

## Is het eindexamen gelijkwaardig tussen scholen? Discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen in het voortgezet onderwijs.

J. Dronkers<sup>1</sup>

### 1. Inleiding

De gelijkwaardigheid van het eindexamen van het voortgezet onderwijs is een belangrijk kenmerk van het Nederlandse onderwijsbestel<sup>2</sup>. Die gelijkwaardigheid, die voortvloeit uit de revolutionaire achtergrond van het Nederlandse onderwijsbestel<sup>3</sup>, heeft twee aspecten. Het eerste heeft betrekking op de gelijkwaardigheid in de tijd. Eindexamens van het ene schooljaar zijn in principe vergelijkbaar met die van een ander schooljaar. De door de minister vastgestelde examenreglementen en de werkzaamheden van het CITO en CEVO (Alberts, 1997) moeten deze gelijkwaardigheid in de tijd garanderen. Deze bijdrage handelt niet over dit temporele aspect, maar gaat over het tweede aspect van de gelijkwaardigheid van het eindexamen van het voortgezet onderwijs: de vergelijkbaarheid van de eindexamenuitslagen tussen scholen. Het eindexamenresultaat van de ene school is in principe vergelijkbaar met dat van de andere school, waardoor het einddiploma voortgezet onderwijs een minimum kwalificatieniveau van de afgestudeerden garandeert, ongeacht de school die de betrokkene heeft bezocht. Opnieuw zijn de door de minister vastgestelde examenreglementen en de werkzaamheden van het CITO en CEVO van belang om deze gelijkwaardigheid te garanderen.

De gelijkwaardigheid van het eindexamen van het voortgezet onderwijs heeft twee belangrijke gevolgen voor het functioneren van het Nederlandse onderwijsstelsel. Het eerste gevolg is dat de noodzaak voor een toelatingsexamen voor het hoger onderwijs minder groot is, want de ondergrens van het kwalificatieniveau van de afgestudeerden is door de gelijkwaardigheid gegarandeerd en de instellingen voor hoger onderwijs behoeven niet bevreesd te zijn voor leerlingen met onvoldoende kwalificaties. Een tweede gevolg is dat op de langere termijn verschillen tussen scholen in het voortgezet onderwijs afnemen. Als alle leerlingen hetzelfde eindexamen moeten afleggen, dwingt dat scholen op de langere termijn een min of meer gelijke maatstaf bij hun onderwijs te gebruiken. Dit proces van vermindering van verschillen tussen scholen wordt in het Nederlandse onderwijs nog versterkt door de gelijkwaardigheid van de financiering van leerkrachten en investeringen. Deze gelijkwaardigheid verklaart bijvoorbeeld waarom in Nederland elitaire kostscholen en privaat onderwijs in de loop van de twintigste eeuw verdwenen zijn: zij hadden een te gering kwaliteitsvoordeel in vergelijking met de scholen die uit algemene middelen werden gefinancierd en die dankzij de gelijkwaardigheid van het eindexamen een min of meer vergelijkbaar kwaliteitsniveau konden garanderen.

Een van de belangrijkste instrumenten om de gelijkwaardigheid van het eindexamen tussen scholen te verzekeren is het centraal vaststellen van dit eindexamen. Maar ook in Nederland kan men niet echt spreken van een centraal vastgesteld eindexamen, omdat het eindcijfer voor elk vak op dat eindexamen uit twee gelijke delen opgebouwd is: het cijfer van het schoolonderzoek en het cijfer van het centraal examen. Daarbij veronderstelt men dat scholen het schoolonderzoek op overeenkomstige wijze becijferen. De bewaking van de gelijkwaardigheid van de becijfering van schoolonderzoek tussen scholen gebeurt door de onderwijsinspectie. Onderzoekers hebben zich echter nauwelijks beziggehouden met de mogelijke verschillen tussen de becijfering van het schoolonderzoek en het centraal eindexamen, met de publicatie van Pijl (1991) als enige uitzondering.

Nu dankzij het beroep van het dagblad *Trouw* op de *Wet Openbaarheid van Bestuur* de nodige inspectiegegevens openbaar zijn geworden, kan de gelijkwaardigheid van de becijfering van beide onderdelen van het eindexamen empirisch getoetst worden. Daarbij gaat mijn belangstelling vooral uit naar de vraag of scholen, al dan niet op goede gronden, verschillen in hun becijfering van het schoolonderzoek. De vragen van deze bijdrage luiden derhalve: *Hoe groot zijn de discrepanties tussen de gemiddelde cijfers voor het schoolonderzoek en het centraal examen? Verschilt de omvang van deze discrepanties tussen scholen? Kunnen deze discrepanties verklaard worden door vak- en schoolkenmerken? Welk gevolg kunnen deze discrepanties voor de gelijkwaardigheid van het eindexamen hebben?*

Het gaat in deze bijdrage niet om de vraag of het centraal schriftelijk examen de beste beoordelingsmaatstaf van scholen is. Het gaat hier alleen om de omvang van de discrepanties en de

verschillen daarin tussen scholen. Uiteraard gaat het ook niet om de integriteit van de individuele leerkrachten bij hun becijfering van examens. Veeleer gaat het hier om een eerste verkenning van een weinig onderzocht maar centraal kenmerk van het Nederlands onderwijs: een naast elkaar bestaand schoolonderzoek en centraal examen, die beide het eindexamen vormen.

## **2. Oorzaken van discrepanties tussen cijfers voor het schoolonderzoek en centraal examen.**

Discrepanties in de becijfering van het centraal examen en het schoolonderzoek zijn nooit het voorwerp geweest van systematische reflectie. Daarom moet ik in het kader van deze bijdrage mij beperken tot enkele voorzetten.

Een eerste reden voor discrepanties tussen de schoolonderzoek- en de centraal-examencijfers kan binnen een bepaald vak liggen. Allereerst kan de verdeling van de lesstof over beide delen van het eindexamen verschillend zijn. De door de minister vastgestelde examenreglementen geven aan welke stof van een bepaald vak in het centraal deel en welke stof in het schoolonderzoek getoetst zal worden. Naar mate de overeenkomst tussen de getoetste stof in de beide examenonderdelen geringer is, neemt de kans toe dat discrepanties tussen de becijfering van beide onderdelen groter wordt. De stof van het ene onderdeel kan immers gemakkelijker zijn dan die van het andere onderdeel. Als de ongelijke verdeling van de stof bij een bepaald vak over de beide onderdelen in de verschillende examenjaren stabiel is, zal het schoolonderzoekcijfer steeds op dezelfde wijze afwijken het centraal-examencijfer. Als echter de verdeling van de ongelijke stof over de beide onderdelen in de verschillende examenjaren niet min of meer stabiel is, zal het schoolonderzoekcijfer in het ene examenjaar op een andere manier afwijken van het centraal-examencijfer dan in een ander examen jaar: bijvoorbeeld in het ene jaar lager en in het andere jaar hoger. Ten tweede kan binnen een bepaald vak een 'becijferingstraditie' bestaan. Het is immers mogelijk dat in de loop der tijd bij een bepaald vak tradities zijn gegroeid bij de becijfering van onderdelen van dat vak, bijvoorbeeld kennis van de grammatica wordt lager becijferd dan de kwaliteit van een vertaling of spreekbeurt. Een dergelijke 'becijferingstraditie' wijst op verschillen in oordeel van het belang van onderdelen binnen een vak. Dit verschillend oordeel over het belang van onderdelen van vakken kan vele jaren voortduren, waardoor het cijfer voor het schoolonderzoek gedurende vele examenjaren op een min of meer stabiele wijze zal afwijken van die voor het centraal examen. Op grond van deze overwegingen kan verondersteld worden dat uiteenlopende discrepanties bij verschillende vakken zullen bestaan.

Een tweede reden voor discrepanties tussen het schoolonderzoek- en het centraal-examencijfer kan een verschil in becijfering van het centraal eindexamen of het schoolonderzoek zijn. Weliswaar probeert het CITO equivalentie van het centraal examen over de verschillende schooljaren te garanderen, dat hoeft niet te lukken (Alberts, 1997; de Gloppe & van Schooten, 1998). Ook is het mogelijk dat scholen hun schoolonderzoek minder laag beoordelen wanneer het een eerste lichte scholieren betreft die een nieuw curriculum gevolgd hebben (bijvoorbeeld basisvorming) het examenjaar bereikt hebben of als slagingspercentages en eindexamencijfers van scholen plotseling in het centrum van de publieke belangstelling zijn gekomen. In het examenjaar 1997/98 werden scholen in het voortgezet onderwijs geconfronteerd met de publicatie van hun intern rendement (Agerbeek et al., 1997), wat gevolgen kon hebben voor hun leerlinginstroom (Dronkers, 1999). Ook bereikten grotere groepen leerlingen die basisvorming hadden gekregen nu het eindexamen. Daarom lijkt het waarschijnlijk dat de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer groter zullen zijn in het examenjaar 1997/98 dan in examenjaar 1996/97.

Een derde reden voor discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen kan een verschil in leerlingpopulatie tussen scholen zijn. Het oordeel van leerkrachten van de prestaties van leerlingen is bijna altijd een relatief oordeel, gebaseerd op de totale leerlingpopulatie (Maas en Meijnen, 1999). Allereerst betekent dit dat leerkrachten de prestaties van een individuele leerling altijd vergelijken met die van medeleerlingen: een bepaalde prestatie in een klas met veel slecht presterende leerlingen wordt hoger becijferd dan een soortgelijke prestatie in een klas met veel goed presterende leerlingen. Ten tweede betreffen leerkrachten bij hun oordeel de condities waaronder de prestaties geleverd zijn: prestaties van leerlingen die door migratie nog een taalachterstand hebben worden bijvoorbeeld vaker hoger becijferd dan die van leerlingen die al hun hele leven in Nederland wonen (Mulder, 1996). Tenslotte interpreteren leerkrachten dezelfde prestaties

van verschillende leerlingen anders, afhankelijk van de achtergrond en geschiedenis van de betrokken leerling: een slechte prestatie van een leerling uit een hooggeschoold gezin wordt sneller geïnterpreteerd als een tijdelijke terugslag, terwijl een goede prestatie van een leerling uit een laaggeschoold gezin sneller verklaard wordt als een 'toevalstreffer' (Hoeven-van Doornum, 1990). Dit relatief oordelen door leerkrachten is geen recent verschijnsel: de 'Wet van Posthumus' (Posthumus, 1940) vloeit uit dit verschijnsel voort, evenals de overadvisering van allochtone leerlingen en leerlingen met hooggeschoolde ouders (Dronkers, van Erp, Robijn & Roeleveld, 1998; Hof & Dronkers, 1993). Ook is het relatief oordelen door leerkrachten niet altijd fout: overgeadviseerde leerlingen aan het eind van de basisschool doen het in de onderbouw van het voortgezet onderwijs niet slechter (Koeslag & Dronkers, 1994). Maar dit relatief oordelen door leerkrachten kan wel leiden tot discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen. Daarom zou men kunnen veronderstellen dat in scholen met veel migranten de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer groter zullen zijn.

Een vierde reden voor discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen kan de didactische of levensbeschouwelijke opvattingen van een school zijn met betrekking tot het belang van eindexamen en de becijfering daarvan. Scholen, die door hun didactische of levensbeschouwelijke opvattingen minder belang hechten aan de kennis en vaardigheden die gemeten worden in het centraal examen, zullen daaraan minder aandacht besteden en dus gemiddeld lagere cijfers voor dat centraal deel halen, waardoor de discrepanties groter kunnen worden. Ook kunnen deze scholen in hun schoolonderzoek meer andere kennis en vaardigheden, waaraan zij meer belang hechten en dus die beter onderwezen hebben, toetsen. Daardoor kunnen discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen ontstaan. Op grond van het belang van didactische en levensbeschouwelijke opvattingen van Montessori en orthodox-protestantse scholen (Dijkstra, Dronkers & Hofman, 1997) zou men kunnen verwachten dat in orthodox-protestantse scholen en Montessorischolen, waar didactische of levensbeschouwelijke opvattingen nog een rol spelen, de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer groter zullen zijn dan in openbare, katholieke of protestantse scholen, waar didactische of levensbeschouwelijke opvattingen veel minder een rol spelen.

Een vijfde reden voor discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen kan de concurrentieverhouding tussen scholen binnen een bepaalde gemeente of regio zijn. Sinds het einde van de inhoudelijke onderwijsverzuiling concurreren scholen met elkaar op de leerlingenmarkt (Roeleveld & Dronkers, 1994; Dijkstra, Dronkers & Hofman, 1997). Een van de argumenten in deze competitie is het succes van leerlingen bij het eindexamen. Een van de manieren om deze competitie te winnen is het schoolonderzoek hoger te becijferen dan het centraal onderzoek, waardoor er meer leerlingen slagen voor hun eindexamen. In gemeenten en provincies waar een grotere concurrentie bestaat omdat het aanbod van scholen groter is, zal daarom de discrepantie tussen het schoolonderzoek- en het centraal-examencijfer groter zijn. Gegeven de verdeling van scholen in het voortgezet onderwijs is het grootste aanbod van scholen in de provincies van de Randstad, in de grote en middelgrote steden en in de forenzengemeenten zou men kunnen verwachten dat in scholen in de randstad, in de grote en middelgrote steden en in de forenzengemeenten de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer groter zullen zijn. Een school met een laag intern rendement (veel zittenblijvers, veel uitval, lage eindexamencijfers, lage slaagpercentages) is in een dergelijke competitie tussen scholen in het nadeel. Teneinde zijn concurrentiepositie te verbeteren, kan een dergelijke school in de verleiding komen zijn schoolonderzoek hoger te becijferen waardoor het eindexamenresultaat gunstiger lijkt te worden. Dit leidt tot de mogelijkheid dat in scholen met een laag intern rendement de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer groter zullen zijn.

Een zesde reden voor discrepanties tussen de cijfers voor het schoolonderzoek en het centraal examen kan de afwijkende "cijfercultuur" binnen een school zijn. Deze afwijkende cultuur kan ontstaan in een stabiele gemeenschap van docenten met weinig doorstroom van nieuwe leerkrachten. Ook kan een afwijkende "cijfercultuur" ontstaan in een school met onvoldoende feedback procedures met betrekking tot de gevolgen van een bepaalde becijfering of met een schoolleiding die onvoldoende in staat is een afwijkende cultuur te corrigeren. Het is daarom mogelijk dat scholen verschillen in hun "cijfercultuur" en deze schoolverschillen zijn niet terug te voeren op verschillen in vakkenpakketten,

examenjaren, leerlingpopulaties, onderwijstypen, interne rendementen, concurrentieverhoudingen of didactische of levenbeschouwelijke opvattingen.

### 3. Gevolgen van discrepanties tussen de schoolonderzoek- en centraal-examencijfers.

Het is uiteraard niet voldoende aan te tonen dat er discrepanties tussen schoolonderzoek- en centraal-examencijfers bestaan. De gelijkwaardigheid van het eindexamen komt pas in gevaar als deze discrepanties tot verschillende slaagpercentages op overigens gelijke scholen zouden leiden. Dat zou betekenen dat op scholen met hogere cijfers voor het schoolonderzoek de ondergrens van het kwalificatieniveau van de afgestudeerden verschilt tussen scholen en dus niet (meer) is gegarandeerd.

### 4. Data

In deze verkenning gebruik ik de inspectiegegevens over de eindexamenresultaten voor de jaren 1996/97 en 1997/98, zoals het dagblad *Trouw* deze heeft verworven. Deze inspectiegegevens zijn beschikbaar op vestigingsniveau en voor elk vak afzonderlijk. Ook zijn ze uitgesplitst naar onderwijsniveau en voor de MAVO naar C en D niveau. In de analyses van deze bijdrage heb ik mij beperkt tot de AVO-vakken. De analyse-eenheid is dus vak per vestiging per onderwijstype per examenjaar. Tabel 1 geeft de betrokken vakken en onderwijstypen weer (N=38062). De tweede *Trouw*-publicatie over schoolprestaties, die het schooljaar 1996/97 bestreek, bevatte ook slaagpercentages, percentages allochtone leerlingen per vestiging en onderwijstype. Deze gegevens heb ik aan de inspectiegegevens gekoppeld. De weinige scholen die in het *Trouw*-bestand ontbraken (bijvoorbeeld als gevolg van fusie) zijn behandeld als missende waarden.

In de analyses spelen de volgende variabelen een rol:

- Discrepancie. Dit is het verschil per vak, vestiging en jaar tussen het gemiddeld cijfer op het centraal examen en het gemiddeld cijfer op het schoolonderzoek<sup>4</sup>. Een positief teken betekent dat het gemiddeld cijfer op het centraal examen hoger ligt dan het gemiddeld cijfer op het schoolonderzoek, een negatief teken betekent dat het gemiddeld cijfer op het centraal examen lager ligt dan het gemiddeld cijfer op het schoolonderzoek<sup>5</sup>. Tabel 1 geeft deze discrepanties voor de verschillende jaren, vakken en onderwijstypen. Gemiddeld liggen de cijfers voor het schoolonderzoek 0,1 punt hoger dan de cijfers van het centraal examen. Tabel 3 (eerste rij, vierde kolom) laat zien dat deze discrepanties bovendien tamelijk scheef verdeeld zijn: hogere schoolonderzoekcijfers komen aanzienlijk vaker voor dan hogere centraal-examencijfers.
- Het percentage allochtone leerlingen. Dit is het percentage 'cumi-leerlingen' op een vestiging en onderwijstype, op grond waarvan OC&W scholen extra subsidieert (Agerbeek, 1998). Het percentage slaat op alle leerlingen van dat onderwijstype van die vestiging en niet alleen op die in de eindexamenklas. Gezien de stabiliteit van dat percentage tussen schooljaren (Dronkers, 1998) lijkt mij dat geen probleem.
- Richting van de school, provincie, urbanisatiegraad van gemeente en de Montessori achtergrond van een school zijn ook ontleed aan Agerbeek (1998).
- De gemeente waarin de vestiging staat kan men karakteriseren op grond van de urbanisatie van die gemeente.
- De toegevoegde waarde van het intern rendement per onderwijstype per vestiging is door mij ten behoeve van de tweede *Trouw*-publicatie over schoolprestaties berekend op grond van het percentage uitvallers, het percentage zittenblijvers, het percentage zij-instromers, het percentage Beta-examens, het gemiddeld cijfer op het centraal examen, het percentage geslaagden voor het eindexamen en het percentage allochtone leerlingen per onderwijstype en per vestiging (Agerbeek, 1998). Deze inspectiegegevens zijn door de scholen als gevolg van de eerste *Trouw*-publicatie extra gecontroleerd. Deze berekende toegevoegde waarde kan beschouwd worden als een goede schatter van het intern rendement van scholen (Dronkers, 1998). Scholen die hoog scoren op deze toegevoegde waarde schaal zijn goed functionerende scholen met hoog presterende leerlingen en een goede sfeer, terwijl scholen die laag scoren op deze schaal minder goed functionerende

scholen met laag presterende leerlingen en een slechte sfeer zijn (Dronkers & Warnaar, 1999; Vink, 1999).

De multivariate analyses, waarin de discrepantie per vak, vestiging en jaar de analyse eenheid vormt, zijn zodanig gewogen dat scholen met weinig of veel vakken (ruwweg categoriale scholen versus brede scholengemeenschappen) een zelfde gewicht hebben. Deze weging is uitgevoerd om te voorkomen dat scholen met veel vakken de analyses zouden domineren.

## 5. De omvang en verklaring van discrepanties

Tabel 1 laat per vak, per examenjaar en per onderwijstype de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer zien. Allereerst valt in deze tabel op dat het cijfer van het schoolonderzoek voor alle talen, op het Turks en Arabisch na, hoger ligt dan het cijfer op het centraal examen. Hetzelfde geldt ook bij alle gamma-vakken. Bij de expressie-vakken zijn de verschillen echter het grootst. Alleen bij Wiskunde is het cijfer van het schoolonderzoek lager dan dat van het centraal examen. Ten tweede valt op dat juist op het VWO het verschil tussen het cijfer van de schoolonderzoeken en dat van het centraal examen groter is dan bij HAVO of MAVO. Tenslotte zijn de verschillen tussen het cijfer van de schoolonderzoeken en dat van het centraal examen in 1997/98 groter dan in 1996/97.

In tabel 2 worden deze verschillen tussen het schoolonderzoek- en het centraal-examencijfer met multivariate regressie<sup>6</sup> geanalyseerd.

De eerste rij laat zien dat de verschillen tussen het schoolonderzoek- en het centraal-examencijfer in 1997/98 inderdaad groter zijn dan in 1996/97: het verschil is bijna 0,2 punt. Uit de stabiliteit van de parameter in de verschillende modellen blijkt dat deze verschillen niet aan andere kenmerken toegeschreven kunnen worden. Dit systematische verschil zou verklaard kunnen worden uit het publiek worden van het intern rendement van scholen gedurende het schooljaar 1997/98, want deze publicatie zou scholen verleid kunnen hebben dat intern rendement te verhogen door het schoolonderzoek hoger te becijferen. Maar uit een vergelijking van de gemiddelde cijfers voor het schoolonderzoek en voor het centraal examen voor beide schooljaren blijkt dat de gemiddelde cijfers voor het centraal voor het schooljaar 1997/98 duidelijk lager zijn dan die voor het schooljaar 1996/97, terwijl de gemiddelde cijfers voor het schoolonderzoek in beide jaren min of meer gelijk zijn<sup>7</sup>. Scholen hebben dus niet via een andere becijfering van het schoolonderzoek hun intern rendement verhoogd. De cijfers van het centraal-examen voor het schooljaar 1997/98 liggen lager dan die van het voorafgaand jaar. Deze lagere becijfering van het centraal-examen past in een trend gedurende de 90-er jaren waarin het genormeerde gemiddelde cijfer voor het centraal-examen voortgezet onderwijs langzaam daalde (Inspectie, 1992: 17-20; Alberts, 1997: 23; Alberts, 1998: 19-20). Ook is het mogelijk dat deze lagere centraal-examencijfers in het schooljaar 1997/98 het grotere percentage eindexamenkandidaten weerspiegelt basisvorming heeft gevolgd. Van der Werf, Lubbers en Kuyper (1999: 46) laten in hun vergelijking van de prestaties van de cohorten 1989 en 1993 zien dat VWO-leerlingen die basisvorming gevolgd hadden consistent lager scoorden voor Wiskunde en tekstbegrip terwijl bij MAVO-leerlingen met basisvorming een consistente stijging optrad.

De parameters van de vakken in de modellen 2 tot 6 laat zien dat bij verschillende vakken uiteenlopende discrepanties bestaan<sup>8</sup>. Controle voor achtergrondkenmerken van scholen maakt deze discrepanties bij verschillende vakken meestal niet geringer: in tegendeel de discrepanties worden af en toe nog groter bij controle voor achtergrondkenmerken. Zo groeit de discrepantie bij Grieks van -0,15 naar -0,25 en bij Turks van 0,82 naar 1,47. Dit voorbeeld laat ook zien dat het niet vanzelfsprekend is dat bij talen het schoolonderzoek hoger becijferd worden dan bij het centraal examen. Het is niet goed te begrijpen waarom eindexamenkandidaten schriftelijk in het Turks of Arabisch vaardiger zouden zijn dan bij een mondeling overhoring. Mogelijk speelt de status van het vak hier een grote rol: door strenger te cijferen dan het centraal examen zouden leraren Turks en Arabisch duidelijk willen maken dat hun nog-niet-gevestigde vakken serieus genomen moeten worden. De status van de andere vreemde talen vakken is blijkbaar gevestigd genoeg dat deze discrepanties niet bedreigend zijn voor die status: het gemiddeld cijfer bij schoolonderzoek is hoger dan bij het centraal schriftelijk. Het verschil tussen het schoolonderzoek- en het centraal-examencijfer

bij de expressievakken (en met name Tekenen en Handvaardigheid) is nog groter<sup>9</sup>. De discrepanties tussen het schoolonderzoek- en het centraal-examencijfer bij de verschillende vakken wekken de indruk van ongelijkwaardige becijfering: bij het schoolonderzoek wordt soepeler gerekend dan bij het centraal examen. Op zich behoeft een discrepantie tussen het schoolonderzoek- en het centraal-examencijfer bij een bepaald vak geen teken van ongelijkwaardige becijfering: beide delen van het eindexamen beslaan vaak verschillende delen van de vereiste kennis en vaardigheden en die een verschillende zwaarte kunnen hebben. Het is echter opvallend dat dit mogelijke verschil bijna steeds uitvalt ten gunste van een hogere becijfering van het schoolonderzoek. Volgens de toevalsregels bij de verdeling van de stof over beide examenonderdelen zou het aantal vakken met een hoger schoolonderzoekcijfer even groot zijn dan het aantal vakken met een hoger centraal-examencijfer. Het is bij een dergelijke verdeling niet logisch dat het gemakkelijkere deel steeds aan het schoolonderzoek zou zijn toegewezen. Ook de wijze van tentamineren van de stof verschilt tussen het schoolonderzoek en het centraal examen: uitgesmeerd over een heel jaar versus geconcentreerd in een relatief korte periode. Het zou gemakkelijker zijn een hoger cijfer in een uitgesmeerde periode te halen dan in een geconcentreerde periode. Het zou echter merkwaardig zijn als een dergelijk vormkwesatie tot zulke drastische discrepanties tussen het gemiddeld cijfer van de schoolonderzoeken en dat van het centraal examen zou leiden, die bovendien per vak nog eens erg verschillen.

Schoolonderzoeken op het VWO worden hoger becijferd in vergelijking met het centraal examen dan op de HAVO en de MAVO-C. Ook deze hogere becijfering kan niet verklaard worden door achtergrondkenmerken, want de parameters worden eerder groter dan kleiner in de opeenvolgende modellen. Een groot verschil tussen het schoolonderzoek- en het centraal-examencijfer is dus kenmerkender voor het meest prestigieuze onderwijstype dan voor de onderkant van het onderwijsgebouw. Op een gemiddelde VWO vestiging is de discrepantie tussen het schoolonderzoek- en het centraal-examencijfer ongeveer 0,2 punt groter dan op de gemiddelde MAVO-C vestiging. Een mogelijke verklaring voor dit grote verschil in discrepanties zou kunnen zijn dat voor het VWO-schoolonderzoek veel zwaardere kennis en vaardigheden worden gevraagd dan voor het VWO-centraal examen, terwijl bij de MAVO-C de zwaarte van kennis en vaardigheden bij schoolonderzoek en centraal examen min of meer gelijk is.

De toe nu toe besproken effecten van vak- en schoolkenmerken op de discrepanties tussen het schoolonderzoek- en het centraal-examencijfers vormen geen bedreiging van de gelijkwaardigheid van het eindexamen van het voortgezet onderwijs. Alle leerlingen binnen een bepaald onderwijstype hebben immers op gelijke wijze last van deze discrepanties waardoor het kwalificatieniveau van leerlingen met een zelfde eindexamenpakket gegarandeerd is. In de modellen 3 tot 6 worden school- en regiokenmerken aan de vergelijking toegevoegd. Deze kenmerken, als zij delen van de variatie in de discrepanties kunnen verklaren, vormen wel een bedreiging van de gelijkwaardigheid van het eindexamen.

Model 3 laat zien dat het percentage allochtone leerlingen op een vestiging de discrepantie tussen schoolonderzoek- en centraal-examencijfer vergroot. Bij 10% meer allochtone leerlingen neemt de discrepantie met 0,1 punt toe. Rohde, Kuyt, Takkenberg en Janssen (1999: 8) en Tesser, Merens & van Praag (1999: 182-184) rapporteren over bijna dezelfde data een soortgelijk verschil: name bij het centraal examen scoren allochtone leerlingen lager dan autochtone leerlingen, terwijl beide groepen op het schoolonderzoek min of meer gelijk scoren. De parameters van model 3 laten zien dat dit verschil niet komt door een verschil in vakkenpakket of onderwijstype tussen allochtone en autochtone leerlingen: ze verschillen niet sterk met die van model 2. Ook verhoogt de toevoeging van de variabele ‘% allochtone leerlingen’ de variantie substantieel: 4%. De parameters van de modellen 4 tot 6 laten tenslotte zien dat de betekenis van het percentage allochtone leerlingen niet verklaard kan worden uit de richting van de school, de regio, de tegenstelling stad-platteland of het intern rendement van de school.

In model 4 voegen wij de richting van de school, de provincie en de urbanisatiegraad van de gemeente waarin de school gevestigd is toe aan de vergelijking. Het gaat hier om school- en regiokenmerken, die geen significant effect op de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer behoren te hebben. Als dat wel het geval is, betekent het immers dat het eindexamen in de ene provincie niet gelijkwaardig is aan dat van in andere of dat het eindexamen in de

ene onderwijsrichting meer vraagt dan in een andere. De parameters van model 4 laten zien dat dit het geval is.

Op algemeen-bijzondere, gereformeerde, reformatorische en montessori scholen worden de schoolonderzoeken in vergelijking met het centraal examen hoger becijferd dan op overeenkomstige protestantse, katholieke en samenwerkingscholen, terwijl op vergelijkbare openbare scholen die schoolonderzoeken juist lager worden becijferd. Scholen van verschillende richtingen verschillen echter ook in hun intern rendement. Daarmee wordt in model 5 daarmee rekening gehouden. Dan blijkt dat montessorischolen op een zelfde wijze cijferen als vergelijkbare protestantse scholen met een zelfde intern rendement. Houdt men ook bij katholieke scholen rekening met hun hoger intern rendement, dan becijferen zij hun schoolonderzoek ook hoger dan vergelijkbare protestantse scholen. De verschillen in discrepanties tussen de meeste onderwijsrichtingen zijn echter klein, behalve die bij de gereformeerde en reformatorische scholen. De discrepanties van scholen van deze twee richtingen zijn meer dan 0,1 groter dan in vergelijkbare protestante scholen.

Ook verschillen de discrepanties per provincie, waarbij de verschillen dus niet geweten kunnen worden aan verschillen tussen stad en platteland, verschillen in percentages allochtonen of in godsdienstige verschillen. Daarvoor is immers al gecontroleerd. Met name scholen in de provincies Friesland, Noord-Holland, Flevoland, Groningen, Utrecht en Zuid-Holland becijferen hun schoolonderzoek hoger in vergelijking met het centraal examen dan vergelijkbare scholen in andere provincies, terwijl scholen in de provincie Zeeland hun schoolonderzoek juist lager becijferen dan vergelijkbare scholen in andere provincies. Een deel van deze verschillen is terug te voeren op het lagere interne rendement van scholen in de eerstgenoemde provincies, maar ook als daarmee rekening wordt gehouden blijven er duidelijke verschillen bestaan, met name bij scholen in Friesland en Zeeland. In verhouding tot de verschillen in discrepanties tussen scholen uit verschillende provincies zijn de verschillen in discrepanties tussen scholen uit verschillende typen gemeenten gering.

Een andere verklaring van grotere discrepanties tussen de becijfering van schoolonderzoek en centraal examen zou zijn dat scholen op deze wijze hun concurrentiepositie in hun regio proberen te verbeteren: er zouden daardoor meer leerlingen slagen. Juist scholen met een gering intern rendement (veel zittenblijvers en uitvallers, veel gezakten voor het eindexamen, lage cijfers op centraal examen) zouden verleid kunnen worden op een hogere becijfering voor hun schoolonderzoek. Model 5 laat zien dat deze verklaring correct is. Scholen met een hoog intern rendement (gecorrigeerd voor leerlingpopulatie en zij-instroom) hebben een kleinere discrepantie tussen de becijfering van schoolonderzoek en centraal examen dan scholen met een laag intern rendement. Het verschil tussen een 0 of één – in de genoemde *Trouw*-publicatie over het intern rendement vergroot de discrepantie met 0,09. Het verschil in discrepantie tussen de becijfering van schoolonderzoek en centraal examen is tussen scholen met een hoog intern rendement (++) en een laag intern rendement (--) 0,36 punten. Voor de goede orde zij er op gewezen dat dit resultaat niet verklaard kan worden door achtergrondkenmerken van scholen, omdat daarvoor zowel in het *Trouw*-cijfer als in mijn analyse wordt gecontroleerd<sup>10</sup>.

## 6. Verschillende discrepanties op verschillende scholen?

In de eerste vijf multivariate analyses van tabel 2 ligt het accent vooral op de algemene verklaringen van discrepanties: schooljaren, vakken, onderwijstypen, percentage allochtone leerlingen, richting, provincie, urbanisatiegraad en intern rendement. Deze algemene kenmerken van scholen en hun omgeving blijken in een aantal gevallen bij te dragen tot grotere of kleinere discrepanties tussen het schoolonderzoek- en het centraal-examencijfer per vak, vestiging en jaar. Maar dat behoeft strikt genomen nog niet te betekenen dat een school of vestiging in zijn geheel grotere of kleinere discrepanties kent dan andere scholen. In deze paragraaf ga ik nader in op verschillen in discrepanties tussen scholen of vestigingen als geheel.

Om de discrepantie per afzonderlijke school of vestiging vast te stellen heb ik de discrepanties per vak en jaar geaggregeerd naar vestiging en op dat niveau gemiddeld. Dit gemiddelde geeft aan hoe groot de gemiddelde discrepantie per school is over alle vakken (ongeacht het onderwijstype) waarover in beide schooljaren door de school examens zijn afgenomen<sup>11</sup>. In de eerste cel van de eerste

rij van tabel 3 staat het gemiddelde van die geaggregeerde discrepanties:  $-12$ . Dat betekent dat het gemiddelde verschil tussen de cijfers op het schoolonderzoek en het centraal examen per school  $.12$  punt is. Daarbij moet opgemerkt worden dat de verschillen in discrepanties binnen scholen aanzienlijk groter zijn dan de verschillen in discrepanties tussen scholen (binnen-vestiging standaarddeviatie  $.54$  versus tussen-vestiging standaarddeviatie  $.22$ ). Dat laat zien dat verschillen binnen scholen zijn gewichtiger dan verschillen tussen scholen. De scheefheid (vierde kolom) laat zien dat de gemiddelde discrepanties per school niet normaal verdeeld zijn: de discrepanties waarbij het schoolonderzoek hoger becijferd wordt dan het centraal examen zijn groter dan de discrepanties waarbij het schoolonderzoek lager becijferd wordt dan het centraal examen. De maximale en minimale scores laten dat nog eens zien: het grootste verschil tussen schoolonderzoek en centraal examen waarbij de laatstgenoemde hoger is dan de eerste genoemde is  $.49$ , terwijl het grootste verschil tussen schoolonderzoek en centraal examen waarbij de laatstgenoemde lager is dan het eerstgenoemde is  $1.80$ . Dus het aantal scholen dat gemiddeld hogere cijfers op het schoolonderzoek heeft dan op het centraal examen is dus aanzienlijk groter dan het aantal scholen met gemiddeld lagere cijfers op het schoolonderzoek dan op het centraal examen.

Op grond van de voorafgaande analyses mogen wij aannemen dat deze verschillen in gemiddelde discrepanties in becijfering van schoolonderzoek en centraal examen tussen scholen gedeeltelijk veroorzaakt worden door school- en regiokenmerken. Daarnaast kunnen echter afzonderlijke scholen ongeacht deze school- en regiokenmerken een eigen “cijfercultuur” kennen, hetzij door een bepaalde traditie hetzij door bewust beleid. Teneinde deze eigen “cijfercultuur” van een school te kunnen vaststellen, gebruiken wij de residuen uit de regressievergelijkingen van tabel 2. Residu betekent in dit verband dat gedeelte van de discrepantie per vak, vestiging en jaar dat niet verklaard kan worden door de onafhankelijke variabele(n) in de betrokken vergelijking. Door vervolgens deze residuen te aggregeren naar vestiging, weten wij welk deel van de gemiddelde discrepanties per vestiging niet verklaard worden door de onafhankelijke variabelen. Deze geaggregeerde residuen zijn weergegeven in de tweede regel van tabel 3. Het gaat hier om de geaggregeerde residuen uit model 1, dat wil zeggen de residuen nadat gecontroleerd is voor schooljaar. Per definitie is het gemiddelde nu  $0$ , maar de binnen-vestiging en tussen-vestiging standaarddeviatie is door deze controle voor schooljaar nauwelijks veranderd (vergelijk regel 1 met regel 2). De scheefheid is door deze controle wel toegenomen (van  $1.28$  naar  $1.43$ ). Maar de zeer hoge correlatie tussen de geaggregeerde residuen en geaggregeerde discrepanties ( $.99$ ) laat zien dat deze controle voor schooljaar zo goed als niets aan de verschillen tussen scholen op het punt van hun gemiddelde discrepanties heeft veranderd.

In model 2 controleer ik ook voor vakkenpakket en onderwijstype. De binnen-vestiging standaarddeviatie wordt door deze controle wat kleiner, maar verder verandert er niet veel in de verschillen tussen scholen op het punt van hun gemiddelde discrepanties. De hoge correlatie tussen de residuen uit model 1 en 2 ( $.99$ ) laat dat zien, evenals de minimale verschillen tussen de scholen met extreme discrepanties voor en na controle voor vakkenpakket en onderwijstype.

In model 3 controleer ik ook voor percentage allochtone leerlingen. De tussenvestiging standaarddeviatie neemt daardoor wat af en de scheefheid in de verdeling van gemiddelde residuen wordt duidelijk geringer. Het percentage allochtone leerlingen op een school lijkt dus vooral verantwoordelijk voor de relatief hoge gemiddelde schoolonderzoekcijfers. Maar de hoge samenhang tussen de geaggregeerde residuen uit model 2 en 3 ( $.84$ ) laat zien dat de eigen “cijfercultuur” van een school in hoofdzaak niet verklaard kan worden uit het percentage allochtone leerlingen. De discrepanties op schoolniveau die in model 3 zijn overgebleven zijn in strijd met de gelijkwaardigheid van het eindexamen, omdat ze niet meer verklaard kunnen worden door wettige verschillen tussen scholen.

In model 4 controleer ik ook voor de richting van de school, de provincie en de urbanisatiegraad van de gemeente. Deze controle vermindert de scheefheid in de verdeling van de gemiddelde residuen nog wat, maar er verandert weinig in de verschillen tussen scholen op het punt van hun gemiddelde discrepanties.

In model 5 controleer ik tenslotte ook voor het intern rendement van scholen, om de betekenis van de eigen “cijfercultuur” vast te stellen die niet verklaard kan worden door enig school- of regiokenmerk. Deze controle vermindert de scheefheid in de verdeling van de gemiddelde residuen verder, maar ook na deze controle blijft de verdeling flink scheef te gunst van hoger



schoolonderzoekcijfers in vergelijking met centraal-examencijfers. Deze scheefheid onderstreept dat de verdeling van de discrepanties geen toevallig proces is, maar veeleer de resultante van keuzen van scholen en leerkrachten bij de becijfering van het schoolonderzoek. De omvang van de tussenvestiging standaarddeviatie van de residuen laat echter zien dat deze vijf extreme scholen slechts een topje van de ijsberg vormen. Bij 10.2% van alle scholen wijkt de eigen “cijfercultuur” zo sterk af dat het gemiddelde schoolonderzoekcijfer op die scholen ten minste 0,2 punt hoger ligt dan het gemiddelde centraal-examencijfer op die scholen, terwijl bij een andere 9,3% van de scholen het gemiddelde centraal-examencijfer ten minste 0,2 hoger is dan het gemiddelde schoolonderzoekcijfer. Bij 20% van de scholen is de eigen “cijfercultuur” zo sterk dat de becijfering van het schoolonderzoek in vergelijking met het centraal schriftelijk examencijfer meer dan een 0,4 verschilt.

Hoe belangrijk is nu deze eigen “cijfercultuur” van een school ter verklaring van de discrepanties tussen het schoolonderzoek- en het centraal-examencijfer per vak, vestiging en jaar? Als die eigen cijfercultuur weinig verklaart van die discrepanties per vak, vestiging en jaar, dan hebben de geaggregeerde residuen uit model 5 het karakter van toevallig toegewezen getallen, die weinig van de variantie kunnen verklaren. In model 6 van tabel 2 voegen wij de geaggregeerde residuen uit model 5, die de mate van eigen “cijfercultuur” weergeven, toe aan de vergelijking<sup>12</sup>. Tabel 2 laat duidelijk zien dat die eigen “cijfercultuur” een zelfstandig effect heeft op de discrepanties tussen schoolonderzoek- en centraal-examencijfers per vak, vestiging en jaar. Hoe groter de discrepantie tussen gemiddeld schoolonderzoek- en centraal-examencijfers op een school hoe groter ook de verschillen tussen schoolonderzoek- en centraal-examencijfers per vak, vestiging en jaar. Deze eigen “cijfercultuur” van een school verklaart nog eens 8% extra variantie en gezien de stabiliteit van de overige parameters staat zij los van de school- en regiokenmerken. Vanuit dit resultaat kan men concluderen dat scholen meer zijn dan een toevallige samenloop van leerkrachten en leerlingen. Tegelijkertijd moet men ook concluderen dat het overgrote deel van de verschillen tussen schoolonderzoek- en het centraal-examencijfer per vak, vestiging en jaar (69%) niet verklaard kan worden door schooljaar, vakkenpakketten, school- of regiokenmerken en een eigen “cijfercultuur” van hun school. Zo groot is blijkaar de autonomie van de afzonderlijke docenten, niet alleen ten opzichte van hun school of regio maar ook ten opzicht van hun eigen vak, dat het overgrote deel van de discrepanties tussen schoolonderzoek- en het centraal-examencijfer per vak, vestiging en jaar niet verklaard kan worden. Voor diegenen die menen dat deze lage verklaarde variantie veroorzaakt wordt door de onvermijdelijk ruis die elke beoordeling met zich meebrengt, wijs ik er op dat deze slecht verklaarbare discrepanties al geaggregeerde gegevens zijn, die minder ruis behoren te bevatten.

## **7. Discrepanties tussen schoolonderzoek- en het centraal-examencijfer en de kans van slagen voor het eindexamen.**

Hoewel uit de vorige paragraaf duidelijk is geworden dat scholen stevig verschillen in hun becijfering van het schoolonderzoek, zowel door school- en regiokenmerken als door een eigen “cijfercultuur”, vloeit daaruit nog niet automatisch voort dat daardoor de gelijkwaardigheid van het eindexamen wordt ondermijnd. Grote discrepanties tussen schoolonderzoek- en het centraal-examencijfer kunnen zodanig samenhangen met andere school- en regiokenmerken dat het feitelijk effect van de omvang van discrepanties per school op de slaagkans per school niet significant is. Daarom onderzoek ik in deze paragraaf of de omvang van discrepanties tussen schoolonderzoek- en het centraal-examencijfer per school de slaagpercentages per school verhoogt, ook als men controleert voor andere school- en regiokenmerken.

Omdat ik alleen over slaagpercentages per vestiging en schooltype uit het jaar 1996/97 beschik, gebruik ik in deze analyse alleen de gegevens die op dat jaar betrekking hebben. De discrepantie tussen schoolonderzoek- en het centraal-examencijfer per vak, vestiging en jaar heb ik naar vestiging en schooltype geaggregeerd, evenals het gemiddeld cijfer voor het centraal examen. Het gaat hierboven dien om de ongecorrigeerde geaggregeerde discrepanties, zoals weergegeven in de eerste rij van tabel 3.

In model 1 van tabel 4 bereken ik het effect van de omvang van discrepanties per vestiging op de slaagpercentages per vestiging. De positieve parameter laat zien dat hoe kleiner het verschil tussen

schoolonderzoek- en het centraal-examencijfer, des te hoger het slaagpercentage. Dit is in lijn met een resultaat uit de vorige paragraaf, waaruit bleek dat scholen met een hoog intern rendement (en dus ook hoge slaagpercentages) kleinere discrepanties tussen schoolonderzoek- en het centraal-examencijfer hadden. In model 2 voeg ik het gemiddeld cijfer voor het centraal examen toe aan de vergelijking: de verklaarde variantie stijgt tot 50% en het effect van de gemiddelde discrepantie per school verandert van richting. Een hogere gemiddelde discrepantie tussen schoolonderzoek- en het centraal-examencijfer verhoogt het slaagpercentage per school. Gegeven het maximale verschil in gemiddelde discrepantie tussen scholen van 2.29 bedraagt het verschil in slaagpercentage van vergelijkbare scholen met de meest uiteenlopende discrepanties 11,7% ( $5.13 \cdot 2.29$ ). Indien ik ook nog rekening houd met andere school- en regiokenmerken die het slaagpercentage per vestiging beïnvloeden, wordt dit verschil in slaagpercentage tussen vergelijkbare scholen met de meest uiteenlopende discrepanties nog groter: 21.9% ( $9.56 \cdot 2.29$ ). Ook als wij deze extreme scholen buiten beschouwing laten zijn de verschillen in slaagpercentage tussen de 196 scholen<sup>13</sup> die ten minste 2 standaarddeviaties in discrepantie van elkaar verschillen (.44) aanzienlijk: 4.2%. Gegeven dat de standaarddeviatie van slaagpercentages van alle AVO scholen 8.2% bedraagt, zijn deze verschillen als gevolg van een afwijkende becijfering van het schoolonderzoek voor scholen, ouders en leerlingen groot. Want dit significante effect van de discrepantie tussen schoolonderzoek en centraal examen op het slaagpercentage betekent dat ruim 4% meer leerlingen op ene school slagen dan op een andere vergelijkbare school uitsluitend door de hogere becijfering van het schoolonderzoek. Dat lijkt een positieve formulering, maar op grond van dezelfde resultaten zou men ook kunnen concluderen dat 4% van de leerlingen op scholen met hoge schoolonderzoekcijfers in verhouding tot hun centraal-examencijfers ten onrechte hun diploma hebben gekregen. Wat ook de juiste conclusie moge zijn, in beide gevallen zijn de einddiploma's niet gelijkwaardig tussen scholen door de uiteenlopende becijfering van het schoolonderzoek.

## 8. Conclusie en consequenties

De eerste conclusie uit de hier gepresenteerde analyse is dat bij verschillende vakken uiteenlopende tradities bestaan, die er in de meeste gevallen toe leiden dat het schoolonderzoek hoger becijferd wordt dan het centraal examen. De gebruikelijke verklaring voor deze traditie, dat in de twee onderdelen van het eindexamen verschillende stof wordt getoetst, snijdt geen hout. Het is niet logisch dat de stof van het schoolonderzoek voor de meeste vakken gemakkelijker is dan de stof van het centraal examen. De scheve verdeling van de discrepanties in becijfering die bij alle vakken wordt gevonden wijst er ook op dat de hogere becijfering niet het resultaat is van toevallige zwaardere of makkelijkere stof in de beide delen van het eindexamen. De afwijkende becijfering van Turks en Arabisch onderstreept nogmaals dat het niet noodzakelijk is dat de becijfering van het schoolonderzoek bij de talen steeds hoger uitvalt. Ook vormverschillen tussen schoolonderzoeken en centraal examen (geconcentreerd versus uitgesmeerd) bieden geen logische verklaring voor de hogere becijfering bij het schoolonderzoek en de scheefheid van de discrepanties. Gezien vanuit de gelijkwaardigheid van het eindexamen tussen scholen zijn deze verschillen tussen vakken niet ernstig zolang scholen onderling maar niet te erg kunnen verschillen in aangeboden vakken.

De tweede conclusie is dat becijferingen niet constant zijn over de tijd. In de twee hier onderzochte jaren verschilt niet de becijfering van het schoolonderzoek, maar varieert de becijfering van het centraal examen. De meest plausibele verklaring lijkt mij dat het centraal examen in 1997/98 voor het geheel zwaarder becijferd is dan in het voorafgaande jaar. Gezien vanuit de gelijkwaardigheid van het eindexamen tussen scholen zijn deze verschillen tussen jaren niet ernstig.

De derde conclusie is dat de discrepantie tussen schoolonderzoek- en centraal-examencijfer het grootst in het VWO en het kleinst MAVO-C. Gezien vanuit de gelijkwaardigheid van het eindexamen tussen scholen zijn deze verschillen tussen onderwijstypen niet ernstig.

De vierde conclusie is dat de schoolonderzoekcijfers op vergelijkbare scholen met veel allochtone leerlingen hoger zijn dan die op scholen met weinig allochtone leerlingen. Uitsluitend door deze hogere becijfering is het slagingspercentage van een school met 10% meer allochtone leerlingen (en dus met 0,1 grotere discrepantie) bijna 1% ( $.1 \cdot 9.56$ ) hoger dan op een vergelijkbare school met minder allochtone leerlingen. Maar tegelijkertijd zakt op scholen met veel allochtone leerlingen een hoger

percentage voor het eindexamen, zoals tabel 4 ook laat zien: 10% meer allochtone leerlingen laat het slaagpercentage van een school met 1,3% ( $10 \cdot 13$ ) dalen. Dit leidt tot de volgende paradox; weliswaar is de becijfering van het schoolonderzoek op scholen met veel allochtone leerlingen hoger waardoor meer leerlingen slagen, maar het grote percentage leerlingen verkleint tegelijkertijd het slaagpercentage van scholen. De optelsom van deze twee tegenstrijdige samenhangen maakt dat totaal het slaagpercentage op vergelijkbare scholen met veel allochtone leerlingen toch daalt (0,4% per 10% allochtone leerlingen). De hogere becijfering van het schoolonderzoek lijkt verwant aan het overadviseren van allochtone leerlingen aan het eind van de basisschool: leerkrachten beoordelen het 'werkelijke' potentieel van migrantenleerlingen hoger. Er bestaan geen data over succes in vervolgopleidingen met betrouwbare gegevens over schoolonderzoek- en centraal-examencijfers. Ik kan dus niet vaststellen of dat dit hoger beoordelen van het schoolonderzoek van allochtone leerlingen terecht of onterecht is.

De vijfde conclusie is dat vergelijkbare scholen van verschillende onderwijsrichtingen het schoolonderzoek verschillend becijferen. Met name gereformeerde en reformatorische scholen geven duidelijk hogere schoolonderzoekcijfers dan vergelijkbare scholen van andere richtingen. Deze afwijkende positie kan niet gemakkelijk verklaard worden uit een groter belang dat deze gereformeerde en reformatorische scholen zouden hechten aan het schoolonderzoek. Wel betekent het dat uitsluitend door deze hogere schoolonderzoekcijfers op deze gereformeerde en reformatorische scholen meer leerlingen slagen (1,3%; 2,1%) dan op vergelijkbare andere scholen. Verklaren deze hogere slaagpercentages niet een deel van de aantrekkingskracht van deze gereformeerde en reformatorische scholen in een verder snel seculariserende Nederlandse samenleving? Deze hogere becijfering van het schoolonderzoek is duidelijk in strijd met de gelijkwaardigheid van eindexamens in Nederland.

De zesde conclusie is dat vergelijkbare scholen in bepaalde provincies het schoolonderzoek duidelijk hoger becijferen dan in andere provincies. Het verschil in discrepantie tussen de twee meest extreme provincies (Friesland, Zeeland) bedraagt 0.16 punt. Dit verschil kan niet verklaard worden uit verschillen in urbanisatiegraad van provincies of uit verschillen in intern rendement van scholen. Dat betekent dat op vergelijkbare scholen in Friesland 1,5% meer leerlingen slagen dan op vergelijkbare scholen in Zeeland, uitsluitend dankzij hun hogere becijfering van het schoolonderzoek. Deze hogere becijfering van het schoolonderzoek is duidelijk in strijd met de gelijkwaardigheid van eindexamens in Nederland.

De zevende conclusie is dat scholen met een lager intern rendement hun schoolonderzoek hoger becijferen dan vergelijkbare scholen met een hoger intern rendement. De meest voor de hand liggende verklaring van dit verschijnsel is dat scholen met een laag intern rendement door deze hogere becijfering van hun school onderzoek toch nog tot redelijke slagingspercentages komen. Gegeven dat de discrepantie tussen schoolonderzoek- en centraal-examencijfer in scholen met een hoog intern rendement (2 standaarddeviaties boven het gemiddeld rendement) 0,36 punten kleiner is dan in vergelijkbare scholen met een laag intern rendement (2 standaarddeviaties onder het gemiddeld rendement), slagen er op laatstgenoemde scholen 3,4% meer leerlingen, uitsluitend dankzij hun hogere becijfering van het schoolonderzoek. Dit verschil van bijna een halve standaarddeviatie als gevolg van een hogere becijfering van het schoolonderzoek is niet onbelangrijk. Deze hogere becijfering van het schoolonderzoek is duidelijk in strijd met de gelijkwaardigheid van eindexamens in Nederland.

De achtste conclusie is dat vergelijkbare scholen sterk kunnen verschillen in hun "cijfercultuur". Deze verschillende "cijfercultuur" is niet herleidbaar tot de beschikbare school- en regiokenmerken, maar lijkt alleen nog verklaard te kunnen worden door historische toevalligheden en unieke omstandigheden. Deze eigen "cijfercultuur" is zo krachtig dat de becijfering van het schoolonderzoek bij 20% van de scholen meer dan 0,4 punten verschilt. Dat betekent dat deze eigen "cijfercultuur" verantwoordelijk is voor een verschil in slaagpercentage van 3.8% tussen overigens volledig vergelijkbare scholen. Dit verschil van bijna een halve standaarddeviatie als gevolg van een hogere becijfering van het schoolonderzoek is duidelijk in strijd met de gelijkwaardigheid van eindexamens in Nederland.

De negende en laatste conclusie is dat het overgrote deel van de verschillen in discrepanties tussen schoolonderzoek- en centraal-examencijfers per vak, vestiging en jaar (69%) niet verklaard kunnen worden uit het vak, uit de school- en regiokenmerken of uit een eigen "cijfercultuur" van scholen. Gegeven dat het hier om geaggregeerde gegevens gaat is, dit onverklaarbare deel groot. Dit

wijst op een grote autonomie van de afzonderlijke docenten, niet alleen ten opzichte van hun school en regio, maar –ernstiger- ook ten opzichte van hun vak en professie. Deze per docent wisselende becijfering van het schoolonderzoek is echter niet in strijd met de wettelijk gegarandeerde gelijkwaardigheid van eindexamens tussen scholen, want zij komt op alle scholen in even grote mate voor en is niet specifiek per school.

De autonomie van scholen bij de becijfering van het schoolonderzoek kan de gelijkwaardigheid van het eindexamen aantasten, want op de ene school slagen meer of andere leerlingen dan op een andere vergelijkbare school, uitsluitend dankzij de hogere becijfering van het schoolonderzoek op de eerste genoemde school. Dat betekent dat het eindexamen niet een gelijke ondergrens van het kwalificatieniveau van afgestudeerden in het voortgezet onderwijs garandeert. Met name de verschillen in becijfering tussen onderwijsrichtingen, tussen scholen in verschillende provincies, tussen scholen met een verschillend intern rendement en tussen scholen met een eigen “cijfercultuur” ondermijnen die gelijkwaardigheid en zijn zo in strijd met het Nederlandse onderwijsbestel. Immers op een denkbeeldige reformatorische school in Friesland met een laag intern rendement en een “cijfercultuur” met een grote discrepantie tussen schoolonderzoek- en centraal-examencijfers slagen 10,8% meer leerlingen dan op een even denkbeeldige openbare school in Zeeland met een hoog intern rendement en een “cijfercultuur” met een kleine discrepantie tussen schoolonderzoek- en centraal-examencijfers. Ook tussen minder extreme scholen zijn deze verschillen nog substantieel: op scholen met een laag intern rendement en een “cijfercultuur” met een grote discrepantie tussen schoolonderzoek- en centraal-examencijfers slagen 7,2% meer leerlingen dan op de overigens vergelijkbare scholen met een hoog intern rendement en een “cijfercultuur” met een geringe discrepantie tussen schoolonderzoek- en centraal-examencijfers. Op de eerstgenoemde scholen hebben afgestudeerden een ander kwalificatieniveau dan op de laatstgenoemde scholen.

Deze uiteenlopende discrepanties tussen schoolonderzoek- en centraal-examencijfers tussen scholen zouden een goede verklaring kunnen zijn voor de matige verklarende kracht van het gemiddeld eindexamencijfer voor succes in het hoger onderwijs. Door deze discrepanties, die geen verschillen in kwalificatieniveaus van leerlingen weerspiegelen, wordt een extra, niet-relevante variantie in de eindexamencijfers geïntroduceerd, die hun verklarende kracht kan aantasten. Indien dit juist is, zou dat er voor pleiten bij een eventuele selectie bij de ingang van het hoger onderwijs alleen rekening te houden met de cijfers voor het centraal examen.

Deze uiteenlopende discrepanties betekent ook dat leerlingen op scholen met een “cijfercultuur” met een grote discrepantie tussen schoolonderzoek- en centraal-examencijfers hogere gemiddelde eindexamencijfer hebben, waardoor zij een grotere kans hebben een plaats te verwerven bij een studierichting met een numerus fixus.

Wat betekent nu deze ongelijkwaardigheid van het eindexamen voor de toekomstige ontwikkeling van het Nederlandse onderwijs? Indien de autonomie van scholen in de toekomst verder vergroot zal worden en de financiering van het voortgezet onderwijs op basis van outputcriteria zal plaatsvinden, kunnen scholen in de verleiding komen de prestaties van hun leerlingen nog eigenzinniger te oordelen en die beoordeling nog meer aan te passen aan hun leerlingpopulatie, de gebruiken in de regio en het functioneren van hun school. Verschillen tussen scholen in hun becijfering van het schoolonderzoek kunnen in dat geval nog verder groeien, waardoor de gelijkwaardigheid van het eindexamen verder ondermijnd wordt en de ondergrens in kwalificatieniveau minder gegarandeerd kan worden (zie ook Slagter, 1999). Het afgelopen decennium is die autonomie van scholen al toegenomen. Helaas ontbreken mij de gegevens om na te gaan of gedurende dit afgelopen decennium de discrepantie tussen schoolonderzoek- en centraal-examencijfers is toegenomen. Maar indien scholen in dit afgelopen decennium meer autonoom zijn geworden, zou dat de langzame daling in het genormeerde gemiddelde cijfer voor het centraal-examen gedurende de 90-er jaren (Inspectie, 1992: 17-20; Alberts, 1997: 23; Alberts, 1998: 19-20) kunnen verklaren. Scholen hoeven zich in het geval van grotere autonomie minder aan te trekken van landelijke criteria. Ook zullen in die omstandigheden centrale kwaliteitscontroleurs (in dit geval de inspectie) minder gemakkelijk scholen corrigeren. Dat kan betekenen dat hetgeen op scholen feitelijk wordt geleerd en beoordeeld steeds meer kan gaan afwijken van hetgeen op het centraal-examen getoetst wordt, waardoor de scores op het centraal-examen dalen bij gelijkblijvende cijfers voor de schoolonderzoeken. Deze steeds grotere afwijking hoeft niet

automatisch te betekenen dat het kwalificatieniveau daalt. De autonomie van scholen kan het aantal verschillende vaardigheden, die onderwezen en in het schoolonderzoek beoordeeld worden, vergroot hebben, terwijl het centraal-examen met deze gegroeide diversiteit geen rekening kan houden (de Glopper & Van Schooten, 1998). Indien deze verklaring van de langzame daling in het genormeerde gemiddelde cijfer voor het centraal-examen gedurende de 90-er jaren correct is, dan zou ook een groeiende discrepantie tussen schoolonderzoek- en centraal-examencijfers gedurende dit decennium gevonden moeten worden. In dat geval heeft de grotere autonomie van scholen al in het verleden bijgedragen tot een afname in de gelijkwaardigheid van het eindexamen.

**Tabel 1: Discrepanties per vestiging tussen het gemiddeld cijfer op het centraal examen (ce) en het gemiddeld cijfer op het schoolonderzoek (so), uitgesplitst naar vak (eventueel per niveau), per opleiding en voor de schooljaren '96/'97 en '97/'98 (positieve discrepantie betekent ce > so).**

	MAVO C		MAVO D		HAVO		VWO		Totaal
	96/97	97/98	96/97	97/98	96/97	97/98	96/97	97/98	
Nederlands	.35	.21	-.22	-.16	-.36	-.33	-.25	-.08	-.20
Fries	-	-	-.90	.00	-	-	-	-	-.30
Latijn	-	-	-	-	-	-	-.42	-.19	-.30
Grieks	-	-	-	-	-	-	-.10	-.45	-.27
Frans	-.29	-.27	-.56	-.64	.04	-.14	-.39	-.59	-.36
Duits	-.25	-.17	-.52	-.58	-.31	-.41	-.02	-.47	-.38
Engels	.11	-.18	.07	.07	.10	.00	-.51	-.56	-.10
Spaans	-	-	-	-.50	-.73	-.63	-.58	-.72	-.66
Turks	-	-	.70	.45	-	-	-	-	.58
Arabisch	-	-	.40	.30	-	-	-	-	.35
Geschiedenis	-.22	-.41	-.21	-.23	-.24	-.18	.09	-.18	-.16
Aardrijkskunde	.22	.05	.59	-.25	-.36	-.37	-.26	-.23	-.10
Aardrijkskunde oud	-	-	-	.40	-	-	-	-	-.15
Wiskunde	.79	-.01	.69	-.48	-	-	-	-	.19
Wiskunde oud	-1.30	-.60	.03	-.47	-	-	-	-	-.40
Wiskunde A	-	-	-	-	.92	.22	-.21	-.21	.18
Wiskunde B	-	-	-	-	.45	-.04	.26	.09	.19
Natuurkunde	.50	-.11	-.09	-.34	.25	.00	.19	-.22	-.03
Scheikunde	-.19	.20	-.26	.13	-.20	-.03	-.05	.31	-.02
Biologie	-.26	.04	-.21	-.14	.03	-.11	.00	-.17	-.11
Economie	.57	-.23	.43	.09	-.07	-.35	-	-	.09
Handelsw. & Recht	-	-	-	-	-	-.28	-	-	-.28
Econ. Weten. I	-	-	-	-	-	-	-.04	.06	.01
Econ. Weten. II	-	-	-	-	-	-	-.16	-.25	-.20
Tekenen	-.20	.40	-1.08	-.75	-.83	-.95	-.50	-.45	-.80
Handvaardigheid I	-.70	-2.00	-1.18	-1.00	-.88	-.99	-.47	-.28	-.93
Handvaardigheid II	-	-	-	-1.14	-	-.79	-	.00	-.85
Muziek	-	-	-.25	-.82	.04	-.56	-.86	-.90	-.53
Filosofie	-	-	-	-	-	-	-.14	-.32	-.24
Maatschappijleer	-.05	.25	-.28	-.62	-.29	.30	-.28	-.18	-.24
Totaal	.20	-.08	-.04	-.24	-.03	-.20	-.13	-.20	-.12

Legenda: alle discrepantie groter dan .10 of kleiner dan -.10 verschillen ruimschoots significant van .00.

**Tabel 2: De ongestandaardiseerde effecten van schooljaar, vak, onderwijstype, school- en regiokenmerken op de discrepantie tussen het gemiddeld cijfer op het centraal examen en het gemiddeld cijfer op het schoolonderzoek per vak, onderwijstype, vestiging en jaar.**

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Constante	16.78	16.72	16.88	16.59	16.70	16.46
'97/'98	-.17	-.17	-.17	-.17	-.17	-.17
Engels (ref.)						
Nederlands		-.13	-.12	-.12	-.11	-.10
Fries		-.13 (ns)	-.20 (ns)	-.14 (ns)	-.20 (ns)	-.18 (ns)
Latijn		-.14	-.15	-.16	-.17	-.19
Grieks		-.15	-.17	-.18	-.19	-.25
Frans		-.29	-.29	-.29	-.29	-.30
Duits		-.32	-.33	-.33	-.33	-.34
Spaans		-.52	-.50	-.47	-.48	-.44
Turks		.82	1.46	1.47	1.38	1.47
Arabisch		.48	1.26	1.31	1.22	1.25
Geschiedenis /Staatsinrichting		-.08	-.08	-.07	-.07	-.08
Aardrijkskunde		.07	.07	.07	.08	.08
Aardrijkskunde (oud)		.31 (ns)	.27 (ns)	.25 (ns)	.25 (ns)	-.05 (ns)
Wiskunde		.30	.29	.29	.29	.28
Wiskunde (oud)		-.03 (ns)	.03 (ns)	.05 (ns)	.06 (ns)	-.06 (ns)
Wiskunde A		.26	.26	.26	.26	.26
Wiskunde B		.29	.28	.28	.28	.27
Natuurkunde		.06	.05	.04	.04	.03
Scheikunde		.03*	.02 (ns)	.02 (ns)	.01 (ns)	.01 (ns)
Biologie		-.04	-.05	-.05	-.05	-.05
Economie		.23	.24	.23	.23	.23
Handelsw. & Recht		-.13	-.13	-.13	-.13	-.14
Econ. Wet. I		.12	.12	.12	.12	.12
Econ. Wet. II		-.07	-.08	-.08	-.07	-.06
Tekenen		-.71	-.72	-.72	-.72	-.73
Handvaardigheid I		-.87	-.89	-.90	-.90	-.91
Handvaardigheid II		-.65	-.69	-.68	-.68	-.71
Muziek		-.40	-.42	-.41	-.41	-.39
Filosofie		-.13 (ns)	-.14 (ns)	-.11 (ns)	-.16 (ns)	-.28\$
Maatschappijleer		-.14	-.14	-.13	-.12	-.05 (ns)
HAVO (ref.)						
VWO		-.05	-.07	-.07	-.08	-.09
MAVO-C		.07	.12	.12	.15	.18
MAVO-D		-.06	-.04	-.04	-.03	.00 (ns)
% allochtone leerlingen			-.01	-.01	-.01	.01
Protestant-Christelijk (ref.)						
Algemeen Bijzonder				-.04	-.03*	-.03*
Gereformeerd				-.13	-.14	-.15
Openbaar				.02	.03	.03
Reformatorisch				-.22	-.21	-.22
Rooms-Katholiek				-.01 (ns)	-.02*	-.02\$
Samenwerking				.00 (ns)	.01 (ns)	.01 (ns)
Montessori				-.04*	.00 (ns)	.01 (ns)
Gelderland (ref)						
Drente				.04*	.03\$	.02 (ns)

Utrecht				-.06	-.05	-.04
Zuid-Holland				-.06	-.03	-.03
Zeeland				.08	.08	.07
Flevoland				-.08	-.05*	-.05*
Friesland				-.12	-.08	-.08
Groningen				-.07	-.03\$	-.03*
Limburg				.00 (ns)	-.01 (ns)	-.01 (ns)
Noord-Brabant				.02\$	.02*	.02\$
Noord-Holland				-.08	-.04	-.04
Overijssel				-.01 (ns)	-.02 (ns)	-.02\$
Middelgrote steden C4 (ref.)						
Plattelandsgemeenten A2				-.012\$	-.014*	-.016
Plattelandsgemeenten A3				.006	.003\$	.002 (ns)
Plattelandsgemeenten A4				-.001 (ns)	.002 (ns)	.003
Verstedelijkt Platteland B1				.002\$	.001 (ns)	.000 (ns)
Verstedelijkt Platteland B2				.000 (ns)	.001\$	.002\$
Verstedelijkt Platteland B3				-.002*	.001 (ns)	.002\$
Plattelandsstadjes				-.002 (ns)	.003*	.004
Kleine steden				-.001 (ns)	.001 (ns)	.001 (ns)
Middelgrote steden C3				.002 (ns)	.001 (ns)	.001 (ns)
Grote steden				-.001 (ns)	.001 (ns)	.001 (ns)
Toegevoegde waarde intern rendement					.09	.09
Geaggregeerd vestiging residue uit model 5						1.02
Adjusted R <sup>2</sup>	.02	.17	.21	.22	.23	.31

Legenda: alle parameters  $p < .01$ , behalve \*  $.05 > p > .01$ , \$  $.10 > p > .05$  & ns  $p > .10$ . Alleen vestigingen met valide kenmerken ( $N_{\text{vestiging}}=826$ ;  $N_{\text{discrepanties}}=38062$ )



**Tabel 3: Kenmerken van de op vestigingsniveau geaggregeerde discrepanties en residuen, behorend bij modellen 1 tot 5 van tabel 2**

	Gemid- Delde	Stand. Dev.		Scheefheid	Maxi- mum	Mini- mum	Correlaties tussen discrepantie en residuen				
		Bin- nen vesti- ging	Tus- sen vesti- ging				Res idue 1	Res idue 2	Res idue 3	Res idue 4	Res idue 5
Discre- pantie	-.12	.54	.22	-1.28	.49	-1.80	.99	.98	.84	.79	.75
Residue 1	.00	.54	.22	-1.43	.54	-1.76		.99	.84	.80	.75
Residue 2	.00	.49	.22	-1.32	.51	-1.62			.84	.80	.75
Residue 3	.00	.49	.18	-.60	.46	-.83				.95	.88
Residue 4	.00	.49	.17	-.50	.49	-.80					.93
Residue 5	.00	.49	.16	-.37	.55	-.72					

N= 826

**Tabel 4: Het ongestandaardiseerde effect van de per vesting en schooltype geaggregeerde discrepantie op het gemiddeld slaagpercentage per vesting en schooltype, gecontroleerd voor het per vesting en schooltype geaggregeerde gemiddeld cijfer op het centraal examen, onderwijstype en school- en regiokenmerken voor het school 1996/1997**

	Model 1	Model 2	Model 3
Constante	90.85	-61.42	-53.72
Geaggregeerde discrepantie ce-so	14.58	-5.13	-9.56
Geaggregeerd cijfer centraal examen		2.35	2.19
% allochtone leerlingen			-.13
Protestant-Christelijk (ref.)			
Algemeen Bijzonder			1.0*
Gereformeerd			.06 (ns)
Openbaar			.51\$
Reformatorisch			-.04 (ns)
Rooms-Katholiek			.88
Samenwerking			.44 (ns)
HAVO (ref.)			
VWO			-1.13
MAVO			8.07
Montessori			-.08 (ns)
Gelderland (ref)			
Drente			.53 (ns)
Utrecht			2.03*
Zuid-Holland			-.62 (ns)
Zeeland			-.05 (ns)
Flevoland			-.07 (ns)
Friesland			-.11 (ns)
Groningen			-.30 (ns)
Limburg			.88\$
Noord-Brabant			.01 (ns)
Noord-Holland			-.37 (ns)
Overijssel			-.19 (ns)
Middelgrote steden C4 (ref.)			
Plattelandsgemeenten A2			-6.76 (ns)
Plattelandsgemeenten A3			-1.23 (ns)
Plattelandsgemeenten A4			-1.37*
Verstedelijkt Platteland B1			-.59 (ns)
Verstedelijkt Platteland B2			-.22 (ns)
Verstedelijkt Platteland B3			-.11 (ns)
Plattelandsstadjes			.36 (ns)
Kleine steden			-.42 (ns)
Middelgrote steden C3			.00 (ns)
Grote steden			-.11
Adjusted R <sup>2</sup>	.20	.50	.73

Legenda: alle parameters  $p < .01$ , behalve \*  $.05 > p > .01$ , \$  $.10 > p > .05$  & ns  $p > .10$ ;

$N_{\text{vestiging} \times \text{onderwijstype}} = 1663$

## Literatuur

- Agerbeek, M., 1998. "Schoolprestaties 1998." *Trouw* 13 juni: ZENZ.
- Agerbeek, M., E. Hagemen, E. Kreulen en H. Lakmaker, 1997. "Trouw-onderzoek schoolprestaties." *Trouw* 25 oktober: ZENZ.
- Alberts, R. V. J. 1997. *Verslag normhandhaving examens vbo/avo/vwo/ 1994-1997*. Arnhem: Cito.
- . 1998. *Examens. vbo/mavo/havo/vwo*. Arnhem: Cito.
- Dijkstra, A. B., J. Dronkers en R. Hofman. 1997. *Verzuiling in het onderwijs. Actuele verklaringen en analyse*. Groningen: Wolters-Noordhoff.
- Dronkers, J., 1998. "Het Trouw-rapportcijfer van scholen in het voortgezet onderwijs; een analyse van de inspectiegegevens over de schooljaren 1995/96 en 1996/97." *Tijdschrift voor Onderwijsresearch* 23:159-176.
- , 1999. "Veranderden leerlingaantallen in het voortgezet onderwijs in het schooljaar 1998-1999 door de publicatie van inspectiegegevens en de berekening van het schoolcijfer door Trouw in oktober 1997? Een nadere analyse." *Tijdschrift voor Onderwijsresearch* 24:in druk.
- Dronkers, J., M. van Erp, M. Robijn en J. Roeleveld, 1998. "Krijgen leerlingen in de grote steden en met name in Amsterdam te hoge adviezen?" *Tijdschrift voor Onderwijsresearch* 23:17-30.
- Dronkers, J. en M. F. Warnaar, 1999. "'Ze leren er niks maar er is een waanzinnig goede sfeer op deze school'. Over de samenhang tussen kwaliteiten van scholen." *VBSchrift* 25:7-16.
- Eisinga, R., P. Scheepers en L. van Snippenburg, 1991. "The standardized effect of a compound of dummy variables or polynominal terms." *Quality and Quantity* 25:105-114.
- Glopper, K. de en E. van Schooten. 1998. *Dalende leerlingprestaties op de centraal schriftelijke examens Duits, Engels en Frans in mavo, havo en vwo; toetsing en verklaring van waargenomen trends (onderzoeksvoorstel)*. Amsterdam: SCO-Kohnstamm Instituut.
- Hoeven-van Doornum, A. A. van der. 1990. *Effecten van leerlingbeelden en streefniveaus op schoolloopbanen*. Instituut voor Toegepaste Sociale Wetenschappen Nijmegen.
- Hof, L. van 't en J. Dronkers, 1993. "Onderwijsachterstanden van allochtonen: klasse, gezin of etnische cultuur?" *Migrantenstudies* 9:2-25.
- Inspectie van het Onderwijs. 1992. *Examens op punten getoetst. Onderzoek naar de ontwikkeling van de normen bij de centrale examens in het Voortgezet Onderwijs*. Zoetermeer: Ministerie van Onderwijs en Wetenschappen.
- Koeslag, M. en J. Dronkers, 1994. "Overadvisering en de schoolloopbanen van migrantenleerlingen en autochtone leerlingen in het voortgezet onderwijs." *Tijdschrift voor Onderwijsresearch* 19:240-258.
- Maas, C. J. M. en G. W. Meijnen, 1999. "Problems students: a contextual phenomenon?" *Social Behavior and Personality* 27:387-406.
- Mulder, L. 1996. *Meer voorrang, minder achterstand? Het onderwijsvoorrangsbeleid getoetst*. Nijmegen: Instituut voor Toegepaste Sociale Wetenschappen.
- Nagel, I., H. Ganzeboom, W. Haanstra en W. Oud. 1996. *Effecten van kunsteducatie in het voortgezet onderwijs*. Amsterdam: SCO-Kohnstamm Instituut, Universiteit van Amsterdam.
- Pijl, Y. J. 1991. *Discrepanties tussen het schoolonderzoek en het centraal schriftelijk examen in havo en vwo*. Groningen: RION.
- Posthumus, K., 1940. "Middelbaar onderwijs en schifting." *De Gids* 104:24-42.
- Roeleveld, J. en J. Dronkers, 1994. "Bijzondere of buitengewone scholen? Verschillen in effectiviteit van openbare en confessionele scholen in regio's waarin hun richting een meerderheids- of minderheidspositie inneemt." *Mens en Maatschappij* 69:85-108.
- Rohde, E., P. Kuyt, D. Takkenberg en G. Janssen, 1999. "Allochtone leerling presteert minder op examen." *Kwartaalschrift Onderwijsstatistiek* :7-10.
- Slagter, M., 1999. "Centraal examen meet kwaliteit van school het beste." *de Volkskrant* :7.
- Tesser, P. T. M., J. G. F. Merens en C. S. van Praag. 1999. *Rapportage minderheden 1999. Positie in het onderwijs en op de arbeidsmarkt*. Den Haag: Sociaal en Cultureel Planbureau.
- Vink, R. 1999. *Schoolorganisatiekenmerken en intern rendement. Een onderzoek naar de relatie tussen organisatiekenmerken en het Trouw-cijfer*. Tilburg: IVA.

Werf, M. P. C. van der, M. J. Lubbers en H. Kuiper. 1999. *Onderwijsresultaten van VOCL '89 en VOCL '93 leerlingen*. Groningen: GION, Rijksuniversiteit Groningen.

## Noten

<sup>1</sup> Dit is een verkorte versie van het college, waarmee ik op 25 november 1999 afscheid nam als hoogleraar onderwijskunde aan de Universiteit van Amsterdam. Een aantal kenmerken van een afscheidscollage zijn daarom in de tekst van dit artikel terug te vinden. Ik dank Marjan Agerbeek voor het ter beschikking stellen van de inspectiegegevens en Wim Meijnen voor zijn altijd verstandig advies over de gevoelige onderdelen van de complete versie. Mijn huidige adres is Afdeling sociologie en antropologie van de Universiteit van Amsterdam, Oudezijds Achterburgwal 185, 1012 DK Amsterdam. E-mail adres: dronkers@psw.uva.nl.

<sup>2</sup> Dat zou men overigens niet zeggen als men onderwijskundige handboeken raadpleegt. Nauwelijks treft men daarin verhandelingen over de inrichting van eindexamens aan.

<sup>3</sup> Het eerste effectieve nationale onderwijsbestel dateert uit het Koninkrijk Holland, waarbij het onderwijs werd gemodelleerd naar Frans revolutionair voorbeeld en bedoeld was de eenheidsstaat mede vorm te geven.

<sup>4</sup> Vakken met minder dan 10 kandidaten per school en onderwijstypen zijn niet in de analyse betrokken.

<sup>5</sup> De relatie tussen schoolonderzoekcijfer en centraal-examencijfer is nagenoeg lineair. Dit betekent dat discrepanties bij zeer lage of zeer hoge schoolonderzoekcijfers niet groter zijn dan die bij gemiddelde schoolonderzoekcijfers. Dat weerlegt de veronderstellingen dat lage schoolonderzoekcijfers leerlingen effectief zouden aanzetten hogere cijfers voor het centraal-examen te halen of dat leerlingen met hoge schoolonderzoekcijfers zich niet meer zouden inzetten voor het centraal-examen.

<sup>6</sup> Een multilevel analyse levert in de praktijk nagenoeg dezelfde resultaten op. In verband met de inzichtelijkheid voor een groter publiek geef ik daarom de voorkeur aan een multivariate aanpak. De referentiecategorieën van de verschillende afhankelijke variabelen zijn zodanig gekozen dat zij een gemiddelde vertegenwoordigen.

<sup>7</sup> De gemiddelde cijfers op het schoolonderzoek en het centraal examen voor de schooljaren 1996/97 en 1997/98 per onderwijstype

	MAVO C	MAVO D	HAVO	VWO
School 96/97	6,14	6,58	6,36	6,63
School 97/98	6,11	6,57	6,34	6,65
Centraal 96/97	6,34	6,54	6,33	6,51
Centraal 97/98	6,03	6,34	6,14	6,45

<sup>8</sup> De schoolvakken alleen verklaren 14% van de variantie in de discrepanties.

<sup>9</sup> Nagel, Ganzeboom, Haanstra & Oud (1996) laten overigens overtuigend zien dat deze grote discrepantie tussen schoolonderzoek- en centraal-examencijfers niet verklaard kan worden door dat kunstvakken “vluchtvakken” zouden zijn voor zwakke leerlingen. Leerlingen met een kunstvak hebben op niet-kunstvakken gemiddeld geen lagere cijfers dan leerlingen zonder een kunstvak.

<sup>10</sup> Voor de begrijpelijkheid heb ik geen nieuw intern rendement berekend, waarin het cijfers voor drie eindexamenvakken niet in opgenomen zou zijn. Dronkers (1998) laat zien dat dit niet tot een wezenlijk andere uitkomst van het intern rendement van scholen zou leiden.

<sup>11</sup> Deze geaggregeerde discrepanties zijn niet gewogen naar leerlingaantal per vak per school, omdat de vakken met minder dan 10 kandidaten reeds uit de analyses verwijderd zijn.

<sup>12</sup> Eisinga, Scheepers & Snippenberg (1991) hebben laten zien dat dit gebruik van residuen als schatting van het effect van een context hetzelfde oplevert als het gebruik van een groot aantal dichotome variabelen voor elke school afzonderlijk.

<sup>13</sup> Deze 196 scholen omvatten 24% van alle 826 scholen.