

Rapport

Vaardigheid Examenkandidaten 2022



Vaardigheid

Examenkandidaten 2022

**Met medewerking van Karin Pilz, Hendrik Straat,
Silvia Rietdijk, Joost Kruis, Maarten de Groot**

Inhoud

Samenvatting	4
1 Inleiding	8
1.1 Aanleiding	8
1.2 Normering 2022	9
1.3 Onderzoeksvragen	10
1.4 Indeling van dit rapport	10
2 Methoden van onderzoek	12
2.1 Gegevens voor het onderzoek	12
2.2 Schatting van vaardigheidsontwikkeling	12
2.3 Berekeningen deelvragen	15
2.3.1 Deelvraag 1 effect van extra herkansing op vaardigheid	15
2.3.2 Deelvraag 2 effect van kandidaten met lage scores	15
2.3.3 Deelvraag 3 effect van eerstekansers in tweede tijdvak	16
2.3.4 Deelvraag 4 spreiding tussen scholen	16
3 Resultaten	18
3.1 Vaardigheidsontwikkeling	18
3.2 Resultaten per deelvraag	20
3.2.1 Effect extra herkansingen op vaardigheid	20
3.2.2 Effect van kandidaten met lage scores	21
3.2.3 Effect van eerstekansers in tweede tijdvak	23
3.2.4 Spreiding tussen scholen	25
4 Conclusie	28
5 Context bij de resultaten	32
6 Literatuur	36
Appendix I Berekeningswijze vaardigheidsverschil tussen 2022 en pre-corona	40

Samenvatting

Net zoals in 2021 hebben kandidaten die in 2022 eindexamen in het voortgezet onderwijs deden, een andere voorbereiding gehad dan voor de coronapandemie, die voor lockdowns, online onderwijs en quarantainesituaties zorgde. Bij het centrale examen in 2021 werden er vanwege de pandemie meerdere maatregelen genomen om kandidaten een haalbare afsluiting van hun voortgezet onderwijs te bieden. Deze maatregelen werden voor het centrale examen in 2022 opnieuw doorgevoerd. Zo konden kandidaten hun centrale examens spreiden over twee tijdvakken, een extra herkansing doen en een vak wegstrepen bij de uitslagbepaling zolang dit geen kernvak betrof. Ook kregen kandidaten de mogelijkheid om de afname van de centrale examens te spreiden over het eerste en het tweede tijdvak.

In 2021 was het technisch niet mogelijk om de examens op de reguliere wijze te normeren. 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 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. In het onderzoek werd de gemiddelde vaardigheid van 2021 vergeleken met de gemiddelde vaardigheid voor de coronapandemie (periode 2015-2019, met nadruk op 2019). Het onderzoek liet gemiddelde dalingen in vaardigheid zien, met uitzondering van de onderwijsniveaus vwo en vmbo-kb (waar de daling niet significant was) en het inhoudelijk cluster Engels (waar sprake was van een stijging). Deze dalingen zijn overal in lijn met bevindingen in ander nationaal en internationaal onderzoek dat leerlingen door de schoolsluitingen een leerachterstand hebben opgelopen.

Het onderzoek naar de daadwerkelijke vaardigheden van de kandidaten werd in opdracht van het CvTE door Stichting Cito in 2022 voortgezet en dit rapport is het resultaat.

Om over jaren heen de prestatie-eisen zo goed mogelijk te handhaven, zijn extra onderzoeken nodig, de zogenaamde normhandhavingsonderzoeken. Deze onderzoeken geven inzicht in de ontwikkeling van de vaardigheid van cohorten kandidaten uit verschillende jaren. De uitkomsten van deze onderzoeken kennen een zekere mate van onzekerheid, die dit jaar, net als in 2021, als gevolg van de omstandigheden groter was dan in andere jaren. Deze onzekerheid verkleint wanneer meerdere onderzoeken worden samengenomen. In dit rapport worden daarom alleen uitspraken gedaan over de vaardigheidsontwikkeling bij clusters van centrale examens. In dit onderzoek zijn twee manieren van clustering toegepast. Enerzijds zijn centrale examens geclusterd die inhoudelijk gerelateerd zijn, en anderzijds zijn centrale examens geclusterd per onderwijsniveau.

In dit onderzoek is de gemiddelde vaardigheid van 2022 vergeleken met de gemiddelde vaardigheid voor de coronapandemie (periode 2015-2019, met nadruk op 2019). De resultaten van deze vergelijking zijn naast de resultaten van 2021 gezet. Bij clustering van de centrale examens is er ten opzichte van voor de coronapandemie een daling van de gemiddelde vaardigheid te zien bij alle onderwijsniveaus. Bij Engels is sprake van een stijging met ruim 0,1 cijferpunt. De overige inhoudelijke clusters van centrale examens laten een gemiddelde vaardigheidsdaling zien in de orde van 0,2 tot 0,7 cijferpunt. Merk op dat dit gemiddelde ontwikkelingen zijn. De vaardigheidsdaling van 2022 ten opzichte van voor corona is in het cluster Nederlands kleiner dan in 2021 ten opzichte van voor corona en groter in het cluster exact. In de overige clusters is geen significante vaardigheidsverandering te zien ten opzichte van 2021. Wel is de vaardigheidsdaling ten opzichte van voor corona voor de onderwijsniveaus vwo en vmbo-kb in 2022 significant waar dat in 2021 niet het geval was.

Een aantal deelvragen die helpen bij de duiding van de gevonden vaardigheidsontwikkelingen is nader onderzocht. De belangrijkste bevinding hieruit is dat de daling in vaardigheid bij sommige clusters, zoals bijvoorbeeld wiskunde, deels verklaard zou kunnen worden door de toename van lage scores voor vakken binnen deze clusters. Andere deelvragen toonden vooral aan dat de mogelijkheid van een extra herkansing en de mogelijkheid van het spreiden van de afname van centrale examens over het eerste en tweede tijdvak geen of weinig impact hebben op de geconstateerde vaardigheidsdalingen. Daarnaast constateerden we dat de relatieve verschillen tussen scholen in vaardigheid gemiddeld niet zijn veranderd. Wat opviel bij het beantwoorden van de deelvragen was dat bij vrijwel alle centrale examens kandidaten die hun eerste afname hadden uitgesteld tot het tweede tijdvak, gemiddeld lagere cijfers haalden dan de kandidaten die al in het eerste tijdvak het centraal examen hadden gedaan. Dit verschil liep bij een aantal examens op tot meer dan 1,3 cijferpunten. Ook de toename ten opzichte van 2021 in lage scores viel op.

Al met al toont dit onderzoek, met uitzondering van Engels, gemiddeld dalingen in vaardigheid ten opzichte van voor de coronapandemie. Deze dalingen mogen volgens statistische criteria middelgrote tot grote dalingen genoemd worden. De dalingen zijn groter dan reguliere jaarlijkse fluctuaties in gemiddelde vaardigheid per examen in pre-corona-omstandigheden.

1 Inleiding

1 Inleiding

1.1 Aanleiding

Begin 2020 werd Nederland, net zoals de rest van de wereld, getroffen door de coronapandemie. Het Nederlands onderwijs kwam in een ongekende situatie terecht. Op 15 maart 2020 sloten de scholen in Nederland en op 24 maart kwam het bericht dat de centrale examens in het voortgezet onderwijs in 2020 geen doorgang zouden vinden. Ook in schooljaar 2020-2021 werd er geen onderwijs gegeven op de reguliere manier. Online onderwijs, maatregelen om besmetting binnen de school te voorkomen en slechts gedeeltelijke fysieke aanwezigheid waren gebruikelijk. 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 konden worden afgenomen. Daarbij werden er meerdere maatregelen genomen om kandidaten een haalbare afsluiting van hun voortgezet onderwijs te bieden. Zo konden kandidaten hun centrale examens spreiden over twee tijdvakken, een extra herkansing doen en één vak, zijnde een niet-kernvak, wegstrepen bij de uitslagbepaling (zogenoemde duimregeling).

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. Daarvoor waren drie hoofdredenen:

- 1 Het was onduidelijk of wel een representatief deel van de kandidaten in het eerste tijdvak zou deelnemen.
- 2 Het was onduidelijk of specifieke onderdelen in de voorbereiding wel evenveel aandacht hadden als in voorgaande jaren.
- 3 Het was onduidelijk of de centrale examens waarbij voorheen normhandhavingsonderzoek werd gedaan in 2021 wel representatief zouden zijn voor alle andere centrale examens.

Vanwege deze redenen was een aangepaste systematiek voor de normering noodzakelijk. Bij deze aangepaste systematiek kon niet worden gegarandeerd dat de prestatie-eisen voor een voldoende (hierna: 'de lat') 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 van 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 hierover¹.

Hoewel het schooljaar 2021-2022 hoopvol was gestart, met de verwachting dat het veel leerlingen 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 nieuwe maatregelen nodig zouden zijn. Hoewel scholen slechts één week – de laatste week voor de kerstvakantie – hoefden te sluiten, hadden veel scholen te kampen met lesuitval door ziekte en quarantaine van zowel

1 Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., & Emons, W. (2021). Vaardigheid examenkandidaten 2021. https://www.examenblad.nl/document/rapport-cito-effect-corona/2021/f=/Rapport_Cito_Effect-Corona_vaardigheid_CE_2021.pdf

leerlingen als leraren. Deze omstandigheden kwamen bovenop 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. Een goede voorbereiding op het examen en het inlopen van eventuele vertragingen werd daardoor moeilijker. Vandaar dat voor de centrale examens in 2022, de maatregelen van 2021 opnieuw werden doorgevoerd².

Het onderzoek naar de vaardigheden van de kandidaten werd in opdracht van het CvTE door Stichting Cito in 2022 opnieuw uitgevoerd en dit rapport is het resultaat. Het brengt de vaardigheid van de examenkandidaten bij de centrale examens van 2022 in beeld, in vergelijking met de jaren 2015-2019 (met nadruk op 2019) en 2021.

1.2 Normering 2022

Net als in 2021 kende de normering van de centrale examens een andere systematiek³ dan in de jaren voor de coronapandemie⁴ zodat de normering ook in 2022 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. Er werd dezelfde informatie opgevraagd bij docenten als in 2021. Docenten werd bijvoorbeeld gevraagd of hun kandidaten, die deelnamen aan het eerste tijdvak, qua vaardigheid gelijk waren aan de groep die in het tweede tijdvak voor het eerst aan dat centraal examen zou gaan deelnemen, en daarmee representatief voor al hun kandidaten. Met behulp van deze antwoorden werd een representatieve subpopulatie van scholen gevormd. Een relatieve N-term werd bepaald waarbij de representatieve scholen gemiddeld een gelijk cijfer zouden krijgen als in voorgaande jaren.

Wanneer alleen deze relatieve gegevens zouden worden gebruikt om te normeren, dan zou de lat niet automatisch op dezelfde hoogte liggen als in voorgaande jaren. De examenkandidaten van 2022 moesten wel allemaal over dezelfde lat springen om een voldoende te halen. Om te zorgen dat de lat ook in 2022 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. Daarnaast werd de N-term vergeleken met de gebruikelijke moeilijkheid van het examen van dat vak, de zogenaamde historische N-term. 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. <https://www.rijksoverheid.nl/documenten/kamerstukken/2022/01/24/aanvulling-besluit-eindexamen-voortgezet-onderwijs-2022>

3 Normering centrale examens 2022 - Examenblad

4 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. Bij zowel de hoofd- als de deelvragen gaan we ook na hoe de resultaten zich verhouden tot de resultaten die we vonden in 2021.

De hoofdvraag van dit onderzoek is hoe de vaardigheid van de examenkandidaten in 2022 zich verhoudt tot de vaardigheid van de examenpopulaties van voor de coronapandemie (2015-2019, met nadruk op 2019).

Om deze vraag goed te kunnen beantwoorden, zijn de volgende deelvragen ook onderzocht:

Deelvraag 1 | effect van extra herkansing op vaardigheid: In welke mate is er vaker gebruik gemaakt van herkansingen in vergelijking met voorgaande jaren? En welke invloed had dit op de vaardigheidsschatting? De gemiddelde vaardigheid van kandidaten, en daarmee hun cijfer, kan na herkansing anders uitvallen. Herkansingen kunnen in 2022 een andere invloed op de uiteindelijke vaardigheid van de populatie hebben gehad dan in voorgaande jaren.

Deelvraag 2 | effect van kandidaten met lage scores: In welke mate is er sprake van 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 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. Omdat kandidaten in 2022 net zoals in 2021 de mogelijkheid kregen om één niet-kernvak weg te strepen bij de uitslagbepaling, zouden lage scores in 2022 vaker voor kunnen komen. De geschatte vaardigheid kan in die gevallen een onderschatting zijn van de werkelijke vaardigheid.

Deelvraag 3 | effect van eerstekansers in tweede tijdvak: In hoeverre was de schatting van de vaardigheid van de kandidaten in het eerste tijdvak representatief voor die van alle kandidaten? De gemiddelde vaardigheid wordt berekend op de kandidaten in het eerste tijdvak. Kandidaten konden net zoals in 2021 ook in 2022 in het tweede tijdvak voor het eerst deelnemen aan een examen.

Deelvraag 4 | spreiding tussen scholen: Is de vaardigheidsontwikkeling gelijk voor alle scholen? Of zijn de verschillen tussen scholen relatief groter geworden door de coronapandemie?

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 een 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 2022 te onderzoeken, worden de gegevens gebruikt die Stichting Cito jaarlijks verzamelt voor het normhandhavingsonderzoek. Het normhandhavingsonderzoek is een schakel in het hele normeringsproces^{5,6} 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 examengegevens⁷), 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 examens⁸), 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 normhandhavingsonderzoek, nagegaan of de gemiddelde vaardigheid van de examenpopulatie in 2022 verschilt van de gemiddelde vaardigheid van de examenpopulaties voor de coronapandemie (2015-2019 met nadruk op 2019) en tijdens de coronapandemie (2021). 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 examens⁹:

- 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 school-examen 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.

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

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

7 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>

8 College van Toetsen en Examens (2017). Examenblad.nl: Wat is Facet? <https://www.examenblad.nl/veelgevraagd/wat-is-facet/2017/vwo?horizon=vq41h1h4i9qe>

9 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.cito.nl/kennis-en-innovatie/kennisbank/cte-rapport-vaardigheidsontwikkelingen-volgens-pisa-en-examens>

- Historische N-termen: volgens deze methode wordt de gemiddelde N-term van de vakken in 2022 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 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 nauwkeurig¹⁰. 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 het rapport over de vaardigheid van examenkandidaten in 2021 is gebruik gemaakt van standaardbepalingen als een methode om de vaardigheid te schatten voor vakken waarbij geen AIP-, pre- of posttest-methode mogelijk was. Standaardbepalingen kennen een grote onnauwkeurigheid en zijn daarnaast een behoorlijk arbeidsintensieve methode om aanvullende gegevens te verzamelen. Daarom is voor dit rapport besloten om in plaats van standaardbepalingen over te stappen op historische N-termen als methodiek om de verandering in vaardigheidsniveau in te schatten. De betrouwbaarheid van de historische N-term is vergelijkbaar met die van een standaardbepaling. Doordat we van meer vakken een historische N-term hebben kan dat wel de betrouwbaarheid van de schatting van de vaardigheidsverandering in een cluster van vakken vergroten. 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. Er zijn meer vakken die meegenomen zouden kunnen worden in een vaardigheidsschatting op basis van een historische N-term. Deze zijn in dit rapport niet allemaal meegenomen om de vergelijkbaarheid van de schatting voor 2022 met 2021 te kunnen waarborgen.

¹⁰ 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.cito.nl/kennis-en-innovatie/kennisbank/cte-rapport-vaardigheidsontwikkelingen-volgens-pisa-en-examens>

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	histN
Duits	Moderne vreemde talen*	histN	AIP	pre/posttest	pre/posttest	pre/posttest
Frans	Moderne vreemde talen			histN	pre/posttest	histN
Engels	Engels	AIP	AIP	pre/posttest	pre/posttest	pre/posttest
Economie	Mens & Maatschappij	histN	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	histN
Natuurkunde/nask1	Exact	AIP	AIP	histN	pre/posttest	pre/posttest
Biologie	Exact	AIP	AIP	pre/posttest	histN	histN
Wiskunde A	Wiskunde				pre-posttest	histN
Wiskunde (B)**	Wiskunde	AIP	AIP	histN	pre/posttest	histN
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.

De ontwikkeling in vaardigheid bij de examenpopulatie van 2022 wordt geschat door prestaties op de centrale examens van 2022 te vergelijken met die op eerdere centrale examens rekening houdend met verschillen in moeilijkheid. Om dezelfde redenen waarom het niet mogelijk was om de gangbare normeringssystematiek te hanteren in 2022, is het niet mogelijk een schatting van de vaardigheidsontwikkeling te doen voor ieder centraal examen afzonderlijk. 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 vaardigheid 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'¹¹.

In reguliere examenjaren leggen kandidaten een centraal schriftelijk examen af in het eerste tijdvak, het tweede tijdvak is voor herkansingen. Net zoals in 2021, mochten kandidaten ook in 2022 kiezen in welk tijdvak ze een examen voor de eerste keer afleggen. In dit rapport heten kandidaten die een examen voor de eerste keer maken eerstekansers. Kandidaten die een herexamen maken, heten herkansers. Bij de digitale centrale examens (vaak aangeduid met CBT-examens, waarbij CBT staat voor Computer Based Testing) in vmbo-bb en vmbo-kb konden kandidaten hun examens maken in de periode van 4 april tot en met 22 juli. In dit onderzoek zijn voor de digitale examens de afnamegegevens van kandidaten tot 3 juni meegenomen. Deze kandidaten worden allen als eerstekansers beschouwd. Het is dezelfde steekproef van kandidaten die ten grondslag lag aan de normering die bekendgemaakt is op 14 juni.

2.3.1 Deelvraag 1 | effect van extra herkansing op vaardigheid

Voor een antwoord op de eerste deelvraag, naar de omvang en het effect van herkansingen, worden voor alle centrale examens in vmbo-gl/tl, havo en vwo de herkansingspercentages vergeleken met die in 2019 en 2021. 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 als eerstekanser 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 eerstekansers bij een vak tussen 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 naar de invloed van lage scores op de vaardigheidsschatting te beantwoorden worden voor alle vakken uit het centraal examen 2022 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.

11 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 2022 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 | effect van eerstekansers in tweede tijdvak

Om de derde deelvraag te beantwoorden, of de vaardigheid van kandidaten in het eerste tijdvak representatief is voor die van alle kandidaten, wordt gekeken of er sprake is van selectieve deelname in het eerste tijdvak. Enerzijds wordt gekeken hoeveel kandidaten centrale examens hebben uitgesteld tot het tweede tijdvak, anderzijds wat de gemiddelde cijfers in het eerste tijdvak waren in vergelijking met de gemiddelde cijfers van eerstekansers in het tweede tijdvak. Ook dit onderwerp is niet voor vmbo-bb en vmbo-kb bestudeerd, omdat herkansers bij de CBT-examens moeilijk te identificeren waren.

2.3.4 Deelvraag 4 | spreiding tussen scholen

De vierde deelvraag is of de vaardigheidsontwikkeling gelijk is voor alle scholen. Om deze vraag te kunnen beantwoorden wordt per centraal examen een meerniveau-model¹² 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 spreiding¹³. Deze ICC zegt iets over de invloed van scholen op de cijfers van hun kandidaten, en kun je berekenen met een meerniveau-model. Om de vierde onderzoeksvraag te beantwoorden zijn voor de examens in Tabel 1 ICC's berekend van zowel 2019, 2021 als 2022. 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.

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

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

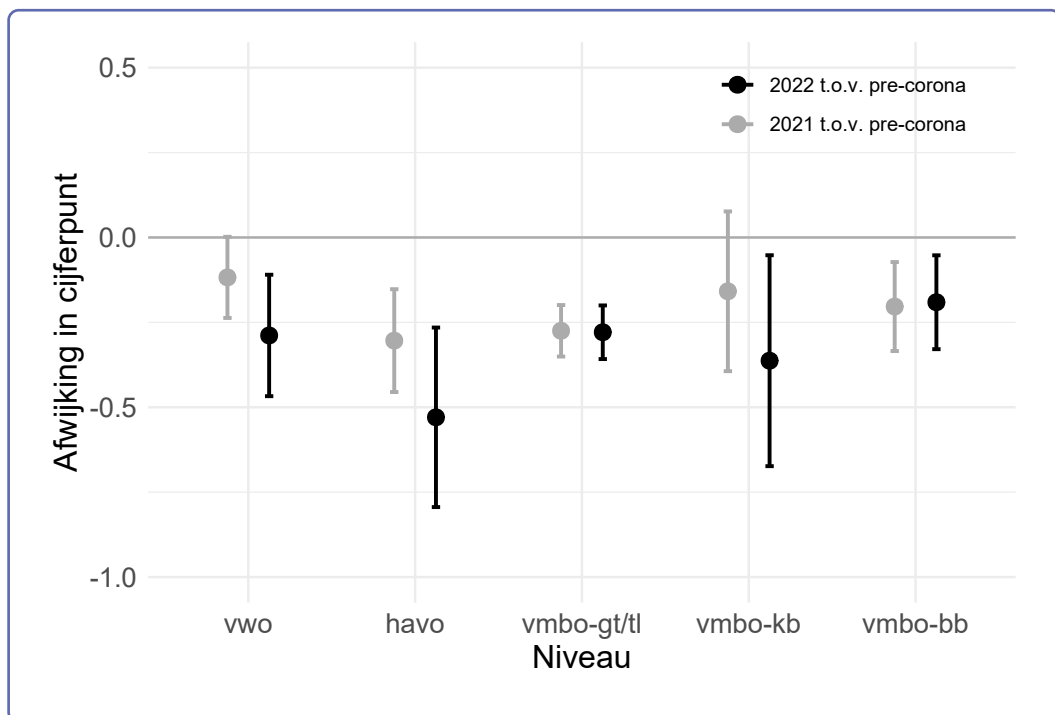
3 Resultaten

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 staat op de verticale as een 90%-betrouwbaarheidsinterval waarbinnen de gemiddelde vaardigheidsontwikkeling binnen dat cluster valt. Als de vaardigheid in 2022 (zwart), resp. 2021 (grijs), over het algemeen lager was dan voor de coronapandemie, dan ligt het interval 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 zegt enerzijds iets over 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 zegt het iets over de onzekerheid van de onderliggende schattingen per examen: bij gelijke vaardigheidsontwikkeling van alle onderliggende examens verwacht je bij een set examens waarbij naar de historische N-terminen is gekeken een groter interval dan bij veel AIP-examens.

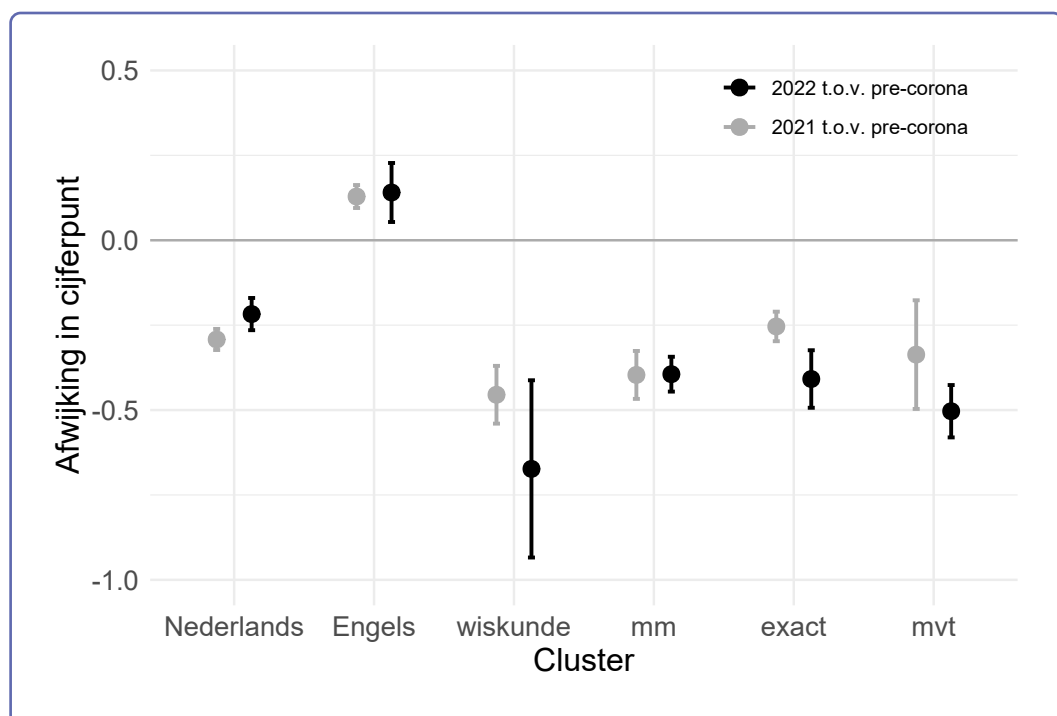
Figuur 1 Geschat vaardigheidsverschil per onderwijsniveau voor 2022 ten opzichte van voor corona (zwart) en 2021 ten opzichte van voor corona (grijs)



Er is bij alle onderwijsniveaus in 2022 ten opzichte van voor de coronapandemie een significante daling in vaardigheid te zien. Dit was in 2021 nog niet het geval voor wvo en vmbo-kb. In vergelijking, lijken de dalingen in vaardigheid voor wvo, havo en vmbo-kb in 2022 ten opzichte van voor corona groter dan in 2021 ten opzichte van voor corona. Het verschil is echter niet significant, omdat de betrouwbaarheidsintervallen van 2021 en 2022 overlappen.

Ook bij inhoudelijke clustering van centrale examens (Figuur 2) is de gemiddelde vaardigheid in alle clusters in 2022 gedaald ten opzichte van voor corona met uitzondering van Engels, waar de vaardigheid is toegenomen. De daling is het sterkst voor wiskunde. Het grote interval maakt duidelijk dat de vaardigheidsontwikkelingen voor de wiskunde-examens sterker variëren dan de vaardigheidsontwikkelingen voor de examens binnen de andere clusters. Als we inzoomen op de vaardigheidsdaling bij wiskunde blijkt dat deze bij vmbo aanzienlijk groter is dan bij havo/vwo¹⁴. De resultaten voor 2022 (Figuur 2, zwart) zijn vergelijkbaar met het verschil in vaardigheden van 2021 ten opzichte van voor corona (Figuur 2, grijs). In vergelijking is de vaardigheidsdaling van 2022 ten opzichte van voor corona in het cluster Nederlands kleiner dan in 2021 ten opzichte van voor corona en groter in het cluster exact. Ook bij wiskunde en moderne en vreemde talen zien we een daling in de puntschatting maar dit verschil is niet significant omdat betrouwbaarheidsintervallen van 2021 en 2022 overlappen.

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



Opmerking: mm = mens en maatschappij, mvt = moderne en vreemde talen exclusief Engels

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

¹⁴ Omdat wiskundevakken alleen kernvakken zijn voor havo en vwo, hebben wij ook gekeken naar een eventueel onderscheidend effect voor wiskunde als kernvak (havo en vwo) en wiskunde als regulier vak (vmbo). Ten behoeve van de vergelijkbaarheid met de resultaten van vorig jaar is ervoor gekozen om in de presentatie ook in dit rapport alle wiskundevakken samen als één cluster te beschouwen.

deels beïnvloed door de onzekerheid waarmee vaardigheidsontwikkelingen geschat kunnen worden. In de analyses per deelvraag, kijken we gericht naar de verschillen tussen centrale examens op basis van herkansingen, lage scores, spreiding van de af te leggen examens over tijdvakken en schoolspreidingen.

3.2 Resultaten per deelvraag

3.2.1 Effect extra herkansingen op vaardigheid

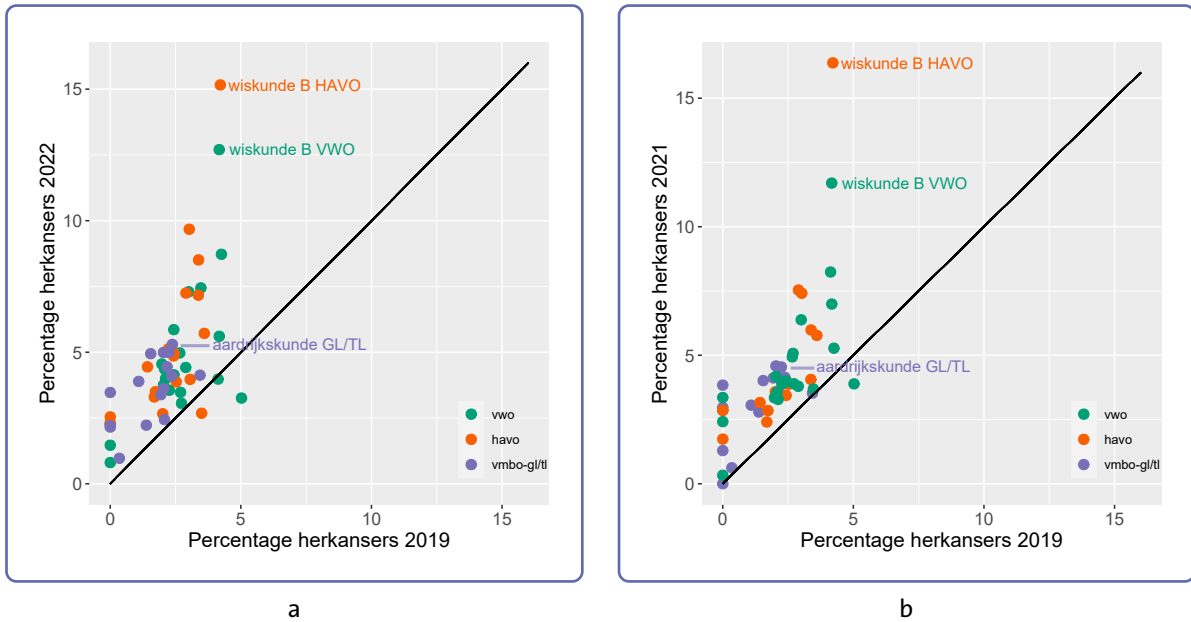
Na herkansing was het cijfer van kandidaten veelal hoger dan daarvoor. Gemiddeld hebben herkansende examenkandidaten zich bij herkansing in vmbo-gl/tl met 0,8 cijferpunten, in havo met 0,8 en in vwo met 0,6 cijferpunt verbeterd (+0,7 gemiddeld over alle onderwijsniveaus). Dit was gemiddeld over alle herkansende kandidaten van deze onderwijsniveaus en alle vakken. In Figuur 3 staat op de horizontale as het percentage herkansers per centraal examen in 2019, op de verticale as het percentage in 2022 (Figuur 3a), resp. 2021 (Figuur 3b). De schuine lijn geeft aan wanneer het percentage herkansers in 2019 gelijk was aan dat in 2022, resp. 2021.

Het eerste dat opvalt in Figuur 3a is dat slechts bij drie centrale examens (wiskunde C en maatschappijwetenschappen in vwo en Frans in havo) het percentage herkansers in 2022 lager is dan voor de coronapandemie in 2019. Gemiddeld is het percentage herkansers per centraal examen in 2022 tweeënhalf procentpunten hoger dan in 2019, een stijging van gemiddeld 2% herkansers in 2019 naar 4,5% herkansers in 2022. Hiermee is het aantal herkansers in het algemeen nog iets groter dan in 2021. Er zijn echter uitschieters als het gaat om het percentage herkansers in 2022. Terwijl in 2019 het percentage tussen de 0 en 5% lag, ligt het percentage in 2022 tussen de 0 en 15%.

In vmbo-gl/tl liggen de percentages herkansers in 2022 weliswaar iets hoger dan in 2019, maar nog steeds zijn dit relatief lage percentages. Bij aardrijkskunde ligt het percentage het hoogste: met 5% herkansers in 2022 terwijl dit in 2019 2,5% was.

Bij havo en vwo zijn er vooral bij de wiskunde-examens hoge percentages herkansers in 2022. Wiskunde B in havo kent het hoogste percentage herkansers, namelijk 15%, gevolgd door wiskunde B in vwo met 12% herkansers, terwijl er voor beide examens in 2019 slechts 4% herkansers waren.

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

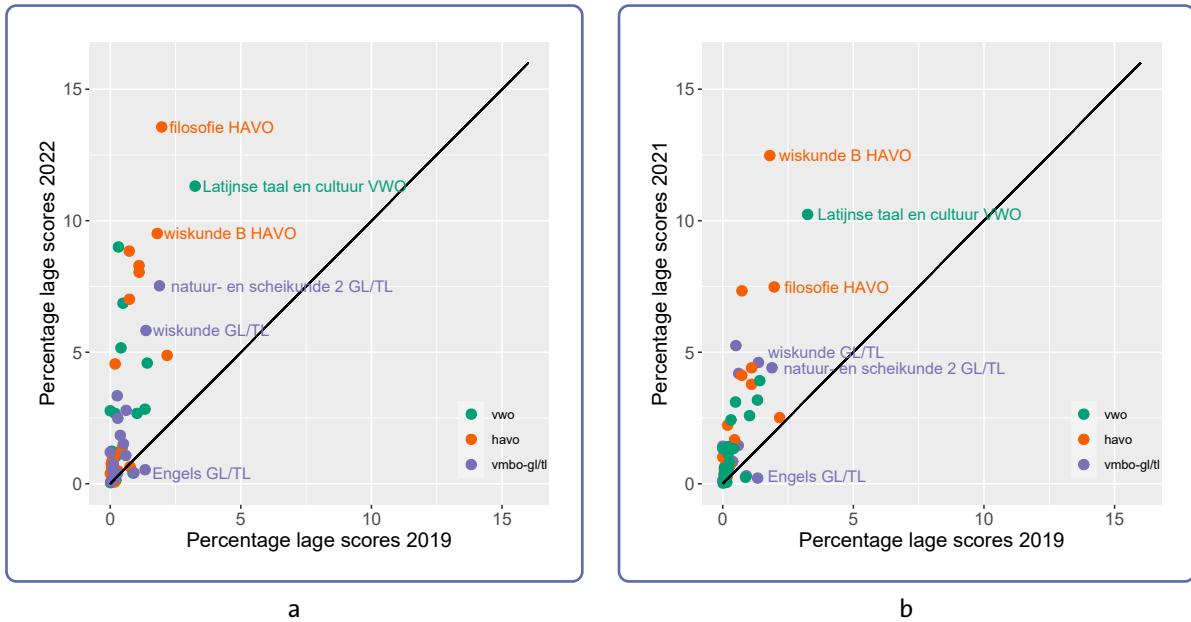


De invloed van deze cijfers op de vaardigheidsschattingen in Figuur 1 en Figuur 2 kunnen we in kaart brengen door een schatting te maken hoeveel kandidaten beter hadden kunnen presteren wanneer het herkansingscijfer al in het eerste tijdvak was behaald. De vaardigheid van de examenpopulatie 2022 zou na herkansing gemiddeld zo'n $4,5\% \cdot 0,7$ cijferpunt = 0,031 cijferpunt hoger zijn dan voor de coronapandemie. Omdat de vaardigheid van de examenpopulatie 2022 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 2022 is dan $0,031 - 0,014 = 0,017$ cijferpunt. Merk op dat dit veel kleiner is dan 0,1 cijferpunt. Daarmee zou het vaardigheidsverschil zoals gepresenteerd in Figuur 1 en Figuur 2 niet veel anders worden 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 eerstekansers 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 2022 (a), resp. 2021 (b). De schuine lijn geeft aan wanneer het percentage in 2022 respectievelijk 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.

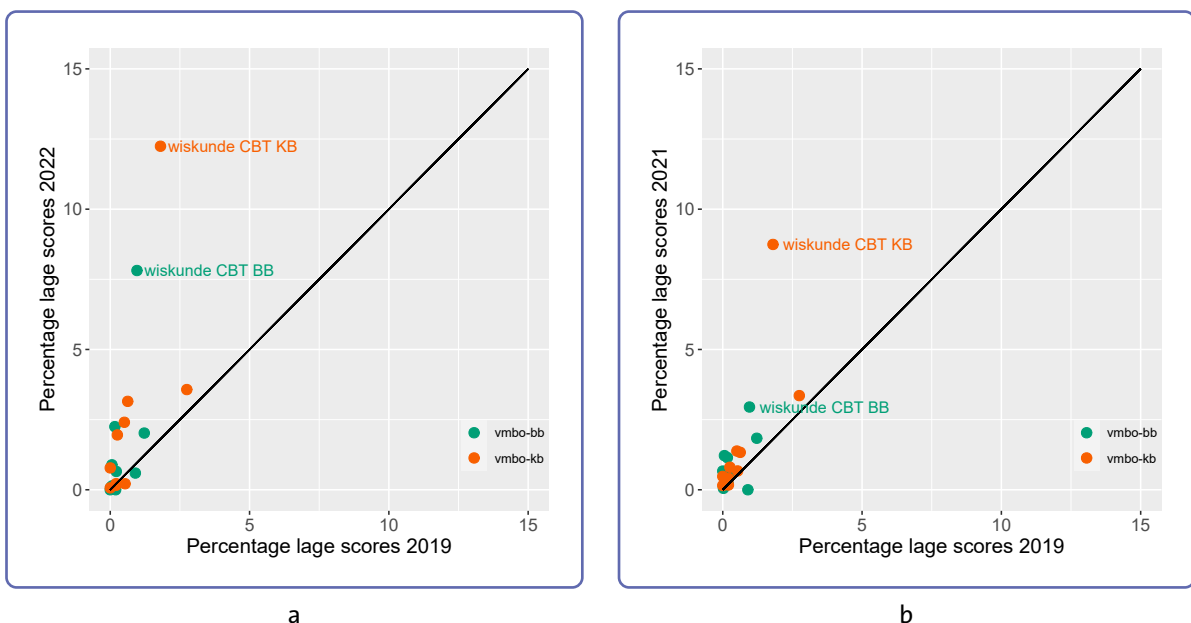
Figuur 4 Percentage eerstekansers met een lage score per CSE in 2019 en 2022 (a) en 2019 en 2021 (b)



Bij vrijwel alle centrale examens, in alle onderwijsniveaus, is het percentage kandidaten met een lage score (minder dan 25% van de maximale score) in 2022 groter dan in 2019 (Figuur 4a en Figuur 5a). In vwo toont Latijnse taal en cultuur een veel hoger percentage examenkandidaten met een lage score (11%) dan in 2019 (3%). Bij havo springt filosofie eruit met 14% kandidaten met een lage score in 2022 (tegenover 2% in 2019). In vmbo-gl/tl liggen in 2022 bij natuur- en scheikunde 2 en wiskunde het percentage kandidaten met een lage score boven de 5%.

In directe vergelijking met de resultaten van 2021 ten opzichte van 2019 (Figuur 4b en 5b) is te zien dat het percentage lage scores in 2022 verder is toegenomen. Het aantal examens dat in 2022 meer dan 5% lage scores heeft (figuur 4a) is groter dan in 2021 (figuur 4b). Er zijn beduidend meer vakken die boven de 5 procent lijn op de y-as liggen.

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



In vmbo-kb en vmbo-bb vertonen de wiskunde-examens in 2022 de grootste toename in kandidaten met een lage score zowel in directe vergelijking met 2019. Wiskunde op kb- en bb-niveau is overigens het enige vak dat meer dan 5% lage scores kende bij de digitale examens, namelijk 12% (kb) en 8% (bb).

De verhoogde frequentie lage scores kan potentieel een groot effect hebben op de gevonden resultaten in figuur 1 en 2. Omdat niet bekend is welke scores leerlingen halen als zij hun beste prestatie neerzetten (als zij hun potentieel volledig laten zien op het centraal examen), moeten we daar aannamen over doen. Een kandidaat die in 2022 een lage score haalt zou onder pre-corona-omstandigheden mogelijk een 5 hebben gehaald. Dit is al snel een verschil van 2 cijferpunten. De verandering van lage scores in 2022 ten opzichte van 2019 varieert van een lichte daling van 0,5% Engels gl/tl tot een grote stijging 12% (filosofie gl/tl). Voor een vijfde van alle centrale examens gaat het om meer dan 5%. Als 5% van de leerlingen een lage score haalt, betekent dit een daling van $(5\% * 2 \text{ cijferpunten}) = 0,1$ cijferpunt. Als dit gemiddeld bij alle centrale examens in een cluster het geval is, dan zou dit effect voor ongeveer 0,1 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 5% dan zou de bijdrage daarvan aan de geschatte vaardigheid nog groter kunnen zijn dan 0,1 cijferpunt. Figuur 4a toont dat met name bij havo- en vwo-vakken examenkandidaten relatief vaak lage scores hebben behaald. Deze percentages zijn ook hoger dan in 2021 het geval was (Figuur 4b). Bij de clusters taal en vmbo-gt komen lage scores niet veel vaker voor dan in eerdere jaren. Bij de clusters exact, mens en maatschappij, havo en vwo komen lage scores wel frequenter voor met een gemiddelde stijging van 2 tot 3 procent per cluster. Een deel van de vaardigheidsdaling in figuur 1 en 2 zou dus verklaard kunnen worden door een toename in lage scores.

3.2.3 Effect van eerstekansers in tweede tijdvak

Net als in 2021, hebben de kandidaten van de centraal schriftelijke examens in 2022 de mogelijkheid gekregen om het examen te spreiden over twee tijdvakken bestaande uit een eerste tijdvak voor eerste afnames en een tweede tijdvak voor eerste afnames en herkansingen. De derde deelvraag was of de kandidaten in het eerste tijdvak wel representatief waren voor alle eerstekansers, aangezien de vaardigheidsontwikkeling geschat is op kandidaten uit het eerste tijdvak. Deze vraag heeft geen betrekking op de digitale centrale examens in vmbo-bb en vmbo-kb, omdat daarbij niet in tijdvakken afgenomen wordt.

In 2021 deden per centraal examen gemiddeld 5% van de kandidaten voor het eerst examen in het tweede tijdvak. In 2022 is het aantal kandidaten die voor het eerst examen in het tweede tijdvak deden naar gemiddeld 3% gedaald. Alleen voor Griekse taal en cultuur, wiskunde B, bedrijfseconomie en kunst in vwo deden relatief meer kandidaten in 2022 dan in 2021 voor het eerst examen in het tweede tijdvak.

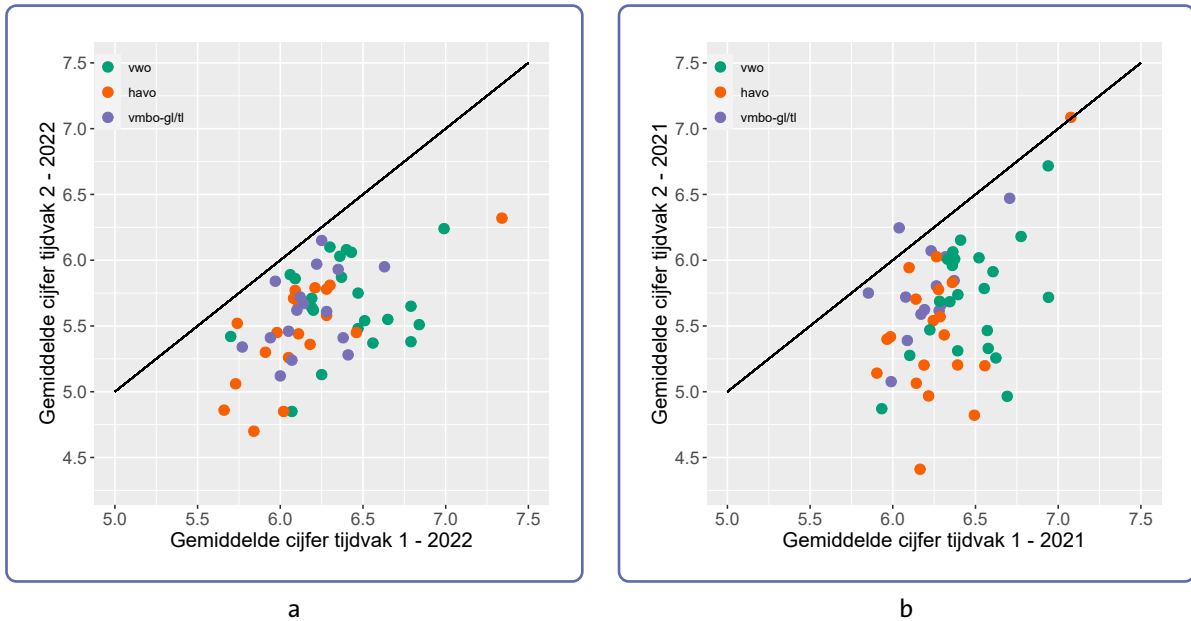
In vmbo-gl/tl hebben examenkandidaten in 2022 minder vaak dan in 2021 gebruik gemaakt van de mogelijkheid om het centraal examen te spreiden. Gemiddeld deed 2% van de vmbo-gl/tl-kandidaten per centraal examen de eerste afname in het tweede tijdvak, terwijl dit in havo en vwo gemiddeld 3,5% per centraal examen was (zie Tabel 2).

Tabel 2 Percentage eerstekeisers dat vak in tweede tijdvak aflegt voor 2022, 2021 en het verschil tussen 2022 en 2021

Vak	2022			2021			verschil 2022-2021		
	Vwo	havo	vmbo-gl/tl	vwo	havo	vmbo-gl/tl	vwo	havo	vmbo-gl/tl
Nederlands	0,9	2,0	1,0	1,9	2,7	1,3	-1,0	-0,7	-0,3
Latijnse taal en cultuur	5,5			11,0			-5,5		
Griekse taal en cultuur	8,9			6,8			2,1		
Frans	1,3	1,9	1,5	4,0	3,1	2,8	-2,7	-1,2	-1,3
Duits	1,6	1,6	1,7	2,2	2,4	2,1	-0,6	-0,8	-0,4
Engels	1,4	1,1	1,3	2,3	2,7	1,4	-0,9	-1,6	-0,1
Spaans	0,7	0,6	2,2	1,7	3,3	1,6	-1,0	-2,7	0,6
Geschiedenis	3,7	4,3	2,6	7,7	6,7	3,8	-4,0	-2,4	-1,2
Aardrijkskunde	2,8	3,3	2,8	4,0	5,9	4,0	-1,2	-2,6	-1,2
Wiskunde A	3,1	3,3		3,7	5,3		-0,6	-2,0	
Wiskunde (B)	3,5	4,0	2,4	3,4	7,3	3,8	0,1	-3,3	-1,4
Wiskunde C	4,2			4,8			-0,6		
Natuurkunde/ nask 1	4,8	4,4	3,1	6,0	6,9	3,9	-1,2	-2,5	-0,8
Scheikunde/ nask 2	2,8	2,9	2,4	5,3	4,3	4,2	-2,5	-1,4	-1,8
Biologie	2,6	4,7	2,0	6,0	8,2	4,1	-3,4	-3,5	-2,1
Economie	2,3	3,1	2,4	3,7	5,4	4,0	-1,4	-2,3	-1,6
Bedrijfseconomie	3,9	2,3		3,7	5,7		0,2	-3,4	
Muziek	3,8	3,3	1,2	7,1	3,6	1,6	-3,3	-0,3	-0,4
Tek., hand., text.	4,6	3,4		7,3	6,4		-2,7	-3,0	
Filosofie	7,0	5,6		12,8	7,6		-5,8	-2,0	
Maatschappijw.	3,3	6,6	2,7	7,9	4,2	4,9	-4,6	2,4	-2,2
Kunst (algemeen)	5,3	6,8		3,2	9,3		2,1	-2,5	
Drama			2,2			11,0			-8,8
Beeldende vakken			1,2			2,1			-0,9

In Figuur 6 zijn de gemiddelde cijfers van de eerstekeisers in het eerste tijdvak (horizontale as) per centraal examen uitgezet tegen de gemiddelde cijfers van de eerstekeisers in het tweede tijdvak (verticale as) voor 2022 (Figuur 6a) en 2021 (Figuur 6b). Bij alle centrale examens op alle drie de onderwijsniveaus behaalden de eerstekeisers in het tweede tijdvak in 2022 gemiddeld een lager cijfer dan de eerstekeisers in het eerste tijdvak. Figuur 6a toont ook dat bij geen van de centrale examens in het eerste tijdvak het gemiddelde cijfer onder de 5,5 (onvoldoende) was, terwijl een groot aantal centrale examens wel een gemiddeld cijfer lager dan 5,5 had in het tweede tijdvak.

Figuur 6 Gemiddelde cijfer eerstekansers in het eerste tijdvak en eerstekansers in het tweede tijdvak voor 2022 (a) en 2021 (b).

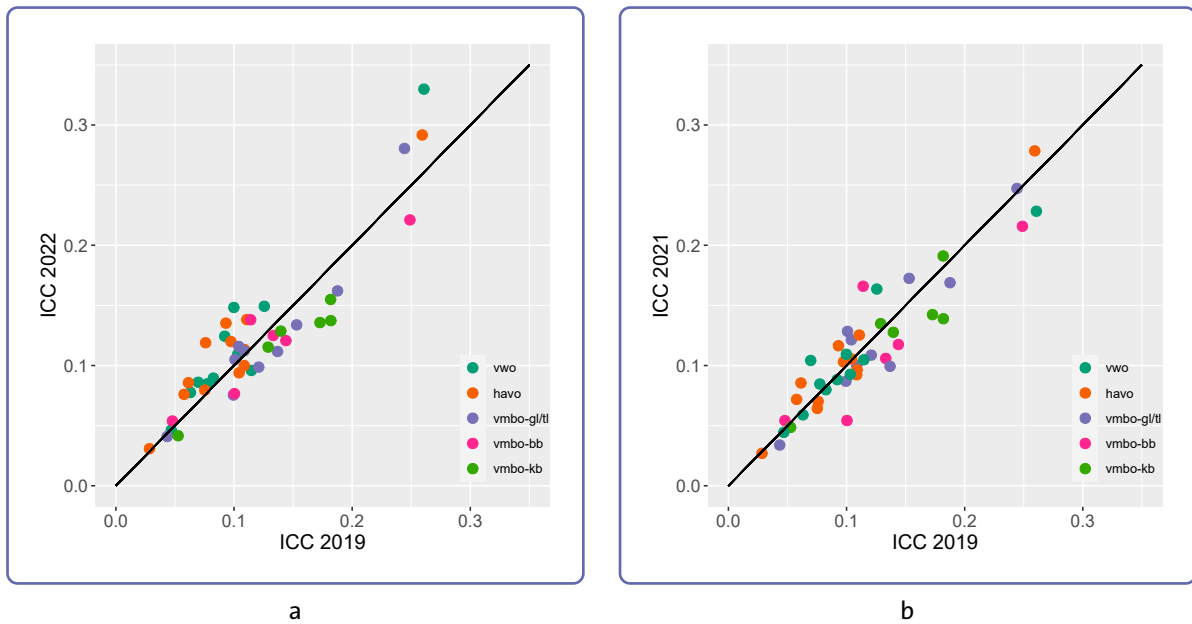


Er is naar gestreefd om de examens van 2022 gemiddeld even streng te normeren in het tweede tijdvak als in het eerste tijdvak. Gemiddeld lagen de cijfers van de eerstekansers in het tweede tijdvak 0,7 cijferpunt lager dan van de kandidaten in het eerste tijdvak. Een dergelijke gemiddelde daling in cijfers duidt op een lagere vaardigheid van de eerstekansers in het tweede tijdvak, dan die van de kandidaten in het eerste tijdvak. De schattingen van de vaardigheidsontwikkeling (Figuur 1 en Figuur 2) zijn gebaseerd op de kandidaten in het eerste tijdvak. Gezien de geringe percentages kandidaten die het per centraal examen betreft, is de vaardigheid van de totale populatie echter niet veel lager dan weergegeven in Figuur 1 en Figuur 2. Een rekenvoorbeeld: als 3% van de populatie 0,7 cijferpunt lager scoort, dan valt de gemiddelde vaardigheid $0,03 \cdot 0,7 = 0,021$ cijferpunt lager uit. Dus als de eerstekansers in het tweede tijdvak wel zouden zijn meegenomen in de berekeningen van Figuur 1 en Figuur 2, dan zouden de gemiddelden maar beperkt, veel minder dan 0,1 cijferpunt, lager uitgevallen zijn dan nu het geval is. Bij een enkel vak zou het effect groter kunnen zijn, maar dan moet zowel het percentage eerstekansers in het tweede tijdvak hoog zijn, als het verschil in gemiddeld cijfer tussen deze kandidaten en de kandidaten in het eerste tijdvak. Het cluster waarbij dit effect het grootst zou zijn is wiskunde met 3,4% van de populatie, maar ook hier valt het gemiddelde ruim onder de 0,1 cijferpunten. Vanwege de onzekerheid bij een enkel vak, is de vaardigheids-schatting juist gedaan op clusters van centrale examens.

3.2.4 Spreiding tussen scholen

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

Figuur 7 ICC (intraklasse correlatie) per centraal examen in 2019 en 2022 (a) en 2019 en 2021 (b)



Per centraal examen is de ICC op de cijfers van 2019 en 2022 (Figuur 7a) en van 2019 en 2021 (Figuur 7b) aangegeven. Elk punt staat voor een centraal examen. De schuine lijn geeft aan wanneer de ICC in 2022 resp. 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 2022 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.

4 Conclusie

4 Conclusie

De hoofdvraag van het onderzoek was hoe de vaardigheid van de kandidaten in 2022 zich verhoudt tot de vaardigheid van de examenpopulaties voor de coronapandemie (2015-2019 met nadruk op 2019). Daarnaast hebben we deze resultaten naast de vaardigheid van de examenpopulaties tijdens in 2021 gelegd. De vaardigheid in 2022 is in alle onderwijsniveaus gedaald ten opzichte van voor de coronapandemie. De vaardigheid voor de clusters exact, mens en maatschappij (inclusief kunst), overige moderne vreemde talen, Nederlands en wiskunde is gedaald. Daarbij vond de grootste gemiddelde daling plaats bij exacte vakken – inclusief wiskunde – en bij mens-en-maatschappijvakken. De gemiddelde vaardigheid in het cluster Engels is daarentegen gestegen.

De resultaten voor 2022 zijn vergelijkbaar met die van 2021, behalve dat de vaardigheidsdaling voor vwo en vmbo-kb vorig jaar niet significant was en dit jaar wel. Het vaardigheidsverschil van 2022 ten opzichte van voor corona is in het cluster Nederlands kleiner dan in 2021 ten opzichte van voor corona en groter in het cluster exact. Voor alle anderen clusters is het verschil niet significant.

De eerste deelvraag betrof de invloed van herkansingen op de gemiddelde vaardigheid van kandidaten. Omdat in 2022, net zoals in 2021, kandidaten twee centrale examens mochten herkansen in plaats van één, was de verwachting dat het aantal herkansers in 2022 hoger zou zijn dan in 2019 maar vergelijkbaar met het aantal herkansers in 2021. Er is inderdaad gemiddeld een kleine toename in het percentage herkansers in vergelijking met 2019, met enkele uitschieters. Bij wiskunde havo en vwo was er een grote toename in het percentage herkansers. Desondanks zou het totale gemiddelde cijfer van een centraal examen als de leerlingen hun prestatie op de herkansing bij de eerste gelegenheid zouden hebben behaald nauwelijks stijgen. De herkansingen hebben dan ook geen noemenswaardig effect op de geschatte vaardigheid van de populatie. Er was geen significante toename in het aantal herkansers in vergelijking met 2021.

Een tweede deelvraag om de daling in vaardigheid te duiden, bekeek een ander aspect, namelijk het percentage kandidaten met een lage score. De geschatte vaardigheid zou beïnvloed kunnen zijn doordat kandidaten de mogelijkheid kregen om één vak weg te strepen bij de uitslagbepaling. Er was geen informatie beschikbaar over welke kandidaten bij welk examen doelbewust van deze mogelijkheid gebruik wilden maken. Wel kwamen kandidaten met lage scores in 2022 vaker voor dan in 2019, wat de gemiddelde geschatte vaardigheid naar beneden trok. De oorzaak van de toename in lage scores is niet af te leiden uit de scores zelf. De toename in lage scores kan veroorzaakt zijn door kandidaten die niet of minder goed hun best hebben gedaan op het centraal examen vanwege de duimregel maar ook door kandidaten die minder goed voorbereid waren vanwege de omstandigheden tijdens de coronapandemie. Bij de kernvakken was de duimregeling niet van toepassing en daarom kan een toename in lage scores bij deze centrale examens niet veroorzaakt zijn door deze regeling. In vergelijking met 2021 is het aantal lage scores in 2022 opnieuw gestegen. De berekening laat zien dat extra lage scores bij sommige examens kunnen leiden tot een extra daling van de vaardigheid binnen de clusters exact, mens en maatschappij, havo en vwo. Bij deze clusters is het denkbaar dat ongeveer 0,1 cijferpunt van de vaardigheidsdaling in Figuren 1 en 2 veroorzaakt is door lage scores.

In 2022 kregen kandidaten de gelegenheid om het centraal examen te spreiden. Kandidaten konden ervoor kiezen om een centraal examen in het tweede tijdvak te doen in plaats van in het eerste tijdvak. De derde deelvraag haakte hierop in door te kijken of de kandidaten waarop de vaardigheidsschatting berekend was, wel representatief waren voor alle kandidaten.

Per centraal examen heeft ongeveer 3% van de kandidaten in 2022 gebruik gemaakt van de mogelijkheid om een afname in het tweede tijdvak te doen in plaats van het eerste tijdvak. Dit varieerde van ongeveer 0,6% bij Spaans in vwo en havo tot 8,9% voor Griekse taal en cultuur in vwo. In vmbo-gl/tl hebben examenkandidaten minder gebruik gemaakt van de mogelijkheid om het centraal examen te spreiden dan in havo en vwo. Bij vrijwel alle centrale examens haalden kandidaten, die hun eerste afname hadden uitgesteld tot het tweede tijdvak, gemiddeld lagere cijfers dan de kandidaten die al in het eerste tijdvak het centraal examen hadden gedaan. Dit verschil liep bij een aantal examens op tot meer dan 1,3 cijferpunt. Bij ongeveer de helft van de centrale examens haalden de eerstekansers in het tweede tijdvak gemiddeld een onvoldoende. Gezien de meestal beperkte percentages eerstekansers in het tweede tijdvak zijn de effecten daarvan op de gemiddelde vaardigheidsschattingen per onderwijsniveau of inhoudelijk cluster van centrale examens echter ook beperkt (gemiddeld minder dan 0,1 cijferpunt).

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 0,2 à 0,4 cijferpunt. Bij de inhoudelijke clusters van examens is sprake van een vaardigheidsdaling tussen de 0,2 en 0,7 cijferpunt, met uitzondering van Engels waar een vaardigheidsstijging van ongeveer 0,1 cijferpunt te zien is. De invloed van herkansende kandidaten en van eerstekansers in het tweede tijdvak op deze schattingen is beperkt. Het aantal lage scores zou de daling in vaardigheden deels kunnen verklaren. 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, zijn bij alle clusters van centrale examens dalingen in vaardigheid van 0,2 tot 0,7 cijferpunt gevonden. De vraag die rijst, is hoe groot deze vaardigheidsdalingen nu eigenlijk zijn. 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 kunnen fluctueren. De fluctuaties zijn in 2022 bijna alle in negatieve richting. Deze negatieve trend is vergelijkbaar met de trend van 2021¹⁶ en deze daling in cijfers is voor een deel van de cluster groter dan verwacht kon worden op basis van gebruikelijke schommeling in gemiddeld cijfer in 2019 en eerder.

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 2022. In 2019 was de gemiddelde standaarddeviatie in cijfers ongeveer 1,1. Daarmee zou een daling van 0,2 tot 0,7 cijferpunt een middelgroot tot groot effect genoemd kunnen worden, volgens de interpretatie van effectmaten van Kraft¹⁷.

Een lichte daling van de vaardigheid of prestaties tussen de jaren 2019 en 2022 is in lijn met bevindingen in ander nationaal en internationaal onderzoek^{18,19,20,21,22,23}. Er zijn echter nog niet veel studies die de langetermijnveranderingen in vaardigheden tijdens de twee jaar corona in Nederland onderzochten. Daarom is het moeilijk de resultaten van dit rapport in verband te brengen met resultaten van andere studies. Uit onderzoek van M-toetsen in het PO blijkt dat er in 2022 nog steeds vertragingen in de leergroei zijn²⁴, wat strookt met de resultaten van dit rapport.

-
- 15 <https://www.cito.nl/kennis-en-innovatie/kennisbank/cte-rapport-vaardigheidsontwikkelingen-volgens-pisa-en-examens>
- 16 Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., en Emons, W. (2021). Vaardigheid examenkandidaten 2021. https://www.examenblad.nl/document/rapport-cito-effect-corona/2021/f=/Rapport_Cito_Effect-Corona_vaardigheid_CE_2021.pdf
- 17 Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241–253. <https://doi.org/10.3102/0013189X20912798>
- 18 Lek, K., Feskens, R., Veldhuizen, J., en Keuning, J. (2021). Leervertragingen basisonderwijs deels ingelopen. Coronarapport 2 (cito.nl)
- 19 Zijlstra, W., Rijtdiek, S., Frissen, S. en Feskens, R. (2021). Leervertraging door schoolsluitingen in het voortgezet onderwijs – deel 2. Title (cito.nl)
- 20 Rapport De Staat van het Onderwijs 2021 | Rapport | Inspectie van het onderwijs (onderwijsinspectie.nl)
- 21 Betthäuser, 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>
- 22 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>
- 23 Bundesministerium für Bildung und Forschung, Deutschland. Pressemitteilung: 36/2022. <https://www.bmbf.de/bmbf/shareddocs/pressemitteilungen/de/2022/05/100522-Aufholen-nach-Corona.html>
- 24 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

Uit dit onderzoek blijkt dat kandidaten de ruimere spreidingsmogelijkheden van examens beperkter hebben ingezet dan in 2021. Het aantal kandidaten dat gebruik maakte van de mogelijkheid om in het tweede tijdvak examens af te leggen voor een vak is in 2022 sterk gedaald. Het percentage herkansers per centraal examen in 2022 is doorgaans gestegen ten opzichte van 2019 maar is vergelijkbaar met het percentage herkansers in 2021. Uit de Examenmonitor²⁵ blijkt dat het percentage kandidaten dat een of twee centrale examens herkanst, in 2022 licht is afgenomen ten opzichte van voor corona (er is geen verschil in 2022 ten opzichte van 2021). Deze schijnbare tegenstrijdigheid komt door een andere berekening van de percentages. In dit onderzoek wordt bij ieder examen afzonderlijk geteld hoeveel eerstekansers er waren, en hoeveel daarvan herkansten. In de examenmonitor wordt er per onderwijsniveau gekeken naar het aantal kandidaten dat één of twee examens herkanst, ten opzichte van het totaal aantal kandidaten in het betreffende onderwijsniveau. Het merendeel van de kandidaten herkanst helemaal niet. Maar doordat de meerderheid van de kandidaten die wel herkanst, in 2022 juist twee centrale examens herkanst, terwijl dat voorheen altijd maar één centraal examen betrof, zijn de percentages herkansers per centraal examen gestegen.

De Examenmonitor laat tevens zien dat het wegstrepen van de resultaten van één vak veruit de meest gebruikte maatregel is. Het gebruik van de regeling is in 2022 licht toegenomen ten opzichte van 2021 en elf procent van de eindexamenkandidaten is geslaagd dankzij de mogelijkheid om een vak buiten beschouwing te laten bij de uitslagbepaling (tien procent in 2021). Dit zou het grotere aantal kandidaten met een lage score bij de niet-kernvakken kunnen verklaren. In dit onderzoek bleek dat van de onderzochte factoren het percentage lage scores de meeste invloed heeft op de vaardigheidsontwikkeling.

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

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 \(examenblad.nl\)](#)

College van Toetsen en Examens (2017). Examenblad.nl: Wat is Facet? <https://www.examenblad.nl/veel-gevraagd/wat-is-facet/2017/vwo?horizon=vg41h1h4i9qe>

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 \(cito.nl\)](#)

[N-term - Examenblad 2019](#)

[Normering centrale examens 2022 - Examenblad](#)

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

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)

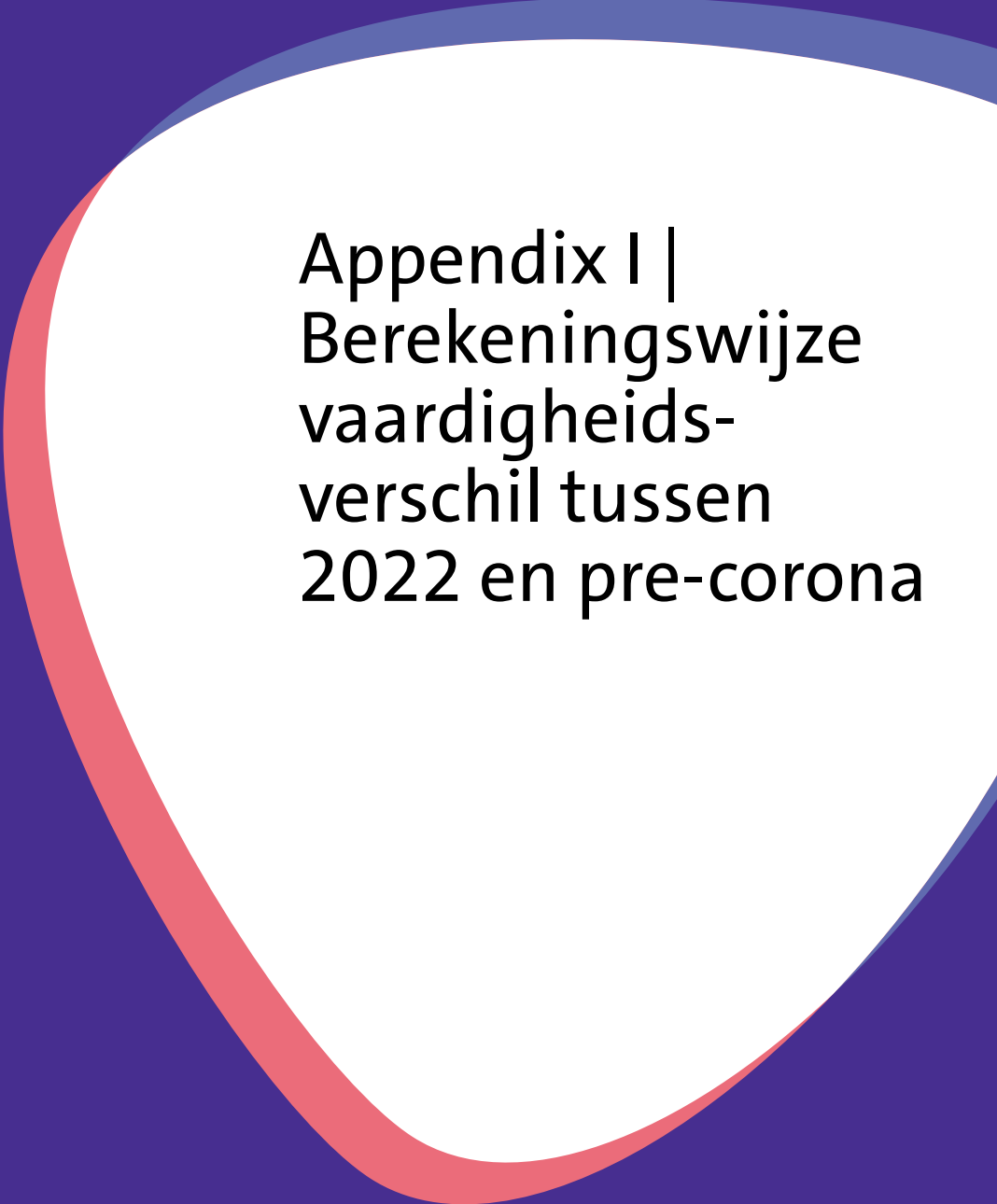
Vaardigheidsontwikkelingen volgens PISA en examens | Rapport | Rijksoverheid.nl

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.cito.nl/kennis-en-innovatie/kennisbank/cte-rapport-vaardigheidsontwikkelingen-volgens-pisa-en-examens>

Van Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., en Emons, W. (2021). *Vaardigheid examenkandidaten 2021*. https://www.examenblad.nl/document/rapport-cito-effect-corona/2021/f=/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*. [Title \(cito.nl\)](https://www.cito.nl)



Appendix I |
Berekeningswijze
vaardigheids-
verschil tussen
2022 en pre-corona

Appendix I | Berekeningswijze vaardigheidsverschil tussen 2022 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 2022 te relateren aan de prestaties in de jaren voor de coronapandemie (2019) en tijdens de coronapandemie (2021).

Gebruikte data

Voor de schatting van de vaardigheid van 2021 en 2022 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 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-omstandigheden 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 2022, 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 2022, 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-analyse²⁶ 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 – 2022 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 2022. Een daling in gemiddelde scores tussen deze twee gemiddelden reflecteerde dus een daling in vaardigheid. Het verschil in gemiddelde scores van 2022 en de pre-corona-periode werd uitgedrukt in een verschil in gemiddeld cijfer²⁷ om de getallen beter te kunnen duiden, en om centrale examens te kunnen aggregeren.

De puntschatting van dit zogenaamde contrast werd voorzien van een nauwkeurighedsmaat²⁸. In de berekening van deze nauwkeurighedsmaat 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 2022 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

26 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.

27 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.

28 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 2022-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 weergaven²⁹. 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

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

²⁹ $w_k = 1 / (SE_k)^2$

© Copyright 2022

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.

Colofon

Redactie: []
Fotografie: []
Illustraties: []
Datum: []

Cito.nl



CITO CTE