Wat was de impact van 1 jaar corona op de kernvakken in het primair onderwijs?

Een statistische analyse van toetsscores voor spelling, rekenen en begrijpend lezen

LEERUNIEK

Juni 2021

Inhoud

Executive Summary	**3**
Inleiding	**5**
Werkwijze en onderzoeksresultaten	**8**
Conclusie en aanbevelingen	**20**
Reflecties tot besluit	**21**
Colofon	**22**

1. Executive Summary

1.1 Context: corona leidde tot leerachterstanden toonde eerdere onderzoeken aan
Het ministerie van OCW heeft toegezegd de Tweede Kamer te informeren over de impact van corona op de leeropbrengsten in het basisonderwijs tot nu toe (mei 2021). Hiertoe gaf OCW aan Leeruniek de opdracht om onderzoek te doen in samenwerking met wetenschappers verbonden aan Oxford. Dit onderzoek bouwt voort op een studie van dezelfde wetenschappers naar de impact van de 1e lockdown, gepubliceerd in het internationale tijdschrift PNAS. Geconstateerd werd dat leerlingen toen, na de 1e lockdown, gemiddeld 3 percentielpunten lager scoorden dan verwacht op halfjaarlijkse Cito-volgtoetsen (rekenen, spelling, taal / begrijpend lezen). Dit stond gelijk aan een verlies van reguliere leergroei van 7-8 lesweken. Velen vertaalden dit in de 'corona-achterstanden'. Dit huidige onderzoek hanteert de term 'leervertraging'.

1.2 Vraag: hoe staat het 1 jaar na het begin van corona met leervertraging en hoe verschilt dit voor subpopulaties?
De nu gestelde onderzoeksvraag van OCW is tweeledig geformuleerd:

1. Wat is de gemiddelde cumulatieve impact van corona op leeropbrengsten tot en met de middentoetsen van schooljaar 2020-2021, 1 jaar na aanvang van de coronacrisis?
2. In hoeverre verschilt deze impact voor verschillende subpopulaties?

Onderzoekers is verzocht de vragen statistisch te beantwoorden. Een onderwijskundige duiding van de resultaten valt buiten de scope van deze onderzoeksopdracht.

1.3 Bevindingen hoofdvraag 1: vertragingen zoals gemeten in zomer 2020 lijken nu - voorjaar 2021 - deels te zijn ingelopen
Het antwoord op de eerste onderzoeksvraag is hoopgevend: de vertragingen van de 1e lockdown lijken deels te zijn ingelopen. De nu gemeten leervertraging ligt waarschijnlijk niet meer op 3 percentielpunten, maar eerder op 1-2 percentielpunten. Dit zou betekenen dat de eerder geduide vertraging van 7-8 weken gehalveerd is.

1.4 Echter, de gemiddeld 6 weken latere toetsing als gevolg van 2e lockdown, bemoeilijkt het trekken van harde conclusies
Er is echter een belangrijke complicerende factor die de duiding van leervertraging over de gehele periode bemoeilijkt: de Cito M-toetsen zijn dit jaar, ten gevolge van de 2e lockdown, gemiddeld 6 weken later afgenomen. Dit komt neer op een verlenging van de eerste helft van het schooljaar met zo'n 30%. Dit maakte een "schone" appels met appels-vergelijking nu niet mogelijk. Analyse van de Cito E-toetsen straks moet meer duidelijkheid bieden. De bevinding dat cumulatieve leervertraging rond de 1-2 percentielpunten ligt, is gemaakt op basis van schattingen die de impact van deze verlate afname van de M-toetsen in 2021 meeneemt.

1.5 Bevindingen hoofdvraag 2: net zoals na de 1e lockdown, verschilt ook nu de impact van corona per subpopulatie
Beantwoording van de tweede onderzoeksvraag laat het volgende zien: verschillen tussen leerlingen met laagopgeleide ouders, vergeleken met de rest van de leerlingenpopulatie, zijn

nog steeds groot (zo'n 55%) maar namen niet verder toe sinds zomer 2020.[1] Verder lijken verschillen tussen scholen met een gelijke schoolweging ook stabiel te zijn gebleven. Een schoolweging[2] (schaal 20-40) staat voor de complexiteit van de leerlingpopulatie op een school. Een hoge weging betekent hoge complexiteit. Tot slot blijven toetsscores van begrijpend lezen achter bij die van spelling en rekenen. Het is onduidelijk of dit komt doordat de verschuiving van de M-toetsen in 2021 een andere effect heeft op toetsresultaten voor begrijpend lezen, of doordat leervertraging in lezen daadwerkelijk nog sterker aanwezig is over de gehele coronaperiode tot nu toe.

1.6 Vervolg: afname E-toetsen op het reguliere moment is essentieel voor definitieve analyse van de impact van corona

Analyse van de E-toetsen van juni 2021 zal moeten uitwijzen wat de impact is geweest van de dan anderhalf jaar durende coronacrisis. Onderzoekers bevelen scholen ten zeerste aan om de E-toetsen sowieso af te nemen, en dat te doen in de reguliere periode. Dit maakt vergelijkbaarheid met eerdere jaren mogelijk, zodat conclusies kunnen worden getrokken over de status van leervertraging (of extra groei) aan het einde van dit schooljaar.

1.7 Ook het NPO en de herstelperiode vragen om continuïteit in analyse en onderwijskundige duiding van volgtoetsen

Analyse van E-toetsen lijkt ook essentieel om te kunnen komen tot een op feiten gebaseerde NPO-schoolscan. Deze vormt weer de basis van een schoolprogramma, en de monitoring daarvan, gedurende de 2½ jaar looptijd van het Nationaal Programma Onderwijs (NPO).

1.8 Werkwijze onderzoekers

Er is naar gestreefd om conclusies te trekken waar mogelijk, maar dat vooral ook *niet* te doen waar dat (nog) niet kan. De precieze methodiek en de werkwijze omtrent geanonimiseerde en geaggregeerde toetsscores vindt u op de website van Leeruniek.

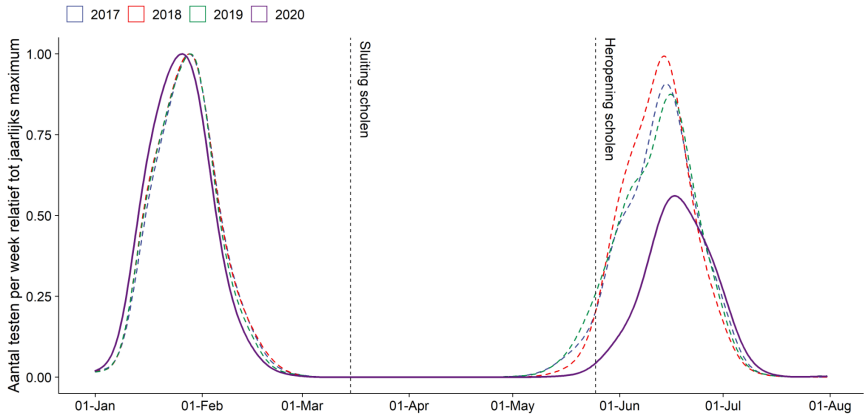
[1] Laagopgeleid is gemeten op basis van de gewichtenregeling (zie de Technische Appendix voor een uitgebreide beschrijving).

[2] https://www.onderwijsinspectie.nl/onderwerpen/onderwijsresultaten-primair-onderwijs/naar-een-nieuw-onderwijsresultatenmodel/de-schoolweging-een-nieuwe-maat-voor-de-leerlingenpopulatie

2. Inleiding

2.1 Onderzoeken: de 1e lockdown ging gepaard met leervertraging

De 1e coronalockdown van voorjaar 2020 liet een consistent beeld zien: uit meerdere onderzoeken bleek dat basisschoolleerlingen op de halfjaarlijkse volgtoetsen van Cito significant lager scoorden dan in reguliere jaren. De toevallige omstandigheid dat de 1e lockdown precies viel tussen de middentoetsen (M-toetsen) van begin februari 2020 en de eindtoetsen (E-toetsen) van medio juni 2020, zorgde voor de mogelijkheid om de schoolsluiting te onderzoeken. Dit maakte het voor onderzoekers mogelijk om een direct verband te leggen tussen de schoolsluiting en de lagere scores op de kernvakken rekenen, spelling en begrijpend lezen. (Zie Figuur 1 voor het moment van de eerste schoolsluiting relatief tot de toetsmomenten.)



Figuur 1: Toetsafnames net voor en net na de eerste schoolsluiting

Zoals verwacht vielen alle toetsen voorafgaand aan de schoolsluiting (M-toetsen) in 2020 rond dezelfde datum als in de jaren daarvoor: tussen medio januari en medio februari. Ook vielen de toetsen na de schoolsluiting (E-toetsen) grotendeels gelijk met voorgaande jaren (juni). Hierdoor konden toetsscores van leerlingen van voor en na de schoolsluiting vergeleken worden in het coronajaar met die van 'normale' jaren. Let wel (paarse lijn) dat er minder toetsen afgenomen zijn na de schoolsluiting dan in een regulier jaar. Hierdoor is er een (methodologisch) risico dat er selectieve afname heeft plaatsgevonden. In de statistische analyses, onderliggend aan het onderzoek naar de eerste schoolsluiting, is hier uitvoerig op gecontroleerd.

2.2 Oxford-studie Nederlandse situatie trekt internationaal aandacht: ongelijkheid nam toe

Drie onderzoekers van de universiteit Oxford onderzochten de impact van de 1e lockdown en deelden als eersten hun bevindingen. Hun publicatie in het internationale tijdschrift PNAS is breed geciteerd door andere wetenschappers en aangehaald in de (inter-)nationale pers. De eerste schoolsluiting had tot gemiddeld 7-8 weken aan leervertraging geleid. Wat verder vooral opviel was dat leerlingen uit laagopgeleide gezinnen meer dan 50% extra vertraging opliepen. Corona leidde dus niet alleen tot gemiddelde vertraging voor alle leerlingen, maar raakte 'kwetsbare' groepen ook nog eens extra hard. Anders gezegd: door corona bleek kansenongelijkheid toe te nemen.

2.3 Corona tekent heel schooljaar 2020-2021 en vraagt veel van scholen en gezinnen

Vanaf dag één is schooljaar 2020-2021 onrustig: reguliere lesweken worden onderbroken door periodes van afstandsonderwijs als gevolg van (gezins-) quarantaines of lokale lockdowns. Met grote inzet, improvisatievermogen en flexibele (hybride) werkvormen weten scholen het onderwijs draaiende te houden, hierin gesteund door ouders, besturen en de leerlingen zelf. Met extra focus op welbevinden en op de kernvakken taal en rekenen beoogt men vertraging in te lopen, ook geholpen door extra OCW-financiering (inhaal- en ondersteuningsprogramma's).

2.4 De 2e lockdown overvalt velen. Men is beter toegerust, maar de 10 weken zijn lang

Medio december 2020 gaan scholen vrij onverwacht weer dicht. Deze 2e lockdown zou tot 8 februari duren, maar door hevige sneeuwval en voorjaarsvakantie gaan veel scholen pas weer open in de week van 22 februari en 1 maart. De tweede lockdown is anders dan de eerste: scholen zijn nu veel beter voorbereid op afstandsonderwijs, maar zijn ook genoodzaakt om op grote schaal noodopvang te verzorgen. Verder speelt vermoeidheid mogelijk ook een rol, net als de grote sociale en economische druk die corona legt op de rest van de samenleving.

2.5 Ministerie van OCW kondigt het Nationaal Programma Onderwijs (NPO) aan

Medio februari 2021 kondigt Minister Slob het Nationaal Programma Onderwijs (NPO) aan. Dit is bedoeld om mogelijke 'leervertragingen en andere problemen door corona' aan te pakken.[3]

Op basis van een analyse en duiding ('schoolscan') van de onderwijskundige behoeften, kunnen scholen kiezen uit een 'menukaart' met kansrijke interventies en deze vertalen in een 'schoolprogramma'. Na instemming van de medezeggenschapsraad en in afstemming met belanghebbenden, zoals de gemeente, komt dit schoolprogramma in aanmerking voor financiering van het NPO. Verdere details van het NPO worden nog bekend gemaakt.

[3] https://www.nponderwijs.nl/, https://www.nponderwijs.nl/po-en-vo/documenten/publicaties/2021/03/23/brief-aan-scholen-over-nationaal-programma-onderwijs

2.6 Vraag OCW aan Leeruniek en onderzoekers Oxford: wat is de impact van corona tot nu toe op de scores van kernvakken?

Het ministerie van OCW heeft toegezegd om de Tweede Kamer vóór zomer 2021 te informeren over de impact van corona op leerontwikkeling in het primair onderwijs. Ter voorbereiding gaf OCW de opdracht aan Leeruniek om hier onderzoek naar te doen, voortbouwend op de eerdere studie van het team van Oxford. De vraagstelling was tweeledig:

1. Wat is de gemiddelde cumulatieve impact van corona op leeropbrengsten tot en met de middentoetsen van schooljaar 2020-2021, 1 jaar na aanvang van de coronacrisis?
2. In hoeverre verschilt deze impact?
    - Afhankelijk van opleidingsniveau ouders
    - Voor meisjes vs. jongens
    - Per vak (begrijpend lezen, spelling, rekenen)
    - Per leerjaar, kijkend naar de huidige groepen 5, 6, en 7
    - Per schoolweging, een indicatie voor de ‘complexiteit’ van de schoolpopulatie. Deze loopt van 20 (geringe complexiteit) tot 40 (hoge complexiteit)

2.7 Doel: statistisch onderzoek voor het objectiveren van het debat en ten behoeve van latere onderwijskundige duiding

De opdracht was om een beschrijvend onderzoek te doen, waarvan bevindingen en analyses ondersteunend zijn aan het publieke debat. Het geven van een complete onderwijskundige duiding valt buiten de scope van de onderzoeksopdracht. Wat er ook buiten valt is onderzoek naar de mechanismen die de impact van corona op toetsresultaten kunnen verklaren, of onderzoek naar specifieke oplossingen voor het inlopen van leervertraging.

2.8 Leeswijzer voor dit rapport

Dit rapport is geschreven voor de ‘deskundige leek’, uitgaande van een aanwezig basiskennis van onderwijskundige en statistische begrippen. Hoofdstuk 3 beschrijft de werkwijze, de methodische uitdagingen en de onderzoeksresultaten. Hoofdstuk 4 verwoordt de conclusies en aanbevelingen. Hoofdstuk 5 bevat enkele extra reflecties en hoofdstuk 6 de colofon.

3. Werkwijze en onderzoeksresultaten

In deze sectie worden zowel de methodologie die gebruikt is om leervertraging te berekenen als de resultaten, besproken. Deze sectie is voor een niet-technisch publiek bestemd. Voor een uitvoerige beschrijving van de gebruikte methodologie wordt de lezer verwezen naar de Technische Appendix (in het engels). De appendix kunt u vinden via deze link[4].

3.1 Basisscholen volgen de ontwikkeling op kernvakken via halfjaarlijkse volgtoetsen

Basisscholen nemen tweemaal per jaar toetsen af voor onder meer de vakken rekenen, spelling en lezen. Deze toetsen zijn door Cito[5] ontwikkeld en 'genormeerd' in percentielscores. Een score van 45 betekent dat de leerling presteert op het 45e percentiel. Anders gezegd: 45% van de leerlingen heeft eenzelfde score of lager behaald, terwijl 55% van de leerlingen juist een hogere score heeft behaald dan deze leerling.

3.2 Vaste normering van volgtoetsen maakt het mogelijk om de impact van corona te meten

Door de laatste toets pre-corona (februari 2020, hierna 'M-2020') als benchmark te gebruiken, en deze te vergelijken met toetsen die later in het schooljaar zijn gemaakt (bijvoorbeeld de E-toets aan het eind van 2020, of de M-toets in 2021), kan een indicatie van leervertraging door toedoen van corona bepaald worden. Merk op dat de normering van toetsen gedaan wordt bij het ontwerpen van de toets zelf, en dat deze dus gelijk blijft over meerdere jaren. Hierdoor kunnen percentielscores van toetsen met elkaar vergeleken worden over verschillende jaren.

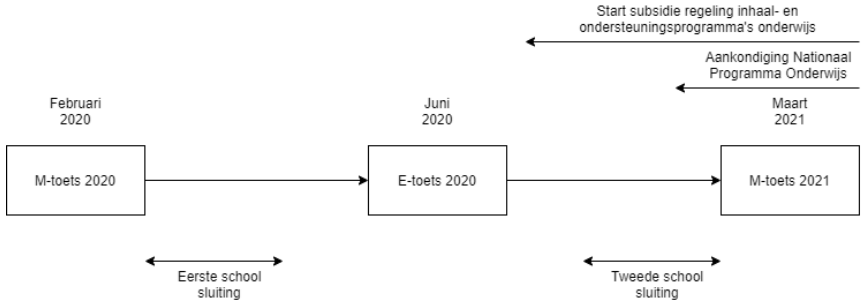
3.3 De 1e lockdown en de hele eerste coronajaar kan men zien als 2 onderzoeksperiodes

Sinds de M-2020 toets hebben er twee reguliere toetsmomenten plaatsgevonden. De eerste vond plaats na afloop van de eerste schoolsluiting, in juni 2020 (hierna 'E-2020'), de tweede vond plaats na afloop van de tweede schoolsluiting in voorjaar 2021 (hierna 'M-2021'). Er kunnen dus twee periodes geïdentificeerd worden: de periode waarbinnen de eerste schoolsluiting plaatsvond (tussen M-2020 en E-2020), en de periode waarbinnen beide schoolsluiting plaatsvonden (tussen M-2020 en M-2021). De tweede helft van de gehele periode (E-2020 tot M-2021) wordt tevens gekarakteriseerd door gerichte stappen door scholen, schoolbesturen en de regering om mogelijke leervertraging te dichten.[6] De toetsmomenten en tijdlijn van school sluitingen zijn in Figuur 2 schematisch weergegeven.

[4] https://www.pnas.org/content/118/17/e2022376118

[5] Recent zijn alternatieven voor de Cito-volgtoetsen beschikbaar gekomen (IEP, Dia, Boom), waarmee een groeiend maar nu nog heel klein deel van de scholen mee werkt. In dit onderzoek is gewerkt met Cito.

[6] https://www.dus-i.nl/subsidies/inhaal--en-ondersteuningsprogrammas-onderwijs



Figuur 2: Tijdlijn schoolsluitingen, toetsmomenten en aankondigingen subsidieprogramma's
De M-toets in 2020 is het laatste 'reguliere' toetsmoment geweest voorafgaand aan de periode van schoolsluiting. Hierdoor vormt M-2020 de logische benchmark om leervertraging tegen af te wegen. De E-toets in 2020 is het eerste toetsmoment waarbij het effect van de eerste schoolsluiting gemeten kan worden. De M-toets in 2021 is het eerste toetsmoment waarbij het effect van beide schoolsluitingen, inclusief dat van de gerichte inhaal- en ondersteuningsprogramma's van scholen gemeten kan worden. Let op dat het niet mogelijk is om het effect van de tweede schoolsluiting te isoleren van de gerichte maatregelen die getroffen zijn door scholen, leraren en het ministerie om leervertraging te adresseren: het is alleen mogelijk om beiden gezamenlijk te onderzoeken.

3.4 Methodologie om vertraging te berekenen tijdens eerste lockdown en de gehéle periode

Door per leerling te kijken naar de prestaties op de E-2020 toets en deze te vergelijken met de prestaties op de M-2020 toets kan een indicatie gemaakt worden van de leervertraging onder leerlingen gedurende de eerste shutdown. Specifiek wordt per leerling het verschil tussen E-2020 en M-2020 berekend per vak en vergeleken met het verschil tussen de E-toets en M-toets in voorgaande jaren (2019, 2018, en 2017). Wanneer leerlingen aanzienlijk slechter presteerden in E-2020 vergeleken met M-2020 is er sprake van leervertraging. Hetzelfde kan gedaan worden voor de M-2020 toets en de M-2021 toets om een indicatie te krijgen van de leervertraging over de periode waar beide schoolsluitingen in vielen.

3.5 Bevindingen uit de eerste schoolsluiting: gemiddelde vertraging van 7-8 lesweken

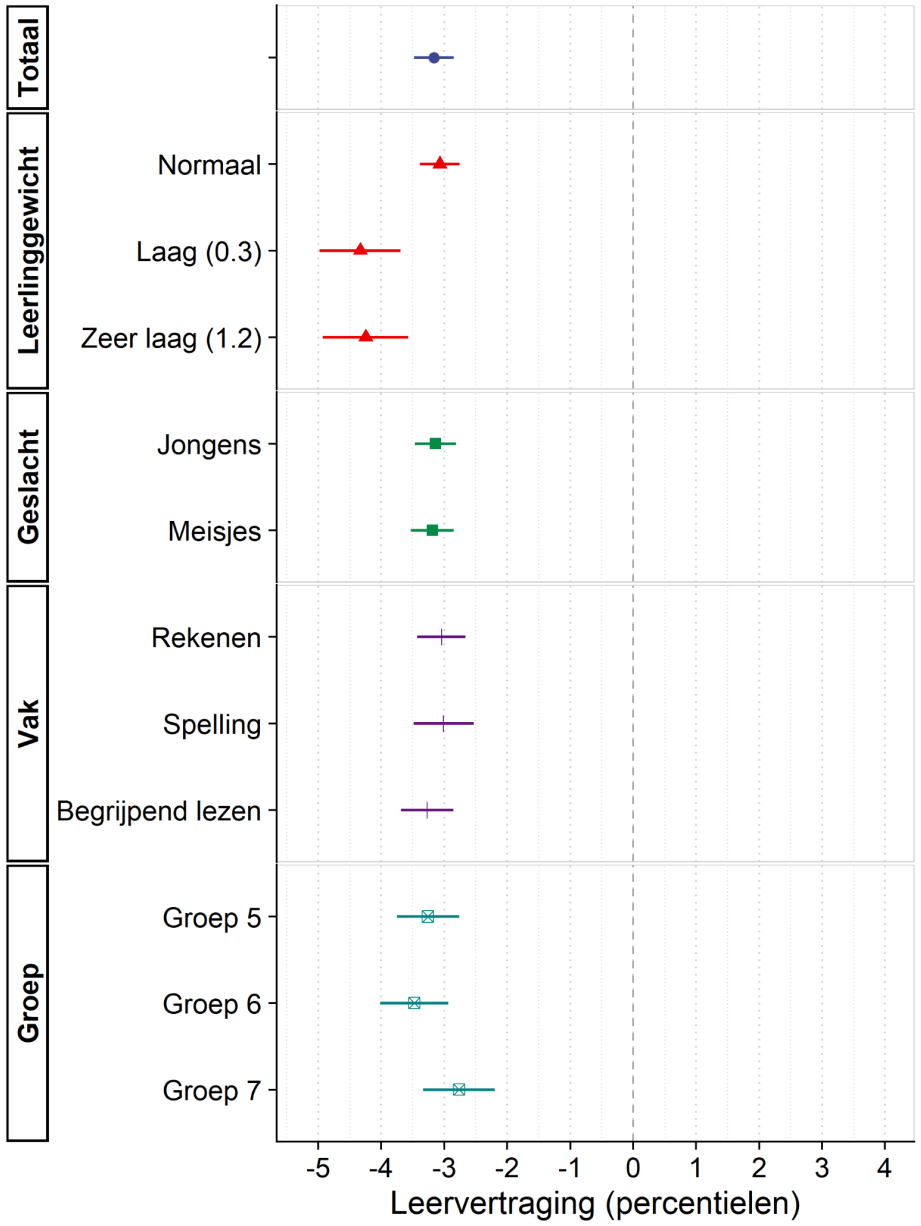
Het vergelijken van toetsscores in E-2020 en M-2020 liet zien dat leerlingen gemiddeld genomen ongeveer 3 percentielpunten lager scoorden dan verwacht na de schoolsluiting. De verwachtingen zijn gebaseerd op de 'normale' progressie in voorgaande jaren. De 3 percentielpunten staan gelijk aan (ontbreken van) de leergroei die een leerling normaal doormaakt in 7 tot 8 lesweken. Oftewel: de toetsscores direct na de 1e lockdown bleken zoveel lager dat men kan concluderen dat de leerlingen in die lockdownperiode amper leerontwikkeling hadden doorgemaakt voor zover meetbaar met de toetsen spelling, lezen en rekenen van Cito.

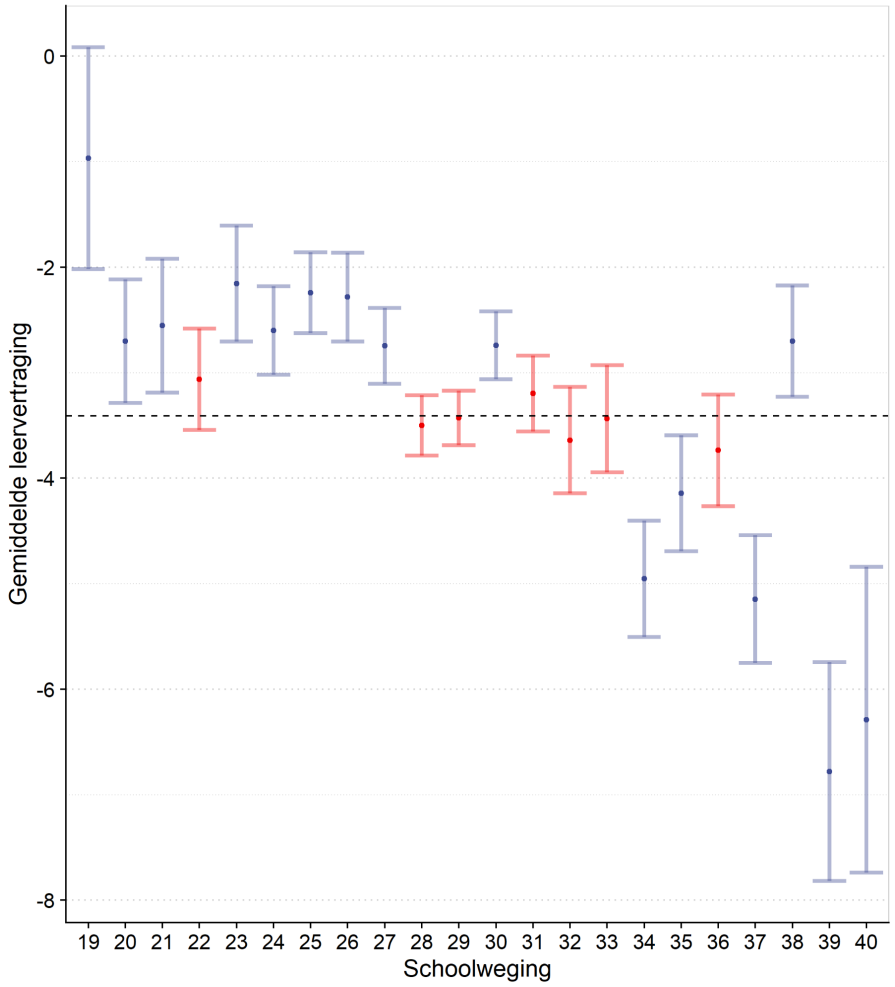
3.6 Andere bevindingen: grotere vertraging bij leerlingen uit laagopgeleide gezinnen

Tevens waren leerlingen, van wie de ouders laag opgeleid zijn (leerlinggewicht 0.3 of 1.2), disproportioneel geraakt. Zij verloren gemiddeld ongeveer 4.5 percentielpunten. Dit betekent dat deze leerlingen achteruit zijn gegaan ten tijde van de eerste schoolsluiting ten opzichte van hun prestaties van vóór corona. Er waren nagenoeg geen verschillen tussen jongens en meisjes, tussen vakken en tussen groepen. Wel was de vertraging van leerlingen op scholen met een lage schoolweging (19-25) relatief gering: deze leerlingen verloren tussen de 1 en 2 percentielpunten. Leerlingen op scholen met een hoge schoolweging (34-40) liepen aanzienlijk meer vertraging op: tussen de 4 en 7 percentielpunten.

Figuur 3: Leervertraging na afloop van de eerste schoolsluiting

In de gehele populatie leerlingen observeren we lagere toetsscores van ongeveer 3 percentielpunten wanneer gekeken wordt naar de prestaties Rekenen, Spelling en Taal (blauwe punt). Per vak zien we relatief weinig verschillen (paarse punten). Ook tussen jongens en meisjes, en groepen zien we weinig variatie (respectievelijk donker groen en turqoise punten). We zien echter verschillen tussen leerlingen uit een lager opgeleid thuismilieu: deze hebben een vertraging opgelopen die 55% groter is dan de rest van de leerlingen. Het thuismilieu wordt hier gemeten aan de hand van het leerlinggewicht, wat aangeeft of ouders van het kind in kwestie hoger onderwijs hebben genoten. Een hoger leerlinggewicht geeft aan dat de ouders lager opgeleid zijn.





Figuur 4: Gemiddelde leervertraging per schoolweging gedurende de eerste schoolsluiting
Stippellijn geeft de gemiddelde leervertraging aan in de gehele populatie. Rood geeft aan dat de gemiddelde leervertraging binnen de betrouwbaarheidsinterval van betreffende schoolweging ligt. Per schoolweging is gekeken wat de gemiddelde leervertraging is voor alle leerlingen op scholen met een bepaalde weging. Zoals te zien zijn er aanzienlijk grotere leervertragingen gemeten voor de hogere schoolwegingen, en lagere leervertragingen voor de scholen met een lage schoolweging.

3.7 Voor de periode waarin beide schoolsluitingen plaatsvonden zijn 5 scenario's denkbaar

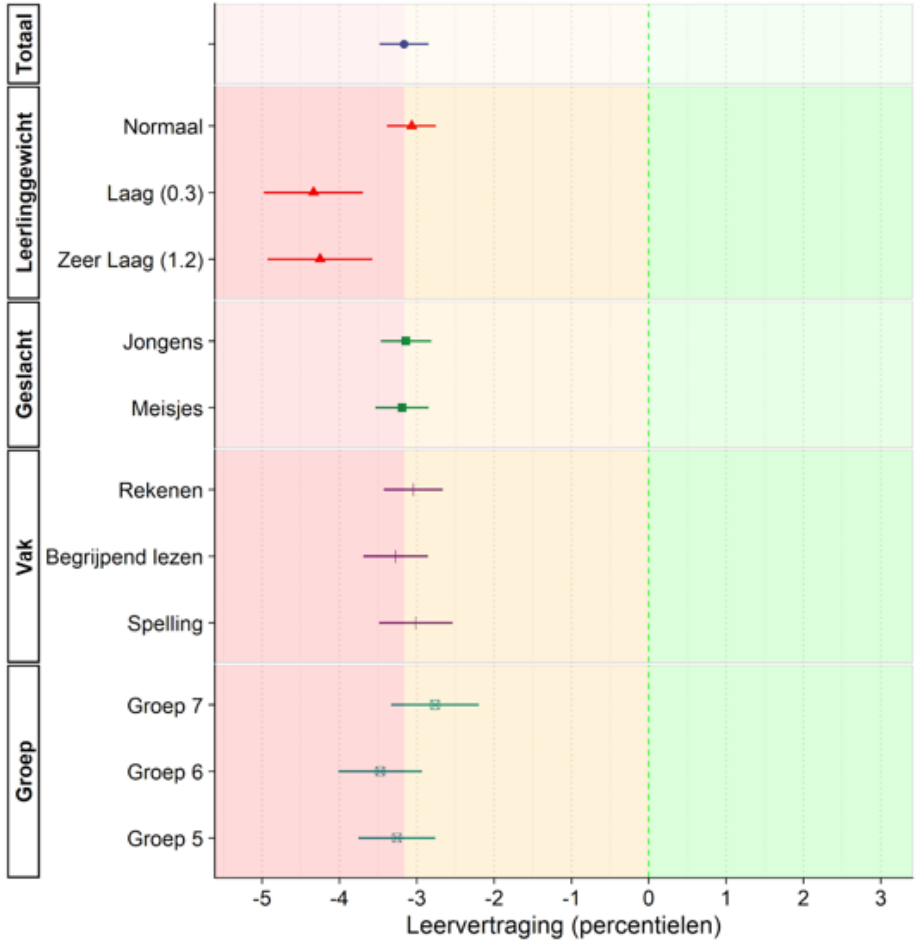
Om het effect te berekenen over de periode waarbinnen beide schoolsluitingen plaatsvonden kan de M-2020 toets wederom als benchmark gebruikt worden, maar nu vergeleken worden met de M-2021 toets. Merk op dat een vijftal scenario's zich kan voordoen wanneer gekeken wordt naar de gehele periode van M-2020 (pre-corona) tot M-2021 (na twee schoolsluitingen, plus gerichte stappen om leervertraging tegen te gaan) vergeleken met de leervertraging die we zagen na afloop van de eerste school sluiting (M-2020 tot E-2020):

1. Leervertraging is vergroot (gemiddelde verschil tussen M-2021 en M-2020 is groter dan 3 percentielpunten)
2. Leervertraging zijn gelijk gebleven (gemiddelde verschil tussen M-2021 en M-2020 is nog steeds 3 percentielpunten)
3. Leervertraging zijn gedeeltelijk ingelopen (gemiddelde verschil tussen M-2021 en M-2020 ligt tussen de 3 en 0 percentielpunten)

4. Leervertraging zijn volledig ingelopen (gemiddelde verschil tussen M-2021 en M-2020 is precies 0 percentielpunten)
5. Leervertraging zijn omgezet in leerverbeteringen (gemiddelde verschil tussen M-2021 en M-2020 is positief)

3.8 Impact van elk van de 5 scenario's is weer te geven met kleur-regio rood, oranje of groen

Deze scenario's zijn weergegeven in Figuur 5. Het eerste scenario zou betekenen dat de gemeten leervertraging over de periode M-2020 tot M-2021 ergens in het rode vlak zou moeten liggen. Het tweede scenario zou betekenen dat het verschil tussen M-2020 en M-2021 precies op de grens tussen het rode en gele vlak zou liggen. Scenario drie zou betekenen dat het verschil tussen M-2020 en M-2021 in het gele vlak ligt. Scenario vier zou betekenen dat het verschil tussen M-2020 en M-2021 precies op de grens tussen het gele en groene vak ligt. Scenario vijf zou betekenen dat het verschil tussen M-2020 en M-2021 in het groene vak ligt.



Figuur 5: Leervertraging van de eerste schoolsluiting en kleur-regio's met mogelijke scenario's voor de analyse over beide schoolsluitingen

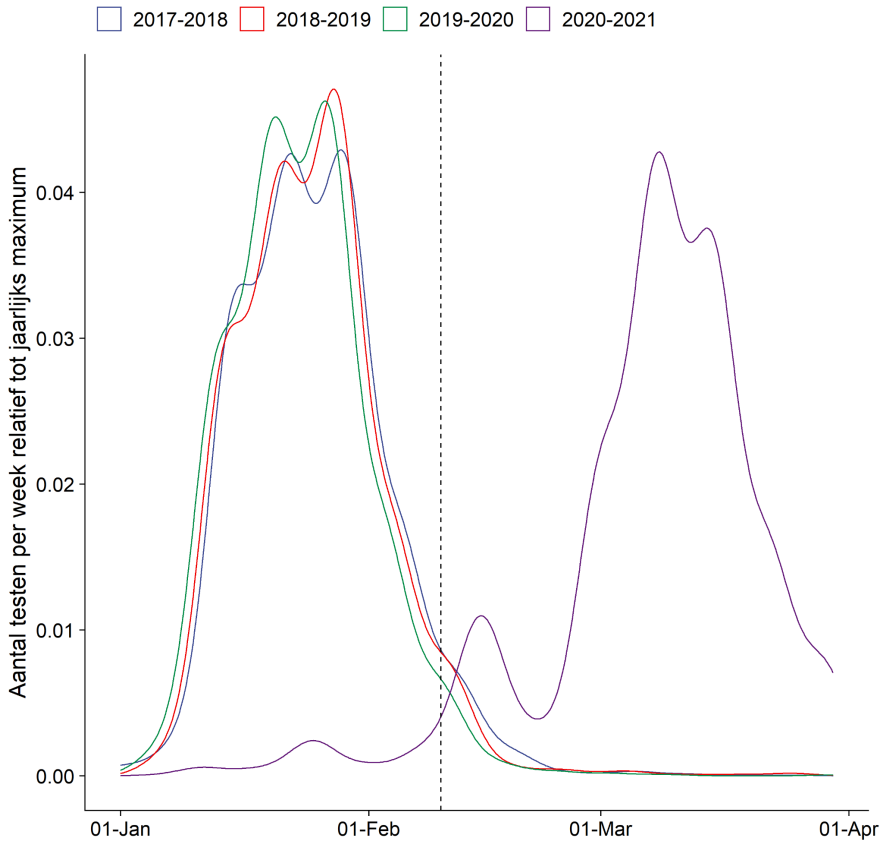
De originele bevindingen uit de eerste schoolsluiting is getoond zoals in Figuur 3, met daarbij gekleurde regio's die aanduiden wat voor scenario's mogelijk zijn voor een analyse over beide schoolsluitingen. Als leervertraging gemeten over beide school sluitingen in het rode gedeelte valt, betekent het dat er nog meer vertraging is ontstaan gedurende de periode waarin de tweede schoolsluiting plaatsvond. Wanneer de leervertraging in het oranje deel zou vallen, is de vertraging iets ingelopen. Wanneer de leervertraging in het groene deel zou vallen, is deze volledig ingelopen, en zijn prestaties over de gehele periode vanaf de eerste schoolsluiting zelfs vooruit gegaan.

3.9 Uitdagingen berekenen van leervertraging hele coronaperiode door verlaat toetsmoment

Zoals aangekondigd in de Executive Summary is er een grote uitdaging in het berekenen van leervertraging over beide periodes. Omdat de tweede schoolsluiting tot medio februari duurde, en het einde van deze tweede schoolsluiting ook nog eens samenviel met de periode van de voorjaarsvakantie, namen scholen de M-2021 toetsen aanzienlijk later af dan normaliter het geval.

Gemiddeld genomen lag het toetsmoment zelfs 6 weken later dan in een regulier jaar (zie Figuur 6). Dit komt overeen met een verlenging van de eerste helft van het schooljaar met zo'n 30%. Hierdoor zou je verwachten dat leerlingen de toets relatief goed maken, omdat ze objectief al verder in het curriculum zitten dan waarvoor de toets bedoeld is of ze in ieder geval meer oefen-/lestijd hebben gehad. Naar verwachting zal het effect van deze extra lestijd tevens afhangen van het vak dat getoetst wordt, van de groep waarin de leerling zit en van keuzes die de school gemaakt heeft gedurende de 'extra' voorbereidingstijd. Dit komt omdat de toetsen van Cito bedoeld zijn om op een bepaald moment in het curriculum de voortgang van een leerling te meten en er verschil kan zijn in de snelheid van de curriculumontwikkeling tussen vakken en groepen. Hierdoor kan zes weken 'extra' tijd anders uitpakken voor bijvoorbeeld Rekenen dan voor Taal, of voor groep 7 leerlingen ten opzichte van groep 5.

Bovenstaande betekent dat er in feite drie separate effecten aan het werk zouden kunnen zijn wanneer we kijken naar de leervertraging over de gehele coronaperiode. Allereerst is er het effect van de twee schoolsluitingen, die naar verwachting negatief zal zijn geweest, op toetsresultaten. Ten tweede zijn er de gerichte stappen van leraren, scholen en de regering, om leervertraging zo veel mogelijk tegen te gaan / in te lopen, gedurende de periode sinds de eerste schoolsluiting. Ten derde is er de mogelijke verbetering van toetsresultaten doordat het toetsmoment aanzienlijk later in het jaar plaatsvond.



Figuur 6: Moment van afname M-toets in 2021 in het coronajaar vergeleken met reguliere jaren

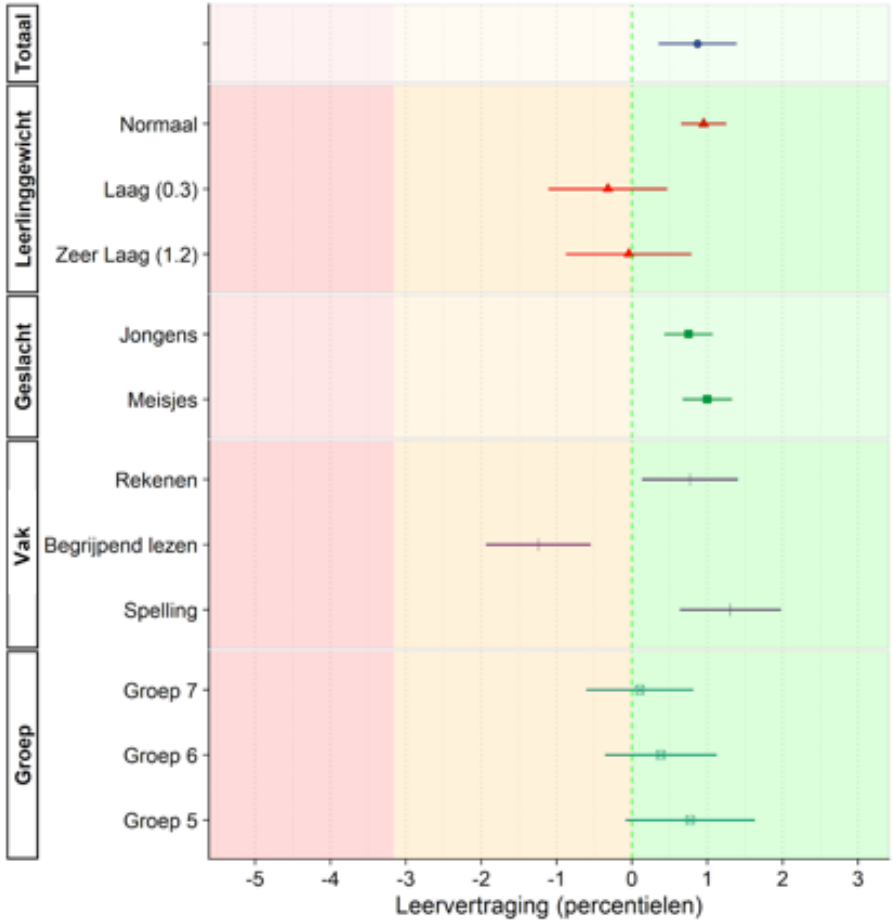
Zoals te zien valt in bovenstaande Figuur valt het gros van de M-toetsen in 2021 ver na het reguliere toetsmoment (medio januari tot medio februari). Dit komt doordat de schoolsluiting tot medio februari aanhield en de meeste scholen kozen om pas weer te toetsen nadat de scholen weer geopend waren. Gemiddeld genomen valt het toetsmoment in M-2021 ongeveer 6 weken later dan in de voorgaande drie jaren.

3.10 Bevindingen over beide schoolsluitingen geven een positief beeld, maar zullen vertekend zijn door het verlate toetsmoment

Wanneer daadwerkelijk gekeken wordt naar de prestaties van leerlingen in M-2021 vergeleken met hun prestaties in M-2020 - m.a.w. de leervertraging over de gehele coronaperiode met beide schoolsluitingen - zien we dat leerlingen over het algemeen juist vooruit zijn gegaan (zie Figuur 7). Dit zou impliceren dat alle leervertraging is ingelopen en leerlingen zelfs hóger zijn gaan scoren gedurende de coronaperiode. Zoals eerder benoemd zal dit niet daadwerkelijk het geval zijn: omdat leerlingen gemiddeld genomen meer dan 6 weken extra tijd hebben gehad, is het heel aannemelijk dat ze op de M-2021 toets relatief hoog gescoord hebben. Hierdoor lijkt het of leerlingen alle leervertraging (en meer) hebben ingelopen.

We zien tevens dat er aanzienlijke verschillen zijn tussen vakken en groepen, zo is vooruitgang met name hoog voor spelling en rekenen en voor leerlingen in lagere groepen. Zoals eerder aangegeven is deze variatie in lijn met de verwachting dat het positieve effect van extra voorbereidingstijd zich anders manifesteert voor verschillende vakken en voor verschillende groepen. Omdat Cito-toetsen ontworpen zijn om op een bepaald moment in het curriculum de voortgang te boeken, en alle toetsen aanzienlijk later zijn afgenomen, zal de toename in toetscijfers door toedoen van de ‘extra’ tijd afhangen van de snelheid waarmee het curriculum zich ontwikkeld heeft in de ‘extra’ tijd. Deze kan verschillen per groep, maar ook per vak.

Tussen jongens en meisjes zijn geen noemenswaardige verschillen gevonden. De verschillen tussen leerlingen met een leerlinggewicht (uit een lager opgeleid milieu), die tevens in de resultaten van de eerste schoolsluiting waren gevonden, blijken opnieuw uit de analyse over de gehele periode met beide schoolsluitingen.



Figuur 7: leervertraging over de periode waarbinnen beide schoolsluitingen plaatsvonden

Zoals hierboven te zien valt zijn leerlingen over het algemeen vooruit gegaan over de gehele periode van voorjaar 2020 tot voorjaar 2021. Dit zou betekenen dat leerlingen niet alleen alle leervertraging uit de eerste school sluiting hebben ingelopen, maar ook dat ze over de gehele periode zich disproportioneel verbeterd hebben ten opzichte van voorgaande jaren. Deze bevinding zullen waarschijnlijk te hoog zijn uitgevallen doordat leerlingen gemiddeld genomen ongeveer 6 weken later de toetsen hebben genomen, vergeleken met de geplande toetsdatum. Hierdoor zullen leerlingen een hogere score hebben gehaald dan de score die bij hun daadwerkelijk progressie hoort.

3.11 Het trekken van sterke conclusies over vertraging is lastig omdat context ontbreekt en extra tijd van zes weken ongekend is

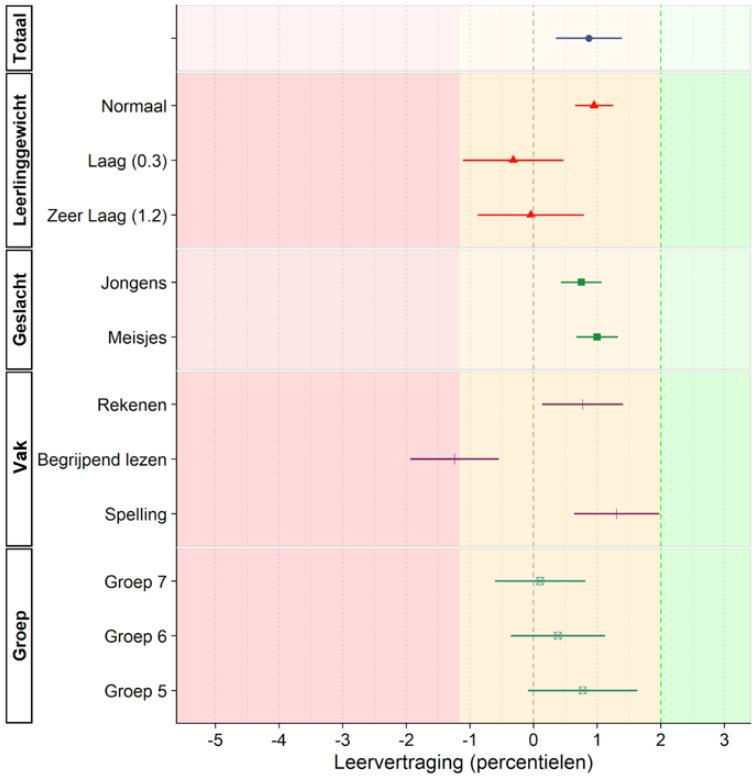
Het is erg lastig om een verwachting te formuleren over hoeveel hoger leerlingen zouden 'moeten' scoren wanneer ze 6 weken extra voorbereidingstijd ontvangen. Anders zou het mogelijk zijn om de geobserveerde resultaten terug te schalen, bijvoorbeeld als we wisten dat 6 weken extra tijd betekent dat leerlingen gemiddeld x percentielpunten hoger scoren. Helaas is deze informatie er niet. Dit komt door een aantal redenen. Allereerst is het onduidelijk wat scholen precies gedaan hebben gedurende de 'extra' tijd: zijn ze doorgegaan met het curriculum, of hebben ze de periode gebruikt om leerlingen voor te bereiden voor de M-toets? Daarnaast is het eigenlijk nooit voorgekomen dat leerlingen een dergelijke hoeveelheid extra tijd hadden voorafgaand aan een toetsmoment. Er zijn wel schattingen te maken over kleine verschillen in tijd, bijvoorbeeld een paar dagen of een week, op basis van natuurlijke variatie in toetsmomenten, maar een periode van 6 weken is ongekend.

3.12 Werkend met 'verwachte leerwinst per week' lijkt de vertraging toch deels ingelopen te zijn, al blijft interpretatie afhankelijk van aannames

Oftewel, er is weinig tot geen houvast om een inschatting te maken van de te verwachten impact van meerdere weken extra voorbereidingstijd. Toch willen we een poging doen om een aantal mogelijke scenario's te schetsen. Op basis van de natuurlijke spreiding in toetsresultaten kan aangenomen worden dat leerlingen 0.33 percentielpunten reguliere leerwinst per week zouden boeken. Dit is tevens hoe in het eerdere onderzoek berekend is dat er 7-8 weken leervertraging had plaatsgevonden (Engzell, Frey & Verhagen, 2021). Dit zou betekenen dat leerlingen gemiddeld ongeveer 2 percentielpunt hoger zouden hebben gescoord op de verlate M-2021 toets dan wanneer ze de toets op het 'normale' moment gemaakt hadden. Hierdoor zou de geobserveerde 'vooruitgang' uit Figuur 7 van ongeveer 0.5 percentielpunt teruggeschaald kunnen worden naar ongeveer -1.5 percentielpunt. Met andere woorden: dit zou betekenen dat leerlingen gemiddeld de helft van de eerder gemeten leervertraging goed hebben weten te maken sinds de eerste schoolsluiting (van -3 percentielpunt naar -1.5 percentielpunt).

3.13 Sensitiviteitsanalyse toont de gevoeligheid voor aannames op 'de' impact van corona

Er is zoals gezegd zeer weinig houvast wat te verwachten is van 6 weken extra lestijd, al lijkt het bijvoorbeeld onrealistisch dat leerlingen meer dan 4 percentielpunten hoger scoorden op de M-2021 toets (dit zou namelijk betekenen dat ze 0.67 percentielpunten per week winnen - zie tevens Van der Werfhorst, 2021). Dit betekent dat leervertragingen waarschijnlijk niet vergroot zijn (scenario 1), maar eerder gelijk zijn gebleven of iets zijn ingelopen (scenario's 2 of 3) gedurende de periode tussen E-2020 en M-2021 (waarbinnen de tweede schoolsluiting viel). In Figuur 8 zijn de drie eerder genoemde scenario's (rood, oranje, groen) met twee punten opgeschoven in lijn met bovenstaande rationale. Hierdoor zien we dat de gemeten leervertraging in het oranje gedeelte valt (scenario 3). In Tabel 1 is sensitiviteit van de resultaten naar de aanname van 'reguliere' leerwinst weergegeven.



Figuur 8: Leervertraging over de gehele periode, waarbij de scenario's zijn opgeschoven met twee percentielpunten
Hierboven zijn de resultaten die eerder beschreven zijn voor de gehele periode waarin beide schoolsluitingen vielen opnieuw getoond, maar zijn de drie scenario's met twee percentielpunten opgeschoven. Hierdoor staat een netto leer-'winst' van 2 percentielpunten gelijk aan scenario 4 (op de rand van het gele en groene vlak): alle leervertraging uit de eerste schoolsluiting is ingelopen. Door de scenario's te verschuiven zien we dat het waarschijnlijk lijkt dat leervertragingen deels zijn ingelopen (scenario 3) ten opzichte van de eerste schoolsluiting, maar niet volledig ingelopen zijn.

Aan-name leer-winst per week	Verwachte verhoging door toedoen 'extra tijd'	Gemiddelde scores		Rekenen		Spelling		Taal	
		Gemeten	Schaling	Gemeten	Schaling	Gemeten	Schaling	Gemeten	Schaling
+0.25 / week	6 x 0.25 = 1.5	+0.8	-0.7	+0.8	-0.7	+1.3	+0.3	-1.2	-2.7
+0.33 / week	6 x 0,33 = 2	+0.8	-1.2	+0.8	-1.2	+1.3	-0.2	-1.2	-3.2
+0.5 / week	6 x 0.5 = 3	+0.8	-2.2	+0.8	-2.2	+1.3	-1.2	-1.2	-4.2
+0.66 / week	6 x 0.66 = 4	+0.8	-3.2.	+0.8	-3.2	+1.3	-2.2	-1.2	-5.2

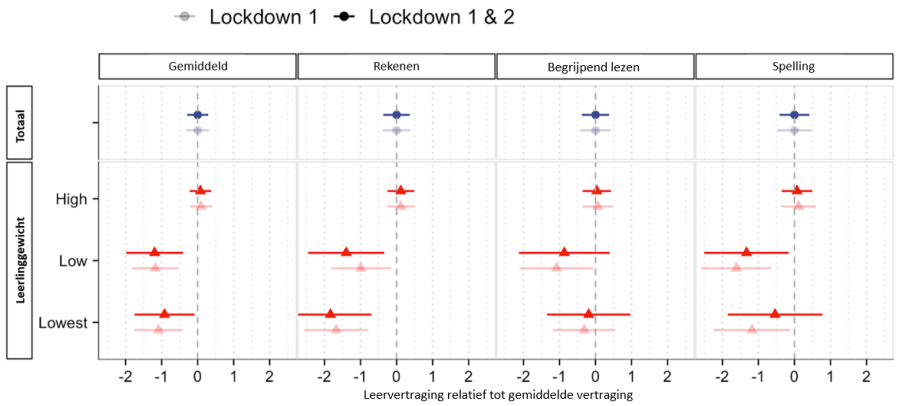
Tabel 1: Impact leervertraging van aannames over hogere toetsscores door 6 weken latere M-2021 toetsmoment
Op basis van de 6 weken 'extra' tijd die leerlingen hadden ter voorbereiding op de M-toets in 2021 kan men voorspellen 'hoeveel hoger' leerlingen hebben gescoord vergeleken met de situatie waarin ze de toets op het reguliere zouden hebben moment gemaakt. Op basis van onderzoek naar de eerste schoolsluiting was een leerwinst van 0.33 percentielpunt per week berekend, maar in dat onderzoek was geen sprake van latere toetsafname. Zoals te zien in tabel 1 hangt de conclusie omtrent leervertraging een-op-een samen met de aannames over de impact van het verlate toetsmoment. Niettemin lijkt het onrealistisch dat leervertraging vergroot is ten opzichte van de eerste schoolsluiting aangezien het onwaarschijnlijk is dat leerwinst per week veel af zal wijken van de aangenomen 0.33 / week. Dat gezegd hebbende is er weinig sturing beschikbaar door de uniciteit van de situatie.

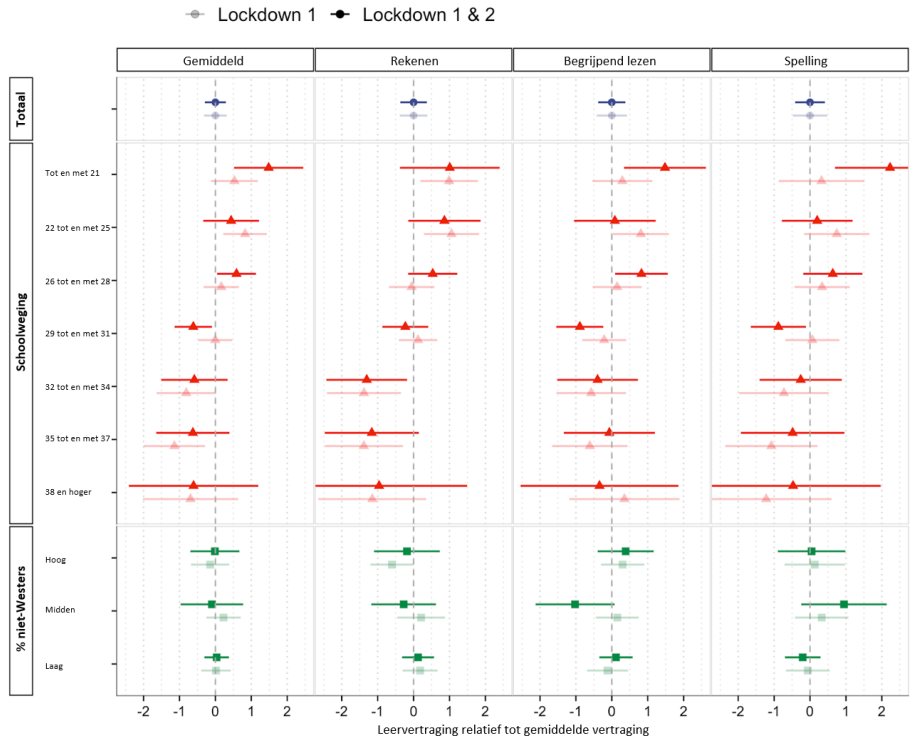
3.14 Analyse van verschillen tussen groepen blijft mogelijk onder aanname dat de impact van de ‘extra’ tijd gelijk is voor alle leerlingen

Er is tevens gekeken of onderlinge verschillen, die eerder zichtbaar waren tussen groepen leerlingen en scholen, zijn veranderd in de periode tussen de E-2020 en M-2020 toetsen. Bijvoorbeeld tussen leerlingen van wie de ouders laagopgeleid zijn en de rest van de leerlingen. Het zou kunnen dat de eerste groep nog verder achter is gaan lopen, of juist dat deze groep iets van de vergrote leervertraging heeft ingelopen op de andere leerlingen. Wanneer we aannemen dat de impact van de ‘extra’ tijd op toetsprestaties gelijk is voor verschillende leerlingen, kan deze analyse uitgevoerd worden.

3.15 Laagopgeleide ouders of hoge schoolweging correleren met grotere vertraging

We kunnen deze vraag beantwoorden door te kijken naar de onderlinge verschillen in leervertraging voor verschillende groepen leerlingen of scholen *ten opzichte van de gehele populatie.* Wanneer we alleen kijken naar de verschillen in vertraging, zien we dat de grotere leervertraging voor leerlingen met laagopgeleide ouders gelijk zijn gebleven over beide schoolsluitingen (Figuur 9, boven). We zien ook dat scholen met een lage schoolweging het relatief beter hebben gedaan over beide sluitingen dan over alleen de eerste (Figuur 9, onder). Deze verschillen tussen scholen met een niet complexe populatie (lage schoolweging) en scholen met een complexe populatie (hoge schoolweging) zijn echter niet statistisch significant.





Figuren 9: Verschil in leervertraging tussen groepen leerlingen en groepen scholen over beide periodes vergeleken met de eerste schoolsluiting

In hiernaast weergegeven figuren wordt alleen gekeken naar het verschil in leervertraging voor sub-populaties ten opzichte van de leervertraging in de gehele populatie. In het bovenste figuur worden leerlingen uit een lager opgeleid milieu vergeleken met leerlingen uit een hoger opgeleid milieu vergeleken met leerlingen uit een hoger opgeleid milieu voor de gehele periode (in donkere kleur) en voor de eerste schoolsluiting (in lichte kleur). Hieruit blijkt dat onderlinge verschillen even groot zijn gebleven. Hetzelfde is gedaan voor verschillende type scholen (onderste figuur). Hieruit blijkt tevens dat schoolverschillen even groot zijn gebleven gedurende de gehele periode ten opzichte van de eerste schoolsluiting.

4.Conclusie en aanbevelingen

4.1 Latere toetsing bemoeilijkt het trekken van harde conclusies

Het feit dat de M-toetsen in 2021 zo'n 6 weken later zijn afgenomen dan normaal (en leerlingen dus zo'n 30% extra lesweken hadden) zorgde voor een serieus methodologisch probleem. Echt harde conclusies zijn daarom eigenlijk niet te trekken. Echter, dankzij enkele onderbouwde aannames durven we hier wel tot een paar (voorzichtige) conclusies te komen:

- De aanwijzingen zijn sterk dat de gemiddelde leervertraging in ieder geval niet is vergroot tussen de zomer 2020 en het einde van de tweede schoolsluiting.
- Onderlinge verschillen tussen leerlingen met laagopgeleide ouders en de rest van de leerlingen zijn nog steeds aanzienlijk (55%), al zijn deze verschillen niet verder toegenomen sinds zomer 2020.
- Tot slot blijft begrijpend lezen sterk achter bij rekenen en spelling, maar is het onduidelijk of dit komt door een verschil in het effect van 'extra' tijd of door een gebrek aan inlopen van leervertraging voor dit vak.

4.2 Afnemen E-toetsen in juni is essentieel om te kunnen sturen - dit geldt op alle lagen

Zowel tijdens een crisis als in de herstelfase is goede informatie essentieel om gedegen te kunnen sturen. Dit geldt op alle lagen: leraren in de klas, de schoolleiding, besturen en de regering. Om echt een zuiver beeld te krijgen van de staat van leerontwikkeling tijdens en na corona is het dan ook essentieel om de E-toetsen in 2021 af te nemen. Afname van E-toetsen op het reguliere moment, medio juni, voorkomt het methodologische probleem van dit huidige onderzoek. Zouden ook de E-toetsen later worden afgenomen dan ontbreekt wederom een zuiver beeld van leerontwikkeling, en bemoeilijkt dit het maken van een NPO-schoolscan als basis voor adequaat herstelbeleid. Hetzelfde geldt voor alternatieve afname in september; de aanbeveling is: niet doen, enkel toetsen in september kan vlot herstel bemoeilijken.

4.3 Een toets oordeelt nooit, hij meet. Zuiver meten en duiden van deze resultaten is waar de onderwijskundige waarde zit

Soms bestaat de vrees dat leerlingen (of leraren) worden 'afgerekend' op een lage toetsscore. Meegaan in deze manier van denken is niet productief. Halfjaarlijkse volgtoetsen zijn ontworpen om te *volgen*, ze (be)oordelen niet. Zeker in/na een crisissituatie, wanneer vergelijken met de 'normale' situatie ten dienste staat van herstel, helpt het om waardeneutraal te blijven. Wat *altijd* nodig is, is het evalueren van resultaten in context, en het samen zoeken naar verklaringen voor toetsresultaten en trends. Juist dat kan helpen om te komen tot adequate maatregelen die elke leerling, groep of school verder helpen in hun ontwikkeling. Door niet te toetsen, onthouden we onszelf (en de leerlingen) van waardevolle informatie die helpt om het gesprek over onderwijs te objectiveren, vertraging te duiden en te werken aan spoedig herstel waar dat nodig is.

4.4 Zet in op duiding of op een extra meting

Een *extra* volgtoets afnemen in september kan natuurlijk altijd, bijvoorbeeld om te evalueren wat het effect is geweest van een zomerschool. Ook in het huidige schooljaar maakten veel scholen gebruik van de extra toetsnormering die Cito had gemaakt voor aparte toetsing in september. Een alternatief is het inzetten van methodetoetsen om te kijken welke leerdoelen extra aandacht behoeven. Ook kan extra duiding van eerdere metingen, of meer aandacht voor de overdracht, helpen om de kennis over een groep of een leerling aan te vullen.

5. Reflecties tot besluit

Als maatschappelijk betrokken onderzoekers nemen we tot slot de vrijheid om nog enkele andere reflecties te delen.

5.1 Dit onderzoek is een eerste stap. Meer onderzoek is nodig voor gericht herstel

Naast het eerder al genoemde belang van het afnemen van E-toetsen is nader onderzoek nodig om te begrijpen wat de verschillen in impact verklaart. Welke mechanismen maken dat we verschillen zien in leervertraging tussen groepen leerlingen en groepen scholen? Welke lessen vallen hieruit te trekken voor ons onderwijsstelsel? Beantwoording hiervan vergt meer inzicht in het verloop van het onderwijs tijdens corona, inclusief inhaalprogramma's. Het actief monitoren van extra interventies tijdens de 2½ jaar NPO maakt het bovendien mogelijk om te leren welke extra interventies effectief zijn in welke situaties. De veelheid en de diversiteit aan interventies tijdens NPO biedt een unieke kans om veel beter te leren wat werkt, ook buiten de context van de coronacrisis. Dit vereist echter gericht inspanningen van onderzoekers, beleidsmakers en scholen.

5.2 Verder is structurele versterking nodig, onderbouwd met (andersoortig) onderzoek

Terwijl de focus bij 'coronaherstel' primair ligt op het inlopen van eventuele achterstanden van de huidige leerlingen, ligt de focus bij structurele versterking op de schoolorganisaties zelf. Juist het feit dat onder meer de Inspectie er in recente jaren telkens op wees hoe groot de verschillen zijn tussen scholen, geeft aan dat er nu enorme kansen liggen. Goed begrijpen waar deze verschillen vandaan komen maakt het mogelijk om de kwaliteit en veerkracht van álle scholen op een hoger niveau te krijgen. Ondanks dat het NPO niet 'even' een oplossing zal bieden voor hardnekkige onderliggende problemen, zoals het structurele lerarentekort, liggen er kansen om nu extra onderzoek te doen om te leren wat werkt en wat nog niet.

5.3 Voorts behoeft naast 'meetbare' ook de 'merkbare' ontwikkeling serieuze aandacht

Veelbesproken is verder de noodzaak om juist nu ook aandacht te schenken aan welbevinden en andere aspecten van onderwijs die niet zo makkelijk te meten zijn als taal en rekenen. Het is belangrijk de ontwikkeling van leerlingen breed te bezien, ook buiten de kernvakken. Hiervoor ontbreekt momenteel nog altijd een duidelijk in- en overzicht. Door gebrek aan data en uitblijven van gecoördineerde actie blijft deze dimensie van de coronacrisis goeddeels onderbelicht.

5.4 Tot slot: de rijke (data-)infrastructuur van ons PO biedt Nederland een voorsprong

Zoals prof. Van der Werfhorst van de Universiteit van Amsterdam onlangs schreef in PNAS, kent Nederland een primair onderwijs een ongekend rijke (data-)infrastructuur. Als de coronacrisis ons iets leert is het wel hoe belangrijk metingen zijn om problemen helder te krijgen, en hier vervolgens adequaat op te reageren. Een parallel valt te trekken met de waarde van zorgvuldig en frequent (medisch) testbeleid op corona, wat diverse landen een voorsprong gaf bij een effectieve pandemiebestrijding. Waar andere landen nog altijd bezig zijn leervertraging überhaupt te duiden - met alle gevolgen van dien voor leerlingen, leraren en het onderwijs - zijn wij in Nederland al bezig met gerichte stappen om deze vertraging in te lopen. De centrale rol van onze rijke toetsinfrastructuur is hierbij van onschatbare waarde geweest.

6. Colofon

6.1 Literatuurverwijzingen

Engzell, E., A. Frey & M.D. Verhagen. 2021. Learning loss due to school closures during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*. Vol 118 (17), 'PNAS'.

Van de Werfhorst, H.G. 2021. Inequality in learning is a major concern after school closures. *Proceedings of the National Academy of Sciences*. Vol 118 (20), 'PNAS'.

6.2 De auteurs

- Mark Verhagen is verbonden aan de Universiteit van Oxford. Zijn onderzoek richt zich op het gebruik van geavanceerde statistische methodiek binnen de sociale wetenschappen.
- Jelte de Jongh werkt bij Leeruniek, een social enterprise die zich sterk maakt voor gelijke kansen in het onderwijs. Dit doet Leeruniek door scholen en besturen te helpen om de ontwikkeling van leerlingen in kaart te brengen, opdat zij het onderwijs kunnen inrichten op basis van de specifieke behoefte van leerlingen. Daarnaast ondersteunt Leeruniek maatschappelijk relevant onderzoek rond dit thema.

6.3 Werkwijze, technische bijlage en Q&A

Een technische appendix met een beschrijving van werkwijze en technische, Engelstalige documentatie over de methodiek en computaties, vindt u op www.leeruniek.nl/technicalappendixcorona. Een Q&A met betrekking tot de inzet van geanonimiseerde en geaggregeerde toetsscores, vindt u op www.leeruniek.nl/onderzoek.