

TECHNISCHE ACHTERGROND

WAAROM WIJ **NIET** CORRIGEREN VOOR
AFNAMEDATUM

(EN SCHOLEN AANRADEN DAT OOK NIET TE DOEN)

Stichting Cito

Kimberley Lek
Remco Feskens
Judith Veldhuizen
Jos Keuning

WAAROM WIJ **NIET** CORRIGEREN VOOR AFNAMEDATUM

EN SCHOLEN AANRADEN DAT OOK NIET TE DOEN

In juni 2021 presenteerden we de resultaten van ons tweede onderzoek naar de gevolgen van de coronacrisis voor het basisonderwijs. Die resultaten presenteerden we zoals we ze vonden: zonder correctie voor – bijvoorbeeld – afnamedatum. Dat we niet corrigeerden voor afnamedatum, was een weloverwogen keuze. Zelfs terwijl wél corrigeren intuïtiever leek. De onderzochte M-toetsen werden dit schooljaar immers gemiddeld zo'n 6 weken later afgenomen dan in voorgaande schooljaren. Daardoor had het huidige cohort leerlingen langer de tijd om zich de leerstof eigen te maken dan het voorgaande cohort. Hieronder leggen we onze fundamentele keuze uit.

WE CORRIGEREN NU NIET:

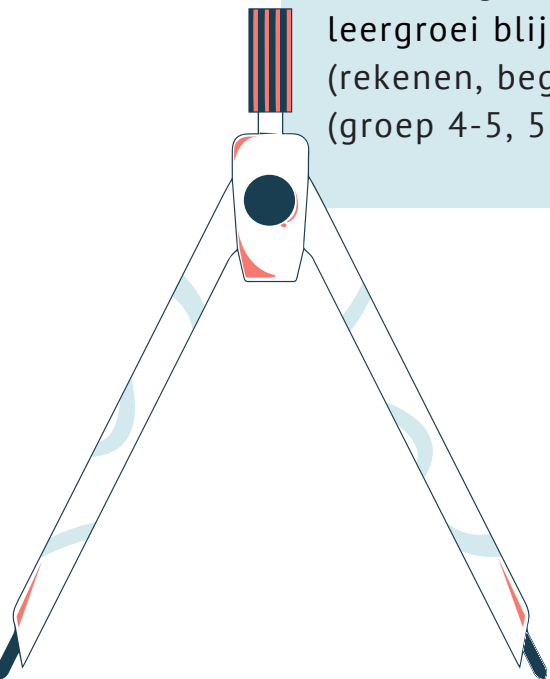
OMDAT ER ALTIJD VERSCHILLEN ZIJN IN AFNAMEDATUM EN LEERTIJD

Ook in 'normale tijden' nemen scholen niet allemaal in dezelfde week de Cito leerlingvolgysteemtoetsen af. In de toetskalender wordt voor de M-toetsen een afnameperiode van half januari tot half februari geadviseerd. Voor de E-toetsen is juni het advies. Of scholen nu aan het begin of aan het einde van deze periode de toetsen afnemen, maakt niet uit. We weten namelijk dat leerlingen die de toets wat later maken, niet systematisch anders scoren dan leerlingen die de toets wat eerder maken. Bovendien verschilt de hoeveelheid leertijd altijd al iets van leerling tot leerling. Door afnamedatum, maar ook doordat het schooljaar vanwege spreiding in schoolvakanties niet voor iedereen op hetzelfde moment start. De eerlijkheid, geldigheid of bruikbaarheid van de toetsresultaten die leerlingen behalen, verandert hier niet door. Niet door een (iets) eerdere of latere afnamedatum en ook niet door een (iets) eerdere of latere start van het schooljaar.

WE CORRIGEREN NU NIET:

OMDAT ER GEEN RELATIE IS TUSSEN VAARDIGHEIDSSCORE EN LATERE AFNAMEDATUM

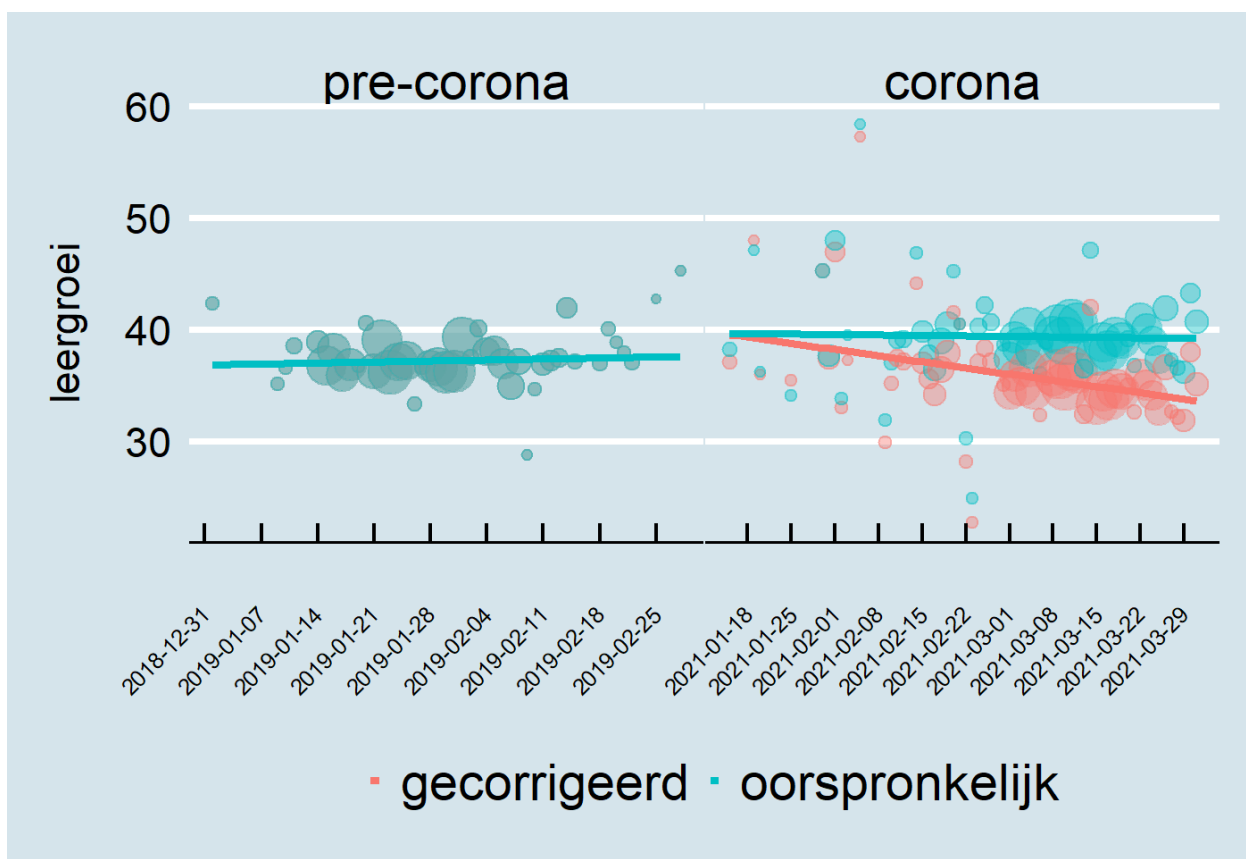
In 2021 viel de afnameperiode voor de M-toets precies in de tweede lockdown. Daardoor werden basisscholen gedwongen de afname behoorlijk wat later te plannen dan de toetskalender adviseert. Namelijk zo'n 6 weken. Is dat zodanig lang dat je de scores van de leerlingen daarvoor moet corrigeren? Je moet dan voorspellen welke score een leerling behaald zou hebben als de toets eerder was afgenomen. Uiteraard hebben we die berekening gemaakt. We hebben in kaart gebracht in hoeverre leerlingen extra leergroei lieten zien ten gevolge van de latere afnamedatum. Die extra leergroei blijkt er niet te zijn. Voor geen van de vakken (rekenen, begrijpend lezen) en in geen van de jaargroepen (groep 4-5, 5-6 en 6-7; zie Bijlage 1).



WE CORRIGEREN NU NIET:

OMDAT 'BAAT HET NIET, DAN SCHAADT HET WÉL'

Wij vinden geen relatie tussen de latere afnamedatum en de scores van leerlingen in coronatijd. En dus hebben wij niet gecorrigeerd voor afnamedatum. Wanneer onderzoekers wél corrigeren, is dit vaak aan de hand van een lineaire correctie. Bij een dergelijke correctie nemen we aan dat leergroei tussen twee toetsmomenten lineair verloopt. Dus als een leerling over een periode van 10 weken een groei doormaakt van 25, dan nemen we bij een lineaire correctie aan dat de vaardigheidsscore elke week met 2.5 toeneemt. Dat dit type correctie een behoorlijk effect heeft op de geschatte leergroei van leerlingen, blijkt uit Figuur 1 (voor rekenen, groep 4-5; figuur is nagenoeg identiek voor andere groepen en begrijpend lezen). Te zien is dat de afnamedatum vóór correctie (de blauwe puntenwolk en lijn) geen effect heeft op de geschatte leergroei. Niet in de periode vóór corona en ook niet in de afgelopen periode met lockdown. De lijn loopt immers nagenoeg horizontaal. Maar passen we een lineaire correctie toe, dan verandert het beeld. We introduceren dan kunstmatig een lineair effect van afnamedatum op de leergroei van leerlingen. Het gevolg is dat de leergroei van leerlingen in het coronacohort flink wordt verlaagd. Waar we vóór de correctie scores hadden die prima bruikbaar waren, hebben we ná correctie scores die hypothetisch zijn, en dus allesbehalve zeker.



FIGUUR 1.

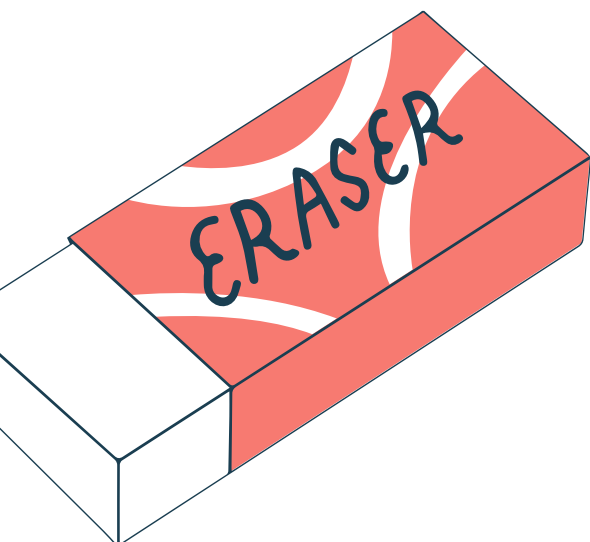
Effect van een lineaire correctie op geschatte leergroei van leerlingen bij rekenen, groep 4-5.

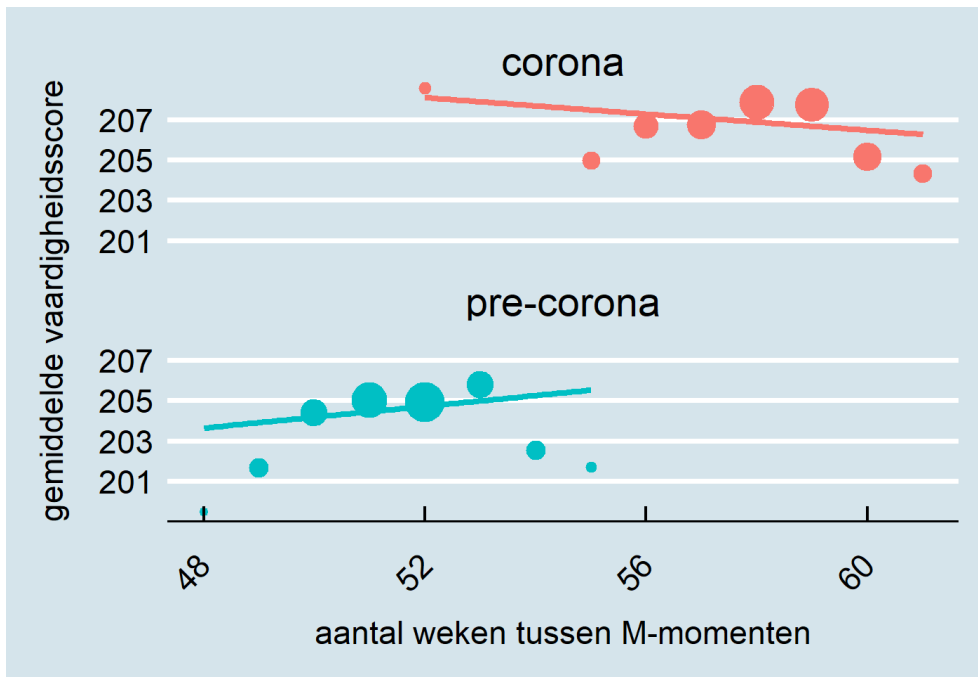
MOET JE DAN NOOIT CORRIGEREN? DIE CONCLUSIE KUN JE NIET TREKKEN

Als Cito vinden we overigens niet dat je *nóóit* moet corrigeren. Want in voorkomende gevallen kan er wel degelijk reden zijn om wél te corrigeren. Bedenk daarbij wel dat het toepassen van een correctie hetzelfde is als het doen van een voorspelling. En in elke voorspelling schuilt onzekerheid. Het is daarom belangrijk dat je onderzoekt hoe je die voorspelling het beste kunt doen. Daar zijn verschillende manieren voor.

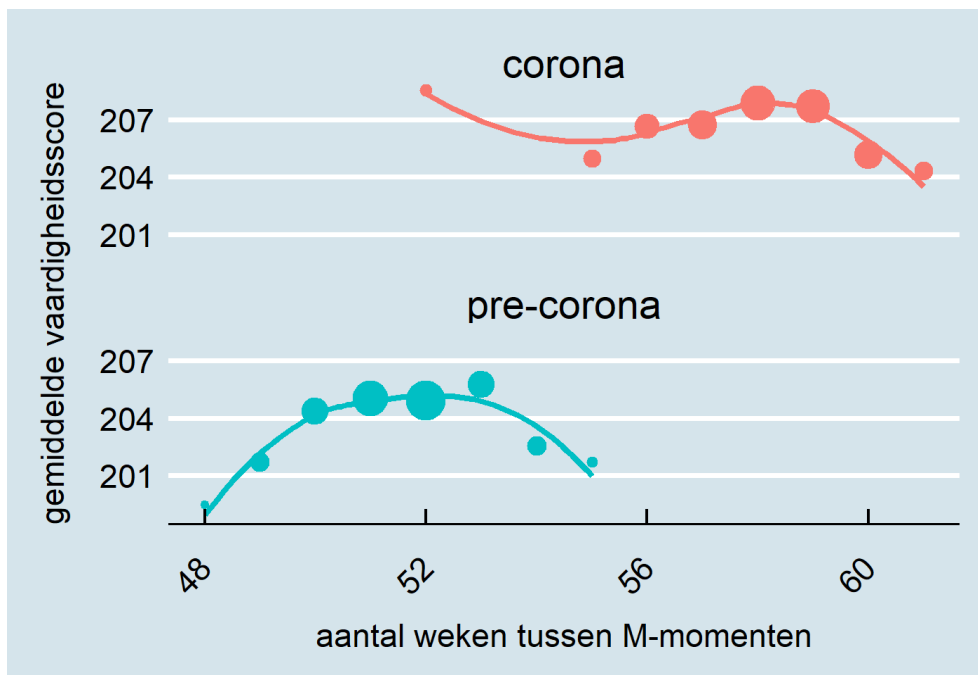
LINEAIRE CORRECTIE NIET OVEREENKOMSTIG MET ONZE DATA

In de eerder geïllustreerde, veelgebruikte correctie wordt aangenomen dat de leergroei tussen twee toetsmomenten lineair verloopt. Een dergelijke lineaire groei zien wij niet terug. Niet bij rekenen en niet bij begrijpend lezen (voor rekenen groep 4-5 geïllustreerd in Figuur 2a). Áls je al zou corrigeren, dan zou een correctie gestoeld op een curvilineaire lijn beter te verdedigen zijn. Die lijkt de eventuele relatie tussen afnamedatum en vaardigheidsscore beter samen te vatten (Figuur 2b).





A. lineair



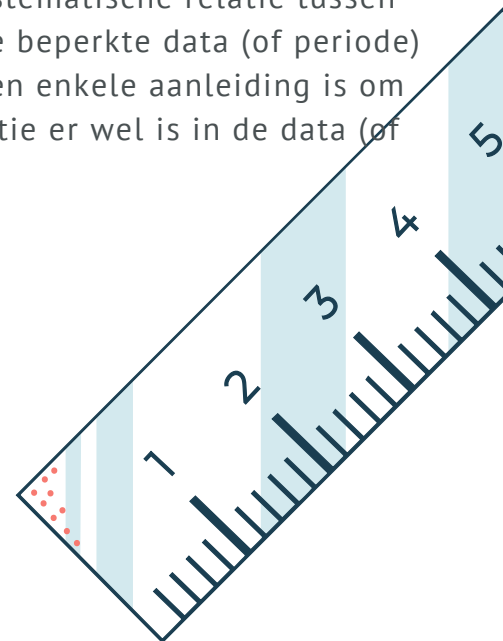
B. curvilineair

FIGUUR 2.

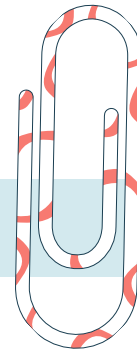
Relatie tussen de gemiddelde vaardigheidsscore en het aantal weken tussen de M-toetsen. Hier getoond voor groep 4-5, rekenen.

RESTRICTION OF RANGE

Toch hebben wij ook niet voor een curvilineaire correctie gekozen. We hebben namelijk te maken met een 'restriction of range'. Dat maakt corrigeren zeer risicovol, zelfs als je een perfecte relatie zou zien tussen afnamedatum en gemiddelde vaardigheidsscore. Je correctie blijft namelijk altijd gebaseerd op een beperkte hoeveelheid data over een korte tijdsperiode. Met andere woorden: omdat je van het coronacohort (in rood) nauwelijks waarnemingen hebt rond de 52 weken, weet je niet hoe de vaardigheidsscores eruit zouden zien als leerlingen de toets wél binnen de gewenste afnameperiode hadden gemaakt. En andersom: omdat geen van de leerlingen in het vergelijkingscohort (in blauw) de toets buiten de gewenste afnameperiode heeft gemaakt, kunnen we niet weten hoe zij hadden gescoord als zij ook extra leertijd hadden gehad. Aannames doen is gevaarlijk als ze vanuit data niet verantwoord kunnen worden. We zien geen enkele systematische relatie tussen afnamedatum en vaardigheidsscore in de beperkte data (of periode) die we wél hebben. Dat maakt dat er geen enkele aanleiding is om aan te nemen dat die systematische relatie er wel is in de data (of periode) die we niet hebben.



TOT SLOT

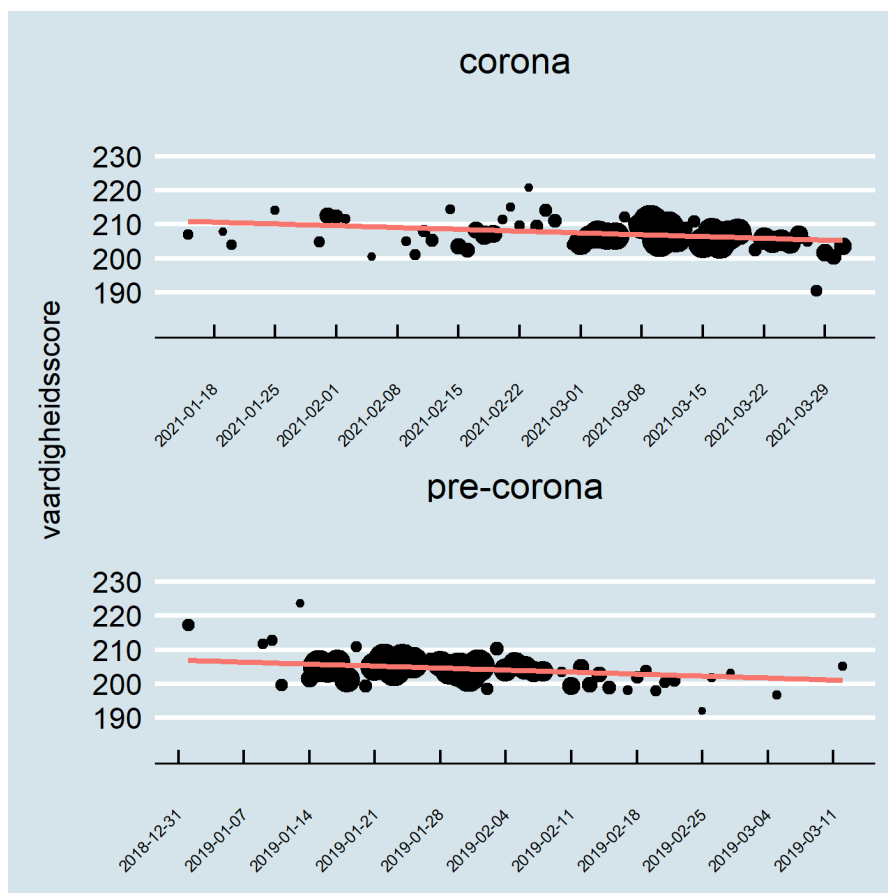


Wij hebben bewust besloten om de resultaten van leerlingen ten tijde van corona niet te corrigeren voor afnamedatum. Er bestaat namelijk géén verband tussen de afnamedatum en de (gemiddelde) vaardigheidsscore. Zouden we een correctie toepassen – welke dan ook – dan introduceren we dat verband. Dat is kunstmatig en niet te verantwoorden vanuit de gegevens waar wij mee werken. Echt verrassend is dit resultaat niet. Ook in normale tijden zijn er verschillen in afnamedatum en leertijd. Dat is logisch en geeft geen enkele aanleiding om aan correcties te denken. Uit onze analyses blijkt dat de verlate M-afname dit schooljaar die aanleiding ook niet geeft. Het is daardoor beter - en eerlijker - om gewoon te werken met de feitelijk verkregen scores van leerlingen. Zonder kunstmatige aanpassingen.

BIJLAGE 1

GEEN EFFECT LATERE AFNAMEDATUM REKENEN

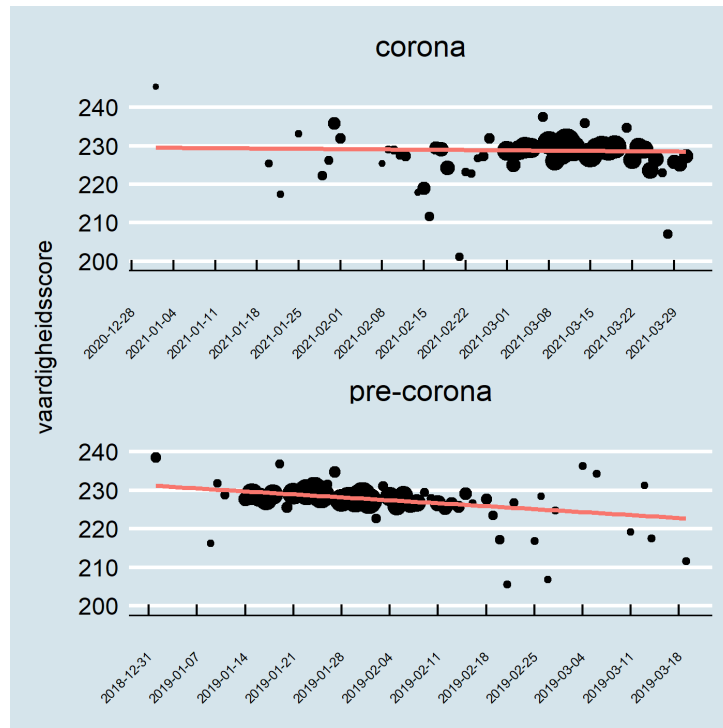
In het figuur hieronder is de gemiddelde vaardigheidsscore (y-as) afgezet tegen de afnamedatum van de M-toets (x-as), voor de leerlingen in groep 4-5 (Figuur 4 en 5 doen hetzelfde voor groep 5-6 en groep 6-7). Daarbij geven de punten de gemiddelde vaardigheidsscore weer over alle leerlingen die op die datum de toets hebben gemaakt. Hoe dikker de punt, des te meer leerlingen hebben op die datum de toets gemaakt. Wat blijkt nu uit deze figuren? Dat leerlingen die later de toets hebben gemaakt niet structureel hoger scoren dan leerlingen die de toets eerder hebben gemaakt, voor zowel de vergelijkingscohort (vóór corona) als de huidige coronacohort. Sterker nog, de oranje lijn die het verband tussen afnamedatum en vaardigheidsscore samenvat loopt zelfs iets naar beneden in plaats van omhoog.



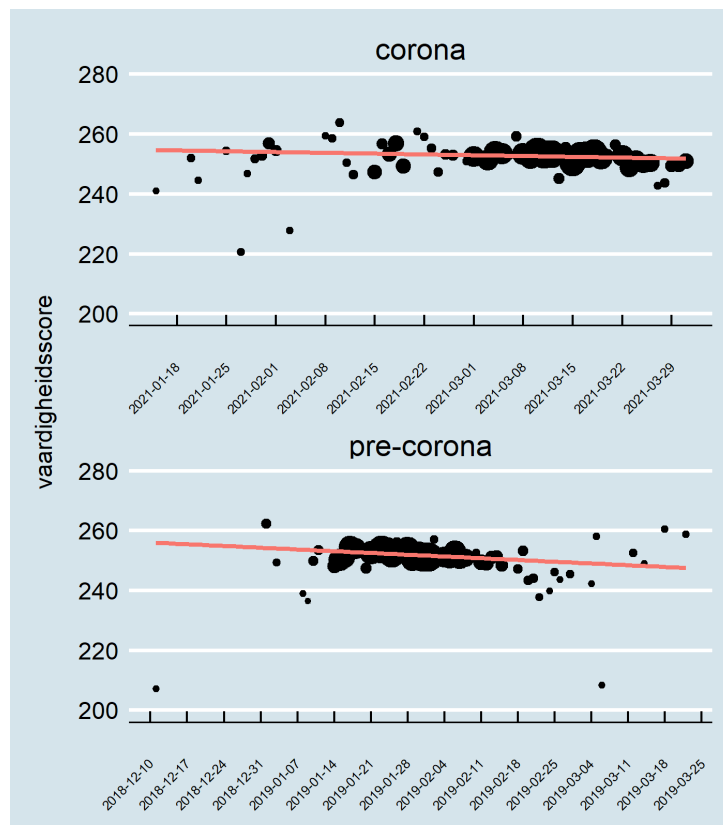
FIGUUR 3.

Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), rekenen, groep 4-5.

BIJLAGE 1

**FIGUUR 4.**

Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), rekenen, groep 5-6.

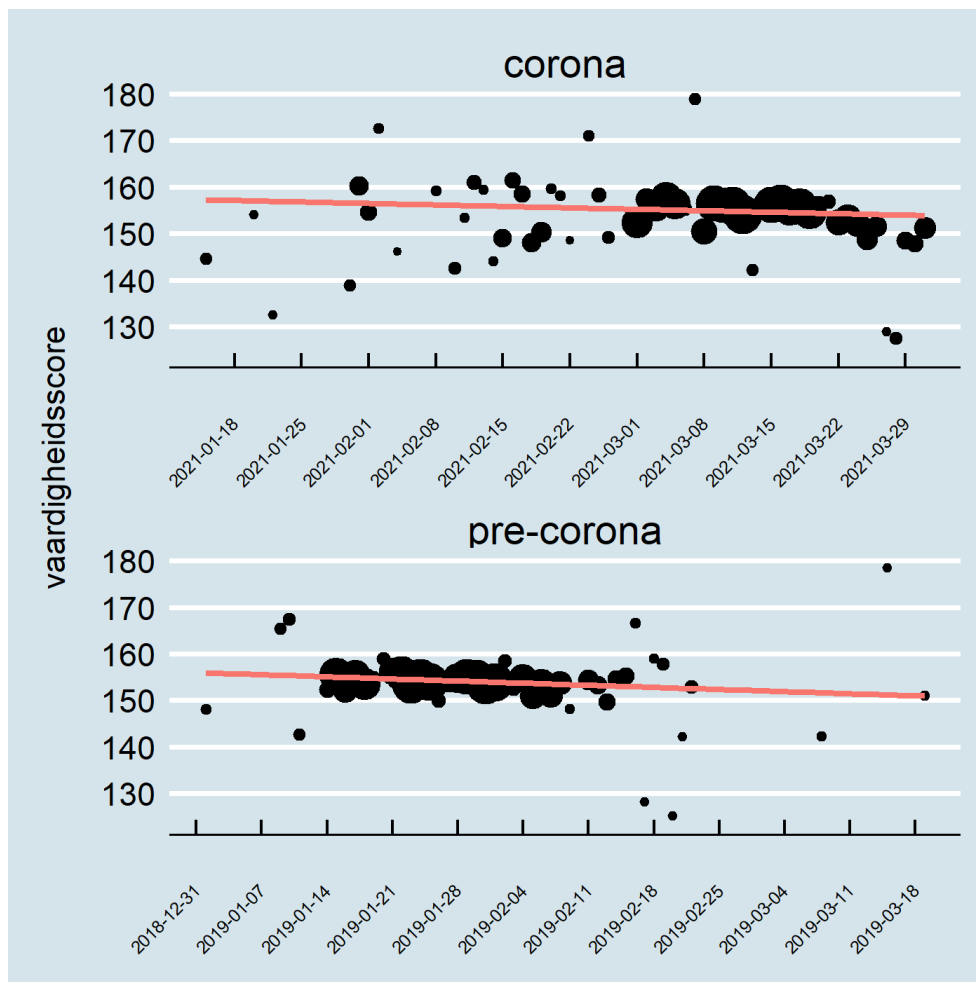
**FIGUUR 5.**

Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), rekenen, groep 6-7.

BIJLAGE 1

GEEN EFFECT LATERE AFNAMEDATUM BEGRIJPEND LEZEN

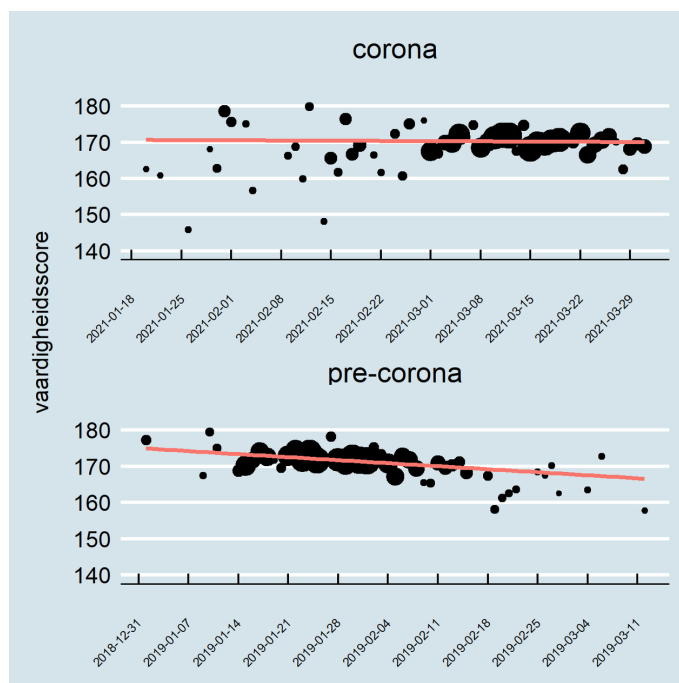
Hetzelfde beeld (dat er geen effect is van afnamedatum op de gemiddelde vaardigheidsscore) is er ook voor begrijpend lezen. Dat is te zien in het onderstaande figuur voor respectievelijk groep 4-5, groep 5-6 en groep 6-7.



FIGUUR 6.

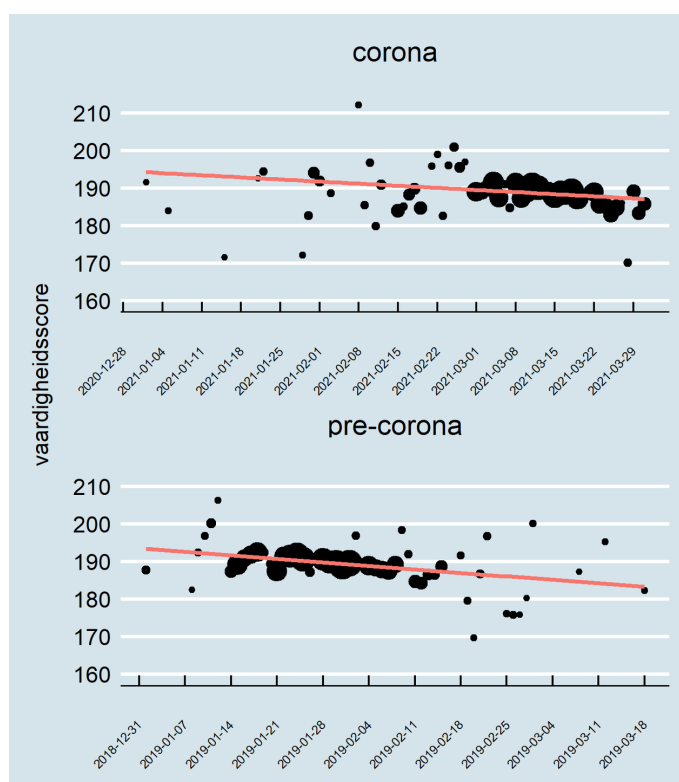
Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), begrijpend lezen, groep 4-5.

BIJLAGE 1



FIGUUR 7.

Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), begrijpend lezen, groep 5-6.



FIGUUR 8.

Geen verband tussen afnamedatum (x-as) en vaardigheidsscore (y-as), begrijpend lezen, groep 6-7.