RR – 85-20

OVER DEFINITIE EN KWANTIFICATIE VAN SCHOOLLOOPBANEN

Jan de Leeuw
Vakgroep Datatheorie FSW/RUL
Ita Kreft
Vakgroep Algemene en Vergelijkende
Onderwijskunde UvA

1 INLEIDING

Aan schoolloopbanen wordt al sinds het begin van deze eeuw aandacht besteed. Het gaat daarbij om het bestuderen van variabelen die verband houden met schoolsucses, dat wil zeggen het vinden van variabelen die niet-triviale samenhangen vertonen met schoolvorderingen en het bestuderen en beschrijven van die samenhangen. Deze beschrijving kan, natuurlijk, kwalitatief en kwantitatief zijn, en theoretisch zowel als empirisch. In dit artikel zijn we geïnteresseerd in empirische en kwantitatieve bestudering van de schoolloopbaan.

De schoolloopbaan, de naam zegt het al, is natuurlijk een in de tijd verlopend proces. Dat wil zeggen dat we de loopbaan kunnen beschouwen als een ontwikkeling, waarbij toestanden elkaar opvolgen. Die toestanden zijn specifiek voor de leeftijd van de leerling en voor het schoolsysteem van het land waarin we loopbanen bestuderen. In Nederland, bijvoorbeeld, kan een kind van drie niet op de basisschool zitten, en kan men van de kleuterschool niet in één keer naar 3 VWO gaan. Zowel toestanden als overgangen zijn dus leeftijd-en systeemgebonden. In de volgende paragrafen zullen we proberen dit wat meer formeel uit te drukken en te beschrijven.

Men kan bij de beschrijving van schoolloopbanen een aantal stadia van gedetailleerdheid onderscheiden. In ouder onderzoek zijn er in het algemeen slechts gegevens over beginniveau en eindniveau, en geen gegevens over het verloop binnen schooltypen. Soms weten we iets over de doubleerstatus in het basisonderwijs, maar dit soort variabelen wordt toch als regel als additionele informatie voor de dwarsdoorsnede van de schoolloopbaan bij bijvoorbeeld het begin van het secundaire onderwijs gebruikt, en niet als dynamische informatie om de loopbaan in het basisonderwijs te beschrijven. Natuurlijk is het aantal mogelijke loopbanen binnen het basisonderwijs ook beperkt. Hetzelfde geldt, eigenlijk, voor het oude categoriale schoolsysteem van voor de Mammoetwet. Maar de grote differentiatie die tegenwoordig mogelijk is in het voortgezet onderwijs, het feit dat overgangen gemakkelijker gemaakt kunnen worden, de rol van het vakkenpakket, dit alles maakt dat de variabelen 'toestand aan het begin van het secundair onderwijs' en 'toestand aan het eind van het secundair onderwijs' steeds minder informatie over de schoolloopbaan gaan bevatten. Daarom ook wordt het steeds belangrijker, en steeds informatiever, om de schoolloopbaan dynamisch te interpreteren en longitudinaal te bestuderen. Voor verklarende, maar ook prediktieve, doeleinden is het ontoelaatbaar om de tussenliggende informatie zonder meer te verwaarlozen. We mogen aannemen dat dit ook één van de redenen is, waarom er de laatste tijd door bijvoorbeeld het CBS longitudinale databestanden opgebouwd zijn.

Bestuderen van de schoolloopbanen in de lengte brengt diverse problemen met zich mee. De belangrijkste is misschien wel, dat de grote mate van gedetailleerdheid die mogelijk is tot allerlei technische complicaties leidt. De schoolloopbaan van een individu, longitudinaal opgevat, is een variabele met duizenden of zelfs miljoenen mogelijke waarden, en bestu - deren van samenhang met andere variabelen is slechts mogelijk nadat het aantal mogelijke waarden aanzienlijk gereduceerd is. Eén reductie is alleen maar bekijken van beginniveau en eindniveau, maar de tegenwoordige databestanden zijn zo rijk dat veel meer soorten reducties mogelijk zijn. Sommige van dit soort reducties komen neer op kwantificatie, een bijzonder soort reductie en systematisering van de loopbaangegevens met bepaalde voor de hand liggende voordelen. Na kwantificatie kan de schoolloopbaan gemakkelijk vergeleken worden met andere gekwantificeerde variabelen, en wordt het proces van bestudering van verbanden ook aanzienlijk eenvoudiger. In de loop van dit artikel zullen we diverse reductiemethoden met elkaar vergelijken, en zullen we in het bijzonder aandacht besteden aan methoden om de schoolloopbaan te kwantificeren. Hiertoe is het in de eerste plaats nodig om wat afspraken te maken over terminologie en notatie.

2 INGREDIÏENTEN VAN SCHOOLLOOPBAANONDERZOEK

2.1 De schoolloopbaan

Bij schoolloopbaanonderzoek gaan we uit van een gegeven onderwijssysteem, in ons geval dus van het Nederlandse. Iedere leerling die deelneemt aan het systeem, en dus ook iedere leerling in het onderzoek, bevindt zich op ieder moment in een bepaalde binnen het systeem bestaande toestand. Het aantal mogelijke toestanden dat binnen een systeem bestaat is in principe eindig, maar in de praktijk buitengewoon groot. Binnen het basisonderwijs valt dit nog wel mee, hoewel de diverse vormen van buitengewoon onderwijs toch voor een grote variatie van toestanden zorgen. In het secundaire onderwijs wordt het aantal verschillende toestanden groter, niet alleen door de vele vormen van voortgezet onderwijs die er bestaan, maar ook omdat men rekening zou kunnen houden met bijvoorbeeld vakkenpakketten. In het tertiair onderwijs is het aantal mogelijke toestanden nog weer groter, omdat er daar helemaal een enorm skala van onderscheidingen is dat men kan hanteren. Omdat het duidelijk onmogelijk is om alle voorkomende variaties in de toestanden in een onderzoek te betrekken, kunnen we schoolloopbanen uitsluitend beschrijven na reductie van het aantal toestanden. Ieder schoolloopbaanonderzoek specificeert dus een eindige lijst van mogelijke toestanden die in de schoolloopbaan onderscheiden worden. Daarmee samenhangend is het ook nodig het tijdsinterval vast te stellen waarover we de loopbaan bestuderen.

Gegeven zijn dus een tijdsinterval (0,T) en een verzameling $S= (s_1, s_2, ..., s_k)$ van toestanden die in de schoolloopbaan op kunnen treden. De schoolloopbaan van leerling i kan nu beschreven worden als een functie f_i die (0,T) afbeeldt in S. Daarbij is dus f_i (t) de toestand waarin leerling i zich op tijdstip t bevindt. Wanneer de leerlingen beschouwd kun - nen worden als een willekeurige steekproef uit een welomschreven populatie van leerlin - gen, dan ligt voor de hand om de individuele schoolloopbanen te zien als realisties van een stochastisch proces op (0,T). Statistische uitspraken over de schoolloopbaan zullen dan uitspraken zijn over het proces f, eventueel in relatie tot stochastische variabelen die de ach - tergrond van de leerling beschrijven.

We moeten bij de feitelijke registratie van schoolloopbanen er natuurlijk rekening mee houden dat het doen van oneindig veel metingen in een eindige tijd niet mogelijk is. Daarom kan de schoolloopbaan ten hoogste in een eindig aantal meetpunten t_j (j=l, ..., m) met $o \leq t_l < ... < t_m \leq T$ vastgesteld worden. Veranderingen tussen twee meetpunten in worden niet adekwaat geregistreerd, maar vanzelfsprekend kunnen de meetpunten willekeurig dicht bij elkaar liggen. De gegevens van een schoolloopbaanonderzoek bestaan dus tenminste uit een n x m matrix F met elementen $f_{ij} = f_i(t_j)$. Daarnaast hebben we in het algemeen nog een aantal metingen op achtergrondsvariabelen. Dit definieert een tweede matrix, zeg X, van afmetingen n x p. Hierbij is x_{is} de waarde van leerling i op achtergrondsvariabele s. Het is in het geheel niet nodig dat de waarden van X en F gehele getallen zijn. We kunnen bijvoorbeeld x_{is} gelijk aan 'Noord-Brabant' hebben en f_{ij} gelijk aan '3 HAVO'. Natuurlijk is het wel mogelijk de categorieën van alle variabelen door middel van nummers te coderen, maar aan deze nummers moet niet zonder meer ordinale of cardinale betekenis worden toegekent.

Naast de boven geschetste puur formele beschrijving van de schoolloopbaan is ook een meer inhoudelijke, en interpretatieve, mogelijk. Kreft en Bronkhorst (1985) onderscheiden bijvoorbeeld aan schoolloopbanen de duur, de structuur, en de kwaliteit. Wat met duur bedoeld wordt is duidelijk. Structuur verwijst naar veranderingen, naar opstroom, af - stroom, en uitstroom. Kwaliteit tenslotte wordt gedefinieerd in termen van de waarde van het behaalde diploma, waarbij het begrip waarde in principe natuurlijk economisch bedoeld is, maar in het kader van schoolloopbaanonderzoek dikwijls ook gedefinieerd kan worden in termen van mogelijke vervolgopleidingen. In het secundair onderwijs, waar we ons in dit onderzoek mee bezig houden, zijn de vier belangrijkste kwaliteits-niveaus LBO-MAVO-HAVO-VWO. In een latere paragraaf van dit paper zullen we trachten de tot nu toe gebruikte operationalisaties van het begrip schoolloopbaan, in Nederlands schoolloopbaan - onderzoek, te bespreken in termen van de begrippen duur, structuur, en kwaliteit, maar ook in termen van het formalisme van functies f_i op (0,T) met waarden in S. Eerst zullen we echter bespreken welke groepen leerlingen over het algemeen in schoolloopbaanonder - zoek bestudeerd worden.

2.2 Schoolloopbaancohorten

Bij de bestudering van de schoolloopbaan gaan we over het algemeen uit van een school - loopbaancohort. Volgens de definitie van Ryder (1965) is een cohort een groep mensen die allen tegelijkertijd een bepaalde belangrijke gebeurtenis meegemaakt hebben. Het kan zijn, dat ze allemaal op hetzelfde tijdstip geboren zijn (een geboortecohort) of dat ze allemaal op hetzelfde tijdstip getrouwd zijn (een huwelijkscohort). Tijdstip is hierbij ruim gedefinieerd, feitelijk als tijdsinterval.

Een schoolloopbaancohort is daarom een groep mensen die allen op een gegeven tijdstip in hetzelfde stadium van hun schoolloopbaan waren. Voor de hand liggende tijdstippen, in het Nederlandse onderwijssysteem, zijn de eerste klas van de basisschool, of de eerste klas van het voortgezet onderwijs. Schoolloopbaancohorten bestaan natuurlijk uit leerlingen uit een bepaalde periode, en de significante gebeurtenis die ze gemeen hebben is gedefinieerd binnen een bepaald onderwijssysteem. Voor verdere uitwijdingen rond het begrip cohort, en voor methodologische problemen die samenhangen met het vergelijken van verschillende cohorten, verwijzen we naar Glenn (1977), Fienberg en Mason (1978), Hagenaars (1985), en Mason en Fienberg (1985). In principe houden we ons in dit artikel niet bezig met vergelijkend schoolloopbaanonderzoek, waarbij verschillende cohorten (op verschillende tijdstippen gestart, of deel uitmakend van verschillende populaties) vergeleken worden. Dronkers (1983) en Peschar (1983) bespreken vergelijkend Nederlands schoolloopbanenonderzoek. Het gaat ons in dit artikel om slechts één enkel cohort van leerlingen, dat gedurende een deel van de schoolloopbaan gevolgd wordt.

De Nederlandse schoolloopbaanonderzoeken, die we hieronder bespreken, verschillen niet alleen in hun definitie en operationalisering van de schoolloopbaan, maar ook in het cohort dat ze gebruiken. Uitgangspunt bij onze bespreking is dus, dat we na willen gaan hoe de diverse onderzoekers geprobeerd hebben, voor hun periode en hun cohort, de enorme verscheidenheid van schoolloopbanen te reduceren ten einde een redelijk overzichtlijke beschrijving mogelijk te maken. Mede op basis van dit eerdere onderzoek zullen we daarna onze eigen vorm van reductie voorstellen. Tesser (1981) geeft een meer omvattend, maar minder specifiek, overzicht van Nederlands schoolloopbanenonderzoek.

3 OPERATIONALISATIE EN METING VAN DE SCHOOLLOOPBAAN

3.1 Inleiding

Veel Nederlands schoolloopbaanonderzoek consentreert zich op één enkel moment in de schoolloopbaan. Bijna zonder uitzondering is dit de overgang van het lager onderwijs naar het middelbaar onderwijs. Aan het eind van de zesde klas zijn drie belangrijke schoolloop - baan variabelen beschikbaar. Dat zijn het advies van de onderwijzer omtrent de meest geschikte vorm van voortgezet onderwijs, de feitelijke keuze van een vervolgopleiding, en in veel gevallen scores op de landelijk of regionaal afgenomen schoolvorderingstoetsen. Het ligt voor de hand om deze variabelen te bestuderen, omdat het in feite de laatste gele - genheid is waarbij vrijwel iedere leerling nog op precies dezelfde manier beoordeeld wordt. In de eerste klas van het secundaire onderwijs begint immers formeel de differentiatie pas binnen ons onderwijssysteem.

Natuurlijk kan men nauwelijks van een loopbaan spreken, wanneer het slechts om één enkel moment gaat. Wij zullen ons dan ook in dit artikel niet of nauwelijks bezighouden met schoolloopbaanonderzoek dat als belangrijkste afhankelijke variabele de keuze van de vorm van secundair onderwijs heeft. Een historisch uitermate belangrijk onderzoek als het Talentenproject (Van Heek e.a., 1968) bespreken we daarom in het geheel niet.

3.2 Het oordeel van experts

Een van de methoden om schoolloopbanen met elkaar te vergelijken werd in het Neder - landse onderzoek geïntroduceerd door Van Weeren (1960). Hij beschreef een groot aantal veel voorkomende schoolloopbanen in woorden ('bevorderd naar 5e klas HBS-A zonder vertraging', 'diploma MMS met 1 jaar vertraging'). In totaal ging het daarbij om 102 schoolloopbanen. Een team van 100 experts (leraren) deelde deze loopbanen in 11 catego - rieën, naar de zwaarte van de eisen die deze loopbanen aan de leerlingen stellen. De oorde - len van de leraren, die een grote mate van homogeneiteit te zien gaven, werden met klas - sieke psychologische schaalmethoden omgezet in schaalwaarden. Op deze manier verkreeg Van Weeren een schaling van de 102 loopbanen. Dezelfde schaal werd tevens gebruikt door Groen (1967). Een soortgelijke procedure werd meer recent gevolgd door Cremers (1980). Het ging bij hem echter om 41 loopbanen ('na 4 jaar voortgezet onderwijs op de 3-jarige MAVO, deze opleiding na de derde klas beeindigd zonder diploma') en om slechts 40 leerkracht-experts. Ook Cremers vond een redelijke mate van consistentie tussen beoordelaars, hoewel het schooltype waarop de leraar werkzaam was wel degelijk invloed had op zijn beoordeling.

Op deze procedure kan men een aantal aanmerkingen maken. De verbale omschrijvingen van loopbanen die gebruikt worden, houden vooral rekening met de duur en de kwaliteit van de loopbaan, en in veel mindere mate met de structuur. Wanneer we de structuur, dat wil zeggen de diverse overgangen tijdens het secundair onderwijs, in overweging nemen, dan neemt het aantal mogelijke loopbanen snel toe. Kreft en Bonkhorst (1985) vinden op basis van 5 meetmomenten al 800 verschillende loopbanen, terwijl zij uitsluitend leerlingen beschouwen die het secundair onderwijs op een scholengemeenschap beginnen. Meester en De Leeuw (1983) vinden in het SMVO-kohort meer dan 1600 loopbanen. Het is duidelijk dat geen enkele expert een dergelijke omvangrijke beoordelingstaak aankan.

Een tweede bezwaar is, dat we niet zozeer de schoolloopbaan als wel het oordeel van een bepaalde groep personen over de schoolloopbaan bestuderen. Door gebruik van het woord 'experts' wordt een zekere mate van objectiviteit gesuggereerd, maar meer dan suggestie is dat natuurlijk niet. Het feit dat verschillende soorten leraren verschillende oordelen geven toont dit al aan. En, meer in het algemeen, is er weinig reden te veronderstellen dat leraren objectiever zijn dan bijvoorbeeld leerlingen. Men kan zich zelfs afvragen in welke zin een VWO-loopbaan eigenlijk zwaarder is dan een LBO-loopbaan. Uiteindelijk vindt men dan natuurlijk dat de procedure gebaseerd is op waarde-oordelen, waarbij met name het oordeel dat hoofdarbeid zwaarder of veeleisender is dan handenarbeid een grote rol speelt.

3.3 Het gebruik van tests

Er is een alternatieve procedure om schoolloopbanen te schalen die in wat mindere mate van oordelen van individuen afhangt. Stel dat van de leerlingen in het kohort testresultaten bekend zijn. Wanneer we deze leerlingen vervolgens onderverdelen in groepen met dezelfde of vrijwel dezelfde schoolloopbaan, dan kunnen we voor iedere groep het gemiddelde test resultaat bepalen. Dit geeft dan vervolgens een schaalwaarde voor de loopbaan.

Diverse varianten van deze procedure zijn in het Nederlandse Loopbaanonderzoek gebruikt. Dirkzwager (1966) en Van Dijk (1977) hebben deze methode toegepast, vooral natuurlijk om de relatie tussen de intelligentietest en het schoolsucses te bestuderen. Dirkzwager geeft de meest uitvoerige resultaten. Hij onderscheidt 42 kriterium groepen (voorbeeld 'meisjes die in november 1960 in 4 MO zaten en in mei 1963 in 6 Gymnasium-B' of 'jongens die in november 1960 in 6 LO zaten en in mei 1963 niet meer op een school'). Van elk van de groepen, die dus geen onderdelen vormen van een goed gedefi -

nieerd cohort, maar die wel op twee vaste tijdstippen gemeten zijn, geeft Dirkzwager mediaan, bovenste kwartiel, en onderste kwartiel van de acht subtests van de DAT. De medianen zouden als loopbaan-scores gebruikt kunnen worden.

Een verwante procedure wordt gevolgd door Meester en De Leeuw (1983, pagina 117-122). De vier meetmomenten 1978-1979-1980-1981 gecombineerd met het al dan niet behaald hebben van een diploma leverden bij kruising meer dan 1600 verschillende loopbanen op. Dit werd door middel van 'moeizame manipulatie' (l.c., pagina 118) gereduceerd tot 30 typen loopbanen, onderverdeeld in vijf hoofdgroepen. Voor ieder van de 30 groepen wordt de verdeling gegeven op de TIB-test, gegroepeerd in 7 categorieën. In principe zou men deze verdeling kunnen gebruiken om de 30 typen loopbanen een schaalwaarde toe te kennen.

Het gebruik van een test om loopbanen te beschrijven heeft een aantal nadelen. In de eerste plaats wordt een nieuwe variabele toegevoegd die, in principe althans, met de loopbaan op zichzelf weinig te maken heeft. In de tweede plaats, hiermee samenhangend, is de keuze van de test nogal willekeurig. Algemene IQ-tests hangen behoorlijk sterk samen, maar wat let ons loopbanen te beschrijven in termen van prestatie-motivatie of in termen van introversie-extraversie. Wanneer het erom gaat de relatie tussen intelligentie en de loopbaan te bestuderen (zoals bij Dirkzwager, Van Dijk, en Meester en De Leeuw), dan is de procedure alleszins gerechtvaardigd. Als maat voor de kwaliteit van een schoolloopbaan is het gemiddelde IQ te eenzijdig en te willekeurig. Met minstens evenveel reden zou men schoolloopbanen kunnen beschrijven in termen van het verwachte inkomen, gegeven de loopbaan.

3.4 Overgangsmatrixen en stroomdiagrammen

De procedures om de schoolloopbaan te beschrijven die we in 3.2 en 3.3 bekeken hebben voldeden niet, omdat ze bij de beschrijving externe criteria betrokken die niet direct relevant waren, te weten de opinie van experts of de score op een test. Achter beide voorstellen zit, als we ons niet vergissen, een illusie van objectiviteit verborgen. De expert wordt geacht de kwaliteit van een loopbaan objectief te kunnen beoordelen, de IQ-test meet de overanderlijke genotypische intelligentie en is daardoor een objectieve maatstaf van begaafdheid. Dat leraren ook maar mensen zijn, en dat scores op de IQ-test sterk beïnvloed worden door scholing, zijn relatief recente ideeen. Het lijkt daarom noodzakelijk om af te zien van externe maatstaven, en om de schoolloopbaan geheel intern te beschrijven. Dit kan op verschillende manieren. In deze paragraaf bespreken we het gebruik van overgangsmatrixen en stroomdiagrammen.

Stroomdiagrammen worden bijvoorbeeld uitvoerig toegepast in het Van Jaar tot Jaar Onderzoek (Kropman en Collaris, 1974, Collaris en Kropman, 1978). Het duidelijkste misschien in bijlage A van Collaris en Kropman (1978). De onderscheiden schooltypen zijn LO, VGLO, ULO, VHMO, LBO, MBO, HBO, WO, en geen verder onderwijs. Ieder individu in de ITS steekproef definieert een reeks typen, bijvoorbeeld LO, LBO, ULO, STOP of LO, VHMO, WO, STOP. Het stroomdiagram laat nu op redelijk inzichtelijke wijze zien welke reeksen voorkomen, en met welke frequentie. Door alleen op dit soort reeksen te letten verwaarlozen we het duur-aspect van de loopbaan, en merken we in ieder geval bepaalde kanten van de structuur niet op (zoals blijven zitten). Natuurlijk is het mogelijk stroomdiagrammen precieser te maken door meer toestanden te onderscheiden. In CBS (1982a) bijvoorbeeld staan in diagram 1 t/m 5 gespecificeerde schoolloopbanen per schooltype, waaruit doorstroom en uitstroom precies uit af te leiden zijn. Dit soort diagrammen brengen een loopbaan zoals LBO1, LBO2, LBO2, LBO2, LBO3, vertrek met diploma prima in beeld. Er wordt op vaste tijdstippen gemeten, dus de duur is precies bepaald, de structuur wordt niet helemaal duidelijk omdat uit de diagrammen niet afgeleid kan worden welke veranderingen van schooltype er precies zijn. In CBS (1982a, tabel 27) staan de veranderingen van type weergegeven in de vorm van aantallen overgangen, maar uit deze overgangstabellen is de loopbaan binnen type weer niet af te leiden. Hetzelfde geldt voor de vergelijkbare stroomdiagrammen en overgangsmatrixen die voor het SMVO-cohort in CBS (1982b) gepresenteerd worden. Op ingenieuze wijze worden grote hoeveelheden informatie gepresenteerd, maar onvermijdelijk gaat een deel van de infor - matie over de loopbaan verloren. Het is bovendien uitermate lastig om zonder nadere ana - lyse wat meer algemene uitspraken over de loopbaan te doen. Er is dus niets mis met deze vorm van presentatie, integendeel zelfs, maar een compleet en inzichtelijk beeld krijgen we er niet door.

3.5 Afgeleide variabelen

Er is ten minste één poging gedaan de loopbaan te beschrijven in temen van een aantal variabelen afgeleid uit de loopbaan. Wanneer men een theorie heeft, die zegt welke aspecten van de loopbaan belangrijk zijn, bijvoorbeeld voor de latere verwerving van inkomen, dan kan men proberen deze aspecten te meten. Diederen (1983) heeft dit geprobeerd. Aan een loopbaan onderscheidt hij het niveau van het aanvangsschooltype, of er al dan niet een omweg gemaakt is, wat het saldo van de niveauveranderingen is, of er ontwikkeling in de beroepswens is, en wat het uiteindelijke bereikte opleidingsniveau is. Het is duidelijk dat aspecten van duur, structuur, en kwaliteit allemaal een rol spelen, maar het is eveneens duidelijk dat uit een complete schoolloopbaan nog wel honderden variabelen van dit type te construeren zijn. De keus van Diederen blijft toch tot op zekere hoogte willekeurig, en in ieder geval wordt zeer veel informatie niet gebruikt. Hoe gecharmeerd men van deze oplossing is, zal samenhangen met de hoogte van de pet die men opheeft van sociaal wetenschappelijke theorievorming. Een wat minder aanvechtbare beschrijving van de loopbanen in termen van een klein aantal variabelen vinden we bij CBS (1982b). Aanvangsniveau is VWO/HAVO/MAVO/LBO. Een leerling kan tijdens de loopbaan afstromen, opstromen, of op hetzelfde schooltype blijven. En een loopbaan kan vertraagd of niet vertraagd zijn. Er zijn dus 4x3x2=24 typen, maar vier typen komen niet voor omdat men vanaf het VWO niet kan opstromen en vanaf het LBO niet kan afstromen. De door het CBS gebruikte 20 typen werden door Meester en De Leeuw (1983) aangevuld tot 30 typen, doordat bij hen ook de aard van de op- en afstroom een rol speelde. Hoewel met name de CBS-procedure netjes en sluitend is, is natuurlijk de hoeveelheid reductie enorm. De nadruk ligt bovendien op kwaliteit, duur en structuur spelen een kleinere rol.

3.6 Individuele geschiedenissen

In onderzoeken met kleine groepen leerlingen, al dan niet deel uitmakend van een cohort, kan men van ieder individu de schoolloopbaan in detail beschrijven. Een klassiek voorbeeld is Wiegersma e.a. (1963). Bij deze aanpak zijn er overigens nog een groot aantal variaties mogelijk in de mate van objectiviteit die nagestreefd wordt. Wiergersma e.a. geven impressionistische beschrijvingen, met zeer weinig poging tot standarisatie, min of meer in de klinische case-history traditie. Voor schoolloopbaan onderzoek doet dit uiter - mate ouderwets en willekeurig aan. Een veel moderne, en ook veel bevredigender, variant vinden we in het fraaie onderzoek van Peschar (1975).

Peschar construeerde, uit het GALO bestand, 112 gematchte paren, waarbij het eerste lid van het paar uit een laag sociaal milieu kwam, en het tweede lid uit een hoog milieu. De paren waren gematched op sexe, IQ, en leeftijd. In hoofdstuk 7 van zijn boek presenteert Peschar enige 'georiënteerde bevindingen', dat wil zeggen school - en beroepsloopbanen van enkele geselecteerde paren. Dit gaat in de vorm van functies van de tijd, precies zoals door ons in sectie 2.1 besproken is. De toestanden zijn schooltypen, waarvoor de negen geordende typen uit de ITS onderzoeken gebruikt worden. Men ziet dus in een bepaald punt van de tijd iemand van niveau 3 naar niveau 4 stappen. Beroepsniveau wordt gedefi - nieerd als het niveau van de opleiding wat voor het beroep nodig is. Hierdoor kunnen we ook voor beroepsloopbaan dezelfde negen categoriën gebruiken. Het geordende karakter van de categoriën maakt het mogelijk er mee te rekenen, en Peschar geeft dan ook gemiddelde loopbanen voor zijn hoge-SES individuen. De methode Peschar werd ook gehanteerd, met wat kleine wijzigingen en met zo mogelijk nog meer virtuositeit, door

Bosman, Louwes, en Van der Meer (1980) in een vergelijking van paren jongens en meisjes. Ongelukkigerwijs vonden zij slechts kleine verschillen tussen loopbanen, terwijl ze grote verschillen verwachtten, en ook op basis van analyse van interviewprotokollen behoorlijke verschillen leken te vinden. Volgens Bosman e.a. kan dit, voor een deel mis - schien, toegeschreven worden aan de specifieke ordinale schaling van de ITS-niveau variabele (l.c. pagina 70-74).

Het spreekt vanzelf dat de individuele geschiedenis geen informatie verwaarloost, mits het aantal toestanden dat onderscheiden wordt maar groot genoeg is. We zijn het met Bosman, Louwes, en Van der Meer eens, dat gebruik van slechts negen toestanden over de jaren heen tot ontoelaatbare simplificaties moet leiden, én dat het toekennen van schaalwaarden 1 t/m 9 aan deze toestanden al helemaal uit de lucht gegrepen is.

3.7 De opkomst van het schoolloopbaanmodel

Gebruik van lineaire structurele modellen (ook wel padmodellen of causale modellen genoemd) werd in het Nederlandse schoolloopbaanonderzoek geintroduceerd door Dronkers en De Jong (1978). Dit soort modellen is sinds 1978 toegepast door vele onder - zoekers, enerzijds omdat ze een model willen vinden dat de complete schoolloopbaan, inclusief achtergrondsvariabelen, van de huidige generatie zo goed mogelijk beschrijft, anderzijds voor vergelijkend schoolloopbaanonderzoek. Recente overzichten van dit soort onderzoek zijn gepubliceerd door Dronkers (1983) en Peschar (1983). Voor de meest moderne verschijningsvorm van het Nederlandse schoolloopbaanmodel verwijzen we naar Tesser (1984, 1985).

In de oudere schoolloopbaanmodellen, die vooral gebaseerd waren op het Van Jaar tot Jaat materiaal, werd de schoolloopbaan geoperationaliseerd door metingen op een klein aantal punten. De klassieke variabelen zijn doubleerstatus op het LO, score op de CITO toets, advies onderwijzer 6e klas LO, keuze VO, en eindniveau VO. Toekennen van schaal - waarden aan de diverse vormen van voortgezet onderwijs was aanvankelijk een groot probleem, maar De Leeuw en Stoop (1979) lieten zien dat de keuze van kwantifikaties min - der willekeurig gemaakt kon worden door gebruik van zogenaamde optimale schaaltechnie - ken. Niettemin is de beschrijving van de schoolloopbaan uiterst gereduceerd, en evenals bij de bespreking van Diederen (1983) moeten we ons afvragen of deze reductie wel zo optimaal is.

Met het beschikbaar komen van informatie uit cohorten die langer of meer intensief ge - volgd zijn, zoals het Groningen-cohort van Meijnen of het SMVO-cohort van het CBS, kwamen er meer meetpunten beschikbaar, en werd het mogelijk de schoolloopbaan meer in detail te beschrijven. Vanzelfsprekend had dit mede tot gevolg, dat het aantal mogelijke schoolloopbanen exponentieel toenam, zodat het reductie-probleem groter en meer urgent werd. Dit is eens te meer zo, omdat de padmodellen die tot nu toe gebruikt zijn afhankelijk variabelen nodig hebben die numeriek zijn. Stroomdiagrammen en overgangsmatrixen waren dus niet voldoende. De diverse meetmomenten in de loopbaan moesten gekwantificeerd worden.

3.8 Stappen

In 3.6 hebben we gezien dat Peschar (1975) de individuele loopbaan uitzette als functie van de tijd. Omdat er slechts een relatief klein aantal meetpunten zijn, ziet de loopbaan eruit als een functie die op een klein aantal tijdstippen van niveau verandert, met slechts een beperkt aantal niveaus. Een stapfunctie dus. In Peschar's onderzoek waren er slechts negen stappen, de schoolniveaus. Tijdens de loopbaan kwamen er daardoor weinig wisselingen voor, de meeste mensen blijven immers op hetzelfde schooltype zitten. Er is dus weinig aandacht voor de structuur van de loopbaan.

Bosker, Hofman, en Van der Velden (1985), zie ook Bosker en Van der Velden (1985), gebruikten aanmerkelijk subtielere coderingen, waardoor de stapfunctie tijdens de schoolloopbaan de kans krijgt aanmerkelijk meer stappen te maken. De loopbaan wordt beschreven in termen van twee variabelen: tussentijds niveau en bereikt eindniveau. Tussentijds niveau wordt vastgesteld bij het begin van het VO. We geven de keuzes IBO, LBO, MAVO, HAVO, VWO achtereenvolgens de scores 2 t/m 6. Wanneer het aantal doublures op het LO één is, dan trekken we hier één vanaf, als er meer doublures zijn op het LO dan trekken we er twee vanaf. Tussentijds niveau loopt nu van 0 (IBO, na twee maal doubleren) tot 6 (VWO, zonder doubleren). Merk op dat mensen die na twee maal doubleren op het LO naar de HAVO gaan een even hoge score krijgen als mensen die zonder doubleren naar het LBO gaan en als mensen die na één jaar doubleren MAVO kiezen. Dat is natuurlijk vreemd, en moeilijk te beargumenteren. Bosker e.a. doen daar dan ook zelfs geen poging toe.

Een soortgelijke procedure wordt gevolg voor eindniveau. IBO, LBO, MAVO, HAVO, VWO zijn weer 2 t/m 6. En voor ieder voltooid leerjaar wordt er één bijgeteld. VWO5 krijgt dus score 11, en IBO1 krijgt score 3. Alweer gelden dingen als IBO5 = LBO4 = MAVO3 = HAVO2 = VWO1. Bosker e.a. definiëren tenslotte het relatieve eindniveau, als verschil van bereikt eindniveau en tussentijds niveau. In hun discussie van deze begrippen zitten een paar merkwaardige inconsistenties. Bovendien gebruiken zij hun scores alleen om eindniveau te coderen. In deze codering zitten wel een aantal aspecten van de schoolloopbaan verpakt, maar het is toch slechts één enkele statische variabele, zoals het loopbaantype in CBS (1982b) of in Meester en De Leeuw (1983). De procedure van Bosker e.a. wordt overigens door Roeleveld, Van den Eeden, en De Jong (1985) dynamisch gebruikt om de hele loopbaan te coderen. De leerling krijgt in ieder jaar een niveauscore, en deze niveauscores worden als kwantitative variabelen in een school - loopbaanmodel gebruikt. In sectie 5.2 gaan we nader op de door hen gevolgde methode in. Bij deze procedure blijft het wat bezwaarlijk, en wat moeilijk te verdedigen, dat LBO3 = MAVO3 = HAVO1 en dergelijke. We kunnen gebruik van dit soort coderingen in feite vergelijken met de coderingen die gebruikt werden door sociaal milieu, advies, keuze, en eindniveau voordat homogeniteits analyse en correspondentie analyse door De Leeuw en Stoop (1979), Gifi (1980, 1981), CBS (1982b), en Meester en De Leeuw (1983) toegepast werden om de coderingen op lineariteit en optimaliteit te onderzoeken. Het ligt dus eigenlijk voor de hand om ook nu te kijken welke coderingen van stappen in de schoolloopbaan met behulp van homogeniteitsanalyse verkregen worden.

3.9 Homogeneitsanalyse van schoolloopbaanpatronen

In deze paragraaf gaan we na hoe Tesser (1985) numerieke loopbaanvariabelen geconstru - eerd heeft. We doen dat wat uitvoeriger, omdat Tesser's procedure veel lijkt op de onze. In zijn hoofdstuk 4.3 gebruikt Tesser de bekende aspecten om de loopbaan gedeeltelijk te beschrijven. Deze aspecten kennen we ook al uit CBS (1982b), en ze worden ook in de - zelfde vorm bijvoorbeeld gebruikt door Faasse, Bakker, Dronkers, en Schijf (1985).

In hoofdstuk 4.4 doet Tesser iets nieuws. Iedere leerling in het cohort (in Tesser's geval een deelbestand van het SMVO cohort van ongeveer 1000 leerlingen) krijgt in de opeenvolgende schooljaren 1977-1982 een positie toegekend. Er zijn 32 posities: LBO1 t/m LBO4, uit LBO met/zonder diploma, MAVO1 t/m MAVO4, uit MAVO met/zonder diploma, MBO1 t/m MBO3, uit MBO zonder diploma, HAVO1 t/m HAVO5, uit HAVO met/zonder diploma, HBO1 en HBO2, VWO1 t/m VWO6, en uit VWO zonder diploma. Omdat er zes meetpunten zijn, kunnen we in theorie $32^6 = 2^{30}$, dat wil zeggen ongeveer een miljard verschillende loopbanen coderen. Verreweg de meesten komen natuurlijk niet voor. Leerlingen beginnen niet in VWO6, en gaan niet van HAVO5 naar LBO1, om maar eens een paar onwaarschijnlijke overgangen te noemen. Het is overigens wel zo, dat bij de codering een paar belangrijke beslissingen genomen moeten worden. In de eerste plaats de brugklassen. Tesser codeert de onderwijspositie van leerlingen die in de brugklas zitten door te kijken naar de vorm van voortgezet onderwijs die ze na de brugperiode kiezen. De brugklas fungeert dus zowel als MAVO1, HAVO1, als VWO1, en de toestand 'brugklas' komt niet voor. Een tweede probleem zijn de schoolverlaters. Zij behouden, in de jaren nadat ze de school verlaten, de laatste code voor het vertrek.

In ieder geval is het na de keuzen van de coderingen mogelijk de schoolloopbaan van iedere leerling vast te leggen in een rij van zes labels, waarbij ieder van de labels één van de 32 mogelijke toestanden is. Op de resulterend data matrix f_i (t_j) wordt nu door Tesser homogeniteitsanalyse toegepast, en dat resulteert in een kwantificatie van de 32 toestanden voor ieder van de zes tijdstippen. In paragraaf 5.3, en vooral in appendix A, bespreken we in detail hoe homogeniteitsanalyse in het algemeen werkt. Omdat we zelf in dit artikel ook homogeniteitsanalyse toe gaan passen, zullen we nog genoeg gelegenheid hebben om terug te komen op de specifieke toepassing op schoolloopbanenonderzoek.

Voor het moment is het voldoende om de uitkomsten van Tesser's homogeniteitsanalyse kort te bespreken. De loopbaan LBO1-LBO2-LBO2-uit LBO zonder diploma wordt ge - kwantificeerd over 6 jaren als -1.23, -1.22, -1.11, -1.36, -1.34, -1.33. Hierbij is bijvoor - beeld -1.34 de kwantificatie van 'uit LBO zonder diploma' in 1981, terwijl -1.11 de kwan - tificatie van LBO2 in 1979 is. De rij van zes kwalitatieve labels wordt vervangen door een rij van zes in principe verschillende getallen. We kunnen nu natuurlijk ook gemakkelijk een globale maat voor de kwaliteit van de schoolloopbaan construeren. Neem gewoon het gemiddelde van de kwantificaties van de zes posities die een leerling doorlopen heeft. In het voorbeeld vinden we voor LBO1-LBO2-LBO2- uit LBO zonder diploma een score van -1.27. Het is duidelijk dat de resultaten van een homogeniteitsanalyse vrijwel geheel gebaseerd zijn op de structuur van de loopbaan, en in feite op de structuur van alle loop - banen in het kohort of bestand. De kwaliteit van de loopbaan telt in principe in het geheel niet mee, alle informatie die we aan het programma geven zijn de labels van de loopbanen, zonder dat we daar van te voren één of andere ordening in aanbrengen. De ordening wordt als het ware aangebracht door de structuur. Het is van belang om op te merken dat de wijze van coderen van de diverse stadia in de loopbaan natuurlijk van belang is.

In latere delen van zijn hoofdstuk 4 gebruikt Tesser de zes met behulp van homogeniteits - analyse geconstrueerde numerieke loopbaanvariabelen in allerlei nadere analyses. Ze wor - den gerelateerd aan achtergrondsvariabelen, en ze worden zelfs gebruikt in een padmodel dat de gehele schoolloopbaan beschrijft. Als belangrijke uitkomst van deze padanalyse noe - men we Tesser's conclusie dat de achtergrond geen rol speelt in de schoolloopbaan gegeven het tussentijds niveau in de eerste klas VO. Het lijkt erop alsof milieu en capacitei - ten slechts een rol spelen bij de keuze van het type secondair onderwijs. Wanneer deze keuze eenmaal gemaakt is, verloopt de schoolloopbaan verder autonoom, dat wil zeggen onafhankelijk van de achtergrond. Ook in de latere fase van zijn onderzoek spelen de schoolniveau variabelen, met name de eindniveau variabele corresponderend met het jaar 1982, een belangrijke rol.

Onafhankelijk van Tesser werd homogeniteits analyse ook gebruikt om de schoolloopbaan te kwantificeren door Kreft en Bronkhorst (1985). Hoewel hun werkwijze veel op die van Tesser lijkt, zijn er toch ook enige belangrijke verschillen. Kreft en Bronkhorst gebruiken ook een deelbestand van het CBS-SMVO cohort, maar ze hebben leerlingen gekozen die hun loopbaan in het VO in een brugklas op een scholengemeenschap voor MAVO-HAVO-VWO beginnen. Dat blijken 5464 leerlingen op 114 scholengemeenschappen te zijn. Om rekentechnische redenen trokken ze hieruit een deelbestand van 2293 leerlingen. Kreft en Bronkhorst coderen op dezelfde manier als Tesser. Het later gekozen schooltype wordt opgezocht, en er wordt teruggecodeerd met dit schooltype als uitgangspunt. Iemand die na een drie-jarige brugperiode naar HAVO4 gaat, krijgt dus een loopbaan HAVO1-HAVO2-HAVO3-HAVO4. Het spreekt vanzelf dat deze codeerbeslissing in het geval van Kreft en Bronkhorst van aanzienlijk groter belang is dan in het geval van Tesser, omdat in het Kreft en Bronkhorst deelbestand alle leerlingen in een brugklas begonnen. Om deze codeerbeslissing toch geen al te grote rol te laten spelen analyseren Kreft en Bronkhorst alleen de jaren 1978-1982, dat wil zeggen dat ze het eerste jaar niet mee laten doen. Een tweede belangrijke beslissing is, wat er met de uitstroom gedaan moet worden. Tesser codeerde door, en gebruikte het laatst bereikte niveau daarvoor. Kreft en Bronkhorst beschouwen alle uitstromers vanaf het moment van uitstroom als ontbrekende gegevens. Omdat we in het vervolg van dit artikel de procedure van Kreft en Bronkhorst zullen volgen, zullen we verderop ook de gevolgen, of de mogelijke gevolgen, van dit soort beslissingen nagaan. Een laatste belangrijk verschil met Tesser is, dat Kreft en Bronkhorst in hun latere analyses naar schooleffecten niet niveau 1982 maar de gemiddelde positie als schoolloopbaan variabele gebruiken. Ook deze keuze heeft natuurlijk bepaalde consequenties.

4 SCHOOLLOOPBANEN IN EEN SMVO-DEELBESTAND

4.1 Het SMVO-bestand

Het CBS-SMVO bestand is een gestratificeerde twee-traps clustersteekproef uit alle leerlingen die deelnemen aan het eerste jaar van het Nederlandse voorgezet onderwijs in het jaar 1977. De steekproef bevatte 37280 leerlingen. De leerlingen werden tijdens hun schoolloopbaan gevolgd. Wij maken gebruik van de versie van het bestand waarin informatie aanwezig is tot en met september 1982. Informatie over het SMVO-cohort en het SMVO-bestand is te vinden in CBS (1982b, 1982c). Deze publikaties behandelen de schoolloopbaan tot en met onderwijselement 1980. Informatie over de later toegevoegde variabelen onderwijselement 1981 en onderwijselement 1982 staat in CBS (1983, 1984). Het aantal secundaire analyses op het CBS-SMVO bestand is ondertussen al te groot om op te noemen. Wij noemen slechts de analyses in de CBS-publikaties zelf, en een aantal publikaties die zich speciaal met het verloop van de schoolloopbaan bezig houden. Dit zijn Tesser (1985), Kreft en Bronkhorst (1985), Faasse, Bakker, Dronkers, en Schijf (1985). Al deze artikelen hebben we reeds besproken of ten minste genoemd in ons hoofdstuk 3.

4.2 Selectie van het deelbestand

We selecteerden in eerste instantie op onderwijselement 1977, dit moet gelijk zijn aan brugklas. Er bleven 15329 leerlingen over. Vervolgens, via de schoolnummers, op scholengemeenschappen voor MAVO-HAVO-VWO. Er bleven 5464 leerlingen op 114 scholengemeenschappen over. De redenen voor een dergelijke selectie zijn dezelfde als bij Kreft en Bronkhorst (1985, pagina 5). Om het aantal mogelijke schoolloopbanen te beperken, en een redelijk homogene groep van scholen te bestuderen, leek het onjuist om categoriale scholen en nog bredere scholengemeenschappen in het bestand op te nemen. Bij smallere scholengemeenschappen (eigenlijk alleen HAVO-VWO) is er scherpere beginselectie, en mogelijk een te groot verlies in variantie in een aantal belangrijke variabelen. We bedanken overigens Hans Bronkhorst (SCO, Amsterdam) voor de selectie van het deelbestand, en voor het verzorgen van een aantal hercoderingen.

4.3 Codering van achtergrondsvariabelen

Als achtergrondsvariabelen gebruiken we in onze analyses geslacht, CITO-score taaltoets, CITO-score rekentoets, provincie waarin de school gevestigd is, sociaal milieu, en advies van de onderwijzer omtrent de meest geschikte vorm van voortgezet onderwijs. De CITO-scores hoeven niet gehercodeerd te worden, we gebruiken gewoon het aantal goed beantwoorde items op de toets. Voor taal ligt dat aantal tussen de 11 en de 45 goede antwoorden, voor 940 mensen ontbreken gegevens. Voor rekenen ligt het aantal goede antwoorden tussen de 1 en de 25, met 941 ontbrekende gegevens. Het gemiddelde op de taaltoets is 31.5, met een standaarddeviatie van 6.2. Op de rekentoets heeft ons deelbestand een gemiddelde van 14.7, met een standaarddevariatie van 5.1. De gemiddelden van de totale CBS-steekproef zijn, respectievelijk, 27.2 en 12.1, wat natuurlijk een heel stuk lager is. In ons deelbestand zitten voor 48.2% jongens en voor 51.8% meisjes. Voor het gehele SMVO-kohort is dit, respectievelijk 48.5% en 51.5%. De verdeling op provincie, sociaal milieu, en advies onderwijzer vergt wat meer studie. Voor dat doel hebben we tabel 1 gemaakt.

De belangrijkste verschillen in tabel 1 tussen ons deelbestand en het gehele CBS-kohort treden op in advies. Dat is natuurlijk logisch, omdat in ons bestand vrijwel geen mensen met LBO advies zitten. Alleen diegenen, die ondanks een LBO advies toch naar een MAVO-HAVO-VWO scholengemeenschap gegaan zijn, en dat zijn er weinig (53, om precies te zijn). Wat de milieu-samenstelling betreft, gebeurt er ook min of meer wat we mogen verwachten. De employes zijn oververtegenwoordigd, hoe hoger hoe meer oververtegenwoordigd zelfs. Zelfstandigen zijn ondervertegenwoordigd, vooral zelfstandigen zonder personeel. Arbeiders zijn ook ondervertegenwoordigd, en natuurlijk vooral ongeschoolde arbeiders. De categorie 'overigen', die nader beschreven wordt in CBS (1982c, hoofdstuk 9) is eveneens ondervertegenwoordigd. Dit komt overeen met eerdere bevindingen dat deze categorie zich aan de lage kant van de sociaal milieu schaal bevindt. Bij de verdeling over provincies valt op dat Friesland en Limburg sterk ondervertegenwoordigd zijn. Drente, Zuid-Holland, en Zeeland zijn oververtegenwoordigd. De reden is, dat in Friesland en Limburg slechts 7.7% en 6.9% voor de schoolsoort 'brugklas met MAVO' kiezen, terwijl dit in Zuid-Holland, Zeeland en Drente respectievelijk 27.5%, 27.1%, en 24.0% is. Vergelijk CBS (1982c, staat 5.7, pagina 87). De CBS-cijfers zijn berekend over de populatie, waarin totaal zo'n 17% voor deze schoolsoort kiest. In de SMVO-steekproef is dit ongeveer 20%.

Samenvattend kunen we zeggen dat onze selectie van MAVO-HAVO-VWO scholengemeenschappen tot gevolg heeft gehad dat de achtergrondsvariabelen in het deelbestand aanzienlijk anders verdeeld zijn dan die in de CBS-steekproef. In ons deelbestand is een fikse oververtegenwoordiging van de 'hogere' milieus, van de VWO/HAVO adviezen, van de hogere CITO-scores, en van provincies als Zuid-Holland, waar blijkbaar veel van dit soort scholengemeenschappen zijn. Dit betekent vanzelfsprekend, dat alle uitspraken in dit artikel gaan over een geselecteerde deelpopulatie van de generatie 1977, en zeker niet over de gehele populatie van leerlingen in het secundaire onderwijs. Wanneer we ons deelbestand vergelijken met de deelbestanden die gebruikt zijn in het andere, meest direct vergelijkbare, schoolloopbanenonderzoek dat we besproken hebben, dat vinden we toch nog aanzienlijke verschillen.

In Roeleveld, Van der Eeden, en De Jong (1985) worden in principe dezelfde leerlingen uit het SMVO-bestand geselecteerd. Kleine scholen, en scholen waarin minder dan 15% van de leerlingen een CITO-toets gedaan heeft, worden echter verwijderd. Er blijven daardoor slechts 4142 leerlingen over. De marginalen verschillen weinig van onze marginalen in tabel 1. Koopman, Van der Eeden, en De Jong (1985) gebruiken het Amsterdamse VOCAL-bestand, jaargang 1975, en wel alleen leerlingen op categoriale MAVO's. Van enige vergelijkbaarheid is natuurlijk geen sprake. Faasse, Bakker, Dronkers en Schijf (1985) gebruiken alleen de Noord-Brabantse leerlingen uit SMVO, omdat ze hun resultaten willen vergelijken met die van een ouder Noord-Brabants cohort. Bosker, Hofman, en Van der Velden (1985), vergelijk ook Bosker en Van der Velden (1985), gebruiken het Groningencohort van Meijnen. Bij dit soort regionale cohorten is het natuurlijk niet helemaal reëel om representativiteit te onderzoeken. Te meer omdat dikwijls de coderingen van de variabelen in verschillende cohorten nogal verschillen. Tesser (1985) gebruikt, zoals we gezien hebben, ook het SMVO-cohort. Zijn deelbestand is gedefinieerd door een nogal ingewikkelde trapsgewijze steekproeftrekking, met daar overheen een niet al te succescolle machtigingsprocedure. Het resultaat zijn 4 deelbestanden van ieder ongeveer 250 leerlingen, met als eerste bekende schooltype LBO, MAVO, HAVO, en VWO. Bij representativiteits onderzoek constateert Tesser een oververtegenwoordiging van de hogere milieus in zijn deelbestand. Een deel van deze oververtegenwoordiging komt door inmiddels beruchte categorie 'overig', waarvan de exacte betekenis en codering eigenlijk niet meer te achterhalen is. Kreft en Bronkhorst (1985), tenslotte, gebruiken een steekproef uit het deelbestand waar wij mee werken. Zij kiezen willekeurig 2293 leerlingen uit onze 5464. De marginalen in hun onderzoek en het onze komen goed overeen.

4.4 Codering van de Loopbaanvariabelen

We volgen de procedure van Kreft en Bronkhorst (1985). Dat wil zeggen dat we drie belangrijke coderingsbeslissingen nemen. In de eerste plaats coderen we terug, dat wil zeggen dat we de brugperiode coderen met gebruikmaking van informatie uit latere jaren. Een loopbaan brugjaar 1, brugjaar 2, brugjaar 2, HAVO 3 wordt dus HAVO 1, HAVO 2, HAVO 2, HAVO 3. Het effect van de codering is dat het aantal mogelijke loopbaanpatronen aanzienlijk teruggebracht wordt. De codering bevordert de homogeneiteit van loopbanen, omdat het schooltype na de brugperiode gebruikt wordt om te bepalen dat de leerling tijdens de brugperiode ook continu op dit schooltype gezeten heeft. In een latere paragraaf zullen we trachten uiteen te zetten wat voor invloed dit soort coderingsbeslis - singen hebben op de kwantificatie van de schoolloopbaan. Een tweede belangrijke code - ringsbeslissing is, dat we de toestand in leerjaar 1 niet coderen. Iedereen zit in het eerste brugjaar, dat wil zeggen feitelijk in dezelfde toestand. Terugcoderen van het latere school - type zou natuurlijk wel differentiatie aanbrengen, maar die zou geheel op codering berusten, te meer natuurlijk omdat ook feitelijk het eerste brugjaar dikwijls ongedeeld is. Dit zou, ons inziens, te veel homogeniteit in de loopbaan suggereren. De derde beslissing is om individuen die uitstromen vanaf het moment van uitstroom te coderen als missing. Voor een deel van hen is dit gemakkelijk te verdedigen, die zijn inderdaad zoek geraakt en niet verder gevolgd. Per september 1982 zijn er van de oorspronkelijke CBS-steekproef 13000 mensen vertrokken, dus ongeveer 35%. De redenen staan, kort aangeduid, in CBS (1984). Van degenen die vertrokken zijn is 4% overleden of naar het buitenland of ziek, 56% is werkend of werkzoekend, 30% is om een andere reden uitgevallen, en van 10% is de reden niet bekend. In CBS (1984, tabel 4) vinden we dat van de leerlingen die in een brugklas beginnen in 1982 19% uitgevallen is. Het gaat daarbij natuurlijk om een paar bredere en om veel smallere scholengemeenschappen dan onze MAVO-HAVO-VWO scho - len. Het is daarom niet verwonderlijk dat ons bestand een uitval heeft van 21% in 1982. We hebben geen informatie over de redenen van vertrek in onze analyse betrokken.

Nogmaals een korte vergelijking met de methode gevolgd door Tesser (1985). Ook Tesser codeert terug, hij hoeft het echter in veel mindere mate te doen omdat hij in principe het meest in categoriale scholen geinteresseerd is. Om dezelfde reden neemt Tesser het eerste leerjaar gewoon op als onderdeel van de schoolloopbaan. Het grootste verschil tussen zijn benadering en de onze zit hem in het feit, dat Tesser doorcodeert. Nadat een leerling vertrokken is, houdt hij op de overblijvende momenten zijn laatste code. Dus HAVO 1, HAVO 2, verdwenen wordt HAVO 1, HAVO 2, HAVO 2, HAVO 2, HAVO 2, HAVO 2. Het is daarom natuurlijk zuiverder om niet te spreken van de schoolloopbaan, maar van het hoogste bereikte onderwijsniveau. Of, beter nog, van het laatst bekende onderwijsniveau. Het spreekt min of meer vanzelf dat deze vorm van doorcoderen kwalitatief veel homogeniteit in de loopbaan introduceert die er eerst niet was, en die eigenlijk in veel gevallen nogal aanvechtbaar is. We weten vaak niet waarom een leerling verdwijnt, of hij nog onderwijs volgt, misschien nog curcussen of vakopleidingen loopt, enzovoorts. Het lijkt ons dat de compactheid van de loopbanen die geintroduceerd wordt met doorcoderen aanzienlijk groter, en aanzienlijk problematischer is, dan de compactheid geintroduceerd met terugcoderen (vooral als men het eerste leerjaar weglaat). Gegeven deze codeerbe - slissingen ligt het nogal voor de hand welke toestanden overblijven om de loopbaan te be - schrijven. We hebben MAVO 1 t/m MAVO 4, HAVO 1 t/m HAVO 5, VWO 1 t/m VWO 6. Dit zijn de basistoestanden. Daarnaast hebben we categorieen nodig voor leerlingen die vanuit de scholengemeenschappen doorstromen (men zegt ook wel afstromen) naar het LBO. Dit zijn er niet erg veel, en we onderscheiden daarom alleen LBO met nog geen diploma en LBO met diploma. Bij de MAVO maken we ondrscheid tussen MAVO 3+ en MAVO 3-, en tussen MAVO 4+ en MAVO 4-, naar gelang de leerling al dan niet een diploma gehaald heeft op het desbetreffende tijdstip. Ook Tesser (1985) gebruikt het examenresultaat bij de codering van de loopbaan. Een extra probleem zijn de leerlingen die doorstromen naar één of andere vorm van beroepsonderwijs. We onderscheiden kort MBO, MBO 1 t/m MBO 3, en HBO 1 als mogelijke toestanden in ons bestand. Gezamelijk leidt dit tot 25 toestanden, die we samen met hun frequenties van voorkomen op de diverse momenten weergeven in tabel 2. We gebruiken dezelfde coderingen als Kreft en Bronkhorst (1985), en de marginalen verschillen dus ook maar marginaal. Tabel 2 geeft weinig aanleiding tot opmerkingen, of het moest zijn dat er één persoon in 1982 nog (of weer) in HAVO 1 zit.

5 KWANTIFICATIE VAN DE SCHOOLLOOPBAAN

5.1 Inleiding

In hoofdstuk 3 hebben we al aangegeven dat er naast vele manieren om de schoolloopbaan te coderen ook verschillende manieren zijn om de schoolloopbaan te kwantificeren. Het is misschien niet overbodig om erop te wijzen, dat het bij codering en kwantificatie gaat om twee verschijningsvormen van hetzelfde proces. Uitgangspunt is dat er ongelooflijk veel verschillende loopbanen zijn, veel meer nog in principe dan in de SMVO steekproef voorkomen. De variabele 'schoolloopbaan' neemt daardoor duizenden verschillende waar - den aan, en dit maakt het onmogelijk deze variabele in ongereduceerde vorm te relateren aan achtergrondsvariabelen. Er is dus reductie nodig, en reductie betekent in dit verband het terugbrengen van het aantal verschillende waarden dat de variabele aanneemt of kan aannemen.

Wanneer we reductie in deze zin gaan toepassen, moeten we in de gaten houden dat toe - kennen van een zelfde categorie of waarde aan twee verschillende loopbanen betekent, dat we in het vervolg van het onderzoek deze twee verschillende loopbanen als identiek zullen beschouwen. Het is duidelijk, dat we er dus naar moeten streven om homogene blokken loopbanen tot categorieen te aggregeren, waarbij onze ideeën over wat homogeen is en wat niet gedeeltelijk bepaald zullen worden door onze vraagstelling en onze theoretische uit - gangspunten. In het overzicht van de diverse procedures hebben we twee fases van redu - cering onderscheiden. In de eerste plaats kan men het aantal nominale categorieen waarin de schoolloopbaan beschreven wordt terugbrengen, en in de tweede plaats kan men, op basis van de kwalitatieve reductie in de eerste fase, een kwantitatieve reductie toepassen door aan de diverse categorieen getalwaarden toe te kennen. Het is duidelijk dat de kwalita - tieve reductie de meest fundamentele is, en dat veel onderzoekers het ook bij kwalitatieve reductie laten.

Bij de kwalitatieve reducties kunnen we onderscheid maken tussen variabelen die de gehele schoolloopbaan of aspecten van de gehele schoolloopbaan in één enkele variabele proberen te geven. De drievoudige codering van Diederen (1983), CBS (1982b), en Faasse, Bakker, Dronkers, en Schijf (1985), is hier een goed voorbeeld van. Ook de variabele bereikt eindniveau, is een globale maat (van kwaliteit, in dit geval). Coderen van de loopbaan als kwalitatief stochastisch proces, door de tijdstippen waarop gemeten wordt als variabelen te coderen, zagen we al bij Peschar (1975) in eenvoudige vorm, gekoppeld aan een a priori kwantificatie methode. Meester en De Leeuw (1983) beschrijven de loopbaan van SMVO-cohort leerlingen vooral door de toestand in 1977 te vergelijken met die in 1981. Faasse, Bakker, Dronkers, en Schijf (1985) coderen ieder tijdstip door middel van schooltype, en vinden daardoor een longitudinale meting van kwaliteit (met iets van structuur, en weinig van duur). Leerjaar speelt hierbij geen rol, en de reductie is daardoor aanzienlijk. Hetzelfde geldt voor de stroomdiagrammen in CBS (1982a, 1982) en in Kropman en Collaris (1974), Collaris en Kropman (1978). Wanneer er te veel toestanden gehandhaafd blijven, dan zijn de stroomdiagrammen onoverzichtelijk en in feite slechts een omslachtige weergave van de ruwe gegevens.

Kwalitatieve reducties die rekening houden met schooltype, leerjaar, uitstroom, en einddiploma zijn we ook al tegengekomen. Deze vorm van reductie, die weinig reducerend is, wordt gebruikt door Bosker e.a. (1984, 1985), Tesser (1985), Koopman, Van der Velden en De Jong (1985), Roeleveld, Van der Eeden, en De Jong (1985), Kreft en Bronkhorst (1985), en, in navolging van de laatsten, door ons. Wij komen op 25 categorieen voor ieder van de vijf tijdstippen uit, dat wil zeggen dat we ruimte hebben voor 25^5 ofwel ongeveer 10 miljoen verschillende loopbanen. Er is, in de eerste fase, dus niet zo zeer sprake van reductie. Daardoor is echter de reductie als het ware uitgesteld, uiteindelijk komt de reductie toch op het tijdstip van de kwantificatie. Al het onderzoek dat zo weinig reduceert in de kwalitatieve fase, maakt gebruik van kwantificatie in de één of andere vorm. Kwantificatie heeft als voordeel dat er op a priori niveau weinig beslissingen genomen hoeven te worden. De benadering is dus relatief theoriearm. In bepaalde kringen wordt dit wel als een nadeel van (empiristische, exploratieve, data masserende) benaderingen beschouwd. Volgens ons is het zo, dat sociaal wetenschappelijke theorie over schoolloopbaanprocessen dusdanig verward en onderontwikkeld is dat men zich er het beste ver van kan houden. Het is onvermijdelijk dat in schoolloopbaanonderzoek vooroordelen een relatief grote rol spelen. Wij zouden er voor willen pleiten om toch te proberen ze in een zo laat mogelijk stadium van het onderzoek te voorschijn te halen. Een ander voordeel van kwantificatie is, vanzelfsprekend, directe vergelijkbaarheid met andere variabelen en andere onderzoeken, en de mogelijkheid tot gebruik van relatief eenvoudige rekenprocedures.

Wellicht ten overvloede nog het volgende. Dat wij voor een kwantificerende benadering kiezen, betekent in geen enkel opzicht dat de reducties die bijvoorbeeld het CBS gebruikt (schooltype, opstromen of afstromen, al dan niet vertraagd) niet nuttig of niet goed zouden zijn. Voor sommige doeleinden zijn ze niet nuttig, er kan bijvoorbeeld niet mee doorgerekend worden in de gebruikelijke zin. Voor andere doeleinden, bijvoorbeeld voor tabellaire of log-lineaire analyse, zijn ze daarentegen superieur. En in de tweede plaats willen we niet beweren dat kwantificatie altijd te verdedigen is, en dat daarbij geen kwaliteitsverschillen of theoretische uitgangspunten een rol kunnen spelen.

5.2 A priori kwantificatie

Vooraf bestaande theorie kan een rol spelen bij de keuze van kwantificaties. We hebben gezien dat Peschar (1975), en in navolging van hem Bosman, Louwes, en Van der Meer (1980), de ITS-schaal voor opleidingsniveau gebruikte. Dit is een variabele die negen waarden aanneemt, gecodeert als 1 t/m 9. In categorie 5 valt bijvoorbeeld een afgemaakte ULO opleiding en enige jaren VHMO, in categorie 8 valt een niet-afgemaakte universitaire opleiding of MO-studie, en een voltooide HBO opleiding. Zie bijvoorbeeld Bosman. Louwes, en Van der Meer (1980, bijlage 2, pagina 182). Het is moeilijk om aan te geven welke theorie hier nu precies een rol bij speelt, het is eerder zo dat een partiele ordening van kwaliteit en duur op nogal willekeurige wijze omgezet is in een numerieke schaal.

In 3.8 hebben we gezien dat Bosker, Hofman, en Van der Velden (1985) en Bosker en Van der Velden (1985) een systematische manier van scoring bedacht hebben voor de variabelen eindniveau en tussentijds niveau, samengesteld uit additieve combinaties van schalen voor kwaliteit en duur. De door Roeleveld, Van der Eeden, en De Jong (1985) voorgestelde aanpassingen van deze procedure zijn voor ons interessanter, en die bespreken we wat meer in detail. Iedere onderwijspositie (een combinatie van schooltype en leerjaar) krijgt een score tussen 10 en 90. Het recept is als volgt. Voor LBO 1, MAVO 1, HAVO 1, en VWO 1 worden, respectievelijk, 10, 20, 30, en 40 punten verstrekt. Brugklas 1 krijgt het gemiddelde der samenstellende typen. Dus brugklas 1 op een LBO-MAVO-HAVO-VWO school krijgt 25, en brugklas 1 op een VWO-HAVO school 35. Voor elk volgend jaar op een zelfde schooltype wordt de onderwijspositie 10 hoger. Een score van 40 krijgen bijvoorbeeld LBO 4, MAVO 3, HAVO 2, VWO 1, HAVO-MAVO-LBO brugklas 3, VWO-HAVO-MAVO brugklas 2. Er komen theoretisch slechts 14 verschillende scores voor, met 5 meetpunten ongeveer 500.000 verschillende loopbanen. De verdeling van voorkomen van de 14 categorieen over de jaren (analoog aan onze tabel 2) staat in bijlage 1 van Roeleveld, Van der Eeden, en De Jong (1985).

Dit is een interessante vorm van a priori kwantificatie. De methode is opgebouwd rond het idee dat iemand die van een lager naar een hoger schooltype wil in het Nederlandse systeem een jaar verliest. Daarom is HAVO 2 hetzelfde als MAVO 3, en als VWO 1. Dit zou men een theoretische overweging kunnen noemen, maar een beetje reflectie laat direct de beperkte toepasbaarheid van dit soort theorie zien. Vooral in het begin van de loopbaan hoeven overgangen van HAVO naar VWO geen jaar te kosten. Het is evenmin zo, dat men een jaar wint wanneer men van een hoger naar een lager schooltype gaat. De diverse vormen van beroepsonderwijs zijn vrijwel niet in de theorie te passen. Het gebruik van hoger en lager, en het gelijkstellen van de intervallen in de codering, suggereren veel meer a priori kennis dan er in feite is. Blijven zitten is hetzelfde als zonder vertraging doorstromen naar een lager type: VWO 3 en dan VWO 3 is hetzelfde als VWO 3 en dan HAVO 4 (het is ook het zelfde als MBO 1 - MBO 1, of als HAVO 4 - MBO 1). Rest ons te vermelden dat Roeleveld, Van der Eeden, en De Jong, evenals Tesser, doorcoderen na vertrek. Wanneer vertrokken is met diploma krijgen de leerlingen 10 punten extra, wanneer ze ziek, overleden, of geemigreerd zijn worden ze missing. Merk op dat we in zekere zin weer terug zijn bij een methode als in 3.2 bespoken werd. De onderwijsposities, en daardoor de loopbanen, worden gekwantificeerd door experts. Alleen zijn de experts nu een klein aantal onderzoekers, niet meer een groot aantal leerkrachten. En de uiteindelijke schaal is niet meer gebaseerd op de mate van overeenstemming tussen experts, maar op de veronderstelling van gelijke intervallen en additiviteit van type en jaarklas. Wat er gebeurd is op geen enkele manier fout natuurlijk, de gevonden maat voor de loopbaan is zeker het bestuderen waard. Maar de a priori gronden waarop de maat gebaseerd is zijn aanvechtbaar, en eigenlijk op geen enkele manier te verdedigen.

5.3 Homogeniteitsanalyse

In dit rapport gebruiken we homogeniteitsanalyse om de schoolloopbaan te kwantificeren, in navolging van Tesser (1985) en Kreft en Bronkhorst (1985). We bedoelen dat niet als een procedure die superieur is aan a priori kwantificatie, of zelfs aan kwalitatieve beschrijving van typische of frequente loopbanen, maar als een alternatieve methoden die uitsluitend gebaseerd is op de overgangen in de gegevens zelfs. En op de keuze van een specifieke data analytische techniek, natuurlijk.

In 3.9 hebben we, in een historische context, het werk van Tesser en van Kreft en Bronkhorst al even kort besproken. In deze paragraaf gaan we eerst wat nader op het algemene idee achter homogeniteitsanalyse in. Voor een meer technische discussie, met referenties naar de data analytische literatuur, verwijzen we naar appendix A.

Er zijn tenminste twee verschillende manieren om homogeniteitsanalyse te introduceren. De erste gaat als volgt. Stel we hebben n leerlingen, en de schoolloopbaan is op m tijdstippen $t_1, ..., t_m$ gemeten. Er zijn dus n functies f_i, en de waarden $f_i\,(t_1)$ zijn elementen van de verzameling S. In ons voorbeeld is n gelijk aan 5464, m is gelijk aan 5, en S heeft de 25 mogelijke toestanden als elementen. Uitgangspunten bij homogeniteitsanalyse is dat we de loopbaan van een leerling willen beschrijven met één enkel getal. Voor leerling i is dat getal, de schaalwaarde van de schoolloopbaan x_i. Hoe kiezen we de 5464 getallen x_i? Dat gebeurt optimaal, waarbij optimaliteit gedefinieert wordt in termen van homogeniteit.

Stel we kiezen de 5464 getallen x_i op een willekeurige wijze. We willen eerst bepalen hoe we de kwaliteit van deze keuze kunnen meten. Daartoe gebruiken we erst de informatie die we hebben op tijdstip t_1 (in 1978 dus). We kijken naar de scores x_i van personen die in 1978 in dezelfde schoolloopbaancategorie zitten, en naar de scores van personen die in verschillende schoolloopbaancategorieen zitten. En we willen dat scores van mensen in dezelfde categorie relatief weinig van elkaar verschillen, terwijl scores van mensen in verschillende categorieen relatief veel van elkaar verschillen. In variantie analytische termi - nologie: we willen dat de tussen-categorieen variantie groot is ten opzichte van de totale variantie. Idealiter willen we dat iedereen in MAVO 1 dezelfde score heeft, iedereen in HAVO 1 ook dezelfde score, enzovoorts. In de praktijk stellen we de bescheidener eis dat de scores van de MAVO 1 leerlingen met weinig spreiding gegroepeerd om het MAVO 1 gemiddelde liggen, de HAVO 1 scores met weinig spreiding om het HAVO 1 gemiddelde, enzovoorts, terwijl de 25 verschillende gemiddelden onderling zo veel mogelijk gespreid zijn. De proportie tussen-groepen variantie noemen we de **homogeniteit** (van de scores x_i op tijdstip t_1). De totale homogeniteit van de scores x_i is de gemiddelde homogeniteit over alle tijdstippen.

We hebben nu aangegeven hoe we van gegeven scores x_i kunnen vaststellen hoe homo - geen ze zijn. Homogeniteitsanalyse maakt de volgende logische stap: bereken de scores die zo homogeen mogelijk zijn. De formules staan in appendix A, we volstaan hier met een klein voorbeeld. Tabel 3 bevat 10 loopbanen, die gedefinieerd zijn door metingen op drie tijdstippen. De categorieen zijn LBO, MAVO, HAVO, VWO. Stel nu dat we onze analyse beginnen met aan de tien loopbanen de scores 1 tot en met 10 toe te kennen. Loopbaan 1 krijgt score 1, loopbaan 2 score 2, enzovoorts. Een volkomen uit de lucht gegrepen keuze, maar we zullen eens bekijken hoe het zit met de homogeniteit van deze scores. De eerste stap in de analyse is, om de scores in afwijkingen van het totaalgemiddelde te brengen en ze kwadratensom gelijk aan één te geven. Het is handig om met genormaliseerde scores te werken, omdat dan in ieder geval de totale variantie niet verandert. Op basis van de genormaliseerde x_i berekenen we nu de gemiddelden voor LBO, MAVO, HAVO, en VWO op ieder van de drie tijdstippen. Deze gemiddelden gebruiken we om een nieuwe gekwanti - ficeerde datamatriks te maken, die in tabel 4a staat. Onderaan 4a staan de kwadraten - sommen van de kolommen, die geven aan hoe groot de homogeniteit van de oor - spronkelijke scores was. In de laatste kolom staan de rijgemiddelden, met hun kwadraten - som. Wanneer we de rijgemiddelden hernormaliseren, zodat ze weer kwadratensom gelijk aan één hebben, dan vinden we een nieuwe x waarmee we procedure kunnen herhalen. Enzovoorts.

Dus nogmaals, voor de duidelijkheid. Bij onze eerste volstrekt willekeurige schatting van de loopbaanscores, in de eerste kolom van 4a, hebben de twee personen die op t=1 in het

LBO zitten scores -.495 en -.385. De LBO-score voor t=1 wordt daarom het gemiddelde van deze twee getallen, dus -.440. We vervangen in de datamatrix LBO in de kolom t=1 door -.440. Op t=3 zitten personen 2, 3, en 4 op de MAVO. Hun scores zijn -.385, -.275, en -.165, en de MAVO-score voor t=3 wordt daarom het gemiddelde van deze drie getal - len, te weten -.275. We zien, in de onderste rij van 4a, dat de homogeniteit van de drie tijd - stippen .567, .780, en .780 is. De gemiddelde homogeniteit, over de drie tijdstippen, is daarom .709. De kwadratensom van de rijgemiddelden in de laatste kolom is .649. Deze kwadratensom is altijd groter als het kwadraat van de gemiddelde homogeniteit, in ons geval $(.709)^2 = .503$.

In tabel 4b herhalen we de berekeningen met de nieuwe x_i, die we verkregen hebben door de laatste kolom van 4a op een kwadratensom van één te normeren. De homogeniteiten schieten omhoog, naar .869, .972, en .976, met een gemiddelde van .939. De kwadraten - som van de rijgemiddelden is .886, wat nog maar nauwelijks hoger is dan $(.939)^2 = .881$. We vinden dat de rijgemiddelden vrijwel proportioneel zijn aan de scores waarmee we de berekeningen in 4b begonnen, maar we herhalen het proces nog een aantal malen tot de tabel volledig stabiliseert. De uitkomst staat in 4c, de gemiddelde homogeniteit is maar liefst .948. We kunnen dus zeggen dat 95% van de variantie tussen schooltypen is, en maar 5% van de variantie is binnen schooltypen.

Tabel 4c is de uitkomst van de homogeniteitsanalyse. We hebben nu scores voor de loopbanen. Die staan, met kwadratensom 1.000, in kolom 1, en met kwadratensom $.898 = (.948)^2$ in kolom 5. We hebben ook kwantificaties voor de categorieen, dus voor de schooltypen op de verschillende tijdstippen. Zo is LBO op t=2 dus -.560, en HAVO op t=3 is .293. In figuur 1a hebben we kwantificaties voor de schooltypen uitgezet tegen de tijdstippen. We zien de typen in de tijd divergeren. Enig nadenken leert dat persoon 9, die van MAVO naar HAVO naar VWO gaat een nogal grote invloed op de oplossing heeft. Omdat hij in drie schooltypen voorkomt, krijgt hij een gemiddelde score. Daardoor haalt hij de MAVO-score in jaar t=1 omhoog, en de VWO-score in jaar t=3 omlaag. Dit is ook de reden waarom in jaar t=3 de HAVO-kwantificatie hoger is dan die van VWO. Bij kleine aantallen vinden we natuurlijk dikwijls een bijzonder hoge homogeniteit, en zo veel mogelijk kapitalisatie op de eigenaardigheden van opvallende individuen. In figuur 1b staat de loopbaan van een paar leerlingen getekend. Homogeniteitsanalyse streeft ernaar deze loopbaan-curven zo horizontaal mogelijk te maken. Voor sommige individuen lukt dit beter dan voor anderen, en met name voor i=9 gaat het niet al te best. De correlatie tussen de kolommen van 4c corresponderend met de drie tijdstippen is bijzonder hoog, kolom 1 correleert .886 met kolom 2 en .882 met kolom 3, kolom 2 correleert zelfs .990 met kolom 3. De grootste eigenwaarde van de correlatiematrix is, zoals altijd bij homogeniteitsanalyse, gelijk aan de gemiddelde homogeniteit, dus .948. De correlaties tussen de uiteindelijke loopbanen-schaal (de eerste of laatste kolom van tabel 4c) en de schalen voor de afzonderlijke tijdstippen (de middelste drie kolommen van 4c) zijn gelijk aan de wortels van de homogeniteiten van de tijdstippen, in dit geval dus $(.900)^{1/2} = .949$, $(.975)^{1/2} = .987$, en $(.971)^{1/2} = .985$.

Natuurlijk is het voor een klein voorbeeld zoals het bovenstaande niet echt nodig homogeniteitsanalyse toe te passen, of welke datareductie techniek dan ook. De tabel 3 is op zichzelf duidelijk genoeg. Maar met 5464 leerlingen is dat natuurlijk wel wat anders. Het is bijvoorbeeld onmogelijk om tussenresultaten, zoals in tabel 4, te laten zien. We volstaan met de kwantificaties van de 25 schooltypen op de 5 tijdstippen. Die staan in tabel 5, terwijl figuur 2 de verschuivingen van een bepaald schooltype in de tijd laat zien. Figuur 2 kan vergeleken worden met de figuur op pagina 14 van Kreft en Bronkhorst (1985). Op de interpretatie van de resultaten in de figuur komen we verderop terug. Voorlopig merken we alleen maar op, dat de cellen in tabel 5 en de grafieken in figuur 2 alleen opgenomen zijn wanneer ze gebaseerd zijn op frequenties groter dan 5. Deze grens is min of meer willekeurig, maar heeft in ieder geval tot gevolg dat de meest instabiele resultaten niet gegeven worden.

We hebben in dit hoofdstuk laten zien, dat homogeniteitsanalyse een voor de hand liggende manier is om schoolloopbanen te kwantificeren. Er is een relatief klein aantal a priori beslissingen nodig, hoewel het natuurlijk zo blijft dat de resultaten mede bepaald worden door de kwalitatieve codering van de tijdstippen die men kiest. Maar gegeven de codeerbeslissingen volgt de kwantificatie vanzelf. We zullen nu nog even, gewapend met de kennis van homogeniteitsanalyse die we inmiddels hebben, onze belangrijkste codeerbeslissingen de revue laten passeren. Het terugcoderen van de brugperiode zal waarschijnlijk tot gevolg hebben, dat de onderscheiden typen MAVO-HAVO-VWO duidelijker gediscrimineerd worden. Immers wanneer men na de brugperiode bijvoorbeeld HAVO kiest, dan wordt gedaan alsof men tijdens de brugperiode ook op de HAVO zat. Dus komen tijdens de brugperiode geen overgangen van het ene schooltype naar het andere voor, wat homogeniteit binnen typen vergroot. We compenseren dit effect gedeeltelijk, door het eerste leerjaar niet in de analyse te betrekken, maar het is niet duidelijk hoe de preciese invloed van terugcoderen is. De andere belangrijke codeerbeslissing die we genomen hebben, is om uitstromers te coderen als missing. Wanneer we ze hun laatste onderwijsniveau laten houden, zoals Tesser deed, dan zullen de individuele loopbaancurven voor het laatste deel van de loopbaan vrijwel horizontaal gaan verlopen, met andere woorden er wordt veel homogeniteit geintroduceerd. Na uitval verandert iemand niet alleen niet meer van schooltype, maar hij verandert zelfs niet meer van leerjaar binnen schooltype. Voor onze smaak introduceert dit artificieel een te grote dosis homogeniteit. Uit dezelfde overwegingen zouden we zelfs kunnen besluiten om ook de brugperiode als ontbrekende gegevens te beschouwen, en in feite hebben we dit al gedeeltelijk gedaan door het eerste leerjaar af te laten vallen.

6 RESULTATEN VAN EN NA HOMOGENITEITSANALYSE

6.1 Kwantificaties van schooltypen en leerjaren

Tabel 5 en figuur 2 verschillen maar weinig van de overeenkomstige resultaten in Kreft en Bronkhorst (1985). We kunnen daardoor hun belangrijkste conclusies overnemen. De LBO-grafiek ligt dicht bij de MAVO-grafiek, omdat in ons deelbestand de LBO-ers grotendeels afstromers uit het MAVO zijn. De curven van de vier belangrijkste schooltypen zijn goed onderscheiden, en nemen in het algemeen monotoon toe. Er zijn wat kleine uitzonderingen op de monotoniciteit, die verklaard kunnen worden door het instromen op bepaalde tijdstippen van leerlingen uit lagere niveaus. Dit gebeurt bijvoorbeeld bij VWO-5 in leerjaar 6 en bij HAVO-4 in leerjaar 5. Leerlingen uit respectievelijk HAVO en MAVO komen aan met een jaar vertraging. De onvertraagde VWO-ers steken ver boven de andere categorieen uit, wat natuurlijk logisch is omdat het onmogelijk is om uit een andere categorie in VWO-onvertraagd te komen. Dit is mogelijkerwijs ook de reden waarom het verschil tussen vertraagd en onvertraagd bij VWO groter is dan bij HAVO en MAVO. Figuur 2 wekt de indruk dat er toch relatief weinig van schooltype gewisseld wordt, en dat zelfs mensen met vertraging toch vooral bij hun eigen schooltype blijven behoren, en niet langzaam afzakken naar een lager type. Naarmate de loopbaan vordert is een opgelopen jaar vertraging dan ook minder belangrijk (in termen van de door ons berekende score, vanzelfsprekend). Figuur 2 laat ook zien, dat gebruik van de gehele loopbaan in combinatie met homogeniteitsanalyse slechts een kleine winst oplevert vergeleken met het gebruik van hetzij eindniveau, hetzij beginniveau. In termen van overgangen gebeurt er blijkbaar te weinig tijdens het secundair onderwijs.

6.2 Externe variabelen en de schoolloopbaan

Het heeft weinig zin om de 5464 scores op de over tijdstippen gemiddelde schoolloopbaanscore, of zelfs de 5464 scores op ieder van de vijf tijdstippen, in detail te gaan bekijken. We zullen ze eerst relateren aan een aantal belangrijke achtergrondsvariabelen. Daartoe standariseren we de vijf gekwantificeerde schoolloopbaanvariabelen, dus datgene wat in ons kleine voorbeeld het binnenwerk van tabel 4 vormde. Door te standariseren geven we ieder van de vijf variabelen variantie één.

We rekenen eerst de gemiddelden uit per provincie, en tegelijkertijd de correlatie-ratio voor provincie (de ratio van de tussen-provincie variantie en de totale variantie). De resultaten staan in de eerste deel-tabel van tabel 6. We zien Limburg, Overijsel, en Friesland nogal laag scoren, en Drente en Zeeland nogal hoog. In 4.3 hebben we gezien, dat in Friesland en Limburg weinig brede scholengemeeschappen zijn, zodat de betere leerlingen waar - schijnlijk vooral naar HAVO-VWO gemeenschappen of naar categoriale vormen van voort - gezet onderwijs gaan. Noord-Holland, Zuid-Holland, en Utrecht hebben de neiging relatief laag te beginnen en relatief hoog te eindigen. Dit lijkt erop te wijzen dat er bij die provincies minder uitval en afstroom is vanuit de brede scholengemeenschappen.

Het verschil tussen jongens en meisjes is niet bijzonder groot. Wel zitten meisjes gemiddeld altijd hoger. Het verschil in gemiddelde blijft ongeveer gelijk. We zien in de latere jaren overigens dat zowel jongens als meisjes een positief gemiddelde hebben. Dit lijkt in tegenspraak met het feit dat het gemiddelde over het gehele bestand nul is, maar de tegenspraak wordt verklaard doordat in de latere jaren veel scores ontbreken. Het gemid - delde is daardoor niet meer nul, en neigt naar het positieve, omdat veel kinderen met een negatieve score op eerdere jaren inmiddels uitgestroomd of uitgevallen zijn.

We zien, in de derde deeltabel van 6, een zeer groot effect van advies, wat over de jaren heen redelijk stabiel blijft. We zien wel dat in de loop der jaren leerlingen met MAVO-ad - vies meer stijgen dan leerlingen met HAVO of VWO advies, terwijl leerlingen met LBO advies (wat ze dus in de wind geslagen hebben) veruit het snelst stijgen wanneer ze vol - houden. Een dergelijk interactie effect zien we ook in de tabel met SES-gemiddelden, hoewel het daar veel minder uitgesproken is.

In het laatste deel van tabel 6 relateren we met behulp van multipele regressie de school - loopbaanvariabele aan de CITO-score. Taal heeft een wat hoger regressie gewicht dan rekenen, en over de jaren heen lijken de gewichten wat af te nemen. Dit laatste kan natuur - lijk een artefact zijn van het feit dat de steekproef steeds meer geselecteerd wordt, en dat hierdoor onder andere het intercept toeneemt. We zien dat CITO de loopbaan beter voor - spelt dan het advies van de onderwijzer.

6.3 De homogeniteit van de schoolloopbaan

De homogeniteit van de schoolloopbaan is maar liefst .918. Dus meer dan 90% van de va - riantie is tussen de 25 schooltypen, minder dan 10% is binnen schooltypen. Door kwanti - ficatie van de schooltypen in combinatie met leerjaren slagen we er redelijk goed in om de loopbanen van het overgrote deel der leerlingen behoorlijk af te beelden als horizontale rechte lijnen. De homogeniteit van de leerjaren is, respectievelijk, 1.006, 1.032, .963, .825, en .765. Dat de homogeniteit voor de eerste jaren boven de één ligt, komt door de manier waarop we ontbrekende gegevens behandelen. Hierdoor zijn de homogeniteits-maten geen ratio's van tussen-variantie en totaal-variantie meer. Het is wel mogelijk, met behulp van het programma PRIMALS, beschreven door Van der Geer en Meulman (1985), de correlaties tussen de vijf tijdstippen te schatten. Deze schattingen blijken te zijn:

1.000				
0.939	1.000			
0.907	0.947	1.000		
0.878	0.914	0.948	1.000	
0.834	0.871	0.905	0.944	1.000

De homogeniteitsmaten die op basis van deze correlatiematriks geschat kunnen worden, zijn de kwadraten van de ladingen op de eerste principale component. Deze zijn:

0.887	0.929	0.941	0.929	0.878

met een gemiddelde van .913. We zien dat de middelste jaren het meest homogeen zijn, wat niet erg verwonderlijk is, omdat de loopbaan daar het meest over de schoolsoorten gespreid is. Wanneer we de resultaten wat pessimistisch interpreteren, zouden we kunnen zeggen dat slechts 10% van de variantie in eindniveau van het secundaire onderwijs maxi-maal toegeschreven kan worden aan de schoolloopbaan, maar liefst 90% van de variantie ligt in feite vast met het beginniveau. Maar natuurlijk is een dergelijke interpretatie mede bepaald door het feit, dat we met onze scoring de loopbaan zo homogeen mogelijk gemaakt hebben. Verderop zullen we een aantal alternatieve, hoewel verwante, kwantificatiemetho-den kort bespreken die althans in principe tot andere resultaten en interpretaties zouden kunnen leiden.

6.4 Structuur van de schoolloopbaan

In deze paragraaf bekijken we diverse padmodellen die de structuur van de schoolloopbaan beschrijven. Voor een meer technische uiteenzetting over de statistische aspecten van de klasse modellen die we gebruiken verwijzen we naar appendix B. De variabelen die we gebruiken in onze modellen zijn de **achtergrond** en de vijf **schoolloopbaanvaria-belen**. We bekijken alleen de 3461 personen die op geen van de variabelen ontbrekende gegevens hebben. Dit is natuurlijk een belangrijke selectie, omdat het betekent dat we ons in feite beperken tot leerlingen die nog op school zitten. Dat zijn natuurlijk vooral

VWO'ers, maar ook HAVO-leerlingen met vertraging, en mensen die naar het beroeps - onderwijs gegaan zijn. Leerlingen die MAVO en HAVO afgemaakt hebben, en die daarna geen onderwijs meer volgen, zijn we kwijt. De achtergrond wordt overigens gecodeerd als 22 dummies, één voor geslacht, 11 voor provincie, 7 voor sociaal milieu, 4 voor advies, plus twee gemeten variabelen (CITO-scores), plus een intercept. In totaal dus met 25 para - meters.

Tabel 7 geeft de diverse multipele regressies die gefit zijn. In appendix B wordt aange - geven hoe complete modellen voor de loopbaan uit de diverse multipele regressies opge - bouwd kunnen worden. Het minst restrictieve globale model is MODEL A. Het beweert dat een loopbaanvariabele, waarvoor we overigens de codering uit homogeniteitsanalyse gebruiken, beinvloed wordt door de achtergrond en alle eerdere loopbaanvariabelen. MODEL B zegt dat een loopbaanvariabele slechts bepaald wordt door de achtergrond en de vorige loopbaanvariabele. MODEL C zegt dat een loopbaanvariabele bepaald wordt door alle vorige loopbaanvariabelen, maar niet door de achtergrond, behalve natuurlijk de positie in leerjaar 2. MODEL D is het meest restrictieve model. Het zegt dat alleen de vorige loopbaanvariabele een rol speelt, behalve bij positie in leerjaar 2, waarbij de achtergrond meedoet. In tabel 8a staan de multipele correlatie-coefficienten, en in tabellen 8b en 8c de chi-kwadraten met hun aantal vrijheidsgraden bij toetsing van de meer restrictieve modellen tegen het nulmodel A.

Hoewel vergelijken van een grote hoeveelheid modellen altijd een riskante bezigheid is, lijkt de conclusie toch vrij duidelijk. De achtergrond speelt geen echte rol meer in de schoolloopbaan op de middelbare school. Model C heeft een bijzonder goede fit (chi-kwadraat is 52.55 met 96 vrijheidsgraden). Model B past niet bijzonder goed, en ook het verschil tussen C en D (19.68 met 6 vrijheidsgraden) is niet onaanzienlijk. Dit lijkt erop te wijzen, dat de schoolloopbaan niet een eerste-orde proces is. Het is niet voldoende om alleen rekening te houden met de vorige klas, ook de eerdere fasen van de loopbaan moeten in de voorspelling betrokken worden.

7 CONCLUSIES

We hebben laten zien, dat er veel mogelijkheden zijn om de schoolloopbanen van leerlingen in het Nederlandse onderwijssysteem te beschrijven. Al dit soort beschrijvingen zijn gebaseerd op data reductie, omdat het aantal mogelijke patronen eenvoudigweg te groot is om ongereduceerd als categorieen van een variabele te kunnen dienen. We hebben beschrijvingen gezien in termen van stroomdiagrammen, overgangsmatriksen, en afgeleide variabelen. We hebben tevens gezien, dat verder rekenen met schoolloopbaanvariabelen, afgezien van hoe ze oorspronkelijk gecodeerd waren, alleen mogelijk is na kwantificatie. Deze kwantificatie kan gebruik maken van het oordeel van experts, van formele eigenschappen van het Nederlandse schoolsysteem, of van homogeniteitsanalyse. Homogeniteitsanalyse is in dit verband eerder gebruikt door Tesser (1985) en Kreft en Bronkhorst (1985). Wij geven, in dit rapport, een wat uitvoeriger inleiding bij het gebruik van homogeniteitsanalyse, en we passen de techniek op een ander deelbestand van het SMVO-bestand toe.

Het grote voordeel van homogeniteitsanalyse is, dat de techniek veel informatie levert voor relatief weinig a priori assumpties. Nadat we een codering van de loopbaan gekozen hebben, worden verder alle keuzes door de techniek gemaakt. Dit impliceert overigens, dat we zelf natuurlijk nog wel de techniek moeten kiezen. Een belangrijk punt, en wel om de volgende redenen. In homogeniteitsanalyse kiezen we kwantificaties van de loopbaancategorieen op zo'n manier dat de loopbanen zo homogeen mogelijk worden. Als functies van de tijd gezien, worden de gekwantificeerde loopbanen dus zo horizontaal mogelijk. Dit maakt het relatief eenvoudig ze met één enkel getal, hun gemiddelde niveau, te representeren. Het criterium dat we met onze keuze optimaliseren is dus volstrekt expliciet, een ander voordeel van de techniek boven gebruik van a priori scoringen of van het oordeel van experts. Dat neemt niet weg, dat we wel eens zouden kunnen vinden (voor sommige toepassingen of verdere analyses) dat dit criterium helemaal niet het juiste is. De resultaten zijn noodzakelijkerwijs wat conservatief, in de zin dat we veranderingen in de schoolloopbaan een zo klein mogelijke rol laten spelen. We bestuderen dus als het ware een soort benedengrens van de mogelijke variabiliteit in de loopbaan.

Er zijn, in principe, veel andere criteria mogelijk, die even goed en dikwijls met ongeveer dezelfde methodes geoptimaliseerd kunnen worden. We kunnen bijvoorbeeld de loopbaancategorieen op zo'n manier kiezen dat de resulterende variabelen zo sterk mogelijk samenhangen met de achtergrond. Dit leidt tot **Kwalitatieve Redundantie Analyse**, recent beschreven door Israels (1985). Met een kleine modificatie vinden we **Kwalitatie-**

ve Pad Analyse, waarin de multipele correlaties in tabel 8a zo groot mogelijk gemaakt worden. We kunnen ook scoren zodat de loopbaanvariabelen zo veel mogelijk lijken op eindniveau, of op beginniveau, of op zo'n manier dat de verschillen tussen scholen zo groot mogelijk worden. In plaats van het kiezen van a priori scoringen moeten we dus kie - zen voor een criterium wat we willen optimaliseren. Dit criterium kunnen we evenwel een - voudig aanpassen aan het doel waarvoor we de schoolloopbaanvariabele later nodig menen te hebben, en de keuze is volstrekt eenduidig en heeft ook volstrekt duidelijke consequen - ties. Dit soort duidelijkheid lijkt ons wenselijk. Discussie over criteria waaraan de loop - baankwantificaties moeten voldoen, lijkt ons aanzienlijk duidelijker dan discussie over vra - gen zoals 'Moeten we HAVO 3 nu dezelfde score geven als MAVO 4, en zo nee, waarom niet?'

APPENDIX A

SCHAALCONSTRUCTIE MET BEHULP VAN HOMOGENITEITSANALYSE

In de tekst, met name in paragraaf 5.3, hebben we homogeniteitsanalyse besproken, en aan de hand van een klein voorbeeld uitgelegd. We lieten daarbij zien hoe de oplossing bere - kend wordt met behulp van zogenaamd **beurtelings middelen** (reciprocal averaging). In deze appendix geven we dezelfde resultaten nogmaals, maar nu door formules te gebrui - ken.

Stel we hebben n individuen en m variabelen (of tijdstippen). In het kleine voorbeeld van tabel 3 geldt n = 10 en m = 3. Om homogeniteitsanalyse elegant te kunnen presenteren hebben we **indicator matrixen** nodig. Iedere variabele j definieert een indicator matrix G_j, met n rijen en k_j kolommen, waarbij k_j het aantal categorieen van variabele j is. De indicator matrix is **binair**, bestaat dus uit enen en nullen, op zo'n manier dat precies één element in iedere rij gelijk is aan één. De rest is nul. De plaats van het element gelijk aan één laat zien in welke categorie van variabele j het individu in de desbetreffende kolom scoort. In tabel 9 staan de drie indicatormatriksen voor ons kleine voorbeeld, naast elkaar.

Als G_1 de indicatormatrix is van variabele 1, G_2 de indicator matrix van variabele 2, dan is $C_{12} = G_1'G_2$ de **kruistabel** van variabele 1 en 2.
de matriksen $C_{11} = G_1'G_1$ en $C_{22} = G_2'G_2$ zijn diagonaal, bevatten de univariate marginalen (de **rechte tellingen**) behorend bij variabelen 1 en 2. Wanner we voor m variabelen de indicator matrixen naast elkaar zetten in één enkele matrix $G = (G_1| \ldots |G_m)$, dan heeft G n rijen en $k_1 + \ldots + k_m$ kolommen. De matrix $C = G'G$ is symmetrisch en vierkant, van de orde $k_1 + \ldots + k_m$, en C bevat als deelmatriksen alle bivariate marginalen (kruistabellen) en univariate marginalen. Zie tabel 10a.C wordt, in de literatuur van homogeniteitsanalyse, we de **Burt matrix** genoemd. Voor de diagonale matrix die alle univariate marginalen bevat, dus voor de diagonaal van C, gebruiken we het symbool D. Dus $D_1 = C_{11} = G_1'G_1$ en $D_2 = G_2'G_2$, enzovoorts. Zie tabel 10b.

Wanneer x een vector van n getallen is, dan kunnen we de gemiddelden van de categorieen van variabele j uitrekenen met $y_j = D_j^{-1}G_j'x$. Immers $G_j'x$ vormt de binnen-categorie kolomsommen, en met D_j^{-1} voorvermenigvuldigen maakt daar gemiddelden van. Wanneer we de categoriemiddelden weer tot lengte m willen opblazen, dan moeten we $G_jy_j = G_jD_j^{-1}G_j'x$ vormen. We berekenen dus eerst categorie gemiddelden voor variabele j, en vervangen vervolgens iedere score door het gemiddelde van de categorie waar hij in valt. Bij perfecte homogeniteit, waarbij alle individuen in een categorie dezelfde score hebben, vinden we dus $x = G_jD_j^{-1}G_j'x$, omdat in dit geval opblazen van de categorie gemiddelden geen wijziging van de scores met zich meebrengt. De matrix $P_j = G_jD_j^{-1}G_j'$ heet de **pro - jector** van variabele j. In tabel 11a, 11b, en 11c zien we de drie projectoren voor ons kleine voorbeeld. Ze zijn allemaal symmetrisch en idempotent. Dit laatste betekent dat $P_j^2 =$

P_j. Om dit in te zien, is het voldoende om je te realiseren dat voor iedere x de vector $P_j x$ perfect homogeen is. Als individuen in dezelfde categorie van variabele j scoren, dan hebben ze in $P_j x$ per definitie dezelfde score. En perfect homogene vectoren, hebben we gezien, worden door P_j niet veranderd. Dus $P_j(P_j x) = P_j x$, en daarom geldt dat $P_j^2 = P_j$.

Nog een andere relatie is van belang. Stel x is een vector van n scores, waarvan de elementen optellen tot nul (x staat **in afwijkingen van het gemiddelde**). Het spreekt vanzelf dat we x kunnen schrijven als $x = P_j x + (I - P_j)x$. Immers $P_j x + (I - P_j)x = P_j x + x - P_j x = x$. Het deel Pjx noemen we nu het tussen-categorieen deel van x, en het deel $(I - P_j)x$ het binnen-categorieen deel. Hierbij slaat 'categorieen' natuurlijk steeds op de categorieen van variabele j. Nogmaals: $P_j x$ vervangt iedere score door zijn groeps- of categoriegemiddelde, en $(I - P_j)x$ zet iedere score dus in afwijkingen van het groepsgemiddelde. Nu geldt dat $x'x = x'P_j x + x'(I - P_j)x$, wat we direct in kunnen zien door de ontbinding $x = P_j x + (I - P_j)x$ met x' voor te vermenigvuldigen. Dus: totale variantie van x is de som van de tussen-categorieen variantie en de binnen-categorieen variantie. Voor iedere variabele j is er een dergelijke opdeling. Als een vector x perfect homogeen is (voor variabele j), dan is de binnen-categorieen variantie (voor variabele j) gelijk aan nul, en dus in totaal gelijk aan tussen. Voor iedere variabele kunnen we de ratio van tussen en totaal gebruiken als maat voor homogeniteit.

Laten we nu eens, met onze nieuwe notatie, het beurtelings middelen uit paragraaf 5.3 be - kijken. We begonnen met een aanvangsschatting x_0, die voldeed aan $u'x_0 = 0$ en $x_0'x_0 = 1$. Hierbij gebruiken we u voor een vector met n elementen gelijk aan één, zodat u'x de som van de elementen van x is. Voor ieder van de drie tijdstippen berekenden we de gemiddelden voor LBO, MAVO, HAVO, en VWO. Dit gaat met $y_j = D_j^{-1}G_j'x$, de kolom - men van tabel 4a voor t=1, 2, 2 zijn dus de vectoren $G_j y_j$, dat wil zeggen P_1x_0, P_2x_0, en P_3x_0. De laatste kolom van 4a is het gemiddelde $(P_1x_0 + P_2x_0 + P_3x_0)/3$. Wanneer we P_* schrijven voor het gemiddelde van de P_j, dan is de laatste kolom dus P_*x_0. Deze laatste kolom wordt opnieuw genormeerd, en definieert onze nieuwe x, laten we zeggen x_1. Dus $x_1 = P_*x_0/(x_0'P_*^2x_0)^{1/2}$. De kolommen van tabel 4b zijn, achtereenvolgens , x_1, P_1x_1, P_2x_1, P_3x_1, en P_*x_1, en hierna berekenen we x_2 door P_*x_1 te normeren.

Wanner we het proces herhalen vinden we op het laatst een stationaire oplossing, dat wil zeggen een oplossing waarvoor P_*x proportioneel is met x. Dit betekent dat als we het algorrithme herhalen met deze x, dat dan de oplossing niet meer verandert. We kunnen schrijven $P_*x = tx$, waarbij t de proportionaliteits-factor is. We noemen t ook wel de **do - minante eigenwaarde** van P_*. Omdat $x'P_*x = tx'x = t$ zien we dat deze eigenwaarde gelijk is aan $x'P_*x = (x'P_1x + x'P_2x + x'P_3x)/3$, dus aan de gemiddelde homogeniteit. Wanneer we $y_j = D_j^{-1}G_j'x$ definieren als de bijbehorende optimale categoriegemiddelden, dan vinden we dus $(G_1y_1 + G_2y_2 + G_3y_3)/3 = tx$. Wanneer we de y_j in één vector met twaalf elementen verzamelen, staat hier ook $Gy = mtx$. Voorvermenigvuldigen met G' geeft $G'Gy = mtG'x$, oftewel $Cy = mtDy$, oftewel $m^{-1}D^{-1}Cy = ty$. Dit laat zien dat y de eigenvector is behorend bij de dominante eigenwaarde van de matrix $m^{-1}D^{-1}C$.

Het algorithme dat we hebben gebruikt berust op beurtelings middelen. Het convergeert naar de eigenvector behorend bij de dominante eigenwaarde van de **gemiddelde projector** P_* of van de **genormeerde Burt matrix** $m^{-1}D^{-1}C$. We zouden deze eigenwaarde problemen ook direct op kunnen lossen, met daartoe geschikte algorithmen, maar omdat in deze speciale toepassing beurtelings middelen voor de hand ligt, gebruiken we dat.

APPENDIX B

AANPASSEN VAN SCHOOLLOOPBAANMODELLEN

Schoolloopbaanmodellen

We bespreken kort een aantal schoolloopbaanmodellen waarin geen latente variabelen voorkomen. De meeste modellen in het Nederlandse schoolloopbaanonderzoek zijn van deze vorm. Onze belangrijkste conclusie is, dat het nogal vreemd is om dit soort modellen met behulp van het programma LISREL aan te passen. Een gewoon multipele regressieprogramma (bijvoorbeeld uit SAS of SPSS) is eenvoudiger aan te sturen, is aanzienlijk goedkoper, en geeft meer informatie van statistische aard. In de besproken klasse modellen zijn de met multipele regressie en met LISREL berekende schatters van pad-coefficienten en residu-varanties exact gelijk (op afrondingsfouten na).

Er zijn n individuen. Achtergrondsinformatie over deze individuen is verzameld in de matrix X. We veronderstellen dat X niet-stochastisch is, daardoor is het eenvoudig in X allerlei dummies en interacties op te nemen (als dat gewenst is). De afmetingen van X zijn n bij p. Er zijn m schoolloopbaan variabelen $\underline{y}_1$, , $\underline{y}_m$, metingen op deze variabelen staan in de n bij m matrix $\underline{Y}$. Veronderstel

$$\underline{Y} = XB + \underline{Y}A + \underline{E}.$$

Hierbij is B een p bij m matrix, terwijl A een m bij m bovendriehoeks matrix is. De storin - gen $\underline{E}$ vormen een n bij m matrix, met rijen die onafhankelijk zijn, verwachte waarde nul hebben, een een diagonale variantie-covariantie matrix V hebben.

In het minst restrictieve model uit deze klasse bevat B mp vrije parameters, en bevat A 1/2m(m-1) vrije parameters. Tenslotte bevat V nog m vrije parameters. In meer restrictieve modellen veronderstellen we dat sommige van de elementen van B gelijk aan nul zijn. Stel er zijn r van dit soort restrictieve nullen in een model M_r.

We passen het minst restrictieve model aan door m multipele regressies te doen. Iedere kolom van $\underline{Y}$ wordt voorspeld uit X en alle eerdere kolommen van $\underline{Y}$. De regressie coefficienten zijn maximale aannemelijkheidsschatters van A en B, de residu varianties zijn maximale aannemelijkheidsschatters van V. Stel R_j^2 is de multipele correlatie in het kwadraat voor het voorspellen van variabele $\underline{y}_j$.

Een meer restricitef model passen we ook aan door m multipele regressies te doen. Iedere kolom van $\underline{Y}$ wordt voorspeld uit de kolommen van A en B met pad-coefficienten die ongelijk aan nul zijn. Stel regressie j heeft t_j voorspellers, bij het volledige model is t_j gelijk aan p + (j - 1). Stel R_j^2 is de gekwadrateerde multipele correlatie voor het restrictieve model.

$$n \sum_{j=1}^{m} \ln (R_j^2/R_j^2)$$

die asymptotisch een chi-kwadraat verdeling heeft met r vrijheidsgraden. Het is interessant dat de statistieken n(ln R_j^2 - R_j^2) wanner M_r waar is onafhankelijke chi-kwadraten te zijn met p + (j - 1) - t_j vrijheidsgraden. Dit geeft een mooie opdeling van de verliesfunctie, die laat zien waar de afwijkingen van het model vooral zitten. Het is overigens van belang op te merken dat de chi-kwadraat toetsing alleen zinvol is, wanneer we aannemen dat de rijen van $\underline{E}$ normaal verdeeld zijn.

In vergelijking met de gebruikelijke manier om schoolloopbaan modellen te toetsen, hebben we verschillende winstpunten. In de eerste plaats maakt onze benadering duidelijk dat we niet hoeven aan te nemen dat de achtergrond normaal verdeeld is, we hebben deze aanname voor de storingstermen nodig. In het tweede plaats hebben we het programma LISREL niet nodig, een gemakkelijk aanstuurbaar multipele regressie programma volstaat. In de derde plaats heoven we niet iedere keer een compleet model te fitten. Het kan per vergelijking, en de vergelijkingen kunnen aan elkaar geplakt worden tot een compleet model. Ook de chi-kwadraat statistiek valt in een aantal stukken uiteen, waarbij ieder stuk met één enkele schoolloopbaan variabele correspondeert.

Tenslotte merken we op dat de standaardfouten die multipele regressie programma's bere - kenen voor de schattingen van A en B asymptotisch correct zijn, zelfs wanneer we niet aannemen dat $\underline{E}$ normaal verdeeld is.

De conclusie is duidelijk. LISREL is in dit soort situaties duur en onhandig. Het is ook duidelijk dat de situatie wezenlijk verandert wanneer het model latente variabelen (naast de storingstermen) bevat.

AANGEHAALDE PUBLIKATIES

Bosker, R.A. Hofman, A., & Velden, R. van der (1985). *Een Beschrijving van de Schoolloopbanen in het Groningen-cohort.* Interimrapport SVO-project 1042. Groningen.

Bosker, R., Hofman, A., & Velden, R. van der (1985). *Een Generatie Geselecteerd*. Groningen: RION.

Bosker, R., en Velden, R. van der (1985). *Onderwijspositie en Selectie.* Paper gepresenteerd op de ORD-85.

Bosman, R., Louwes, W., & Meer, A. van der (1980). *Sexe, School, Beroep*. Doctoraalscriptie Onderwijskunde/Sociologie, RUG.

CBS, (1982a). *Schoolkeuze en Schoolloopbaan bij het Voortgezet Onderwijs,* Cohort 1964/'65. 's Gravenhage: Staatsuitgeverij.

CBS, (1982b). *Schoolloopbaan en Herkomst van leerlingen bij het Voortgezet Onderwijs. Deel I: Bestand 1977.* 's Gravenhage: Staatsuitgeverij.

CBS, (1982c). *Schoolloopbaan en Herkomst van Leerlingen bij het Voortgezet Onderwijs. Deel II: Cohort 1977, Schoolkeuze.* 's Gravenhage: Staatsuitgeverij.

CBS, (1983), *Schoolloopbaan en Herkomst van Leerlingen bij het Voortgezet Onderwijs. (Cohort 1977, Peildatum September 1981).* CBS-mededelingen no 7797.

CBS, (1984). *Schoolloopbaan en Herkomst van Leerlingen bij het Voortgezet Onderwijs. (Cohort 1977, Peildatum September 1982).* CBS-mededelingen no 7818.

Collaris, J.W.M., en Kropman, J.A. (1978). *Van Jaar tot Jaar. Tweede Fase.* Nijmegen: ITS.

Cremers, P.G.J. (1980). Constructie van een Schaal voor Bereik van Voortgezet Onderwijs. *Tijdschrift voor Onderwijsresearch,* 5, 80-91.

Diederen, J. (1983). *Van Jaar tot Jaar. Derde Fase.* Nijmegen, ITS.

Dirkzwager, A. (1966). *Intelligentie en Schoolprestatie.* Amsterdam: Swets en Zeitlinger.

Dronkers, J. (1983). Have Inequalities in Educational Opportunity Changed in the Netherlands? A Review of Empirical Evidence. *The Netherlands Journal of Sociology,* 19-2, 1-18.

Dronkers, J., en De Jong, U. (1978). Jencks en Fägerlind op zijn Hollands. *Sociologische Gids,* 25, 4-30.

Dijk, H. van (1977). De Relatie tussen Intelligentie en Schoolsucces. *Beroepskeuze,* 24, 197-215, 242-258.

Faasse, J.H., Bakker, B., Dronkers, J., en Schijf, H. (1985). *Vergelijking van Individuele Schoolloopbanen in het Voortgezet Onderwijs in Noord-Brabant voor en na de Invoering van de Mammoetwet.* Paper gepresenteerd op de ORD-85.

Fienberg, S.E., en Mason, W.M. (1978). Identification and Estimation of Age-Period-Cphort Models in the Analysis of Discrete Archival Data. In K.F. Schuessler (red.), *Sociological Methodology* 1979. San Fransisco: Jossey-Bass.

Geer, J.P. van der, en Meulman, J. (1985). *PRINCALS*. Leiden: Vakgroep Datatheorie FSW/RUL.

Gifi, A. (1980). *Niet-lineaire Multivariate Analyse*. Leiden: Vakgroep Datatheorie FSW/-RUL.

Gifi, A. (1981). *Nonlinear Multivariate Analysis*. Leiden: Vakgroep Datatheorie FSW/-RUL.

Glenn, N.D. (1977). *Cohort Analysis*. Beverley Hills: Sage.

Groen, M. (1967). *De Voorspelbaarheid van Schoolcarrières in het Voortgezet Onderwijs* . Groningen: Wolters.

Hagenaars, J.A.P. (1985). *Loglineaire Analyse van Herhaalde Surveys*. Proefschrift KHT.

Heek, F. van, e.a. (1968). *Het Verborgen Talent. Milieu, Schoolkeuze, en Schoolgeschiktheid*. Meppel: Boom.

Israels, A.Z. (1984). *Path Analysis for mixed Qualitative and Quantitative Exogeneous and Intervening Variables*. Rapport 7998-84-M1. CBS, Voorburg.

Koopman, P., Eeden, P. van der, en De Jong, U. (1985). *Categorale MAVO-scholen en Schoolloopbanen in Amsterdam*. Paper gepresenteerd op de ORD-85.

Kreft, I., en Bronkhorst, H. (1985). *Maken Scholen Verschil?* Paper gepresenteerd op de ORD-85.

Kropman, J.A. , en Collaris, J.W.M. (1974). *Van Jaar tot Jaar*. Nijmegen: ITS.

Leeuw, J. de, en Stoop, I. (1979). Secondaire Analyse 'Van Jaar tot Jaar' met behulp van Niet-lineaire Multivariate Technieken. In J.L. Peschar (red.), *Van Achteren naar Voren*. 's Gravenhage: Staatsuitgeverij.

Meester, A.C., en Leeuw, J. de (1983). *Intelligentie, Sociaal Milieu, en de Schoolloopbaan*. Leiden: Vakgroep Datatheorie FSW/RUL.

Peschar, J.L. (1975). *School, Milieu, Beroep*. Groningen: Tjeenk Willink.

Peschar, J.L. (1983). *Schoolloopbanen in het Voortgezet Onderwijs*. 's Gravenhage: Staatsuitgeverij.

Roeleveld, J., Eeden, P. van den, Jong, U. de (1985). *Scholengemeenschappen met MAVO en Schoolloopbanen*. Paper gepresenteerd op de ORD-85.

Ryder, N.B. (1965). The Cohort as a Concept in the Study of Social Change. *American Sociological Review*, 30, 843-861.

Tesser, P.Th.M. (1981). *Schoolloopbaan onderzoek in Nederland*. Nijmegen: ITS.

Tesser, P.Th.M. (1984). De Overgang van Basisonderwijs naar Voortgezet Onderwijs in 1977. *Mens en Maatschappij*, 57, 26-54.

Tesser, P.Th.M. (1985).*Schoolloopbanen en Sociale Herkomst*. Concept Eindverslag SVO-project 0636.

Weeren, P. van (1960). *Project Nationale Differentiatie test. Rapport No 3: Ontwerp Ontwikkeling Criteriumschaal.* Leiden: RUL

Wiegersma, S., Swiebel, M., Groen, M., en Dommerholt, I. (1963). *School en Toekomst*. Haarlem: De Toorts.

Mason, W.M., en Fienberg, S.E. (red.) (1985). *Cohort Analysis in Social Research.* New York: Springer-verlag.

TABELLEN
EN
GRAFIEKEN

TABEL 1 Vergelijking marginalen populatie 1977, SMVO-steekproef, en ons deelbe - stand. Respectievelijk N = 247286, N = 5464.

	Pop	CBS	Ons
Groningen	4.1	3.9	4.2
Friesland	4.6	4.8	1.4
Drente	3.7	3.4	4.6
Overijssel	7.8	7.4	9.3
Gelderland	12.2	12.1	10.7
Utrecht	6.2	5.6	4.0
Noord-Holland	15.5	16.3	15.1
Zuid-Holland	20.3	20.6	33.3
Zeeland	2.3	2.1	3.2
Noord-Brabant	15.1	15.2	11.5
Limburg	8.0	8.1	2.6
IJsselmeerpolders	0.3	0.3	0.1
VWO advies	11.6	13.5	22.1
HAVO advies	16.8	19.5	35.1
MAVO advies	33.4	37.4	37.9
LBO advies	31.7	24.8	1.0
onbekend	5.9	4.7	3.9
hogere employes	9.1	10.4	14.3
middel employes	15.9	17.6	21.7
lagere employes	10.4	11.2	13.5
zelfst met pers	4.9	5.2	4.8
zelfst zon pers	7.4	7.0	5.9
gesch arbeiders	17.0	16.5	14.7
onges arbeiders	14.3	12.3	8.4
overigen	21.0	19.8	16.6

TABEL 2 Schoolloopbaan categorieen, verdeling over de meetpunten.

	1978	1979	1980	1981	1982
LBO-	384	408	279	95	132
LBO+			115	132	
MAVO 1	384	2			
MAVO 2	1695	725	64	2	
MAVO 3-		1345	829	133	8
MAVO 3+		11	28	11	
MAVO 4-			377	262	276
MAVO 4+			589	460	
HAVO 1	107				1
HAVO 2	1329	289		8	
HAVO 3		1154	469	26	1

TABEL 2 Vervolg

	1978	1979	1980	1981	1982
HAVO 4			970	1011	472
HAVO 5-				233	928
HAVO 5+				336	
VWO 1	34				
VWO 2	1449	106			
VWO 3		1276	175	5	
VWO 4			1071	215	22
VWO 5			1	952	462
VWO 6				1	792
KMBO				1	22
MBO 1			12	426	608
MBO 2				16	426
MBO 3					18
HBO 1				5	154
uitval	82	148	477	1142	1142
totaal	5464	5464	5464	5464	5464

TABEL 3 Tien korte schoolloopbanen.

1	LBO	LBO	LBO
2	LBO	MAVO	MAVO
3	MAVO	MAVO	MAVO
4	MAVO	MAVO	MAVO
5	HAVO	HAVO	VWO
6	HAVO	VWO	HAVO
7	VWO	VWO	VWO
8	VWO	VWO	HAVO
9	MAVO	HAVO	VWO
10	HAVO	HAVO	HAVO

TABEL 4 Homogeneity analysis iterations
4a: initial solution
4b: first iteration
4c: final solution

x	t=1	t=2	t=3	x
-.495	-.440	-.495	-.495	-.477
-.385	-.440	-.275	-.275	-.330
-.275	-.018	-.275	-.275	-.190
-.165	-.018	-.275	-.275	-.190
-.055	.165	.275	.165	.202
.055	.165	.165	.275	.202
.165	.220	.165	.165	.183
.275	.220	.165	.275	.220
.385	-.018	.275	.165	.141
.495	.165	.275	.275	.238
1.000	.567	.780	.780	.649

x	t=1	t=2	t=3	x
-.592	-.501	-.592	-.592	-.562
-.410	-.501	-.294	-.294	-.363
-.236	-.099	-.294	-.294	-.229
-.236	-.099	-.294	-.294	-.299
.251	.266	.241	.218	.242
.251	.266	.250	.273	.263
.227	.250	.250	.218	.239
.273	.250	.250	.273	.258
.175	-.099	.241	.218	.120
.296	.266	.241	.273	.260
1.000	.869	.972	.976	.886

x	t=1	t=2	t=3	x
-.560	-.471	-.560	-.560	-.530
-.381	-.471	-.306	-.306	-.361
-.268	-.150	-.306	-.306	-.254
-.268	-.150	-.306	-.306	-.254
.232	.267	.195	.199	.220
.302	.267	.298	.293	.286
.279	.296	.298	.199	.264
.312	.296	.298	.293	.296
.086	-.150	.195	.199	.081
.266	.167	.195	.293	.252
1.000	.900	.975	.971	.898

TABEL 5 PRIMALS kwantificaties van de schoolloopbaancategorieen

	leerjaar 2	leerjaar 3	leerjaar 4	leerjaar 5	leerjaar 6
lbo-	-1.598	-1.530	-1.440	-1.332	-1.319
lbo+			-1.393	-1.338	
mavo 1	-1.146				
mavo 2	-0.832	-1.085	-1.158		
mavo 3-		-0.763	-0.968	-1.051	-1.185
mavo 3+		-0.793	-1.021	-1.045	
mavo 4-			-0.691	-0.903	-0.972
mavo 4+			-0.698	-0.861	
havo 1	-0.334				
havo 2	+0.019	-0.255	-0.257		
havo 3		+0.124	-0.118	-0.139	
havo 4			+0.246	-0.145	-0.495
havo 5-				+0.298	-0.017
havo 5+				+0.296	
vwo 1	+0.452				
vwo 2	+1.395	+0.718			
vwo 3		+1.520	+1.008		
vwo 4			+1.633	+1.271	+0.696
vwo 5				+1.665	+1.167
vwo 6					+1.686
kmbo					-1.080
mbo 1			-0.071	-0.863	-0.830
mbo 2				-0.136	-0.697
mbo 3					+0.035
hbo 1					+0.341

TABEL 6 Relaties van schoolloopbaanvariabelen met achtergrondsvariabelen

	II	III	IV	V	VI
Groningen	-.247	-.199	-.098	.018	.015
Friesland	-.074	-.032	-.097	-.037	-.035
Drente	.016	.040	.090	.220	.177
Overijssel	-.177	-.219	-.156	-.062	-.062
Gelderland	.006	-.016	.007	.064	.055
Utrecht	.065	.006	.002	.169	.206
Noord-Holland	-.085	-.052	.052	.189	.176
Zuid-Holland	-.099	-.075	-.008	.088	.083
Zeeland	.192	.236	.307	.407	.468
Noord-Brabant	.114	-.095	-.047	.057	.085
Limburg	-.217	-.180	-.072	-.032	-.029
IJsselmeerpolders	-.177	-.141	-.096	.367	.053
eta^2 provincie	.0137	.0115	.0068	.0175	.0197

TABEL 6 Vervolg

	II	III	IV	V	VI
jongens	-.150	-.142	-.075	.023	.028
meisjes	-.017	.000	.058	.168	.163
eta^2 sexe	.0044	.0047	.0042	.0051	.0047
vwo advies	.728	.769	.819	.868	.817
havo advies	.064	.065	.111	.171	.152
mavo advies	-.673	-.680	-.631	-.534	-.475
lbo advies	-1.054	-1.053	-.991	-.782	-.747
eta^2 advies	.2950	.2998	.2971	.2891	.2656
hogere employes	.249	.271	.315	.378	.365
middel employes	.103	.119	.162	.220	.217
lagere employes	-.114	-.101	-.066	-.002	.015
zelfst met pers	-.231	-.193	-.123	-.035	-.059
zelfst zon pers	-.168	-.153	-.098	-.032	-.033
gesch arbeiders	-.196	-.200	-.144	-.017	-.014
onges arbeiders	-.424	-.445	-.360	-.183	-.222
overigen	-.244	-.216	-.118	.021	.045
eta^2 SES	.0466	.0453	.0359	.0346	.0371
intercept	-2.919	-2.968	-2.892	-2.733	-2.543
beta taal	.064	.065	.064	.061	.057
beta rekenen	.055	.058	.058	.057	.053
R^2 CITO	.3428	.3458	.3375	.3214	.3028

TABEL 7 Afhankelijke en onafhankelijke variabelen voor diverse regressies (back = achtergrond, pos 2 tm pos 5 zijn schoolloopbaan variabelen)

afh var	model A	model B	model C	model D
pos 2	back	back	back	back
pos 3	back pos 2	back pos 2	pos 2	pos 2
pos 4	back pos 2 pos 3	back pos 3	pos 2 pos 3	pos 3
pos 5	back pos 2 pos 3 pos 4	back pos 4	pos 2 pos 3 pos 4	pos 4

TABEL 7 Vervolg

afh var	model A	model B	model C	model D
pos 6	back pos 2 pos 3 pos 4 pos 5	back pos 5	pos 2 pos 3 pos 4 pos 5	pos 5

TABEL 8a Multipele correlaties, gekwadrateerd, voor regressies in tabel 7

afh var	model A	model B	model C	model D
pos 2	.4561	.4561	.4561	.4561
pos 3	.8724	.8724	.8662	.8662
pos 4	.8949	.8929	.8928	.8905
pos 5	.8871	.8869	.8848	.8845
pos 6	.8706	.8683	.8679	.8655

TABEL 8b: chi-kwadraat waarden bij vergelijking met model A

afh var	model A	model B	model C	model D
pos 2	---	---	---	---
pos 3	---	---	24.68	24.68
pos 4	---	7.74	8.13	17.06
pos 5	---	.78	8.99	10.16
pos 6	---	9.16	10.75	20.33
totaal	---	17.68	52.55	72.23

TABEL 8c: vrijheidsgraden voor de chi-kwadraten in tabel 8b

afh var	model A	model B	model C	model D
pos 2	0	0	0	0
pos 3	0	0	24	24
pos 4	0	1	24	25
pos 5	0	2	24	26
pos 6	0	3	24	27
totaal	0	6	96	102

TABEL 9 drie indicator matriksen

	L	M	H	V	L	M	H	V	L	M	H	V
1	1	0	0	0	1	0	0	0	1	0	0	0
2	1	0	0	0	0	1	0	0	0	1	0	0
3	0	1	0	0	0	1	0	0	0	1	0	0
4	0	1	0	0	0	1	0	0	0	1	0	0
5	0	0	1	0	0	0	1	0	0	0	0	1
6	0	0	1	0	0	0	0	1	0	0	1	0
7	0	0	0	1	0	0	0	1	0	0	0	1
8	0	0	0	1	0	0	0	1	0	0	1	0
9	0	1	0	0	0	0	1	0	0	0	0	1
10	0	0	1	0	0	0	1	0	0	0	1	0

TABEL 10a Burt Matrix met kruistabellen

	L	M	H	V	L	M	H	V	L	M	H	V
L	2	0	0	0	1	1	0	0	1	1	0	0
M	0	3	0	0	0	2	1	0	0	2	0	1
H	0	0	3	0	0	0	2	1	0	0	2	1
V	0	0	0	2	0	0	0	2	0	0	1	1
L	1	0	0	0	1	0	0	0	1	0	0	0
M	1	2	0	0	1	0	0	0	1	0	0	0
H	0	1	2	0	0	0	3	0	0	0	1	2
V	0	0	1	2	0	0	0	3	0	0	2	1
L	1	0	0	0	1	0	0	0	1	0	0	0
M	1	2	0	0	0	3	0	0	0	3	0	0
H	0	0	2	1	0	0	1	2	0	0	3	0
V	0	1	1	1	0	0	2	1	0	0	0	3

TABEL 10b: diagonale matrix met rechte tellingen

	L	M	H	V	L	M	H	V	L	M	H	V
L	2	0	0	0	0	0	0	0	0	0	0	0
M	0	3	0	0	0	0	0	0	0	0	0	0
H	0	0	3	0	0	0	0	0	0	0	0	0
V	0	0	0	2	0	0	0	0	0	0	0	0
L	0	0	0	0	1	0	0	0	0	0	0	0
M	0	0	0	0	0	3	0	0	0	0	0	0
H	0	0	0	0	0	0	3	0	0	0	0	0
V	0	0	0	0	0	0	0	3	0	0	0	0
L	0	0	0	0	0	0	0	0	1	0	0	0
M	0	0	0	0	0	0	0	0	0	3	0	0
H	0	0	0	0	0	0	0	0	0	0	3	0
V	0	0	0	0	0	0	0	0	0	0	0	3

TABEL 11a, 11b, 11c: drie projectoren. nb: we gebruiken ? voor het getal 1/3

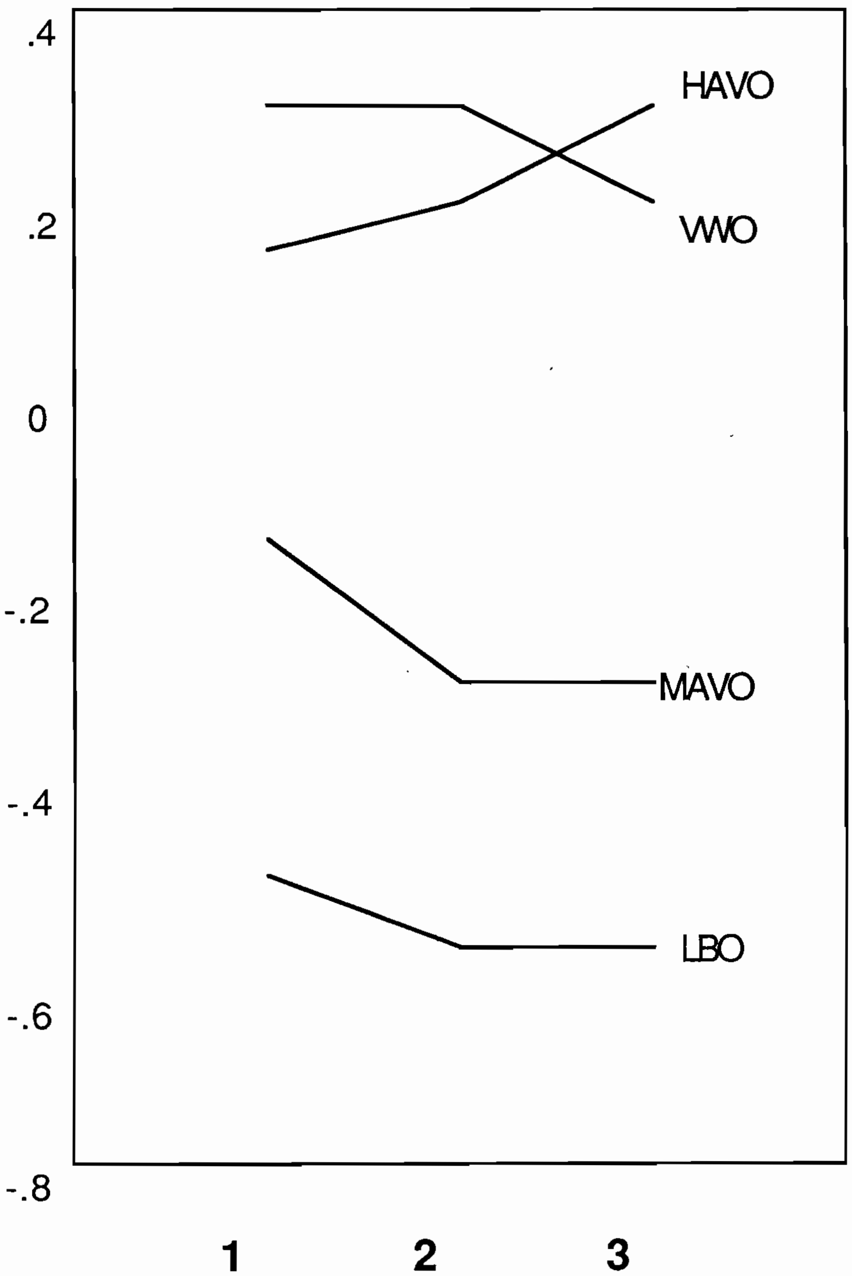
1/2	1/2	0	0	0	0	0	0	0	0
1/2	1/2	0	0	0	0	0	0	0	0
0	0	?	?	0	0	0	0	?	0
0	0	?	?	0	0	0	0	?	0
0	0	0	0	?	?	0	0	0	?
0	0	0	0	?	?	0	0	0	?
0	0	0	0	0	0	1/2	1/2	0	0
0	0	0	0	0	0	1/2	1.2	0	0
0	0	?	?	0	0	0	0	0	0

1	0	0	0	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	0	0	0	?	0	0	0	?	?
0	0	0	0	0	?	?	?	0	0
0	0	0	0	0	?	?	?	0	0
0	0	0	0	0	?	?	?	0	0
0	0	0	0	?	0	0	0	?	?

1	0	0	0	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	?	?	?	0	0	0	0	0	0
0	0	0	0	?	0	?	0	?	0
0	0	0	0	0	?	0	?	0	?
0	0	0	0	?	0	?	0	?	0
0	0	0	0	?	0	?	0	?	0
0	0	0	0	0	?	0	?	0	?

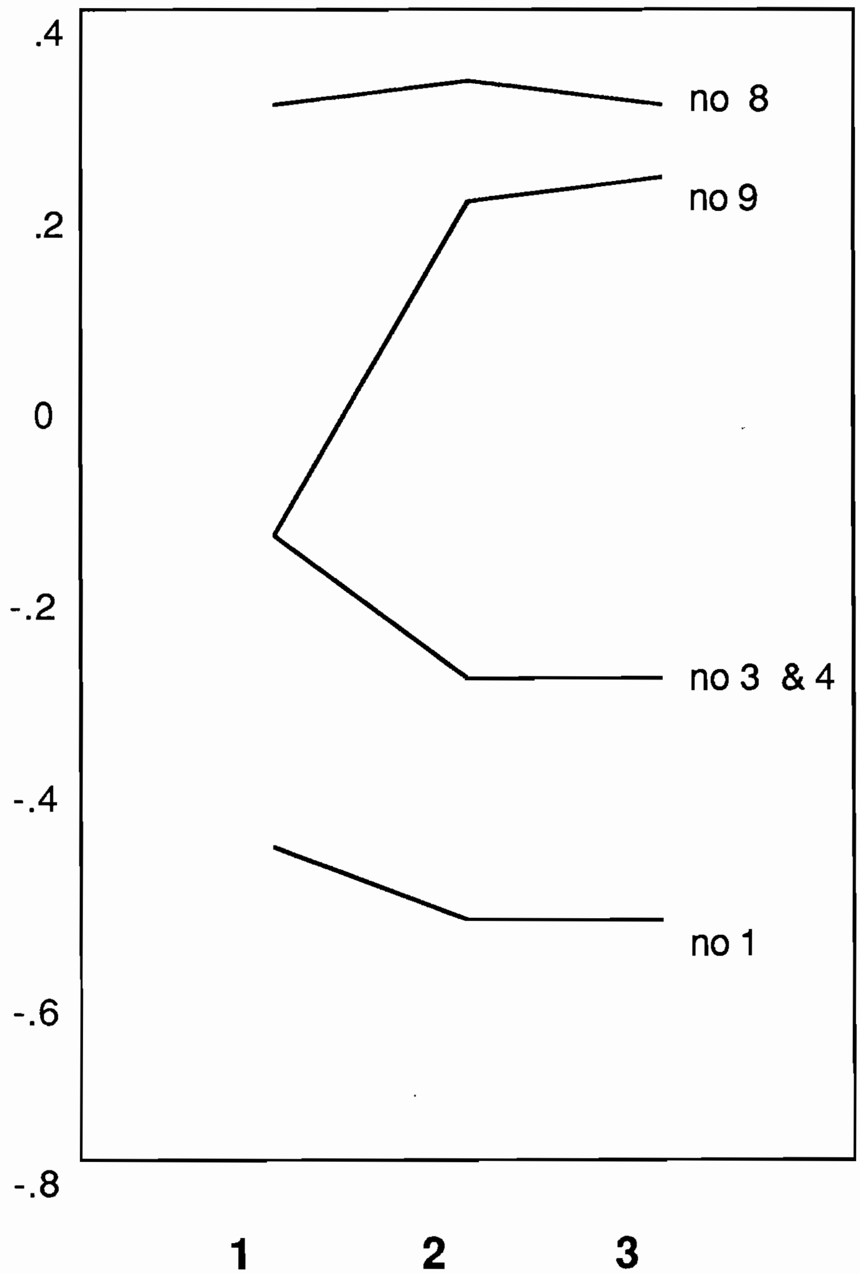
TABEL 11d gemiddelde projector

5/6	1/6	0	0	0	0	0	0	0	0
1/6	7/18	2/9	2/9	0	0	0	0	0	0
0	2/9	1/3	1/3	0	0	0	0	1/9	0
0	2/9	1/3	1/3	0	0	0	0	1/9	0
0	0	0	0	1/3	1/9	1/9	0	2/9	2/9
0	0	0	0	1/9	1/3	1/9	2/9	0	2/9
0	0	0	0	1/9	1/9	7/18	5/18	1/9	0
0	0	0	0	0	2/9	5/18	7/18	0	1/
0	0	1/9	1/9	2/9	0	1/9	0	1/3	1/9
0	0	0	0	2/9	2/9	0	1/9	1/9	1/3



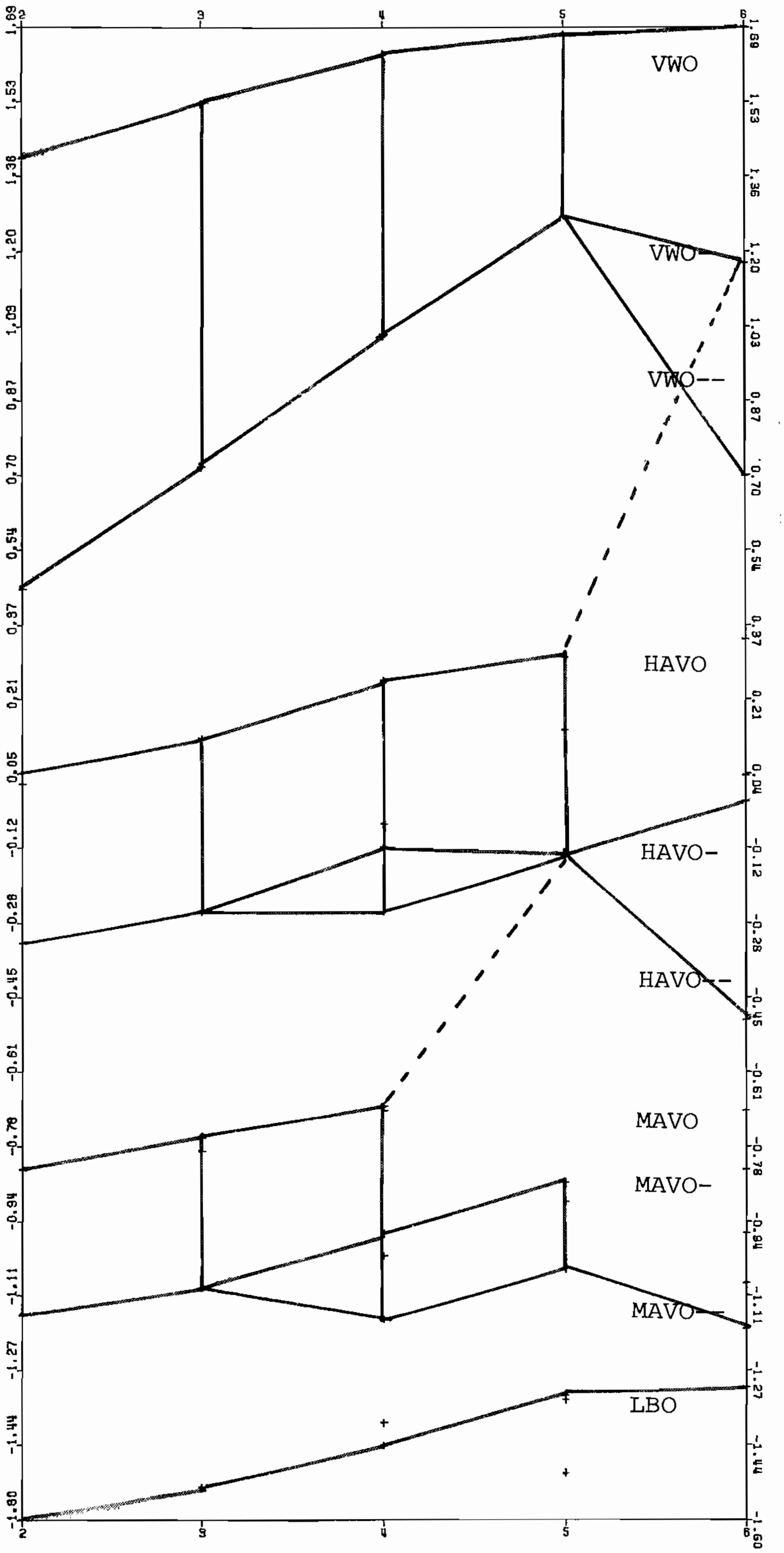
figuur 1a

loopbaanpatronen



figuur 1b

individuele carrieres



figuur 2.
schooltypen in de tijd
dikke lijnen: doorstroom
stippellijnen: opstroom