toegenomen scorespreiding *tussen* scholen zou namelijk kunnen wijzen op verschillen in hoe goed de scholen konden omgaan met het afstandsonderwijs en andere coronamaatregelen. Leerlingen presteren dan meer of minder goed afhankelijk van de school waar ze op zitten. De spreiding *binnen* scholen daarentegen geeft een indicatie van verschillen tussen individuele leerlingen, die bijvoorbeeld te maken kunnen hebben met de thuissituatie of de motivatie van de leerlingen. Uit onze resultaten bleek dat de spreiding tussen scholen op de verschillende onderdelen in 2020/2021 meestal was toegenomen of gelijk was gebleven; van een afname was zelden sprake. Soms was de spreiding tussen scholen heel fors toegenomen; dit zagen we wat vaker bij havo en vwo dan bij het vmbo. De spreiding binnen scholen was vaak stabiel, of enigszins toegenomen. Wanneer die was toegenomen ging het echter meestal om een relatief bescheiden toename vergeleken met de toename in de spreiding tussen scholen. Samengenomen wijzen deze resultaten dus in de richting van toegenomen schooleffecten op de leerlingprestaties in 2020/2021.

In dit onderzoek is ook gekeken naar mogelijke differentiële effecten van de coronacrisis en de lockdownperiodes op subpopulaties van leerlingen. We onderscheidden daarbij subpopulaties op basis van het geslacht van de leerling of de achterstandsscore van de school. De verschillen tussen jongens en meisjes lijken op de meeste onderdelen weinig veranderd te zijn in 2020/2021. Meisjes presteren gemiddeld meestal beter op Nederlands leesvaardigheid, en jongens op Nederlands woordenschat, Engels woordenschat, Rekenen en Wiskunde. Op het onderdeel Nederlands woordenschat waren de verschillen in 2020/2021 soms iets kleiner geworden doordat de gemiddelde vaardigheid van meisjes was gedaald. Op Engels woordenschat zagen we juist dat de meisjes soms wat vooruit waren gegaan, waardoor het geslachtsverschil kleiner werd. Maar er waren weinig effecten die voor alle leerjaren of schooltypen golden, dus al met al kunnen we hier weinig conclusies aan ophangen. Voor de achterstandsscore op schoolniveau gold dat deze geen samenhang leek te vertonen met verschuivingen in 2020/2021 of met trendverschillen. Sowieso leek de indeling in vier categorieën van lage naar hoge achterstandsscores eigenlijk alleen in het vmbo bbkb enigszins gerelateerd aan de gemiddelde leerlingprestaties. Bij de andere schooltypen leek de achterstandsscore niet samen te hangen met de gemiddelde vaardigheid.

Een belangrijke beperking van deze studie is dat er gebruik gemaakt is van observationele data om het effect van de coronacrisis op de leerlingprestaties in kaart te brengen. In tegenstelling tot een experiment, betekent dit dat we niet *voorafgaand aan* de dataverzameling controle hebben uitgeoefend op welke leerlingen of scholen op welk moment

een toets zouden maken of juist niet, en dat er geen controlegroep van leerlingen was die in dezelfde periode regulier onderwijs op school heeft kunnen volgen zonder beperkende maatregelen. We hebben wel gebruik gemaakt van speciale statistische technieken om rekening te houden met mogelijke verschillen in de afnamepopulatie per jaar. Dit probleem hebben we namelijk geaddresseerd door te werken met matching op achtergrondkenmerken, om zo vergelijkbaar mogelijke groepen te verkrijgen voor de analyses. Een ander probleem wat echter blijft bestaan, is dat we op basis van observationele data een gevonden verandering in de leerlingprestaties niet met zekerheid kunnen toeschrijven aan het effect van de coronacrisis en het afstandsonderwijs; er kan immers ook sprake zijn geweest van andere invloeden of trends, waarvoor we niet kunnen corrigeren.

Samenvattend kan uit deze studie geconcludeerd worden dat de coronacrisis en de periodes van schoolsluitingen in 2020 en 2021 vooral een negatief effect lijken te hebben gehad op de leerlingprestaties op Nederlands leesvaardigheid en Rekenen, terwijl de leerlingprestaties op Engels woordenschat alleen maar (extra) verbeterd lijken te zijn. In het vmbo en havo lijkt er bovendien wat vaker sprake te zijn van leerachterstanden in 2020/2021 dan in het vwo. De spreiding tussen scholen was in 2020/2021 vaak opmerkelijk groter dan in voorgaande jaren, en was dan meestal veel sterker toegenomen dan de spreiding tussen leerlingen binnen scholen. Dit duidt op aanzienlijke effecten van hoe verschillende scholen het onderwijs in de coronaperiode hebben vormgegeven, en een relatief bescheiden invloed van individuele leerlingkenmerken op hoe het onderwijs voor hen heeft uitgepakt. Dit patroon in de spreiding kwam vooral naar voren bij de schooltypen havo en vwo.

6. Referenties

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

Centraal Bureau voor de Statistiek (2021). Achterstandsscores Nationaal Programma Onderwijs. Leidschenveen: Centraal Bureau voor de Statistiek.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.

Embretson, S.E. & Reise, S.P. (2000). *Item Response Theory for Psychologists*. Psychology Press.

Emons, W., Frissen, S. & Straat, H. (2021). Effecten van de Schoolsluitingen op de Leerresultaten voor Nederlands en Engels Leesvaardigheid & Woordenschat, en Rekenen-Wiskunde in het Voortgezet Onderwijs. Arnhem: Cito.

Hambleton, R.K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Press.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. `https://www.R-project.org/`.

Rosenbaum, P.R. & Rubin, D.B. (1983). The Central Role of the Propensity Score in Observational Studies for Causal Effect. *Biometrika*. 70(1): 41–55.

Schouten, B., Cobben, F., & Bethlehem, J. (2009), Indicators for the representativeness of survey response, *Survey Methodology*, 35 (1), 101 – 113.

Zijlstra, W., Frissen, S., Koops, J., van Onna, M., & Feskens, R. (2021). Leervertraging door schoolsluitingen in het voortgezet onderwijs. Effecten van de eerste lockdown periode. Arnhem: Cito.

A. Appendix

Deze Appendix bevat de volgende aanvullende gegevens.

In Sectie A.1 staan tabellen met, per onderdeel, de effectgroottes voor de gemiddelde verschillen in het schooljaar 2018/2019 tussen aansluitende leerjaren binnen een schooltype, of tussen aansluitende schooltypen binnen een leerjaar. Deze kunnen dienen als vergelijkingsmateriaal om de effectgroottes in dit onderzoek nader te duiden.

In de Secties A.2 en A.3 staan de Figuren met de trendanalyse per vaardigheid voor subpopulaties op basis van geslacht of de achterstandsscore van de school.

In Sectie A.4 presenteren we de resultaten van additionele sensitiviteitsanalyses op basis van de gegevens van Toets 3. Voor de leerlingen die Toets 3 in de periode na de schoolsluitingen hebben afgenomen, zijn er namelijk toetsresultaten van Toets 1 beschikbaar die *voor* de Corona-periode zijn verzameld. Deze gegevens gebruiken we om een de vergelijkingsgroepen ook op vaardigheid zoals deze voor de Corona-periode is gemeten te matchen. In dit geval gebruiken we - naast de andere kenmerken die in de matchingsprocedure zijn gebruikt - dus als vergelijkingsgroepen leerlingen die dezelfde vaardigheidsscore op Toets 1 hadden in 2016/2017 en 2018/2019. We vergelijken dan de uitkomsten van deze leerlingen op Toets 3 in respectievelijk 2018/2019 en 2020/2021. Op deze manier hebben we een extra controle op de robuustheid van de uitkomsten zoals deze in Hoofdstuk 4 gepresenteerd zijn[1].

[1]Voor Toets 1 en Toets 2 is deze sterkere vorm van matching niet mogelijk, omdat de gegevens verzameld in 2019/2020 vanwege verlate afnames niet zijn opgenomen in dit onderzoek.

A.1 Effectgroottes 2018/2019 tussen aansluitende leerjaren of schooltypen

NLV	**Cohen's *d* in 2018/2019**		
Contrast:	**Toets 1 - Toets 2**	**Toets 2 - Toets 3**	***Gemiddelde***
vmbo bbkb	0.15	0.39	0.27
vmbo gt	0.26	0.40	0.33
havo	0.17	0.66	0.42
vwo	0.16	0.92	0.54
Gemiddelde	0.19	0.59	0.39

Tabel A.1: Effectgroottes van leerjaar op Nederlands leesvaardigheid

NLV	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 1	0.71	0.71	0.71	0.71
Toets 2	0.84	0.69	0.80	0.78
Toets 3	0.77	0.88	1.02	0.89
Gemiddelde	0.78	0.76	0.84	0.79

Tabel A.2: Effectgroottes van schooltype op Nederlands leesvaardigheid

NWS	**Cohen's *d* in 2018/2019**		
Contrast:	**Toets 1 - Toets 2**	**Toets 2 - Toets 3**	***Gemiddelde***
vmbo bbkb	0.40	0.51	0.46
vmbo gt	0.56	0.55	0.56
havo	0.64	0.70	0.67
vwo	0.54	0.49	0.52
Gemiddelde	0.53	0.57	0.55

Tabel A.3: Effectgroottes van leerjaar op Nederlands woordenschat

NWS	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 1	0.73	0.51	0.74	0.66
Toets 2	0.87	0.62	0.76	0.75
Toets 3	0.84	0.78	0.59	0.74
Gemiddelde	0.81	0.64	0.70	0.72

Tabel A.4: Effectgroottes van schooltype op Nederlands woordenschat

ELV	**Cohen's *d* in 2018/2019**		
Contrast:	**Toets 1 - Toets 2**	**Toets 2 - Toets 3**	***Gemiddelde***
vmbo bbkb	0.45	0.36	0.41
vmbo gt	0.49	0.42	0.45
havo	0.45	0.47	0.46
vwo	0.56	0.31	0.43
Gemiddelde	0.49	0.39	0.44

Tabel A.5: Effectgroottes van leerjaar op Engels leesvaardigheid

ELV	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 1	0.66	0.63	0.55	0.62
Toets 2	0.73	0.58	0.70	0.67
Toets 3	0.84	0.64	0.61	0.69
Gemiddelde	0.74	0.62	0.62	0.66

Tabel A.6: Effectgroottes van schooltype op Engels leesvaardigheid

EWS	**Cohen's *d* in 2018/2019**		
Contrast:	**Toets 1 - Toets 2**	**Toets 2 - Toets 3**	***Gemiddelde***
vmbo bbkb	0.47	0.49	0.48
vmbo gt	0.53	0.55	0.54
havo	0.50	0.65	0.58
vwo	0.52	0.60	0.56
Gemiddelde	0.50	0.57	0.54

Tabel A.7: Effectgroottes van leerjaar op Engels woordenschat

EWS	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 1	0.49	0.47	0.41	0.46
Toets 2	0.55	0.47	0.49	0.50
Toets 3	0.39	0.48	0.33	0.40
Gemiddelde	0.48	0.47	0.41	0.45

Tabel A.8: Effectgroottes van schooltype op Engels woordenschat

REK	**Cohen's *d* in 2018/2019**		
Contrast:	**Toets 1 - Toets 2**	**Toets 2 - Toets 3**	***Gemiddelde***
vmbo bbkb	0.47	0.55	0.51
vmbo gt	0.38	0.57	0.47
havo	0.28	0.59	0.43
vwo	0.53	0.96	0.75
Gemiddelde	0.41	0.67	0.54

Tabel A.9: Effectgroottes van leerjaar op Rekenen

REK	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 1	1.12	0.98	0.68	0.92
Toets 2	0.93	0.83	0.94	0.90
Toets 3	0.86	0.84	1.30	1.00
Gemiddelde	0.97	0.88	0.97	0.94

Tabel A.10: Effectgroottes van schooltype op Rekenen

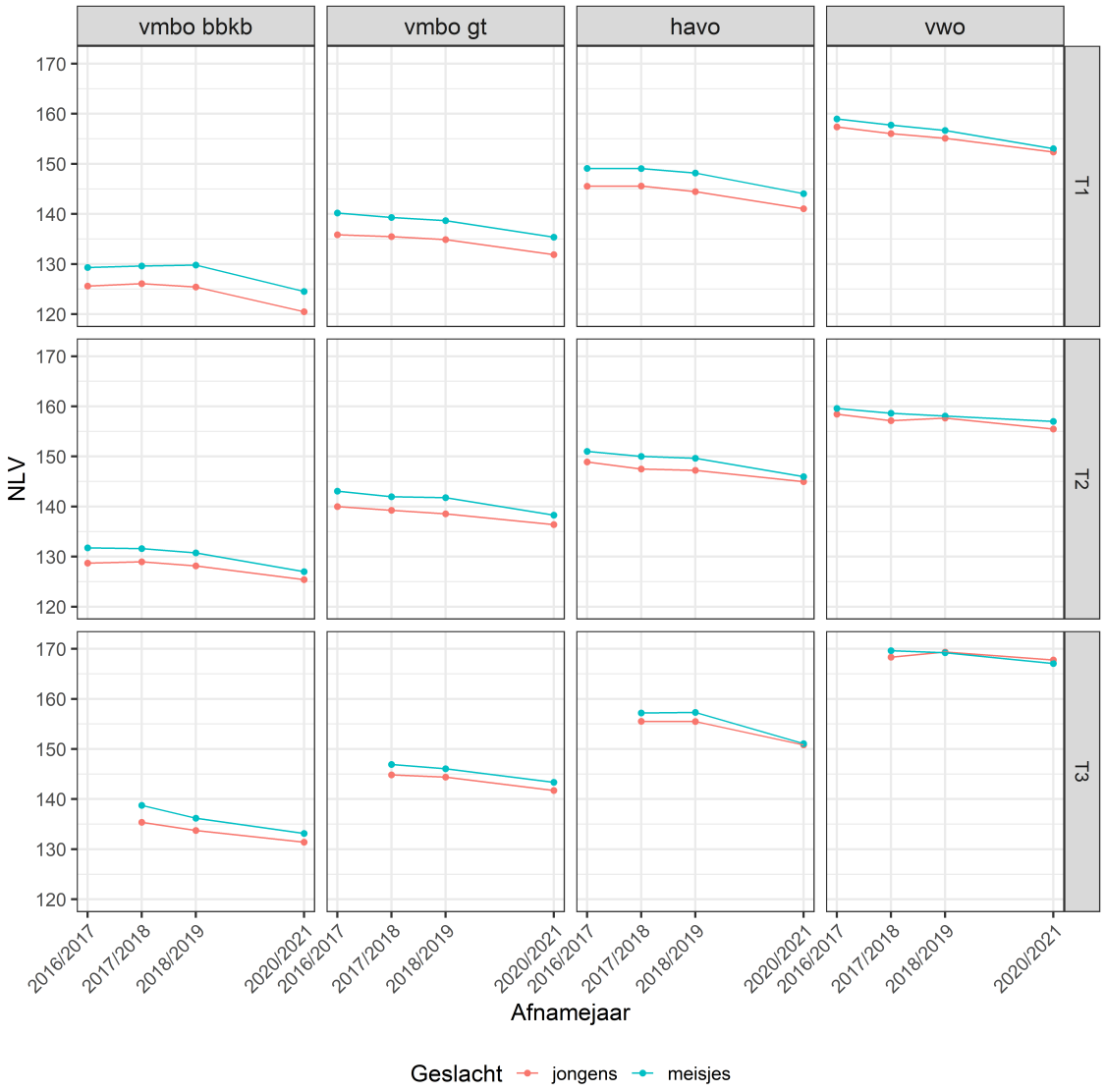
WIS	**Cohen's *d* in 2018/2019**
Contrast:	**Toets 2 - Toets 3**
vmbo bbkb	0.26
vmbo gt	0.54
havo	0.52
vwo	0.93
Gemiddelde	0.56

Tabel A.11: Effectgroottes van leerjaar op Wiskunde

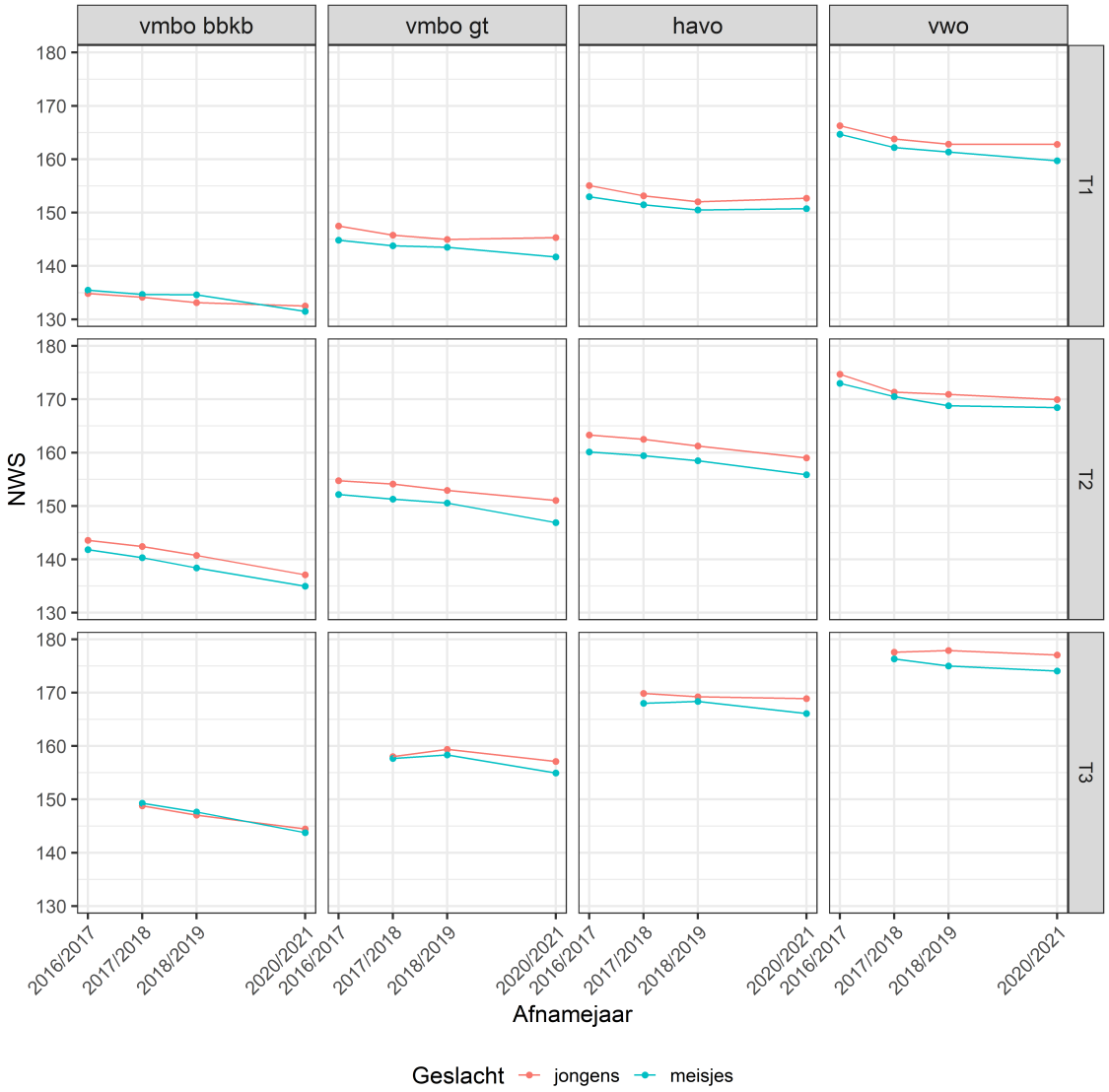
WIS	**Cohen's *d* in 2018/2019**			
Contrast:	**bbkb - gt**	**gt - havo**	**havo - vwo**	***Gemiddelde***
Toets 2	1.08	0.90	0.90	0.96
Toets 3	1.10	0.76	1.12	0.99
Gemiddelde	1.09	0.83	1.01	0.98

Tabel A.12: Effectgroottes van schooltype op Wiskunde

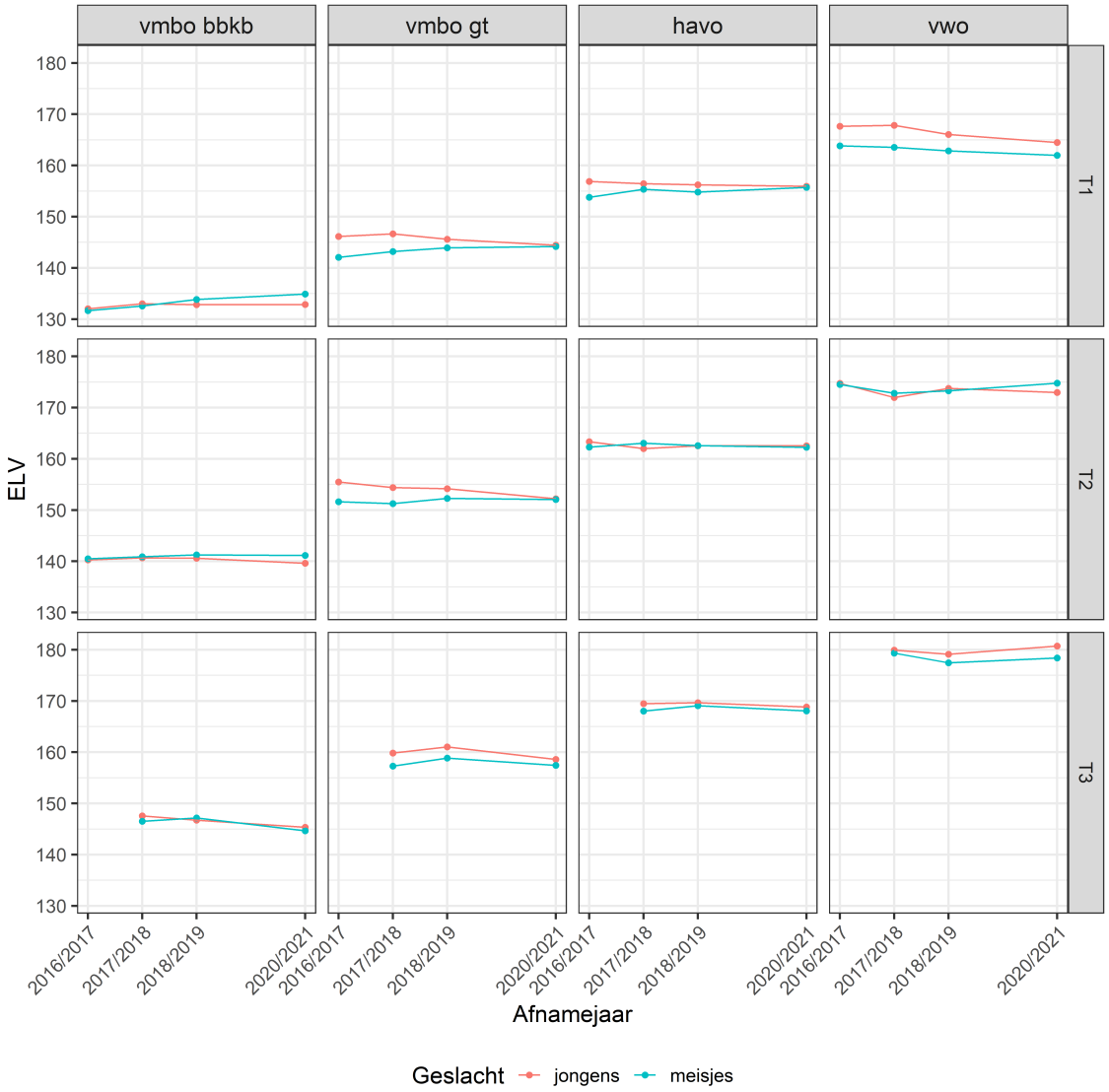
A.2 Figuren trendanalyse naar geslacht



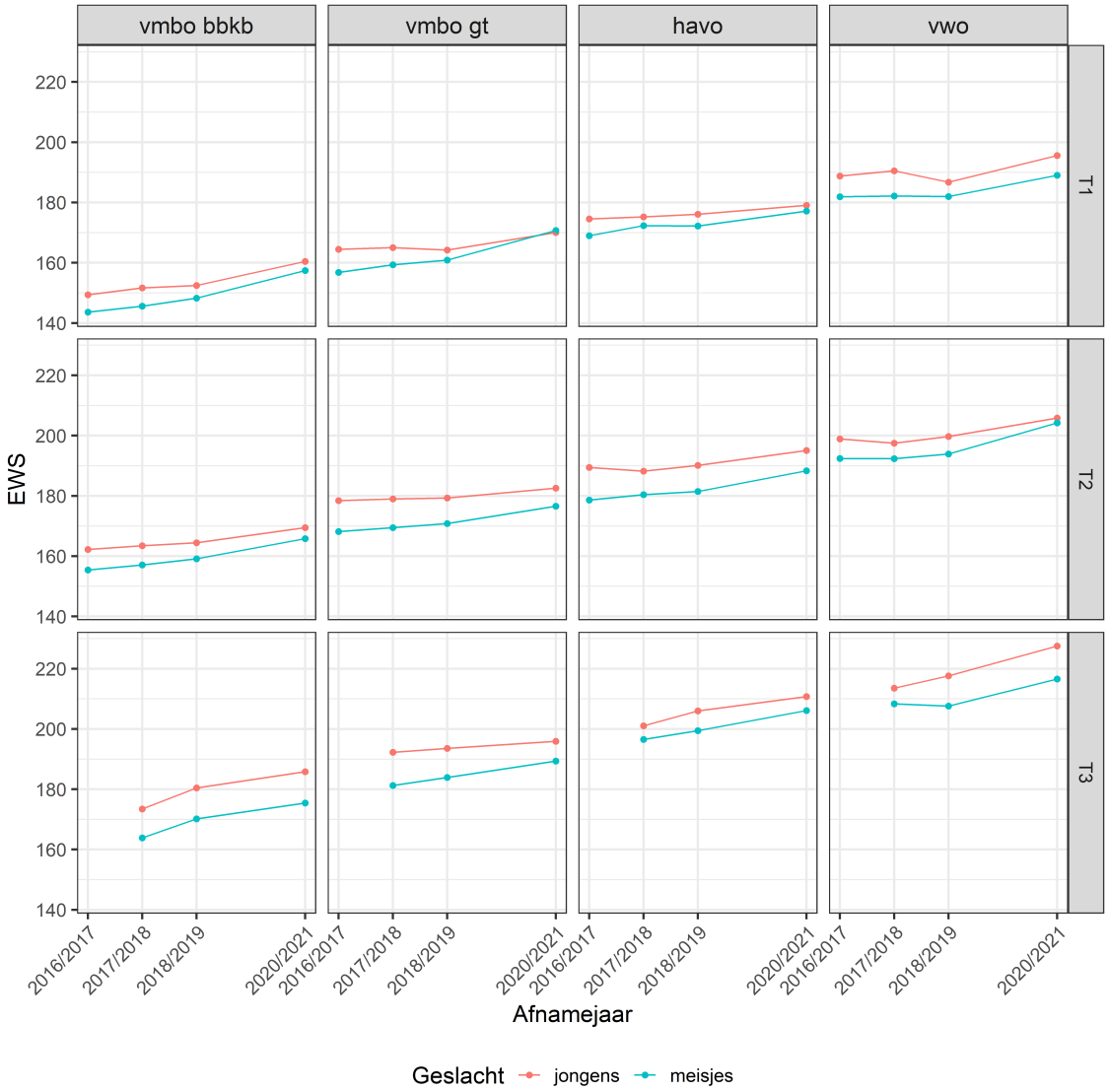
Figuur A.1: Trends op Nederlands leesvaardigheid naar geslacht



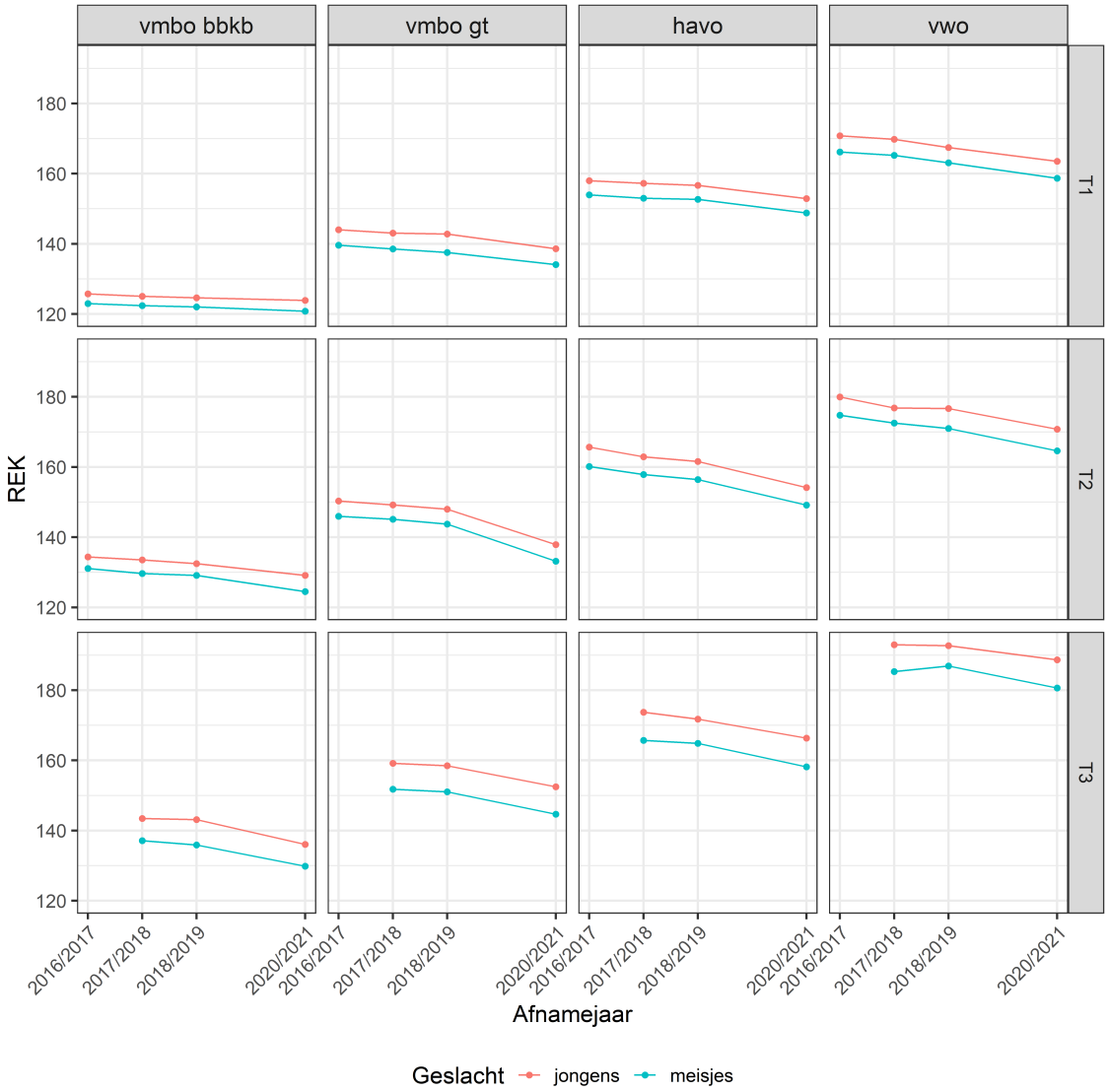
Figuur A.2: Trends op Nederlands woordenschat naar geslacht



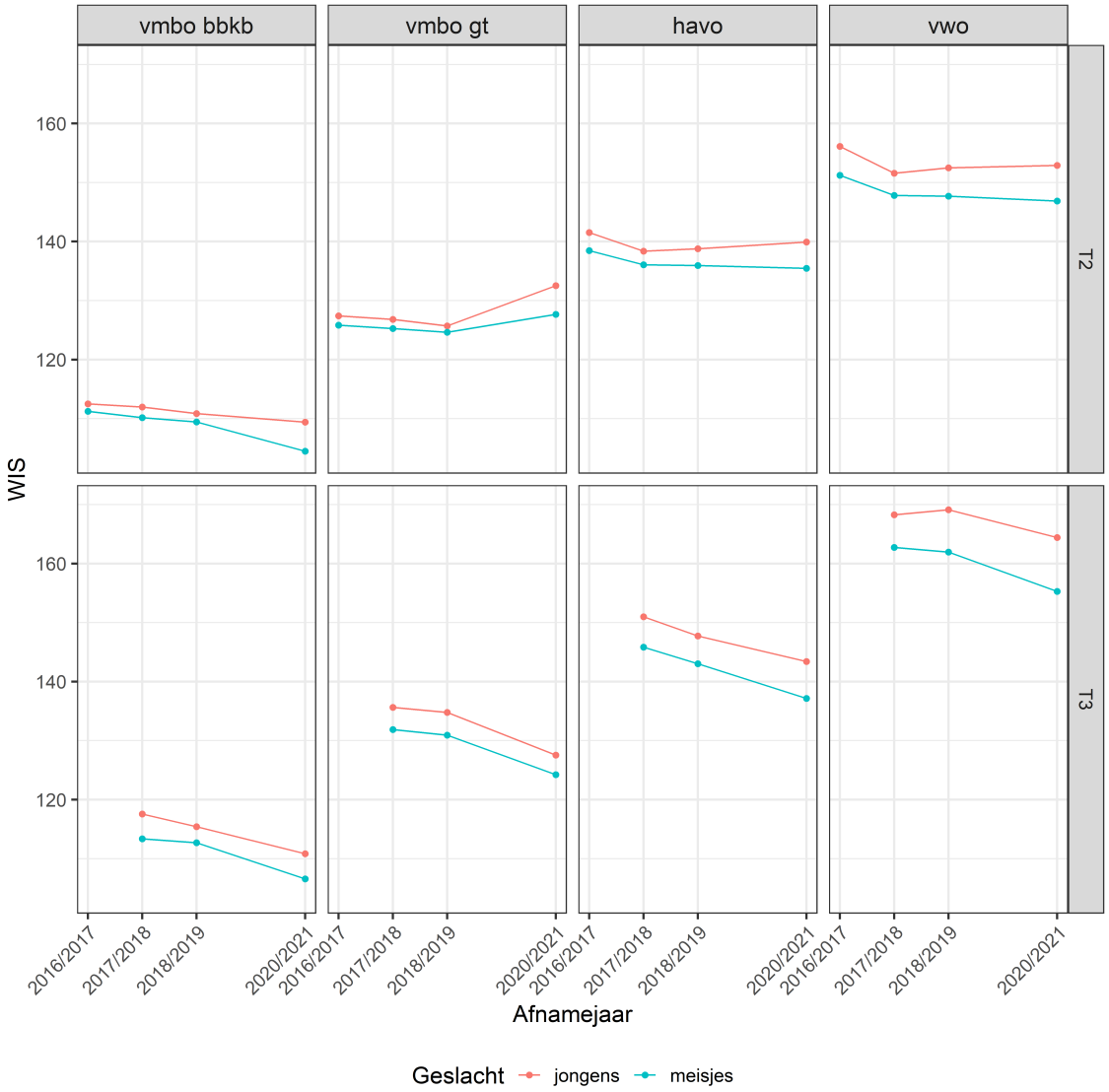
Figuur A.3: Trends op Engels leesvaardigheid naar geslacht



Figuur A.4: Trends op Engels woordenschat naar geslacht

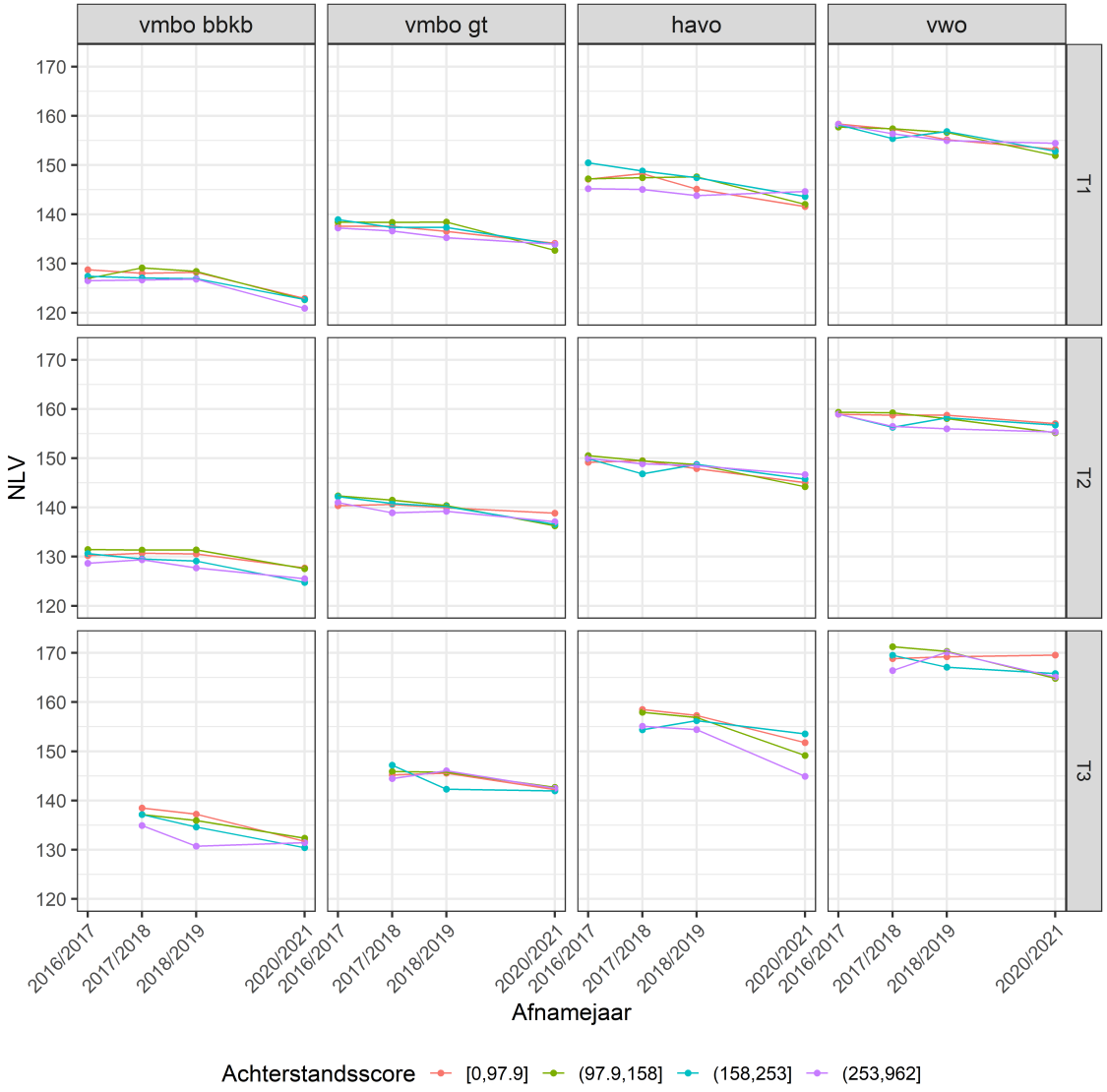


Figuur A.5: Trends op Rekenen naar geslacht

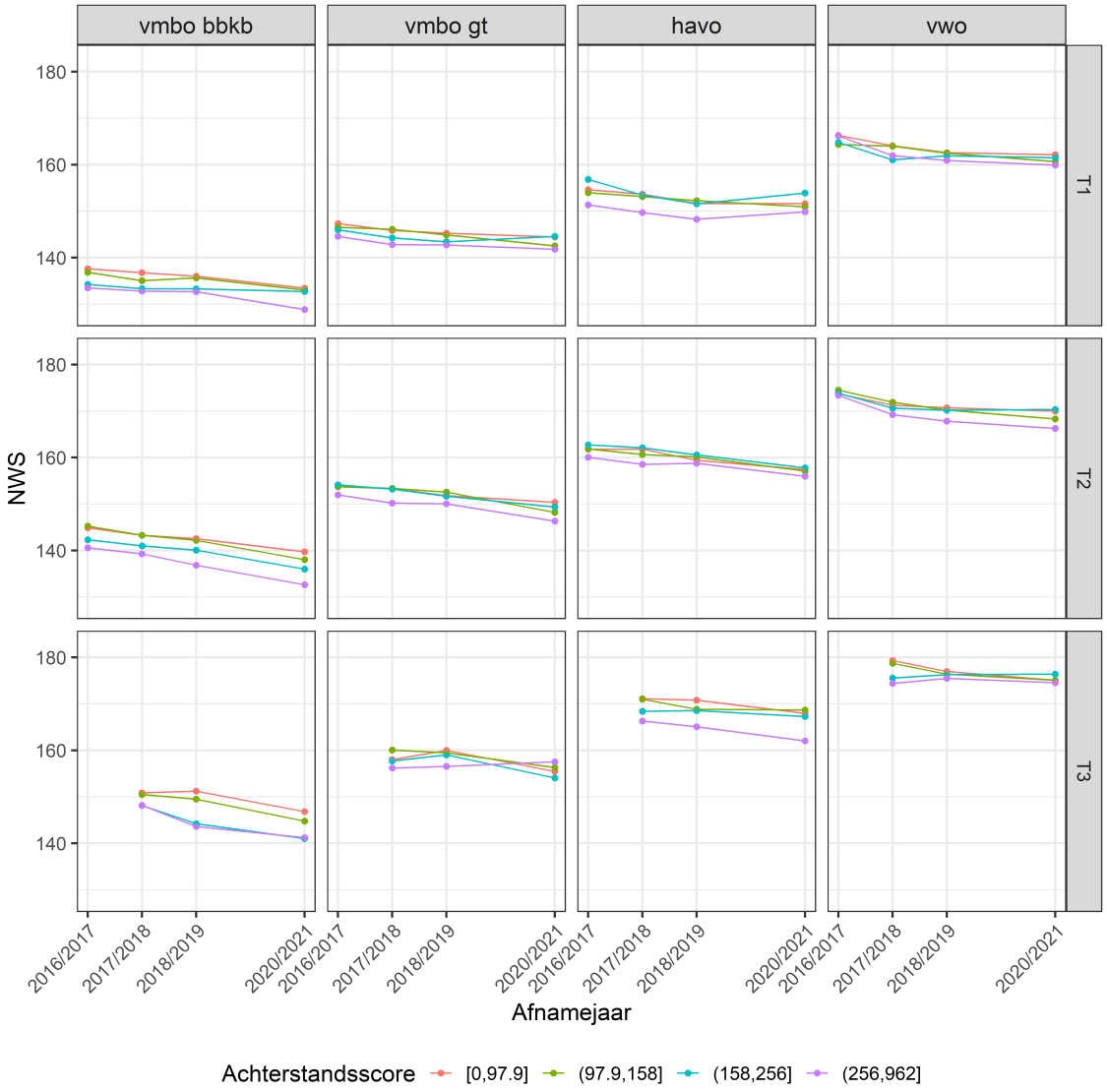


Figuur A.6: Trends op Wiskunde naar geslacht

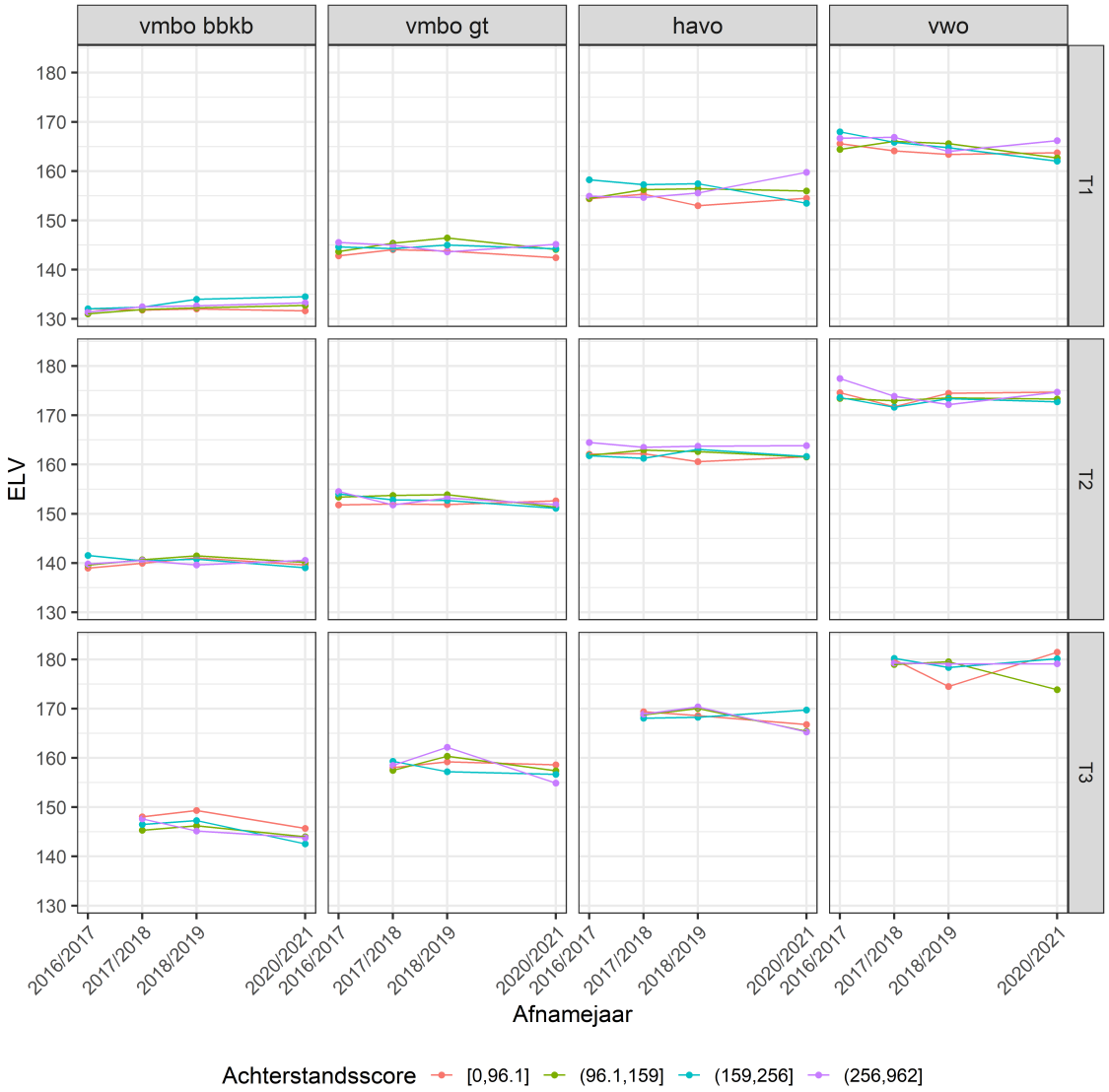
A.3 Figuren trendanalyse naar achterstandsscore van de school



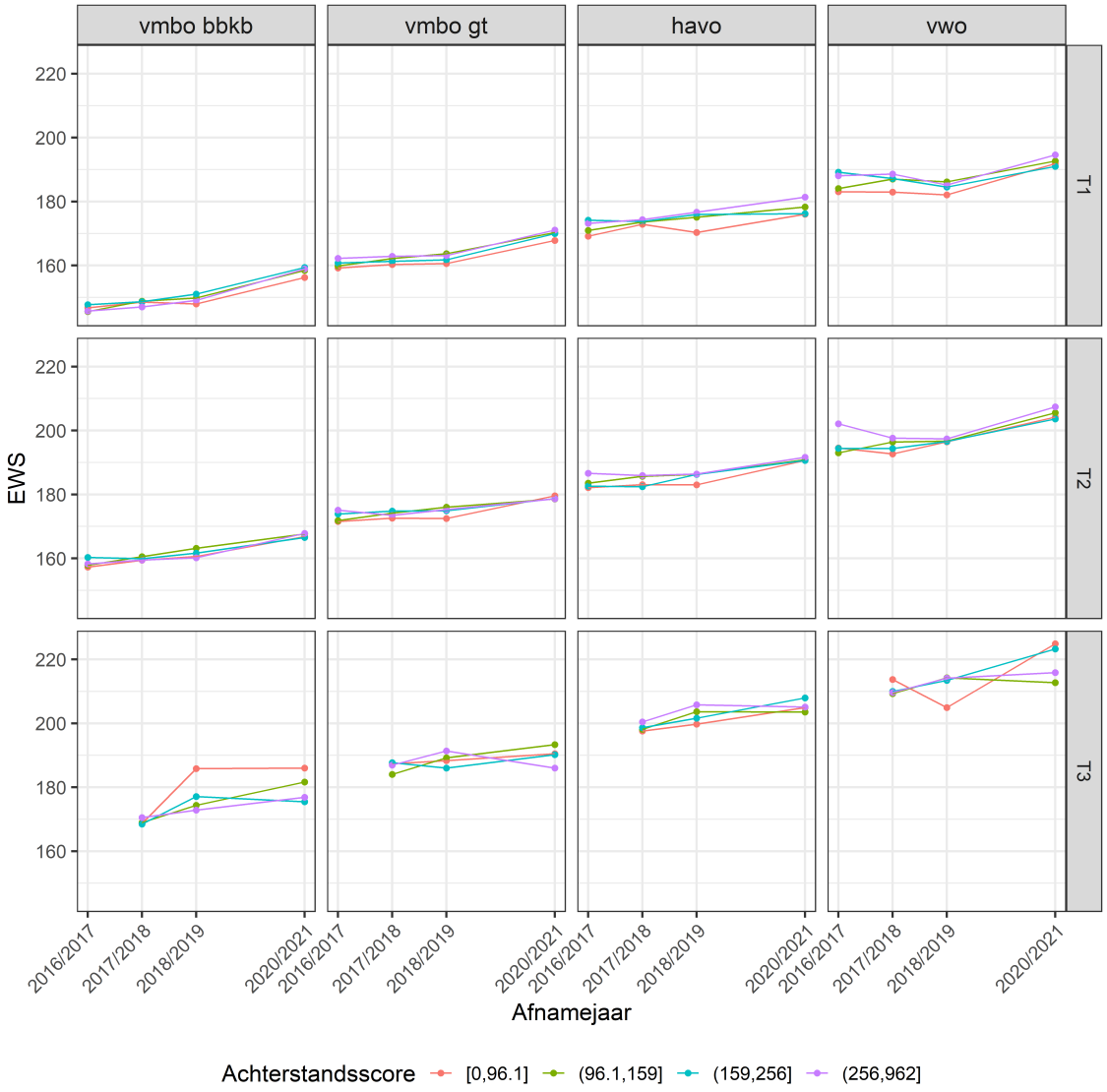
Figuur A.7: Trends op Nederlands leesvaardigheid naar achterstandsscore school



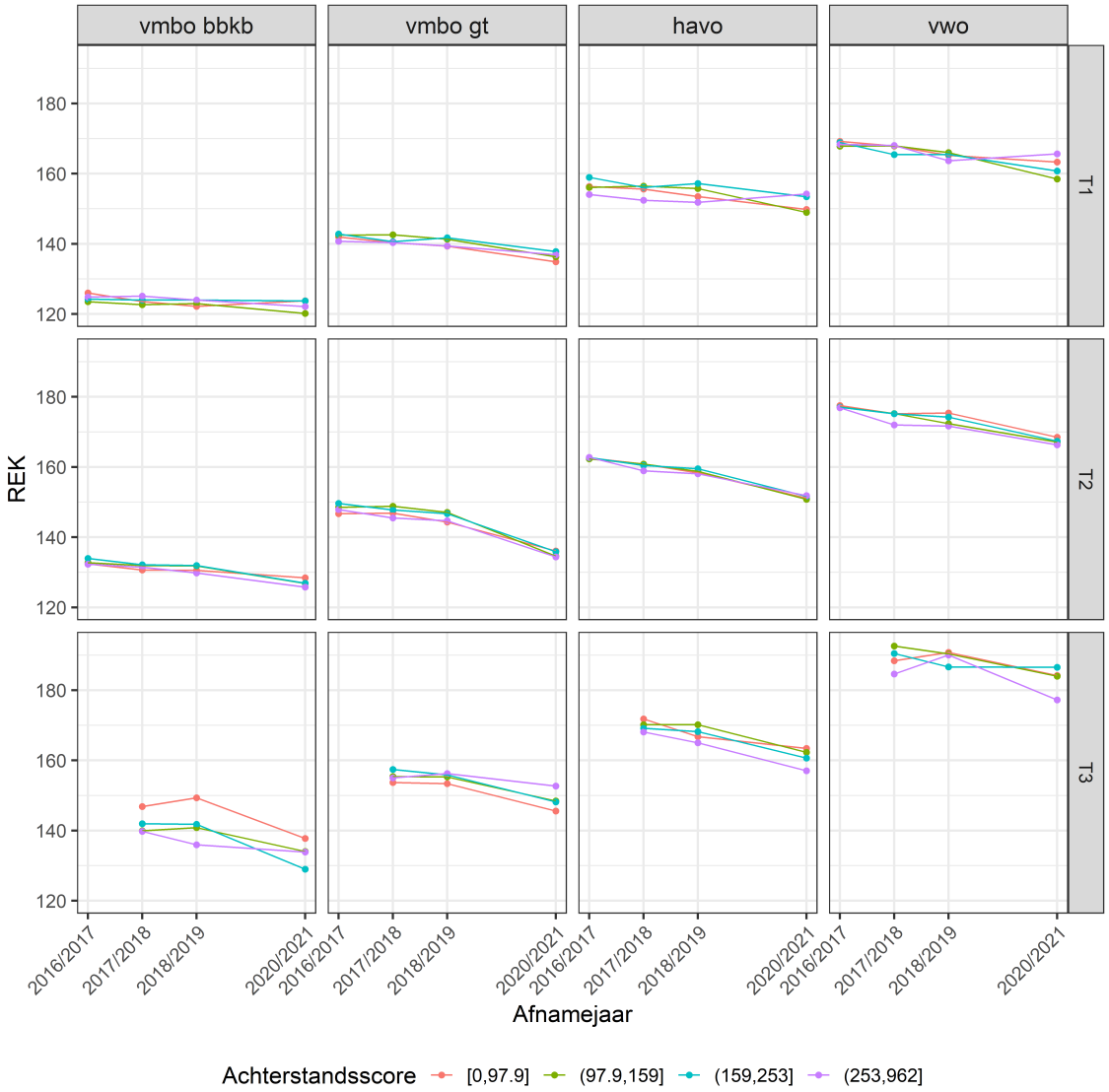
Figuur A.8: Trends op Nederlands woordenschat naar achterstandsscore school



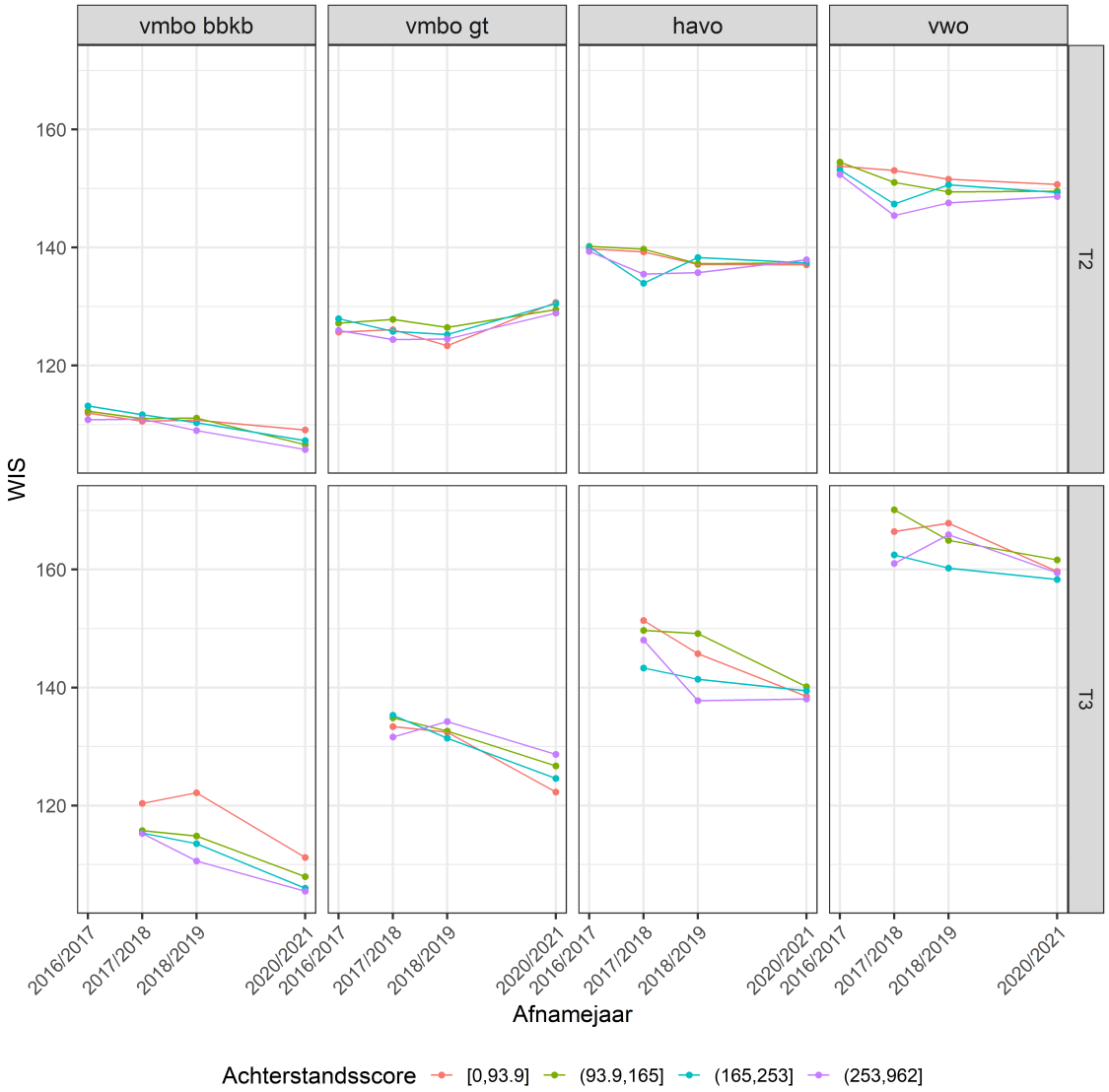
Figuur A.9: Trends op Engels leesvaardigheid naar achterstandsscore school



Figuur A.10: Trends op Engels woordenschat naar achterstandsscore school

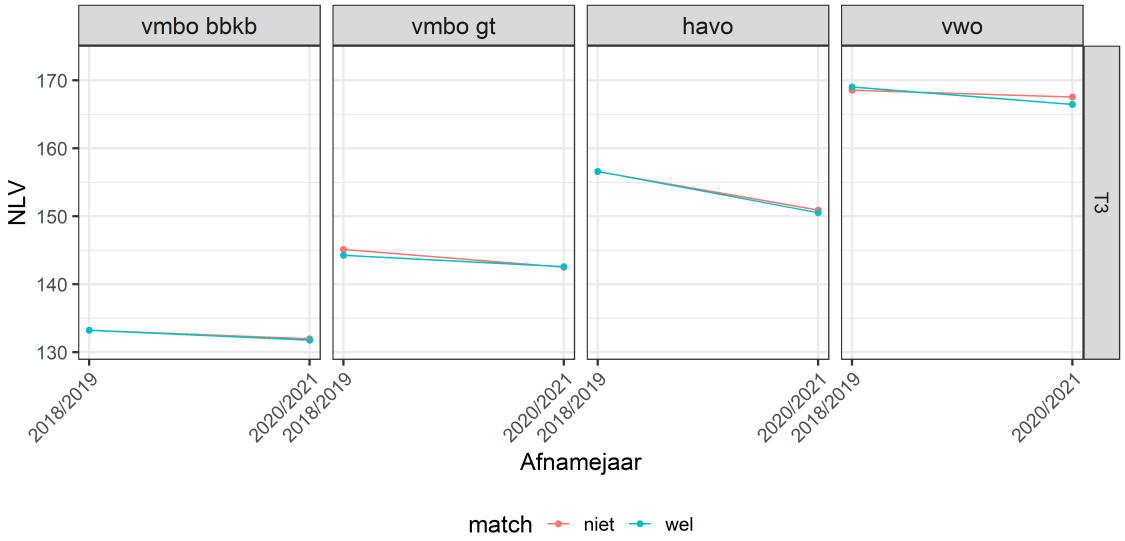


Figuur A.11: Trends op Rekenen naar achterstandsscore school

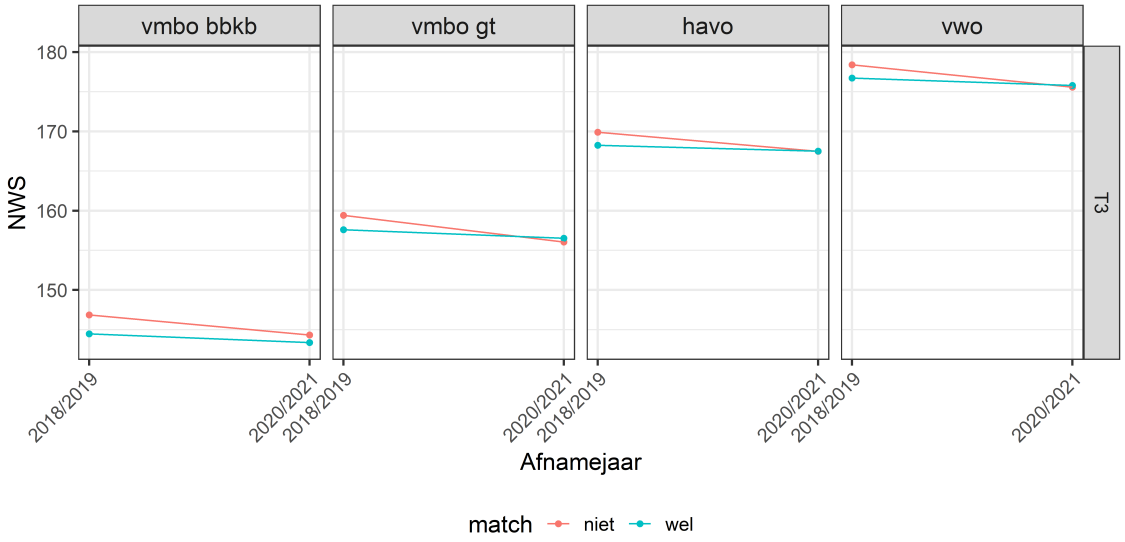


Figuur A.12: Trends op Wiskunde naar achterstandsscore school

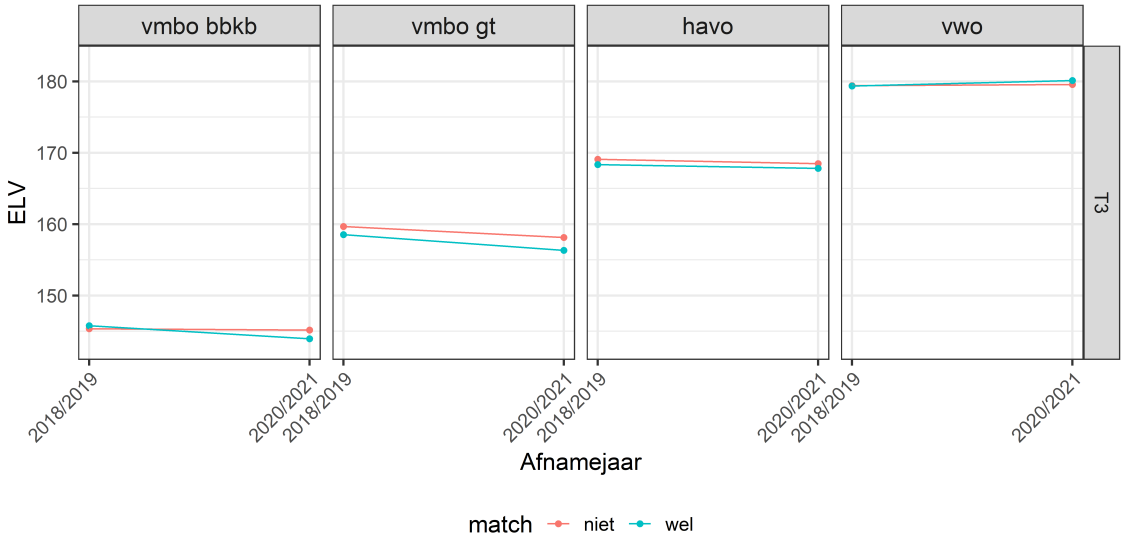
A.4 Sensitiviteitsanalyses



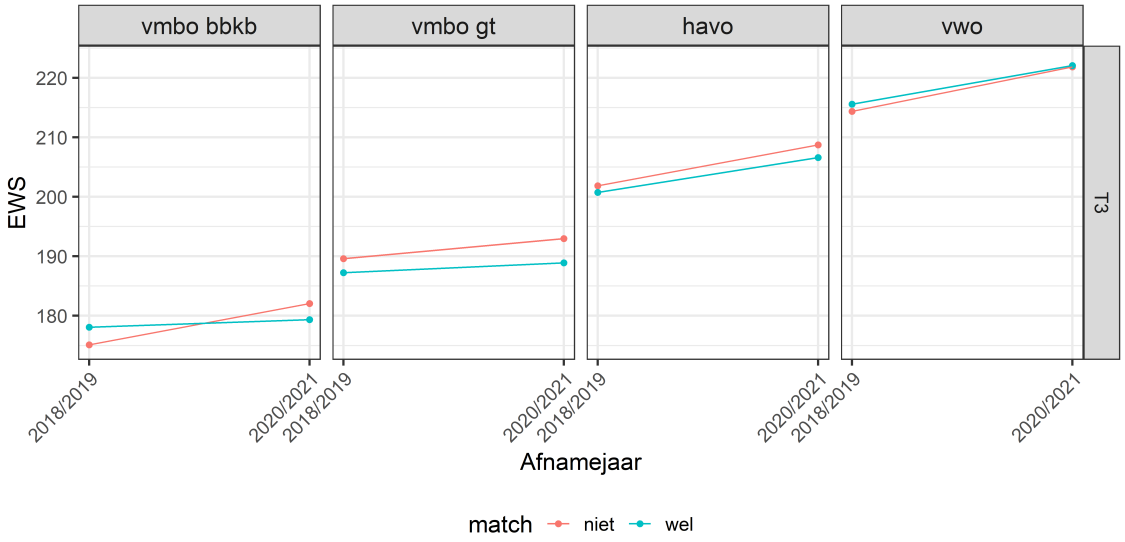
Figuur A.13: Sensitiviteitsanalyses Nederlands leesvaardigheid



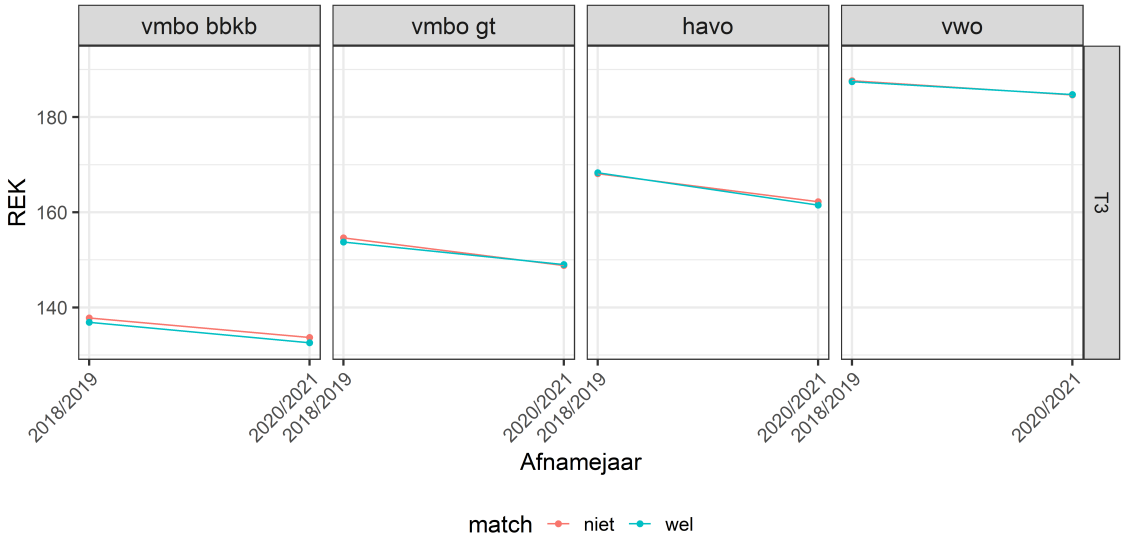
Figuur A.14: Sensitiviteitsanalyses Nederlands woordenschat



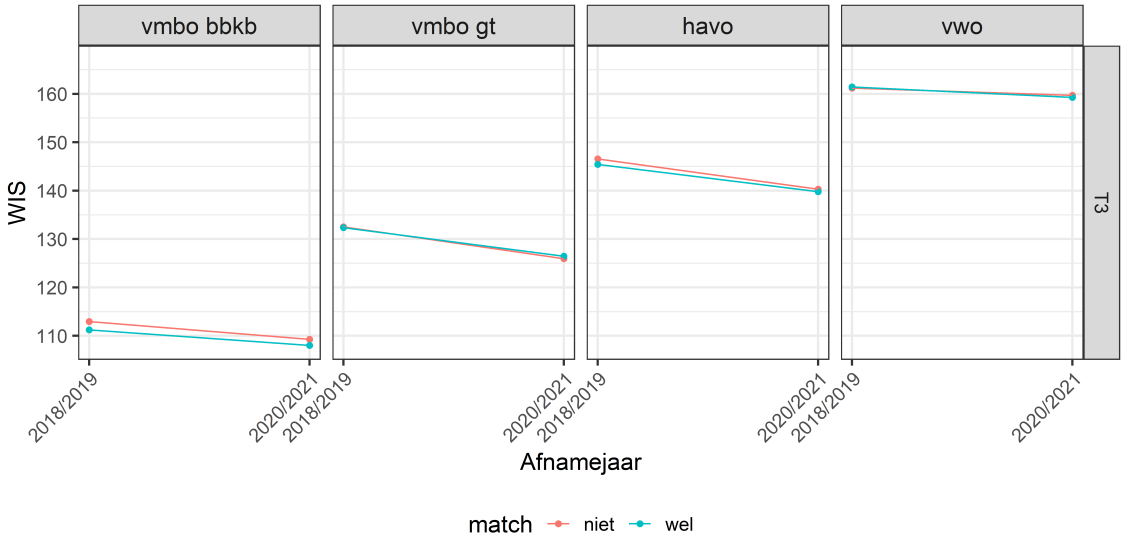
Figuur A.15: Sensitiviteitsanalyses Engels leesvaardigheid



Figuur A.16: Sensitiviteitsanalyses Engels woordenschat



Figuur A.17: Sensitiviteitsanalyses Rekenen



Figuur A.18: Sensitiviteitsanalyses Wiskunde

Lijst van figuren

1 Overzicht effectgroottes 4

2.1 Aantal deelnemende leerlingen 13
2.2 Aantal deelnemende scholen 14

3.1 Centraal Examen resultaten in de populatie en de responsgroep 19
3.2 Achterstandsscore in de populatie en de responsgroep 20
3.3 Stedelijkheid in de populatie en de responsgroep 21
3.4 Schoolgrootte in de populatie en de responsgroep 22
3.5 Stedelijkheid in de populatie en de responsgroep 23
3.6 Aantal leerlingen per schooltype in de populatie 23
3.7 Aantal leerlingen per schooltype in de responsgroep 24
3.8 Voortgezet onderwijsscholen in Nederland 25
3.9 Voortgezet onderwijsscholen in de respons groep 25

4.1 Gemiddelde vaardigheidsscores Nederlands leesvaardigheid 30
4.2 Verdeling over referentieniveaus Nederlands leesvaardigheid 32
4.3 Gemiddelde vaardigheidsscores Nederlands woordenschat 33
4.4 Gemiddelde vaardigheidsscores Engels leesvaardigheid 34
4.5 Verdeling over ERK-niveaus Engels leesvaardigheid 35
4.6 Gemiddelde vaardigheidsscores Engels woordenschat 36
4.7 Gemiddelde vaardigheidsscores Rekenen 38
4.8 Verdeling over referentieniveaus Rekenen 39
4.9 Gemiddelde vaardigheidsscores Wiskunde 40
4.10 Spreiding leeruitkomsten Nederlands leesvaardigheid 44
4.11 Spreiding leeruitkomsten Nederlands woordenschat 44
4.12 Spreiding leeruitkomsten Engels leesvaardigheid 45
4.13 Spreiding leeruitkomsten Engels woordenschat 46
4.14 Spreiding leeruitkomsten Rekenen 46
4.15 Spreiding leeruitkomsten Wiskunde 47

A.1 Trends op Nederlands leesvaardigheid naar geslacht 59
A.2 Trends op Nederlands woordenschat naar geslacht 60
A.3 Trends op Engels leesvaardigheid naar geslacht 61
A.4 Trends op Engels woordenschat naar geslacht 62
A.5 Trends op Rekenen naar geslacht 63
A.6 Trends op Wiskunde naar geslacht 64
A.7 Trends op Nederlands leesvaardigheid naar achterstandsscore school . 65
A.8 Trends op Nederlands woordenschat naar achterstandsscore school . . 66

A.9 Trends op Engels leesvaardigheid naar achterstandsscore school 67
A.10 Trends op Engels woordenschat naar achterstandsscore school 68
A.11 Trends op Rekenen naar achterstandsscore school 69
A.12 Trends op Wiskunde naar achterstandsscore school 70
A.13 Sensitiviteitsanalyses Nederlands leesvaardigheid 71
A.14 Sensitiviteitsanalyses Nederlands woordenschat 71
A.15 Sensitiviteitsanalyses Engels leesvaardigheid 72
A.16 Sensitiviteitsanalyses Engels woordenschat 72
A.17 Sensitiviteitsanalyses Rekenen 73
A.18 Sensitiviteitsanalyses Wiskunde 73

Lijst van tabellen

3.1 Gemiddelde effectgroottes tussen aansluitende leerjaren (T1-T2 of T2-T3) of schooltypen (bv. vmbo gt-havo) in 2018/2019 17
3.2 R-Indicator waarden 26

4.1 Effecten op Nederlands leesvaardigheid (o.b.v. gematchte groepen) . . 31
4.2 Effecten op Nederlands woordenschat (o.b.v. gematchte groepen) . . . 33
4.3 Effecten op Engels leesvaardigheid (o.b.v. gematchte groepen) 35
4.4 Effecten op Engels woordenschat (o.b.v. gematchte groepen) 37
4.5 Effecten op Rekenen (o.b.v. gematchte groepen) 38
4.6 Effecten op Wiskunde (o.b.v. gematchte groepen) 41

A.1 Effectgroottes van leerjaar op Nederlands leesvaardigheid 56
A.2 Effectgroottes van schooltype op Nederlands leesvaardigheid 56
A.3 Effectgroottes van leerjaar op Nederlands woordenschat 56
A.4 Effectgroottes van schooltype op Nederlands woordenschat 56
A.5 Effectgroottes van leerjaar op Engels leesvaardigheid 57
A.6 Effectgroottes van schooltype op Engels leesvaardigheid 57
A.7 Effectgroottes van leerjaar op Engels woordenschat 57
A.8 Effectgroottes van schooltype op Engels woordenschat 57
A.9 Effectgroottes van leerjaar op Rekenen 58
A.10 Effectgroottes van schooltype op Rekenen 58
A.11 Effectgroottes van leerjaar op Wiskunde 58
A.12 Effectgroottes van schooltype op Wiskunde 58