



UvA-DARE (Digital Academic Repository)

Beelden van tijd: De rol van historisch tijdsbewustzijn bij het leren van geschiedenis

Wilschut, A.H.J.

Publication date
2011

[Link to publication](#)

Citation for published version (APA):

Wilschut, A. H. J. (2011). *Beelden van tijd: De rol van historisch tijdsbewustzijn bij het leren van geschiedenis*. [Thesis, externally prepared, Universiteit van Amsterdam]. Koninklijke Van Gorcum.

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

**Empirisch onderzoek naar
oriëntatie in de historische
tijd bij leerlingen**

Het in het vorige hoofdstuk beschreven psychologisch en didactisch onderzoek suggereert dat tijdsbewustzijn in eerste instantie berust op associatieve contexten en pas in tweede instantie op chronologische datering. Toegepast op geschiedenisonderwijs zou dat betekenen dat leerlingen - om een historisch referentiekader te ontwikkelen - eerder gebruik maken van associatieve contexten dan van jaartallen en tijdbalken. De hypothese die in dit hoofdstuk met behulp van empirisch onderzoek wordt getest, is daarop gebaseerd. Deze luidt als volgt: *Aanduidingen van tijd die gebruik maken van jaartallen en eeuwen zijn moeilijker te leren en te onthouden dan aanduidingen van tijd die gebruik maken van beeldend-associatieve contexten; beeldend-associatieve contexten hebben een positieve invloed op het opbouwen van een historisch referentiekader.* Met een historisch referentiekader wordt bedoeld: een (tijds)kader waarmee niet alleen bekende gebeurtenissen en verschijnselen in verband kunnen worden gebracht, maar waarin ook (voor de leerling) nieuwe gebeurtenissen en verschijnselen kunnen worden geplaatst.

Het ontwikkelen van een historisch referentiekader blijkt in de praktijk van het geschiedenisonderwijs geen eenvoudige zaak. Om effectief met zo'n kader om te kunnen gaan moeten leerlingen:

- weten welke perioden kunnen worden onderscheiden en in welke volgorde zij moeten worden geplaatst;
- voorbeelden van gebeurtenissen en verschijnselen uit elke periode kunnen noemen;
- kenmerkende aspecten van een periode kennen en herkennen, zodanig dat 'nieuwe' gebeurtenissen en verschijnselen juist in de tijd kunnen worden geplaatst.

Onderzoek wijst uit dat leerlingen dit slechts met moeite leren beheersen. In het kader van het project *Usable Historical Past*s werd aan een aantal Britse leerlingen van ca. 14-jarige leeftijd gevraagd of zij in grote lijnen konden omschrijven wat zich in de laatste 2000 jaar in Groot-Brittannië had voorgedaan. Een aanzienlijk deel van hen weigerde de opdracht uit te voeren omdat zij geen idee hadden hoe zo iets aan te vangen. Anderen noemden lukraak gebeurtenissen en verschijnselen, zonder verband of volgorde. Slechts een kleine minderheid lukte het iets samenhangends op papier te krijgen.¹ Ook uit Duits onderzoek blijkt dat de mate waarin leerlingen de grote lijnen van de geschiedenis onder de knie krijgen niet meevalt.² Bij gebrek aan een behoorlijk overzicht van perioden met hun kenmerkende aspecten zal al helemaal geen sprake zijn van wat in het Engels wordt aangeduid als *'sense of period'*, een gevoel voor perioden dat wordt opgebouwd op basis van uiteenlopende soorten kennis zoals: hoe de wereld er in een bepaalde periode uitzag (gebouwen, kleding, gebruiksvoorwerpen), wat de houdingen, overtuigingen en gedachten van mensen waren, hoe het gesteld was met overheidszegg, oorlogen, sociale ongelijkheid, enzovoort.³ Voordat daaraan kan worden toegeko-

1 Howson (2007), p. 40-41.

2 Popp (2006), Popp (2009).

3 Dawson (2009), p. 51.

hoe past hij in een verloop van tijd ten opzichte van andere perioden, wie leefden er toen en wat gebeurde er?

Het experiment dat in dit hoofdstuk wordt omschreven, gaat over de vraag hoe leerlingen zich het gemakkelijkst een beeld van een stuk historisch verleden vormen dat antwoord geeft op deze laatste soort vragen. Het gaat om een klein experiment dat onderzoekt of leerlingen verschillende perioden in een onbekend stuk leerstof het best onderscheiden met behulp van jaartallen en eeuwen, of met behulp van beeldend-associatieve contexten en of zij gebeurtenissen het best met jaartallen en eeuwen, of met beeldend-associatieve contexten in de juiste tijd kunnen plaatsen. Dat is dus geen onderzoek naar volwaardig contextualiseren van historische verschijnselen en gebeurtenissen, waarbij ook zaken komen kijken als het leggen van inhoudelijke verbanden tussen bekende contexten en nieuwe gebeurtenissen en verschijnselen. Ook is het geen onderzoek naar de manier waarop leerlingen een gevoel voor perioden ontwikkelen.

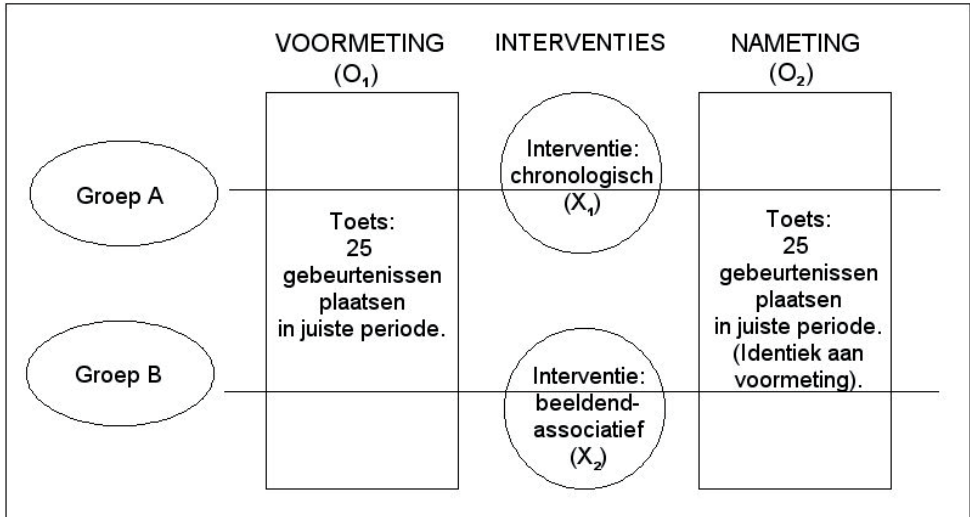
4.1 Onderzoeksontwerp

Uit de voorgaande hoofdstukken is gebleken dat de chronologische component van een historisch referentiekader relatief vreemd is aan de wijze van zich oriënteren in de tijd die mensen bij voorkeur ontwikkelen. In het geschiedenisonderwijs wordt echter veelal verondersteld dat de beheersing van chronologie een belangrijk aangrijpingspunt dient te zijn bij het bestuderen van een hoeveelheid leerstof. In één van de eerste geschiedenislessen leren leerlingen vaak tijdbalken maken, de jaartelling hanteren met jaren vóór en ná Christus, en eeuwen te onderscheiden. In schoolboeken voor het geschiedenisonderwijs worden hoofdstukken vaak begonnen met een tijdbalk waarop de belangrijkste gebeurtenissen geplaatst zijn. Bij het leren van stof na de behandeling van een onderwerp worden jaartallen vaak als ankerpunten gebruikt.

Om na te gaan welke rol chronologie (jaartallen, eeuwen, tijdbalken) en associatieve contexten spelen bij het ontwikkelen van een historisch referentiekader, is een experiment ontworpen waarbij twee parallelle groepen leerlingen onderwijs kregen over hetzelfde thema volgens twee verschillende didactische aanpakken. Deze beide didactische aanpakken zijn als volgt te typeren:

- *De chronologische aanpak* die zich in de eerste plaats oriënteert op jaartallen en eeuwen. De plaatsing van gebeurtenissen in de juiste volgorde en de juiste eeuw speelt bij deze aanpak de hoofdrol. Jaartallen en eeuwen zijn de ruggengraat volgens welke de leerstof wordt geordend en aangeboden. De koppeling van nieuwe gegevens aan de geleerde context verloopt in dit geval dus via de jaartallen en eeuwen, die vaste geheugenlocaties moeten gaan vormen.
- *De beeldend-associatieve aanpak* die zich in de eerste plaats richt op het creëren van bepaalde locaties in het geheugen die gekoppeld zijn aan beelden en voorstellingen waarmee geassocieerd kan worden. Daartoe worden in verhalende vorm historische situaties aangeboden die een beeld kunnen oproepen. Tijden worden niet van elkaar onderscheiden door nummering – zoals bij eeuwen het geval is – maar door typerende benamingen. Deze benamingen hebben te maken met een belangrijk ver-

schijnsel dat kenmerkend is voor een tijdvak. De vraag is of op zo'n manier een geheugenlocatie wordt gecreëerd waaraan ook nieuwe gegevens te koppelen zijn, zaken die niet door de benaming van een tijdvak worden gedekt.



Figuur 4.1 Ontwerp van het experiment

Gebaseerd op deze omschrijvingen zijn twee onderwijseenheden ontworpen die gaan over eenzelfde historische inhoud die zich over een langere tijdsperiode uitstrekt, namelijk enige honderden jaren. Een experiment over een korter tijdsverloop zou een onvoldoend beroep doen op het vermogen van leerlingen zich in de historische tijd te oriënteren. De beide inhouden zijn aangeboden aan twee parallelle groepen leerlingen van het voortgezet onderwijs van dezelfde leeftijd en hetzelfde onderwijsniveau: interventie X₁ en interventie X₂ (zie figuur 4-1). Het effect van de interventies is vastgesteld door middel van een nameting (O₂), die bestond uit de opdracht om 25 gebeurtenissen te plaatsen in de historische tijd. Om na te gaan wat de beginsituatie van de leerlingen was, is gebruik gemaakt van een voormeting (O₁) waarbij leerlingen voorafgaand aan de interventie de opdracht kregen 25 gebeurtenissen juist in de tijd te plaatsen; het ging hierbij om dezelfde gebeurtenissen die ook tijdens de nameting werden gebruikt. De voormeting en de nameting waren voor beide onderzoeksgroepen (groep A en groep B) gelijklopend.

Het gaat bij dit onderzoek om een *quasi-experimentele* aanpak, omdat de beide groepen leerlingen (groep A en groep B) geen aselechte steekproeven van leerlingen waren; het onderwijs is in bestaande schoolklassen gegeven, waardoor niet uit te sluiten is dat de groepen onderling enigszins verschilden. Op deze mogelijke verschillen tussen de groepen wordt verderop in dit hoofdstuk nader ingegaan. De quasi-experimentele aanpak is in onderwijsonderzoek zeer gebruikelijk, omdat laboratoriumsituaties rond leer-

lingen moeilijk te creëren zijn en om ethische redenen niet wenselijk.⁴ De hier besproken opzet is een voorbeeld van een quasi-experimenteel design dat gebruik maakt van een voormeting, een nameting en vergelijkingsgroepen.⁵ Verschillend ten opzichte van de gebruikelijke quasi-experimenten is dat in beide groepen een interventie is toegepast, waarvan de effecten met elkaar worden vergeleken. Meer gebruikelijk is een aanpak met een interventie in de ene groep en achterwege laten van die interventie in de andere groep (de 'controlegroep'), waardoor het effect van de interventie zuiver gemeten kan worden. Het achterwege laten van een interventie was in het hier beschreven geval echter geen optie, zoals uit het vervolg duidelijk zal worden. Zuivere vergelijking van de effecten wordt hierdoor echter een cruciaal aspect van het onderzoeksdesign.

Een probleem bij elk quasi-experimenteel onderzoek is dat verschillende uitkomsten op de nameting mogelijk niet (geheel) toe te schrijven zijn aan het effect van de interventie, maar aan verschillen tussen groepen die er al van aanvang af waren. Daarom is bij de uitvoering van dit experiment veel aandacht besteed aan de samenstelling van de groepen en het meten van mogelijke verschillen tussen de groepen.

Samenstelling van de deelnemersgroepen

De groepen A en B zijn samengesteld uit klassen van vier verschillende scholen uit verschillende streken in Nederland (Leiden, Amsterdam, Alkmaar en Leeuwarden), waardoor twee groepen zijn ontstaan van resp. 135 en 132 leerlingen. Deze grootte en samenstelling van de groepen vergrootte de kans op onderlinge vergelijkbaarheid en op representativiteit van de steekproef voor de totale leerlingpopulatie (van de betreffende leeftijd en het betreffende niveau) in Nederland (externe validiteit⁶). Groep A en groep B werden samengesteld uit telkens parallelle klassen van leerlingen van vier scholen: twee tweede klassen havo uit Leiden, twee tweede klassen vwo uit Leiden, twee tweede klassen vwo uit Amsterdam, twee tweede klassen havo uit Leeuwarden, twee derde klassen havo uit Alkmaar en twee derde klassen vwo uit Alkmaar. Zo konden groep A en groep B elk worden gevormd uit leerlingen van zes klassen afkomstig van vier verschillende scholen. De parallelle klassen waren telkens vergelijkbaar wat betreft samenstelling; ze waren in die zin 'aselect' dat de klassen op de scholen niet van elkaar waren gescheiden op grond van verschillende leerlingkenmerken, maar alleen omdat het aantal leerlingen in een bepaald leerjaar te groot was voor één klas. In drie van de zes gevallen had een tweetal parallelklassen – buiten de tijd van dit experiment - geschiedenisles van dezelfde leraar of lerares. Tijdens het experiment was dat bij alle parallelklassen het geval.

De toedeling van klassen aan groep A of groep B is willekeurig geweest. Noch de onderzoeker, noch de eigen docent van de klassen, noch de school heeft daarop invloed gehad. De interventie werd uitgevoerd door een speciaal daarvoor gevraagde gastdocent, die zelf bepaalde of hij in een bepaalde klas aanpak A of aanpak B zou toepassen.

4 Cohen, Manion & Morrison (2007), p. 282.

5 Shadish, Cook & Campbell (2002), p. 136-144.

6 Cohen, Manion & Morrison (2007), p. 136-137.

Deze keuze is willekeurig geweest, aangezien de bezoekende gastdocent de klassen vóór het moment van toepassing van de interventie niet heeft bezocht of gezien. De gastdocent overlegde niet met de onderzoeker welke keuze hij waar zou maken. Aan de klassen is niet meegedeeld wat de verschillen tussen de didactische aanpakken waren waarmee zij te maken kregen (ze kregen niet te horen wat de andere, parallelle groep meemaakte en hadden dus ook geen reden aan bepaalde zaken speciaal aandacht te besteden). Deze ‘afschermende’ maatregelen hebben waarschijnlijk een positief effect gehad op de validiteit van het experiment.⁷

De samenstelling van de beide op deze wijze tot stand gekomen groepen wat betreft havo- en vwo-niveau, jongens en meisjes, allochtoon en autochtoon, en leeftijden blijkt uit tabel 4-1. Als allochtoon zijn die leerlingen aangemerkt die aangaven zelf buiten Nederland geboren te zijn of van wie één van de ouders of beide ouders buiten Nederland was/waren geboren.

Zoals blijkt uit tabel 4-1, telde groep A verhoudingsgewijs iets meer vwo-leerlingen dan groep B. In groep A was het aantal jongens en meisjes vrijwel exact gelijk, in groep B was het aantal meisjes groter dan het aantal jongens. De aantallen allochtonen en autochtonen stemden vrijwel exact overeen, evenals de verdeling over leeftijden (zoals te zien is in de tabel, heeft een aantal leerlingen geen leeftijd opgegeven).

Aan de leerlingen werd gevraagd of ze geschiedenis een leuk vak vonden (ja = 2, een beetje = 1, nee = 0). Van de leraren werden de laatste rapportcijfers opgevraagd. De gemiddelden van deze gegevens zijn te vinden in tabel 4-2.

De verschillen zijn klein, maar voor zover er van verschil kan worden gesproken, valt het in het voordeel van groep A uit. Bij toepassing van een t-test bleken de verschillen tussen de groepen echter statistisch niet significant (waardering voor geschiedenis: $t = 0,308$; $df = 256$; $p = 0,758$; rapportcijfers: $t = 1,290$; $df = 239,04$; $p = 0,198$). De standaarddeviatie van de rapportcijfers was in groep B groter dan in groep A, wat inhoudt dat de verschillen tussen hogere en lagere cijfers in deze groep groter waren dan in de andere groep.

Tabel 4.1 Waardering voor geschiedenis en rapportcijfers

	Groep	N	Gemiddelde	Std. deviatie
Vindt geschiedenis leuk	A	133	1,19	0,947
	B	125	1,15	0,925
Rapportcijfer voor geschiedenis	A	135	6,696	0,9641
	B	127	6,520	1,2270

7 Shadish, Cook & Campbell (2002), p. 158.

Tabel 4.2 Kenmerken van deelnemende leerlingen

Niveau			
Groep		aantal	percentage
A	havo	62	45,9
	vwo	73	54,1
	Totaal	135	100,0
B	havo	69	52,3
	vwo	63	47,7
	Totaal	132	100,0
Jongen of meisje			
Groep		aantal	percentage
A	jongen	67	49,6
	meisje	68	50,4
	Totaal	135	100,0
B	jongen	56	42,4
	meisje	76	57,6
	Totaal	132	100,0
Allochtoon			
Groep		aantal	percentage
A	nee	118	87,4
	ja	17	12,6
	Totaal	135	100,0
B	nee	115	87,1
	ja	17	12,9
	Totaal	132	100,0
Leeftijd			
Groep		aantal	percentage
A	13	23	17,7
	14	72	55,4
	15	30	23,1
	16	5	3,8
	Totaal	130	100,0
B	13	23	17,7
	14	72	55,4
	15	30	23,1
	16	5	3,8
	Totaal	130	100,0

Beschrijving van de interventies

De interventies in groep A en groep B bestonden uit één lesuur (50 minuten) waarin de leerlingen geconfronteerd werden met een onderwerp waarover ze naar verwachting niets tot vrijwel niets zouden weten. Als onderwerp is gekozen voor 'de geschiedenis van Aruba vanaf het jaar 1400 tot heden' om de volgende redenen:

- Het is een onderwerp waarover in het gebruikelijke curriculum geen onderwijs wordt gegeven en waarover ook buiten het onderwijs om weinig bekendheid en algemene indrukken bestaan. Te verwachten was dat de nulsituatie wat betreft kennis van dit onderwerp onder de leerlingen nihil of vrijwel nihil zou zijn. De 'verstoring' invloed van in het gebruikelijke onderwijs verworven historische kennis was bij dit onderwerp waarschijnlijk minimaal.
- Het is een onderwerp dat zich over zes eeuwen uitstrekt en waarbij oriëntatie in de tijd dus een belangrijke rol speelt.
- Het is een onderwerp dat verhoudingsgewijs weinig gecompliceerd en 'uitgebreid' is, zodat in een beknopt onderwijsaanbod toch een redelijk 'compleet' overzicht ervan kon worden aangeboden.
- Het is een onderwerp dat een rijksdeel van het Koninkrijk der Nederlanden betreft, waardoor - ondanks de grote mate van onbekendheid ermee - toch een redelijke mate van interesse bij de leerlingen kon worden verondersteld, waardoor het tot stand komen van een referentiekader niet onevenredig negatief door desinteresse werd beïnvloed.

Gekozen werd voor één volledig lesuur van 50 minuten, te geven door een ervaren geschiedenisdocent, waarin de geschiedenis van Aruba volgens een chronologische dan wel een beeldend-associatieve methode aan de leerlinggroepen werd aangeboden. Deze keuze is gemaakt om het onderwijs zoveel mogelijk op 'normaal geschiedenisonderwijs' te laten lijken. Hierdoor werd de 'ecologische validiteit' van het experiment vergroot, wat inhoudt dat de experimentsituatie niet zo zeer afweek van de gebruikelijke onderwijssituatie dat daardoor negatieve effecten zouden optreden op het te meten leerresultaat.⁸ Een keuze voor alleen een schriftelijk aanbod van de leerstof zou ieder risico van (een mogelijk verschillende) invloed van een docent op het resultaat van het leerproces hebben uitgesloten. Daar staat tegenover dat de confrontatie met de leerstof, en dus de beide didactische aanpakken, dan veel minder intensief zou zijn geweest. Om verschillen tussen docenten en hun relatie met de klas geen rol te laten spelen, zijn gastdocenten ingezet die speciaal voor deze gelegenheid om hun medewerking werd gevraagd. De gastdocenten hadden ruime ervaring als geschiedenisleraar, maar waren in hun normale bestaan géén docent aan de scholen waar het experiment werd uitgevoerd en dus voor de deelnemende leerlingen onbekend. Voor een les in twee parallelgroepen volgens twee verschillende aanpakken werd steeds één en dezelfde docent ingezet. Twee docenten hebben hieraan meegewerkt: één van hen bediende de scholen in Leiden en Leeuwarden, de ander die in Amsterdam. In Alkmaar gaven beide docenten les, de één in het tweetal havo-klassen en de ander in het tweetal vwo-klassen.

8 Cohen, Manion & Morrison (2007), p. 138-139. Bryman (2008), p. 33.

Voor beide lessen is een powerpointpresentatie ontwikkeld (zie appendix 2, bijlage 1) en een protocol van de inhoud van de lessen (zie appendix 2, bijlage 2).⁹ Van de lessen zijn geluidsopnamen gemaakt. Hieruit bleken geen belangrijke afwijkingen van de tevoren bepaalde lesinhoud. Uit gesprekken met de uitvoerende docenten na afloop van de experimentele lessen is gebleken dat zij de lessen – afgezien van het verschil in didactische aanpak – als gelijkwaardig hebben ervaren: de ene les was niet veel beter, leuker of aangenamer dan de andere les.

Bij de *beeldend-associatieve* aanpak werd elk tijdvak geïntroduceerd met de naam (tijd van de indianen, het goud, paarden en piraten, etc.) waarop een verhalend stuk leerstof volgde dat deze naam kon illustreren. Deze verhalende context was bedoeld als geheugenlocatie waaraan zaken zouden kunnen worden gekoppeld. Vervolgens werden nog enkele gebeurtenissen en verschijnselen behandeld dit bij het betreffende tijdvak hoorden. Bij de *chronologische* aanpak werd elk tijdvak geïntroduceerd als genummerde eeuw en werden aan de hand van jaartallen enkele ankerpunten besproken om dit tijdvak op de chronologische lijn te plaatsen. Het nummer van de eeuw en de jaartallen waren bedoeld als geheugenlocatie waaraan zaken zouden kunnen worden gekoppeld. Vervolgens werden nog enkele gebeurtenissen en verschijnselen behandeld dit bij het betreffende tijdvak hoorden. Deze laatste gebeurtenissen en verschijnselen waren gelijkloidend aan die welke bij de beeldend-associatieve aanpak werden behandeld. Figuur 4-2 laat zien welke indelingen van tijden in beide aanpakken werden gebruikt. Een verduidelijking van de betekenis van de associatieve benamingen van aanpak B is te vinden in de tekst van het lesprotocol behorend bij aanpak 2, telkens aan het begin van een nieuw tijdvak (appendix 2, bijlage 2).

<p>Aanpak A: chronologisch:</p> <ul style="list-style-type: none"> - de vijftiende eeuw - de zestiende eeuw - de zeventiende eeuw - de achttiende eeuw - de negentiende eeuw - de twintigste eeuw 	<p>Aanpak B: beeldend-associatief:</p> <ul style="list-style-type: none"> - de tijd van de indianen (tot ca 1500) - de tijd van het nutteloze eiland (ca 1500 - 1600) - de tijd van paarden en piraten (ca 1600 - 1800) - de tijd van het goud (ca 1800 - 1900) - de tijd van de olie (ca 1900 - 1980) - de tijd van het toerisme (ca 1980 tot heden)
---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

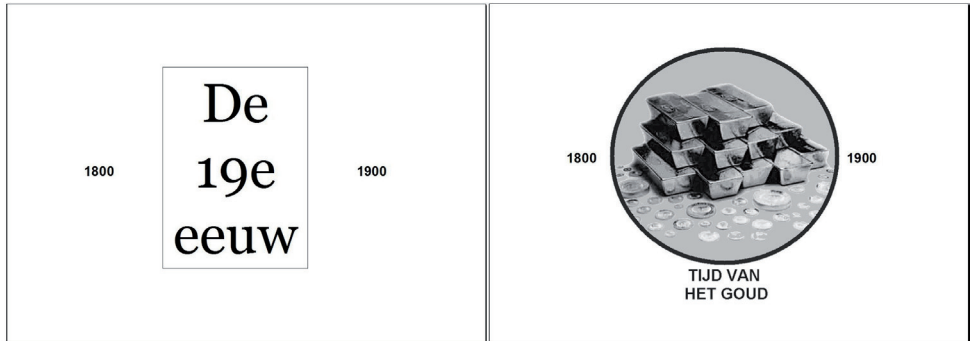
Figuur 4.2 Zes 'tijden' volgens twee aanpakken

Het gebruik van uitsluitend genummerde eeuwen en jaartallen als ankerpunten bij de chronologische aanpak is ten opzichte van het gebruikelijke geschiedenisonderwijs enigszins kunstmatig, omdat gewoonlijk in het onderwijs naast jaartallen en eeuwen

⁹ Voor de historische inhoud van de lessen en de toetsen is gebruik gemaakt van Hartog (1980), Hartog (1993), Daal (1992) en Dalhuisen e.a. (2009).

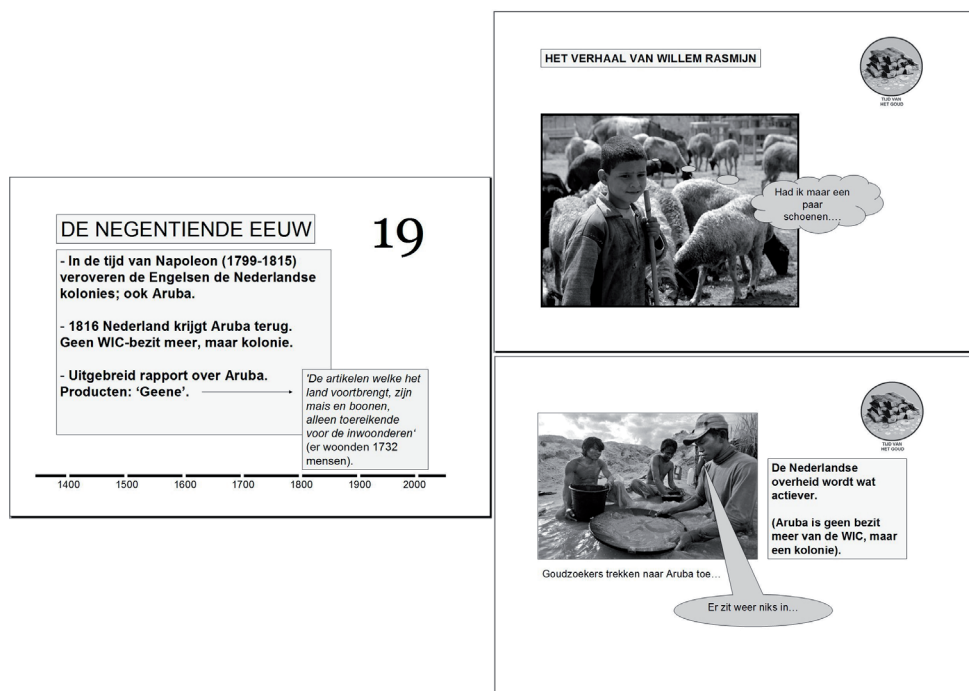
ook benamingen worden gebruikt als Renaissance en Verlichting, waarmee associaties kunnen worden opgeroepen. In het kader van dit onderzoek was het echter zuiver om een scherp onderscheid te maken tussen alleen chronologische en alleen beeldend-associatieve middelen om tijden van elkaar te onderscheiden. De vraag was immers welke middelen de oriëntatie in de tijd het best ondersteunen: chronologische of beeldend-associatieve.

Aan de hand van het voorbeeld van *de negentiende eeuw / de tijd van het goud* volgt nu een nadere uiteenzetting van het verschil tussen de twee aanpakken.



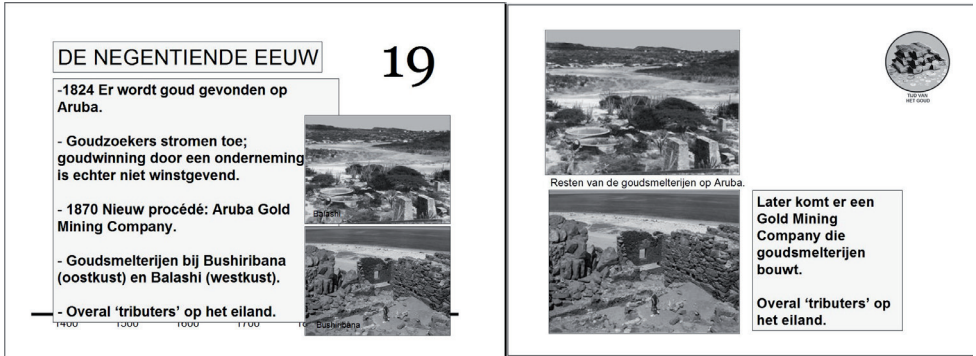
Figuur 4.3

De titeldia van een tijdvak (fig. 4-3) bestond in de *chronologische* aanpak uit groot en duidelijk aangekondigd de eeuw waarover het ging, in dit geval de 19e eeuw. Het getal negentien stond als kenmerk in alle volgende dia's rechtsboven in de hoek gedurende de behandeling van deze eeuw. Uitgelegd werd dat dit de tijd is die van 1800 tot 1900 loopt. Bij de *beeldend-associatieve* aanpak verscheen in het midden een beeldmerk van het tijdvak, in dit geval een stapel goudstaven en gouden munten. Aangekondigd werd dat nu de 'tijd van het goud' aan de beurt was, en dat deze tijd liep van 1800 tot 1900. Dit is de enige keer dat in de beeldend-associatieve aanpak van (alleen ronde) jaartallen gebruik gemaakt werd: telkens aan het begin van een tijdvak. Het beeldmerk van het tijdvak bleef rechtsboven in de hoek van de dia's zichtbaar zolang het tijdvak werd besproken.



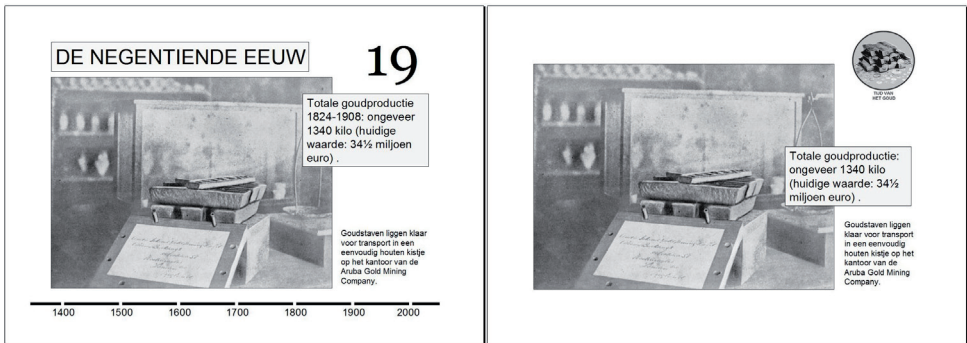
Figuur 4.4

Via de tweede dia (fig. 4-4) werd bij de *chronologische* aanpak de tijd geïntroduceerd door enkele chronologische gegevens: Aruba is in de tijd van Napoleon veroverd geweest door de Engelsen, maar werd nu als kolonie teruggegeven aan Nederland. De Nederlandse overheid liet een onderzoek instellen naar wat er in haar nieuwe kolonie te vinden was. Het resultaat was dat er eigenlijk niets te vinden was; het eiland bracht nauwelijks genoeg op om de inwoners te onderhouden. Er is een tijdbalk in beeld gekomen waarop de negentiende eeuw wordt aangegeven; deze bleef gedurende de behandeling van deze eeuw in beeld (de negentiende eeuw was met een contrastkleur op de tijdbalk aangegeven, wat bij de zwartwitte reproductie van de dia niet tot uitdrukking komt). Bij de *beeldend-associatieve* aanpak werden hier tegenover twee dia's gebruikt om een verhaal te vertellen. Een herdersjongetje wordt zonder schoenen de bergen in gestuurd om schapen te hoeden en vindt bij toeval een klompje goud, waardoor er meer dan genoeg geld is om schoenen voor hem te kopen. Als gevolg van de ontdekking komt er een stormloop van goudzoekers naar het eiland op gang, maar ze vinden maar weinig. Het verhalende van deze dia's werd ondersteund door gebruik van 'wolkjes' met uitspraken en gedachten van mensen (wat bij de chronologische aanpak niet voorkwam). De interesse van de Nederlandse overheid in haar kolonie – waarvan in het voorbijgaan werd meegedeeld dat die niet meer van de WIC was – werd groter door het goud.



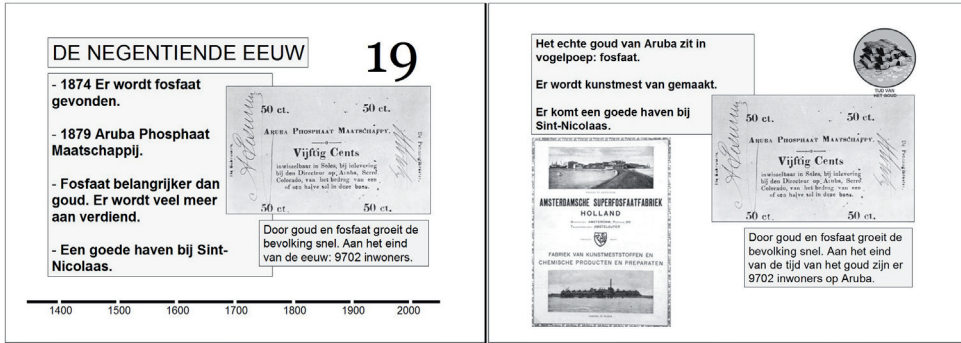
Figuur 4.5

Zowel in de *chronologische* aanpak als de *beeldend-associatieve* werd nu de informatie over de goudwinning uitgelegd (fig. 4-5): bij de chronologische aanpak aan de hand van twee jaartallen, bij de beeldend-associatieve aanpak globaler. De afbeeldingen waren dezelfde, maar bij de beeldend-associatieve aanpak konden ze wat groter zijn, omdat er minder chronologische informatie hoefde te worden gegeven.



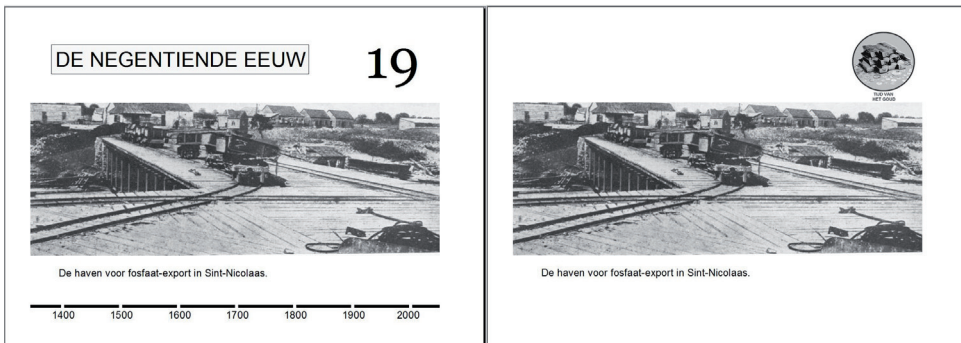
Figuur 4.6

De balans van de goudproductie werd opgemaakt (fig. 4-6). De informatie in beide aanpakken was hier precies gelijk, alleen de tijdbalk, het eeuwnummer, resp. het beeldmerk van het tijdvak verschilden.



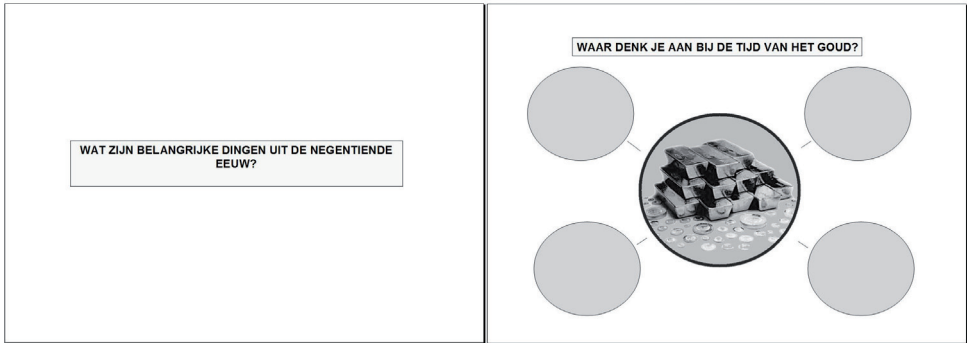
Figuur 4.7

Het tweede aspect van dit tijdvak werd geïntroduceerd (fig. 4-7): de fosfaatproductie die economisch veel belangrijker is geweest dan de goudproductie, omdat hij veel meer opbracht. Bij de *chronologische* aanpak werden weer enkele jaartallen gegeven rond de ontdekking en exploitatie van het fosfaat. Bij de *beeldend-associatieve* aanpak werd een koppeling gelegd met het goudverhaal door (enigszins expressief) te zeggen: 'het *echte* goud van Aruba zat in de vogelpoep'. De afbeelding van het muntbiljet van de Aruba Fosfaat Maatschappij en de informatie over de gevolgen van de fosfaatproductie waren gelijkkluidend. Bij de *beeldend-associatieve* aanpak was er ruimte voor een extra afbeelding (de Amsterdamse fosfaatfabriek) in plaats van de chronologische informatie op de corresponderende andere dia.



Figuur 4.8

(Fig. 4-8) Bij beide aanpakken gelijkkluidend: als gevolg van de fosfaatproductie werd een haven aangelegd, wat van belang was, omdat daar in het volgende tijdvak de olietankers konden aanleggen.



Figuur 4.9

De laatste dia (fig. 4-9) van een tijdvak was er altijd één waarbij de balans werd opgemaakt: wat zijn de belangrijkste dingen uit de negentiende eeuw, respectievelijk: waar denk je aan bij de tijd van het goud? In het laatste geval werd het associatieve denken benadrukt door het beeldmerk van het tijdvak in het centrum te plaatsen en daaromheen associërende ‘wolkjes’ waarin leerlingen hun associaties konden invullen (dit werd alleen mondeling gedaan tijdens de les).

Het verschil in aanpak kwam ook tot uitdrukking in de samenvatting op papier die de leerlingen na afloop van de les meekregen. Bij de *chronologische* aanpak werd een overzicht gegeven van de zes eeuwen met in elk daarvan de belangrijkste jaartallen en gebeurtenissen. Bij de *beeldend-associatieve* aanpak werden de zes tijdvakken gepresenteerd via hun beeldmerken en enkele afbeeldingen met losse blokjes informatie. De introducerende verhalende onderdelen die voor elk tijdvak tijdens de les werden gebruikt, kwamen in deze samenvattingen niet voor. Zie appendix 2, bijlage 3 voor de aan de leerlingen verstrekte schriftelijke samenvattingen.

Een week voor de experimentele les heeft de onderzoeker de deelnemende klassen bezocht om uit te leggen wat er de volgende week zou gebeuren en wat daarvan de bedoeling was. Ook heeft hij enige algemene informatie over Aruba gegeven (waar ligt het, wat is het voor een eiland) om daarmee de vraagstelling die de volgende week zou worden beantwoord (‘Hoe is Aruba het eiland geworden dat het nu is?’) in te leiden. Tijdens dit lesuur werd bovendien de voormeting afgenomen. Leerlingen moesten enkele vragen over chronologie, hun kennis van en affiniteit met Aruba beantwoorden en vervolgens proberen vijftientig gebeurtenissen uit de geschiedenis van Aruba op de juiste manier te plaatsen in de tijd. De leerlingen werden aangemoedigd zoveel mogelijk in te vullen wat ze dachten te weten. Dat was niet zo eenvoudig, omdat veel leerlingen bij het maken van een toets over een totaal onbekend onderwerp waarover ze niets hebben geleerd, de reflex vertonen dat ze er ‘dus’ niets van kunnen weten en dus ook niets hoeven in te vullen. Om te stimuleren dat ze dit toch zouden doen, werd, voor het invullen van de voormeting, klassikaal geoefend met twee gebeurtenissen die op grond

van algemeen bekende kennis redelijkerwijs juist in de tijd te plaatsen zouden moeten zijn. Aan leerlingen werd op die manier getoond: jullie weten misschien méér dan jullie denken. Gehoopt werd dat de leerlingen er hierdoor zoveel mogelijk toe zouden worden aangezet de voormeting zo serieus mogelijk in te vullen. Zie appendix 2, bijlage 4 voor de vragen die voorafgaand aan de voormeting aan de leerlingen werden gesteld en de vijftientig gebeurtenissen die zij in de tijd moesten proberen te plaatsen.

In de week volgend op de voormeting werd de les, volgens aanpak A dan wel B, aan de leerlingen gegeven door de gastdocent. Van de lesstof werd na de les de schriftelijke samenvatting aan de leerlingen verstrekt (zie appendix 2, bijlage 3). De leerlingen werden ertoe aangezet hiervan kennis te nemen voor de nameting die een week na de les zou plaatsvinden. Het bleek echter niet mogelijk deze leerstof serieus te laten meetellen voor een 'echte toets met een cijfer', omdat dat een te grote inbreuk zou maken op de bestaande toetsprogramma's van de scholen. Dat heeft invloed gehad op de mate waarin de leerlingen de leerstof serieus hebben bestudeerd. Op de rol die dit al dan niet bestuderen van de samenvatting heeft gespeeld, wordt bij de rapportage van de resultaten van dit experiment nader ingegaan.

Een week na de experimentele les werd de nameting afgenomen. Hiertoe heeft de onderzoeker zelf opnieuw alle deelnemende klassen bezocht en op de afname van de nameting toegezien. Dezelfde vijftientig gebeurtenissen werden gebruikt die ook bij de voormeting werden ingezet. Daarbij kan sprake zijn geweest van het test-hertest-effect, een effect waarbij leerlingen van de eerste test hebben 'geleerd' en daardoor beter scoren op de hertest.¹⁰ Indien dat het geval geweest is, kan aangenomen worden dat het bij beide groepen in gelijke mate een rol heeft gespeeld en dus niet van invloed is geweest op de gemeten verschillen in leereffect tussen beide groepen.

De nameting werd voorafgegaan door enkele vragen die beoogden vast te stellen of de leerlingen na het volgen van de experimentele les nog iets aan het leren van de stof hadden gedaan: hadden zij de uitgedeelde samenvatting nog eens doorgelezen of deze zelfs 'geleerd'? (zie appendix 2, bijlage 4E). De leerlingen werden aangemoedigd deze vragen naar waarheid te beantwoorden, omdat het voor hen geen consequenties zou hebben als ze niets meer aan de leerstof hadden gedaan.

De nameting werd gevolgd door een viertal evaluatieve vragen waarmee leerlingen hun mening konden geven over de les en de leraar: was de les leuk, was de les duidelijk en was de leraar aardig? (zie appendix 2, bijlage 4F). Op deze manier konden eventuele verschillen tussen de wijze waarop leerlingen het onderwijs volgens verschillende didactische aanpakken hadden ervaren, in kaart worden gebracht. Bijlage 4D in appendix 2 geeft de vijftientig gebeurtenissen die bij de voormeting en de nameting aan de leerlingen werden voorgelegd.

De beginsituatie van de groepen

In het voorgaande werd opgemerkt dat bij een quasi-experimenteel onderzoeksdesign extra zorgvuldig moet worden omgegaan met mogelijke verschillen tussen de groepen.

¹⁰ Cohen, Manion & Morrison (2007), p. 146. Bryman (2008), p. 150.

Daarom is onderzocht hoe het in beide groepen gesteld was met de beheersing van de chronologie, met hun eventuele voorkennis over het onderwerp Aruba, met hun affiniteit met en hun houding ten opzichte van het krijgen van onderwijs over dit onderwerp.

Met het oog op de beheersing van de chronologie werden aan de leerlingen van beide groepen vijf meerkeuzevragen voorgelegd (zie appendix 2, bijlage 4A voor deze vragen). De eerste vraag richtte zich op hun beheersing van het algemene beeld van de grote perioden (Oudheid, Middeleeuwen, Vroegmoderne Tijd, Moderne Tijd), de tweede vraag op hun vaardigheid in het bepalen van lengten van tijdsduren, de derde vraag op hun vaardigheid in het plaatsen van een gebeurtenis in een bepaalde eeuw (bijvoorbeeld: 1672 hoort in de 17^e eeuw), de vierde op het bepalen van een volgorde en de laatste op het inschatten van tijdsduur op grond van de hoeveelheid aandacht die in een boek aan een bepaalde tijd wordt besteed. De scoring van elk van deze vijf vragen was 1 punt bij een juist antwoord en 0 punten bij een onjuist antwoord. Bij elkaar opgeteld gaf dit een variabele 'beheersing van chronologie' uitgedrukt in een waarde van minimaal 0 en maximaal 5. Cronbach's alfa voor deze vijf items was 0,324: een lage waarde. Het is daarom de vraag of met deze vijf items hetzelfde werd gemeten. De gemiddelde resultaten van beide groepen zijn te vinden in tabel 4-3.

Tabel 4.3 Gemiddelde scores van de beide groepen op een vijftal chronologische vragen

	Groep	N	Gem.	Std. Deviatie
Beheersing chronologie	A	134	4,1567	0,96444
	B	127	4,0394	1,01106

Bij beide groepen is de gemiddelde score hoog: leerlingen wisten gemiddeld ruim vier van de vijf vragen goed te beantwoorden. Door de lage waarde van alfa kan daarvan niet méér worden gezegd dan dat beide groepen op ongeveer gelijk gemiddeld niveau vijf uiteenlopende vragen over chronologie beantwoordden. Bij toepassing van een t-toets bleek het verschil tussen de gemiddelden van beide groepen statistisch niet significant ($t = 0,960$; $df = 259$; $p = 0,338$). Het grootste verschil was er bij de eerste vraag waarop groep A gemiddeld 0,86 (st. dev. 0,350) en groep B gemiddeld 0,73 (st. dev. 0,447) scoorde.

Met het oog op eventuele voorkennis van de geschiedenis van Aruba werden aan de leerlingen van beide groepen vijf open vragen voorgelegd (zie appendix 2, bijlage 4B). Deze vijf open vragen waren:

- *Wanneer ongeveer werd Aruba door de Nederlanders veroverd?* Het exact juiste antwoord is 1636, maar antwoorden als (begin) zeventiende eeuw, of gouden eeuw zijn hier goed gerekend. Niet goed gerekend werd 'in de VOC-tijd', omdat die zich uitstrekte van 1602 tot 1799, een te lange tijdsperiode.
- *Wie waren de baas op Aruba toen de Nederlanders het eiland veroverden?* Het juiste antwoord is de Spanjaarden. Veel leerlingen dachten dat autochtonen de baas waren. Ook het antwoord 'de Engelsen' werd een aantal malen gegeven.

- Welke delfstof (bijv. steenkool, aardolie, of iets anders dat je in de grond vindt) is ooit op Aruba gevonden? Juiste antwoorden zijn goud of fosfaat. De antwoorden 'olie' en 'aloe' werden niet goed gerekend, omdat olie niet op Aruba is gevonden, maar geraffineerd en omdat aloë geen delfstof is.
- Voor welk doel werd Aruba vóór 1800 vooral door de Nederlanders gebruikt? Het juiste antwoord is: als een 'rancho' voor paarden en geiten.
- Hoe heette de Arubaanse leider die het eiland een aparte eigen regering bezorgde? Het juiste antwoord is Betico Croes.

De scoring op deze vijf vragen was 1 punt bij een juist antwoord en 0 punten bij een onjuist antwoord. Bij elkaar opgeteld gaf dit een variabele 'kennis van Arubaanse geschiedenis' uitgedrukt in een waarde van minimaal 0 en maximaal 5. Bij vraag 4 en 5 bleek echter geen enkele leerling uit beide groepen een juist antwoord te hebben gegeven, zodat de werkelijke schommeling van deze variabele zich afspeelt tussen minimaal 0 en maximaal 3. De gemiddelde scores van beide groepen op deze vijf vragen worden weergegeven in tabel 4.4.

Tabel 4.4 Scores van de beide groepen op een vijftal vragen over Arubaanse geschiedenis

	Groep	N	Gem.	Std. deviatie
Kennis van Arubaanse geschiedenis	A	135	0,3926	0,56113
	B	131	0,2824	0,50038

Bij beide groepen is de gemiddelde score zeer laag. Berekening van Cronbach's alfa heeft bij dergelijke zeer lage scores weinig zin. Bij beide groepen lijkt het niveau van voorkennis van Arubaanse geschiedenis zeer laag te zijn. Bij toepassing van een t-toets bleek het verschil tussen de groepen statistisch niet significant ($t = 1,688$; $df = 264$; $p = 0,093$). Het grootste verschil was er bij de eerste kennisvraag (groep A gem. 0,444, st. dev. 0,38, groep B gem. 0,388, st.dev. 0,34) over de verovering van Aruba in het begin van de zeventiende eeuw. Dat groep A hierop iets beter – zij het statistisch niet significant beter – scoorde dan groep B, zou kunnen duiden op een overeenkomst met de iets betere algemene beheersing van de chronologie in deze groep blijkend uit de iets betere beantwoording van de eerste chronologievraag. Het juist plaatsen van de verovering van Aruba in (het begin van) de zeventiende eeuw berust op beheersing van de algemene Nederlandse geschiedenis, bijvoorbeeld die over de Verenigde Oost-Indische Compagnie en de West-Indische Compagnie, en dus niet op specifieke kennis van de Arubaanse geschiedenis.

Een klein aantal leerlingen in beide groepen (13 in groep A en 8 in groep B) antwoordde op vraag 3 terecht met 'goud'. Of dat op specifieke kennis duidt, is onzeker. Misschien hebben ze dit afgeleid uit hun algemene kennis over spreekwoordelijke goudzoekers in de Nieuwe Wereld. Op de overige punten was de kennis zoals die bleek uit gegeven antwoorden vrijwel nihil tot nihil in beide groepen. Iemand die behoorlijk op de hoogte is van de Arubaanse geschiedenis, bijvoorbeeld schoolkinderen van Aruba

zelf, zou zeker de naam van de Arubaanse leider Betico Croes gekend hebben. Dat bleek hier in beide groepen bij geen enkele leerling het geval. Geconcludeerd kan worden dat in beide groepen geen specifieke voorkennis over de Arubaanse geschiedenis aanwezig was.

Met het oog op hun mogelijke affiniteit met het onderwerp werd aan de leerlingen van beide groepen de vraag voorgelegd of ze ooit op Aruba of één van de andere eilanden van de Antillen geweest waren. Hierop antwoordden in beide groepen 13 leerlingen met 'ja' en de overige met 'nee'. Verder werd gevraagd of ze mensen afkomstig van Aruba of één van de andere Antillen kenden, en zo ja of ze die 'gewoon gemiddeld' of 'heel goed' kenden. In groep A antwoordden 35 leerlingen dat ze 'gewoon gemiddeld' mensen van de eilanden kenden, in groep B waren dat 40 leerlingen. 'Heel goed' kenden de leerlingen van groep A in 18 gevallen iemand van de eilanden en in groep B in 16 gevallen. Deze verschillen zijn zeer klein en bleken statistisch niet significant.

Met het oog op hun houding ten opzichte van het onderwerp werd aan de leerlingen van beide groepen de vraag voorgelegd of de geschiedenis van Aruba voor Nederlanders een belangrijk onderwerp is. Daarop konden ze antwoorden met 'ja' of 'nee', maar sommigen schreven erbij dat dat 'een beetje' zo was. Bij de scoring werd aan de vraag 0 punten toegekend voor 'nee', 1 punt voor 'een beetje' en 2 punten voor 'ja'. Verder werd gevraagd of het de leerlingen leuk leek les te krijgen over de geschiedenis van Aruba. Daarop kon geantwoord worden met 'nee, niet leuk' (0 punten), 'ik weet het niet / het maakt me niet uit' (1 punt) of met 'ja, best leuk' (2 punten). Door optelling van de scorepunten op deze twee vragen kon een variabele 'houding ten opzichte van het onderwerp Aruba' worden bepaald die kon variëren tussen 0 en 4. De gemiddelde scores worden weergegeven in tabel 4-5.

Tabel 4.5 Houding ten opzichte van (les over) het onderwerp Aruba

	Groep	N	Gem.	Std. deviatie
Houding t.o.v. onderwerp Aruba	A	124	2,2419	1,38147
	B	116	2,2586	1,42719

Bij beide groepen is de gemiddelde score hoger dan 2, wat een in het algemeen positieve houding ten opzichte van het onderwerp en het lesexperiment impliceert. De verschillen tussen de groepen zijn uiterst klein en bleken dan ook bij toepassing van een t-toets statistisch niet significant ($t = 0,092$; $df = 238$; $p = 0,927$)

Conclusie. De verschillen tussen de twee groepen tot aan het moment van afnemering van de voormeting zijn gering gebleken. Bij de samenstelling van de groepen bleek dat in groep A het aandeel vwo-leerlingen iets groter was dan in groep B, maar het verschil was statistisch niet significant. Ook in andere opzichten leek groep A in een net iets betere uitgangspositie te verkeren dan groep B, zij het dat ook die verschillen statistisch niet significant waren. De algemene beheersing van de chronologie leek in groep A iets beter te zijn dan in groep B, wat overeenstemt met het gegeven dat in groep A iets meer

leerlingen de Nederlandse verovering van Aruba juist in de tijd wisten te plaatsen. De verschillen tussen de groepen waren klein, maar voor zover ze van invloed zijn geweest op het resultaat van de voor- en nameting, zou te verwachten zijn dat groep A wat beter zou presteren dan groep B.

4.2 De gebruikte toets bij de voor- en nameting

Tijdens de voormeting en de nameting moesten de leerlingen van 25 gebeurtenissen uit de Arubaanse geschiedenis aangeven wanneer deze hadden plaatsgevonden (zie appendix 2, bijlage 4D). Van deze vijftientig waren negen tijdens beide lessen expliciet door de leraar genoemd, de overige zestien moesten de leerlingen door logisch redeneren met behulp van hun historische referentiekader aan het juiste tijdvak of de juiste eeuw koppelen. Bij de voormeting maakte dit verschil uiteraard niets uit, omdat de leerlingen nog niet beschikten over een referentiekader betreffende de geschiedenis van Aruba. Bij de evaluatie van de nameting is echter een verschil gemaakt tussen de categorie reproductie (gebeurtenissen die tijdens de lessen waren genoemd) en de categorie toepassing (gebeurtenissen die niet tijdens de lessen waren genoemd).

Een ander onderscheid dat in verband met dit experiment van belang is, is de aanwezigheid van een woord in de omschrijving van een gebeurtenis dat ook voorkomt in de naam van een tijdvak bij de beeldend-associatieve aanpak. Dat was bij de omschrijving van acht gebeurtenissen het geval: 'indianen' bij gebeurtenis 15 en 17, 'paarden' bij gebeurtenis 3, 17 en 18, 'piraten' bij gebeurtenis 8 en 'olie' bij gebeurtenis 6, 16 en 22. Bij de voormeting maakte dit onderscheid niets uit, maar bij de nameting konden de leerlingen van groep B antwoorden op basis van woordherkenning. Deze bracht hen in twee gevallen op een dwaalspoor: bij gebeurtenis 15, waarin een 'indiaanse' hoofdman wordt genoemd, die ten onrechte een plaatsing in de 'tijd van de indianen' zou kunnen oproepen, en bij gebeurtenis 17, waarbij zowel indianen als paarden worden genoemd, terwijl de gebeurtenis noch in de tijd van de indianen, noch in die van paarden en piraten thuishoort. In de zes andere gevallen kon woordherkenning voor de leerlingen van groep B tot een plaatsing van een gebeurtenis in het juiste tijdvak leiden. Door bij de verwerking van de gegevens van de nameting onderscheid te maken tussen gebeurtenissen waarbij wél en níét gebruik gemaakt wordt van een sleutelwoord in de omschrijving, kon worden nagegaan in hoeverre woordherkenning bij de leerlingen van groep B een rol heeft gespeeld. In tabel 4-6 wordt een overzicht gegeven van de toetsitems in de verschillende categorieën.

Tabel 4.6 Overzicht van de toetsitems van verschillende categorieën in de tests

	reproductie	toepassing
met sleutelwoord	6, 8	3, 15, 16, 17, 18, 22
zonder sleutelwoord	2, 5, 9, 12, 14, 24, 25	1, 4, 7, 10, 11, 13, 19, 20, 21, 23

Scoring van de toetsen

Op het toetsformulier was aangegeven dat leerlingen van diverse mogelijkheden gebruik konden maken voor het plaatsen van een gebeurtenis in de tijd: jaartallen, eeuwen of aanduidingen als 'in de tijd van'. Ook konden de leerlingen het vakje 'ik weet het niet' aankruisen in het geval waarin dat van toepassing was. Dat laatste is gedaan om te voorkomen dat leerlingen allerlei 'wilde gokken' zouden maken met jaartallen en eeuwen 'om maar iets op te schrijven', waardoor de zuiverheid van de meting zeer negatief zou worden beïnvloed. Een gevolg is echter waarschijnlijk ook geweest dat leerlingen al te gemakkelijk gebruik maakten van deze mogelijkheid, ook als het om zaken ging waarvan ze redelijkerwijs konden vermoeden wanneer die hadden plaatsgevonden. Zo ligt het bijvoorbeeld voor de hand dat de aanleg van een vliegveld en de bouw van een winkelcentrum in de twintigste eeuw hebben plaatsgevonden (gebeurtenis 23 en 13). Niettemin heeft meer dan de helft van de leerlingen op de voormeting aangegeven dat niet te weten of heeft hier een onjuist antwoord ingevuld.

Bij de scoring zijn voor de voormeting en de nameting dezelfde regels gebruikt. Een volledig juist antwoord werd gewaardeerd met 2 punten, een volledig onjuist of 'ik weet het niet'-antwoord met 0 punten. Daar tussenin is een categorie 'enigszins juist' onderscheiden van 1 punt waar antwoorden in de goede richting waren, maar niet volledig juist. Deze categorie is vooral gebruikt voor de elf gebeurtenissen uit de twintigste eeuw (nrs 2, 5, 6, 7, 10, 13, 14, 16 en 22). In dit geval werd een nadere specificatie vereist om de 2 punten te scoren; alleen de vermelding 'twintigste eeuw' leverde 1 punt op. Dit is zo gedaan omdat de twintigste eeuw in de beeldend-associatieve aanpak twee tijdvakken kende (tijd van de olie en tijd van het toerisme) waaruit de juiste moest worden gekozen. Daar staat tegenover dat de zeventiende en de achttiende eeuw in de beeldend-associatieve aanpak als één periode werden beschouwd (de tijd van paarden en piraten), zodat voor gebeurtenissen uit die periode (nrs 3, 8, 18 en 19) zowel 'zeventiende eeuw' als 'achttiende eeuw' (of jaartallen uit die periode) met 2 punten werden gescoord. Ook een aanduiding als 'VOC-tijd' werd in dit geval zo gescoord, omdat deze vrijwel samenvalt met deze periode. Behalve voor de perioden 1600-1800 en die vóór 1500 zijn tijds-aanduidingen die verwezen naar een periode van langer dan honderd jaar, maar waarin het juiste tijdstip wel voorkwam, gescoord met één punt.

Gebeurtenissen die een 'moderne indruk' maken, zoals het bouwen van een winkelcentrum (13), het aanleggen van een vliegveld (23) of het bouwen van een installatie die drinkwater uit zeewater destilleert (24) zijn met nul punten gescoord als ze in het 'heden', of 'nu' of na het jaar 2000 werden geplaatst, omdat ze op deze manier geassocieerd werden met het heden in plaats van met een historisch verleden. Uit zo'n plaatsing spreekt geen historisch besef, geen poging om ze in een historische tijd te plaatsen. Voor gebeurtenis 4 (de verdwijning van het Amerikaanse meisje Natalee Holloway in 2005) geldt het omgekeerde: omdat deze gebeurtenis alleen uit de actualiteit bekend kon zijn, moesten de leerlingen deze plaatsen in de periode van 'nu' of de tijd ná 2000. Andere plaatsingen zijn met nul punten gescoord.

Gebeurtenis 1 moest worden geplaatst in de zestiende eeuw of de tijd van het nutteloze eiland. Het gaat om de tijd nadat de Spanjaarden de gehele populatie van Aruba hebben weggevoerd. Gebeurtenis 9 slaat op de goudkoorts tijdens de negentiende eeuw

en moet dus in die eeuw, of in de tijd van het goud, worden geplaatst. Gebeurtenis 11 moet slaan op de tijd waarin het christendom nog geen invloed had op Aruba, dus elke aanduiding vóór de zestiende eeuw (of: 'de tijd van indianen') levert twee punten op. Aanduidingen tijdens de zestiende eeuw zijn met één punt gescoord, omdat het niet uitgesloten is dat nog niet alle indianen op Aruba in die eeuw zich tot het christendom hadden bekeerd. Gebeurtenis 12 moest worden geplaatst in de zestiende eeuw of de tijd van het nutteloze eiland; het gaat om het deporteren van de indiaanse bevolking door de Spanjaarden in die periode. Gebeurtenis 15 moest worden geplaatst in de zestiende eeuw of de tijd van het nutteloze eiland omdat het christendom blijkbaar reeds was aangekomen (dus na de komst van de Spanjaarden) maar nog niet tot alle indianen doorgedrongen (dus voor de verovering door de Nederlanders). Gebeurtenis 20 slaat op mijnbouw en dus op goudwinning. De juiste plaatsingen zijn de negentiende eeuw of de tijd van het goud. Hetzelfde geldt voor gebeurtenis 21. Gebeurtenis 25 slaat op het begin van de koloniale tijd, de tijd ná die van de West-Indische Compagnie, en moest dus begin negentiende eeuw worden geplaatst. 'Rond 1800', 'negentiende eeuw' of 'tijd van het goud' zijn hier met twee punten gescoord. In appendix 2, bijlage 5 is een overzicht van de gebruikte scoringsvoorschriften te vinden. Twintig via een random-generator willekeurig gekozen combinaties van voormetingen en nametingen werden gescoord door een onafhankelijke tweede beoordelaar. Daarbij bleek een geobserveerde overeenkomst bij de scoringen van 98,2% op de voormeting en 93,8% op de nameting, Cohen's kappa resp. 0,98 en 0,96.

Betrouwbaarheid van de toetsen

De interne betrouwbaarheid van de toets (de mate waarin de items hetzelfde meten, in dit geval: het vermogen van leerlingen om gebeurtenissen in de juiste tijd te plaatsen) is bepaald voor de toepassing tijdens de voormeting en de toepassing tijdens de nameting. In het eerste geval was Cronbach's alfa 0,807, in het tweede geval 0,880. Dat zijn hoge waarden. In sociaal-wetenschappelijk onderzoek wordt doorgaans de grens 0,80 gebruikt als de limiet waarboven een test een hoge interne betrouwbaarheid vertoont; door sommige onderzoekers worden ook lagere waarden geaccepteerd.¹¹ Door het weglaten van telkens één toetsitem en vervolgens opnieuw Cronbach's alfa voor de voormeting en de nameting te bepalen is nagegaan of bepaalde items een onevenredige invloed hadden op de interne betrouwbaarheid van de toets. Zowel bij de voormeting als de nameting was Cronbach's alfa bij weglating van een item doorgaans lager, met vijf uitzonderingen: bij de voormeting maakte het weglaten van item 1, 4, 12, 15 en 19 de waarde van alfa hoger, bij de nameting het weglaten van item 1, 4, 17, 19 en 25. Bij zowel de voormeting als de nameting hadden items 1, 4 en 19 dus een negatieve invloed op Cronbach's alfa. Een verklaring daarvoor bij item 1 zou kunnen zijn dat een 'leeg eiland' sterke associaties oproept met de periode vóór de Europese kolonisatie, wat echter onjuist is. Item 4 is van een andere aard dan de overige, omdat het over kennis van actualiteit gaat. Item 19 is een erg lastig item waarbij grondige kennis en het leggen van veel

11 Cohen, Manion & Morrison (2007), p. 506. Bryman (2008), p. 151.

verbanden komt kijken: de Nederlanders zijn de baas, die zijn protestants en willen dus geen katholieke priesters, maar de indiaanse bevolking van Aruba is (door de Spanjaarden) katholiek geworden en moet dus in het geheim bediend worden. Dat kan alleen gelden voor de 17^e en 18^e eeuw, maar er is heel wat redeneerkunst voor nodig om dat vast te stellen.

De items 12 en 15 riepen tijdens de voormeting associaties op met resp. de tijd van de slavernij en de tijd van de ontdekkingen, wat in beide gevallen onjuist is. Item 17 was tijdens de nameting één van de moeilijkste, omdat hij valt in de tussenperiode van het einde van het Spaanse gezeg en het moment waarop de Nederlanders het overnamen. Item 25 werd vaak in verband gebracht met het moment waarop de Nederlanders voor het eerst Aruba veroverden (en dus de balans opmaakten van wat er daar te vinden was), maar in werkelijkheid gebeurde dit – zoals tijdens de les vermeld is – aan het begin van de negentiende eeuw. De items die Cronbach's alfa negatief beïnvloedden waren items die tijdens de voormeting inspeelden op onjuiste associaties, en tijdens de nameting meer vergden dan enkelvoudige associaties met kenmerken van bepaalde tijdvakken of perioden.

Als bij de voormeting alle vijf de items met negatieve invloed worden weggelaten, blijkt Cronbach's alfa voor de overige twintig uit te komen op 0,827. Bij de nameting is alfa bij weglating van de vijf items met negatieve invloed voor de overige twintig 0,894. Dat zijn weliswaar hogere waarden dan voor de gehele test van 25 items, maar niet zodanig hoger dat besloten is de bewuste items in de verdere analyses niet mee te nemen.

De interne betrouwbaarheid van de test bij de vier eerder omschreven subcategorieën van toetsitems (zie tabel 4-6) is bepaald voor de nameting. Voor de voormeting heeft dit geen zin, aangezien de categorisering betrekking heeft op de leerstof die tijdens de interventie aan de leerlingen is aangeboden. De alfa-waarden voor de diverse subcategorieën zijn te vinden in tabel 4-7. De interne betrouwbaarheid van alle vier de subcategorieën ligt op een acceptabel niveau.

Tabel 4.7 Cronbach's alfa voor subcategorieën van toetsitems tijdens de nameting

Subcategorie	Cronbach's alfa
Mét sleutelwoorden	0,782 / 0,817*
Zonder sleutelwoorden	0,806
Reproductie	0,758
Toepassing	0,813

* De waarde van alfa is bepaald voor alle acht gebeurtenissen met sleutelwoorden en voor de zes die de leerlingen van groep B op grond van woordherkenning in de juiste richting leidden. Het tweede getal is de laatste waarde.

4.3 Resultaten

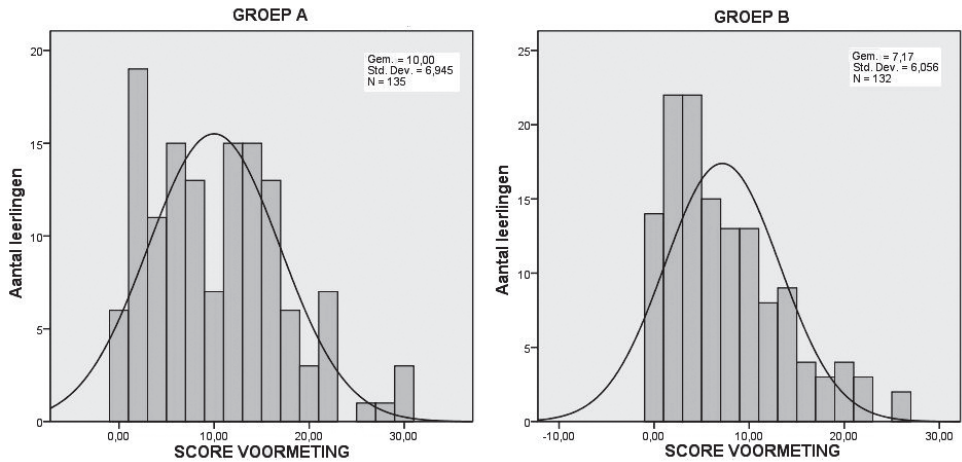
De voormeting

Een verschil in scores tussen groep A en groep B tijdens de voormeting zou kunnen inhouden dat de ene groep bij voorbaat méér in staat was om gebeurtenissen juist in de tijd te plaatsen dan de andere. Daarmee zou dan rekening moeten worden gehouden bij de interpretatie van de resultaten van de nameting. Voor elk item was de maximale score 2, zoals eerder uiteengezet. De hoge waarde van Cronbach's alfa rechtvaardigde het optellen van de scores op alle vragen van de voormeting bij elkaar, omdat in belangrijke mate hetzelfde met deze vragen werd gemeten. Daarom werd de totale score van de voormeting bepaald via optelling van de scores op elke afzonderlijke vraag, waardoor een maximale score van 50 (25 maal 2) mogelijk werd. Uit het resultaat van de voormeting bleek dat de gemiddelde score van alle 267 leerlingen 8,60 was, de maximale score 29 en de minimale 0. Tabel 4-8 toont de gemiddelde resultaten uitgesplitst over groep A en groep B.

Tabel 4.8 Gemiddelde scores op de voormeting

	Groep	N	Gem.	Std. deviatie
Voormeting	A	135	10,0000	6,94542
	B	132	7,1667	6,05593

Bij vergelijking van de gegevens van de twee groepen blijkt dat het verschil aanzienlijk en volgens een toegepaste t-toets statistisch significant is ($t = 3,550$; $df = 265$; $p < 0,001$). De verwachting dat groep A beter zou scoren dan groep B, wordt door deze gegevens bevestigd. Het verschil is veel groter dan te verwachten zou zijn op grond van de eerder geconstateerde kleine verschillen in de beginsituatie. In figuur 4-10 is te zien hoe de scores in beide groepen verdeeld zijn: in beide gevallen een 'scheve' verdeling naar de linkerkant van het histogram, wat wordt aangeduid als 'positief scheef'. In groep A bedraagt deze positieve scheefheid 0,640 en in groep B 0,997, wat inhoudt dat het aantal scores in het lagere gebied links op het histogram in groep B relatief groter is dan in groep A.



Figuur 4.10 Scores op de voormeting

Eenzestig leerlingen scoorden nul, één of twee punten, wat aangeeft dat er inderdaad nogal vaak voor de verleiding is gezicht om maar in te vullen 'ik weet het niet', ook als dat mogelijk niet overeenstemde met wat werkelijk het geval was. Het aantal leerlingen dat op de voormeting 0 tot 2 punten scoorde (en dus relatief vaak gezicht was voor de verleiding om overall 'ik weet het niet' in te vullen) bedroeg in groep A 25 (18,5%) en in groep B 36 (27,3%). Dat is een aanzienlijk verschil. Een verklaring voor de verschillen tussen beide groepen zou kunnen zijn dat groep A tijdens de voormeting beter in staat was gebeurtenissen juist te plaatsen dan groep B. Groep A heeft in dat geval een aanmerkelijk beter uitgangsniveau gehad dan groep B, reden waarom de score op de voormeting als mogelijke beïnvloedende covariabele bij de interpretatie van de resultaten van de nameting wordt betrokken. Aangenomen wordt dat de eerder gebleken geringe verschillen in beginsituatie tussen groep A en B kunnen worden beschouwd als een oorzaak van het verschil op de voormeting, zodat die geringe verschillen via het meewegen van de voormeting in de analyses worden meegenomen.

De nameting

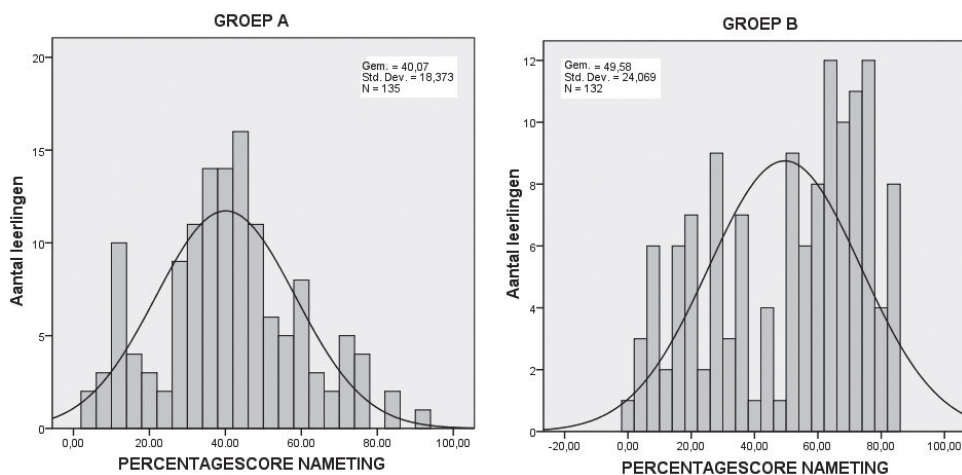
Voor de berekening van de resultaten van de nameting werden de totaalscores omgezet in percentagescores (het percentage van de maximale score 50 dat behaald was). Dat is gedaan om vergelijking met deelscores, ook uitgedrukt in percentages, beter mogelijk te maken. De hoge waarde van Cronbach's alfa rechtvaardigde het optellen van de scores op alle vragen van de nameting bij elkaar, omdat in belangrijke mate hetzelfde met deze vragen werd gemeten. Daarom werd de totale score van de nameting bepaald via optelling van de scores op elke afzonderlijke vraag en vervolgens het percentage van de maximaal mogelijke score te berekenen. Uit het resultaat van de nameting bleek dat de gemiddelde percentagescore van alle 267 leerlingen 44,8% was, de maximale percenta-

gescore 90,0 % en de minimale 0,0%. Tabel 4-9 toont de gemiddelde resultaten uitgesplitst over groep A en groep B.

Tabel 4.9 Gemiddelde percentagescores op de nameting

	Groep	N	Gem.	Std. deviatie
NAMETING	A	135	40,0741	18,37326
	B	132	49,5758	24,06864

Het verschil tussen beide groepen is omgekeerd ten opzichte van het verschil op de voormeting: daar was groep A de groep die hoger scoorde, hier is het groep B. Het verschil is aanzienlijk en volgens een uitgevoerde t-test statistisch significant ($t = 3,620$; $df = 245.041$; $p < 0,001$). In figuur 4-11 is te zien hoe de scores in beide groepen verdeeld zijn: in het geval van groep A een ongeveer normale verdeling met één uitschieter aan de linkerkant, in groep B een scheve verdeling naar rechts. De scheefheid is ten opzichte van de voormeting in groep A veranderd van 0,640 naar 0,232, en in groep B van 0,997 naar -0,452. Dat impliceert een veel grotere verschuiving in groep B dan in groep A. Bij de voormeting was er in beide groepen een overhelling naar het gebied van de lagere scores, meer in groep B dan in groep A. Bij de nameting was er in groep A nog steeds een geringe overhelling naar het gebied van de lagere scores, minder dan op de voormeting. In groep B was de overhelling omgeslagen van het gebied van de lagere scores naar dat van hogere scores (van positieve scheefheid naar negatieve scheefheid).



Figuur 4.11 Scores op de nameting

Bij vergelijking van de scores op de afzonderlijke gebeurtenissen (appendix 2, bijlage 6) is te zien dat tijdens de voormeting groep A op 24 van de 25 gebeurtenissen

hoger scoorde dan groep B. Tijdens de nameting is dat beeld omgekeerd: groep B scoorde op 23 van de 25 gebeurtenissen hoger dan groep A. Voor de nameting is met een paarsgewijze t-toets vastgesteld op welke items significant hoger werd gescoord op de nameting ten opzichte van de voormeting. In groep A bleek dat bij zestien items het geval, in groep B bij tweeëntwintig items. In groep B was de t-waarde van het verschil in drieëntwintig gevallen groter dan in groep A (appendix 2, bijlage 7).

Voorafgaand aan de vijftientwintig toetsitems waarbij leerlingen gebeurtenissen in het juiste tijdvak moesten plaatsen, werd op de nameting aan de leerlingen gevraagd of ze in de week tussen de experimentele les en de nameting de uitgereikte schriftelijke samenvatting nog hadden gelezen of geleerd (zie appendix 2, bijlage 4E). Het aantal leerlingen dat zei te hebben *geleerd* was in beide groepen heel klein. Bovendien is het achteraf de vraag of er een scherp onderscheid tussen de categorieën gelezen en geleerd kan worden gemaakt. Daarom werd op basis van deze vragen een nominale variabele gecreëerd: gelezen/geleerd 'ja' of 'nee'. Deze blijkt een aanzienlijk effect te hebben op de scores van de nameting. In tabel 4-10 zijn de aantallen leerlingen te vinden die in beide groepen zeiden te hebben gelezen of geleerd en de gemiddelde scores van de op deze wijze te onderscheiden deelgroepen.

Tabel 4.10 [Tabel 4-10] Resultaten van de nameting voor de groepen die zeiden de uitgereikte samenvatting van de leerstof wel of niet te hebben gelezen of geleerd

	Groep	N	Gem.	Std. deviatie
Nameting bij NIET gelezen / geleerd	A	40	30,90	15,60
	B	51	34,78	22,09
	Groep	N	Gem.	Std. deviatie
Nameting bij WEL gelezen / geleerd	A	95	43,94	18,15
	B	80	59,30	20,19

Uit de gegevens blijkt duidelijk de grote invloed van het al dan niet kennismaken van de schriftelijke samenvatting in beide groepen. Dat effect is in groep B het grootst. In beide gevallen blijkt het gemiddelde resultaat van groep B hoger te zijn dan dat van groep A, maar in het geval van de groep leerlingen die zeiden te hebben gelezen of geleerd was het verschil veel groter dan in het andere geval. Bij toepassing van een t-toets bleek het verschil tussen groep A en B bij de leerlingen die zeiden *niet* te hebben geleerd statistisch niet significant ($t = 0,982$; $df = 88,129$; $p = 0,329$), bij de leerlingen die zeiden dat *wel* te hebben gedaan was het verschil tussen groep A en B statistisch significant ($t = 5,298$; $df = 173$; $p < 0,001$). Daarom leek het verstandig bij de verdere verwerking van de gegevens onderscheid te maken tussen leerlingen die *wel* of *niet* zeiden te hebben gelezen of geleerd. Daardoor zijn na de interventie vier *condities* tot stand gekomen:

- leerlingen die de les volgens de chronologische aanpak hadden gevolgd en *wel* zeiden te hebben gelezen of geleerd (conditie 1);

- leerlingen die de les volgens de beeldend-associatieve aanpak hadden gevolgd en *wel* zeiden te hebben gelezen of geleerd (conditie 2);
- leerlingen die de les volgens de chronologische aanpak hadden gevolgd en *niet* hadden gelezen of geleerd (conditie 3);
- leerlingen die de les volgens de beeldend-associatieve aanpak hadden gevolgd en *niet* hadden gelezen of geleerd (conditie 4).

In tabel 4-11 worden de totaalscores en de deelscores op subcategorieën van vragen voor alle leerlingen en voor de vier condities bijeengebracht.

Tabel 4.11 Overzicht van percentagescores op de nameting

		Alle leerlingen					
		Nameting			Nameting zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	N	135	135	135	135	135	135
	Gem.	40,07	37,49	41,53	40,19	36,24	42,96
	Std. deviatie	18,37	21,51	19,28	18,28	21,98	19,63
B	N	132	132	132	132	132	132
	Gem.	49,58*	48,95*	49,93*	47,93*	45,29*	49,77
	Std. deviatie	24,07	27,30	23,86	22,72	25,79	23,74

		Leerlingen die zeiden te hebben gelezen of geleerd					
		Nameting			Nameting zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	N	95	95	95	95	95	95
	Gem.	43,94	41,99	45,03	43,44	40,83	45,26
	Std. deviatie	18,15	22,03	18,76	18,43	22,68	19,40
B	N	80	80	80	80	80	80
	Gem.	59,30*	59,10*	59,41*	56,54*	54,46*	58,00*
	Std. deviatie	20,19	23,94	20,22	18,89	23,47	20,29

		Leerlingen die zeiden niet te hebben gelezen of geleerd					
		Nameting			Nameting zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	N	40	40	40	40	40	40
	Ge.	30,90	26,81	33,20	32,50	25,36	37,50
	Std. deviatie	15,60	15,96	18,12	15,59	15,76	19,32
B	N	52	52	52	52	52	52
	Gem.	34,62	33,33	35,34	34,67	31,18	37,12
	Std. deviatie	21,90	24,85	21,70	21,85	22,81	23,25

* Significant verschil van groep B t.o.v. groep A ($p < 0,005$), zie appendix 2 bijlage 8.

In het overzicht is met vetgedrukte getallen en het teken * aangegeven waar volgens een toegepaste t-toets sprake is van een significant verschil tussen de resultaten van groep B ten opzichte van groep A. Bij deze significante verschillen scoort groep B telkens hoger dan groep A. Geen significant verschil kan worden aangetoond voor subcategorie 'toepassing zonder sleutelwoorden' bij de meting over alle leerlingen; wel voor de leerlingen die zeiden te hebben gelezen of geleerd. In het geval van de leerlingen die zeiden *niet* te hebben gelezen of geleerd is het verschil tussen groep A en groep B nergens statistisch significant.

Het effect van de vier condities op de nameting kan alleen zuiver worden bepaald als het gecorrigeerd wordt voor de invloed van de voormeting. Uit de voormeting bleken immers significante verschillen tussen de groepen. In hoeverre kan het verschillende resultaat van de nameting worden toegeschreven aan de verschillende condities (dus aan de interventies) en in hoeverre aan de verschillen tussen de groepen die er reeds voor de interventie waren (het resultaat van de voormeting)? Via een variantie-analyse met de toegepaste *lesmethode* en de factor *gelezen/geleerd* als onafhankelijke variabelen (deze bepaalden immers de conditie), *voormeting* als *covariabele*, en *nameting* als afhankelijke variabele kon worden vastgesteld wat het effect van de interventies is geweest gecorrigeerd voor het verschil op de voormeting. Deze ANCOVA toonde significante effecten van zowel de toegepaste *lesmethode* ($F(1, 261) = 28,969, p < 0,001, \eta_p^2 = 0.100$), als de factor *gelezen/geleerd* ($F(1, 261) = 54,708, p < 0,001, \eta_p^2 = 0.173$), als voormeting ($F(1, 261) = 56,308, p < 0.001, \eta_p^2 = 0.177$) op de nameting. Ook was er sprake van een significant interactie-effect tussen de toegepaste *lesmethode* en de factor *gelezen/geleerd* ($F(1, 261) = 10,832, p = 0,001, \eta_p^2 = 0.040$), dat wil zeggen dat het al dan niet hebben gelezen of geleerd significant invloed had op het effect van de toegepaste *lesmethode*. Dat effect was in groep B groter dan in groep A: het voor de invloed voormeting gecorrigeerde resultaat van de nameting kwam in groep A uit op 31,73% (SD 2,77) voor de groep die zei niet te hebben gelezen of geleerd en op 41,13% (SD 1,83) voor de groep die zei dat wél te hebben gedaan. In groep B waren deze getallen resp. 36,61% (SD 2,46) en 61,06% (SD 1,97). De twee verschillende *lesmethoden* alsook de verschillen in het al dan niet hebben gelezen of geleerd (de interventies) hebben dus aantoonbaar invloed gehad op de nameting, ook als rekening gehouden wordt met de verschillen op de voormeting. Deze verschillen werden niet veroorzaakt door mogelijke afwijkingen in de groepsgemiddelden veroorzaakt door verschillen *binnen* groepen, maar door verschillen *tussen* groepen.

Ook op een andere manier kon het effect van de voormeting en dat van de interventies op de nameting worden vastgesteld, namelijk door de Pearson-correlaties te bepalen tussen voormeting en nameting. Zijn deze aanzienlijk en significant, dan is er sprake van een relatief sterke relatie tussen voormeting en nameting. Zijn ze kleiner en niet significant, dan kan worden verondersteld dat de score op de nameting moet worden toegeschreven aan iets anders dan de voormeting, en dat kan eigenlijk alleen de interventie zijn. In tabel 4-12 zijn de Pearson-correlaties verzameld voor alle leerlingen

en afzonderlijk voor de vier de condities, totalen en uitgesplitst over zes verschillende categorieën van toetsitems.

Tabel 4.12 Correlaties tussen voormeting en nameting

Correlaties tussen voormeting en nameting Alle leerlingen							
		Nameting			Nameting items zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	Pearson-correlatie	0,543*	0,385*	0,566*	0,501*	0,330*	0,535*
	p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
	N	135	135	135	135	135	135
B	Pearson-correlatie	0,298	0,255	0,305*	0,311*	0,242	0,321*
	p	0,001	0,003	< 0,001	< 0,001	0,005	< 0,001
	N	132	132	132	132	132	132

Correlaties tussen voormeting en nameting Leerlingen die zeiden te hebben gelezen of geleerd							
		Nameting			Nameting items zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	Pearson-correlatie	0,545*	0,392*	0,565*	0,516*	0,361*	0,539*
	p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
	N	95	95	95	95	95	95
B	Pearson-correlatie	0,164	0,138	0,164	0,146	0,124	0,131
	p	0,146	0,222	0,146	0,197	0,275	0,248
	N	80	80	80	80	80	80

Correlaties tussen voormeting en nameting Leerlingen die zeiden niet te hebben gelezen of geleerd							
		Nameting			Nameting items zonder sleutelwoorden		
Groep		Tot.	Repr.	Toep.	Tot.	Repr.	Toep.
A	Pearson-correlatie	0,435	0,199	0,487*	0,350	0,029	0,463*
	p	0,005	0,218	0,001	0,027	0,860	0,003
	N	40	40	40	40	40	40
B	Pearson-correlatie	0,598*	0,506*	0,618*	0,630*	0,500*	0,664*
	p	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001	< 0,001
	N	52	52	52	52	52	52

De correlaties zijn in het geval van alle leerlingen en in het geval van leerlingen die zeiden te hebben gelezen of geleerd voor groep A in alle subcategorieën van toetsitems groter dan die voor groep B, en voor groep A zijn ze overal significant ($p < 0,001$, in de tabel aangegeven met * en vetgedrukte getallen), in groep B alleen in de beide subcategorieën 'toepassing' en de subcategorie 'totaal, items zonder sleutelwoorden, reproduc-

tie' bij alle leerlingen samen. Ook hier was de correlatie echter aanzienlijk minder sterk dan die in groep A. Dat wil zeggen dat in het geval van alle leerlingen samen en het geval van de leerlingen die zeiden te hebben gelezen of geleerd de voormeting in grotere mate samenhang met de nameting in groep A dan in groep B, m.a.w.: verondersteld kan worden dat bij de leerlingen van groep B het resultaat van de nameting in grotere mate toe te schrijven is aan iets anders dan de voormeting, dus aan de interventie, het duidelijkst bij de groep leerlingen die zeiden de schriftelijke samenvatting te hebben gelezen of geleerd. In het geval van de leerlingen die zeiden *niet* te hebben gelezen of geleerd ligt dat precies andersom: daar is de correlatie tussen voormeting en nameting sterker in B dan in groep A, en dus waarschijnlijk het effect van de interventie verhoudingsgewijs minder. Wat reeds eerder bleek bij de ANCOVA, blijkt ook hier: het effect van de beeldend-associatieve aanpak is het duidelijkst vast te stellen onder de leerlingen die de schriftelijke samenvatting zeiden te hebben gelezen of geleerd.

Waardering van beide aanpakken door leerlingen

Eén van de vragen die voorafgaand aan de nameting moesten worden beantwoord was de vraag of leerlingen tijdens de experimentele les hadden opgelet of niet (antwoord ja (=1) of nee (= 0)). De gemiddelde score van groep A op deze vraag was 0,85 en die van groep B 0,92, een klein en statistisch niet significant verschil ($t = 1,639$; $df = 252,523$; $p = 0,102$). Aan de mate waarin leerlingen tijdens de les hadden opgelet kan een eventueel verschil in waardering en/of verschil in prestatie dus niet worden toegeschreven.

Aan het einde van de nameting konden de leerlingen hun mening geven over de les die ze gevolgd hadden (zie appendix 2, bijlage 4F). Gevraagd werd of ze de les leuk en duidelijk vonden. Ook hun mening over de leraar werd gevraagd: vonden ze de leraar aardig of niet, en welk rapportcijfer zouden ze hem geven? Bij de vragen over leuk, duidelijk en aardig kon gescoord worden op een vijfpuntsschaal, variërend tussen 0 (helemaal niet leuk, helemaal niet duidelijk en helemaal niet aardig) en 4 (erg leuk, erg duidelijk en erg aardig). Het gemiddelde resultaat van deze meting is te zien in tabel 4-13.

Tabel 4.13 Waardering van leerlingen van de les over Aruba en de leraar.

	Groep	N	Gem.	Std. deviatie
Vond les leuk	A	134	2,18	0,93
	B	130	2,37	0,92
Vond les duidelijk	A	134	2,51	0,96
	B	130	3,00	0,77
Vond leraar aardig	A	134	2,75	0,69
	B	130	3,03	0,61

Het verschil tussen groep A en B bij de vraag 'vond les leuk' is statistisch niet significant ($t = 1,678$; $df = 262$; $p = 0,95$). In beide groepen vonden de leerlingen de les meer dan gemiddeld leuk. De leerlingen van groep B vonden de les wel significant duidelijker

dan de leerlingen van groep A ($t = 4,623$; $df = 253,229$; $p < 0,001$). Blijkbaar vertaalde dat zich ook in de waardering voor de leraar. Het verschil is hier overigens weer kleiner, maar nog wel statistisch significant ($t = 3,461$; $df = 262$; $p = 0,001$). Beide groepen vonden de leraar overigens in behoorlijke mate aardig. Het gemiddelde rapportcijfer dat leerlingen de leraar gaven kwam in groep A uit op 7,0, en groep B op 7,6.

4.4 Bespreking van de resultaten

Uit de bespreking van de verschillen tussen groep A en B in 4.1 bleek enige voor-sprong van groep A ten opzichte van groep B (meer vwo-leerlingen, iets betere beheersing van chronologie) op grond waarvan werd verwacht dat groep A tijdens voor- en nameting iets beter zou presteren dan groep B. Wat betreft de voormeting is deze verwachting bevestigd, zij het dat het verschil veel groter was dan het geringe gemeten verschil in beginsituatie deed vermoeden: groep A scoorde een aanzienlijk hoger resultaat dan groep B, met een statistisch significant verschil. Het lijkt dus duidelijk dat voorafgaand aan de interventie groep A een hoger niveau had dan groep B. Na de interventie bleek het beeld omgekeerd: groep B scoorde nu aanmerkelijk hoger dan groep A met een statistisch significant verschil. De verwachting werd hier dus niet bevestigd, wat impliceert dat groep B tijdens de interventie een zodanig beter referentiekader heeft opgebouwd dan groep A, dat een beter resultaat kon worden gerealiseerd. Aan iets anders dan de interventie is dit resultaat moeilijk toe te schrijven, gezien de factoren waarvoor is gecontroleerd. Het resultaat van de variantie-analyse maakte duidelijk dat de invloed van de resultaten op de voormeting (hoger uitgangsniveau) aanmerkelijk is geweest, maar dat verschillen tussen de groepen niet alleen aan de voormeting konden worden toegeschreven. Verschillen moeten volgens deze analyse minstens voor een deel zijn veroorzaakt door de interventie.

De verdeling van de resultaten zoals die blijkt uit de histogrammen van figuur 4-10 en 4-11 wijst in dezelfde richting. Op de voormeting is in beide groepen – zoals te verwachten is – sprake van een positief scheve verdeling, d.w.z.: er zijn relatief veel resultaten in het gebied van de lagere scores; de leerlingen wisten immers nog weinig of niets van het onderwerp. In groep B is deze scheefheid groter dan in groep A, wat overeenstemt met het gemiddeld lagere resultaat van groep B op de voormeting. Tijdens de nameting is er in groep A nog steeds een licht positief-scheve verdeling, waarschijnlijk veroorzaakt door de ene – moeilijk verklaarbare – uitschieter in het gebied rond de 10%. Voor het overige is de verdeling van de resultaten in groep A op de nameting redelijk 'normaal'. In groep B is duidelijk iets anders aan de hand. De positief-scheve verdeling is omgeslagen in een negatief-scheve, wat een duidelijke omslag suggereert. De scheve verdeling in de richting van het gebied tussen 60% en 80% laat zien dat een behoorlijk deel van de leerlingen in groep B een relatief hoog resultaat heeft geboekt. Daar staat echter tegenover dat aan de andere zijde van het histogram ook enkele uitschieters te constateren zijn. Opmerkelijk is dat in het middengebied – waar bij een normale verdeling de hoogste aantallen resultaten te verwachten zijn – bij groep B een

inzinking is te constateren. Blijkbaar heeft de beeldend-associatieve aanpak vaak óf heel goed gewerkt, óf niet goed, maar niet vaak gemiddeld.

Binnen de interventie was sprake van twee ‘media’: de les die volgens de beeldend-associatieve dan wel de chronologische aanpak werd gegeven, en de schriftelijke samenvatting die conform deze twee aanpakken was vormgegeven. Uit de analyses is gebleken dat het al dan niet bestuderen van deze samenvatting een relatief grote invloed heeft gehad op het resultaat van de nameting. Behalve het verschil tussen genummerde eeuwen en tijdvakken met beeldende namen vertonen de schriftelijke samenvattingen ook het verschil tussen enerzijds systematisch opgesomde chronologische informatie, anderzijds meer beeldmateriaal en losse blokjes informatie met minder systematische samenhang (zie appendix 2, bijlage 3). Een vraag is hoe groot de invloed van dit beeldmateriaal – dat overigens wel past bij een oriëntatie in de tijd volgens een beeldende methode – geweest is. Verondersteld zou kunnen worden dat het op het eerste gezicht aantrekkelijkere uiterlijk van het beeldend-associatieve materiaal positief heeft gewerkt op de motivatie daar kennis van te nemen. Dat blijkt niet uit de aantallen leerlingen in groep A en B die zeiden de samenvatting te hebben gelezen of geleerd: in groep A waren dat 95 leerlingen (70,4%) en in groep B 80 leerlingen (61,1%). Dit wijst eerder op grotere ijver en motivatie in groep A – overeenstemmend met hun betere beginsituatie. Desondanks heeft groep A op de nameting een minder goed resultaat behaald. Hoe groot de invloed van het beeldmateriaal is geweest, is hier niet vast te stellen.

Het verschil tussen de beide groepen in de gevallen waarin wél was gelezen of geleerd, was veel groter dan het verschil tussen beide groepen in de gevallen waarin niet was gelezen of geleerd, en ook groter dan het verschil in het algemeen tussen beide groepen. Het verschil tussen de groepen leerlingen die zeiden niet te hebben gelezen of geleerd, was statistisch niet significant. Dat zou kunnen inhouden dat de les op zichzelf geen sterk genoeg middel is geweest om een verschil tussen beide aanpakken te genereren. Het groepje leerlingen dat niet had gelezen of geleerd was echter ook klein, en naarmate een groepje kleiner is, wordt het moeilijker een significant verschil aan te tonen – anders gezegd: worden er hogere ‘eisen’ aan het verschil gesteld om nog significant te kunnen zijn.¹² Ook kan worden verondersteld dat de leerlingen die zeiden niet te hebben gelezen of geleerd minder ijverige, geïnteresseerde leerlingen zijn geweest, op wie wellicht ook de les een minder grote invloed heeft gehad. In elk geval is bij de minderheid van leerlingen die de schriftelijke samenvattingen niet zei te hebben gelezen of geleerd, geen verschil tussen de aanpakken aangetoond. De leeropbrengst van de lessen, zo zou men kunnen concluderen, is dus nogal gering geweest. Dat is begrijpelijk omdat het slechts om één les ging die bovendien in beide gevallen de vorm had van een lang hoorcollege – niet direct een optimale omstandigheid om veel kennis op te doen. Het is daarom gerechtvaardigd te veronderstellen dat een beeldend-associatieve aanpak in een reguliere onderwijssituatie, waarin meer ruimte is voor een gevarieerd onderwijsaanbod, leerlingen hun huiswerk leren en zich grondig voorbereiden op toetsen, tot meer leerwinst leidt dan een op chronologie gebaseerde aanpak.

12 Cohen, Manion & Morrison (2007), p. 516-517.

Hoe groot is het effect is geweest van het voorkomen van bepaalde overeenkomende woorden in de namen van tijdvakken en de omschrijvingen van gebeurtenissen? Het lijkt logisch dat leerlingen van groep B een gebeurtenis waarbij (bijvoorbeeld) 'olie' in de beschrijving voorkomt relatief gemakkelijk associeerden met 'de tijd van de olie'. Groep A beschikte niet over deze gemakkelijke associaties. In hoeverre was het grotere leereffect in groep B aan dit soort mechanismen toe te schrijven? Indien in grote mate, dan is relatief minder sprake geweest van echte historische oriëntatie en meer van identificatie op grond van woordherkenning. Zoals al eerder is opgemerkt, is woordherkenning niet in alle gevallen voor de leerlingen van groep B positief geweest. Bij twee gebeurtenissen brachten woorden in de omschrijving hen op een dwaalspoor: bij gebeurtenis 15, waarin een 'indiaanse' hoofdman wordt genoemd, die ten onrechte een plaatsing in de 'tijd van de indianen' zou kunnen oproepen, en bij gebeurtenis 17, waarbij zowel indianen als paarden voorkomen, terwijl de gebeurtenis noch in de tijd van de indianen, noch in die van paarden en piraten thuishoort. Gebeurtenis 17 is dan ook één van de gebeurtenissen waarop groep A een hoger resultaat scoort dan groep B (zie appendix 2, bijlage 6 en 7). In zes andere gevallen werkte deze koppeling echter wél in het voordeel van de leerlingen van groep B. Nu is het wegcijferen hiervan wellicht niet helemaal terecht, omdat de beeldend-associatieve wijze van zich oriënteren in de tijd mede op dit soort associaties gebaseerd is en dus met reden een voordeel voor groep B heeft opgeleverd. Geconstateerd is echter dat ook bij weglating van de acht gebeurtenissen met sleutelwoorden in de omschrijving groep B nog steeds aanmerkelijk hoger scoorde dan groep A met een statistisch significant verschil. Het verschil was echter wat minder groot dan bij de nameting over alle 25 gebeurtenissen: groep A haalde een resultaat dat vrijwel exact overeenkwam met de meting over alle 25 gebeurtenissen, terwijl groep B een resultaat scoorde dat iets onder het totaalresultaat over 25 gebeurtenissen lag. Bij paarsgewijze t-toetsen op het verschil tussen voor- en nameting blijkt in de acht gevallen waarin groep B wél een significante vooruitgang boekt en groep A níét er slechts één te zijn waarbij deze vooruitgang deels aan de herkenning van een sleutelwoord kan worden toegeschreven (item 8), terwijl er vier gevallen zijn van items uit de categorie toepassing zonder sleutelwoorden (items 4, 7, 10 en 11). Het grotere leereffect van de leerlingen in groep B is dus in geringe mate gebaseerd geweest op het herkennen van woorden uit de namen van tijdvakken. Echt 'associatieve' factoren zullen waarschijnlijk een grotere rol hebben gespeeld.

Als leerlingen zich gebeurtenissen herinnerden die in de lessen waren genoemd en deze juist konden plaatsen, kwam dat neer op reproductie van kennis. Bij gebeurtenissen die niet in de lessen voorkwamen, ging het bij juiste plaatsing om toepassing van kennis, een hogere categorie van onderwijsdoelstellingen dan reproductie. Geconstateerd is dat leerlingen van groep B in alle gevallen een hoger resultaat scoorden op de nameting dan leerlingen van groep A, zij het dat het verschil in het geval van het niveau 'toepassing' alleen significant is als dat gemeten wordt over de leerlingen die zeiden te hebben gelezen en geleerd (zie tabel 4-11). Dat betekent dat de beeldend-associatieve aanpak leerlingen niet alleen beter helpt bij het zich herinneren van gebeurtenissen die in de les zijn behandeld, maar ook bij het plaatsen van 'nieuwe' gebeurtenissen. De

vraag is in hoeverre dat laatste in het geval van dit experiment bepaald werd door het voorkomen van een overeenkomstig woord in de naam van een tijdvak en de beschrijving van een gebeurtenis. Uit de gegevens van tabel 4-11 voor de totale groep en de groep die zei te hebben gelezen en geleerd (de groep die zei dat niet te hebben gedaan laten we hier buiten beschouwing, omdat het effect van de interventie daar sowieso gering is geweest) blijkt dat dat in geringe mate het geval is. Groep A scoort in de situatie van weglating van toetsitems met sleutelwoorden op een ongeveer vergelijkbaar niveau als bij alle 25 toetsitems, wat ook te verwachten is, omdat de sleutelwoorden voor groep A uiteraard geen sleutelwoorden waren. In het geval van groep B zien we bij weglating van de toetsitems met sleutelwoorden een licht lagere score dan bij alle 25 gebeurtenissen, vooral in de categorie reproductie. Het verschil tussen de beide groepen wordt dus minder groot, wat erop wijst dat woordherkenning bij groep B enige rol heeft gespeeld, maar zeker niet in een zodanige mate dat het gehele verschil erdoor kan worden verklaard. Zowel bij reproductie van kennis als bij toepassing hebben waarschijnlijk echt associatieve factoren bij groep B een rol gespeeld. Bij een uitbreiding van dit onderzoek met (bijvoorbeeld) hardop denkende leerlingen of interviews zou duidelijk kunnen worden of en zo ja op welke wijze zulke associaties plaats vinden.

Merkwaardig is dat de scores wat betreft reproductie vergeleken met die betreffende toepassing in alle gevallen lager uitvallen, zowel in groep A als in groep B, terwijl reproductie van stof geacht wordt een lager niveau te vertegenwoordigen. Waaraan dat ligt is moeilijk vast te stellen. Misschien is de confrontatie met de leerstof in één les en op grond van niet zeer grondige kennisname van de schriftelijke samenvatting daarvan te oppervlakkig geweest om een goed reproductief resultaat te genereren.

Op grond van zowel het resultaat van de ANCOVA als van de vergelijking van de Pearson-correlaties wordt het meer dan aannemelijk dat de geconstateerde verschillen tussen groep A en groep B werkelijk aan de interventie, en niet aan iets anders zijn toe te schrijven. De ANCOVA toont een significant effect van zowel de toegepaste lesmethode als het al dan niet lezen of leren na correctie daarvan voor de invloed van de voormeting. Er zijn, met andere woorden, door de interventie significante verschillen *tussen* groepen ontstaan die niet slechts verklaarbaar zijn uit vreemde verdelingen van verschillen *binnen* groepen en ook niet op grond van een betere beginsituatie.

Dit wordt nog eens bevestigd door de Pearson-correlaties tussen voormeting en nameting. Uit tabel 4-12 bleek dat de correlaties tussen voormeting en nameting, bepaald voor alle leerlingen en voor alleen de leerlingen die zeiden te hebben gelezen en geleerd, in groep A voor alle subcategorieën van toetsitems sterker is dan in groep B. Hierdoor kan worden verondersteld dat het resultaat in groep B in die gevallen in sterkere mate door iets anders dan de voormeting is bepaald, dus door de (in het geval van groep B beeldend-associatieve) interventie. Dat geldt zowel voor reproductie als voor toepassing, en zowel voor de situatie waarin wél als de situatie waarin níét van woordherkenning gebruik kon worden gemaakt. Bij het groepje leerlingen dat zei niet te hebben gelezen of geleerd, was in groep B de relatie tussen voormeting en nameting sterker dan in groep A. Dat beklemtoont nog eens dat het effect van de beeldend-associatieve interventie niet aan te tonen is voor de groep leerlingen die zei niet te hebben

gelezen of geleerd. Al eerder werd geconstateerd dat dat bij het volgen van één les in de vorm van een lange, door powerpoint ondersteunde voordracht, ook niet zo vreemd is, zeker niet als verondersteld wordt dat de leerlingen die niet de moeite hebben genomen de samenvatting nog eens in te zien, waarschijnlijk ook tot de minder goede, minder gemotiveerde leerlingen hebben behoord. Dat laatste kan overigens niet worden vastgesteld op grond van gegevens die tijdens dit experiment zijn gemeten. De groep die zei niet te hebben gelezen of geleerd was ook een aanmerkelijk kleinere groep, waarin moeilijker significante verschillen of effecten kunnen worden aangetoond.

Wat betreft de waardering van de leerlingen voor beide aanpakken is gebleken dat leerlingen van beide groepen de les in behoorlijke mate leuk en de leraar in behoorlijke mate aardig vonden. De leerlingen van beide groepen zeiden ook in ongeveer dezelfde mate dat ze tijdens de les goed hadden opgelet. De leerlingen van groep B vonden de les echter in veel grotere mate duidelijk. Dat stemt overeen met het feit dat ze een veel groter leereffect hebben bereikt. De standaarddeviatie van de antwoorden op de vraag over duidelijkheid is bij groep B veel lager dan bij groep A, wat erop duidt dat er grote overeenstemming was in de meningen van de leerlingen over de duidelijkheid van de beeldend-associatieve aanpak. Aan een verschil in het optreden van de leraren kan dit niet worden toegeschreven, aangezien telkens in parallelle groepen sprake was van dezelfde gastdocent die de les aan de klassen heeft gegeven.

4.5 Conclusies

Om te kunnen vaststellen of oriëntatie in de historische tijd meer ondersteund wordt door chronologische hulpmiddelen of meer door beeldend-associatieve hulpmiddelen, is een parallel onderwijsaanbod gegeven aan twee vergelijkbare groepen leerlingen. De vergelijkbaarheid van de twee groepen was essentieel om conclusies te kunnen trekken. Aangetoond werd dat de twee groepen in hoge mate vergelijkbaar waren, maar dat de groep die geconfronteerd werd met de chronologische aanpak hogere prestaties vertoonde op de voormeting, wat overeenstemde met een aantal andere kenmerken van die groep (meer vwo-leerlingen en een iets betere score op chronologische vaardigheden). Niettemin bereikte deze groep op de nameting geen beter resultaat dan de andere groep. Tijdens de nameting is aangetoond dat de groep die met de beeldend-associatieve aanpak is geconfronteerd op vrijwel alle punten een hoger leereffect bereikte dan de groep die met de chronologische aanpak werd geconfronteerd. De beeldend-associatieve hulpmiddelen lijken dus beter te werken bij een oriëntatie in de historische tijd dan de chronologische. Dat is niet alleen het geval waar sprake is van woordherkenning in de naam van een tijdvak en de beschrijving van een gebeurtenis, maar ook bij toepassing van het aangeleerde kader op onbekende gebeurtenissen. De beeldend-associatieve tijdvakken kunnen dienen als geheugenlocaties waarmee zaken in verband kunnen worden gebracht zonder dat sprake is van woordherkenning of identificatie op grond van jaartallen. Dit geldt zowel op het reproductieve niveau als op het niveau van toepassing: het niveau van correct in de tijd plaatsen van gebeurtenissen die niet in de leerstof

zijn behandeld. De beeldend-associatieve aanpak werd door leerlingen duidelijker gevonden dan de chronologische en meer door hen gewaardeerd. De hier beschreven effecten zullen waarschijnlijk sterker zijn als sprake is van regulier onderwijs waarin leerlingen hun huiswerktaken serieus bestuderen om zich voor te bereiden op een toets.

Geconstateerd kan worden dat de hypothese die aan het begin van dit hoofdstuk is gesteld, is bevestigd op beide punten, zowel wat betreft het zich herinneren van informatie als wat betreft het gebruik van de informatie als referentiekader. Merkwaardig was echter wel dat op de categorie reproductie relatief minder werd gescoord dan op de categorie toepassing, wat zou kunnen betekenen dat het zich herinneren van informatie door de beeldend-associatieve aanpak minder wordt bevorderd dan het gebruik van informatie als referentiekader.

Op grond van de theorie van de 'tijdseilanden' van Friedman en op grond van het onderzoek dat door Levstik en Barton is verricht onder basisschoolleerlingen kon worden verwacht dat de beeldend-associatieve methode beter zou werken dan de chronologische. Het werk van Friedman gaat echter over het autobiografisch geheugen. Het was tot dusverre niet zeker of soortgelijke mechanismen ook bij het leren van geschiedenis, waarbij sprake is van andere tijdsdimensies en waarbij geen sprake is van herinneringen aan zaken die men persoonlijk heeft meegemaakt, ook zouden opgaan. Het in dit hoofdstuk beschreven experiment lijkt daarin verandering te hebben gebracht. Het onderzoek van Barton en Levstik maakte uitsluitend gebruik van beelden. Van gebeurtenissen was bij hen geen sprake. De leerlingen die de afbeeldingen van Barton en Levstik in de tijd lokaliseerden, maakten gebruik van een aantal strategieën die bij het hier beschreven experiment nauwelijks een rol hebben kunnen spelen.¹³ Alleen de strategie van 'vooruitgang' ('wat meer op ons lijkt moet in een recentere tijd worden geplaatst') is door een aantal leerlingen waarschijnlijk gebruikt waar zij modern lijkende zaken als het openen van een winkelcentrum en het aanleggen van een vliegveld in het 'nu' of de jaren na 2000 plaatsten, wat in dit geval geen succes opleverde. Aannemelijk is gemaakt dat leerlingen van het voortgezet onderwijs van de leeftijdsgroep 13 tot 16 jaar op beeldend-associatieve wijze geheugenlocaties creëren waaraan zij historische gebeurtenissen met meer succes kunnen koppelen dan in het geval van het gebruik van jaartallen en eeuwen. Dit stemt overeen met de constatering dat oriëntatie in de tijd met chronologische middelen relatief ver af staat van de wijze waarop mensen zich doorgaan met tijd omgaan.

Het ontwerp van de tijdvakken voor de geschiedenis van Aruba dat bij het hier beschreven experiment is gebruikt, doet denken aan dat van de tien tijdvakken die in 2001 werden voorgesteld in het rapport van de Commissie Historische en Maatschappelijke Vorming.¹⁴ Dat is geen toeval. Toen de Commissie HVM het op tien tijdvakken gebaseerde curriculum voorstelde, werd verondersteld dat leerlingen zich hierdoor gemakkelijker in de tijd zouden kunnen oriënteren en dat het geleerde referentiekader bruikbaar zou zijn en beter zou bekijken dan de tot dan toe gebruikelijke kaders van het

13 Barton (2002), p. 171-174. Zie ook hoofdstuk 3.

14 CHMV (2001), p. 32, 44 en 98.

geschiedenisonderwijs.¹⁵ Een wetenschappelijk fundament voor deze veronderstelling ontbrak. Dit gemis werd onderkend. Zowel de Commissie HVM als de aan haar voorafgaande Commissie Geschiedenisonderwijs wezen op het gebrek aan geschiedenisdidactisch onderzoek en bepleitten een versterking van de wetenschappelijke basis van de geschiedenisdidactiek.¹⁶

Het voorstel voor een geschiedeniscurriculum gebaseerd op tien tijdvakken werd in de loop van de jaren 2005, 2006 en 2007 ingevoerd in de Nederlandse programma's van kerndoelen voor basisonderwijs en onderbouw voortgezet onderwijs en in het eind-examenprogramma geschiedenis voor havo en vwo. De tien tijdvakken hebben sindsdien een algemene geldigheid en bekendheid gekregen. Het is daarom niet langer mogelijk het effect van het gebruik van dat curriculum te meten zonder te maken te krijgen met de storende omstandigheid dat er op de ene school anders aandacht aan wordt besteed dan op de andere. In ieder geval zal de onderzoeker te maken krijgen met leerlingen met voorkennis, nooit met leerlingen die met een schone lei beginnen. Daarom werd in het hier beschreven experiment gekozen voor leerstof uit een totaal onbekende context en voor daarop geënte beeldend-associatieve tijdvakken. Een van de bedoelingen van dit experiment was ondertussen wel te testen of de veronderstellingen waarop de Commissie HVM in 2001 haar curriculum baseerde, juist waren. Het resultaat van dit onderzoek suggereert dat dat het geval was.

15 CHMV (2001), p. 20-21.

16 Commissie Geschiedenisonderwijs (1998), p. 34. CHMV (2001), p. 156.