



# Doorstroom- onderzoek eindtoetsen 2022

Bevindingen en impact op  
het doorstroomtoetsstelsel

Dit onderzoek is door  
Stichting Cito uitgevoerd

# Inhoud

<b>Managementsamenvatting</b>	<b>3</b>
<b>Achtergrond</b>	<b>4</b>
De toetsadviesgrenzen in de eind- en doorstroomtoetsen	4
De toetsadviesgrenzen evalueren met doorstroomonderzoek	4
Het huidige doorstroomonderzoek	5
De toetsadviesgrenzen van de Centrale Eindtoets	5
De normering van de eindtoetsen in 2022	5
<b>Doelgroep en data</b>	<b>7</b>
Doelgroep	7
Data	7
Exclusiecriteria	7
<b>Beschrijvende statistieken</b>	<b>8</b>
<b>Onderzoeksvraag 1: Hoe hangen de toetsscores van de Centrale Eindtoets van 2022 samen met het onderwijsniveau in vo leerjaar 3?</b>	<b>11</b>
Conclusies	14
<b>Onderzoeksvraag 2a: Hoe sterk voorspellen de toetsadviezen en de voorlopige schooladviezen uit 2022 de plaatsing in het vo in leerjaar 1, 2 en 3?</b>	<b>17</b>
Conclusies	20
<b>Onderzoeksvraag 2b: Wat is de relatie tussen het onderwijsniveau in leerjaar 1, het toetsadvies en de op- en afstroom tussen leerjaar 1 en 3 op het vo?</b>	<b>22</b>
Conclusies	23
<b>Discussie</b>	<b>24</b>
Beperkingen van de huidige evaluatie	24
Implicaties voor het doorstroomtoetsstelsel	24
<b>Verwijzingen</b>	<b>25</b>
<b>Appendix A: Toelichting landelijke normering commerciële eindtoetsen 2022</b>	<b>26</b>
<b>Appendix B: Toelichting pseudonimisering en dataverwerking</b>	<b>28</b>
Pseudonimisering	28
Dataverwerking	28
Ontbrekende data	30
Gegevenskoppeling	31
<b>Appendix C: Beschrijvende statistieken</b>	<b>33</b>
<b>Appendix D: Simulatiestudie misclassificaties normering</b>	<b>35</b>
Deelstudie 1: Inschatting van classificatiefout toetsadvies door meetonzekerheid	35
Deelstudie 2: Effect van IRT2-normering op de classificatiefout van het toetsadvies	36

# Managementsamenvatting

Basisschoolleerlingen maken in groep 8 een doorstroomtoets (voorheen: eindtoets). Het toetsadvies wordt, naast het voorlopig schooladvies, gebruikt als objectief tweede gegeven over het best passende onderwijsniveau in het voortgezet onderwijs (vo). Met doorstroomonderzoek volgen we leerlingen in de eerste drie jaar van het vo. Hiermee kan de voorspellende waarde van de toetsadviezen geëvalueerd worden. Daarvoor kijken we of de gegeven toetsadviezen overeenkomen met het onderwijsniveau van leerlingen in de eerste drie jaar van het vo. Op basis daarvan kunnen eventueel de gehanteerde grenzen tussen de toetsadviescategorieën worden bijgesteld.

Stichting Cito heeft, in opdracht van het College van Toetsen en Examens (CvTE), een verkennend doorstroomonderzoek in het vo uitgevoerd op basis van de eindtoetsafname van schooljaar 2021-2022, het moment waarop gestart is met dataverzameling voor doorstroomonderzoek. Dit rapport beschrijft de resultaten van het onderzoek dat uitgevoerd is op leerlingen die in 2022 een eindtoets maakten, in het vorige stelsel, maar de resultaten zijn ook relevant voor het huidige doorstroomtoetsstelsel. De grensscores tussen de toetsadviezen die in het doorstroomtoetsstelsel worden gebruikt zijn een voortzetting van de grensscores zoals gebruikt bij de Centrale Eindtoets in het eindtoetsstelsel. Met dit onderzoek, en ook met toekomstig doorstroomonderzoek, kunnen we deze toetsadviesgrenzen evalueren. In schooljaar 2026-2027 kunnen we ook de doorstroom onderzoeken van leerlingen die de eerste doorstroomtoets maakten in 2024.

Uit dit rapport blijkt dat de toetsadviesgrenzen van de Centrale Eindtoets, die de basis vormden voor de grenzen in het doorstroomtoetsstelsel, in 2022 een goede voorspelling waren van het onderwijsniveau in het derde jaar van het vo. Een kleine aanpassing in de toetsadviesgrenzen zou de voorspelling iets verbeteren. De aanbeveling is om de toetsadviesgrenzen op dit moment nog niet bij te stellen maar de evaluaties tot en met het eerste leerlingcohort dat de doorstroomtoets maakte (2024) af te wachten om te zien of dit effect doorzet. Hiermee voorkomen we de situatie dat toetsadviesgrenzen jaar op jaar bij geringe verschillen die kunnen worden veroorzaakt door ruis moeten worden aangepast.

Het blijkt dat zowel de toetsadviezen als de voorlopige schooladviezen een goede voorspeller zijn van plaatsing van leerlingen in vo leerjaar 3. Zo verbleef 82% van de leerlingen in het derde jaar van het vo op een onderwijsniveau passend bij het gegeven toetsadvies en 79% op een niveau passend bij het voorlopig schooladvies. Samen onderschrijven deze uitkomsten de voorspellende validiteit van zowel het toetsadvies als het voorlopig schooladvies. Ook zien we dat leerlingen vaker afstromen als ze op een hoger onderwijsniveau in het vo starten dan het toetsadvies. Leerlingen stromen ook vaker op als ze op een lager onderwijsniveau dan het toetsadvies starten. Dit zijn leerlingen wiens voorlopig schooladvies niet is bijgesteld naar het toetsadvies of die om een andere reden op een lager onderwijsniveau zijn gestart. Hieruit blijkt de meerwaarde van het toetsadvies in het eindtoetsstelsel, en dit onderstreept het belang van het toetsadvies als objectief tweede gegeven naast het schooladvies voor de plaatsing van leerlingen in het vo.

Tot slot blijkt voor iedere eindtoets een vergelijkbare relatie te bestaan tussen het voorlopig schooladvies en de plaatsing van leerlingen. Dit betekent dat de scholen die voor een bepaalde toets kiezen vergelijkbaar zijn in de manier waarop voorlopige schooladviezen worden gegeven aan leerlingen met een gelijke vaardigheid. Deze bevinding is interessant in de context van de waargenomen prestatieverschillen tussen leerlingen met hetzelfde voorlopig schooladvies die verschillende toetsen maken in zowel het eind- als doorstroomtoetsstelsel. Een selectie-effect op schoolniveau met betrekking tot de voorlopige schooladviezen kan hier slechts van beperkte invloed zijn.

# Achtergrond

## De toetsadviesgrenzen in de eind- en doorstroomtoetsen

De eind- en doorstroomtoetsen hebben als doel leerlingen, ouders en leerkrachten te informeren over een passend onderwijsniveau voor de leerling in het vo. Dit noemen we het toetsadvies. Na deelname aan de toets krijgt iedere leerling één van zes mogelijke toetsadviezen (**tabel 1**). Welk toetsadvies een leerling krijgt, wordt bepaald door de prestaties van de leerling op de toets en de vastgestelde toetsadviesgrenzen. De toetsadviesgrenzen, ook cesuren genoemd, geven aan welke toetsprestatie tot welk toetsadvies leidt. Zo kreeg een leerling in 2022 met een standaardscore van 510 op de Centrale Eindtoets een pro/vmbo bb toetsadvies en een leerling met score 544 een havo/vwo toetsadvies.

**Tabel 1: Toetsadviesgrenzen op de standaardscoreschaal van de Centrale Eindtoets, ongecorrigeerd en met de coronacorrectie zoals toegepast in 2022. De maximumscore was 550.**

toetsadvies	ondergrens Centrale Eindtoets	
	ongecorrigeerd	gecorrigeerd 2022
pro/vmbo bb	501	501
vmbo bb/vmbo kb	511	511
vmbo kb/vmbo gl-tl	525	524
vmbo gl-tl/havo	533	532
havo/vwo	540	540
vwo	545	545

## De toetsadviesgrenzen evalueren met doorstroomonderzoek

De toetsadviesgrenzen moeten periodiek geëvalueerd worden om ervoor te zorgen dat ze de best mogelijke voorspelling geven van een passend onderwijsniveau. Die evaluatie doen we met doorstroomonderzoek. In doorstroomonderzoek wordt gekeken of de gegeven toetsadviezen uit de eind- en doorstroomtoets goed samenhangen met de daadwerkelijke plaatsing van leerlingen in leerjaar 3 van het vo. Als bijvoorbeeld blijkt dat veel leerlingen met een toetsadvies vmbo gl-tl/havo terecht zijn gekomen op vmbo kb, dan hebben die leerlingen een te hoog toetsadvies gekregen. Er kan dan overwogen worden om de toetsadviesgrens tussen vmbo kb/vmbo gl-tl en vmbo gl-tl/havo te verhogen. Als echter blijkt dat veel van de leerlingen met een toetsadvies vmbo gl-tl/havo terecht zijn gekomen op vwo, dan hebben die leerlingen een te laag toetsadvies gekregen. Dan kan worden overwogen om de grens tussen vmbo gl-tl/havo en havo/vwo te verlagen. Op die manier kan het doorstroomonderzoek helpen om de grenzen bij te stellen zodat zo veel mogelijk leerlingen een passend toetsadvies krijgen.

De bevindingen uit doorstroomonderzoek worden alleen gebruikt voor *toekomstige* toetsafnames. We moeten namelijk eerst drie jaar lang leerlingen volgen vanaf het maken van de eind- of doorstroomtoets tot en met de plaatsing in het derde jaar van het vo. Alhoewel het eerste en tweede leerjaar ook onderzocht worden, is vooral het derde leerjaar van belang. Het derde leerjaar is een gunstig moment om de doorstroom van leerlingen te onderzoeken omdat er vanaf dat leerjaar geen brede klassen meer bestaan. Daarnaast krijgen leerlingen drie jaar de tijd om op een voor hen passend niveau uit te komen, waardoor de invloed van een eventueel minder goed passend schooladvies of startniveau wordt verkleind. Als uit het doorstroomonderzoek blijkt dat de toetsadviesgrenzen moeten worden bijgesteld, zal dat enkel een impact hebben op de leerlingen die hun doorstroomtoets nog moeten maken.

## Het huidige doorstroomonderzoek

Dit onderzoek betreft leerlingen die in schooljaar 2021-2022 de eindtoets hebben gemaakt. Het CvTE verzamelt namelijk de resultaten van de eind- en doorstroomtoetsen van de leerlingen vanaf dat schooljaar. Ook verzamelt het CvTE de plaatsingsgegevens uit het vo van deze leerlingen. In schooljaar 2024-2025 heeft het eerste cohort leerlingen sinds het begin van de gegevensverzameling het derde leerjaar in het vo bereikt. Daarom is het vanaf dat schooljaar mogelijk om periodiek doorstroomonderzoek uit te voeren.

Dit rapport is opgedeeld in twee onderwerpen. Samen geven die onderwerpen inzicht in hoe geschikt de toetsadviesgrenzen in de eind- en doorstroomtoets zijn op basis van de daadwerkelijke plaatsing van leerlingen in het vo. Ten eerste evalueren we de toetsadviesgrenzen. Specifiek onderzoeken we de grenzen van de Centrale Eindtoets, omdat die toetsadviesgrenzen ook worden gebruikt in het doorstroomtoetsstelsel vanaf schooljaar 2023-2024. Ten tweede onderzoeken we de toegepaste normering in het eindtoetsstelsel van 2022. We kijken naar de voorspellende waarde van de school- en toetsadviezen van alle toetsaanbieders en de daarmee samenhangende op- en afstroom van leerlingen in het vo. Hieronder lichten we beide onderwerpen verder toe.

## De toetsadviesgrenzen van de Centrale Eindtoets

De toetsadviesgrenzen van het doorstroomtoetsstelsel vanaf schooljaar 2023-2024 zijn een voortzetting van de toetsadviesgrenzen die in 2019 zijn vastgesteld voor de Centrale Eindtoets<sup>1</sup>. Dat betekent dat een eventuele bijstelling van de grenzen van de Centrale Eindtoets op basis van doorstroomonderzoek ook kan leiden tot een bijstelling van de grenzen van het doorstroomtoetsstelsel. Daarom evalueren we of de toetsadviesgrenzen die destijds gehanteerd zijn (**tabel 1**) goed passen bij de doorstroom van leerlingen die in 2022 de Centrale Eindtoets hebben gemaakt. Omdat in 2022 de gevolgen van de leerachterstanden van de COVID-19-pandemie nog merkbaar waren, werd voor alle eindtoetsen een *coronacorrectie* toegepast op de toetsadviesgrenzen tussen vmbo bb/vmbo kb en vmbo kb/vmbo gl-tl en tussen vmbo kb/vmbo gl-tl en vmbo gl-tl/havo (zie ook CvTE, 2022). Voor alle toetsadviesgrenzen bepalen we of een lichte verschuiving van een grens zou leiden tot een betere of slechtere voorspelling van de plaatsing. De eerste onderzoeksvraag is daarom:

### 1. Hoe hangen de toetsscores van de Centrale Eindtoets uit 2022 samen met het onderwijsniveau in vo leerjaar 3?

Als een verschuiving van de grenzen zou leiden tot een betere voorspelling van de plaatsing van leerlingen in vo leerjaar 3, dan kan dit leiden tot een bijstelling van de toetsadviesgrenzen in het doorstroomtoetsstelsel.

## De normering van de eindtoetsen in 2022

In 2022 bestonden er zes eindtoetsen van vijf toetsaanbieders waar een school uit kon kiezen: de AMN Eindtoets van AMN, de papieren en de digitale Centrale Eindtoets (CET) van CvTE, Dia van Diataal, IEP van Bureau ICE en ROUTE 8 van A-vision. Van deze toetsen waren alleen de papieren Centrale Eindtoets en de IEP op papier. De andere toetsen waren digitaal. Net als in het doorstroomtoetsstelsel hadden deze eindtoetsen allen als doel leerlingen een zo goed mogelijk tweede, onafhankelijk advies naast het voorlopig schooladvies te geven over een passend niveau in het vo<sup>2</sup>. De beschikbare doorstroomgegevens stellen ons in staat om de voorspellende waarde van de gegeven toetsadviezen van alle eindtoetsaanbieders te onderzoeken en te vergelijken. De eindtoetsafname in 2022 kende wel een aantal bijzonderheden die we hieronder toelichten.

In 2022 werden door de Centrale Eindtoets en de commerciële eindtoetsen verschillende normeringsmethoden voor de toetsadviezen gebruikt. De Centrale Eindtoets gebruikte een absolute normering gebaseerd op doorstroomonderzoek dat in de periode 2015-2018 is uitgevoerd. Ankeropgaven zorgen er vervolgens in deze normering voor dat ieder afnamejaar op dezelfde, vergelijkbare schaal genormeerd kan worden. Deze absolute normeringsmethode is vergelijkbaar met de methode zoals die nu nog gebruikt wordt in het doorstroomtoetsstelsel.

---

1 Koppeling via een equipercientiel-equivalering, zie voor een toelichting CvTE (2024)

2 Daarnaast werd de eindtoets ook gebruikt om de beheersing van de referentieniveaus Lezen, Rekenen en Taalverzorging te meten.

De commerciële eindtoetsen gebruikten in 2022 de IRT2-normering. In deze normeringsmethode werd een inschatting gemaakt van de vaardigheid van de populatie van iedere eindtoets ten opzichte van de Centrale Eindtoets. Deze inschatting vond plaats op basis van de verdeling van voorlopige schooladviezen van de deelnemende leerlingen. In appendix A is een uitgebreide beschrijving van de normeringsmethode van 2022 terug te vinden. De reden om in 2022 de IRT2-normering toe te passen bij de commerciële eindtoetsen was dat uit analyses van voorgaande afnamejaren bleek dat er niet goed verklaarbare prestatieverschillen bestonden tussen de verschillende eindtoetsen als via het gezamenlijk anker werd geëquivalereerd (Brouwer & Veldkamp, 2022). In de IRT2-normering zijn de toetsadviezen echter niet meer helemaal onafhankelijk van de voorlopige schooladviezen. Daarnaast was de verwachting dat met relatief simpele ingrepen in de toetsopbouw en de manier waarop het gezamenlijk anker ingezet wordt de verschillen tussen de eindtoetsen verkleind zouden worden. De IRT2-normering is daarom destijds ingezet als een tijdelijke maatregel.

Met de beschikbare doorstroomgegevens kunnen we de toepassing van de IRT2-normering evalueren. Voor de IRT2-normering wordt namelijk aangenomen dat er geen verschillen zijn in de manier waarop voorlopige schooladviezen worden gegeven aan leerlingen met gelijke vaardigheid door scholen die voor een bepaalde eindtoets kiezen. Als deze verschillen er wel zijn, zou dit tot systematische verschillen moeten leiden in de voorspellende waarde van de voorlopige schooladviezen en toetsadviezen bij de verschillende eindtoetsen.

Daarnaast kunnen we met deze gegevens mogelijk een deel van de waargenomen prestatieverschillen tussen toetsen in het doorstroomtoetsstelsel verklaren. In het doorstroomtoetsstelsel is het namelijk nog onduidelijk of de prestatieverschillen van leerlingen met gelijke voorlopige schooladviezen worden veroorzaakt door selectie-effecten op schoolniveau of door interacties tussen leerlingen en doorstroomtoetsen (CvTE, 2025). Een selectie-effect op schoolniveau kan ontstaan wanneer scholen die op elkaar lijken zich vaker bij dezelfde toetsaanbieder inschrijven, bijvoorbeeld scholen die optimistisch zijn in het geven van voorlopige schooladviezen of scholen die minder gemotiveerd zijn bij het maken van toetsen. Als dergelijke selectie- of clustereffecten bestaan, dan zal de voorspellende waarde van het voorlopig schooladvies of toetsadvies systematisch verschillen tussen toetsen. Het ligt niet voor de hand dat deze selectiedynamiek op schoolniveau in het eindtoetsstelsel anders was dan in het doorstroomtoetsstelsel.

Om zowel de voorspellende waarde in algemene zin als eventuele verschillen tussen eindtoetsen, al dan niet als gevolg van selectie-effecten op schoolniveau, te onderzoeken kijken we naar de overeenstemming tussen voorlopige schooladviezen en toetsadviezen en de plaatsing van leerlingen in het vo:

**2a. Hoe sterk voorspellen de toetsadviezen en de voorlopige schooladviezen uit 2022 de plaatsing van leerlingen in het vo in leerjaar 1, 2 en 3?**

Onderzoeksvraag 2a richt zich voornamelijk op de voorspellende waarde van beide adviezen voor de plaatsing in leerjaar 3. Daarnaast kunnen we de voorspellende waarde onderzoeken door te kijken naar de op- en afstroom van leerlingen tussen de leerjaren:

**2b. Wat is de relatie tussen het onderwijsniveau in leerjaar 1, het toetsadvies en de op- en afstroom tussen leerjaar 1 en 3 op het vo?**

# Doelgroep en data

## Doelgroep

Dit onderzoek betreft de leerlingen in het vo die in kalenderjaar 2022 (schooljaar 2021-2022) een eindtoets hebben gemaakt. Na het maken van de eindtoets zijn deze leerlingen aan het begin van schooljaar 2022 begonnen aan het vo waar we ze drie kalenderjaren hebben gevolgd. Aan het eind van kalenderjaar 2022 en 2023 volgde het merendeel van deze leerlingen leerjaar 1 respectievelijk 2 van het vo. Aan het eind van kalenderjaar 2024 zat het merendeel van deze leerlingen in leerjaar 3 van het vo.

## Data

Voor dit onderzoek zijn twee databronnen gecombineerd. Enerzijds levert de Dienst Uitvoering Onderwijs (DUO) voorlopige en definitieve schooladviezen en gevolgde onderwijsniveaus in het vo van leerlingen die in 2022 een eindtoets hebben gemaakt. Anderzijds leveren alle eindtoetsaanbieders data over de resultaten van de eindtoets, inclusief toetsscores en toetsadviezen. We hebben data van DUO en de toetsaanbieders gekoppeld met een pseudoniem. Dit pseudoniem zorgt ervoor dat er geen persoonsgegevens hoeven te worden opgeslagen en verwerkt in dit onderzoek. Zie hoofdstuk 5 van de Regeling beoordelingsnormen (Ministerie van Onderwijs, Cultuur en Wetenschap, 2025) en appendix B voor een toelichting op de pseudonimisering en dataverwerking.

## Exclusiecriteria

De volgende groepen leerlingen vallen niet binnen de doelgroep van dit onderzoek en zijn als exclusiecriteria gehanteerd:

- Leerlingen met een ontheffing. Deze leerlingen hoeven vanwege verschillende redenen geen eindtoets te maken, bijvoorbeeld omdat ze pas kort in Nederland wonen.
- Leerlingen in het speciaal basisonderwijs (sbo) of speciaal onderwijs (so). Dit is een heterogene en complexe populatie waarbij in de doorstroom andere aspecten een rol kunnen spelen dan bij leerlingen uit het regulier basisonderwijs.
- Leerlingen met een voorlopig of definitief schooladvies voor voortgezet speciaal onderwijs (vso). Het onderwijsniveau en leerjaar van deze leerlingen in vo was niet voldoende eenduidig geregistreerd om gegronde uitspraken op te kunnen baseren.

Daarnaast zijn gegevens van leerlingen uit de dataset verwijderd:

- van wie het pseudoniem niet voorkwam in de data van DUO of de toetsaanbieder;
- die hetzelfde pseudoniem kregen omdat ze twee- of meerling zijn met dezelfde toetsscore;
- van wie de gegevens fouten of inconsistenties bevatten.

Na het koppelen van de data van DUO en de eindtoetsaanbieders, bleek er voor 94,8% van de leerlingen een overlappend pseudoniem aanwezig (172 222 van de 181 657 leerlingen). Daarna zijn de exclusiecriteria zoals hierboven toegepast, waarna 160 912 leerlingen overbleven. Na het toepassen van de exclusiecriteria hebben we gekeken naar ontbrekende data. De ontbrekende data waren gelijkmatig verdeeld over de toetsaanbieders en betroffen ongeveer 0,2% van de leerlingpopulatie. Zie appendix B voor een toelichting op de ontbrekende data. In totaal bleven er 160 059 leerlingen over voor de analyses zonder ontbrekende data. Tenzij anders aangegeven kijken we voor iedere analyse alleen naar leerlingen die een nominaal onderwijstraject volgen (het verblijfjaar en leerjaar zijn gelijk).

# Beschrijvende statistieken

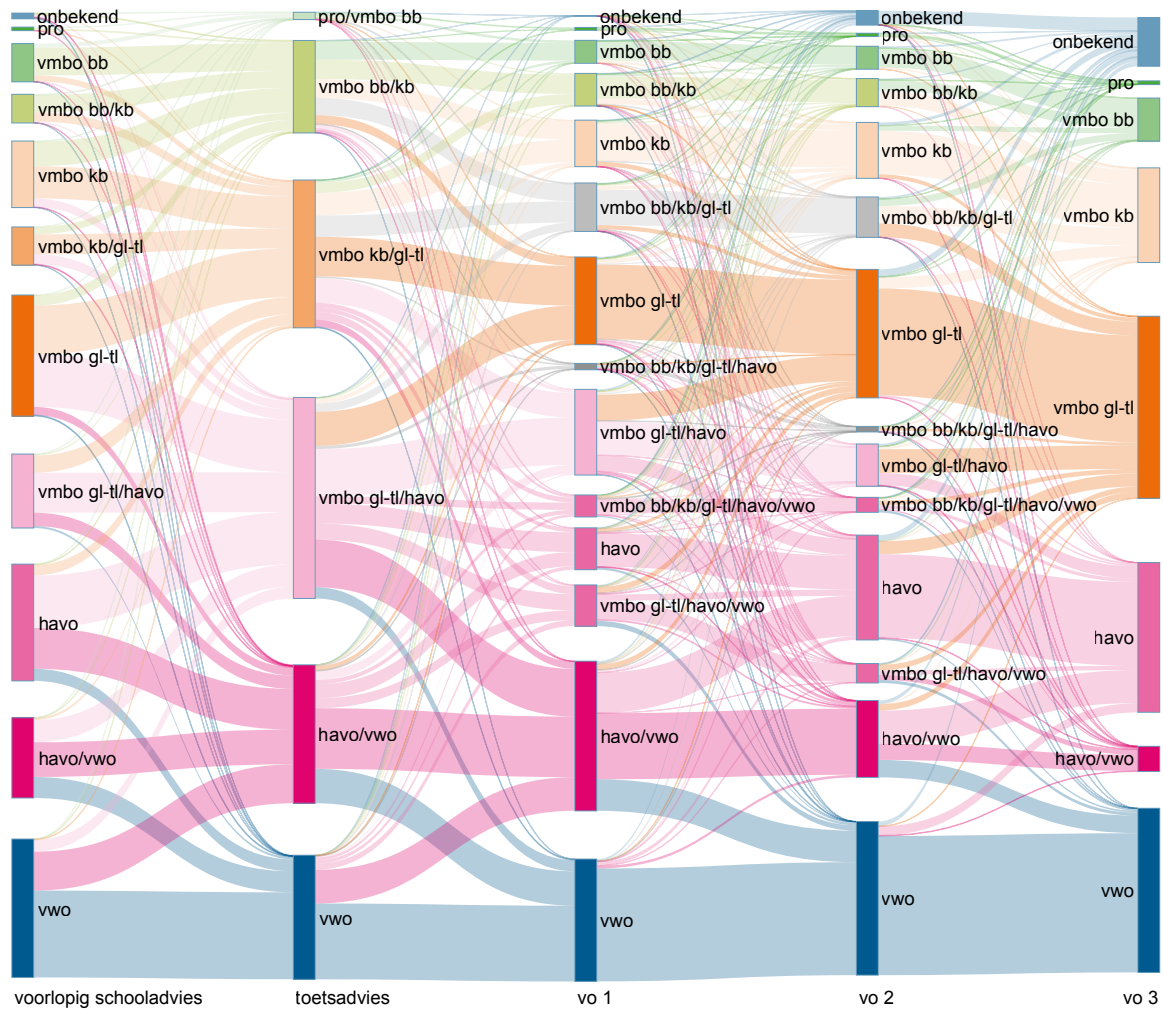
In totaal omvatte dit onderzoek gegevens van 160 059 leerlingen die in 2022 de eindtoets hebben gemaakt. **Tabel 2** beschrijft het aantal leerlingen per eindtoets die in dit onderzoek gevolgd zijn. In appendix C beschrijven **tabellen C1 tot en met C4** de gegeven voorlopige schooladviezen, de gegeven toetsadviezen, het aantal leerlingen dat ooit doubleert of versnelt en het aantal bijstellingen van de gegeven schooladviezen per eindtoets.

**Tabel 2:** Aantal leerlingen per gemaakte eindtoets in 2022 die in dit onderzoek gevolgd zijn.

eindtoets	aantal leerlingen
AMN	3 158
CET (digitaal)	21 098
CET (papier)	51 057
Dia	6 702
IEP (papier)	56 861
ROUTE 8	21 183
<b>totaal</b>	<b>160 059</b>

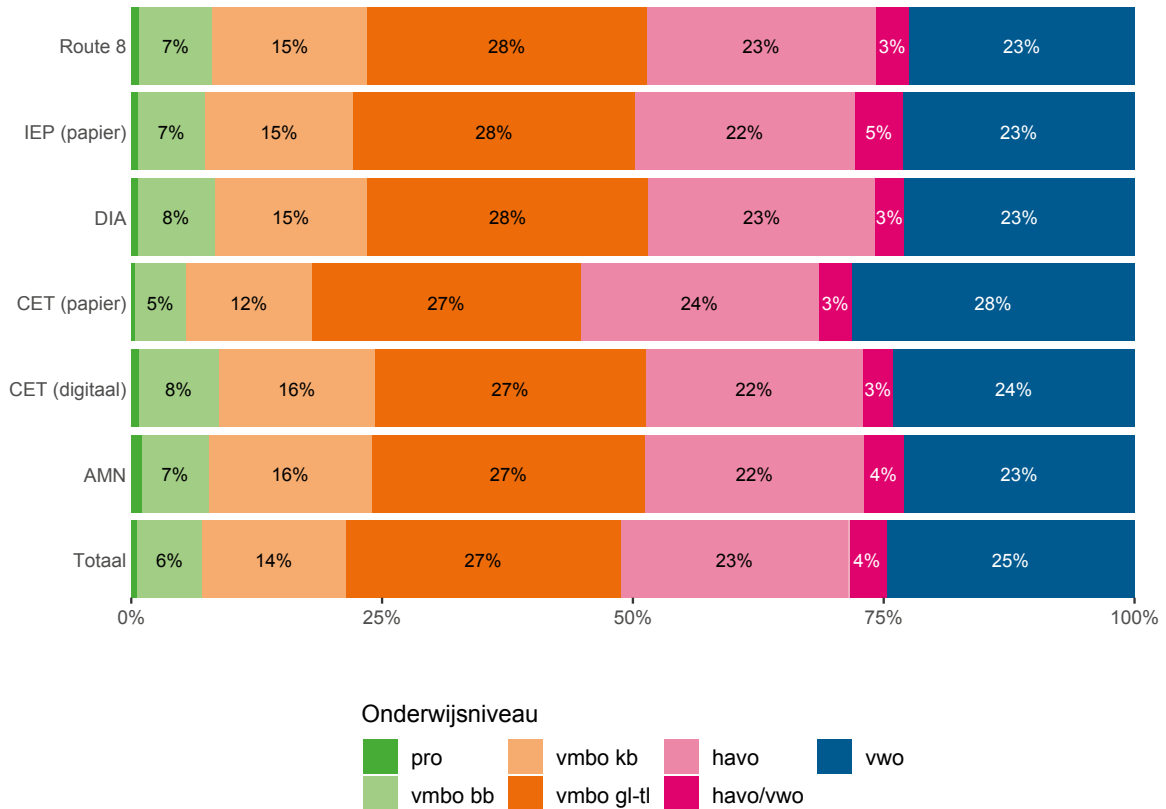
**Figuur 1** geeft de leerlingstromen weer tussen voorlopig schooladvies en toetsadvies en onderwijsniveau in het vo in leerjaar 1, 2 en 3. De hoogte van de balk geeft het aandeel leerlingen in dat onderwijsniveau weer en laat zien waar zij in de volgende stap naartoe gaan. Logischerwijs komen er in de kolom met toetsadviezen meerdere stromen samen vanuit de voorlopige schooladviezen, omdat de toetsadviezen bredere combinaties van onderwijsniveaus zijn. Vervolgens is te zien dat in vo leerjaar 1 door de introductie van brede brugklassen er veel beweging lijkt te zijn naar verschillende niveaus. Deze breedte neemt af in leerjaar 2 en leerjaar 3, omdat leerlingen naar een specifiek niveau stromen. Verder valt op dat leerlingen gelijkmatig doorstromen van het toetsadvies naar de verschillende onderwijsniveaus binnen dat toetsadvies. Zo is bijvoorbeeld te zien dat de leerlingen met een havo/vwo-toetsadvies zich verspreiden over brede brugklassen, havo/vwo- en vwo-classes en dat leerlingen met een vmbo kb/vmbo gl-tl toetsadvies zich voornamelijk en gelijkmatig verspreiden over vmbo kb-, vmbo gl-tl-, en vmbo gl-tl/havo-classes.

**Figuur 1:** Leerlingstromen van voorlopig schooladvies naar toetsadvies naar onderwijsniveau in het vo in leerjaar 1, 2 en 3, voor alle eindtoetsen samen.

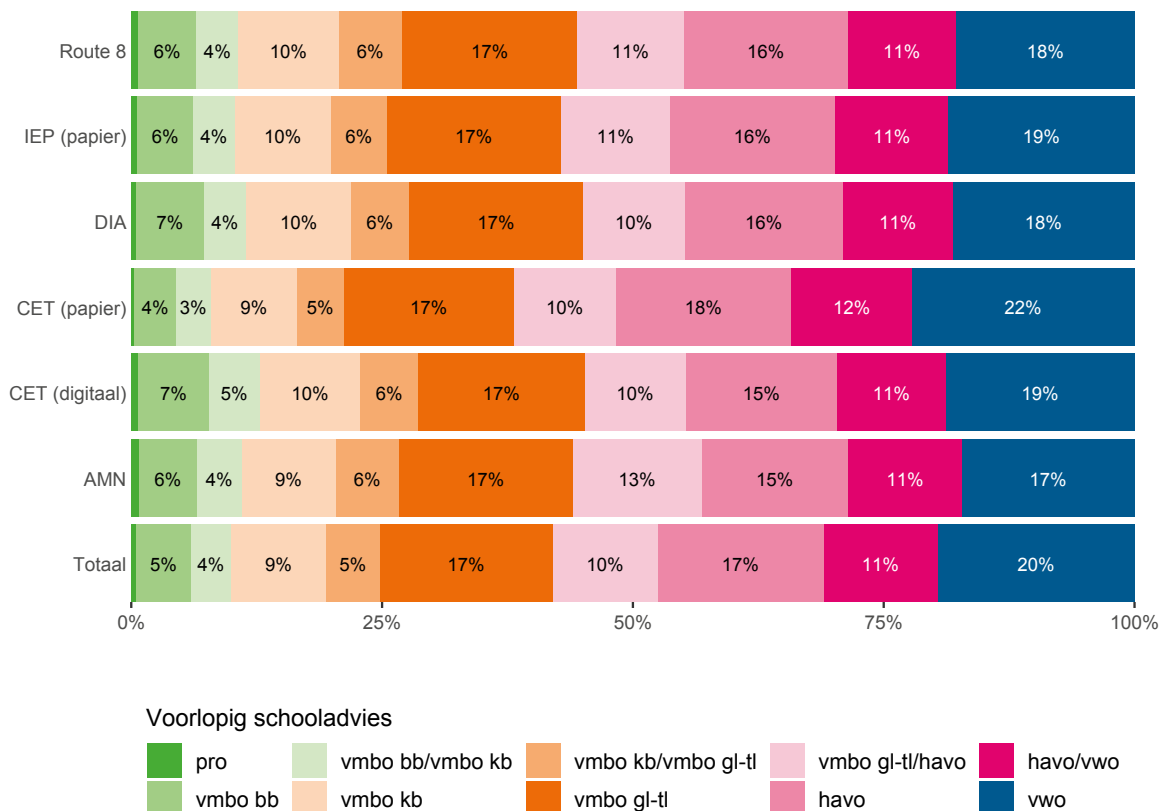


**Figuur 2** laat de verdeling zien van onderwijsniveaus die in leerjaar 3 worden gevolgd door de leerlingen van iedere eindtoets. **Figuur 3** laat dezelfde verdeling zien, maar dan voor het voorlopige schooladvies. Beide figuren geven een impressie van de samenstelling van de leerlingpopulaties van iedere eindtoets. Over het algemeen is de samenstelling van de leerlingpopulaties redelijk vergelijkbaar tussen de eindtoetsen. Alleen de leerlingpopulatie van de papieren Centrale Eindtoets is enigszins vaardiger dan de populaties van de andere eindtoetsen. Dit is te zien aan het aandeel leerlingen met onderwijsniveau of schooladvies vwo dat hoger is en het aandeel vmbo dat lager is.

**Figuur 2:** De samenstelling van de populatie per eindtoets uitgedrukt in percentages onderwijsniveau in leerjaar 3.



**Figuur 3:** De samenstelling van de populatie per eindtoets uitgedrukt in percentages voorlopige schooladviezen.

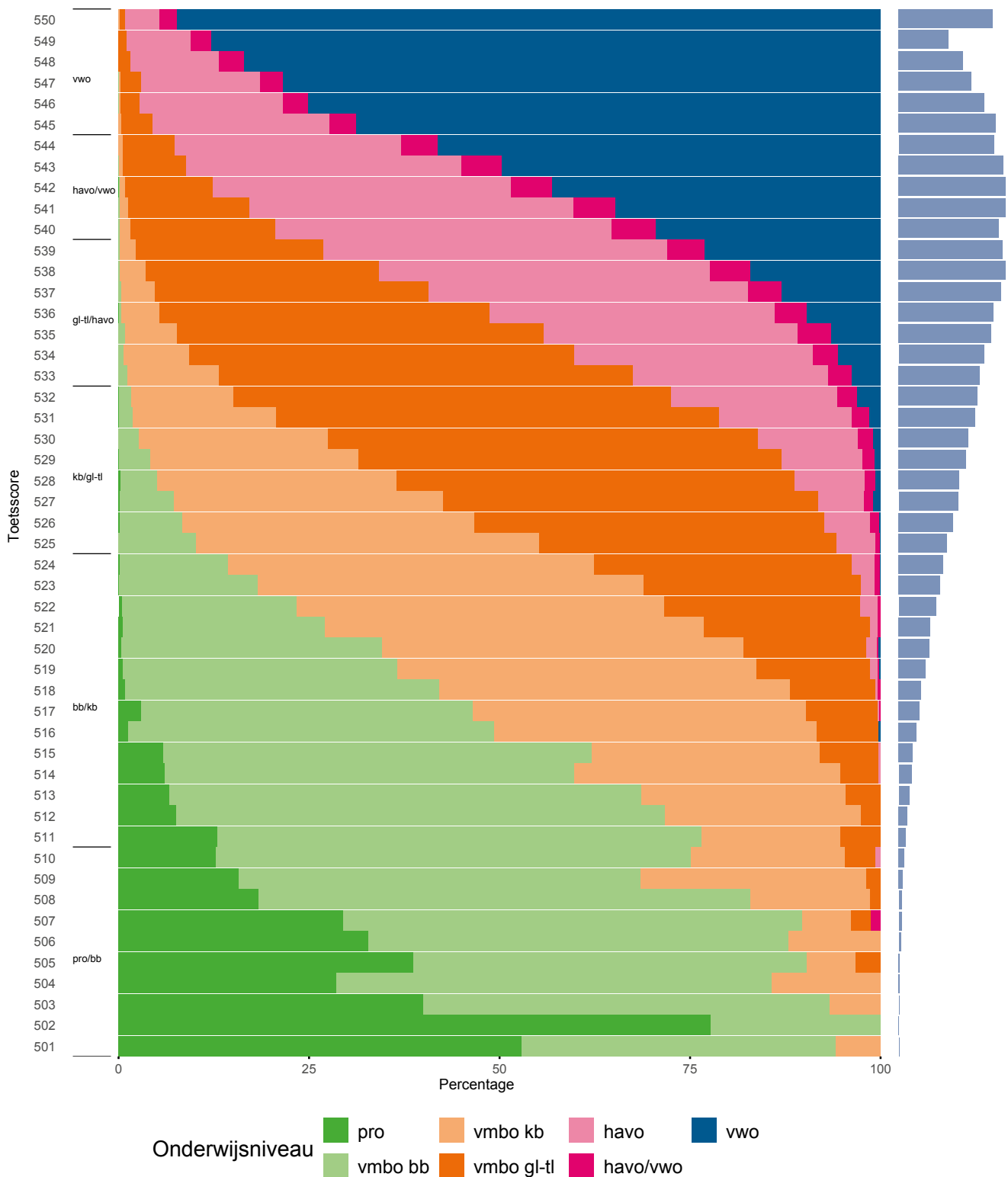


# Onderzoeksvraag 1:

## Hoe hangen de toetsscores van de Centrale Eindtoets van 2022 samen met het onderwijsniveau in vo leerjaar 3?

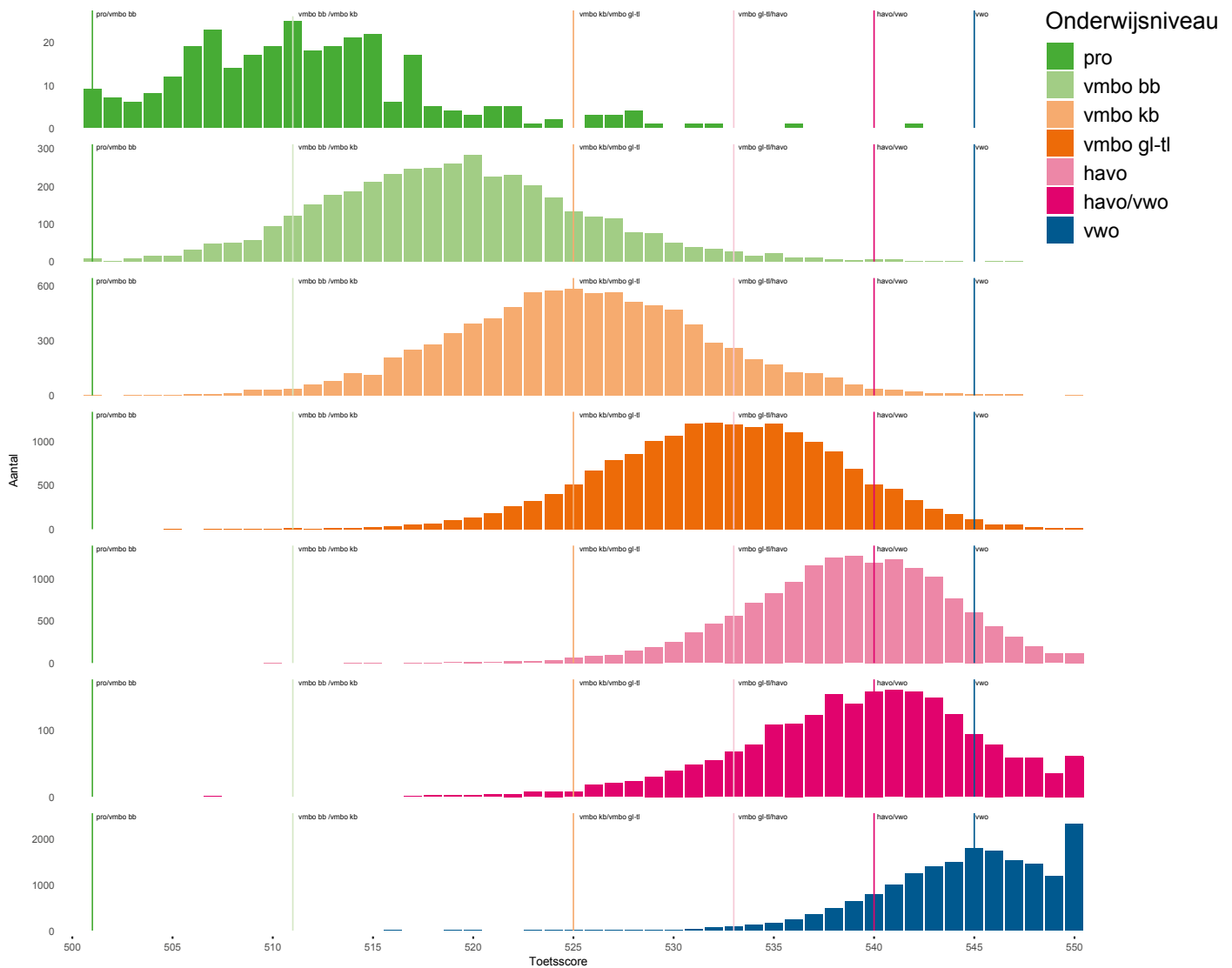
Voor de papieren en digitale Centrale Eindtoets samen kijken we naar de relatie tussen de toetsscores, de bijbehorende toetsadviesgrenzen en de gevolgde onderwijsniveaus. **Figuur 4** laat de relatie zien tussen de toetsscores en het onderwijsniveau van leerlingen in leerjaar 3, die de CET maakten. Bijvoorbeeld, van de leerlingen met een toetsscore 546 volgde ongeveer 75% onderwijs op onderwijsniveau vwo en 15% onderwijs op onderwijsniveau havo. De verhouding tussen de verschillende onderwijsniveaus verschuift tussen de verschillende toetsscores. Te zien is dat, naarmate de toetsscore stijgt, het aandeel leerlingen met bijvoorbeeld een onderwijsniveau vmbo bb daalt en het aandeel met een onderwijsniveau vwo stijgt. Iedere toetsscore is gekoppeld aan een toetsadvies, wat te zien is rechts van de toetsscores (zonder coronacorrectie). Voor iedere toetsadviescategorie is te zien dat het aandeel leerlingen dat een gelijk onderwijsniveau volgt groter is dan het aandeel leerlingen dat een ander onderwijsniveau volgt. Aan de rechterkant is in grijs de verdeling te zien van het aantal leerlingen per toetsscore. Te zien is dat het zwaartepunt van de populatie ligt rond de toetsadviescategorieën vmbo gl-tl/havo, havo/vwo en vwo.

**Figuur 4:** De samenstelling van de onderwijsniveaus in leerjaar 3 per toetsscore op de Centrale Eindtoets. Met daarin: de mogelijk te behalen toetsscores op de Centrale Eindtoets, de daaraan gekoppelde toetsadviescategorieën (zonder coronacorrectie), de samenstelling van de leerlingpopulatie voor iedere toetsscore uitgedrukt als percentage, en als laatste een histogram van het aandeel van de leerlingpopulatie met die toetsscore.



**Figuur 5** laat per onderwijsniveau zien hoeveel leerlingen elke mogelijke toetsscore gehaald hebben.

**Figuur 5:** Histogram van de leerlingaantallen per toetscore op de Centrale Eindtoets, uitgesplitst per onderwijsniveau in leerjaar 3. De verticale lijnen geven de toetsadviesgrenzen weer zonder coronacorrectie. Let op dat de schaal van de y-as anders is per onderwijsniveau om de data zichtbaar te houden.



**Figuur 4 en 5** laten samen zien dat de toetscores en bijbehorende toetsadviescategorieën samenhangen met het uiteindelijk gevolgde onderwijsniveau in leerjaar 3. Naast deze visualisaties, hebben we onderzocht of de voorstellende waarde van de Centrale Eindtoets hoger zou zijn als de toetsadviesgrenzen anders waren geweest. Bijvoorbeeld, wanneer een leerling niet 511 maar 510 moet scoren voor een toetsadvies vmbo bb/vmbo kb. Daarvoor groeperen we de leerlingen op het behaalde onderwijsniveau in leerjaar 3 en berekenen we de frequentie van iedere toetscore binnen de groep. Vervolgens bepalen we of het verschuiven van de grens leidt tot een hoger aantal correct geclassificeerde leerlingen in vo leerjaar 3. Deze analyse hebben we uitgevoerd zonder de coronacorrectie op de toetsadviesgrenzen (zie ook **tabel 1**). Deze analyse betreft zowel de populaties van de papieren als de digitale Centrale Eindtoets, omdat we daarmee de breedste dekking van de vaardigheidsschaal hebben.

**Tabel 3** is de kruistabel die volgt uit deze analyse. Daarin staat per onderwijsniveau het aantal leerlingen dat een bepaalde toetscore heeft gehaald. Bijvoorbeeld, er zijn 57 leerlingen met een toetscore 509 die in leerjaar 3 in het vmbo bb zitten. De score 509 valt binnen de grenzen (501-510) voor leerlingen die het toetsadvies pro/vmbo bb krijgen. Als de toetsadviesgrens zou worden verlegd, en bijvoorbeeld op 508 in plaats van 510 zou komen te liggen, dan zouden deze 57 leerlingen het advies vmbo bb/vmbo kb hebben gehad. In beide scenario's, met en zonder verlegging van de grens, zouden alle 57 leerlingen een toetsadvies krijgen dat overlapt met hun daadwerkelijk gevolgde onderwijs in leerjaar 3 (vmbo bb zit in beide toetsadviezen). Er zijn echter ook 17 leerlingen met toetscore 509 die in leerjaar 3 pro-onderwijs volgen. Als we de grens verleggen naar 508, zouden ook deze leerlingen een toetsadvies vmbo bb/vmbo kb hebben gekregen. Voor deze leerlingen zou het toetsadvies dan *niet* meer in overeenstemming zijn met het gevolgde onderwijsniveau in leerjaar 3.

Om te bepalen of de toetsadviesgrenzen moeten worden verlegd, kijken we naar het effect van het verleggen van de grenzen op het aantal overeenstemmend geclassificeerde leerlingen. In **tabel 3** zijn de leerlingen blauw gekleurd die met de huidige toetsadviesgrenzen een onderwijsniveau volgen dat overeenstemt met hun toetsadvies. In groen staan de leerlingen die onder de huidige grenzen niet overeenstemmend worden geclassificeerd, maar bij een aanpassing van de grenzen wel overeenstemmend worden geclassificeerd (winst). In rood staan de leerlingen die bij een verlegging van de grenzen niet meer overeenstemmend zouden worden geclassificeerd (verlies). Als de classificatiewinst groter is dan het verlies, dan beschouwen we die grens als een betere voorspeller voor de hele groep leerlingen dan de huidige toetsadviesgrens. Hierbij moet worden opgemerkt dat deze methode uitgaat van een hypothetisch scenario, waarin een andere grens wordt gehanteerd dan in de praktijk is gebruikt. Bij andere toetsadviesgrenzen zou het startniveau en mogelijk ook doorstroom naar leerjaar 3 in het vo anders uitgedrukt hebben voor deze leerlingen.

Te zien is dat de toetsadviesgrens van pro/vmbo bb naar vmbo bb/vmbo kb wat te hoog ligt, maar de classificatiewinst zou een zeer klein aantal leerlingen betreffen (<30). Met dezelfde vuistregel van >30 leerlingen winst, zou de grens tussen vmbo bb/vmbo kb en vmbo kb/vmbo gl-tl te hoog liggen met ongeveer 2 punten. De coronacorrectie heeft voorzien in de tijdelijke verlaging van die grens met 1 punt. Voor de grens tussen vmbo kb/vmbo gl-tl en vmbo gl-tl/havo ligt de grens met 1 punt te hoog. Ook daar heeft de coronacorrectie voorzien met een tijdelijke verlaging van de grens met 1 punt. Bij de grens tussen vmbo gl-tl/havo en havo/vwo leidt een verschuiving van de grens niet tot een betere classificatie.

## Conclusies

De toetsadviesgrenzen van de Centrale Eindtoets blijken sterk te overlappen met de plaatsing van leerlingen in vo leerjaar 3. Er zijn daarom geen sterke aanwijzingen dat de toetsadviesgrenzen moeten worden bijgesteld. Enkel bij de toetsadviesgrenzen tussen vmbo bb/vmbo kb en vmbo kb/vmbo gl-tl en tussen vmbo kb/vmbo gl-tl en vmbo gl-tl/havo zou een verlaging van de grens van 2 respectievelijk 1 punt tot een betere classificatie van leerlingen leiden. In 2022 is tijdelijk de coronacorrectie toegepast, waardoor voor beide toetsadviescategorieën een grens van 1 punt lager is gehanteerd. Deze bijstelling blijkt dus gerechtvaardigd te zijn voor de destijds geldende situatie waarin leerlingen nadelige gevolgen hebben ervaren van de coronapandemie op hun onderwijs. Enkel voor de grens tussen vmbo bb/vmbo kb en vmbo kb/vmbo gl-tl blijft de mogelijkheid dat de grensscore met nog een extra punt naar beneden moet worden bijgesteld los van de coronacorrectie.

We adviseren echter om dit jaar de grens tussen vmbo bb/vmbo kb en vmbo kb/vmbo gl-tl nog niet bij te stellen. De situatie in 2022 was bijzonder door de nasleep van de coronapandemie. Daarnaast speelt ook mee dat de grenzen inherent gevoelig zijn voor meetonzekerheid waardoor een geobserveerde classificatiewinst van geringe grootte volgend afnamejaar mogelijk anders kan zijn (zie ook appendix D). Daarom stellen we voor om niet direct de toetsadviesgrenzen in het doorstroomtoetsstelsel aan te passen na de observatie van dit verschil in dit afnamejaar, maar de grenzen te monitoren in meerdere afnamejaren. Een bijstelling van de grenzen op basis van een kleine classificatiewinst is beter te rechtvaardigen wanneer uit meerdere evaluaties blijkt dat een bijstelling structureel leidt tot een betere voorspelling van het onderwijsniveau in leerjaar 3. We blijven dit grenspunt monitoren in toekomstige evaluaties, die we verwachten te publiceren in het najaar van 2026 (over afnamejaar 2023) en in het voorjaar van 2027 (over afnamejaar 2024, de eerste afname van de doorstroomtoets).

Een bijzondere toetsadviescategorie is vwo. Door de overlap tussen havo/vwo (toetsscore 540) en vwo (toetsscore 545), kunnen we de grens van het toetsadvies vwo niet beoordelen volgens de gebruikte methode. Immers, we bereiken de beste classificatie als we het enkelvoudige toetsadvies vwo niet meer zouden geven en iedereen boven de 540 een havo/vwo advies geven. Het is uiteraard wel wenselijk om de grens van het toetsadvies vwo te kunnen beoordelen. We stellen daarom voor om te kijken naar de verdeling van vwo leerlingen over de toetsscores (onderste rij van **figuur 5**). De grens voor het toetsadvies vwo kan dan gelegd worden bij de piek van de populatieverdeling, oftewel bij de toetsscore die de meeste vwo-leerlingen hebben behaald<sup>3</sup>. In deze evaluatie is dat 545, waardoor een bijstelling niet nodig is. Met behulp van deze methode zullen we ook in komende jaren deze grens monitoren.

---

3 Met uitzondering van 550 omdat daar veel leerlingen komen puur door het bereiken van het einde van de schaal. 550 bevat daardoor ook de leerlingen die eigenlijk 551, 552, 553 enzovoorts behalen omdat die scores niet toegekend worden.

**Tabel 3:** Aantal leerlingen per onderwijsniveau (kolommen) en per toetscore (rijen) op de Centrale Eindtoets (papier en digitaal samen). Blauw = overeenstemming tussen toetsadvies en onderwijsniveau; groen = classificatiewinst bij mogelijke grensverhuiving; rood = classificatieverlies bij mogelijke grensverhuiving. Zodra er meer winst dan verlies mogelijk was door een grens te verleggen zijn die cellen ingekleurd in de tabel. De cellen zijn lichter gekleurd wanneer 1) de coronacorrectie al voorziet in de juiste classificatie en 2) de classificatiewinst 30 of minder leerlingen betreft. De scoregrens van de coronacorrectie is weergegeven met de stippellijnen.

toetscore	pro	vmbo bb	vmbo kb	vmbo gl-tl	havo	havo/vwo	vwo	grens 2022	verlegging
501	9	7	1	0	0	0	0		
502	7	2	0	0	0	0	0		
503	6	8	1	0	0	0	0		
504	8	16	4	0	0	0	0		
505	12	16	2	1	0	0	0	pro/ vmbo bb	pro/ vmbo bb
506	19	32	7	0	0	0	0		
507	23	47	5	2	0	1	0		
508	14	49	12	1	0	0	0		
509	17	57	32	2	0	0	0		
510	19	93	30	6	1	0	0		
511	25	122	35	10	0	0	0		
512	18	152	61	6	0	0	0		
513	19	176	76	13	0	0	0		
514	21	186	121	17	1	0	0		
515	22	212	113	29	1	0	0		vmbo bb/ vmbo kb
516	6	231	204	39	0	0	1		
517	17	245	247	53	1	1	0	vmbo bb/ vmbo kb	
518	5	249	278	68	2	2	0		
519	4	260	340	107	7	2	1		
520	3	283	392	134	11	2	2		
521	5	224	422	184	9	3	0		
522	5	229	482	257	23	4	0		
523	1	202	562	317	20	7	1		
524	2	169	572	402	36	8	1		
525	0	132	582	505	66	7	1	vmbo kb/ vmbo gl-tl	
526	3	119	558	670	87	18	2		vmbo kb/ vmbo gl-tl
527	3	114	562	786	95	20	15		
528	4	78	510	849	150	23	10		
529	1	74	493	1 005	191	30	13		
530	0	50	467	1 057	246	39	17		

toetscore	pro	vmbo bb	vmbo kb	vmbo gl-tl	havo	havo/ vwo	vwo	grens 2022	verlegging
531	1	38	387	1 199	357	48	30		
532	1	34	285	1 211	462	54	65		
533	0	27	260	1 187	559	67	83		
534	0	16	197	1 158	716	77	127		
535	0	21	170	1 194	827	107	161	vmbo gl-tl/ havo	vmbo gl-tl/ havo
536	1	10	126	1 105	955	109	246		
537	0	10	121	990	1 154	121	357		
538	0	5	98	880	1 248	152	491		
539	0	4	61	687	1 263	139	645		
540	0	5	37	512	1 188	157	793		
541	0	5	31	457	1 225	159	999		
542	1	2	23	331	1 126	157	1 239	havo/vwo	havo/vwo
543	0	1	14	234	1 021	148	1 402		
544	0	1	13	175	761	123	1 485		
545	0	0	8	109	602	93	1 790		
546	0	1	5	57	432	78	1 727		
547	0	1	5	52	307	58	1 537		
548	0	0	0	27	201	58	1 446	vwo	vwo
549	0	0	0	15	113	35	1 178		
550	0	0	4	18	113	60	2 329		

Bovenstaande is een vervolg van tabel 3.

# Onderzoeksvraag 2a:

## Hoe sterk voorspellen de toetsadviezen en de voorlopige schooladviezen uit 2022 de plaatsing in het vo in leerjaar 1, 2 en 3?

Om de overeenstemming tussen het toetsadvies en het onderwijsniveau van de totale populatie te bepalen, berekenen we Cohen's kappa ( $\kappa$ ) (Cohen, 1960). Daarvoor bepalen we eerst welke toetsadviezen en onderwijsniveaus kunnen worden aangemerkt als overeenkomend. In dit onderzoek beschouwen we adviescategorieën overeenkomend zodra één van de niveaus overlapt. Zo komt bijvoorbeeld het toetsadvies vmbo bb/vmbo kb overeen met plaatsing in het vo in een vmbo bb-, vmbo kb-, vmbo bb/vmbo kb-klas of een brede brugklas die één of beide niveaus aanbiedt. In appendix B staat een volledig overzicht van overeenkomende combinaties. Na het bepalen van iedere overeenkomende combinatie tussen toetsadvies en onderwijsniveau, berekenen we Cohen's  $\kappa$ . Cohen's  $\kappa$  is een voor kans gecorrigeerde uitkomstmaat die de overeenkomst tussen indelingen weergeeft, in dit geval de overeenstemming tussen toetsadvies en onderwijsniveau. De  $\kappa$  kan variëren van -1 tot +1, waarbij een negatieve  $\kappa$  betekent dat de indelingen slechter overeenkomen dan je op basis van toeval zou verwachten. Een  $\kappa$  van 0 betekent dat de overeenkomst gelijk is aan toeval en een  $\kappa$  van +1 betekent dat de indelingen een perfecte overeenkomst kennen. Als vuistregel wordt aangenomen dat een Cohen's  $\kappa$  van 0,60-0,79 een redelijke/gemiddelde overeenstemming weergeeft, van 0,80-0,90 een sterke overeenstemming en boven 0,90 een bijna perfecte overeenstemming is (McHugh, 2012).

In **tabel 4** staat de overeenstemming tussen toetsadviezen en de plaatsing in leerjaar 1, 2 en 3. Deze overeenstemming is uitgedrukt in Cohen's  $\kappa$  en als absolute overeenstemming. De absolute overeenstemming ligt bij alle toetsaanbieders boven de 80% en de voor kans gecorrigeerde overeenstemming ligt overal boven de 68%. Beide percentages zijn het hoogst kort na afname van de toets en dalen naarmate leerlingen langer op het vo zitten. We zien geen noemenswaardige verschillen tussen de verschillende eindtoetsen in de overeenstemming tussen het gegeven toetsadvies en het onderwijsniveau. Ook zien we vrijwel geen verschillen tussen toetsen in het percentage leerlingen dat een hoger of lager onderwijsniveau volgt dan het gegeven toetsadvies. De enige toets die opvalt is de digitale Centrale Eindtoets, met het laagste percentage leerlingen met een lager onderwijsniveau dan het toetsadvies en het hoogste percentage leerlingen met een onderwijsniveau hoger dan het toetsadvies. Met andere woorden, vergeleken met de andere eindtoetsen kregen de leerlingen van de digitale Centrale Eindtoets over het algemeen lagere toetsadviezen dan hun onderwijsniveau in vo leerjaar 3 zou rechtvaardigen.

**Tabel 4:** Cohen's  $\kappa$  en percentages correcte, lagere en hogere classificatie tussen het toetsadvies en onderwijsniveau in vo leerjaar 1, 2 en 3. De absolute overeenstemming staat in de kolom "% vo gelijk aan toetsadvies".

vo jaar	toets	n	$\kappa$	% vo gelijk aan toetsadvies	% vo lager dan toetsadvies	% vo hoger dan toetsadvies
1	Totaal	159 879	0,813	91,5	3,0	5,5
2	Totaal	156 806	0,765	88,1	5,1	6,8
3	Totaal	149 056	0,694	82,2	9,8	8,0
1	AMN	3 153	0,815	92,1	2,7	5,2
2	AMN	3 092	0,767	88,0	5,2	6,8
3	AMN	2 922	0,707	82,8	10,1	7,1
1	CET (papier)	51 007	0,814	91,7	2,8	5,5
2	CET (papier)	50 090	0,759	87,9	5,1	6,9
3	CET (papier)	47 707	0,689	82,0	9,8	8,2
1	CET (digitaal)	21 081	0,823	91,6	1,7	6,8
2	CET (digitaal)	20 631	0,779	88,4	2,8	8,8
3	CET (digitaal)	19 682	0,717	83,1	5,9	10,9
1	Dia	6 687	0,830	92,4	3,0	4,6
2	Dia	6 586	0,781	88,9	5,3	5,8
3	Dia	6 260	0,702	82,5	10,3	7,2
1	IEP (papier)	56 794	0,809	91,4	3,4	5,2
2	IEP (papier)	55 632	0,764	88,2	5,7	6,1
3	IEP (papier)	52 756	0,691	82,1	10,7	7,2
1	ROUTE 8	21 157	0,803	91,0	3,6	5,4
2	ROUTE 8	20 775	0,757	87,5	5,8	6,7
3	ROUTE 8	19 729	0,684	81,4	10,9	7,6

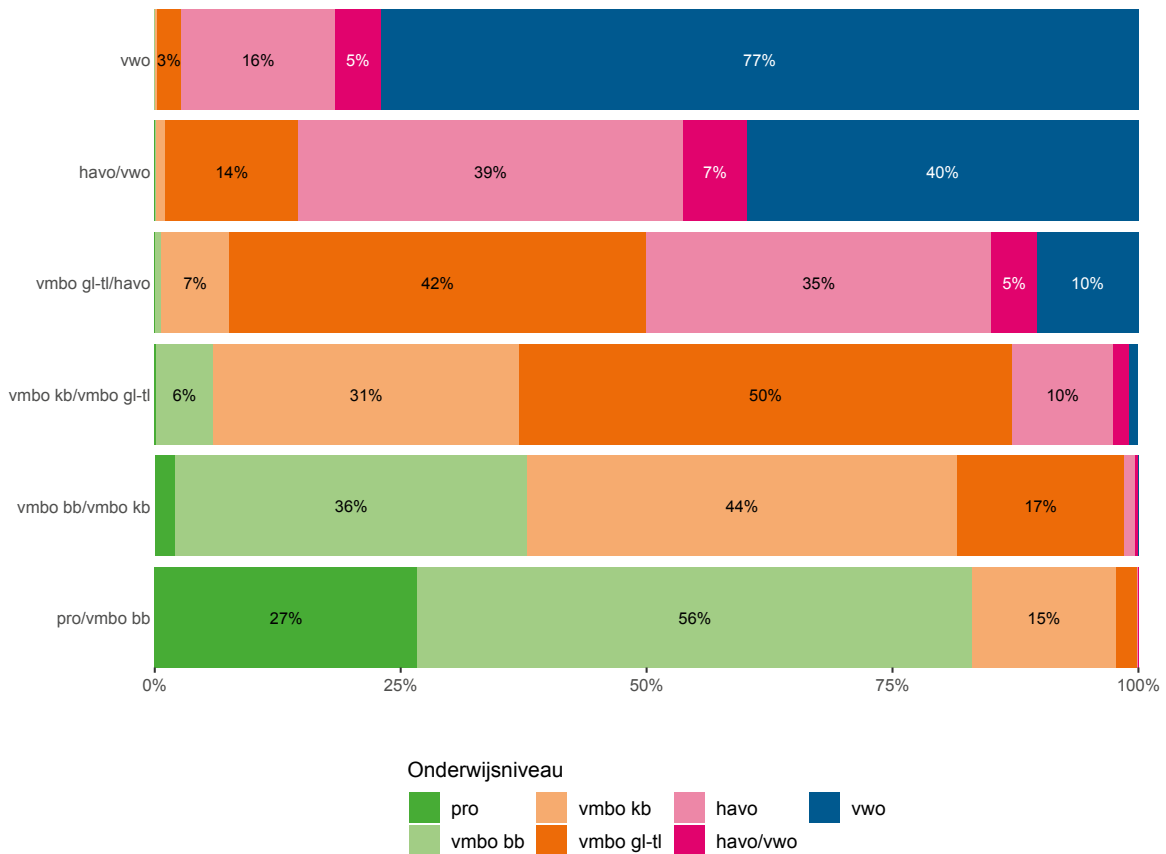
Dezelfde analyse hebben we ook uitgevoerd met het voorlopig schooladvies in plaats van het toetsadvies. In **tabel 5** staan de resultaten van de overeenstemming tussen het voorlopige schooladvies en het onderwijsniveau in leerjaar 1, 2 en 3. Hier zien we percentages absolute overeenstemming boven de 77% en gecorrigeerde overeenstemming boven 68%. We zien geen betekenisvolle verschillen tussen toetsen in de mate van overeenstemming tussen voorlopig schooladvies en onderwijsniveau in het vo, ook niet wat betreft het percentage leerlingen dat hoger of lager is geplaatst dan het voorlopig schooladvies.

**Tabel 5:** Cohen's  $\kappa$  en percentages correcte, lagere en hogere classificatie tussen het voorlopig schooladvies en onderwijsniveau in vo leerjaar 1, 2 en 3.

vo jaar	toets	n	$\kappa$	% vo gelijk aan voorlopig schooladvies	% vo lager dan voorlopig schooladvies	% vo hoger dan voorlopig schooladvies
1	Totaal	158 594	0,913	95,3	1,1	3,6
2	Totaal	155 574	0,807	88,4	4,3	7,3
3	Totaal	147 889	0,695	78,9	10,7	10,4
1	AMN	3 059	0,914	95,6	1,1	3,3
2	AMN	3 000	0,807	88,2	4,3	7,5
3	AMN	2 832	0,704	79,6	10,0	10,3
1	CET (papier)	50 550	0,918	95,7	1,0	3,3
2	CET (papier)	49 644	0,806	88,5	4,3	7,3
3	CET (papier)	47 288	0,695	79,3	10,4	10,3
1	CET (digitaal)	20 879	0,919	95,4	1,1	3,5
2	CET (digitaal)	20 434	0,818	88,7	3,9	7,4
3	CET (digitaal)	19 493	0,701	78,9	10,0	11,1
1	Dia	6 637	0,909	95,1	1,3	3,6
2	Dia	6 541	0,794	87,5	4,8	7,7
3	Dia	6 216	0,683	77,8	11,1	11,1
1	IEP (papier)	56 383	0,911	95,1	1,2	3,7
2	IEP (papier)	55 247	0,807	88,5	4,3	7,2
3	IEP (papier)	52 392	0,693	78,9	10,9	10,2
1	ROUTE 8	21 086	0,906	94,9	1,2	3,9
2	ROUTE 8	20 708	0,803	88,0	4,3	7,8
3	ROUTE 8	19 668	0,690	78,4	11,1	10,5

Daarnaast hebben we de overeenstemming tussen iedere toetsadviescategorie en het onderwijsniveau in vo leerjaar 3 gevisualiseerd in **figuur 6**. Per gegeven toetsadvies is te zien welk aandeel van de leerlingen een bepaald onderwijsniveau volgt. We zien dat voor iedere toetsadviescategorie minstens 80% van de leerlingen op een onderwijsniveau is ingeschreven passend bij het gegeven toetsadvies.

**Figuur 6:** Voor ieder toetsadvies is de samenstelling van de leerlingpopulatie in leerjaar 3 weergegeven voor alle eindtoetsen samen. De samenstelling wordt uitgedrukt in het percentage leerlingen dat een bepaald onderwijsniveau volgt.



## Conclusies

Bij 82,2% van de leerlingen komt het toetsadvies overeen met het gevolgde onderwijs in leerjaar 3. Dit hoge percentage, samen met de redelijke overeenstemming op basis van Cohen's  $\kappa$ , ondersteunt de voorspellende validiteit van het toetsadvies. Eerder onderzoek vond ook al hoge percentages overeenstemming tussen het dubbele toetsadvies en de plaatsing in vo leerjaar 3 (Douma et al., 2025; Lek & Van de Schoot, 2019). De overeenstemming die in het huidige onderzoek wordt gevonden is nog hoger, wat een aanwijzing kan zijn dat de doorontwikkeling van de eindtoetsen heeft geleid tot een betrouwbaardere meting en/of dat het toetsadvies in 2022 meer bepalend is geworden bij de plaatsing van leerlingen in het vo. Een deel van de leerlingen (9,8%) volgt na 3 jaar een lager onderwijsniveau dan het gegeven toetsadvies en een kleiner deel (8,0%) volgt een hoger onderwijsniveau.

Voor het voorlopig schooladvies zien we dat dit bij 78,9% van de leerlingen overeenkomt met het onderwijsniveau in leerjaar 3, wat in lijn ligt met eerder onderzoek naar oudere eindtoetsgegevens (Lek & Van de Schoot, 2019). Als we de overeenkomst corrigeren voor kans, dan zijn zowel het toetsadvies als het voorlopig schooladvies even goede voorspellers van het gevolgde onderwijsniveau 3 jaar later ( $\kappa = 0,69$ ). Voor het voorlopig schooladvies valt op dat de verhouding tussen leerlingen die uiteindelijk lager (10,7%) eindigen en die hoger eindigen (10,4%) anders is dan voor het toetsadvies.

We zien geen noemenswaardige verschillen in bovenstaande analyses tussen de toetsen in de overeenstemming van het toetsadvies en het gevolgde vo-niveau. Ook zien we geen belangrijke verschillen tussen de toetsen wat betreft op- en afstroom. De enige toets die opvalt, is de digitale Centrale Eindtoets. Zoals **tabel 4** laat zien heeft deze toets een relatief laag aandeel leerlingen dat een lager onderwijsniveau volgt dan het toetsadvies én een relatief hoog aandeel leerlingen dat een hoger onderwijsniveau volgt dan het toetsadvies.

De oorzaak hiervan ligt waarschijnlijk in de toegepaste normeringsmethode. Bij de digitale Centrale Eindtoets werd de eigen absolute normering gebruikt, terwijl de commerciële eindtoetsen gebruikmaakten van de IRT2-normering. Omdat de populatie die de papieren Centrale Eindtoets maakt vaardiger is dan de populaties van de commerciële eindtoetsen, kan dit bij toepassing van de IRT2-normering leiden tot overschattingen van de toetsadviezen bij de commerciële eindtoetsen (zie appendix D). Deze overschatting onderstreept één van de argumenten om de IRT2-normering in het eindtoetsstelsel alleen als tijdelijke maatregel in te zetten. Daarnaast zorgt de absolute normering bij de digitale Centrale Eindtoets ervoor dat een prestatieverschil tussen papier en digitaal (ook wel moduseffect genoemd) zichtbaar zou kunnen zijn. In de IRT2-normering van de commerciële eindtoetsen zijn dergelijke verschillen niet waarneembaar, omdat via de populatie wordt geëquivalereerd in plaats van via de prestaties op gezamenlijke ankeropgaven.

Omdat de equivalering van de IRT2-normering gebruikmaakt van voorlopige schooladviezen, is het ontbreken van betekenisvolle verschillen tussen de commerciële eindtoetsen in de overeenkomst van toetsadvies met plaatsing in jaar 3 (**tabel 4**) een belangrijke indicatie dat de scholen die voor een bepaalde toets kiezen vergelijkbaar zijn in de manier waarop voorlopige schooladviezen worden gegeven aan leerlingen met een gelijke vaardigheid. Wanneer dit namelijk niet het geval zou zijn, bijvoorbeeld als er een eindtoets is waarbij de deelnemende scholen beduidend kansrijker of minder kansrijk adviseren, zou dit automatisch leiden tot hogere of lagere toetsadviezen via de IRT2-normering. Dit zou dan waarschijnlijk tot uiting komen via een afwijkende op- of afstroom. Dit is ook een sterke aanwijzing dat een selectie-effect op schoolniveau in de manier waarop voorlopige schooladviezen gegeven worden slechts van beperkte invloed is op de waargenomen prestatieverschillen tussen toetsen in het huidige doorstroomtoetsstelsel. Dit beeld wordt ook bevestigd door de relatie tussen het voorlopig schooladvies en de plaatsing in jaar 3 (**tabel 5**), waarin ook geen noemenswaardige verschillen tussen de eindtoetsen naar voren komen.

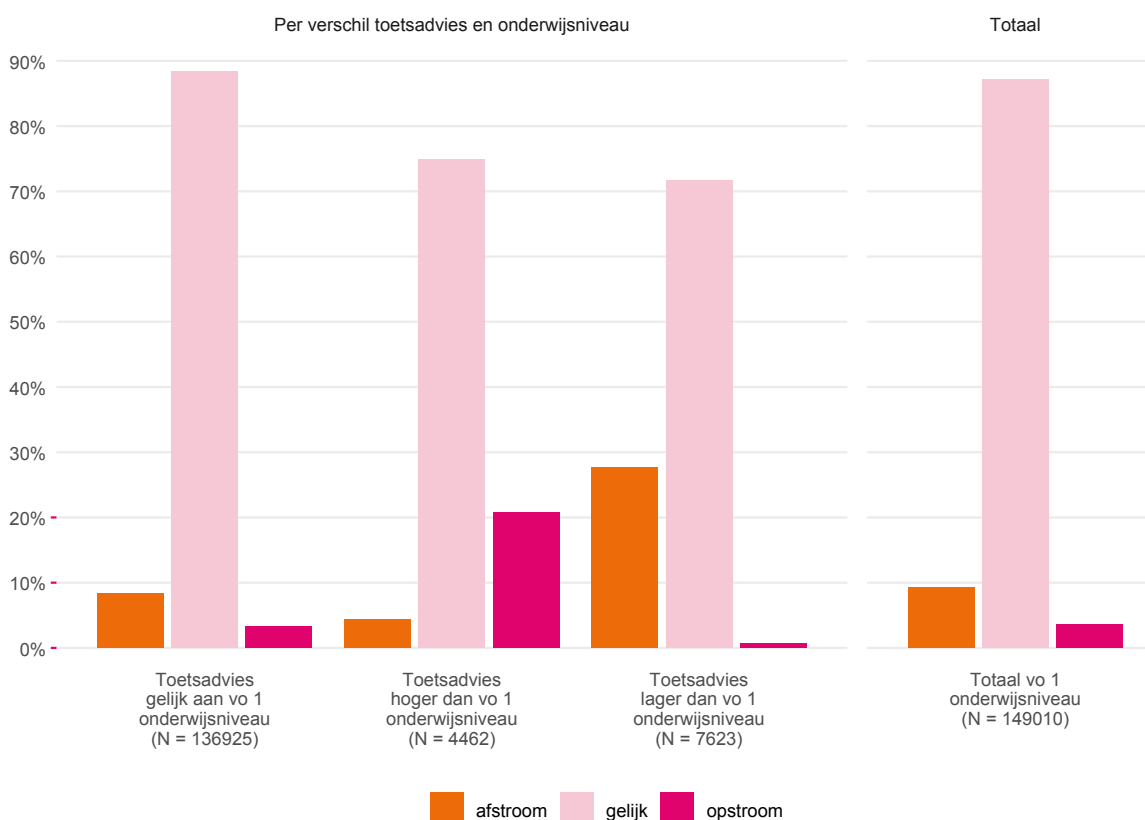
Als we vooruitkijken naar toekomstig doorstroomonderzoek, dan verwachten we dat de resultaten zoals we die hier beschrijven anders zullen uitpakken vanaf afnamejaar 2024. In het doorstroomtoetsstelsel is immers een andere normeringsmethode toegepast waardoor de relatie tussen het toetsadvies en de plaatsing in vo leerjaar 3 anders zal uitpakken. Er zullen daardoor ook grotere verschillen ontstaan tussen de verschillende doorstroomtoetsen. Ook is door de verplichte bijstellingsmaatregel de dynamiek tussen voorlopig schooladvies, toetsadvies en de plaatsing in vo anders dan in het eindtoetsstelsel. Dit zal impact hebben op zowel de relatie tussen toetsadvies als voorlopig schooladvies en de plaatsing in leerjaar 3 van het vo.

# Onderzoeksvraag 2b:

## Wat is de relatie tussen het onderwijsniveau in leerjaar 1, het toetsadvies en de op- en afstroom tussen leerjaar 1 en 3 op het vo?

Voor deze onderzoeksvraag analyseren we het verschil tussen het gevolgde onderwijsniveau van leerlingen in vo leerjaar 1 en 3. Op- en afstroom definiëren we als het volgen van een hoger of lager onderwijsniveau in leerjaar 3 vergeleken met leerjaar 1. **Figuur 7** laat de op- en afstroom zien voor de totale leerlingpopulatie (uiterst rechts). In deze kolom (Totaal) is te zien dat de meerderheid van de leerlingen niet op- of afstroomt. Een klein deel stroomt af en een nog kleiner deel stroomt op. De totale groep leerlingen is opgesplitst in drie subgroepen: leerlingen die een toetsadvies hadden dat gelijk, hoger of lager was dan hun onderwijsniveau in vo leerjaar 1. Voor de leerlingen met een toetsadvies gelijk aan hun onderwijsniveau in jaar 1 (meest linker groep) is te zien dat de op- en afstroom ongeveer dezelfde verhouding heeft als dat van de totale groep. Dit is logisch omdat dit ook de grootste groep leerlingen betreft. Als we kijken naar de leerlingen met een toetsadvies hoger dan hun plaatsing in jaar 1, dan zien we dat er relatief meer opstroom is dan afstroom (tweede groep van links). Voor de leerlingen met een toetsadvies lager dan hun plaatsing in jaar 1, zien we relatief meer afstroom dan opstroom (derde groep van links).

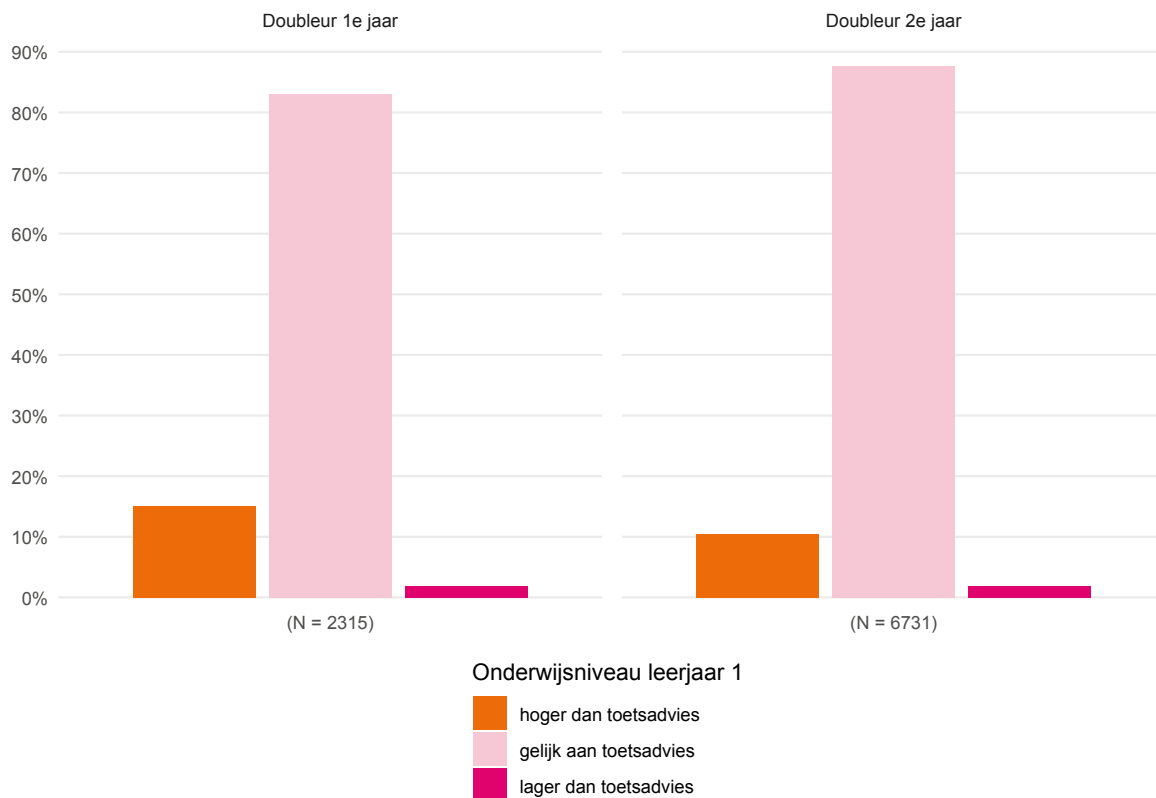
**Figuur 7:** Op- en afstroom van vo leerjaar 1 naar leerjaar 3, totaal en uitgesplitst naar drie groepen: was het toetsadvies gelijk, hoger of lager dan het gevolgde onderwijsniveau in leerjaar 1. Deze grafiek betreft alle eindtoetsen samen.



In **figuur 8** zien we hoeveel leerlingen er doubleren in leerjaar 1 (links) en in leerjaar 2 (rechts). Te zien is dat de meeste leerlingen die doubleren een onderwijsniveau volgden dat gelijk was aan hun toetsadvies (donkerroze staven). Echter, ook valt op dat naar verhouding meer leerlingen doubleren die zijn gestart op een onderwijsniveau dat hoger lag dan het gegeven toetsadvies (oranje staven) in vergelijking tot leerlingen die op een lager onderwijsniveau zijn gestart (paarse staven). Net als uit de analyse van op- en afstroom in **figuur 7** blijkt hieruit de voorspellende waarde van het

toetsadvies met betrekking tot schoolprestaties.

**Figuur 8:** Aantal leerlingen dat doubleert in leerjaar 1 en leerjaar 2, uitgesplitst naar drie groepen: was het onderwijsniveau in leerjaar 1 hoger, gelijk aan of lager dan het gegeven toetsadvies?



## Conclusies

Bij onderzoeksvraag 2a zagen we al dat bij meer dan 75% van de leerlingen het onderwijsniveau in jaar 3 van het vo overlapt met het gegeven toets- en schooladvies. Als een leerling toch op- of afstroomt, dan zien we dat opstroom relatief vaker voorkomt bij leerlingen die begonnen zijn op een onderwijsniveau lager dan hun toetsadvies. Dit ligt in lijn met eerder onderzoek naar oudere eindtoetsresultaten (Lek & Van de Schoot, 2019; Strouken et al., 2008). Afstroom zien we relatief vaker bij leerlingen die begonnen zijn op een onderwijsniveau hoger dan het ontvangen toetsadvies. Hetzelfde patroon zien we voor leerlingen die doubleren, waarbij dat aandeel groter is voor leerlingen met een lager toetsadvies dan het onderwijsniveau in leerjaar 1. Deze bevindingen samen ondersteunen de voorspellende validiteit van de eindtoetsen.

# Discussie

Leerlingen ontwikkelen zich sterk in de eerste drie jaar van het vo. Zowel de voorlopige schooladviezen als de toetsadviezen proberen een goede voorspelling te geven van de meest waarschijnlijke ontwikkeling van een leerling om zo tot een passend onderwijsniveau te komen. Uit de analyses blijkt dat zowel het voorlopige schooladvies als het toetsadvies goed in staat is om deze voorspelling te maken: een groot deel van de leerlingen eindigt na drie jaar vo op het onderwijsniveau dat ze geadviseerd is.

## Beperkingen van de huidige evaluatie

Deze evaluatie was niet zonder beperkingen. Ten eerste was het niet voor iedere leerling mogelijk om de doorstroomgegevens uit het vo van DUO aan de eindtoetsresultaten van de toetsaanbieders te koppelen. Alhoewel we geen patroon zien in het aantal geslaagde gegevenskoppelingen tussen de eindtoetsen zien we wel dat het aantal gegevens dat niet gekoppeld kon worden het grootst is voor de leerlingen met een pro/vmbo bb toetsadvies en steeds kleiner wordt naarmate het toetsadvies naar vwo gaat (**tabel B3**). Deze evaluatie heeft daarmee een kleine selectiebias.

Ten tweede is er een kleine informatiebias in dit onderzoek. Een informatiebias houdt in dat we mogelijk een klein deel van de informatie missen van specifieke groepen. Dit komt onder andere doordat we keuzes hebben gemaakt in de gegevensverwerking, bijvoorbeeld door leerlingen te verwijderen die niet gekoppeld konden worden tussen beide datasets of waarvan gegevens incompleet waren. We hebben geprobeerd de aannames in de dataverwerking zo helder mogelijk te omschrijven en laten in de appendix zien welke data zijn behouden.

Ten derde hebben we gekeken naar een zeer complex schoolsysteem. Een leerling doorloopt een eigen, uniek pad vanaf het begin van het po tot en met het vo. Met het voorlopig schooladvies, het toetsadvies en de plaatsing in het vo, hebben we geprobeerd deze paden op groepsniveau te omschrijven en te evalueren. De gebruikte gegevens zijn echter niet geschikt om alternatieve verklaringen voor de gevonden effecten uit te sluiten. Denk hierbij aan het self-fulfilling prophecy-effect<sup>4</sup>, de mogelijkheid om te wisselen van onderwijsniveau binnen een vo-school en het schoolaanbod in de regio.

## Implicaties voor het doorstroomtoetsstelsel

Alhoewel deze evaluatie betrekking heeft op het eindtoetsstelsel in 2022, zijn de bevindingen ook relevant voor het doorstroomtoetsstelsel.

- We hebben laten zien dat de toetsadviesgrenzen van de Centrale Eindtoets, en daarmee die van het doorstroomtoetsstelsel, op een accurate plek liggen. De grenzen moeten de komende evaluaties gemonitord worden, maar dit onderzoek geeft geen aanleiding om op dit moment de grenzen bij te stellen.
- Uit de plaatsing in jaar 3 van het vo blijkt dat de leerlingpopulatie van de papieren Centrale Eindtoets in het eindtoetsstelsel vaardiger was dan de leerlingpopulaties van andere eindtoetsen. Omdat veel scholen van de papieren Centrale Eindtoets zijn overgestapt op de papieren variant van de Leerling in Beeld doorstroomtoets<sup>5</sup>, kan dat deels verklaren waarom bij de Leerling in Beeld doorstroomtoets in 2024 en 2025 meer leerlingen de referentieniveaus behaalden dan bij andere doorstroomtoetsen.
- Er lijken geen systematische verschillen tussen eindtoetsen te zijn in de toekenning van voorlopige schooladviezen aan leerlingen met vergelijkbare vaardigheid. Als we aannemen dat de manier waarop scholen een toets kiezen in het huidige stelsel niet fundamenteel verschilt van het eindtoetsstelsel, kunnen de waargenomen prestatieverschillen van leerlingen met gelijk schooladvies tussen de doorstroomtoetsen in het huidige stelsel dus waarschijnlijk niet verklaard worden door verschillen in de schooladvisering.

---

4 Het verschijnsel waarbij bepaalde verwachtingen het gedrag, de motivatie en/of de prestaties van een leerling beïnvloeden.

5 70,7% van de scholen die in 2023 de papieren Centrale Eindtoets maakten, maakte in 2025 de papieren Leerling in Beeld doorstroomtoets.

# Verwijzingen

Brouwer, A. J., & Veldkamp, B. P. (2022). *Onderzoek naar de kwaliteit van de normering eindtoetsen primair onderwijs 2022*. Opgehaald van <https://www.rcec.nl/nieuws/onderzoek-naar-de-kwaliteit-van-de-normering-eindtoetsen-primair-onderwijs-2022/>

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*(20), pp. 37-46. doi:<https://psycnet.apa.org/doi/10.1177/001316446002000104>

CvTE. (2022). *Terugblik Centrale Eindtoets 2022*. Opgehaald van <https://www.cvte.nl/documenten/2024/09/23/terugblik-centrale-eindtoets-2022>

CvTE. (2024). *Terugblik normering doorstroomtoetsen 2024*. Opgehaald van <https://www.cvte.nl/documenten/rapporten/2024/07/16/terugblik-normering-doorstroomtoetsen-2024>

CvTE. (2025). *Terugblik normering doorstroomtoetsen 2025*. Opgehaald van <https://www.cvte.nl/actueel/nieuws/2025/07/01/terugblik-normering-doorstroomtoetsen-2025>

Douma, N. H., Warrens, M. J., Fleur, E., Dijks, M. A., & Korpershoek, H. (2025). The predictive power of Dutch end of primary school tests for educational attainment in ninth grade. *British Educational Research Journal*(51), pp. 280-298. doi:<https://doi.org/10.1002/berj.4074>

Lek, K., & Van de Schoot, R. (2019). Wie weet het beter, de docent of de centrale eindtoets. *De Psycholoog*(4), pp. 10-21. Opgehaald van <https://www.tijdschriftdepsycholoog.nl/wp-content/uploads/2019/04/Lek-K.-van-de-Schoot-R.-2019.-Wie-weet-het-beter-de-docent-of-de-centrale-eindtoets.-De-Psycholoog-4-10-21..pdf>

McHugh, M. L. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica*(22), pp. 276-282.

Ministerie van Onderwijs, Cultuur en Wetenschap. (2025). *Staatscourant 2025, 7385*. Opgehaald van <https://zoek.officielebekendmakingen.nl/stcrt-2025-7385.html>

Stroucken, L., Takkenberg, D., & Béguin, A. (2008). Citotoets en de overgang van basisonderwijs naar voortgezet onderwijs. *CBS Sociaaleconomische trends*. Opgehaald van <https://www.cbs.nl/nl-nl/achtergrond/2008/22/citotoets-en-de-overgang-van-basisonderwijs-naar-middelbaar-onderwijs>

# Appendix A:

## Toelichting landelijke normering commerciële eindtoetsen 2022

In 2022 werd door alle commerciële eindtoetsen de IRT2-normering toegepast. Voor iedere toets werd uitgerekend wat de streefpercentages voor iedere toetsadviescategorie zouden moeten zijn, gebaseerd op een populatie-equivalering met de papieren Centrale Eindtoets op basis van de voorlopige schooladviezen. Door deze methode wordt in feite berekend wat de verwachte percentages toetsadviezen bij iedere eindtoets zouden zijn, indien de toetspopulatie de papieren Centrale Eindtoets gemaakt zou hebben. Deze streefpercentages worden vervolgens vertaald naar grenspunten op de schaal van de gemiddelde latente vaardigheid (GLV), de gewogen gemiddelde vaardigheid die per leerling wordt berekend over de drie verplichte toetsonderdelen. Hierdoor had iedere eindtoets zijn eigen grenspunten op de GLV-schaal. De schaal zelf werd wel via een item-equivalering met het gezamenlijk anker gefixeerd, waardoor een gelijke toetsprestatie tot een gelijke GLV leidde.

In het eindtoetsstelsel werd het gebruik van zowel het eenparameter (1pl) als het tweeparameter (2pl) psychometrische model toegestaan in de kalibratie en normering van de commerciële eindtoetsen. Beide modellen hanteren een andere schaal, en de cesuurpunten van de GLV zijn in beide modellen anders, maar zijn wel beide op dezelfde manier geëquivaerd. Iedere toetsaanbieder had zijn eigen voorkeur voor een bepaald model, wat vervolgens in de normering werd toegepast. **Tabel A1** geeft een overzicht van de voorkeursmodellen per toets in 2022. De Centrale Eindtoets gebruikte een eigen absolute normering die onafhankelijk was van de landelijke normering van de commerciële eindtoetsen. Deze normering was gebaseerd op grenzen die zijn bepaald via doorstroomonderzoek. In de normering van de Centrale Eindtoets werd als psychometrisch model het OPLM gebruikt (Brouwer & Veldkamp, 2022).

**Tabel A1:** Voorkeursmodel voor de normering van iedere commerciële eindtoets in 2022.

eindtoets	voorkeursmodel
AMN	2pl
Dia	2pl
IEP (papier)	1pl
ROUTE 8	1pl

Bovenop de toetsadviesgrenzen zoals bepaald in de IRT2-normering werd ook nog een correctie uitgevoerd op het grenspunt tussen pro/vmbo bb en vmbo bb/vmbo kb. De reden hiervoor was het lage aantal leerlingen dat doorgaans een toetsadvies pro/vmbo bb kreeg bij de papieren Centrale Eindtoets. Hierdoor kon dit grenspunt via de equivalering van de IRT2-normering niet goed bepaald worden. In plaats hiervan werd op basis van de voorlopige schooladviezen voor iedere eindtoets een nieuw grenspunt berekend, waarbij de toetsadviezen zo goed mogelijk aansluiten bij de voorlopige schooladviezen. Een overzicht van alle grenspunten op de GLV-schaal van de commerciële eindtoetsaanbieders zoals toegepast tijdens de normering van 2022 staat in **tabel A2**.

**Tabel A2:** Grenspunten op de GLV-schaal in ieder model voor iedere commerciële eindtoets, na toepassing van IRT2-normering met coronacorrectie en correctie van pro/vmbo bb-grenspunt.

toetsadvies	AMN		Dia		IEP (papier)		ROUTE 8	
	1pl	2pl	1pl	2pl	1pl	2pl	1pl	2pl
pro/vmbo bb								
vmbo bb/vmbo kb	-2,509	-2,198	-2,403	-2,105	-2,106	-2,091	-1,962	-2,124
vmbo kb/vmbo gl-tl	-1,401	-1,361	-1,325	-1,271	-1,128	-1,177	-1,191	-1,340
vmbo gl-tl/havo	-0,681	-0,763	-0,603	-0,705	-0,479	-0,564	-0,699	-0,807
havo/vwo	-0,009	-0,155	0,149	-0,088	0,216	0,111	-0,195	-0,208
vwo	0,489	0,295	0,711	0,398	0,744	0,645	0,179	0,268

# Appendix B:

## Toelichting pseudonimisering en dataverwerking

### Pseudonimisering

De doorstroomgegevens uit het vo van DUO en de toetsresultaten vanuit de toetsaanbieders moeten worden gecombineerd op leerlingniveau. Dat betekent dat persoonsgegevens van leerlingen moeten worden verwerkt. Om te voorkomen dat Stichting Cito en het CvTE persoonsgegevens in beheer moesten nemen, is besloten om alle leerlingen te voorzien van een pseudoniem. Daarvoor hebben zowel DUO als iedere toetsaanbieder los van elkaar een algoritme toegepast. Dat algoritme creëert een uniek pseudoniem voor iedere leerling op basis van de instellingscode van de school, achternaam, geboortedatum en toetsscore van een leerling. Op basis van dit pseudoniem kan Stichting Cito de gegevens van DUO en toetsaanbieders aan elkaar koppelen. Noodzakelijkerwijs zijn twee- of meerlingen met dezelfde toetsscore uitgesloten van dit onderzoek, omdat hun gegevens leiden tot identieke pseudoniemen en deze leerlingen daarmee niet uniek volgbaar zijn in het vo. Ook zijn leerlingen uitgesloten van het onderzoek bij wie één van de vier gegevens voor het pseudonimiseren ontbreken. Daarnaast zijn alle leerlingen verwijderd van wie het pseudoniem alleen voorkomt in de gegevens van DUO of de toetsaanbieder, oftewel zijn alleen gegevens van leerlingen behouden die in beide datasets voorkwamen. Een dergelijke mismatch tussen de pseudoniemen van DUO en de toetsaanbieder kan veroorzaakt worden door complexe leerlingnamen of administratieve fouten in bijvoorbeeld de instellingscode van de school of handmatige registratie van toetsscores.

### Dataverwerking

**Tabel B1** bevat een overzicht en omschrijving van alle variabelen die beschikbaar zijn in de gegevens zoals aangeleverd door DUO en de toetsaanbieders. Na het koppelen van leerlingen uit beide datasets zijn leerlingen verwijderd die: drie jaar op rij hetzelfde leerjaar hadden, leerlingen die meermaals hetzelfde verblijfjaar hadden, leerlingen die onwaarschijnlijk versnelde of vertraagde trajecten doorliepen (bijvoorbeeld een leerling die van havo 2 naar vwo 5 gaat) en leerlingen met inconsistente data (bijvoorbeeld twee verschillende definitieve schooladviezen tussen verblijfsjaren). Voor praktijkonderwijs (pro) zijn er geen gegevens geregistreerd omtrent het leerjaar. Voor deze leerlingen is aangenomen dat het leerjaar gelijk staat aan het verblijfjaar. Leerlingen die door DUO geadministreerd zijn met een schooladvies met leerwegondersteunend onderwijs (lwoo) zijn gecategoriseerd op basis van het schooladvies, dus een leerling met advies vmbo bb lwoo wordt in alle analyses beschouwd als een leerling met schooladvies vmbo bb.

**Tabel B1:** Omschrijving van de gegevens zoals aangeleverd door DUO en toetsaanbieders.

omschrijving variabele	waarden in data	aangeleverd door DUO	aangeleverd door toetsaanbieders
<b>pseudoniem:</b> leerlingcode gebaseerd op 4 variabelen: instellingscode, achternaam, geboortedatum, toetsscore	hash van 64 cijfers/letters	X	X
<b>type onderwijs</b> gevolgd door de leerling in het po	basisonderwijs, speciaal basisonderwijs, speciaal onderwijs	X	X
<b>onthefing:</b> indicatie of leerling verplicht is een eindtoets te maken	J / N	X	
<b>kalenderjaar:</b> afnamejaar eindtoets	2022 t/m 2024	X	
<b>toets:</b> afgenomen eindtoets	AMN, Centrale Eindtoets, Dia, IEP, ROUTE 8	X	X
<b>afnamemodus:</b> medium waarop de toets is afgenomen	papier / digitaal		X
<b>toetsscore:</b> behaalde score op specifieke eindtoets	AMN: 300-500 Centrale eindtoets: 501-550 Dia: 321-390 IEP: 50-100 ROUTE 8: 100-300	X	X
<b>toetsadvies:</b> geadviseerd type onderwijs in het vo op basis van de behaalde toetsscore op de eindtoets	pro/vmbo bb, vmbo bb/vmbo kb, vmbo kb/vmbo gl-tl, vmbo gl-tl/havo, havo/vwo, vwo		X
<b>verblijfsjaar:</b> aantal jaar dat leerling onderwijs volgt op het vo	1, 2, 3	X	X
<b>onderwijsniveau:</b> vo onderwijsniveau dat de leerling volgt (ook wel plaatsingsniveau)	pro, vmbo bb, vmbo bb/vmbo kb, vmbo kb, vmbo kb/vmbo gl-tl, vmbo gl-tl, vmbo gl-tl/havo, havo, havo/vwo, vwo, breed*	X	
<b>leerjaar:</b> jaargang die de leerling volgt	1, 2, 3, 4, 5	X	
<b>voorlopig schooladvies</b> afgegeven door de school vóór afname van de eindtoets	mis, vso, pro, vmbo bb, vmbo bb/vmbo kb, vmbo kb, vmbo kb/vmbo gl, vmbo gl, vmbo kb/vmbo tl, vmbo gl/vmbo tl,	X	
<b>definitief schooladvies</b> afgegeven door de school ná afname van de eindtoets	vmbo tl, vmbo gl/havo, vmbo tl/havo, havo, havo/vwo, vwo**	X	

\* Onder breed vallen de brede brugklassen: vmbo bb/vmbo kb/vmbo gl-tl, vmbo bb/vmbo kb/vmbo gl-tl/havo, vmbo bb/vmbo kb/vmbo gl-tl/havo/vwo, vmbo gl-tl/havo/vwo

\*\* In dit onderzoek zijn vmbo gl en vmbo tl samengevoegd tot vmbo gl-tl wat betreft het gevolgde onderwijsniveau maar niet wat betreft de schooladviesdata.

## Ontbrekende data

De ontbrekende data betreffen enkel de gegevens die ontbreken na het toepassen van de exclusiecriteria en de filters zoals beschreven onder “Exclusiecriteria”. De exclusiecriteria en de filters bepalen namelijk de doelgroep van deze evaluatie en moeten vooraf worden toegepast. Na het toepassen daarvan bleven er data over van 160 912 leerlingen. Voor iedere analyse was een minimaal criterium dat er geen ontbrekende waarden waren op de variabelen toets, toetsadvies, verblijfjaar en leerjaar. Bij ~0,2% van de populatie (853 leerlingen) ontbrak het verblijf- of leerjaar, waardoor de uiteindelijke steekproef 160 059 leerlingen betrof. Deze 853 leerlingen waren gelijkmatig verdeeld over de toetsaanbieders. Voor de volledigheid presenteren we hieronder de aantallen leerlingen die zijn verloren tijdens het koppelen op basis van de pseudoniemen, tijdens het toepassen van de exclusiecriteria en filters en tijdens het verwijderen van leerlingen met ontbrekende data.

**Tabel B2:** Aantal leerlingen per toetsaanbieder na het koppelen, na het toepassen van exclusiecriteria en waar ontbrekende data voorkwamen.

toets	DUO	toetsaanbieder	overlap pseudoniem	onthefing	s(b)o	vso	subtotaal	ontbrekend	totaal	% verloren
AMN	4 208	4 179	4 004	28	829	68	3 176	18	3 158	0,43
CET (totaal)	79 171	79 086	75 163	333	1 176	217	72 527	372	72 155	0,47
Dia	7 803	7 892	7 335	24	552	33	6 730	28	6 702	0,35
IEP (papier)	61 319	61 921	58 864	326	1 021	150	57 137	276	56 861	0,45
ROUTE 8	28 493	28 579	26 856	167	5 650	637	21 342	159	21 183	0,56
Totaal	180 994	181 657	172 222	878	9 228	1105	160 912	853	160 059	0,47

DUO = aantal leerlingen zoals aangeleverd door DUO.

toetsaanbieder = aantal leerlingen zoals aangeleverd door toetsaanbieders.

overlap pseudoniem = aantal overlappende leerlingen (pseudoniemen) tussen data DUO en toetsaanbieder.

onthefing = aantal leerlingen met een ontheffing (exclusie criterium).

s(b)o = aantal leerlingen die het speciaal (basis)onderwijs volgen (exclusie criterium).

vso = aantal leerlingen met een voorlopig of definitief schooladvies voortgezet speciaal onderwijs (exclusie criterium).

subtotaal = aantal leerlingen die gekoppeld konden worden en niet in de exclusiecriteria vielen. Let op: het subtotaal is niet ‘overlap pseudoniem’ minus de exclusiecriteria kolommen, omdat sommige leerlingen in meerdere categorieën van exclusie vielen. Daarnaast zijn ook de filters zoals beschreven toegepast, bijvoorbeeld geen verandering in leerjaar. Vanwege de beperkte omvang van deze groepen zijn deze niet los gerapporteerd.

ontbrekend = aantal leerlingen van het subtotaal dat een of meer ontbrekende gegevens had in toets, leerjaar of verblijfjaar.

totaal = aantal leerlingen dat overbleef na koppeling, exclusie, filtering en verwerking van ontbrekende gegevens. Deze worden gebruikt voor de analyses.

% verloren = percentage leerlingen dat verloren is tussen de aangeleverde data van de toetsaanbieder(s) en de uiteindelijke dataset gebruikt in de analyses.

De administratie rondom de leerlingen die praktijkonderwijs volgen is anders, bijvoorbeeld omdat er voor deze leerlingen geen leerjaren worden vastgelegd. Desondanks hebben we de leerlingen die pro volgen op dezelfde manier meegenomen in alle dataverwerking en analysestappen om een volledig beeld te kunnen geven van de beschikbare data van de leerlingen in het regulier vo. De pro-groep omvat naar verhouding een klein aantal leerlingen en zal daarmee weinig impact hebben op de analyses, maar bij de interpretatie van de resultaten moet hiermee rekening worden gehouden.

## Gegevenskoppeling

**Tabel B3** geeft extra inzicht in de mate waarin leerlinggegevens van DUO en de eindtoetsaanbieders succesvol gekoppeld konden worden op basis van het pseudoniem.

**Tabel B3:** Geslaagde koppelingen per toetsadvies en eindtoets. Het aantal leerlingen aangeleverd door de eindtoets (n toets) is als leidend beschouwd, met daarnaast het aantal en percentage leerlingen dat is overgebleven in de dataset na koppeling.

toetsadvies	toets	n toets	gekoppeld		gem. % toetsadvies
			n	%	
pro/vmbo bb	AMN	376	338	89,9	86,6
	CET (totaal)	1 269	1 082	85,3	
	Dia	368	329	89,4	
	IEP (papier)	1 057	854	80,8	
	ROUTE 8	2 542	2 258	88,8	
vmbo bb/vmbo kb	AMN	800	762	95,2	92,3
	CET (totaal)	11 445	10 533	92,0	
	Dia	1 222	1 117	91,4	
	IEP (papier)	8 450	7 832	92,7	
	ROUTE 8	5 579	5 129	91,9	
vmbo kb/vmbo gl-tl	AMN	834	799	95,8	94,2
	CET (totaal)	16 473	15 554	94,4	
	Dia	1 532	1 405	91,7	
	IEP (papier)	12 958	12 221	94,3	
	ROUTE 8	5 529	5 177	93,6	
vmbo gl-tl/havo	AMN	1 035	998	96,4	95,4
	CET (totaal)	20 702	19 777	95,5	
	Dia	2 332	2 161	92,7	
	IEP (papier)	17 898	17 134	95,7	
	ROUTE 8	6 893	6 578	95,4	
havo/vwo	AMN	598	578	96,7	96,1
	CET (totaal)	15 572	14 999	96,3	
	Dia	1 248	1 187	95,1	
	IEP (papier)	11 406	10 978	96,2	
	ROUTE 8	4 229	4 035	95,4	
vwo	AMN	536	529	98,7	96,9
	CET (totaal)	13 625	13 218	97,0	
	Dia	1 190	1 136	95,5	
	IEP (papier)	10 152	9 845	97,0	
	ROUTE 8	3 807	3 679	96,6	

**Tabel B4:** Overeenkomende (1) en niet-overeenkomende (0) toetsadviezen en onderwijsniveaus.

toetsadvies	pro	vmbo bb	vmbo bb/vmbo kb	vmbo kb	vmbo bb/vmbo kb/ vmbo gl-tl	vmbo kb/vmbo gl-tl	vmbo bb/vmbo kb/ vmbo gl-tl/havo	vmbo gl-tl	vmbo bb/vmbo kb/ vmbo gl-tl/havo/vwo	vmbo gl-tl/havo	havo	vmbo gl-tl/havo/vwo	havo/vwo	vwo
pro/vmbo bb	1	1	1	0	1	0	1	0	1	0	0	0	0	0
vmbo bb/vmbo kb	0	1	1	1	1	1	1	0	1	0	0	0	0	0
vmbo kb/vmbo gl-tl	0	0	1	1	1	1	1	1	1	1	0	1	0	0
vmbo gl-tl/havo	0	0	0	0	1	1	1	1	1	1	1	1	1	0
havo/vwo	0	0	0	0	0	0	1	0	1	1	1	1	1	1
vwo	0	0	0	0	0	0	0	0	1	0	0	1	1	1

# Appendix C:

## Beschrijvende statistieken

**Tabellen C1 tot en met C4** beschrijven de gegeven voorlopige schooladviezen, de gegeven toetsadviezen, het aantal doubleurs en versnellers, en het aantal bijstellingen per eindtoets in 2022.

**Tabel C1:** Aantal leerlingen per eindtoets en voorlopige schooladviezen. Kolom "mis" geeft het aantal leerlingen waarvan dit gegeven ontbreekt.

eindtoets	mis	pro	vmbo bb	vmbo bb/vmbo kb	vmbo kb	vmbo kb/vmbo gl-tl	vmbo gl-tl	vmbo gl-tl/havo	havo	havo/vwo	vwo
AMN	94	22	179	136	289	192	532	392	449	346	527
CET (digitaal)	202	144	1 468	1 062	2 096	1 204	3 471	2 098	3 156	2 262	3 935
CET (papier)	458	150	2 099	1 760	4 356	2 334	8 600	5 121	8 873	6 069	11 237
Dia	50	32	448	285	694	389	1 151	673	1 050	723	1 207
IEP (papier)	412	311	3 165	2 327	5 421	3 130	9 816	6 137	9 283	6 380	10 479
ROUTE 8	72	129	1 240	878	2 110	1 333	3 692	2 236	3 451	2 274	3 768

**Tabel C2:** Aantal leerlingen per eindtoets en gegeven toetsadviezen.

eindtoets	pro/vmbo bb	vmbo bb/vmbo kb	vmbo kb/vmbo gl-tl	vmbo gl-tl/havo	havo/vwo	vwo
AMN	51	420	664	942	562	519
CET (digitaal)	429	3 684	4 727	5 722	3 910	2 626
CET (papier)	228	5 906	10 295	13 524	10 772	10 332
Dia	122	884	1 310	2 099	1 163	1 124
IEP (papier)	521	7 087	11 855	16 793	10 858	9 747
ROUTE 8	310	2 828	4 418	6 157	3 891	3 579

**Tabel C3:** Aantal en percentage leerlingen per eindtoets dat nominaal onderwijs volgt, doubleert of versnelt per verblijfsjaar (1, 2 en 3).

toets	tempo	n jaar 1	n jaar 2	n jaar 3	% jaar 1	% jaar 2	% jaar 3
AMN	nominaal	3 153	3 092	2 922	100,0	98,7	94,3
AMN	versneller	0	4	4	0,0	0,1	0,1
AMN	doubleur	0	37	174	0,0	1,2	5,6
CET (digitaal)	nominaal	21 081	20 631	19 682	100,0	98,3	94,6
CET (digitaal)	versneller	2	8	9	0,0	0,0	0,0
CET (digitaal)	doubleur	0	340	1 106	0,0	1,6	5,3
CET (papier)	nominaal	51 007	50 090	47 707	100,0	98,6	94,5
CET (papier)	versneller	7	22	38	0,0	0,0	0,1
CET (papier)	doubleur	0	704	2 762	0,0	1,4	5,5
Dia	nominaal	6 687	6 586	6 260	99,9	98,7	94,6
Dia	versneller	5	5	7	0,1	0,1	0,1
Dia	doubleur	0	79	349	0,0	1,2	5,3
IEP (papier)	nominaal	56 794	55 632	52 756	100,0	98,4	94,1
IEP (papier)	versneller	10	24	30	0,0	0,0	0,1
IEP (papier)	doubleur	0	891	3 266	0,0	1,6	5,8
ROUTE 8	nominaal	21 157	20 775	19 729	100,0	98,7	94,6
ROUTE 8	versneller	6	9	13	0,0	0,0	0,1
ROUTE 8	doubleur	0	269	1 115	0,0	1,3	5,3

**Tabel C4:** Aantal leerlingen per eindtoets met een bijstelling van het voorlopige schooladvies.

toets	bijstelling	geen bijstelling	% bijgesteld
AMN	211	2 947	6,7
CET (digitaal)	1 468	19 630	7,0
CET (papier)	3 776	47 281	7,4
Dia	500	6 202	7,5
IEP (papier)	4 344	52 517	7,6
ROUTE 8	1 749	19 434	8,3

# Appendix D:

## Simulatiestudie misclassificaties normering

In deze appendix worden twee deelstudies uitgewerkt met behulp van een simulatie. De eerste deelstudie betreft de grootte van misclassificatie rondom de toetsadviesgrenzen als gevolg van meetonzekerheid en de tweede deelstudie betreft het effect van populatieverschillen op de IRT2-normering.

### Deelstudie 1: Inschatting van classificatiefout toetsadvies door meetonzekerheid

In de normering wordt het toetsadvies bepaald op basis van de GLV (gemiddelde latente vaardigheid). Dit is een gewogen gemiddelde van de latente vaardigheden op de onderdelen Lezen (gewicht = 0,3), Taalverzorging (gewicht = 0,25) en Rekenen (gewicht = 0,45). De inschatting van de latente vaardigheid van een leerling kent een bepaalde onzekerheidsmarge omdat we niet volledig betrouwbaar kunnen meten met een toets. Deze drukken we uit in de standaardmeetfout (SE). De standaardmeetfout op de GLV noemen we  $SE_{GLV}$ . Voor elke leerling die de papieren CET in 2022 heeft gemaakt worden 1 000 mogelijke GLV's getrokken (gesimuleerd) uit een individuele verdeling (normaalverdeling met gemiddelde = gemeten GLV en  $sd = SE_{GLV}$ ). Voor elke gesimuleerde GLV wordt het toetsadvies bepaald. Een afwijking in het gesimuleerde toetsadvies ten opzichte van de gemeten GLV is een indicatie van de invloed van de meetonzekerheid op het toetsadvies.

In **tabel D1** staat een overzicht van de gevonden misclassificaties als gevolg van deze meetonzekerheid. Bij toetsadviezen vmbo bb/vmbo kb en vmbo kb/vmbo gt-tl is de misclassificatie eerder richting een hoger toetsadvies terwijl bij het toetsadvies havo/vwo de misclassificatie eerder richting een lager toetsadvies is. Dit verschil kan verklaard worden doordat binnen de categorie vmbo bb/vmbo kb meer leerlingen een vaardigheid in de buurt van de bovengrens hebben dan bij de ondergrens, en binnen de categorie havo/vwo meer leerlingen een vaardigheid in de buurt van de ondergrens hebben dan bij de bovengrens. De totale misclassificatie is 23,3% met 12,3% een lager toetsadvies en 11,0% een hoger toetsadvies.

**Tabel D1:** Gemiddelde percentage (over 1 000 replicaties) leerlingen CET papier 2022 die een lager, gelijk of hoger toetsadvies kan krijgen door meetonzekerheid.

toetsadvies	n	% lager	% gelijk	% hoger
pro/vmbo bb	154		67,9	32,1
vmbo bb/vmbo kb	6 582	2,3	85,1	12,6
vmbo kb/vmbo gl-tl	11 475	9,6	76,9	13,5
vmbo gl-tl/havo	15 092	12,3	74,8	12,9
havo/vwo	11 215	18,1	67,1	14,8
vwo	10 088	15,5	84,5	
Totaal	54 606	12,3	76,7	11,0

De papieren CET had in 2022 de meest vaardige leerlingpopulatie als we uitgaan van de voorlopige schooladviezen en de plaatsing in jaar 3 van het vo (zie ook **figuur 2 en 3**). De andere eindtoetsen, waaronder ook de digitale CET, hadden een minder vaardige populatie. Zoals hierboven geïllustreerd hangt het aantal misclassificaties af van de verdeling van leerlingvaardigheden in relatie tot de toetsadviesgrenzen. Om een idee te hebben hoe de classificatiefout door meetonzekerheid is bij een minder vaardige populatie is een GLV gesimuleerd uit een mindervaardige verdeling (effectgrootte = -0,5,  $n = 10\ 000$ ) en is er een voorspelde  $SE_{GLV}$  gekoppeld op basis van CET papier 2022. De resultaten van deze simulatie zijn te zien in **tabel D2**. De verdeling van toetsadviezen is verschoven (er zijn dus minder leerlingen in de categorie vwo en meer in de categorie vmbo bb/vmbo kb), maar de grootte en richting van misclassificatie is vergelijkbaar met CET papier 2022. De populatieverschillen tussen de

verschillende toetsen hebben dus slechts beperkt invloed op het totaal aantal misclassificaties als gevolg van de meetonzekerheid, zeker omdat de effectgrootte van -0,5 vrij groot is in vergelijking met de daadwerkelijke verschillen tussen de populaties van de verschillende eindtoetsen. Als wordt ingezoomd op categorieën aan de randen (pro/vmbo bb en vwo), dan is wel een duidelijk verschil zichtbaar.

**Tabel D2:** Gemiddelde percentage (over 1 000 replicaties) van 10 000 gesimuleerde GLV's uit een minder vaardige populatie (effectgrootte = -0,5) dat een lager, gelijk of hoger toetsadvies kan krijgen door meetonzekerheid.

toetsadvies	n	% lager	% gelijk	% hoger
pro/vmbo bb	465		85,1	14,9
vmbo bb/vmbo kb	2 123	4,6	85,6	9,9
vmbo kb/vmbo gl-tl	2 434	10,9	77,0	12,2
vmbo gl-tl/havo	2 457	13,0	75,0	12,0
havo/vwo	1 537	18,9	66,8	14,3
vwo	984	18,1	81,9	
Totaal	10 000	11,5	77,6	10,9

## Deelstudie 2: Effect van IRT2-normering op de classificatiefout van het toetsadvies

In de IRT2-normering werden voor iedere commerciële eindtoets streefpercentages voor de toetsadviezen uitgerekend door te kijken naar de toetsresultaten van leerlingen met een gelijk voorlopig schooladvies in de papieren CET. Uit de papieren CET werden leerlingen getrokken (met teruglegging) zodat de verhouding voorlopig schooladvies gelijk is aan de verdeling van voorlopig schooladviezen van de leerlingpopulatie die deelneemt aan een bepaalde toets. Vervolgens werd gekeken in de getrokken steekproef uit de papieren CET hoe de verdeling van toetsadviezen is. Deze verdeling werd dan overgenomen door de toets voor het bepalen van de grensscores tussen de toetsadviezen (zie ook appendix A).

Een belangrijke aanname voor de populatie-equivalering van de IRT2-normering is dat de vaardigheidsverdelingen binnen de voorlopig schooladviescategorieën gelijk zijn tussen de toets die wordt genormeerd en de papieren CET. Dit is niet het geval als de toetspopulatie van de genormeerde toets verschilt van de papieren CET.

Omdat alle commerciële eindtoetsen een minder vaardige populatie hebben dan de papieren CET, wordt de aanname van gelijke vaardigheidsverdelingen binnen voorlopig schooladviescategorieën geschonden en leidt dit tot classificatiefouten bij het toekennen van toetsadviezen in de IRT2-normering. Om dit beter te begrijpen en een idee te krijgen van de grootte van dit effect, voeren we een simulatiestudie uit.

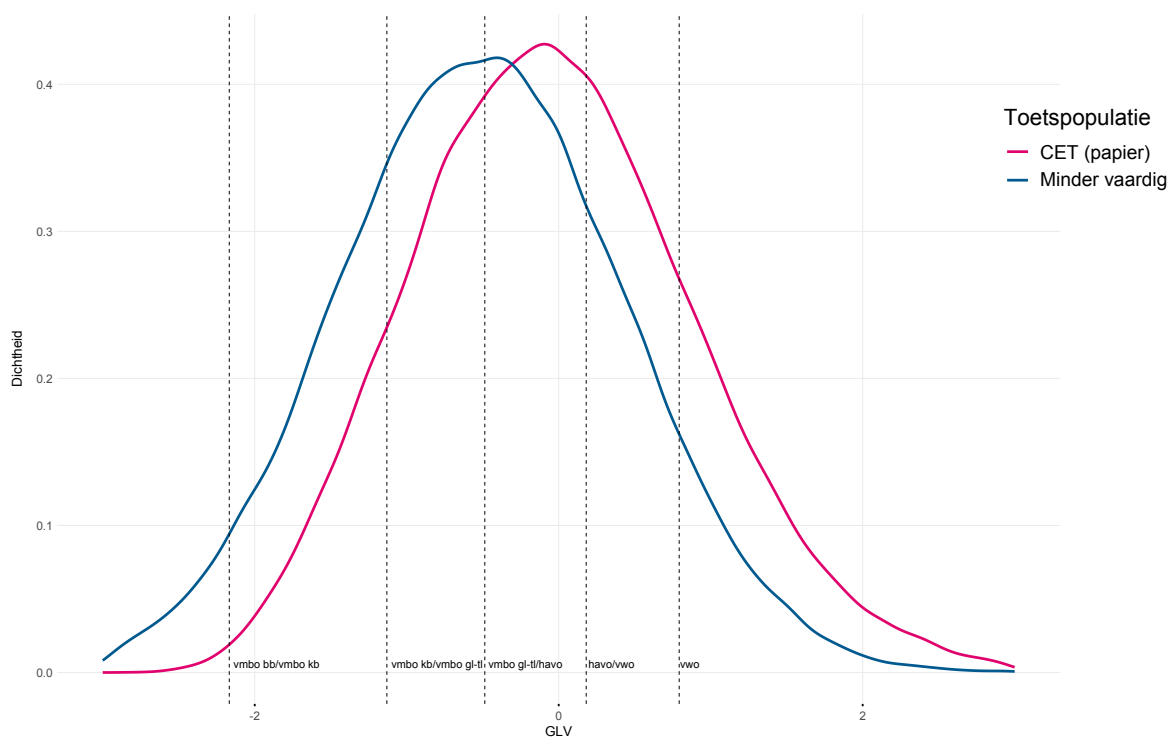
Hiervoor simuleren we een leerlingpopulatie die minder vaardig is dan de papieren CET uit 2022 (effectgrootte = -0,5, n = 10 000). Dit is een relatief groot effect in vergelijking met de daadwerkelijke populatieverschillen tussen de toetsen, maar is geschikt om de effecten duidelijk te illustreren. Deze gesimuleerde leerlingen hebben enkel een GLV, maar hun schooladvies is nog onbekend. Daarom imputeren we voor deze leerlingen een schooladvies op basis van de schooladviezen van de CET papier 2022<sup>6</sup>. Op de geïmputeerde schooladviezen en gesimuleerde GLV's is vervolgens de IRT2-normering toegepast. De toetsadviezen volgens de IRT2-normering worden vergeleken met de toetsadviezen die toegekend kunnen worden op basis van de GLV en de toetsadviesgrenzen van de papieren CET (de absolute IRT-normering). Het verschil in de classificatie volgens beide normeringen kan worden gezien als de classificatiefout

6 Dit geïmputeerde schooladvies wordt als volgt bepaald: de GLV-schaal wordt in 100 bins opgedeeld, bijvoorbeeld 1,0-1,1; 1,2-1,3; 1,3-1,4; et cetera. Voor iedere bin kunnen we bepalen welke leerlingen uit de papieren CET 2022 in die bin vallen, gebaseerd op hun bekende GLV. Van de CET leerlingen binnen 1 bin kunnen we de schooladviesverdeling bepalen. Om schooladviezen te imputeren, zoeken we voor iedere gesimuleerde leerling (met een GLV) op in welke bin de leerling valt. Daarna trekken we willekeurig een schooladvies uit de schooladviesverdeling die hoort bij die bin zoals die is bepaald op de leerlingen uit de papieren CET 2022. Op die manier volgen de geïmputeerde schooladviezen dezelfde verdeling als de schooladviezen van de papieren CET 2022, maar dan verschoven op de GLV-schaal.

geïntroduceerd door de IRT2-normering. Deze simulatie is 100 keer herhaald.

De verdeling van de GLV van de papieren CET uit 2022 en de gesimuleerde minder vaardige toetspopulatie zijn weergegeven in **figuur D1**. Hierin is ook te zien hoe dit binnen categorieën (in dit geval toetsadvies) leidt tot een verschillende vaardigheidsverdeling. Dit is het meest duidelijk te zien in de toetsadviescategorie vmbo gl-tl/havo, waar de piek van de minder vaardige leerlingen dichter tegen de grens ligt dan de piek van de CET leerlingen.

**Figuur D1:** Verdeling van de GLV voor de papieren CET uit 2022 (rode lijn) en de gesimuleerde minder vaardige populatie (blauwe lijn, effectgrootte = -0,5) met toetsadviesgrenzen uit de papieren CET.



In **tabel D3** is het gemiddelde en de spreiding van de GLV in iedere categorie van het voorlopig schooladvies uitgerekend voor de gesimuleerde toets met een minder vaardige populatie en de papieren CET weergegeven uit één trekking van de simulatie. Door het verschil in populatie is binnen iedere categorie van het voorlopig schooladvies de gemiddelde GLV lager bij de gesimuleerde toets met minder vaardige populatie dan in de papieren CET.

**Tabel D3:** Gemiddelde GLV per categorie voorlopig schooladvies van de gesimuleerde toets met minder vaardige populatie (effectgrootte = -0,5; n = 10 000) en CET papier 2022 (voorbeeld van één trekking).

voorlopig schooladvies	n	CET papier gem. GLV	minder vaardige populatie gem. GLV
pro	94	-1,88	-2,11
vmbo bb	1 030	-1,50	-1,79
vmbo bb/vmbo kb	657	-1,23	-1,47
vmbo kb	1 316	-0,98	-1,16
vmbo kb/vmbo gl-tl	594	-0,70	-0,82
vmbo gl-tl	1 865	-0,51	-0,63
vmbo gl-tl/havo	976	-0,18	-0,30
havo	1 468	0,13	-0,02
havo/vwo	868	0,45	0,29
vwo	1 132	1,05	0,79

Puur door de vorm van de verdelingen heeft de minder vaardige populatie in dezelfde schooladviescategorie dus

gemiddeld genomen een lagere GLV dan de CET-populatie. In **tabel D4** is te zien dat hetzelfde verschil in gemiddelde GLV te zien is bij de toekenning van de toetsadviezen met de IRT2-normering: de minder vaardige populatie heeft een lagere GLV dan de CET papier binnen hetzelfde toetsadvies. Met andere woorden: de IRT2-normering zorgt voor een overschatting bij het toekennen van de toetsadviezen als de populatie van de genormeerde toets minder vaardig is.

**Tabel D4:** Gemiddelde GLV van de gesimuleerde toets met minder vaardige populatie (effectgrootte = -0,5; n = 10 000) en de papieren CET na toekenning van toetsadviezen op basis van de IRT2-normering (voorbeeld van één trekking).

toetsadvies IRT2	n	CET Papier gem. GLV	minder vaardige populatie gem. GLV
pro/vmbo bb	77	-2,31	-3,12
vmbo bb/vmbo kb	2 045	-1,50	-1,75
vmbo kb/vmbo gl-tl	2 568	-0,80	-0,92
vmbo gl-tl/havo	2 692	-0,16	-0,29
havo/vwo	1 532	0,45	0,30
vwo	1 086	1,34	1,07

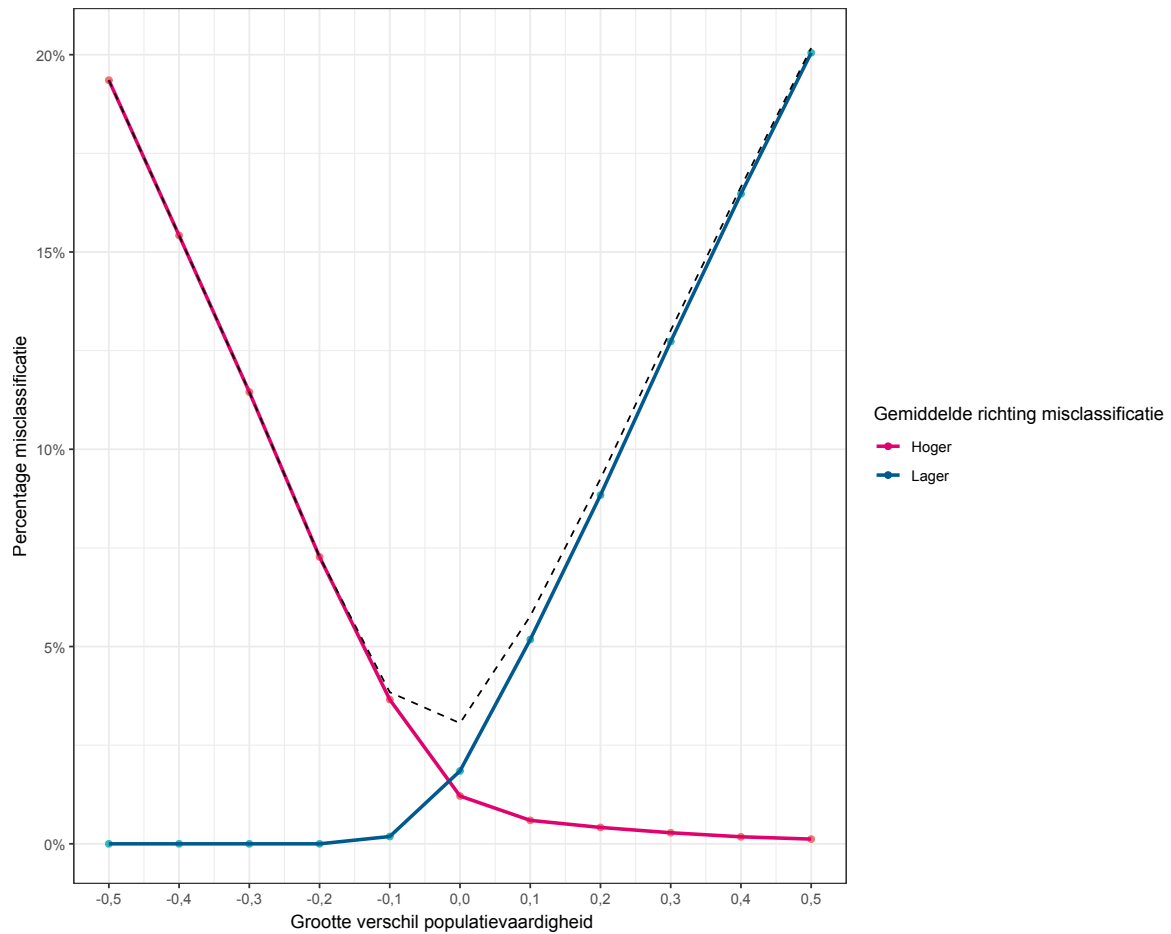
De kruistabel in **tabel D5** geeft de misclassificaties in de IRT2-normering weer voor het gesimuleerde populatieverschil (effectgrootte -0,5). Er is een systematische neiging tot hogere toetsadviezen (19,6% krijgt hoger toetsadvies) bij de IRT2-normering ten opzichte van de situatie waarin de absolute toetsadviesgrenzen van de papieren CET toegepast zouden worden.

**Tabel D5:** Percentage misclassificaties van toetsadviezen in de gesimuleerde toets met minder vaardige populatie (effectgrootte -0,5) door toepassing van de IRT2-normering (rij) in vergelijking met de toepassing van de absolute toetsadviesgrenzen van de papieren CET (kolom), gemiddeld over 100 replicaties.

toetsadvies IRT2	totaal	toetsadvies normering CET					
		pro/vmbo bb	vmbo bb/vmbo kb	vmbo kb/vmbo gl-tl	vmbo gl-tl/havo	havo/vwo	vwo
pro/vmbo bb	0,7%	0,7%					
vmbo bb/vmbo kb	20,5%	3,3%	17,3%				
vmbo kb/vmbo gl-tl	25,8%		4,7%	21,1%			
vmbo gl-tl/havo	26,3%			4,6%	21,8%		
havo/vwo	15,5%				4,0%	11,5%	
vwo	11,1%					3,0%	8,1%
<b>Totaal</b>		<b>4,0%</b>	<b>21,9%</b>	<b>25,7%</b>	<b>25,8%</b>	<b>14,6%</b>	<b>8,1%</b>

Het percentage misclassificaties en de richting van de misclassificaties hangen in de IRT2-normering af van de grootte van het populatieverschil tussen de toets die genormeerd wordt en de papieren CET. In het uitgewerkte voorbeeld hierboven wordt alleen de situatie met een effectgrootte van -0.5 beschreven. In **figuur D2** is het percentage misclassificaties voor verschillende populatieverschillen weergegeven zoals berekend met deze simulatiestudie. Zo is het eerste datapunt bij de effectgrootte -0.5 gelijk aan de 19,6% misclassificatie zoals getoond in **tabel D5**. Te zien is dat het percentage misclassificaties afneemt als het effect dichter bij 0 komt. De misclassificatie van 3% bij effectgrootte = 0 komt door het trekken met teruglegging in schooladviescategorieën in de IRT2-normering, waardoor de toetsadviezen niet exact overeenkomen. Als een toets genormeerd wordt via de IRT2-normering die een populatie heeft die juist sterker is dan de papieren CET zal het effect omkeren, dus zal dit leiden tot een onderschatting van de toetsadviezen.

**Figuur D2:** Percentage misclassificaties in de IRT2-normering (misclassificatie hoger = rood, lager = blauw, totaal = gestreepte lijn) als functie van het verschil in leerlingpopulatie per effectgrootte van minder vaardig -0,5 tot vaardiger 0,5.



Colofon

Namens Stichting Cito in opdracht van het

**College voor Toetsen en Examens (CvTE)**

© 2026 CvTE, Utrecht. Alle rechten voorbehouden, met uitzondering van het recht het werk openbaar te maken of te vermenigvuldigen.

Juni 2026