

Rapport

Vaardigheid Examenkandidaten 2024



Vaardigheid Examenkandidaten 2024

**Met medewerking van Karin Pilz, Marieke van Onna,
Harco Weemink**

Inhoud

Samenvatting	5
1 Inleiding	6
1.1 Aanleiding	6
1.2 Normering 2024	6
1.3 Onderzoeksvragen	7
1.3.1 Effect van kandidaten met lage scores	7
1.3.2 Vaardigheid 2024 ten opzichte van 2023	7
1.4 Indeling van dit rapport	7
2 Methoden van onderzoek	8
2.1 Gegevens voor het onderzoek	8
2.2 Schatting van vaardigheidsontwikkeling	8
2.3 Effect van kandidaten met lage scores	10
2.4 Vaardigheid 2024 ten opzichte van 2023	10
3 Resultaten	12
3.1 Vaardigheidsontwikkeling	12
3.2 Effect van kandidaten met lage scores	13
3.3 Vaardigheid 2024 ten opzichte van 2023	15
4 Conclusie	16
5 Literatuur	17
Bijlage Berekeningswijze vaardigheidsverschil tussen 2024 en pre-corona	18

Samenvatting

Net zoals in voorgaande jaren, hebben kandidaten die in 2024 eindexamen in het voortgezet onderwijs deden, tijdens hun opleiding te maken gehad met de gevolgen van de coronapandemie. Om in kaart te brengen of de vertragingen ten gevolge van corona zijn verminderd, heeft het ministerie het CvTE verzocht om de vaardigheden van examenkandidaten in 2024 te onderzoeken. Het onderzoek is in opdracht van het CvTE door Stichting Cito uitgevoerd.

De buitengewone omstandigheden tijdens de coronapandemie maakten het normeren van de centrale examens in 2021 tot en met 2023 op traditionele wijze technisch onmogelijk. Door een aangepaste normeringssystematiek reflecteerde de hoogte van de cijfers niet precies de vaardigheid van de kandidaten. Onderzoek uitgevoerd door Stichting Cito in opdracht van het CvTE in 2021 en 2022, toonde aan dat er gemiddelde dalingen waren in vaardigheid in alle onderzochte clusters van vakken, behalve voor het vak Engels. Deze bevindingen werden ondersteund door nationaal en internationaal onderzoek dat wees op leerachterstanden door de pandemie. Ook in 2023, het eerste jaar zonder examenmaatregelen, waren zowel bij clustering op het niveau van vakken (Nederlands, wiskunde, moderne vreemde talen (excl. Engels), mens en maatschappij, exact) als op het niveau van schoolsoorten (havo) nog dalingen in vaardigheid ten opzichte van voor corona te zien. Dit rapport bevat de resultaten van een onderzoek naar de vaardigheidsontwikkeling van de kandidaten in 2024. De gemiddelde vaardigheid van kandidaten in 2024 wordt net zoals in de voorafgaande jaren vergeleken met die van de periode voor de pandemie (2015-2019, met een focus op 2019). Deze vergelijking wordt steeds moeilijker, omdat het referentiejaar door de jaren heen moeilijker te relateren is aan de actuele afname. Voor een vergelijking van de vaardigheid van de examenkandidaten in 2024 met die van de examenkandidaten in 2023 wordt gekeken naar de gemiddelde cijfers.

Bij clustering van centrale examens op schoolsoort is voor vwo, vmbo-kb en bb geen significant verschil in vaardigheid te zien in 2024 ten opzichte van voor de coronapandemie. Voor vmbo-gl/tl laten de vaardigheidsschattingen gemiddeld een significante daling zien van 0,2 cijferpunten en voor havo van 0,15 cijferpunt. Bij de inhoudelijke clustering van examens is sprake van een vaardigheidsdaling van 0,3 cijferpunten voor wiskunde, 0,4 cijferpunten voor moderne vreemde talen. Bij Engels is een vaardigheidsstijging van ongeveer 0,1 cijferpunt te zien en bij de exacte vakken is geen significant verschil met pre-corona te zien.

Het aantal lage scores is gedaald ten opzichte van het afgelopen jaar, maar is nog steeds bij enkele vakken hoger dan voor de coronapandemie. Het is mogelijk dat het aantal lage scores, vooral voor het cluster havo een klein effect heeft op de geschatte gemiddelde vaardigheden.

Het verschil in gemiddelde cijfers tussen 2023 en 2024 weerspiegelt de bevindingen uit het vaardigheidsonderzoek van dit jaar en die van 2023. Bij vwo, vmbo-kb en vmbo-bb lijken de populatie-vaardigheden sinds 2023 weer op het niveau van voor corona. Bij havo is de vaardigheid nog niet op het niveau van voor corona maar is deze wel gestegen sinds 2023. Bij vmbo-gl/tl leek de vaardigheid in 2023 niet significant af te wijken van die voor corona, maar is deze in 2024 weer gedaald.

1 Inleiding

1.1 Aanleiding

In 2020 leidde de coronapandemie tot ongekende uitdagingen in het Nederlandse onderwijs, waaronder de sluiting van scholen en de annulering van de centrale examens. In 2021 werden de centrale examens wel afgenomen met maatregelen zoals een extra herkansing, een extra afnamemoment en de mogelijkheid om één examenresultaat weg te strepen bij de uitslagbepaling. Ook moest een aangepaste normering worden ingevoerd¹. Dit leidde tot een tweetal onderzoeken naar de vaardigheidsontwikkeling van examenkandidaten in 2021 en 2022, uitgevoerd door Stichting Cito in opdracht van het CvTE.^{2,3} In 2023 werd grotendeels teruggekeerd naar de gang van zaken voor de pandemie. Alleen de normeringssystematiek was nog gelijk aan die in 2021 en 2022. Het onderzoek naar de vaardigheden van kandidaten werd in 2023 voortgezet.⁴

In 2024 is de normeringssystematiek wederom aangepast met als doel om, net als voor corona, de eisen die gesteld worden aan leerlingen op een centraal examen van jaar tot jaar gelijk te houden. De gemiddelde cijfers representeren de ontwikkeling van de vaardigheid van de leerlingen dan weer beter.

Examenkandidaten in 2024 hebben, afhankelijk van de schoolsoort waarin zij examen deden, in groep 8, hun onderbouwjaren en/of het 4^e leerjaar te maken gehad met schoolsluitingen en afstandsonderwijs. Om te onderzoeken of zij in staat zijn geweest de door corona opgelopen onderwijsvertragingen in te lopen, heeft het ministerie het CvTE opnieuw gevraagd de vaardigheden van examenkandidaten te onderzoeken. Dit rapport is het resultaat van het onderzoek en is uitgevoerd door Stichting Cito in opdracht van het CvTE.

1.2 Normering 2024

In 2024 waren de prestatie-eisen van 2023 het uitgangspunt van de normering van de centrale examens. Een peiling onder docenten in het najaar van 2023 leidde bij een aantal vakken tot een aanpassing van de prestatie-eisen.⁵

In de normeringssystematiek van 2024 werd zowel gebruik gemaakt van equivaleringsgegevens, zoals voor de corona-periode, als van informatiebronnen, die in de corona-periode gebruikt werden. Deze nieuwe systematiek had als doel de prestatie-eisen van 2023 te handhaven in 2024. Afgezien van de beperkte aanpassingen aan de prestatie-eisen geven verschillen in gemiddelde cijfers tussen 2024 en 2023 dus weer hoe de vaardigheid tussen 2023 en 2024 wijzigde.

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

2 [Rapport Vaardigheid Examenkandidaten 2021 \(bijlage bij 31289,nr.506\) - Parlementaire monitor](#)

3 [Vaardigheid Examenkandidaten 2022 | Rapport | Rijksoverheid.nl](#)

4 [Vaardigheid examenkandidaten 2023 | Rapport | Rijksoverheid.nl](#)

5 [Normering centrale examens | 2024 | Examenblad.nl](#)

De prestatie-eisen van 2023 waren niet altijd in lijn met die van voor de corona-periode als gevolg van de normeringssystematiek die in de jaren 2021 tot en met 2023 werd gehanteerd. Door handhaving van de eisen van 2023 waren de prestatie-eisen in 2024 daarom ook niet in lijn met die van voor corona. Dit betekent dat verschillen in gemiddelde cijfers tussen 2024 en 2019 niet goed weergeven hoe de vaardigheid tussen 2024 en 2019 wijzigde.

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 vaardigheidsonderzoek gebruikt om de vaardigheidsontwikkeling te bestuderen voor alle vijf schoolsoorten, bij minimaal zes centrale examens per schoolsoort.

De hoofdvraag van dit onderzoek is of het vaardigheidsniveau van de examenleerlingen op het centraal examen weer op het niveau is van voor corona.

1.3.1 Effect van kandidaten met lage scores

In aansluiting op deelvraag 2 uit de onderzoeken van de afgelopen 3 jaar, onderzochten we nu ook het effect van kandidaten met lage scores. In vorige jaren hadden lage scores een duidelijke invloed op de vaardigheidsdaling, daarom is opnieuw onderzocht in welke mate er sprake is van extreem lage scores in vergelijking met voorgaande jaren. 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, wat een sterke toename in het aantal lage scores tot gevolg had. De mogelijkheid om een niet-kernvak weg te strepen is sinds 2023 niet meer van toepassing.

1.3.2 Vaardigheid 2024 ten opzichte van 2023

Met de hoofdvraag van dit rapport hebben wij onderzocht of het vaardigheidsniveau van de examenleerlingen op het centraal examen terug was op het niveau van vóór corona. Daarmee weten we echter nog niet of er sprake is van herstel ten opzichte van de examenleerlingen in de afgelopen drie jaren. Daarom vergelijken we ook de vaardigheid van de examenleerlingen in 2024 met die van 2023.

1.4 Indeling van dit rapport

Hoofdstuk 2 bevat een beschrijving van de onderzoeksmethoden. Daarna volgen in hoofdstuk 3 de resultaten van de vaardigheidsschattingen voor 2024 t.o.v. 2019, de invloed van kandidaten met lage scores daarop en een vergelijking van de vaardigheid in 2024 t.o.v. 2023. In de conclusie (hoofdstuk 4) worden de resultaten samengevat.

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

2 Methoden van onderzoek

2.1 Gegevens voor het onderzoek

Om de vaardigheid van de examenpopulatie 2024 te onderzoeken, worden de gegevens gebruikt die Stichting Cito jaarlijks verzamelt voor het normhandhavingsonderzoek. Het normhandhavingsonderzoek is een schakel in het hele normeringsproces^{6,7} 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 examengegevens⁸), 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. Net zoals in de afgelopen drie jaar zijn de cspe's niet meegenomen in dit onderzoek, hoewel ze zowel in 2019 als 2023 en 2024 zijn afgenomen. De profielvak-cspe's zoals die in 2023 en 2024 werden afgenomen, zijn echter in 2019 pas voor het eerst afgenomen. Verschillen in vaardigheid tussen 2024 en 2019 zijn daarom, in vergelijking met de andere examens in dit onderzoek, minder het gevolg van de coronapandemie maar ook van onderwijs en examens die nog in ontwikkeling waren.

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 2024 verschilt van de gemiddelde vaardigheid van de examenpopulaties voor de coronapandemie (2015-2019 met nadruk op 2019). Er zijn twee 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 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

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

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

8 Cito (geen datum). Tools en informatie voor docenten: Wolf. [Wolf – uitwisselprogramma voor examengegevens \(cito.nl\)](#).

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

verschillende centrale examens, gecorrigeerd voor een verschil in moeilijkheidsgraad, levert informatie over de vaardigheidsontwikkeling.

De twee methodes zijn niet even nauwkeurig¹⁰. Anchor in Package (AIP) levert de meest nauwkeurige schatting op, gevolgd door pre- en posttest. In de afgelopen jaren werd voor de vaardigheidsschattingen ook nog een derde methode gehanteerd, de historische N-term. Omdat voor de bepaling van de N-termen in 2024 het jaar 2023 als referentiepunt wordt gebruikt, terwijl in de voorgaande jaren 2019 als referentiepunt gold, zijn de N-termen tussen 2019 en 2024 niet langer vergelijkbaar. De resterende twee methoden werken niet bij regelmatige wisselingen van onderwerpen in het examenprogramma, of bij de introductie van een nieuw examenprogramma. Aangezien 2019 steeds verder in het verleden ligt en er dus meer examens zijn die veranderingen hebben ondergaan, zijn er steeds meer examens waarvoor het niet meer mogelijk is een vergelijking met 2019 te maken. Daarnaast kunnen AIP en pre- en posttest niet worden toegepast bij centrale examens met weinig examenkandidaten.

In Tabel 1 is te zien welke methoden in dit onderzoek zijn ingezet bij de verschillende examens. Tevens is te zien welke vakken in dit rapport zijn samengenomen in een cluster. In de tabel zijn dezelfde examens opgenomen die ook zijn meegenomen in 2021, 2022 en 2023 met uitzondering van Nederlands vwo, biologie vmbo-bb en vmbo-kb en wiskunde vmbo-kb. De vakken waarbij in de afgelopen jaren de historische N-term is gehanteerd zijn niet meegenomen omdat de methode dit jaar niet kon worden toegepast voor dit onderzoek.

Tabel 1 | Gehanteerde gegevens om vaardigheids- en moeilijkheidsverschillen te onderzoeken in 2024

Vak	Cluster	BB-CBT	KB-CBT	GL/TL	HAVO	VWO
Nederlands	Nederlands	AIP			histN	
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			histN	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

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

10 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 2024 wordt geschat door prestaties op de centrale examens van 2024 te vergelijken met die op eerdere centrale examens rekening houdend met verschillen in moeilijkheid. De examens werden op dezelfde wijze geclusterd als in de rapporten van de afgelopen drie jaar, met uitzondering van het cluster 'Nederlands' en het cluster 'Mens & Maatschappij' omdat in die clusters te weinig gegevens beschikbaar zijn. De schatting van vaardigheid bevat onzekerheden. Door de informatie van meerdere centrale examens te combineren wordt deze onzekerheid verkleind en is het enigszins 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 schoolsoort. Meer details omtrent deze berekeningen staan in Appendix I.

Met dit onderzoek wordt de vaardigheidsontwikkeling van de examenleerlingen in 2024 ten opzichte van de examenleerlingen in 2019 in beeld gebracht. Omdat deze vaardigheidsontwikkeling op een heel andere set vakken is gebaseerd dan in het onderzoek van 2023, 2022 en 2021 zijn de resultaten niet langer te vergelijken met de resultaten van die jaren. In hoofdstuk 3 zullen dan ook alleen de resultaten voor 2024 gepresenteerd worden en niet die van de afgelopen jaren.

2.3 Effect van kandidaten met lage scores

Voor de examenscores in dit rapport kijken we steeds naar de scores zoals deze door de eerste corrector zijn toegekend. 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. Om de effecten van kandidaten met lage scores in kaart te brengen, worden voor alle vakken uit het centraal examen 2024 van iedere schoolsoort 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 wordt per centraal examen uitgerekend, zodat het totaal aantal kandidaten dat aan het examen deelnam geen invloed heeft op de resultaten.

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 2024 andere percentages lage scores voorkomen dan in 2019, kan hiermee de geschatte vaardigheidsontwikkeling tussen verschillende jaren deels statistisch verklaard worden.

2.4 Vaardigheid 2024 ten opzichte van 2023

Omdat de vaardigheidsontwikkeling van de examenleerlingen van 2024 t.o.v. 2019 niet meer te vergelijken is met de vaardigheidsontwikkeling zoals gepresenteerd in de rapporten van 2023 tot en met 2021 kunnen we hiermee geen conclusies trekken over de vaardigheidsontwikkeling ten opzichte van de coronajaren. Daarvoor kijken we naar de gemiddelde cijfers van 2024 in vergelijking met 2023.

De gemiddelde cijfers geven de verschillen in vaardigheid weer tussen de kandidaten van 2023 en 2024. De gebruikte cijfers zijn gebaseerd op de normeringssteekproef en de scores van de

11 Stcrt, 2016, 4817; Stcrt, 2021,18168: [wetten.nl - Regeling - Regeling omzetting scores in cijfers VO - BWBR0037590](https://wetten.nl/Regeling-Regeling%20omzetting%20scores%20in%20cijfers%20VO-BWBR0037590) (overheid.nl)

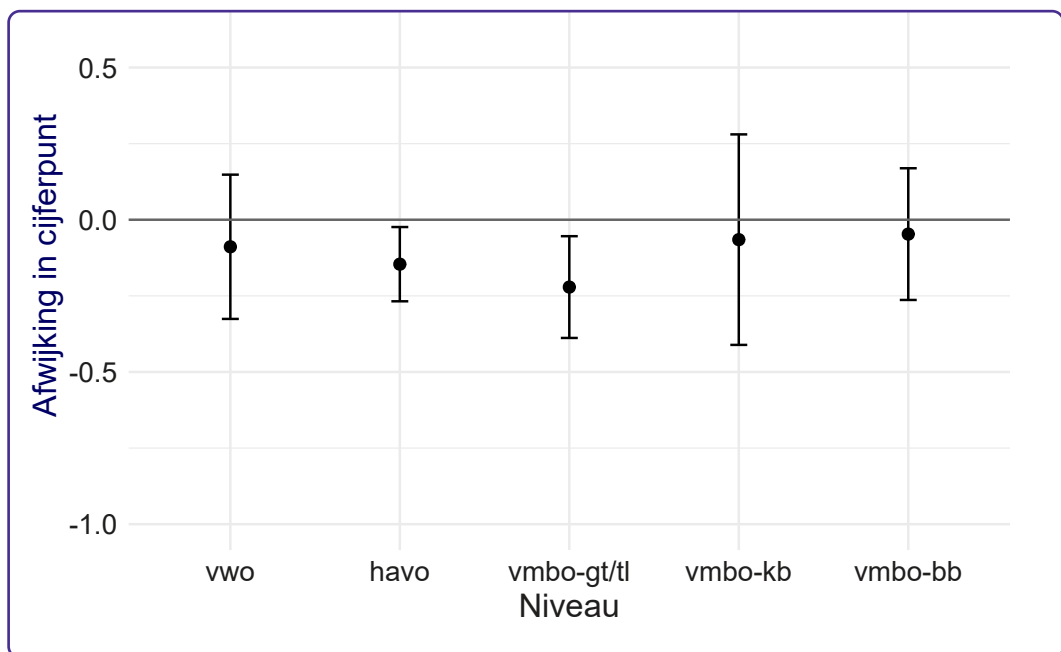
eerste corrector. De cijfers van 2023 zijn gecorrigeerd voor de aanpassingen aan de prestatie-eisen zodat deze vergelijkbaar zijn met die van 2024.

Compensaties via de N-term voor onvolkomenheden in een examen zijn bedoeld om (soms beperkte) subgroepen van kandidaten niet te benadelen. Deze compensaties kwamen slechts bij een deel van de examens voor. Echter, de vaardigheid van de gehele groep wordt beter weergegeven met het gemiddelde cijfer waarbij de compensaties niet meegerekend zijn. De gebruikte gemiddelde cijfers zijn dus zonder eventuele compensaties, en gecorrigeerd voor aanpassingen in prestatie-eisen.

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 van dat cluster valt. Als de vaardigheid in 2024 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 pre-corona had geresulteerd in cijfers die gemiddeld 0,2 cijferpunt lager zouden zijn. 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. Het is belangrijk op te merken dat zowel de gebruikte methode voor het berekenen van de vaardigheid als het aantal vakken die bij een cluster mee zijn genomen, invloed kunnen hebben op zowel de puntschatting als het betrouwbaarheidsinterval.

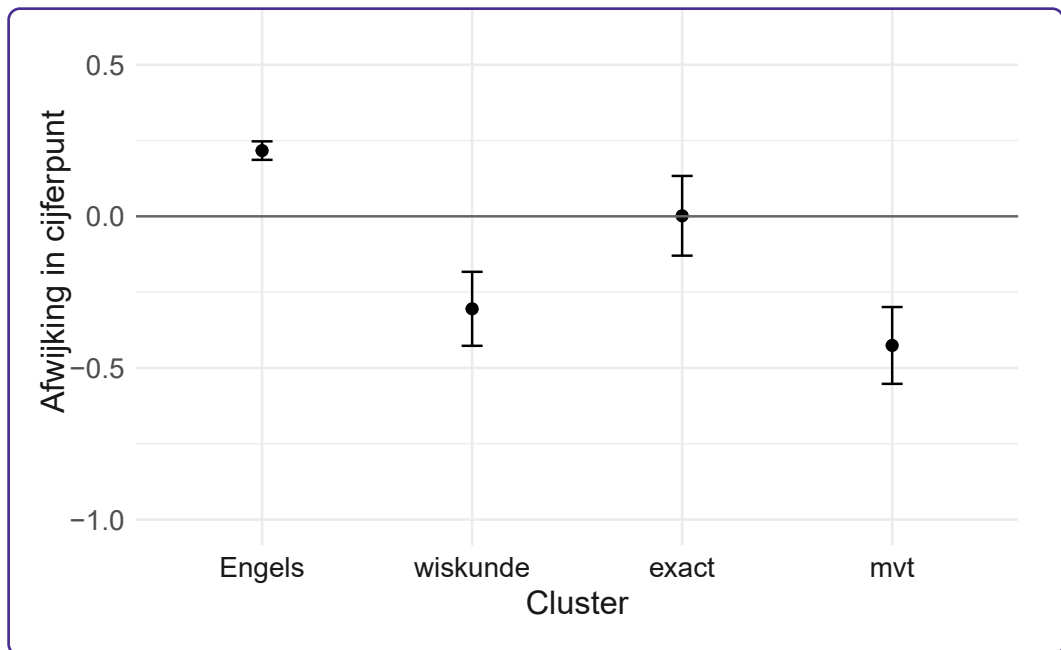


Figuur 1 | Geschat vaardigheidsverschil per schoolsoort voor 2024 ten opzichte van voor corona

Er is bij alle schoolsoorten in 2024 ten opzichte van voor de coronapandemie een daling in vaardigheid te zien. Het verschil is echter alleen statistisch significant voor havo en vmbo-gt/tl, omdat bij alle andere niveaus de betrouwbaarheidsintervallen de nullijn omvatten.

Bij inhoudelijke clustering van centrale examens (Figuur 2) is de gemiddelde vaardigheid in de clusters wiskunde en moderne vreemde talen in 2024 statistisch significant gedaald ten opzichte van voor corona. Bij Engels is de vaardigheid significant toegenomen. Bij exact is er geen significant verschil tussen 2024 en pre-corona te zien. Bij Nederlands omvatte het cluster

informatie over slechts één van de vijf schoolsoorten, en bij mens en maatschappij drie van de voormalig veertien datapunten voor vijf vakken over 5 schoolsoorten (zie tabel 1). Daarom was het niet mogelijk een schatting te geven van de gemiddelde vaardigheid binnen deze clusters van vakken waardoor deze niet zijn opgenomen in de grafiek.

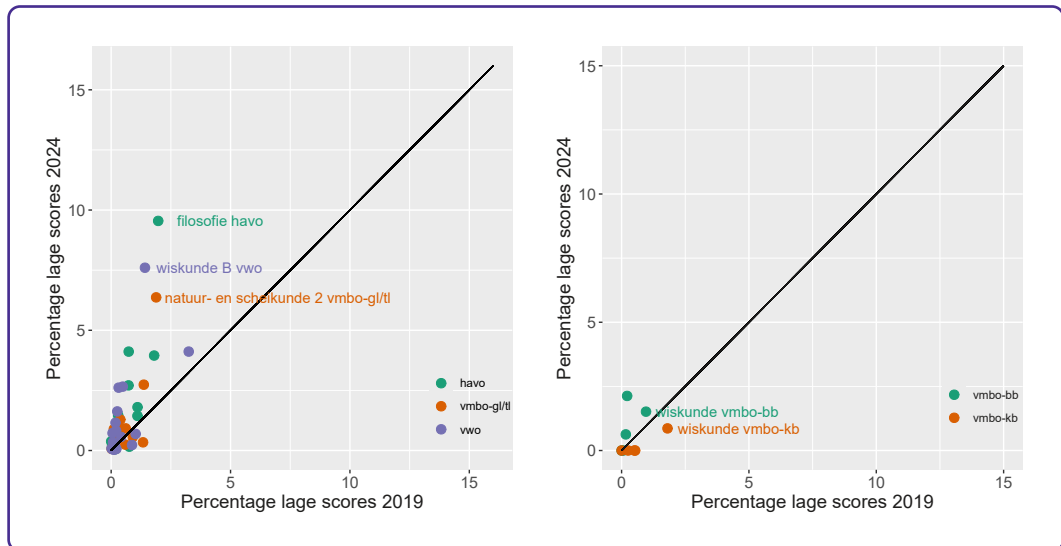


Opmerking: mvt = moderne vreemde talen exclusief Engels

Figuur 2 | Geschat vaardigheidsverschil per cluster voor 2024 ten opzichte van voor corona

3.2 Effect van kandidaten met lage scores

In Figuur 3 (links: vwo, havo en vmbo-gl/tl; rechts: vmbo-bb, vmbo-kb) staan per centraal examen de percentages kandidaten in het eerste tijdvak 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 2024. De schuine lijn geeft aan wanneer het percentage in 2024 gelijk is aan het percentage in 2019. De punten in Figuur 3 (links) liggen grotendeels boven de diagonaal. 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 3 | Percentage kandidaten in het eerste tijdvak met een lage score per CSE in 2019 en 2024 (links) en percentage kandidaten in het eerste tijdvak met een lage score per CBT in 2019 en 2024 (rechts)

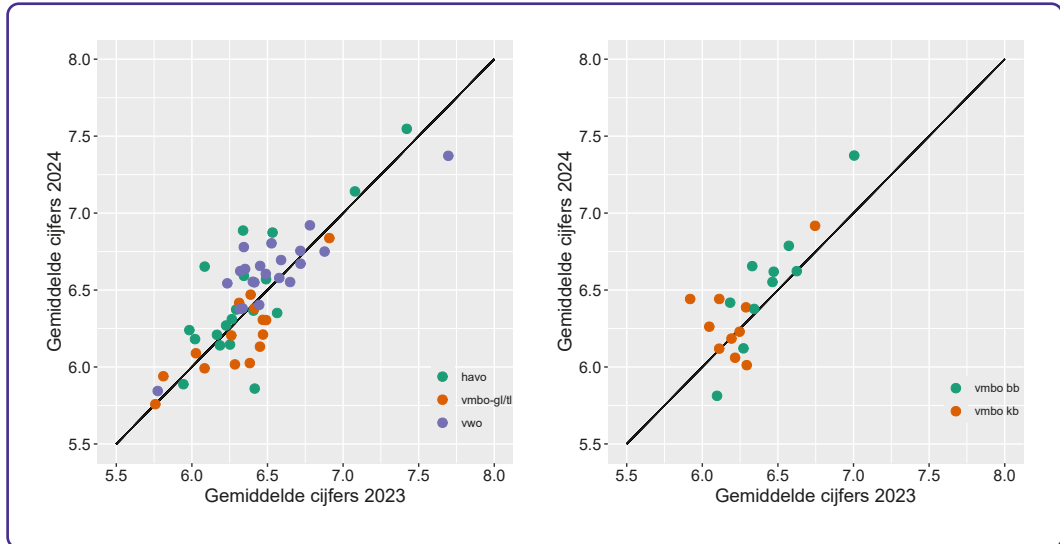
Bij veel centrale examens voor vwo, havo en vmbo-gl/tl is het percentage kandidaten met lage scores in 2024 en 2019 onder de 2% en daarmee vergelijkbaar (Figuur 3, links). Bij sommige centrale examens is in 2024 ten opzichte van 2019 echter nog steeds een verschil te zien. De vakken met meer dan 5% lage scores zijn filosofie havo (9,6% kandidaten met een lage score in 2024 tegenover 2% in 2019), wiskunde B vwo (7,6 % kandidaten met een lage score in 2024 tegenover 1,4% in 2019) en natuur- en scheikunde 2 vmbo-gl/tl (6,4% kandidaten met een lage score in 2024 tegenover 1,9% in 2019). Deze examens waren in 2024 relatief moeilijk en kenden een hoge N-term (resp. 1,9, 2,1, en 1,9), wat het relatief hoge percentage kandidaten met een lage score deels kan verklaren. In 2019 waren de N-termen van deze examens respectievelijk 1,3, 1,8, en 1,6.

Voor een vijfde van alle examens gaat het om een verschil van meer dan 2%. Dit is vergelijkbaar met het aantal lage scores in 2023. Als 2% van de 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 3 (links) toont aan dat met name op havo examenkandidaten nog relatief veel lage scores hebben behaald. Dit zou dus gedeeltelijk de daling in vaardigheid in het examencluster havo kunnen verklaren (Figuur 1).

Voor vmbo-kb en vmbo-bb (Figuur 3, rechts) is er geen examen met meer dan 2,5% lage scores in 2024. Bovendien is het percentage lage scores voor de meeste examens weer in lijn met de percentages van voor de coronapandemie.

3.3 Vaardigheid 2024 ten opzichte van 2023

In Figuur 4 (links: vwo, havo en vmbo-gl/tl; rechts: vmbo-bb, vmbo-kb) staan per centraal examen de gemiddelde cijfers voor alle kandidaten van tijdvak 1. Op de horizontale as staan de gemiddelde cijfers van 2023 en op de verticale as de gemiddelde cijfers van 2024. De schuine lijn geeft aan wanneer het gemiddelde cijfer in 2024 gelijk is aan het gemiddelde cijfer in 2019.



Figuur 4 | Gemiddelde cijfers per CSE in 2023 en 2024 (links) en gemiddelde cijfers per CBT in 2023 en 2024 (rechts)

De punten in Figuur 4 (links) liggen voor havo grotendeels boven de diagonaal in de grafiek. Dit betekent dat de gemiddelde cijfers in 2024 hoger zijn dan in 2023. Voor vwo is het gemiddelde cijfer voor de helft van de vakken in 2024 hoger dan in 2023 en voor vmbo-gl/tl is slechts voor 3 vakken een lichte vooruitgang te constateren.

De punten in Figuur 4 (rechts) liggen grotendeels boven de diagonaal in de grafiek. De gemiddelde cijfers voor vmbo-kb en vmbo-bb zijn dus hoger in 2024 ten opzichte van 2023.

4 Conclusie

De hoofdvraag van het onderzoek was of -het vaardigheidsniveau van de examenleerlingen op het centraal examen weer op het niveau is van voor corona. De resultaten van dit onderzoek laten zien dat voor de schoolsoorten vwo, vmbo-kb en vmbo-bb de vaardigheid in 2024 niet significant verschilt ten opzichte van de vaardigheid voor de coronapandemie. Voor havo en vmbo-gl/tl is een significant lager vaardigheidsniveau ten opzicht van voor de coronapandemie. Bij de vaardigheden voor de clusters moderne vreemde talen (exclusief Engels) en wiskunde zijn de vaardigheden onder het niveau van voor de coronajaren. De gemiddelde vaardigheid in het cluster Engels is hoger dan voor de coronapandemie. Bij het cluster exact is er geen significant verschil in vaardigheid ten opzichte van voor de coronapandemie.

Het aantal lage scores is bij enkele vakken hoger dan voor de coronapandemie. Het is mogelijk dat het aantal lage scores, vooral voor havo een klein effect heeft op de geschatte gemiddelde vaardigheden.

Op basis van de vergelijking van de gemiddelde cijfers in 2024 ten opzichte van 2023 blijkt dat de vaardigheid voor havo, vmbo-bb en vmbo-kb over het algemeen hoger is. Voor vwo geldt voor de helft van de vakken een stijging en voor de andere helft een daling in gemiddeld cijfer. Voor vmbo-gl/tl is voor de meeste vakken geen verschil of een achteruitgang ten opzichte van 2023 waar te nemen.

De bevindingen op basis van zowel de vaardigheidsschattingen tussen 2019 en 2024 en de gemiddelde cijfers van 2023 en 2024 zijn consistent met de resultaten van het vaardigheidsonderzoek van 2023. Bij vwo lijkt het vaardigheidsniveau sinds 2023 weer op het niveau van voor corona. Bij havo is het vaardigheidsniveau nog niet op het niveau van voor corona, maar stijgend ten opzichte van 2023. Bij vmbo-gl/tl week het vaardigheidsniveau in 2023 niet significant af van het niveau voor corona, maar is er in 2024 weer een daling te zien. Bij vmbo-bb en vmbo-kb week het vaardigheidsniveau in 2023 ook niet significant af van het niveau voor corona, en met een verdere stijging in 2024 wordt het herstel daar bevestigd.

5 Literatuur

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

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

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

Cito (geen datum). Tools en informatie voor docenten: Wolf. [Wolf – uitwisselprogramma voor examengegevens \(cito.nl\)](#).

College van Toetsen en Examens (geen datum). Examenblad.nl: [Normering centrale examens | Examenblad.nl](#)

van Onna, M., Schouwstra, S., van der Molen, P., de Groot, M., & Emons, W. (2021). [Rapport Vaardigheid Examenkandidaten 2021 \(bijlage bij 31289,nr.506\) - Parlementaire monitor](#)

Pilz, K., Straat, H., Rietdijk, S., Kruis, J. & de Groot, M. (2022) [Vaardigheid Examenkandidaten 2022 | Rapport | Rijksoverheid.nl](#).

Pilz, K., Straat, H. & Weemink, H. (2023). [Vaardigheid examenkandidaten 2023 | Rapport | Rijksoverheid.nl](#)

[Normering centrale examens | 2024 | Examenblad.nl](#)

Bijlage | Berekeningswijze vaardigheidsverschil tussen 2024 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 2024 te relateren aan de prestaties in de jaren voor de coronapandemie (2019).

Gebruikte data

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.

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) gebruikt is. Bij Anchor in package en pre- en posttest examens is een trendschattingwijze gebruikt om de vaardigheid voor en na corona te vergelijken. In alle gevallen resulteerde een verschil in gemiddeld cijfer tussen de pre-corona-populatie en de populatie in 2024, 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 2024, 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-analyse¹² 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 – 2024 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 2024. Een daling in gemiddelde scores tussen deze twee gemiddelden reflecteerde dus een daling in vaardigheid. Het verschil in gemiddelde scores van 2024 en de pre-corona-periode werd uitgedrukt in een verschil in gemiddeld cijfer¹³ om de getallen beter te kunnen duiden, en om centrale examens te kunnen aggregeren.

12 Vanwege het grote aantal centrale examens zijn er geen uitgebreide analyses op modelfit gedaan. Dit is eveneens een reden waarom de afzonderlijke trendschattingen niet gepresenteerd worden.

13 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.

De puntschatting van dit zogenaamde contrast werd voorzien van een nauwkeurighedsmaat¹⁴. 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.

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 alle examen-specifieke verschillen op dezelfde cijferpuntenschaal waren uitgedrukt. Enerzijds zijn er berekeningen per schoolsoort 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 2024-populaties werd berekend als een gewogen gemiddelde van de examen-specifieke verschillen D_k in vaardigheid,

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

waarbij de gewichten w_k de nauwkeurigheid van de examen-specifieke schatting weergaven¹⁵.

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.

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

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

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

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

15 $w_k = 1 / (SE_k)^2$

CITO CTE ontwikkelt en adviseert bij wettelijke toetsen

Cito

Amsterdamseweg 13
6814 CM Arnhem
Postbus 1034
6801 MG Arnhem
T (026) 352 11 11
www.cito.nl

Fotografie: Gijs Versteeg
© Stichting Cito Arnhem (2024)

