

Rapport

Vaardigheid Examenkandidaten 2023



Vaardigheid

Examenkandidaten 2023

**Met medewerking van Karin Pilz, Hendrik Straat,
Harco Weemink**

Inhoud

Samenvatting	4
1 Inleiding	8
1.1 Aanleiding	8
1.2 Normering 2023	9
1.3 Onderzoeksvragen	10
1.4 Indeling van dit rapport	11
2 Methoden van onderzoek	14
2.1 Gegevens voor het onderzoek	14
2.2 Schatting van vaardigheidsontwikkeling	14
2.3 Berekeningen deelvragen	16
2.3.1 Deelvraag 1 effect van herkansingen op vaardigheid	16
2.3.2 Deelvraag 2 effect van kandidaten met lage scores	16
2.3.3 Deelvraag 3 spreiding tussen scholen	17
3 Resultaten	20
3.1 Vaardigheidsontwikkeling	20
3.2 Resultaten per deelvraag	22
3.2.1 Effect extra herkansingen op vaardigheid	22
3.2.2 Effect van kandidaten met lage scores	24
3.2.3 Spreiding tussen scholen	27
4 Conclusie	30
5 Context bij de resultaten	32
6 Literatuur	36
Appendix I Berekeningswijze vaardigheidsverschil tussen 2023 en pre-corona	40

Samenvatting

Net zoals in voorgaande jaren, hebben kandidaten die in 2023 eindexamen in het voortgezet onderwijs deden, te maken gehad met een andere voorbereiding op de centrale examens als gevolg van de coronapandemie. Daarnaast was er nog sprake van een aangepaste normering. Daarom is in 2023 ook weer onderzoek gedaan naar de vaardigheden van deze populatie ten opzichte van de populaties voor corona en die van 2021 en 2022.

In het jaar 2021 werden er verschillende maatregelen genomen om kandidaten tegemoet te komen tijdens de pandemie, waaronder het spreiden van examens over meerdere tijdvakken, extra herkansingsmogelijkheden en de optie om niet-kernvakken te schrappen bij de uitslagbepaling. Deze maatregelen werden in 2022 voortgezet. In 2023 gold voor leerlingen dat de exameneisen weer teruggingen naar het niveau van voor corona en er dus behalve van een verlengde tweede tijdvak en langere afnameperiodes voor de cbt's en cspe's geen extra maatregelen van kracht waren. Het was niet langer mogelijk een vak te schrappen, de examens te spreiden of een extra examen te herkansen.

De buitengewone omstandigheden in 2021 en 2022 maakten het normeren van de centrale examens op traditionele wijze technisch onmogelijk. Hierdoor reflecteerde de hoogte van de cijfers mogelijk niet precies de vaardigheid van de kandidaten. Daarom heeft de minister in 2021 het CvTE verzocht om na afloop van de centrale examens onderzoek te laten uitvoeren naar de daadwerkelijke vaardigheid van de kandidaten bij de centrale examens in het voortgezet onderwijs. Het onderzoek is in opdracht van het CvTE door Stichting Cito uitgevoerd. Het resultaat van dit onderzoek, uitgevoerd in 2021 en 2022, toonde aan dat er gemiddelde dalingen waren in vaardigheid, behalve voor het vak Engels. Deze bevindingen werden ondersteund door nationaal en internationaal onderzoek dat wees op leerachterstanden door de pandemie.

In 2023 kreeg het onderzoek naar de vaardigheden van kandidaten een vervolg, met een specifieke focus op het bepalen of de vaardigheden zich hebben hersteld naar het niveau van voor de pandemie. Dit rapport vergelijkt de gemiddelde vaardigheid van kandidaten in 2023 opnieuw met die van de periode voor de pandemie (2015-2019, vooral 2019) om te onderzoeken of de vaardigheden van examenkandidaten weer terug zijn op het niveau van voor de coronapandemie.

Bij clustering van de centrale examens op onderwijsniveau is er ten opzichte van voor de coronapandemie een vaardigheidsdaling te zien. Het verschil is echter alleen statistisch significant voor havo, omdat de betrouwbaarheidsintervallen voor alle andere niveaus de nullijn omvatten. De dalingen ten opzichte van voor corona zijn kleiner dan in 2022 (m.u.v. vmbo bb) en voor vwo en vmbo-gl/tl zelfs kleiner dan in 2021. Het verschil in vaardigheid tussen 2023 en de jaren tijdens de coronapandemie (2021 en 2022) is echter niet significant, omdat de betrouwbaarheidsintervallen overlappen.

Bij Engels is er net zoals in 2021 en 2022 sprake van een vaardigheidsstijging ten opzichte van voor de coronapandemie met ruim 0,1 cijferpunt. De overige inhoudelijke clusters van centrale examens laten een gemiddelde vaardigheidsdaling zien in de orde grootte van 0,2 tot 0,6 cijferpunt ten opzichte van voor corona. Hierbij moet opgemerkt worden dat dit gemiddelde vaardigheidsontwikkelingen per cluster van vakken beschrijft. De vaardigheidsdaling van 2023 vergeleken met 2022 ten opzichte van voor corona is in alle clusters behalve moderne vreemde talen kleiner. De verschillen zijn echter niet significant omdat betrouwbaarheidsintervallen van 2021, 2022 en 2023 voor alle clusters overlappen.

Net zoals in de rapporten van 2021 en 2022 is een aantal deelvragen bij de gevonden vaardigheidsontwikkelingen bestudeerd. Er is specifiek onderzocht welke invloed het afschaffen van de maatregelen heeft gehad op het percentage lage scores, het aantal herkansingen en de spreiding van scores over de scholen. Het aantal herkansers is niet alleen afgenomen in vergelijking met de jaren tijdens de coronapandemie, maar ook in vergelijking met 2019. Het aantal lage scores is in 2023 nog steeds hoger dan in 2019, maar lijkt gedaald ten opzichte van 2021 en 2022. De relatieve verschillen in vaardigheid tussen scholen zijn gemiddeld genomen niet veranderd.

Al met al toont dit onderzoek, met uitzondering van het inhoudelijke cluster Engels, dalingen in vaardigheid ten opzichte van de periode voor de coronapandemie. Hoewel de dalingen in het algemeen kleiner zijn dan tijdens de coronapandemie, zijn de verschillen met voor corona nog steeds middelgroot tot groot en groter dan reguliere jaarlijkse fluctuaties in gemiddelde vaardigheid per examen in pre-corona-omstandigheden.

1 Inleiding

1 Inleiding

1.1 Aanleiding

In het begin van 2020 werd Nederland, net zoals de rest van de wereld, getroffen door de uitbraak van de coronapandemie. Dit leidde tot een ongekenne situatie binnen het Nederlandse onderwijs. Op 15 maart 2020 werden de scholen in Nederland gesloten en op 24 maart volgde het bericht dat de centrale examens in het voortgezet onderwijs dat jaar geannuleerd werden. Gedurende het schooljaar 2020-2021 werd het onderwijs niet op de gebruikelijke manier gegeven. Onlineonderwijs, preventieve maatregelen om besmetting te voorkomen en slechts gedeeltelijke fysieke aanwezigheid van leerlingen werden normaal. De centrale examens van 2021 vonden wel plaats, met uitzondering van de centraal schriftelijke en praktische examens (cspe's) in het vmbo, die als schoolexamen werden afgenomen. Om de examenkandidaten een passende afsluiting van hun voortgezet onderwijs te bieden, werden diverse maatregelen genomen, waaronder het spreiden van examens over twee tijdvakken, het verstrekken van een extra herkansingsmogelijkheid en de optie om één niet-kernvak te laten vervallen bij de uitslagbepaling.

De verwachting was dat de vaardigheid van de eindexamenkandidaten van 2021 zou verschillen van die van de kandidaten die in de jaren voor de coronapandemie eindexamen deden. De gebruikelijke normeringswijze van de centrale examens kon niet zonder aanpassingen gehanteerd worden en was een aangepaste systematiek voor de normering noodzakelijk. Het was bijvoorbeeld onduidelijk of wel een representatief deel van de kandidaten in het eerste tijdvak zou deelnemen, of specifieke onderdelen in de voorbereiding wel evenveel aandacht hadden als in voorgaande jaren en of de centrale examens waarbij voorheen normhandavingsonderzoek werd gedaan in 2021 representatief zouden zijn voor alle andere centrale examens.

Bij de aangepaste systematiek kon niet worden gegarandeerd dat de prestatie-eisen voor het halen van een voldoende op een centraal examen gelijk zouden blijven aan de eisen in voorgaande jaren. Daarmee kon ook niet worden gegarandeerd dat de cijfers zonder meer te vergelijken zijn met die van voorgaande jaren, ofwel er kon niet worden gegarandeerd dat de cijfers de vaardigheid van de populatie getrouw zouden weergeven. De minister voor Basis- en Voortgezet Onderwijs en Media heeft het CvTE verzocht om na afloop van het eindexamen 2021 onderzoek te laten uitvoeren naar de vaardigheid van de examenkandidaten bij de centrale examens in het voortgezet onderwijs. Het CvTE heeft vervolgens Stichting Cito opdracht gegeven om onderzoek te doen naar de vaardigheidsontwikkeling bij de centrale examens, resulterend in een eerste rapport hierover.¹

Hoewel het schooljaar 2021-2022 hoopvol was gestart, met de verwachting dat het veel kandidaten zou lukken om eventuele vertraging in te lopen en zich beter voor te bereiden op het eindexamen in 2022, werd in het najaar duidelijk dat er toch maatregelen nodig waren. Hoewel scholen slechts één week – de laatste week voor de kerstvakantie – hoefden te sluiten, hadden veel scholen te kampen met fysieke lesuitval door ziekte en quarantaine van zowel leerlingen als leraren. Deze omstandigheden kwamen boven op de gevolgen van corona op het onderwijs die examenkandidaten in 2022 al tweeënhalf jaar hadden ervaren en hadden hun weerslag ook op de mentale gezondheid van kandidaten. Een goede voorbereiding op het examen en het inlopen van eventuele vertragingen werd daardoor moeilijker. Vandaar dat voor de centrale examens in

1 Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., & Emons, W. (2021). Vaardigheid examenkandidaten 2021.

2022, de maatregelen van 2021 opnieuw werden doorgevoerd.² Dit resulteerde in een tweede rapport hierover.³

Eind 2022 is besloten door de minister voor Primair en Voortgezet Onderwijs dat er geen maatregelen meer zouden worden genomen om kandidaten tijdens de afname van de examens in het schooljaar 2022/2023 tegemoet te komen, met uitzondering van de verlengde tweede tijdvakken en de langere afnameperiodes voor de cbt's en cspe's. Hoewel de coronapandemie in 2023 geen directe invloed meer had op de centrale examens, hebben examenkandidaten in de jaren voorafgaand aan deze examens andere omstandigheden ervaren dan kandidaten van voor de coronapandemie. Ook werd nog wel de aangepaste normeringssystematiek vanuit de coronajaren gehanteerd. Het onderzoek naar de vaardigheden van de kandidaten werd in opdracht van het CvTE door Stichting Cito in 2023 opnieuw uitgevoerd met dit rapport als resultaat. In 2023 gaat het onderzoek naar de vaardigheden van kandidaten verder en heeft als doel te bepalen of de vaardigheden van de examenkandidaten zich hebben hersteld naar het niveau van voor de pandemie. Het rapport zal de ontwikkeling van vaardigheden bij de centrale examens in 2023 in kaart brengen en vergelijken met de periode voor de pandemie. De opzet van het rapport is zoveel mogelijk hetzelfde gehouden als de rapporten van vorige jaren. Wanneer dit niet mogelijk is, lichten we de reden daarvoor toe. Net zoals in de afgelopen twee jaar zijn cspe's niet meegenomen in dit onderzoek, hoewel ze in 2023 zijn afgenomen. Omdat cspe's in 2021 en 2022 niet zijn afgenomen, zou een vergelijking met deze jaren dus niet mogelijk zijn. Ook zijn de profielvak-cspe's zoals die in 2023 werden afgenomen in 2019 voor het eerst afgenomen.

1.2 Normering 2023

Net als in 2021 en 2022 kende de normering van de centrale examens in 2023 een aangepaste systematiek⁴ ten opzichte van voor de coronapandemie.⁵ Dit zorgt ervoor dat de normering ook in 2023 recht doet aan de prestaties van kandidaten, waardoor kandidaten terecht een voldoende of onvoldoende blijven behalen, en de eisen die gesteld worden aan kandidaten zo veel mogelijk gelijk blijven aan de eisen van voorgaande jaren. Dit jaar hoefden we, in tegenstelling tot vorig jaar, geen vraag te stellen over de representativiteit van kandidaten uit een klas in het eerste tijdvak. Dit komt doordat de mogelijkheid om de centrale examens te spreiden over het eerste en tweede tijdvak is komen te vervallen in het schooljaar 2022/2023.

In eerste instantie werd een N-term bepaald waarbij de examenpopulatie voor een examen hetzelfde gemiddeld cijfer zou krijgen als voor corona. Wanneer alleen deze relatieve gegevens zouden worden gebruikt om te normeren, dan zou de lat niet automatisch op dezelfde hoogte liggen als voor corona. De examenkandidaten van 2023 moesten wel allemaal over dezelfde lat springen om een voldoende te halen. Om te zorgen dat de lat ook in 2023 zoveel mogelijk op dezelfde hoogte lag als in de jaren voor de coronapandemie, werd docenten gevraagd om ook een oordeel over de moeilijkheidsgraad van het examen te geven, net zoals in 2021 en 2022. Daarnaast werd de gebruikelijke moeilijkheid van het examen van dat vak, de zogenaamde historische N-term, beschouwd. Het docentoordeel en de historische N-term vormden zo twee instrumenten om te zorgen dat de lat zoveel mogelijk op dezelfde hoogte zou komen te liggen als in de jaren voor de coronapandemie.

2 Wiersma, D. (24-1-2022). [Kamerbrief over aanvulling besluit eindexamen voortgezet onderwijs 2022](#).

3 Pilz, K., Straat, H., Rietdijk, S., Kruis, J., & de Groot, M. (2022). [Vaardigheid examenkandidaten 2022](#)

4 [Normering centrale examens | 2023 | Examenblad.nl](#)

5 [N-term - Examenblad 2019](#)

1.3 Onderzoeksvragen

Stichting Cito verricht jaarlijks het zogenaamde normhandavingsonderzoek. In dit onderzoek worden de verschillen in moeilijkheidsgraad van de examens uit opeenvolgende jaren en de vaardigheidsontwikkelingen van de examenpopulatie in kaart gebracht. Deze gegevens zijn in dit onderzoek gebruikt om de vaardigheidsontwikkeling te bestuderen in alle vijf onderwijsniveaus bij minimaal zes centrale examens per niveau.

De hoofdvraag van dit onderzoek is of de vaardigheden van de examenkandidaten in 2023 weer op hetzelfde niveau zijn als voor de coronapandemie (2015-2019, met nadruk op 2019) en zijn gestegen ten opzichte van het niveau tijdens de coronapandemie (2021 en 2022).

Naar aanleiding van de rapporten van de afgelopen twee jaren tijdens de coronapandemie zijn ook de volgende deelvragen onderzocht:

Deelvraag 1 – effect van herkansingen op vaardigheid: In welke mate is er gebruik gemaakt van herkansingen, in vergelijking met voorgaande jaren? In de jaren tijdens de coronapandemie hadden examenkandidaten de mogelijkheid voor een extra herkansing. Omdat deze maatregel in 2023 niet meer van toepassing was, is een verandering in het aantal herkansingen ten opzichte van 2021 en 2022 te verwachten.

Deelvraag 2 – effect van kandidaten met lage scores: In welke mate is er sprake van extreem lage scores, in vergelijking met voorgaande jaren? En in hoeverre heeft dit de gemiddelde vaardigheid beïnvloed? De gemiddelde vaardigheid van examenkandidaten kan vertekend zijn door kandidaten met extreem lage scores. Er zijn meerdere mogelijke verklaringen voor lage scores: kandidaten zijn bijvoorbeeld onvoldoende voorbereid op het examen, zij kunnen een black-out tijdens het examen hebben gekregen, of zij hebben geen serieuze poging gedaan. Kandidaten kregen in 2021 en 2022 de mogelijkheid om één niet-kernvak weg te strepen bij de uitslagbepaling. De mogelijkheid om een niet-kernvak weg te strepen was in 2023 niet meer van toepassing, zodat een verschil in percentage lage scores in 2023 ten opzichte van 2021 en 2022 te verwachten is.

Deelvraag 3⁶ – spreiding tussen scholen: Hier wordt onderzocht of de verschillen in vaardigheden tussen scholen in 2023 afwijken van die voor de coronapandemie.

6 In de rapporten over de vaardigheden van 2021 en 2022, werden 4 deelvragen onderzocht. De vierde deelvraag – deelvraag 3 in de vorige rapporten - betrof de gemiddelde cijfer eerstekansers in het tweede tijdvak. Tijdens de coronapandemie konden de examenkandidaten hun examens spreiden over twee tijdvakken waardoor zij in het tweede tijdvak voor het eerst deelnemen aan een examen. Vandaar dat in de rapporten van 2021 en 2022 werd onderzocht in hoeverre de schatting van de vaardigheid van de kandidaten in het eerste tijdvak representatief was voor die van alle kandidaten. Omdat deze maatregel in 2023 niet meer van toepassing was, is deze deelvraag in dit rapport niet onderzocht.

1.4 Indeling van dit rapport

In hoofdstuk 2 zullen de methoden worden toegelicht waarmee de onderzoeksvragen beantwoord worden. Daarna volgen in hoofdstuk 3 de resultaten van de vaardigheidsschattingen en van de berekeningen naar aanleiding van de deelvragen. In de conclusie (hoofdstuk 4) worden de resultaten samengevat, en in hoofdstuk 5 worden de resultaten van context voorzien en vergeleken met ander onderzoek naar het effect van de maatregelen ten gevolge van de coronapandemie op onderwijsprestaties.

Het rapport kent één appendix waarin meer in detail wordt ingegaan op de gebruikte methode om vaardigheidsontwikkeling te schatten.



2 Methoden van onderzoek

2 Methoden van onderzoek

2.1 Gegevens voor het onderzoek

Om de vaardigheid van de examenpopulatie 2023 te onderzoeken, worden de gegevens gebruikt die Stichting Cito jaarlijks verzamelt voor het normhandavingsonderzoek. Het normhandavingsonderzoek is een schakel in het hele normeringsproces^{7,8} dat jaarlijks wordt uitgevoerd onder leiding van het CvTE. De gegevens betreffen de scores op de centraal schriftelijke examens die door docenten zijn ingevoerd in Wolf (uitwisselprogramma voor examengegevens⁹), de scores op de digitale centrale examens in vmbo-bb/kb die afkomstig zijn uit Facet (een digitaal systeem voor het afnemen van toetsen en examens), en gegevens die vooraf en tijdens de examencampagne worden verzameld om de moeilijkheidsgraad van centrale examens te onderzoeken. De gegevens van centrale examens met minder dan 50 kandidaten zijn niet meegenomen in dit onderzoek. Dit betreft voornamelijk maar niet uitsluitend de gegevens van de centraal schriftelijke examens in vmbo-bb en vmbo-kb.

2.2 Schatting van vaardigheidsontwikkeling

Om de hoofdvraag te beantwoorden is voor clusters van centrale examens, die deel uitmaken van het normhandavingsonderzoek, nagegaan of de gemiddelde vaardigheid van de examenpopulatie in 2023 verschilt van de gemiddelde vaardigheid van de examenpopulaties voor de coronapandemie (2015-2019 met nadruk op 2019) en tijdens de coronapandemie (2021 en 2022). Er zijn drie methodes gehanteerd om het verschil in vaardigheid van de examenkandidaten uit opeenvolgende jaren te onderzoeken, waarbij rekening wordt gehouden met het verschil in moeilijkheid tussen centrale examens¹⁰:

- Anchor in Package (AIP): deze methode wordt ingezet bij de digitale centrale examens vmbo-bb en kb. Als kandidaten in verschillende jaren dezelfde set opgaven voorgelegd krijgen, dan is een verschil in prestatie op deze set een belangrijke indicatie van een verschil in vaardigheid van de kandidaten uit de verschillende jaren.
- Pretest en posttest: volgens deze methodes krijgt een groep kandidaten in één schoolexamen zowel opgaven uit het actuele centrale examen voorgelegd als opgaven uit een ankerexamen. Door verschillende centrale examens met hetzelfde ankerexamen te vergelijken, kan de moeilijkheidsgraad van de centrale examens van verschillende jaren vergeleken worden. Een vergelijking van de prestaties van examenpopulaties op de verschillende centrale examens, gecorrigeerd voor een verschil in moeilijkheidsgraad, levert informatie over de vaardigheidsontwikkeling.
- Historische N-termen: volgens deze methode wordt de gemiddelde N-term van de vakken in 2023 met de gemiddelde N-term van dezelfde vakken in de historie vergeleken (2015-2019). We beschouwen de N-termen die we zouden krijgen bij de aanname dat de huidige populatie even vaardig is als populaties van 2019 en eerder. Als deze N-termen bij aanname van gelijke

7 College voor Toetsen en Examens (30 november 2016). De normeringssystematiek van de Centrale Examens VO. De normeringssystematiek van de centrale examens vo (examenblad.nl)

8 College van Toetsen en Examens (geen datum). Examenblad.nl: Normering centrale examens. <https://www.examenblad.nl/onderwerp/normering-centrale-examens>

9 Cito (geen datum). Tools en informatie voor docenten: Wolf. <https://www.cito.nl/onderwijs/voortgezet-onderwijs/centrale-examens-voortgezet-onderwijs/tools-en-informatie-voor-docenten/wolf>.

10 Van der Molen, P., Schouwstra, S., Feskens, R., & Van Onna, M. (2019). Vaardigheidsontwikkelingen volgens PISA en examens. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

populaties significant hoger zijn dan historisch gebruikelijk, is dit een indicatie dat de huidige populatie mogelijk minder vaardig is.

De drie methodes zijn niet even nauwkeurig¹¹. Anchor in Package (AIP) levert de meest nauwkeurige schatting op, gevolgd door pre-posttest. De historische N-term levert de minst nauwkeurige schatting op. Alle drie de methoden werken niet bij regelmatige wisselingen van onderwerpen in het examenprogramma, of bij de introductie van een nieuw examenprogramma. De historische N-term is ook niet bruikbaar als het examen zodanig veranderd is dat daardoor de moeilijkheidsgraad niet meer automatisch te vergelijken is met de moeilijkheidsgraad voor de verandering. AIP en pre- en posttest kunnen niet worden toegepast bij centrale examens met weinig examenkandidaten.

In Tabel 1 is te zien welke methoden zijn ingezet bij de verschillende examens. Tevens is te zien welke vakken in dit rapport zijn samengenomen in een cluster. In dit rapport zijn dezelfde vakken meegenomen als in 2021 en 2022.

Tabel 1 Gehanteerde aanvullende gegevens om vaardigheids- en moeilijkheidsverschillen te onderzoeken in 2022

Vak	Cluster	BB-CBT	KB-CBT	GL/TL	HAVO	VWO
Nederlands	Nederlands	AIP			pre/posttest	pre/posttest
Duits	Moderne vreemde talen*	histN	AIP	pre/posttest	pre/posttest	pre/posttest
Frans	Moderne vreemde talen			pre/posttest	pre/posttest	histN
Engels	Engels	AIP	AIP	pre/posttest	pre/posttest	pre/posttest
Economie	Mens & Maatschappij	AIP	AIP	pre/posttest	histN	
Geschiedenis	Mens & Maatschappij	histN	histN	histN	histN	histN
Aardrijkskunde	Mens & Maatschappij				histN	histN
Kunst (algemeen)	Mens & Maatschappij				histN	histN
Beeldende vakken	Mens & Maatschappij			histN		
Scheikunde/nask2	Exact			histN	pre/posttest	pre/posttest
Natuurkunde/nask1	Exact	histN	AIP	histN	pre/posttest	histN
Biologie	Exact	AIP	AIP	pre/posttest	pre/posttest	pre/posttest
Wiskunde A	Wiskunde				pre-posttest	histN
Wiskunde (B)*	Wiskunde	AIP	AIP	pre/posttest	pre/posttest	pre/posttest
Wiskunde C	Wiskunde					histN

Opmerking: AIP = Anchor in Package, histN = historische N-term

* exclusief Engels

** In vmbo bestaat het vak wiskunde zonder letter, in havo en vwo zowel wiskunde A als wiskunde B. In vwo bestaat ook wiskunde C.

11 Van der Molen, P., Schouwstra, S., Feskens, R., & Van Onna, M. (2019). Vaardigheidsontwikkelingen volgens PISA en examens. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.

De ontwikkeling in vaardigheid bij de examenpopulatie van 2023 wordt geschat door prestaties op de centrale examens van 2023 te vergelijken met die op eerdere centrale examens rekening houdend met verschillen in moeilijkheid. Om dit rapport te kunnen vergelijken met de twee voorafgaande rapporten, hebben we dezelfde methoden gebruikt en werden examens op dezelfde wijze geclusterd. De schatting van vaardigheid bevat onzekerheden. Door de informatie van meerdere centrale examens te combineren wordt deze onzekerheid verkleind en is het wél mogelijk een schatting van de trend te geven. In dit onderzoek zijn twee manieren van clustering toegepast. Enerzijds zijn centrale examens geclusterd die inhoudelijk gerelateerd zijn, en anderzijds zijn examens geclusterd per onderwijsniveau. Meer details omtrent deze berekeningen staan in Appendix I.

2.3 Berekeningen deelvragen

Voor de examenscores in dit rapport kijken we bij de deelvragen steeds naar de scores door de eerste corrector, die op de centrale examens behaald zijn. De bijbehorende cijfers zijn gebaseerd op de vastgestelde N-termen per centraal examen. Bij de centrale examens krijgen kandidaten scorepunten voor correcte antwoorden. Deze scorepunten worden omgezet naar cijfers. De basisregels voor het omzetten van scores in cijfers zijn vastgelegd in de ‘Regeling omzetting scores in cijfers VO’¹².

2.3.1 Deelvraag 1 | effect van herkansingen op vaardigheid

Voor een antwoord op de eerste deelvraag, worden voor alle centrale examens in vmbo-gl/tl, havo en vwo de herkansingspercentages vergeleken met die in 2019. Dit betreft dan per centraal examen het percentage van alle herkansers ten opzichte van alle eerstekansers. Bij de CBT-examens waren herkansers en de daarbij horende eerste afname moeilijk eenduidig te identificeren, omdat de afnames flexibel konden plaatsvinden. Daarom is deze deelvraag niet bestudeerd in vmbo-bb en vmbo-kb. In vmbo-gl/tl, havo en vwo is een grove inschatting gedaan van de toename in cijferpunt van de herkansende kandidaten om de invloed van de herkansingen op de vaardigheid te schetsen.

De herkansingspercentages zijn steeds berekend ten opzichte van het aantal kandidaten dat het vak in tijdvak 1 maakte, en niet ten opzichte van het totaal aantal kandidaten per onderwijsniveau. Centrale examens met relatief weinig kandidaten kunnen daarom een hoog percentage herkansers hebben. Schommelingen in aantal kandidaten dat het vak in tijdvak 1 maakte tussen 2023, 2022, 2021 en 2019 hebben met deze berekeningswijze geen invloed op de resultaten.

2.3.2 Deelvraag 2 | effect van kandidaten met lage scores

Om de tweede deelvraag te beantwoorden worden voor alle vakken uit het centraal examen 2023 van de vijf onderwijsniveaus de percentages lage scores vergeleken met de percentages in 2019. Een lage score wordt daarbij gedefinieerd als een score lager dan 25% van de maximale score. Het percentage werd weer per centraal examen uitgerekend, zodat het totaal aantal kandidaten dat aan het examen deelnam geen invloed had op de resultaten.

¹² Stcrt, 2016, 4817; Stcrt, 2021,18168: wetten.nl - Regeling - [Regeling omzetting scores in cijfers VO - BWBR0037590 \(overheid.nl\)](#)

Een gemiddelde vaardigheid wordt beïnvloed door de verdeling van vaardigheid in de populatie. Als deze disproportioneel meer lage scores krijgt, zakt de gemiddelde vaardigheid. Als er in 2023 andere percentages lage scores voorkomen dan in 2019, kan hiermee de geschatte vaardigheidsontwikkeling tussen verschillende jaren deels statistisch verklaard worden.

2.3.3 Deelvraag 3 | spreiding tussen scholen

De derde deelvraag is of de vaardigheidsontwikkeling gelijk is voor alle scholen. Om deze vraag te kunnen beantwoorden wordt per centraal examen een meerniveau-model¹³ toegepast. In een meerniveau-model zijn de scholen en de kandidaten daarbinnen de twee analyseniveaus. Per analyseniveau kijkt een meerniveau-model hoe groot de spreiding in cijfers is. Dus hoeveel gemiddelde cijfers van scholen onderling verschillen (schoolniveau), en hoeveel de cijfers van kandidaten binnen scholen van elkaar verschillen (leerlingniveau). De totale spreiding in de cijfers is een optelsom van de spreiding tussen scholen en de spreiding binnen scholen.

De intraklasse correlatie (ICC) is een gangbare maat voor de relatieve omvang van de verschillen tussen scholen ten opzichte van de totale spreiding¹⁴. Deze ICC zegt iets over de invloed van scholen op de cijfers van hun kandidaten, en kun je berekenen met een meerniveau-model. Voor de examens in Tabel 1 zijn ICC's berekend van zowel 2019, 2021, 2022 als 2023 om de vierde onderzoeksvraag te beantwoorden. Als de ICC van een vak tussen twee jaren gelijk zou zijn, dan is de vaardigheidsontwikkeling tussen die twee jaren min of meer gelijk geweest voor alle scholen.

13 Zie bijvoorbeeld Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling* (sec. ed.). London, Sage.

14 R-library lme4 is gebruikt om het meerniveau-model te schatten; de R-library performance om de ICC's te schatten.

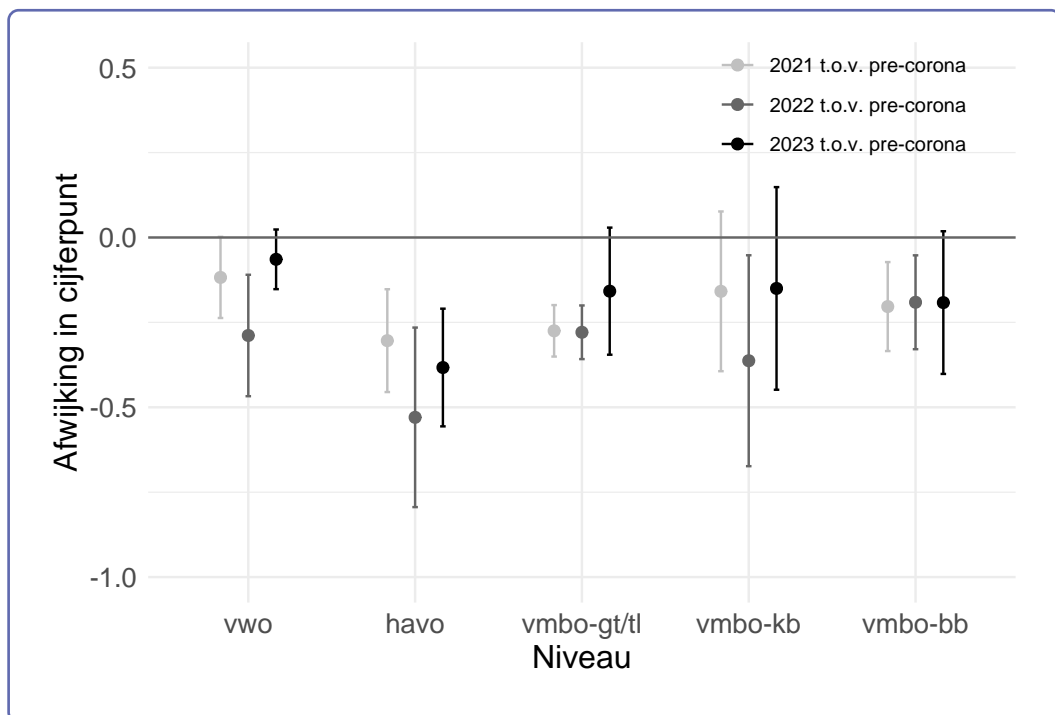
3 Resultaten

3 Resultaten

3.1 Vaardigheidsontwikkeling

Figuur 1 en Figuur 2 tonen op de horizontale as clusters van centrale examens, zoals in Tabel 1 gespecificeerd. Per cluster van centrale examens is een 90%-betrouwbaarheidsinterval weergegeven waarbinnen de gemiddelde vaardigheidsontwikkeling binnen dat cluster valt. Als de vaardigheid in 2023 (zwart), 2022 (donkergrijs) en 2021 (licht grijs), voor een bepaald cluster lager was dan voor de coronapandemie, dan ligt de puntschatting onder 0. De grootte van de vaardigheidsverschillen is uitgedrukt in cijferpunten. Dus een verschil van -0,2 betekent dat de populatie gemiddeld een vaardigheid had, die op de oude examens had geresulteerd in cijfers die gemiddeld 0,2 cijferpunt lager zouden zijn dan pre-corona. De lengte van een interval wordt enerzijds bepaald door de verschillen in vaardigheidsontwikkeling van de examens binnen een cluster. Hoe groter de verschillen in vaardigheidsontwikkeling tussen examens binnen een cluster, hoe groter het interval. Anderzijds wordt de lengte van het interval bepaald door de onzekerheid van de onderliggende schattingen per examen: bij gelijke vaardigheidsontwikkeling van alle onderliggende examens is bij een set examens waarbij naar de historische N-termen is gekeken een groter interval dan bij veel AIP-examens te verwachten.

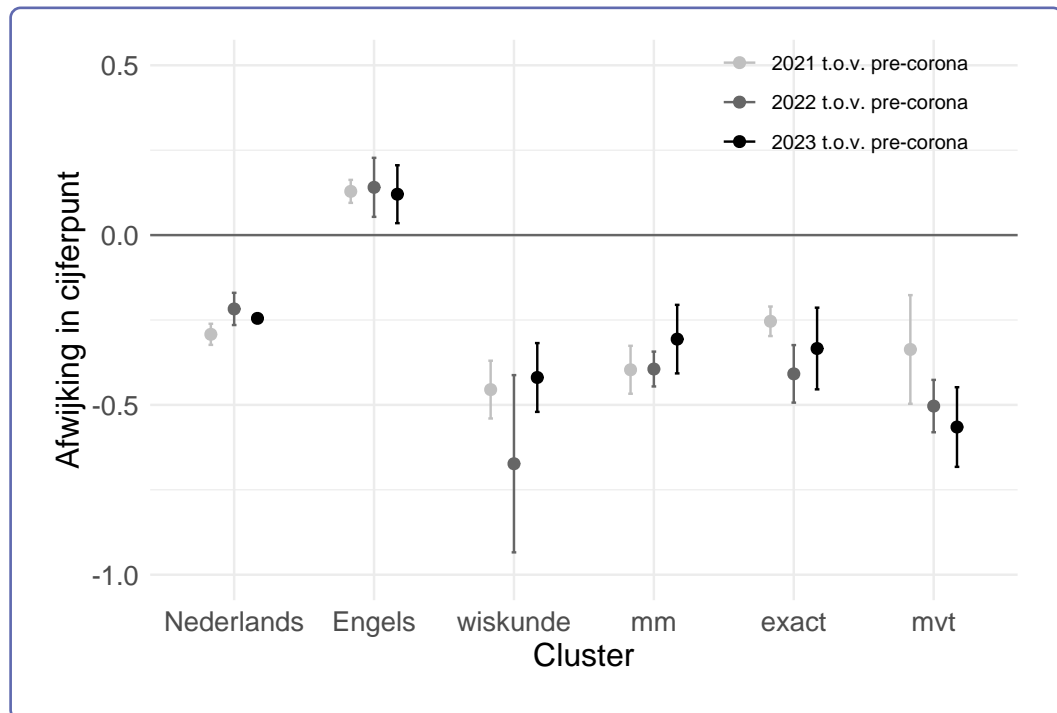
Figuur 1. Geschat vaardigheidsverschil per onderwijsniveau voor 2023 ten opzichte van voor corona (zwart), 2022 ten opzichte van voor corona (donkergrijs) en 2021 ten opzichte van voor corona (licht grijs)



Er is bij alle onderwijsniveaus in 2023 ten opzichte van voor de coronapandemie een daling in vaardigheid te zien. Het verschil is echter alleen statistisch significant voor havo, omdat bij alle andere niveaus de betrouwbaarheidsintervallen de nullijn omvatten. Het is ook te zien dat de dalingen kleiner zijn dan in 2022 (m.u.v. vmbo-bb) en voor vwo en vmbo-gl/tl zelfs kleiner dan in 2021. Het verschil tussen 2023 en 2022 en 2021 is echter niet significant, omdat de betrouwbaarheidsintervallen overlappen.

Bij inhoudelijke clustering van centrale examens (Figuur 2) is de gemiddelde vaardigheid in alle clusters in 2023 statistisch significant gedaald ten opzichte van voor corona met uitzondering van Engels, waar de vaardigheid is toegenomen. Dit resultaat is hetzelfde als in de afgelopen twee jaar. In vergelijking is de vaardigheidsdaling van 2023 ten opzichte van voor corona niet significant omdat betrouwbaarheidsintervallen van 2021, 2022 en 2023 voor alle clusters overlappen.

Figuur 2 Geschat vaardigheidsverschil per cluster voor 2023 ten opzichte van voor corona (zwart), 2022 ten opzichte van voor corona (donkergrijs) en 2021 ten opzichte van voor corona (licht grijs)



Opmerking: mm = mens en maatschappij, mvt = moderne en vreemde talen exclusief Engels; het betrouwbaarheidsinterval bij Nederlands is in dit onderzoek zodanig klein dat alleen de puntschatting te zien is.

In dit onderzoek kijken we met name naar de gemiddelde vaardigheidsontwikkeling. Een clustering van centrale examens helpt om te kijken of deze vaardigheidsontwikkeling op alle onderwijsniveaus of inhoudelijke clusters gelijk is. De verschillen tussen de clusters willen niet zeggen dat de examens binnen één cluster allemaal dezelfde vaardigheidsontwikkeling laten zien. Er is een diversiteit in vaardigheidsontwikkeling bij de examens binnen de clusters, al is dit deels beïnvloed door de onzekerheid waarmee vaardigheidsontwikkelingen geschat kunnen worden. In de analyses per deelvraag, kijken we naar de invloed van herkansingen en lage scores op de vaardigheidsontwikkelingen.

3.2 Resultaten per deelvraag

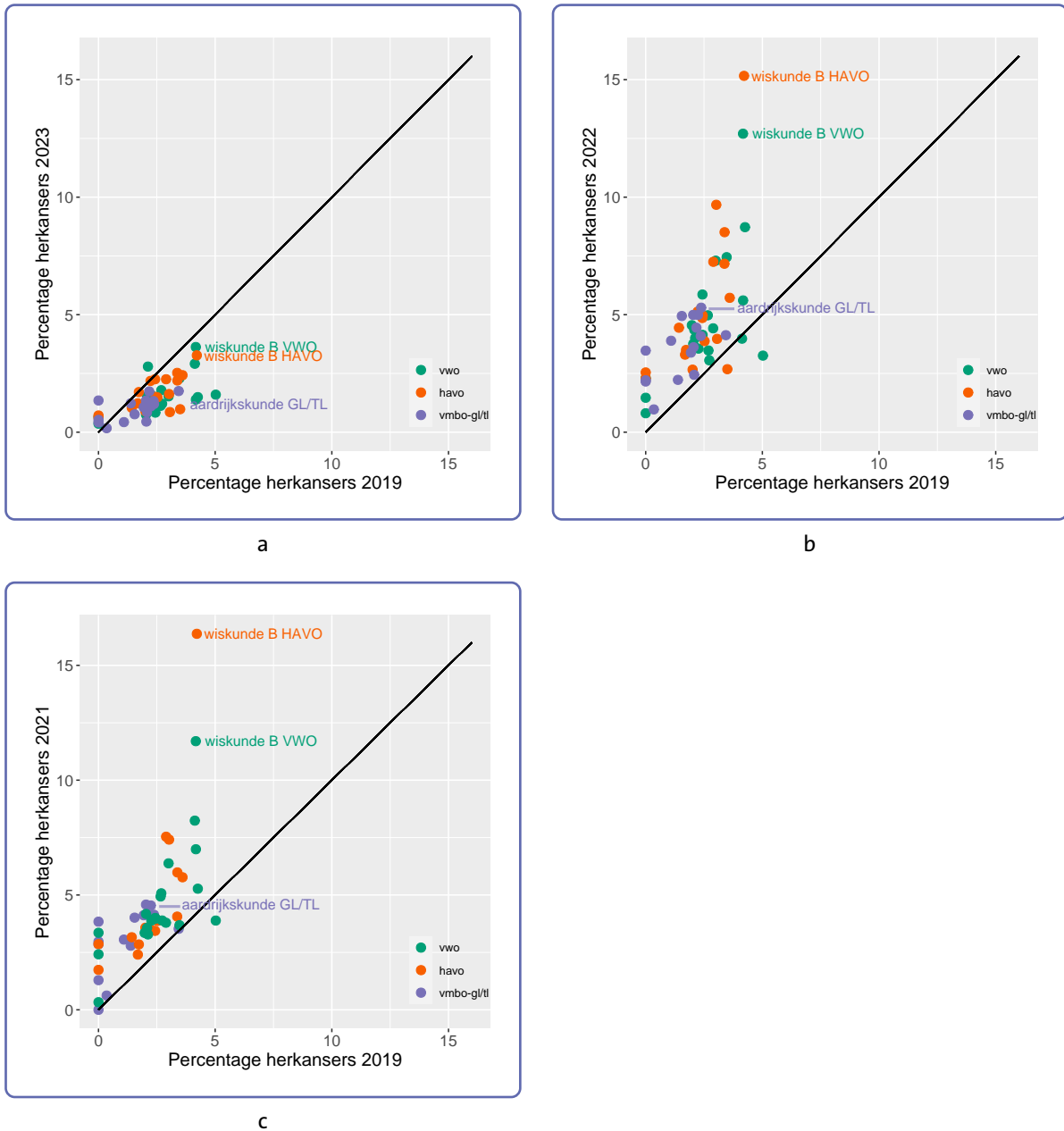
3.2.1 effect van herkansingen op vaardigheid

In Figuur 3 wordt op de horizontale as het percentage herkansers per centraal examen in 2019 weergegeven, terwijl op de verticale as het overeenkomstige percentage voor de jaren 2023 (Figuur 3a), 2022 (Figuur 3b) en 2021 (Figuur 3c) wordt getoond. De schuine lijn markeert de lijn waarop het percentage herkansers in 2019 gelijk is aan het percentage in respectievelijk 2023, 2022 of 2021.

De herkansingen in het tweede tijdvak leidden veelal tot verbetering van het cijfer in het eerste tijdvak. Gemiddeld hadden herkansende examenkandidaten zich bij herkansing in vmbo-gl/tl met 1 cijferpunt, in havo met 0,8 en in vwo met 0,5 cijferpunt verbeterd (+0,8 gemiddeld over alle onderwijsniveaus). Dit was gemiddeld over alle herkansende kandidaten van deze onderwijsniveaus en alle vakken.

Het valt op dat het percentage herkansers in 2023 vergeleken kan worden met, of zelfs iets lager is dan, het percentage in 2019. Dit vertegenwoordigt een omgekeerde trend in vergelijking met 2021 en 2022, toen het percentage herkansers in bijna alle vakken aanzienlijk hoger lag dan voor de coronapandemie. Het lage aantal herkansers in 2023 ten opzichte van voor corona heeft mogelijk te maken met het feit dat het in 2023 moeilijker was om herkansers in het systeem te herkennen, omdat voor een grotere groep leerlingen dan in 2019 het onduidelijk was of zij in het eerste tijdvak al het examen hadden gemaakt. Het percentage herkansers ligt in alle centrale examens onder de 4%, met de laagste percentages voor vmbo-gl/tl.

Figuur 3 Percentage herkansers per centraal examen in 2019 en 2023 (a), 2019 en 2022 (b) en 2019 en 2021 (c)

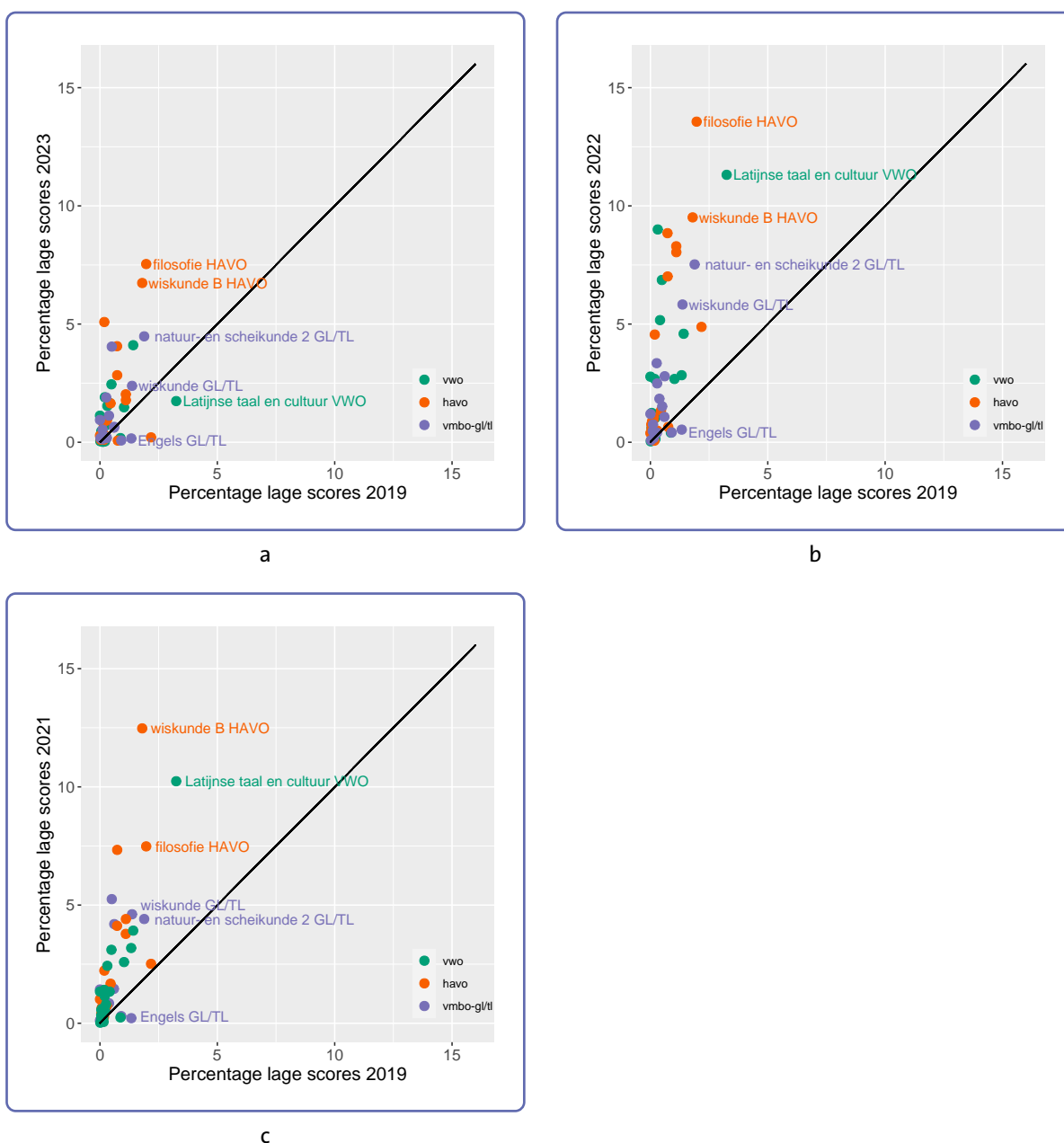


De invloed van deze cijfers op de vaardigheidsschattingen in Figuur 1 en Figuur 2 kunnen we in kaart brengen door een schatting te maken hoeveel hoger het gemiddeld cijfer in het eerste tijdvak zou zijn geweest wanneer de herkansingscijfers al in het eerste tijdvak was behaald. De vaardigheid van de examenpopulatie 2023 zou na herkansing gemiddeld zo'n $1,5\% \cdot 0,8$ cijferpunt = 0,012 cijferpunt hoger zijn dan voor de coronapandemie. Omdat de vaardigheid van de examenpopulatie 2023 geschat wordt ten opzichte van die in de jaren voor de coronapandemie, moet ook het effect van herkansingen in die jaren worden meegenomen. De vaardigheid van de populatie van 2019 zou ook stijgen na herkansing, gemiddeld zo'n $2\% \cdot 0,7$ cijferpunt = 0,014 cijferpunt. Het verschil met het herkansingseffect in 2023 is dan $0,012 - 0,014 = -0,002$ cijferpunt en dus te verwaarlozen. Daarmee zou het vaardigheidsverschil zoals gepresenteerd in Figuur 1 en Figuur 2 niet veel anders zijn geweest wanneer het effect van de extra herkansingen meegerekend zou worden.

3.2.2 Effect van kandidaten met lage scores

In Figuur 4 (vwo, havo en vmbo-gl/tl) en Figuur 5 (vmbo-bb, vmbo-kb) staan per centraal examen de percentages eerste-kansers met een lage score, dat wil zeggen een score lager dan 25% van de maximale score. Op de horizontale as staan de percentages kandidaten in 2019 en op de verticale as de percentages in 2023 (a), 2022 (b) en 2021 (c). De schuine lijn geeft aan wanneer het percentage in 2023, 2022 en 2021 gelijk was aan dat in 2019. De punten in Figuur 4 en Figuur 5 liggen grotendeels aan de linkerkant van de grafiek. Dit betekent dat lage scores voor de coronapandemie maar beperkt voorkwamen bij de centrale examens, gemiddeld bij minder dan 1% van de kandidaten, en hooguit 3% bij een enkel examen. Het is opvallend dat hoewel de lage scores voor Latijnse taal en cultuur vwo in 2021 en 2022 ten opzichte van voor corona nog redelijk hoog waren, deze in 2023 weer zijn gezakt.

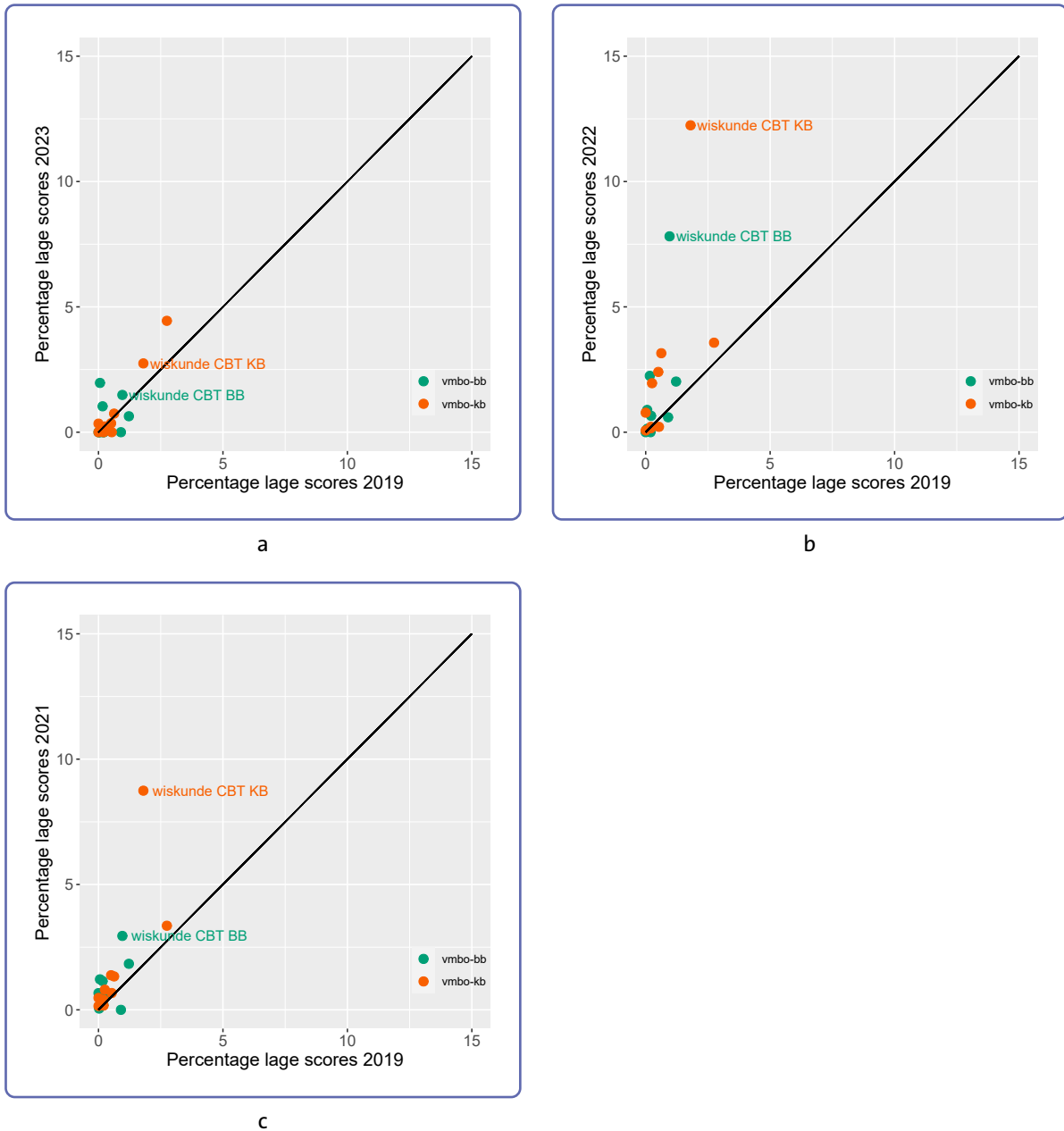
Figuur 4 Percentage eerste-kansers met een lage score per CSE in 2019 en 2022 (a) en 2019 en 2021 (b)



Bij vrijwel alle centrale examens voor vwo, havo en vmbo-gl/tl is het percentage kandidaten met een lage score (minder dan 25% van de maximale score) in 2023 groter dan in 2019 (Figuur 4a). De vakken met meer dan 5% lage scores zijn te vinden bij havo met filosofie (7.5% kandidaten met een lage score in 2023 tegenover 2% in 2019), wiskunde B (6.7 % kandidaten met een lage score in 2023 tegenover 2% in 2019) en economie (5% kandidaten met een lage score in 2023 tegenover 0% in 2019).

Voor een vijfde van alle examens gaat het om een verschil van meer dan 2%. Als 2% kandidaten een lage score haalt, betekent dit een daling van $(0,02 * 2 \text{ cijferpunten} =) 0,04$ cijferpunten. Als dit gemiddeld bij alle centrale examens in een cluster het geval is, dan zou dit effect voor ongeveer 0,04 cijferpunt hebben bijgedragen aan de geschatte vaardigheid van dat cluster. Als er in een cluster echter meerdere vakken voorkomen met een percentage lage scores hoger dan 2% dan zou de bijdrage daarvan aan de geschatte vaardigheid nog groter kunnen zijn dan 0,04 cijferpunt. Figuur 4 toont aan dat met name op havo-examenkandidaten relatief veel lage scores hebben behaald. Dit zou dus gedeeltelijk de daling in vaardigheid in het examencluster havo kunnen verklaren (Figuur 1).

Figuur 5 Percentage eerstekansers met een lage score per CBT in 2019 en 2023 (a), 2019 en 2022 (b) en 2019 en 2021 (c)



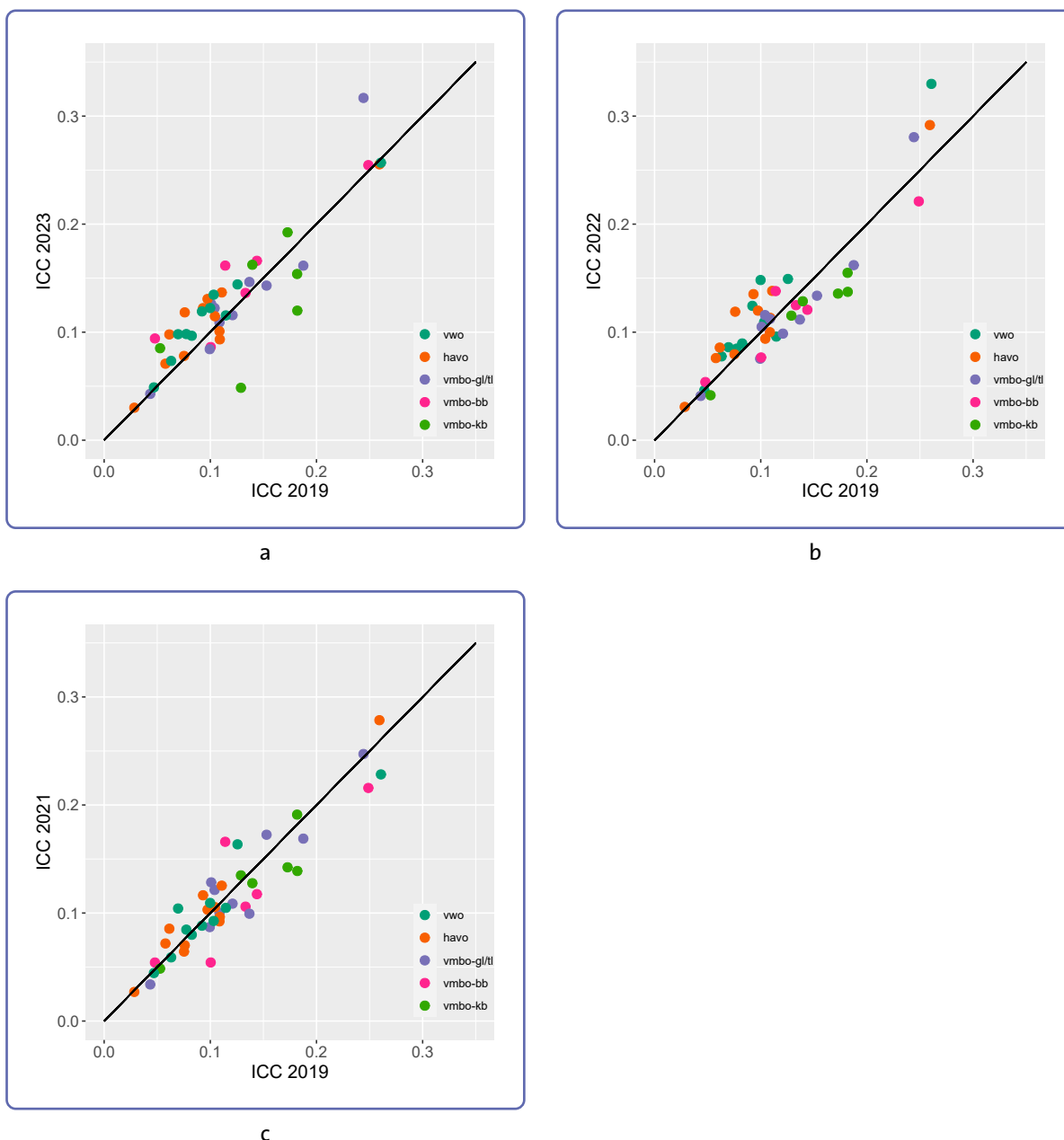
In vmbo-kb en vmbo-bb (Figuur 5) is er geen examen met meer dan 5% lage scores in 2023. Bovendien is het percentage lage scores voor de meeste examens weer in lijn met de percentages van voor de coronapandemie (Figuur 5a). Geschiedenis vmbo-kb heeft met 4% het hoogste aandeel aan lage scores, maar dit percentage lag ook al op 3% in 2019, voor de coronapandemie.

In directe vergelijking met de resultaten van 2022 (Figuur 4b en 5b) en 2021 ten opzichte van 2019 (Figuur 4c en 5c) is te zien dat het percentage lage scores in 2023 is afgenomen. Alleen twee van de vmbo-kb examens tonen een percentage lage scores hoger dan 2%. Daarom is het onwaarschijnlijk dat de percentages lage scores de vaardigheidsverschillen in vmbo-bb en vmbo-kb kunnen verklaren.

3.2.3 Spreiding tussen scholen

De derde deelvraag spitst zich toe op de vraag of de vaardigheidsontwikkeling gelijk was op alle scholen. Daartoe werd zowel voor de cijfers van 2023 als voor die van 2019, 2021 en 2022 een maat (ICC) berekend die de relatieve invloed van een school op de cijfers van haar kandidaten aangeeft.

Figuur 6 ICC (intraklasse correlatie) per centraal examen in 2019 en 2023 (a), 2019 en 2022 (b) en 2019 en 2021 (c)



Per centraal examen is de ICC op de cijfers van 2019 en 2023 (Figuur 6a), van 2019 en 2022 (Figuur 6b) en van 2019 en 2021 (Figuur 6c) aangegeven. Elk punt staat voor een centraal examen. De schuine lijn geeft aan wanneer de ICC in 2023, respectievelijk 2022 en 2021 gelijk was aan die in 2019. Opvallend is dat er verschillen in ICC zijn tussen centrale examens, maar

dat gemiddeld genomen er geen verschil is tussen ICC's in 2023 ten opzichte van die in 2019. De vaardigheidsontwikkeling lijkt dus in min of meer gelijke mate bij alle scholen te hebben plaatsgevonden, want de relatieve verschillen tussen scholen zijn niet structureel groter of kleiner geworden. We zien wel wat meer stippen boven de lijn maar het verschil is zodanig klein dat we geen structureel verschil in vaardigheidsontwikkeling tussen scholen zien.

4 Conclusie

4 Conclusie

De hoofdvraag van het onderzoek was of de vaardigheid van de kandidaten in 2023 weer terug is op het niveau van voor de coronapandemie (2015-2019 met nadruk op 2019). Daarnaast hebben we deze resultaten naast de vaardigheid van de examenpopulaties in 2021 en 2022 gelegd. De vaardigheid in 2023 is in alle onderwijsniveaus gedaald ten opzichte van voor de coronapandemie. Het verschil is echter alleen significant voor havo. De vaardigheid voor de clusters exact, mens en maatschappij (inclusief kunst), moderne vreemde talen (exclusief Engels), Nederlands en wiskunde is gedaald ten opzichte van voor corona. Daarbij vond de grootste gemiddelde daling plaats bij moderne vreemde talen. De gemiddelde vaardigheid in het cluster Engels is daarentegen gestegen.

In vergelijking met de vaardigheid in 2021 en 2022 tijdens de coronapandemie, is de punt-schatting van de vaardigheid van kandidaten in 2023 in alle niveaus gestegen, maar deze verschillen zijn niet statistisch significant. Een stijging is ook te zien voor de clusters mens en maatschappij, exact en wiskunde, terwijl het cluster moderne vreemde talen in 2023 een daling laat zien ten opzichte van 2021 en 2022. Ook deze verschillen zijn niet significant. Voor de clusters Nederlands en Engels verschilt de gemiddelde vaardigheid in 2023 nauwelijks met die in 2021 en 2022.

In 2023 waren er geen extra maatregelen voor examenkandidaten. De mogelijkheid een vak te schrappen bij de uitslagbepaling, de extra herkansing en de mogelijkheid de examens te spreiden waren niet meer van toepassing. Zo is te zien dat het aantal herkansingen weer op, of zelfs iets onder, het niveau van voor corona is gekomen. Vandaar dat de herkansingen dan ook geen noemenswaardig effect op de geschatte vaardigheid van de populatie hebben. Het aantal lage scores is nog steeds hoger dan voor de coronapandemie, maar is afgenomen in vergelijking met de periode tijdens de coronapandemie in 2021 en 2022. Het is mogelijk dat het aantal lage scores, vooral voor het cluster havo een klein effect zou kunnen hebben op de geschatte gemiddelde vaardigheden.

De laatste deelvraag betrof de spreiding tussen scholen. De relatieve verschillen tussen de scholen in cijfers zijn gemiddeld niet veranderd, wat erop duidt dat de vaardigheidsontwikkeling bij alle scholen in min of meer gelijke mate plaats vond.

Samenvattend laat de vaardigheid op onderwijsniveau gemiddeld een daling zien van rond de 0,1 à 0,3 cijferpunt in 2023 ten opzichte van voor de coronapandemie. Deze daling is alleen statistisch significant voor havo. Bij de inhoudelijke clusters van examens is sprake van een vaardigheidsdaling tussen de 0,2 en 0,6 cijferpunt, met uitzondering van Engels waar een vaardigheidsstijging van ongeveer 0,1 cijferpunt te zien is. Er is geen invloed van herkansende kandidaten of lage scores op deze schattingen. Scholen zijn niet meer van elkaar gaan verschillen in gemiddelde vaardigheid op het centraal examen dan voor de coronapandemie.



5 Context bij de resultaten

5 Context bij de resultaten

Met uitzondering van het cluster Engels, is het vaardigheidsniveau in 2023 bij de inhoudelijke clusters van centrale examens nog steeds 0,2 tot 0,6 cijferpunt lager ten opzichte van voor de coronapandemie. De vraag die rijst is hoe groot deze vaardigheidsdalingen nu eigenlijk zijn ten opzichte van jaarlijkse schommelingen. Het eerdere onderzoek van Stichting Cito naar de vaardigheidsontwikkelingen volgens PISA en centrale examens¹⁵ laat zien dat de resultaten op de centrale examens sinds 2015 vrij stabiel zijn, maar wel jaarlijks rond de 0,1 tot 0,2 cijferpunt fluctueren. Het valt op dat in 2023 voor vrijwel alle centrale examens een daling te zien is. Dus de fluctuaties zijn in 2023 bijna allemaal in dezelfde richting. Deze negatieve trend van lagere cijfers voor bijna alle vakken is vergelijkbaar met de trend van 2021¹⁶ en 2022¹⁷.

Een alternatieve manier om de omvang van een daling te beoordelen, is door een vergelijking te maken met de spreiding in cijfers tussen kandidaten per centraal examen voor de coronapandemie. Dit geeft een idee over het aantal kandidaten dat in 2019 per centraal examen lager scoorde dan het gemiddelde in 2023. In 2019 was de gemiddelde standaarddeviatie in cijfers ongeveer 1,1. Daarmee zou een daling van 0,2 tot 0,6 cijferpunt een middelgroot tot groot effect genoemd kunnen worden, volgens de interpretatie van effectmaten van Kraft¹⁸.

Omdat dit jaar geen maatregelen van kracht waren, behalve van een verlengd tweede tijdvak en langere afnameperiodes voor de cbt's en cspe's, was de verwachting dat we veranderingen zouden zien in het aantal herkansingen en in het aantal lage scores. Het aantal lage scores is inderdaad afgenomen ten opzichte van de coronajaren 2021 en 2022 maar is nog steeds hoger dan voor corona. Het aantal herkansingen is in dit onderzoek gedaald ten opzichte van de coronajaren en ook ten opzichte van voor corona. Dit is ook in overeenstemming met de data uit de examenmonitor¹⁹ die ook een daling in herkansingen in 2023 ten opzichte van voor corona laat zien.

De Examenmonitor laat ook zien dat het slagingspercentage in 2023 ten opzichte van voor corona is gedaald. Deze resultaten komen overeen met de constatering in dit rapport dat de vaardigheid op de centrale examens voor alle niveaus en vooral voor havo (Figuur 1), lager is ten opzichte van voor corona. Ook laat de examenmonitor zien dat het gemiddelde cijfer voor de centrale examens in 2023 ongeveer gelijk is aan dat in 2019. Echter, uit resultaten van dit vaardigheidsonderzoek blijkt dat de vaardigheid van de kandidaten gemiddeld genomen in 2023 lager is dan in 2019. In 2023 is een aangepaste normeringssystematiek toegepast, vergelijkbaar met die in 2021 en 2022. In deze tijdelijke systematiek is door het gebruik maken van bronnen als historische N-termen en docentoordeelen getracht om de lat zoveel als mogelijk op dezelfde plaats te houden als tot 2019. Omdat door de omstandigheden de beschikbare bronnen van normering iets minder betrouwbaar waren, hebben we ook het uitgangspunt gehanteerd dat leerlingen minimaal het cijfer zouden halen als ze in de jaren voor corona voor een gelijke prestatie gehaald zouden hebben, waar we normaal streven naar hetzelfde cijfer voor een gelijke prestatie. Dit zou het verschil kunnen verklaren tussen de gemiddelde cijfers

15 [Vaardigheidsontwikkelingen volgens PISA en examens | Rapport | Rijksoverheid.nl](#)

16 Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., en Emons, W. (2021). [Vaardigheid examenkandidaten 2021](#).

17 Pilz, K., Straat, H., Rietdijk, S., Kruis, J., & de Groot, M. (2022). [Vaardigheid examenkandidaten 2022](#)

18 Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>

19 [Examenmonitor Voortgezet Onderwijs 2023 | Rapport | Rijksoverheid.nl](#)

enerzijds en de vaardigheid van de populatie anderzijds. De N-termen zijn gemiddeld echter zo'n 0,0-0,1 cijferpunt hoger dan in 2019. Dit verschil had slechts een klein effect op de gemiddelde cijfers kunnen hebben. Daarom kan de aangepaste beoordelingssystematiek het verschil tussen cijfers en vaardigheden niet volledig verklaren.

Een daling van de vaardigheid of prestaties tussen de jaren 2019 en 2023 is in lijn met bevindingen in ander nationaal en internationaal onderzoek^{20,21,22,23,24,25,26}. Ten tijde van het schrijven van dit rapport zijn er echter nog geen studies over in hoeverre de vaardigheidsveranderingen tijdens corona in Nederland zijn ingelopen. Daarom is het nog niet mogelijk de resultaten van dit rapport in verband te brengen met resultaten van andere studies.

-
- 20 Lek, K., Feskens, R., Veldhuizen, J., en Keuning, J. (2021). Leervertragingen basisonderwijs deels ingelopen. Coronarapport 2 (cito.nl)
- 21 Zijlstra, W., Rietdijk, S., Frissen, S. en Feskens, R. (2021). Leervertraging door schoolsluitingen in het voortgezet onderwijs – deel 2. Title (cito.nl)
- 22 Rapport De Staat van het Onderwijs 2021 | Rapport | Inspectie van het onderwijs (onderwijsinspectie.nl)
- 23 Haelermans, C., Abbink, H., Baumann, S., Bijlsma, I., Havermans, W., Jacobs, M., Van Vugt, L., & Van Wetten, S. (2022). Balans na twee jaar: nog steeds vertraging in de leergroei, maar nadeel kwetsbare kandidaten verkleind. NCO Factsheet No. 2022-1. https://www.nationaalcohortonderzoek.nl/sites/nco/files/media-files/factsheet2022_01_nco_algemeen_2_jaar_leergroei_v2.pdf
- 24 OECD and Harvard Graduate School of Education (2020), A framework to guide an education response to the COVID-19 Pandemic of 2020. https://globaled.gse.harvard.edu/files/geii/files/framework_guide_v2.pdf
- 25 Kuhfeld, M., Soland, J. & Lewis, K (2022). Test Score Patterns Across Three COVID-19-impacted School Years. (EdWorkingPaper: 22-521). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/ga82-6v47>
- 26 Bundesministerium für Bildung und Forschung, Deutschland. Pressemitteilung: 36/2022. <https://www.bmbf.de/bmbf/shareddocs/pressemitteilungen/de/2022/05/100522-Aufholen-nach-Corona.html>

6 Literatuur

6 Literatuur

Bethhäuser, B. A., Bach-Mortensen, A., & Engzell, P. (2022). A systematic review and meta-analysis of the impact of the COVID-19 pandemic on learning. *SocArXiv*. April 20. <https://doi.org/10.31235/osf.io/g2wuy>

Bundesministerium für Bildung und Forschung, Deutschland. Pressemitteilung: 36/2022. <https://www.bmbf.de/bmbf/shareddocs/pressemitteilungen/de/2022/05/100522-Aufholen-nach-Corona.html>

Cito (geen datum). Tools en informatie voor docenten: Wolf. <https://www.cito.nl/onderwijs/voortgezet-onderwijs/centrale-examens-voortgezet-onderwijs/tools-en-informatie-voor-docenten/wolf>.

College voor Toetsen en Examens (30 november 2016). De normeringssystematiek van de Centrale Examens VO. [De normeringssystematiek van de centrale examens vo \(examenblad.nl\)](https://www.examenblad.nl),

College van Toetsen en Examens (2017). Examenblad.nl: Wat is Facet? <https://www.examenblad.nl/veel-gevraagd/wat-is-facet/2017/vwo?horizon=v41h1h4i9qe>

College van Toetsen en Examens (geen datum). Examenblad.nl: Normering centrale examens. <https://www.examenblad.nl/onderwerp/normering-centrale-examens>

DUO (2022). Examenmonitor VO 2022. Examenmonitor Voortgezet Onderwijs 2022 | Kamerstuk | Rijksoverheid.nl

Haelermans, C., Abbink, H., Baumann, S., Bijlsma, I., Havermans, W., Jacobs, M., Van Vugt, L., & Van Wetten, S. (2022). Balans na twee jaar: nog steeds vertraging in de leergroei, maar nadeel kwetsbare leerlingen verkleind. NCO Factsheet No. 2022-1. https://www.nationaalcohortonderzoek.nl/sites/nco/files/media-files/factsheet2022_01_nco_algemeen_2_jaar_leergroei_v2.pdf

Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>

Kuhfeld, M., Soland, J. & Lewis, K (2022). Test Score Patterns Across Three COVID-19-impacted School Years. (EdWorkingPaper: 22-521). Retrieved from Annenberg Institute at Brown University: <https://doi.org/10.26300/ga82-6v47>

Lek, K., Feskens, R., Veldhuizen, J., en Keuning, J. (2021). Leervertragingen basisonderwijs deels ingelopen. [Coronarapport 2 \(cito.nl\)](https://www.cito.nl)

[N-term - Examenblad 2019](https://www.examenblad.nl)

[Normering centrale examens 2022 - Examenblad](https://www.examenblad.nl)

OECD and Harvard Graduate School of Education (2020), A framework to guide an education response to the COVID-19 Pandemic of 2020. https://globaled.gse.harvard.edu/files/geii/files/framework_guide_v2.pdf

Pilz, K., Straat, H., Rietdijk, S., Kruis, J., en Groot de, M. (2022). Vaardigheid examenkandidaten 2022. <https://open.overheid.nl/documenten/ronl-b6cdb73c6a9a7bb7b555e468ec9b8d807bf97a69/pdf>

Rapport De Staat van het Onderwijs 2021 | Rapport | Inspectie van het onderwijs (onderwijsinspectie.nl)

Snijders, T.A.B. & Bosker, R.J. (2012). *Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling (sec. ed.)*. London, Sage.

Stcrt, 2016, 4817; Stcrt, 2021,18168: wetten.nl - Regeling - Regeling omzetting scores in cijfers VO - BWBR0037590 (overheid.nl)


Vaardigheidsontwikkelingen volgens PISA en examens | Rapport | Rijksoverheid.nl

Van der Molen, P., Schouwstra, S., Feskens, R., & Van Onna, M. (2019). *Vaardigheidsontwikkelingen volgens PISA en examens*. Arnhem: Stichting Cito Instituut voor Toetsontwikkeling.' https://www.nro.nl/sites/nro/files/migrate/Cito_Vaardigheidsontwikkelingen_volgens_PISA_en_examens_maart_2019.pdf

Van Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., en Emons, W. (2021). *Vaardigheid examenkandidaten 2021*. https://www.examenblad.nl/system/files/2022/Rapport_Cito_Effect-Corona_vaardigheid_CE_2021.pdf

Wiersma, D. (24-1-2022). *Kamerbrief over aanvulling besluit eindexamen voortgezet onderwijs 2022*. <https://open.overheid.nl/repository/ronl-cb3f1b3803a593c4353b99c906b2be9434fc36a0/1/pdf/aanvulling-besluit-eindexamen-voortgezet-onderwijs-2022.pdf>

Zijlstra, W., Rijtdiek, S., Frissen, S. en Feskens, R. (2021). *Leervertraging door schoolsluitingen in het voortgezet onderwijs – deel 2*. <https://open.overheid.nl/documenten/ronl-31ef6390-f8c1-4e6d-b676-80f8a276d632/pdf>



**Appendix I |
Berekeningswijze
vaardigheids-
verschil tussen
2023 en pre-corona**

Appendix I | Berekeningswijze vaardigheidsverschil tussen 2023 en pre-corona

Onderliggend aan Figuur 1 en Figuur 2 zijn berekeningen uitgevoerd. In deze appendix worden de details daarvan besproken. Het doel was om de prestaties op de examens in 2023 te relateren aan de prestaties in de jaren voor de coronapandemie (2019) en tijdens de coronapandemie (2021 en 2022).

Gebruikte data

Voor de schatting van de vaardigheid van 2021, 2022 en 2023 zijn alleen de kandidaten meegenomen die deelnamen aan het eerste tijdvak, en voor de digitale examens alleen de kandidaten die t/m 9 juni 2021 respectievelijk t/m 3 juni 2022, respectievelijk t/m 30 mei hebben deelgenomen. De eerstekansers in het tweede tijdvak zijn buiten beschouwing gelaten omdat er geen geschikte equivaleringsgegevens beschikbaar waren voor het tweede tijdvak. Omdat het aantal eerstekansers bij het tweede tijdvak substantieel lager was, is de aanname dat de prestaties van alle eerstekansers niet veel verschillen van de kandidaten die opgenomen zijn in deze analyses. Met één van de deelvragen wordt onderzocht of deze kandidaten van het eerste tijdvak representatief zijn voor alle kandidaten.

Om de pre-corona-vaardigheid te schatten werd gekozen welke afnamejaren de pre-corona-conditie het best weerspiegelen. Een schatting van vaardigheid wordt beter naarmate deze op meer afnamegegevens gebaseerd is. Dan moeten deze afnamegegevens wel een weerspiegeling zijn van eenzelfde vaardigheid. Het laatste pre-corona-jaar was 2019. Bij de examens met een AIP, pre-testen of posttesten is gekozen om altijd 2019 als pre-corona-jaar te gebruiken. Voor de examens met een historische N-term zijn de trends berekend op basis van de historische N-termen van 2015-2019 mits er in die jaren geen sprake was van een aanpassing van het examenprogramma en/of de syllabus en de moeilijkheidsgraad van de examens niet doelbewust is aangepast.

Methoden

In Tabel 1 wordt aangegeven welke centrale examens in de berekening voor de vaardigheidsschattingen per niveau en cluster zijn meegenomen. Ook wordt daar aangegeven welke methode (Anchor in package, pre- en posttest of historische N-term) gebruikt is. Bij Anchor in package en pre- en posttest examens is een trendschattingwijze gebruikt om de vaardigheid voor en na corona te vergelijken. Bij de historische N-term is op een andere manier de daling in cijfer tussen pre-corona en post-corona berekend. In alle gevallen resulteerde een verschil in gemiddeld cijfer tussen de pre-corona-populatie en de populatie in 2023, met daarbij een schatting van de nauwkeurigheid van dit verschil.

Trendschatting op basis van AIP, pre- en posttest

Bij een trendschatting zijn alle afnamegegevens van 2015 tot en met 2023, plus eventuele pre- of posttesten opgenomen in één IRT-analyse. Daarvoor moest de moeilijkheid van de opgaven in deze afnamejaren wel aan elkaar gerelateerd kunnen worden via direct hergebruik van opgaven (Anchor in package) of via hergebruik van ankeropgaven in pre- of posttestboekjes. De IRT-analyse²⁷ leverde naast parameters die de moeilijkheid van de opgaven weergaven, ook schattingen van vaardigheid op de latente trek. Deze vaardigheden op de latente trek werden vervolgens uitgedrukt als scores op een (grote) set van items. Dus iedere kandidaat in de periode 2015 – 2023 had ‘virtueel’ een score behaald op dezelfde mega-toets; een score tussen 0 en de maximum score op de totale set van items.

Vervolgens werd per centraal examen een representatieve pre-corona-periode aangewezen, zoals hierboven beschreven bij de gebruikte data. Zowel van de kandidaten in de pre-corona-periode werd de gemiddelde score berekend als van de kandidaten in 2023. Een daling in gemiddelde scores tussen deze twee gemiddelden reflecteerde dus een daling in vaardigheid. Het verschil in gemiddelde scores van 2023 en de pre-corona-periode werd uitgedrukt in een verschil in gemiddeld cijfer²⁸ om de getallen beter te kunnen duiden, en om centrale examens te kunnen aggregeren.

De puntschatting van dit zogenaamde contrast werd voorzien van een nauwkeurigheidsmaat²⁹. In de berekening van deze nauwkeurigheidsmaat werd rekening gehouden met de nauwkeurigheid van de schatting van het IRT-model op de data. Bij een centraal examen met relatief veel hergebruik van opgaven en grotere aantallen afnames (Anchor in package), is deze nauwkeurigheid groter dan bij centrale examens waarbij hergebruik beperkt is tot kleine ankerstetjes in pre- of posttest boekjes met relatief weinig afnames.

Historische N-term

Bij vakken waarbij we gebruik maken van de historische N-term vergelijken we de gemiddelde N-term van de vakken in 2023 wanneer we aan zouden nemen dat de kandidaten even vaardig zijn als voor de coronapandemie met de gemiddelde N-term van dezelfde vakken in de historie (2015-2019). Ook wordt de standaarddeviatie bepaald van deze 5 jaargemiddelden. Zo gaat ieder cluster met een eigen verschil in N-term en meetfout mee in de analyses.

Examen-overstijgende berekening

Omdat de nauwkeurigheid bij ieder centraal examen niet bijzonder hoog was, en er bij ieder examen jaarlijks wel schommelingen voorkomen in vaardigheid, zijn de resultaten van meerdere centrale examens geaggregeerd om een vaardigheidsschatting te doen. Dit was mogelijk omdat

27 Vanwege het grote aantal centrale examens zijn er geen uitgebreide analyse op modelfit gedaan.

Dit is eveneens een reden waarom de afzonderlijke trendschattingen niet gepresenteerd worden.

28 Uit de gebruikelijke manier waarop scores in cijfers worden omgezet, namelijk met een N-term, volgt dat een verschil in cijfer gelijk is aan een verschil in score vermenigvuldigd met een constante. Deze constante is gelijk aan 9 gedeeld door de maximale score.

29 Standard Error, deze SE is omgekeerd gerelateerd aan de nauwkeurigheid: een grotere SE wijst op minder nauwkeurigheid.

alle examen-specifieke verschillen op dezelfde cijferpuntschaal waren uitgedrukt. Enerzijds zijn er berekeningen per onderwijsniveau uitgevoerd (Figuur 1), anderzijds per inhoudelijk cluster van centrale examens (Figuur 2). De clustering van de centrale examens voor de examen-overstijgende berekeningen staat weergegeven in Tabel 1.

De examen-overstijgende schatting van het verschil in vaardigheid tussen de pre-corona- en 2023-populaties werd berekend als een gewogen gemiddelde van de examen-specifieke verschillen D_k in vaardigheid,

$$\bar{D}_{overstijgend} = \frac{\sum w_k D_k}{\sum w_k}$$

waarbij de gewichten w_k de nauwkeurigheid van de examen-specifieke schatting weergaven³⁰. Omdat de examen-specifieke verschillen onderling kunnen variëren, is de variatie daartussen gebruikt om een betrouwbaarheidsinterval rondom het examen-overstijgende gemiddelde te berekenen. Ook hierbij is rekening gehouden met de gewichten/nauwkeurigheid van de examen-specifieke verschillen.

$$V(\bar{D}_{overstijgend}) = \frac{1}{(\sum w_k)^2} \sum w_k^2 (D_k - \bar{D}_{overstijgend})^2$$

Het 90%-betrouwbaarheidsinterval in Figuur 1 en Figuur 2 loopt van

$$\bar{D}_{overstijgend} - 1,645 * \sqrt{V(\bar{D}_{overstijgend})} \text{ tot } \bar{D}_{overstijgend} + 1,645 * \sqrt{V(\bar{D}_{overstijgend})}$$

³⁰ $w_k = 1 / (SE_k)^2$

© Copyright 2023

Dit is een uitgave van divisie
Centrale toetsen en examens,
Stichting Cito.

Alle rechten voorbehouden.
Niets uit dit werk mag zonder
voorafgaande schriftelijke
toestemming van Stichting Cito
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.

Cito.nl



CITO CTE