

Twee ontwerpen voor waardevormend geschiedenisonderwijs: een effectstudie¹

J. A. Schuitema, W. Veugelers, G. Rijlaarsdam en G. ten Dam

Samenvatting

Dit onderzoek toetst de effectiviteit van groepswork in het kader van dialogisch waardevormend onderwijs binnen het vak geschiedenis. Er werden twee verschillende leerarrangementen ontwikkeld en beproefd in het tweede leerjaar van het vwo. Beide beoogden leerlingen te stimuleren zich een mening te vormen, rekening houdend met conflicterende waarden en verschillende perspectieven. De arrangementen verschilden in de hoeveelheid groepswork en klassikaal onderwijs. De kwaliteit van de meningsvorming werd gemeten via korte essays die leerlingen schreven over een moreel dilemma. Deze essays werden beoordeeld op het gebruik van waarden en multiperspectiviteit, en op de kwaliteit van historisch redeneren. De resultaten laten zien dat leerlingen die relatief veel in kleine groepjes werkten, vergeleken met leerlingen die verhoudingsgewijs meer klassikaal les kregen, in hun essays vaker refereerden aan waarden en zich beter rekenschap gaven van verschillende perspectieven.

1 Inleiding

Er is veel geschreven over het bevorderen van de morele en sociale ontwikkeling van leerlingen, maar er is betrekkelijk weinig bekend over de effectiviteit van verschillende didactische benaderingen van waardevormend onderwijs (Solomon, Watson, & Battistich, 2001; Schuitema, Ten Dam, & Veugelers, in druk). Onderzoek richt zich vooral op de invloed van de schoolcultuur op de sociale en morele ontwikkeling van leerlingen (bijvoorbeeld Power, Higgins, & Kohlberg, 1989) en op het morele handelen van de docent (bijvoorbeeld Oser, 1994; Hansen, 2001).

Waarom is er zo weinig didactisch onderzoek naar waardevormend onderwijs? Een mogelijke reden is de scheiding die men

veelal maakt tussen de aandacht voor de morele en sociale vorming van leerlingen enerzijds en het onderwijzen van kennis en vaardigheden binnen de vakdomeinen anderzijds (Ten Dam, Wardekker, Miedema, & Veugelers, 2004). Aandacht voor de waardevormende taak van het onderwijs betreft dan de dagelijkse omgang en regels in de school of speciale lessen. Het 'vakonderwijs' blijft buiten beeld. Wij stellen ons echter op het standpunt dat het verwerven van kennis en vaardigheden niet los gezien kan worden van de morele en sociale vorming van leerlingen. Met de verworven kennis en vaardigheden zouden leerlingen in staat moeten zijn zelfstandig te oordelen en te handelen, opdat zij actief kunnen participeren in de samenleving. Dit betekent dat leerlingen zouden moeten leren te reflecteren op waarden in de leerstof, op de vraag hoe zij zich tot die waarden verhouden en op de – per definitie waardegebonden – maatschappelijke toepassingen van vakken (zie ook Yore, Bisanz, & Hand, 2003; Sadler & Zeidler, 2005). Waardevormend onderwijs is zo onverbrekelijk onderdeel van het vakonderwijs.

Geschiedenis is een van de disciplines waarin een levendige discussie gaande is over de relatie tussen vakonderwijs en waardevormend onderwijs (Wrenn, 1999; Phillips, 2002; Barton & Levstik, 2004). Bij het interpreteren en reconstrueren van het verleden spelen morele waarden een belangrijke rol. Iedere samenleving heeft sterk de behoefte om moreel te oordelen over zaken in het verleden (Barton & Levstik, 2004). Bij leerlingen is dat niet anders: ook zij hebben doorgaans een mening over historische onderwerpen. Het vakonderwijs zou die morele oordelen kunnen benutten door leerlingen zich bewust te maken van hun waarden en hen te leren hun oordeel te staven. Het geschiedenisonderwijs biedt dus in principe goede mogelijkheden om vorm te geven aan waardevormend onderwijs. Onderzoek dat met name in de jaren tachtig is verricht laat

echter zien dat er in de geschiedenisles over het algemeen weinig ruimte is voor de inbreng van leerlingen, voor probleemoplossend leren en voor leeractiviteiten die kritisch denken stimuleren (Wilson, 2001).

Het onderzoek dat we in dit artikel rapporteren is erop gericht inzicht te verkrijgen in welke leeractiviteit het meest effectief is in waardevormend onderwijs binnen het vak geschiedenis. Alvorens de opzet van het onderzoek te bespreken, gaan we eerst in op de vraag wat het doel van waardevormend onderwijs zou moeten zijn en wat uit de literatuur bekend is over didactische richtlijnen.

1.1 Leerdoelen van waardevormend onderwijs

Het doel van waardevormend onderwijs is leerlingen voor te bereiden op deelname aan de samenleving. Constituerende elementen van de huidige westerse samenleving zijn democratie en pluriformiteit (Parker, 1997a; Banks, 2004). Volgens Gutmann en Thompson (1996) wordt een democratische samenleving gekenmerkt doordat waarden als rechtvaardigheid en het algemeen belang ter discussie staan in een voortdurend debat, waarvan de uitkomst niet bij voorbaat vast staat. Participeren in de samenleving betekent dat burgers aan dat debat kunnen bijdragen; het gaat om 'meedenken en meepraten' over kwesties aangaande rechtvaardigheid en het algemeen belang (Parker, 2001; Haste, 2004). Waardevormend onderwijs kan zich daarom niet beperken tot het overdragen van waarden, maar zal vooral gericht moeten zijn op het bevorderen van de attitudes en vaardigheden die leerlingen nodig hebben om hun morele afwegingen te maken en daarover verantwoording af te leggen. Gedurende de laatste decennia wint deze visie op het stimuleren van waardevorming in het onderwijs terrein (zie Veugelers & Vedder, 2003). De waarden van de democratische samenleving dienen in die visie als normatieve context.

Wat betekent dit voor de onderwijspraktijk? Hoe kan het geschiedenisonderwijs bijdragen aan waardevorming? Het gaat erom het vermogen van leerlingen te vergroten om zelfstandig een standpunt in te nemen met betrekking tot waardegebonden kwesties, rekening houdend met conflicterende waarden,

in een pluriforme samenleving (Veugelers, 2000). Multiperspectiviteit is dan een kernbegrip: het is essentieel voor waardevormend onderwijs dat leerlingen leren dat er verscheidene perspectieven mogelijk zijn op morele en sociale vraagstukken en dat hun eigen perspectief een van de mogelijke perspectieven is (Roegholt, 1995; Banks et al., 2005). Leerlingen moeten de bereidheid en de vaardigheden ontwikkelen om verschillende perspectieven te onderzoeken en die te betrekken bij de vorming en verantwoording van hun mening (zie ook Oser, 1986). Vanzelfsprekend zouden leerlingen die vaardigheden voor meningsvorming ook moeten kunnen en willen toepassen buiten de context waarin zij die verworven, bij nieuwe onderwerpen.

1.2 Didactiek voor waardevormend onderwijs

Solomon e.a. (2001) onderscheiden *directe* en *indirecte* didactische benaderingen van waardevormend onderwijs. De directe methode is erop gericht waarden over te dragen. De docent onderwijst waarden, draagt deze expliciet uit en moedigt gebruik aan via beloning en straf. De docent is hier voorbeeld en rolmodel (Doyle, 1997; Elkind & Sweet, 1997). Bij indirecte benaderingen gaat het erom een dialoog op gang te brengen (Noddings, 1995; Grant, 1996; Parker, 1997b; Tappan, 1998). Verondersteld wordt dat leerlingen via de dialoog met anderen over voor hen betekenisvolle onderwerpen zich vaardigheden en attitudes eigen maken, zoals bijvoorbeeld kritische denkvaardigheden voor het oplossen van morele dilemma's (Blatt & Kohlberg, 1975; Ten Dam & Volman, 2004; Frijters, Ten Dam & Rijlaarsdam, in druk), communicatieve vaardigheden (Parker, 1997b; Preskill, 1997) en attitudes als tolerantie en respect voor de mening van anderen (Grant, 1996; Saye, 1998). Vaak wordt een probleemgestuurde aanpak bepleit, waarbij leerlingen tot op zekere hoogte zelfstandig aan het werk zijn en controle krijgen over hun eigen leerproces (Hicks, 2001; Schultz, Barr & Selman, 2001; Beane, 2002), met didactische werkvormen die samenwerken vereisen (Saye, 1998; Murray, 1999; Covell & Howe, 2001). Auteurs die een indirecte

benadering van waardevormend onderwijs voorstaan, benadrukken doorgaans het belang van het klassenklimaat en de schoolcultuur. Leerlingen moeten voldoende zelfvertrouwen hebben en zich vrij voelen om hun mening te uiten, ook als deze afwijkt van die van docenten en andere leerlingen. Het is daarom belangrijk dat docenten en leerlingen in school op een respectvolle manier met elkaar omgaan (Power et al., 1989) en dat er in de klas een klimaat heerst waarin leerlingen gestimuleerd worden hun mening te geven (Covell & Howe, 2001).

Op basis van een uitgebreide analyse van onderzoek naar waardeonderwijs, concluderen Solomon en collega's (2001) dat waardevormend onderwijs via een indirecte didactiek meer effect heeft op de sociale en morele ontwikkeling van leerlingen dan onderwijs via de directe didactiek van waardeoverdracht. Meer specifiek zijn er aanwijzingen dat een indirecte benadering, waarbij leerlingen in groepjes werken, discussies voeren en perspectieven wisselen, onder andere een positief effect hebben op attitudes ten opzichte van de universele rechten van de mens (Covell & Howe, 2001), leiden tot minder racistische attitudes (Schultz et al., 2001) en minder redeneren vanuit eigen belang (McQuaide, Leinhardt, & Stainton, 1999).

Onderzoek naar de effectiviteit van waardevormend onderwijs is niet alleen schaars, ook de kwaliteit van het onderzoek laat te wensen over (Solomon et al., 2001; Schuitema et al., in druk). Veel van de voorgestelde didactische ontwerpen zijn onvoldoende uitgewerkt. Er wordt nauwelijks aandacht besteed aan de attitudes en vaardigheden die nodig zijn voor dialogisch leren (zie bijvoorbeeld Van der Linden & Renshaw, 2004). Daarnaast schiet de kwaliteit van de gehanteerde onderzoeksdesigns veelal tekort: voormetingen of een controlegroep ontbreken. Solomon en collega's (2001) concluderen verder dat in onderzoeken waarbij wel sprake is van een controlegroep over het algemeen een heel programma vergeleken wordt met de afwezigheid van dat programma. Hierdoor is het moeilijk te bepalen wat de effectiviteit is van de afzonderlijke componenten van dat programma en wat de meest effectieve combinatie is.

1.3 Meningsvorming door dialoog

Het onderzoek dat we in dit artikel presenteren, past in de traditie van de indirecte benaderingen van waardevormend onderwijs. Dialogisch leren zien wij als een mogelijk effectieve manier om de meningsvorming van leerlingen te stimuleren in het licht van waarden en multiperspectiviteit. Een dialoog trekt leerlingen bij de leerstof, stimuleert participatie, maakt leerlingen bewust van hun waarden en perspectieven, en confronteert leerlingen met het perspectief van anderen.

Om het gewenste resultaat te realiseren, moet een dialoog aan bepaalde voorwaarden voldoen. Een voor waardevorming effectieve dialoog moet erop gericht zijn overeenstemming te bereiken of de ander te begrijpen (zie bijvoorbeeld Grant, 1996; Parker, 1997b). In een goede dialoog staan de deelnemers open voor elkaars ideeën, bouwen ze daarop voort en gaan ze op een constructieve manier om met verschillende perspectieven. Er is onderzoek waaruit blijkt dat dialoog gericht op co-constructie van ideeën een positief effect heeft op de redeneervaardigheden van leerlingen en op het vermogen om op waarden te reflecteren (Rojas-Drummond & Mercer, 2003; Frijters et al., in druk).

Voor een kwalitatief goede dialoog zijn specifieke attitudes en vaardigheden nodig. Waardevormend onderwijs zou erop gericht moeten zijn deze te ontwikkelen. De volgende vaardigheden en attitudes zijn belangrijk voor het voeren van een effectieve dialoog (Frijters et al., in druk):

- **Uitwisselen:** het kunnen en willen uitwisselen van persoonlijke meningen met anderen;
- **Co-construeren:** het kunnen en willen vormen van een eigen mening in dialoog door gebruik te maken van de meningen van medeleerlingen, de leraar en de leerstof en het kunnen en willen bijdragen aan de meningsvorming van anderen;
- **Verantwoorden:** de eigen mening kunnen en willen verantwoorden naar anderen, rekening houdend met morele waarden.

1.4 Groepswerk versus klassikaal onderwijs

Er zijn verschillende manieren om dialogisch leren vorm te geven. Werken in kleine groep-

jes is een veel gebruikte methode om interactie tussen leerlingen te stimuleren (Covell & Howe, 2001; Beane, 2002). In groepjes kunnen meer leerlingen actief aan de dialoog deelnemen dan in klassikaal verband. Ook hebben leerlingen in kleine groepjes meer verantwoordelijkheid voor het verloop van de dialoog. Ze moeten zelf de discussie reguleren en eventuele conflicten oplossen. Daarnaast wordt in de literatuur ook klassikale discussie aanbevolen als manier om dialoog in de klas vorm te geven (zie bijvoorbeeld Tredway, 1995; Saye, 1998). In een klassikale discussie geeft de docent leiding aan de dialoog. De docent kan door het stellen van vragen leerlingen laten nadenken over hun mening en hen tot een diepere, beter door-dachte mening brengen dan wanneer zij zonder leiding van de docent discussiëren.

1.5 Vraagstellingen

Het in dit artikel gepresenteerde onderzoek heeft als doel het effect te onderzoeken van dialogisch waardevormend onderwijs op de meningsvorming van leerlingen en daarin het effect van de mate van groepswork. Omdat het hier gaat om waardevorming geïntegreerd in het vak geschiedenis, waren we daarnaast geïnteresseerd in het effect van beide lessenseries op historisch redeneren. In hoeverre gaat waardevormend geschiedenis-onderwijs ten koste van het vak? Het onderzoek richtte zich op de volgende drie vragen:

1. Levert aandacht voor dialogisch waardevormend geschiedenisonderwijs een bijdrage aan de kwaliteit van de meningsvorming van leerlingen (gebruik van waarden en multiperspectiviteit)?
2. Draagt de mate van groepswork bij dialogisch waardevormend geschiedenisonderwijs bij aan het effect op de kwaliteit van de meningsvorming van leerlingen (gebruik van waarden en multiperspectiviteit)?
3. Gaat aandacht voor dialogisch waardevormend geschiedenisonderwijs ten koste van het vermogen van leerlingen tot historisch redeneren, en is hierop het didactisch arrangement van invloed?

2 Methode

2.1 Ontwerp lessenserie

Met een vakdidacticus geschiedenis ontwierpen we twee lessenseries van dertien lessen van 45 minuten voor het vak geschiedenis in het tweede leerjaar van het vwo of havo/vwo (exclusief metingen). Beide lessenseries werden getest in een pilotstudy waaraan twee ervaren docenten deelnamen met ieder twee klassen. De lessen werden geobserveerd en de docenten gaven feedback na elke les. Daarnaast werden leerlingen geïnterviewd. Deze ervaringen leidden tot verdere ontwikkelingen (zie voor een uitgebreide beschrijving van de beide lessenseries Schuitema, Van Riessen, Veugelers, & Ten Dam, ingediend).

Voor beide lessenseries werd aanvullend materiaal ontwikkeld bij een bestaande en veelgebruikte geschiedenismethode (MeMo) die weinig aandacht aan meningsvorming besteedt. Het aanvullende lesmateriaal bestond uit een docentenhandleiding en werkboekjes voor de leerlingen. Centraal stond de geschiedenis van de Verenigde Staten van Amerika van de eerste kolonisten tot aan het begin van de negentiende eeuw. Aan de orde kwamen het ontstaan van de VS, de situatie van de indianen, de immigratie naar de VS, de slavernij en de burgeroorlog. Om de betekenisvolheid van de leerstof voor leerlingen te vergroten, werd er geappelleerd aan hun belevingswereld. Dit was vooral het geval in de laatste lessen, met leerstofthema's als immigratie en de hedendaagse Nederlandse multiculturele samenleving.

Begrip van waarden

In beide lessenseries werd systematisch aandacht besteed aan morele waarden, in de lessenserie omschreven als: "opvattingen, wensen of idealen over hoe mensen met elkaar om zouden moeten gaan". Leerlingen leerden waardegeladen meningen en meningen die niet gebaseerd zijn op een morele waarde van elkaar te onderscheiden. Dit gebeurde onder andere met behulp van een kaartjesopdracht. Leerlingen kregen een set kaartjes met daarop verschillende uitspraken. Vervolgens moesten zij de kaartjes verdelen in meningen waarin een morele waarde wordt uitgedrukt

en in meningen waarbij dat niet zo is. Regelijk bijvoorbeeld “Iedereen in Nederland zou dezelfde rechten en plichten moeten hebben” (op waarde gebaseerde uitspraak) met “Geschiedenis is één van de moeilijkste vakken op school” (uitspraak zonder morele waarde). Vervolgens leerden zij waarden te herkennen in de leerstof. Leerlingen bestudeerden bijvoorbeeld de tekst van de Amerikaanse onafhankelijkheidsverklaring uit 1776 en stukken uit de grondwet van 1788 en moesten aangeven welke waarden de stukken bevatten. Daarnaast onderzochten ze welke waarden ze zelf en andere leerlingen belangrijk vonden.

Multiperspectiviteit

De lessen waren gericht op de ontwikkeling van vaardigheden en attitudes die leerlingen nodig hebben om verschillende perspectieven op morele kwesties in de leerstof en de daarmee verbonden waarden te onderzoeken. Leerlingen kregen bijvoorbeeld (primaire) bronnen, die verschillende perspectieven op bepaalde gebeurtenissen weergaven. Door leerlingen hiermee te confronteren konden zij ervaren hoe het perspectief de waarneming beïnvloedt. Daarnaast werd er gewerkt met ‘perspectieven wisselen’. Zo werd er bij het onderdeel van de lessenserie dat gaat over de situatie van de indianen gewerkt met twee bronnensets. Een deel van de leerlingen bestudeerden bronnen die het perspectief van de indianen weergaven, terwijl andere leerlingen een gebeurtenis benaderden met behulp van bronnen die het perspectief van de kolonisten weergaven. Vervolgens gingen leerlingen met elkaar in dialoog over een aantal stellingen. Leerlingen werd gevraagd te bedenken wat een indiaan en een kolonist van zo’n stelling gedacht zouden hebben.

Vaardigheden en attitudes voor dialoog

De vaardigheden en attitudes die leerlingen nodig hebben voor het voeren van een dialoog (uitwisselen, co-construeren en verantwoorden) kregen systematisch aandacht. Leerlingen werden vanaf de eerste lessen gestimuleerd hun mening uit te wisselen met anderen, bijvoorbeeld via oefeningen waar zij elkaars mening opschreven zonder dat er direct overeenstemming hoefde te zijn.

Gaandeweg kwamen daar activiteiten bij gericht op co-construeren en verantwoorden. Bij co-construeren kwamen leerlingen zo mogelijk tot een gezamenlijk standpunt en gaven zij aan waarover ze het eens waren geworden en waarover niet. Vervolgens stelden zij vast wat belangrijke waarden waren die een rol speelden bij het bepalen van hun standpunt (verantwoorden).

Verschillen tussen de twee lessenseries

In de ene lessenserie vond de dialoog vaker plaats in kleine groepjes tussen leerlingen onderling, in de andere, rivaliserende, lessenserie had de dialoog vaker de vorm van een klassikale discussie onder leiding van de docent. In de handleiding voor docenten was per onderdeel aangegeven hoeveel tijd er besteed moest worden aan de betreffende werkvorm. In de lessenserie gericht op groepswork was in totaal 65% van de tijd ingepland voor het werken in groepjes (drie tot vijf leerlingen of duo’s). 15% was in deze lessenserie gepland voor klassikale dialoog. In de lessenserie gericht op klassikaal onderwijs was twintig 20% van de tijd gepland voor groepswork en 40% voor klassikale dialoog.

In de lessenserie gericht op groepswork werden docenten geïnstrueerd zo min mogelijk inhoudelijke hulp te geven en vooral te sturen op het proces van samenwerken. Als leerlingen er zelf niet uitkwamen, moest de docent hulp geven aan de hele groep. In de lessenserie gericht op klassikaal onderwijs werden docenten geïnstrueerd om een klasgesprek te leiden via het stellen van vragen. Daarbij werd hun gevraagd te stimuleren dat zoveel mogelijk leerlingen deelnemen aan de dialoog (uitwisselen), op elkaar reageren (co-constructie) en hun mening onderbouwen (verantwoording).

2.2 Procedure

Het onderzoek is opgezet als een quasi-experimenteel design, waarin twee experimentele condities en een controleconditie met elkaar worden vergeleken. In de eerste experimentele conditie werkten docenten met het lesmateriaal met het accent op het werken in kleine groepen (deze conditie wordt verder aangeduid met de groepsworkconditie), in de tweede experimentele conditie werkten

docenten met het lesmateriaal met het accent op klassikale dialoog. Daarnaast was er een controleconditie (alleen voor- en nametingen). In minimaal tien lessen (gemiddeld 10,6) behandelden de docenten dezelfde leerstof als de docenten in de experimentele condities. Zij werkten met de methode MeMo zoals ze dit normaliter doen (dus zonder systematische aandacht voor meningsvorming).

Voor deelname aan het onderzoek werden geschiedenisdocenten op scholen in de Randstad benaderd met tweede klassen vwo of havo/vwo, die werkten met het geschiedenis-methode MeMo. Negen docenten van acht scholen (in totaal veertien klassen) waren bereid met het aanvullende lesmateriaal te werken. De docenten werden aselekt toegewezen aan één van beide experimentele condities. Vijf docenten van drie scholen – vijf klassen – deden mee in de controleconditie. De deelnemende klassen zijn gecombineerde havo/vwo-klassen (verlengde brugklas), atheneumklassen en gymnasiumklassen. Tabel 1 geeft een overzicht van de verdeling van de verschillende schooltypen over de condities. In de tabel is te zien dat de schooltypen niet gelijkmatig verdeeld zijn over de condities. Door uitval is er geen gymnasium-klas in de groepswerkconditie.

In totaal participeerden 482 leerlingen in het onderzoek. 49% van leerlingen was vrouwelijk, met een aselekt verdeling over de condities ($\chi^2 = 0,876$, $\nu = 2$, $p = ,65$). 15% van de leerlingen definieerde de eigen etnici-

teit als niet-Nederlands of gemengd, met een aselekt verdeling over de condities ($\chi^2 = 4,95$, $\nu = 2$, $p = ,08$). Als maat voor de sociaal economische status (SES) werd de hoogste opleiding van de ouders genomen. 21% van de leerlingen had een lage SES (hoogste opleiding is vmbo). 7% had een gemiddelde SES (mbo) en 72% een hoge SES (havo/vwo of meer). Deze groepen waren aselekt verdeeld over de condities ($\chi^2 = 5,263$, $\nu = 4$, $p = ,26$).

Voorafgaand aan de lessen werd er in alle condities één les besteed aan voormetingen die bestonden uit vragenlijsten. De nameting nam eveneens één lesuur in beslag. De leerlingen werkten aan een essayopdracht, waarbij zij hun mening moesten opschrijven aan de hand van een stelling. De klas werd hierbij aselekt in tweeën gedeeld. De ene helft leerlingen kreeg een leerstofgebonden onderwerp, de andere helft een onderwerp dat niet in de lessenserie was behandeld (transferopdracht, zie verder paragraaf 2.3). Aan de voormetingen deed 7% van de leerlingen niet mee, aselekt verdeeld over de condities ($\chi^2 = 0,163$, $\nu = 2$, $p = ,92$). Bij de eindmeting deed 9% niet mee. Deze uitval was niet gelijkmatig over de condities verdeeld ($\chi^2 = 10,258$, $\nu = 2$, $p = ,01$). In de groepswerkconditie vielen meer leerlingen uit (14%) dan in de klassikale conditie (7%) en de controleconditie (4%). In Tabel 2 is een overzicht weergegeven van het aantal leerlingen dat aan de voor- en nametingen meedeed.

Tabel 1

Aantal klassen per conditie en schooltype

	Gymnasium	Atheneum	Havo/vwo	Totaal
Groepswerk	0	2	6	8
Klassikaal	2	3	1	6
Controle	2	2	1	5
Totaal	4	7	8	19

Tabel 2

Aantal deelnemende leerlingen aan de voor- en nametingen per conditie

	Voormetingen		Nametingen	
		Leerstofgebonden	Transfer	
Groepswerk	214	85	112	
Klassikaal	131	57	73	
Controle	108	45	65	
Totaal	453	187	250	

Training docenten

De docenten in de experimentele condities kregen vooraf een training. Alle docenten namen hieraan deel. De training bestond uit een toelichting op het werken met de docentenhandleiding en conditiespecifieke instructies over de rol van de docent.

Implementatie

Om na te gaan hoe de beide lessenseries zijn geïmplementeerd, werd aan de docenten gevraagd een logboek bij te houden. Docenten hebben daarin per les per onderdeel aangegeven of dat onderdeel was uitgevoerd zoals omschreven in de docentenhandleiding en hoeveel tijd er aan besteed was. Met behulp van deze logboeken werd berekend hoeveel procent van de tijd de leerlingen in groepjes hadden gewerkt en hoeveel tijd er aan klassikale dialoog werd besteed. In de groepswerkconditie was het gemiddelde percentage groepswerk 59 (variërend van 55 tot 64) en in de klassikale conditie 24 (variërend van 19 tot 31) ($t = 11,471, p < ,001$). Vervolgens werden beide condities vergeleken op de hoeveelheid klassikale dialoog. In de groepswerkconditie gaven docenten aan gemiddeld 15% (variërend van 11 tot 18%) van de tijd te hebben besteed aan klassikale dialoog. In de klassikale conditie was dit 31% (variërend van 29 tot 34) ($t = 9,492, p < ,001$). Daarnaast werd gekeken hoeveel onderdelen uit de lessenserie waren uitgevoerd zoals dat in de docentenhandleiding was beschreven. In de klassikale conditie varieerden die percentages van 77 tot 90% met een gemiddelde van 79%. In de groepswerkconditie lagen de percentages iets hoger. Het gemiddelde was hier 87% en de percentages varieerde van 77 tot 92%. Een t -toets liet geen significante verschillen zien tussen beide condities. Op basis hiervan concluderen we dat de lessenseries voldoende waren geïmplementeerd en dat beide condities duidelijk verschilden in de hoeveelheid groepswerk en klassikale dialoog.

2.3 Meetinstrumenten

Afhankelijke variabele: essayopdrachten

Het effect van de beide lessenseries op de kwaliteit van de meningen van leerlingen is onderzocht via twee verschillende essay-

opdrachten. Beide opdrachten bestonden uit een korte introductie van een moreel dilemma en een stelling. In de eerste opdracht (leerstofgebonden opdracht) stond een onderwerp centraal dat in de lessenserie was behandeld. Daarin moesten leerlingen de belangen afwegen van de inheemse bevolking van Amerika tegen die van immigranten uit Europa en andere continenten. De stelling was: "Om de manier van leven van de indianen te beschermen, had de Amerikaanse regering veel eerder de komst van nieuwe immigranten moeten tegenhouden". De leerlingen hadden zich in de lessen uitgebreid kunnen verdiepen in de materie en konden hierbij de kennis opgedaan tijdens de lessen gebruiken. Een tweede essayopdracht beoogde vast te stellen of leerlingen in staat waren het geleerde toe te passen bij een voor hen nieuw onderwerp. Die essayopdracht (transferopdracht) ging daarom over een onderwerp dat niet in de lessen was behandeld. De stelling was: "Schooluniformen terug in de klas!". Voorgesteld werd om schooluniformen in te voeren om te voorkomen dat leerlingen gepest kunnen worden wegens hun kleding.

Binnen elke klas werden leerlingen aselekt aan een van beide essayopdrachten toegewezen. De ene helft leerlingen kreeg de leerstofgebonden essayopdracht, de andere helft de transferopdracht. Leerlingen kregen tien minuten de tijd om in groepjes van vier leerlingen de stelling te bespreken. Daarna schreven zij individueel hun mening op aan de hand van de stelling. Aan de leerlingen is vooraf gemeld dat niet de mening op zich, maar de onderbouwing ervan zou worden beoordeeld voor een cijfer dat zou meetellen voor het rapport.

Een team van onafhankelijk werkende beoordelaars beoordeelden de essays van leerlingen op *morele waarden* en *multiperspectiviteit*. Om na te gaan wat de invloed is van aandacht voor meningsvorming op historische kennis en inzicht werd het leerstofgebonden essay daarnaast ook beoordeeld op *historisch redeneren*.

Bij het eerste aspect, *morele waarden*, is het aantal argumenten waarin gerefereerd wordt aan een morele waarde gescoord, alsmede de mate waarin aan een morele waarde

expliciet gerefereerd werd: hoe duidelijker een leerling refereert aan een algemeen geldende waarde die de specifieke context overstijgt, hoe hoger de score. Een essay over schooluniformen waarin een leerling bijvoorbeeld aangeeft dat iedereen zelf moet kunnen kiezen wat hij of zij aantrekt, wordt hoger gescoord dan een essay waarin een leerling zegt dat zij zelf wil kunnen kiezen. Een nog hogere score krijgt een essay waarin een leerling aanvoert dat het kunnen kiezen van je kleding past bij de vrijheid van meningsuiting.

Met het aspect *multiperspectiviteit* beoogden we na te gaan in hoeverre leerlingen bij de onderbouwing van een mening aandacht besteden aan verschillende perspectieven. Verplaatsen zij zich in andere zienswijzen? In de beoordeling werd behalve op het aantal perspectieven, gelet op de uitwerking van de perspectieven. Er werd gekeken naar het aantal (sub)argumenten voor beide perspectieven. Vaak vielen argumenten vóór en tegen de stelling samen met twee perspectieven. Dit was echter niet altijd het geval. Wanneer een leerling aangeeft dat het nadeel van een schooluniform voor haar is dat ze dan niet haar nieuwe kleren op school kan tonen en als voordeel dat ze dan niet meer 's ochtends hoeft te bedenken wat ze aan moet, is dat allebei vanuit hetzelfde perspectief geredeneerd.

Het aspect *historisch redeneren* is beoordeeld door te bepalen in welke mate leerlingen rekening houden met de context van die tijd (zie bijvoorbeeld Lee & Ashby, 2001). Maken leerlingen in het leerstofgebonden essay gebruik van de historische kennis uit de leerstof? Blijken leerlingen zich bijvoorbeeld bewust van de historiciteit van waarden? En gebruiken leerlingen de historische kennis ook bij de argumentatie en bij het trekken van hun conclusie?

Elk essay is voor ieder aspect apart in zijn geheel beoordeeld volgens een methode die afgeleid is van de *comparison method* (Blok, 1986). Er werd gewerkt met *ankers*. Dit zijn voorbeeldessays waarmee de beoordelaars de andere essays vergelijken. We hebben (per aspect) drie typische en representatieve voorbeeldessays gekozen van een goed, een gemiddeld en een zwak essay. De drie ankers hebben een vastgesteld aantal punten, respectievelijk 150, 100 en 50. Aan de beoordelaars

is vervolgens gevraagd elk essay een score te geven tussen de 0 en 200, in vergelijking met de ankers. Vindt een beoordelaar een essay bijvoorbeeld beter dan het gemiddelde voorbeeldessay, maar minder goed dan het goede voorbeeld, dan krijgt dit essay een score tussen de 100 en 150 punten. Deze methode beoogt een normaalverdeelde interval schaal te creëren met als gemiddelde 100 en een standaardafwijking van 50.

Om een schatting te kunnen geven van de betrouwbaarheid van de scores op de essays is gewerkt volgens de *slangmethode* (Van den Bergh & Eiting, 1989). De essays werden aselekt verdeeld in evenveel steekproeven als het aantal beoordelaars. Elke beoordelaar beoordeelde twee steekproeven. Beoordelaar 1 beoordeelde steekproef 1 en 2, beoordelaar 2 beoordeelde steekproef 2 en 3, en zo verder. De laatste beoordelaar gaf oordelen over de essays in de laatste en de eerste steekproef. Op deze manier werd elk essay door twee beoordelaars beoordeeld en was het mogelijk om voor elke beoordelaar een betrouwbaarheid te schatten gebaseerd op alle andere beoordelaars.

Per aspect werden verschillende beoordelarteam geformeerd om te voorkomen dat de beoordelaars zich zouden laten beïnvloeden door criteria van een ander aspect. De teams bestonden uit vijf (voor waarden en voor multiperspectiviteit) en uit drie beoordelaars (drie historici voor historisch redeneren).

De beoordelaars werden getraind in trainingssessies van twee tot drie uur. Tijdens deze training werden het betreffende aspect en de bijbehorende criteria toegelicht. Verder kregen de beoordelaars een overzicht van het soort redeneringen en uitspraken dat ze konden verwachten met voorbeelden van goede en minder goede fragmenten. Eerst werd de transferopdracht getraind en beoordeeld. Daarna volgde een nieuwe training, nu met voorbeelden van de leerstofgebonden opdracht.

De betrouwbaarheden van het beoordelen van de essays zijn geschat met behulp van een LISREL-model (Jöreskog & Sörbom, 2002). Door de slangmethode konden we de betrouwbaarheid van elke beoordelaar schatten op basis van alle andere beoordelaars. De

scores op de essays zijn berekend door steeds het gemiddelde te nemen van de twee beoordelaars die het betreffende essay hadden beoordeeld. De betrouwbaarheid van de gemiddelde scores op de essays kan voor elke steekproef afzonderlijk worden berekend met de Spearman-Brown formule voor testverlenging. Deze varieerde van ,76 tot ,93 (zie bijlage 1). Dat betekent dat de beoordelingen voldoende betrouwbaar zijn.

Controlevariabelen

Er werd een aantal voormetingen verricht voor variabelen waarvan een correlatie verwacht werd met de eindmetingen. Deze voormetingen kunnen worden gebruikt om te controleren voor eventuele relevante verschillen tussen condities aan het begin van de lessenserie.

- Redeneervaardigheden. Er mag verwacht worden dat redeneervaardigheden een belangrijke rol spelen bij het kunnen formuleren van een mening. Om te controleren voor het niveau van argumenteren van leerlingen zijn drie relevante schalen (21 items) van de *Alvavavotoets* (algemene vaardigheden in de basisvorming) gebruikt (Meijer, Elshout-Mohr, & Van Hout-Wolters, 2001): (1) meningsvorming; (2) opvattingen en overtuigingen (3) feiten en meningen. Cronbach's alpha over de drie schalen was ,63.
- Citoscores. De citoscores van de leerlingen zijn gebruikt om bij de essayopdrachten en de kennistoets te kunnen controleren voor het algemene prestatieniveau van leerlingen.
- Attitudes ten opzichte van dialoog. In de lessenserie is dialoog een belangrijke didactische strategie om de meningsvorming van leerlingen te stimuleren. De attitudes van leerlingen t.o.v. dialogisch leren bij aanvang van de lessenserie kunnen van invloed zijn op de afhankelijke variabelen. Deze werden gemeten met behulp van de 'Attitudes to dialogic learning scale' (Frijters et al., in druk). Deze schaal meet de attitudes ten aanzien van (1) uitwisselen van meningen, (2) co-construeren van meningen, en (3) het verantwoord worden van meningen. Cronbach's alpha over de drie schalen was ,82.

- Sociale betrokkenheid. Het vormen van een mening waarbij rekening wordt gehouden met anderen, kan worden beïnvloed door de mate van betrokkenheid bij anderen. Om te kunnen controleren voor verschillen in sociale betrokkenheid is gebruik gemaakt van een schaal van het Child Development Project ('Concern for others scale') (Solomon et al., 2001). Cronbach's alpha was ,73.
- Schoolcultuur. De manier waarop leerlingen en docenten in school met elkaar omgaan, kan eveneens van invloed zijn op de meningsvorming van leerlingen. De schoolcultuur is gemeten met de 'School culture scale' (Higgins-D'Allesandro & Sath, 1997) waarin wordt gevraagd of leerlingen onderling, en docenten en leerlingen, op een respectvolle manier met elkaar omgaan. Cronbach's alpha was ,80.
- Klassenklimaat. Een laatste factor die van invloed kan zijn op de meningsvorming van leerlingen, is het klimaat in de klas als het gaat om het bespreken van sociale en politieke kwesties. Om het klimaat te meten werd de 'IEA Measure of classroom climate for discussion' (Torney-Purta, Lehmann, Oswald & Schulz, 2001) gebruikt. Deze schaal meet of leerlingen zich vrij voelen om hun mening te uiten in de klas en of de docent over het algemeen leerlingen aanmoedigt hun mening te uiten. Cronbach's alpha was ,69.

2.4 Analyses

De essayopdrachten werden in groepjes voorbereid. Leerlingen discussieerden eerst met hun groepje over de stelling en wisselden zo argumenten uit. De groepjes maakten deel uit van een klas. Vanwege deze hiërarchische geneste structuur van de data werden de scores op de essays geanalyseerd met behulp van multilevelregressieanalyse. Hiervoor werd *MLwiN 2,02* (Rasbash, Browne, Cameron & Charlton, 2005) gebruikt. Er werden modellen vergeleken met drie niveaus (leerling, groep en klas). Het opnemen van een extra niveau met docent of school, leidde niet tot het beter passen van het model. In de verdere analyse werden deze niveaus dan ook buiten beschouwing gelaten.

Het beoordelen van de essays had geresul-

teerd in vijf verschillende scores: *waarden*, *multiperspectiviteit* en *historisch redeneren* bij de leerstofgebonden opdracht, en *waarden* en *multiperspectiviteit* bij de transferopdracht. Het effect van conditie werd onderzocht in een multivariate multilevelanalyse met deze vijf scores als afhankelijke variabelen. Op klassenniveau en groepsniveau werden de varianties voor de verschillende afhankelijke variabelen samengenomen.

Omdat de gemiddelde scores op de essays per schooltype verschilden en de schooltypes ongelijk over de condities waren verdeeld, hebben we het effect van conditie onderzocht binnen de schooltypen. In het eerste model is conditie samen met schooltype opgenomen. Om te kunnen controleren voor individuele verschillen in sekse, SES en etnische identiteit, zijn via een stapsgewijze procedure deze persoonsvariabelen toegevoegd, waarbij alleen significante variabelen in het model werden opgenomen (model 2). Voor model 3 werd op dezelfde manier gekeken of de controlevariabelen konden bijdragen; ook hier werden alleen significante variabelen in het model opgenomen.

Elke variabele werd in eerste instantie in het model gebracht met voor elke afhankelijke variabele aparte regressiegewichten. Vervolgens werd gekeken of deze regressiegewichten voor elke afhankelijke variabele significant van elkaar verschilden. Wanneer dit niet zo is, mag worden aangenomen dat het effect van de betreffende variabele hetzelfde is voor elke afhankelijke variabele. Om het model zo zuinig mogelijk te houden, werden in dat geval de regressiegewichten vervangen door één regressiegewicht voor alle afhankelijke variabelen tezamen.

3 Resultaten

In deze paragraaf worden de resultaten van de multilevelanalyses gepresenteerd: in hoeverre bevordert dialogisch waardevormend geschiedenisonderwijs de kwaliteit van de meningsvorming van leerlingen, en in hoeverre is de mate van groepswerk hierop van invloed? De kwaliteit van meningsvorming wordt in dit onderzoek geoperationaliseerd als het gebruik van waarden en multiperspec-

tiviteit door leerlingen. Tevens wordt in deze paragraaf de onderzoeksvraag beantwoord in hoeverre dialogisch waardevormend geschiedenisonderwijs van invloed is op het historisch kunnen redeneren van leerlingen.

In Tabel 3 zijn de resultaten voor de essayscores weergegeven. In model 1 zijn alleen conditie en schooltype opgenomen. Het gemiddelde van de havo/vwo-leerlingen in de controleconditie is het intercept en groepswerkconditie en klassikale conditie zijn als afwijking hiervan weergegeven. De gemiddelde score op het aspect *waarden* bij de leerstofgebonden opdracht is voor havo/vwo-leerlingen in dit model geschat op 54,2. De leerlingen in de groepswerkconditie scoorden bij deze opdracht hoger op het refereren aan waarden dan de leerlingen in de controleconditie (het verschil is 23,4 punten). De leerlingen in de klassikale conditie refereerden in hun essays minder aan waarden dan de leerlingen in de controleconditie (-9,8), maar dit verschil is niet significant. Het verschil tussen de groepswerkconditie en de klassikale conditie kan met de gegevens uit de tabel berekend worden. Het verschil is 33,2 en significant ($\chi^2 = 9,2$, $v = 1$, $p = ,002$).

Tabel 3 laat verder zien dat de havo/vwo-leerlingen in de groepswerkconditie meer verschillende perspectieven in hun essay betrokken dan de leerlingen in de controleconditie. Deze leerlingen hadden zowel bij de leerstofgebonden opdracht als bij de transferopdracht een significant hogere score op *multiperspectiviteit*. De leerlingen in de groepswerkconditie scoorden hoger dan die in de klassikale conditie. Bij de leerstofgebonden opdracht is het verschil op multiperspectiviteit 39,88 ($\chi^2 = 12,79$, $v = 1$, $p < ,001$). Het verschil op multiperspectiviteit bij de transferopdracht is 33,13 ($\chi^2 = 8,78$, $v = 1$, $p = ,002$). De gemiddelde scores op waarden bij de transferopdracht verschilden niet significant tussen de drie condities.

Wat *historisch redeneren* betreft, is in Tabel 3 te zien dat de leerlingen in de groepswerkconditie het even goed deden als de leerlingen in de controleconditie. De leerlingen in de klassikale conditie scoorden evenwel significant lager dan de leerlingen in de controleconditie (-25,1). Het verschil tussen de groepswerkconditie en de klassikale conditie

moet ook hier weer apart berekend worden. Dit verschil is 31,32 en significant ($\chi^2 = 8,20$, $\nu = 1$, $p = ,004$). De verschillen tussen de groepswerkconditie en de controleconditie zijn niet significant.

Het effect van schooltype bleek niet te verschillen over de verschillende afhankelijke variabelen. Voor de verschillende schooltypen werd daarom één regressiegewicht in het model opgenomen voor alle afhankelijke

variabelen. Voor atheneum is alleen een hoofdeffect opgenomen in het model, waardoor het beeld voor de atheneumleerlingen hetzelfde als voor de havo/vwo-leerlingen. Alle gemiddeldes voor atheneum zijn in dit model 25,8 punten hoger dan voor havo/vwo.

Aangezien er in de groepswerkconditie geen gymnasiumklas was, kun je niet spreken van een hoofdeffect voor gymnasium. Het effect van gymnasium is daarom ge-

Tabel 3
Resultaten van de multilevelanalyse op de essayscores

			Model 1		Model 2		Model 3	
			Gem.	SE	Gem.	SE	Gem.	SE
Leerstofgebonden								
Waarden	havo/vwo	controle	54,2	(10,9)	52,5	(10,7)	52,3	(10,6)
		groepswerk ¹	23,4	(11,6)	23,7	(11,2)	24,2	(11,1)
		klassikaal ¹	-9,8	(12,5)	-9,5	(12,1)	-9,0	(11,9)
		sekse (meisjes) ²			3,8	(4,7)	3,7	(4,6)
		dialogo					7,6 (5,2)	
Multipers.	havo/vwo	controle	56,4	(11,0)	50,9	(10,9)	50,7	(10,8)
		groepswerk ¹	25,6	(11,9)	26,1	(11,5)	26,8	(11,3)
		klassikaal ¹	-14,3	(12,7)	-13,9	(12,3)	-13,4	(12,1)
		sekse (meisjes) ²			11,3	(5,0)	11,2	(5,0)
		dialogo					10,6 (5,6)	
Historisch redeneren	havo/vwo	controle	61,9	(10,9)	62,1	(10,7)	62,2	(10,6)
		groepswerk ¹	6,2	(11,6)	6,4	(11,2)	6,3	(11,1)
		klassikaal ¹	-25,1	(12,5)	-24,9	(12,1)	-24,4	(11,9)
		sekse (meisjes) ²			-0,1	(4,7)	-0,2	(4,7)
		dialogo					-1,0 (5,2)	
Transfer								
Waarden	havo/vwo	controle	70,9	(10,9)	66,5	(10,7)	67,1	(10,5)
		groepswerk ¹	15,6	(11,6)	15,3	(11,2)	16,2	(11,0)
		klassikaal ¹	2,8	(12,5)	2,2	(12,1)	2,0	(11,9)
		sekse (meisjes) ²			10,2	(4,6)	8,6	(4,6)
		dialogo					14,2 (5,8)	
Multipers.	havo/vwo	controle	58,2	(10,6)	51,2	(10,3)	51,2	(10,2)
		groepswerk ¹	40,1	(11,2)	39,5	(10,8)	39,3	(10,6)
		klassikaal ¹	6,9	(12,1)	5,8	(11,6)	6,3	(11,5)
		sekse (meisjes) ²			16,2	(4,0)	11,2	(4,0)
		dialogo					-0,53 (4,9)	
		atheneum ³	25,8	(8,4)	24,7	(8,1)	24,7	(8,0)
		gym controle ³	21,35	(14,5)	22,1	(13,9)	22,0	(13,7)
		gym klassikaal ³	68,6	(14,1)	68,7	(13,6)	67,2	(13,4)
Variantie								
		level 3 (klas)	165,6	(69,1)	150,0	(63,6)	143,7	(61,8)
		level 2 (groep)	143,5	(42,3)	139,8	(41,1)	144,4	(41,4)
		level 1 (leerling)						
	waarden	leerstof	665,4	(79,6)	663,9	(79,3)	655,2	(78,4)
	multipers.	geb.	817,5	(96,1)	800,1	(94,1)	784,4	(92,4)
	hist. red.		663,5	(79,5)	662,4	(79,3)	656,7	(78,7)
	waarden	transfer	1068,1	(107,2)	1045,1	(104,7)	1017,9	(102,2)
	multipers.		768,2	(79,4)	720,5	(74,6)	717,4	(74,3)
Fit			9494,0		9471,3		9459,2	
		verbetering	59,6		22,7		12,1	
		verschil in vrijheidsgraden	14		5		5	
		p-waarde	$p < ,001$		$p < ,001$		$p = ,034$	

Vet gedrukt is significant ($p < 0,05$)

¹ Verschil t.o.v. controlegroep

² Verschil tussen jongens (= 0) en meisjes (= 1)

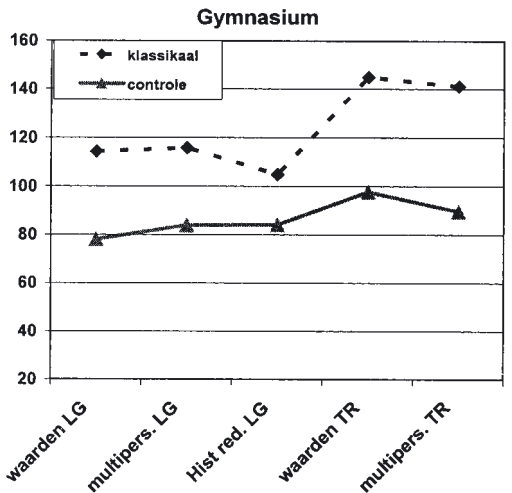
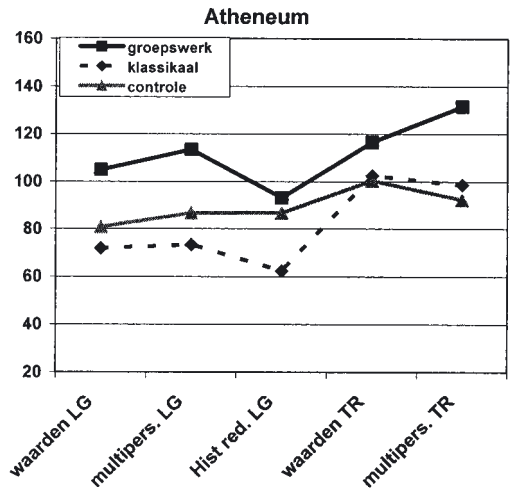
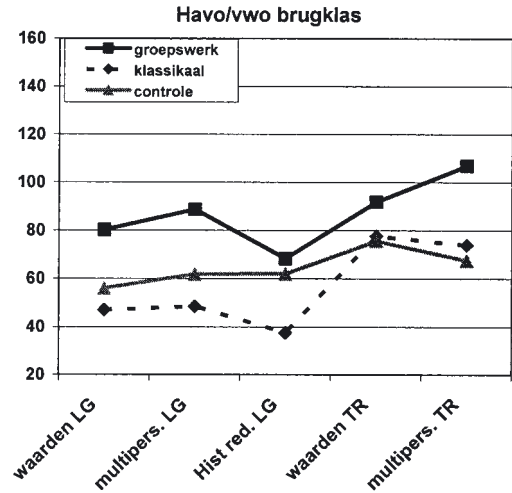
³ Verschil t.o.v. havo/vwo

splitst en er zijn aparte regressiegewichten toegevoegd voor de leerlingen in de klassikale conditie en de controleconditie. Uit Tabel 3 is af te lezen dat gymnasiumleerlingen in de klassikale conditie op alle afhankelijke variabelen significant hoger scoorden (68,6) dan de havo/vwo-leerlingen in dezelfde conditie. In de controleconditie daarentegen scoorden gymnasium leerlingen niet hoger dan de havo/vwo-leerlingen. Toevoeging van termen voor interactie van de andere schooltypen (atheneum en havo/vwo) met conditie leidde niet tot het beter passen van het model.

In het tweede model werd sekse toegevoegd. De regressiegewichten voor sekse verschilden significant per afhankelijke variabele en konden niet worden samengenomen. Etnische identiteit en SES bleken niet significant en zijn uit het model verwijderd. In tabel 3 is te zien dat meisjes significant hoger scoorden op het betrekken van verschillende perspectieven bij de onderbouwing van hun mening bij beide essayopdrachten. Daarnaast refereerden meisjes bij de transferopdracht meer aan waarden dan jongens. De effecten van conditie en schooltype bleven nagenoeg hetzelfde.

In model 3 werd de voormeting van attitudes ten opzichte van dialogisch leren toegevoegd. Ook hier konden de regressiegewichten voor de verschillende afhankelijke variabelen niet worden samengenomen. De overige controlevariabelen bleken niet significant. Het effect van attitudes ten opzichte van dialogisch leren bleek alleen significant bij het aspect waarden bij de transferopdracht. Na het toevoegen van deze controlevariabelen verdween het significante effect van sekse op waarden bij de transferopdracht. Wat de overige effecten betreft veranderde het toevoegen van deze variabele weinig.

In figuur 1 zijn de gemiddelden van de verschillende condities op de vijf afhankelijke variabelen per schooltype grafisch weergegeven zoals die geschat zijn in model 3. Alleen de gemiddelden voor meisjes zijn gegeven. Voor jongens is het patroon hetzelfde, maar dan liggen de beide scores op het aspect multiperspectiviteit 11,2 punten lager. In de figuur is te zien dat het patroon bij havo/vwo hetzelfde is als bij atheneum, met het verschil dat de scores bij atheneum hoger liggen. De



Figuur 1. Geschatte gemiddelden van de scores op de essays voor meisjes in havo/vwo, atheneum en gymnasium.

leerlingen in de groepswerkconditie scoorden op alle variabelen hoger dan de leerlingen in de klassikale conditie en de controleconditie, maar alleen de verschillen op het aspect waarden bij de leerstofgebonden opdracht en op het aspect multiperspectiviteit bij zowel de leerstofgebonden opdracht als de transferopdracht waren significant. De leerlingen in de klassikale conditie scoorden hetzelfde als de leerlingen in de controleconditie, behalve op historisch redeneren. Hier scoorden de leerlingen in de klassikale conditie lager.

Bij gymnasium zien we een ander patroon. Hier is zoals gezegd geen groepswerkconditie aanwezig. Verder zijn de gemiddelde scores van de gymnasiasten in de klassikale conditie op alle afhankelijke variabelen hoger dan die van de leerlingen in de controleconditie. Ook het kleinste verschil tussen deze twee groepen, op historisch redeneren, is significant ($\chi^2 = 5,15$, $\nu = 1$, $p = 0,023$).

4 Discussie

In dit onderzoek zijn we nagegaan wat het effect is van dialogisch waardevormend geschiedenisonderwijs op de meningsvorming van leerlingen. Daarnaast hebben we meer specifiek gekeken naar de invloed van de mate van groepswerk. Ten behoeve van het onderzoek hebben we twee lesmethoden ontwikkeld. Beide lesmethodes waren er op gericht leerlingen te stimuleren een mening te vormen en daarbij rekening te houden met conflicterende waarden en verschillende perspectieven. De lessenseries verschilden in de mate van groepswerk en klassikaal onderwijs.

De resultaten laten allereerst zien dat het loont om systematisch aandacht te besteden aan meningsvorming in het vakonderwijs. Het geschiedenisonderwijs biedt goede mogelijkheden om vorm te geven aan waardevormend onderwijs. De effectiviteit van de twee verschillende didactische benaderingen van waardevormend onderwijs, verschilt echter. Leerlingen die relatief veel in kleine groepjes werken, refereren in hun essays vaker en explicieter aan waarden en zijn beter in staat zich rekenschap te geven van verschillende

perspectieven, in vergelijking met leerlingen die meer klassikaal hebben gewerkt.

Het effect van dialogische waardevormend onderwijs verschilt tussen de schooltypes, havo/vwo-brugklas, atheneum en gymnasium. Bij de havo/vwo-brugklassen en het atheneum zien we dat de hoeveelheid groepswerk een belangrijke rol speelt bij het bevorderen van de meningsvorming. De leerlingen uit de havo/vwo- en atheneumklassen in de klassikale conditie hadden geen hogere scores dan in de controleconditie. De leerlingen in de groepswerkconditie hadden daarentegen op een aantal aspecten hogere scores dan de leerlingen in beide andere condities. Deze leerlingen scoorden hoger op multiperspectiviteit. Dit gold zowel voor de leerstofgebonden opdracht als voor de transferopdracht. Deze resultaten wijzen erop dat wanneer leerlingen relatief veel in kleine groepjes werken, zij beter in staat zijn verschillende perspectieven te betrekken bij de verantwoording van hun mening, dan leerlingen die minder vaak in groepjes werkten.

Wat het gymnasium betreft, blijkt dat de leerlingen in de klassikale conditie het op alle gemeten aspecten van de essays beter doen dan leerlingen in de controle conditie. Omdat er geen gymnasiumklas in de groepswerkconditie zat, kunnen we alleen hypothetisch wat zeggen over de invloed van de mate van groepswerk voor deze groep leerlingen. Aangezien bij de andere schooltypen de leerlingen in de groepswerkconditie op een aantal punten overtuigend hoger scoorden, veronderstellen we dat een hoge mate van groepswerk ook bij gymnasiumleerlingen niet nadelig zou zijn.

De vraag of het integreren van meningsvorming in het vakonderwijs al dan niet ten koste gaat van "het vak", kan niet in algemene zin beantwoord worden. Het hangt af van het gekozen didactische arrangement. De leerlingen in de havo/vwo- en atheneumklassen die relatief veel in kleine groepjes hebben gewerkt, bleken even hoog te scoren op historisch redeneren als de leerlingen in de controleconditie. Binnen deze schooltypen scoorden de leerlingen in de klassikale conditie echter lager op historisch redeneren dan de leerlingen in de groepswerkconditie en de controle conditie.

Wat het aspect *waarden* betreft moet er voor de leerlingen in de havo/vwo- en atheneumklassen een onderscheid gemaakt worden tussen de leerstofgebonden opdracht en de transferopdracht. Alleen bij de leerstofgebonden opdracht scoorden de leerlingen in de groepswerkconditie hoger dan de leerlingen in de controle en klassikale conditie.

Dat de effecten van dialogisch waardevormend geschiedenisonderwijs op de kwaliteit van de meningsvorming van leerlingen in de havo/vwo- en atheneumklassen in de groepswerkconditie zijn gevonden, kan mede veroorzaakt zijn door de vorm van de eindopdracht. Alvorens individueel hun mening op te schrijven, kregen leerlingen immers tien minuten de tijd om in groepjes van vier leerlingen de betreffende stelling te bespreken. De leerlingen in de groepswerkconditie waren meer gewend aan het werken in groepjes dan de leerlingen in de klassikale conditie. Dit roept de vraag op in hoeverre de gevonden resultaten toe te schrijven zijn aan de in de lessenserie opgedane ervaring met samenwerken in plaats van aan de specifieke didactische wijze waarop aandacht is besteed aan de vaardigheden en attitudes van leerlingen gericht op het genereren van argumenten. Het verdient dan ook aanbeveling in toekomstig onderzoek tevens aandacht te schenken aan de invloed van de kwaliteit van de gevoerde groepsdialogen op de kwaliteit van de meningsvorming van leerlingen.

Het uitblijven van een effect op waarden bij de transferopdracht bij de leerlingen in de havo/vwo- en atheneumklassen is een resultaat dat we vaak zien in onderwijskundig onderzoek, maar desalniettemin teleurstellend. De leerlingen die relatief veel in kleine groepjes hadden gewerkt refereerden meer aan waarden bij een voor hen bekend, dat wil zeggen in de lessen onderwezen, onderwerp. Maar geconfronteerd met een nieuw onderwerp, bleek de kwaliteit van de meningsvorming van deze groep leerlingen niet hoger te zijn dan die van de leerlingen die geen waardevormend onderwijs hadden gevolgd. Veel leerlingen zijn niet in staat een dilemma als schooluniformen uit de concrete context van hun eigen alledaagse omgeving te halen en te beschouwen in een bredere, maatschappelijke context. Hun vermogen om in hun me-

ningsvorming aan waarden te refereren, blijft met andere woorden gebonden aan het onderwezen leerdomein. Er heeft geen (verre) transfer plaatsgevonden. We leiden hieruit de aanbeveling af om in toekomstig onderzoek expliciet aandacht te besteden aan het integreren van transferbevorderende kwaliteiten van onderwijsleerprocessen (zie bijvoorbeeld Elshout-Mohr, van Hout-Wolters & Broekamp, 1999) in een op groepswerk gerichte aanpak van dialogisch waardevormend onderwijs.

Een andere verklaring voor het uitblijven van transfer van het vermogen om aan waarden te refereren, heeft te maken met morele sensitiviteit (Tirri & Pehkonen, 2002). Hebben leerlingen ingezien dat het in de transferopdracht om een moreel dilemma handelde? Terwijl in de lessenserie aandacht is besteed aan waarden via oefeningen waarin geëxpliciteerd was dat het herkennen van morele waarden het doel van de opdracht was, stond dit niet in de instructie van de (essay)eindopdrachten. Het was nadrukkelijk de bedoeling dat leerlingen uit zichzelf ook bij een voor hen nieuw onderwerp zouden gaan refereren aan waarden. Dit was wellicht voor veel leerlingen een brug te ver.

Noten

- 1 De auteurs danken Wilfried Admiraal en Huub van den Bergh voor hun steun en advies bij de opzet van de beoordelingen en de multilevelanalyses, en de reviewers voor hun kritische en opbouwende opmerkingen.

Literatuur

- Banks, J. A. (2004). *Diversity and Citizenship Education*, San Francisco: Yossey-Bass.
- Banks J.A., McGee Banks, C. A., Cortes, C. E., Hahn, L. C., Merryfield, M. M., Moodley, et al. (2005). *Democracy and diversity: Principles and concepts for educating citizens in a global age*. Center for Multicultural Education, College of Education. University of Washington, Seattle.
- Barton, K. C., & Levstik, L. S. (2004). *Teaching History for the Common Good*. Mahwah, NJ:

- Lawrence Erlbaum.
- Beane, J. (2002). Beyond self-interest: A democratic core curriculum. *Educational Leadership*, 59(7), 25-28.
- Bergh, H. van den & Eiting, M. (1989). Estimating rater reliability. *Journal of Educational Measurement*, 26, 29-40.
- Blatt, M., & Kohlberg, L. (1975). The effects of classroom moral discussions upon children's level of moral judgement. *Journal of Moral Education*, 4(2), 129-161.
- Blok, H. (1986). Essay rating by the comparison method. *Tijdschrift voor onderwijsresearch*, 11, 169-76.
- Covell, K., & Howe, R. (2001). Moral education through the 3 Rs: rights, respect and responsibility. *Journal of Moral Education*, 30(1), 29-44.
- Dam, G. ten, & Volman, M. (2004). Critical thinking as a citizenship competence: teaching strategies. *Learning and Instruction* 14(4), 359-379.
- Dam, G. ten, Wardekker, W., Miedema, S., & Veugelers, W. (2004). De kern van het onderwijs en de lerarenopleiding in theoretisch perspectief. In G. ten Dam, W. Veugelers, W. Wardekker & S. Miedema (Eds.), *Pedagogisch Opleiden. De pedagogische taak van de lerarenopleidingen* (pp. 21-50). Amsterdam: Uitgeverij SWP.
- Doyle, D. (1997). Education and character: a conservative view. *Phi Delta Kappan*, 78(6), 440-443.
- Elkind, D., & Sweet, F. (1997). The Socratic approach to character education. *Educational Leadership*, 54(8), 56-59.
- Elshout-Mohr, M., Van Hout-Wolters, B., & Broekamp, H. (1999). Mapping situations in classroom and research: Eight types of instructional-learning episodes. *Learning and Instruction*, 9, 57-75.
- Frijters, S., Dam, G. ten, & Rijlaarsdam, G. (in press). Effects of dialogic learning on value-loaded critical thinking. *Learning and Instruction*.
- Grant, R. (1996). The ethics of talk: classroom conversation and democratic politics. *Teachers College Record*, 97(3), 470-482.
- Gutmann, A. & Thompson, D. (1996). *Democracy and disagreement*. London, Harvard University Press.
- Hansen, D. (2001). Teaching as a moral activity. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 566-603). Washington, D.C.: AERA.
- Haste, H. (2004). Constructing the citizen. *Political Psychology*, 25(3), 413-439.
- Hicks, D. (2001). Re-examining the future: The challenge for citizenship education. *Educational Review*, 53(3), 229-240.
- Higgins-D'Alessandro, A., & Sadh, D. (1997). The dimensions and measurement of school culture. *International Journal of Educational Research*, 27, 553-569.
- Jöreskog, K. G., & Sörbom, D. (2002). LISREL (Version 8.54) [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.
- Lee, P., & Ashby, R. (2001). Empathy, perspective taking and rational understanding. In O.L. Davis Jr., E. A. Yeager, & S. J. Foster (Eds.), *Historical Empathy and Perspective Taking in the Social Studies* (pp. 21-50). Lanham, MD: Rowman & Littlefield.
- Linden, J. van der, & Renshaw, P. (Eds.) (2004). *Dialogic learning*. Dordrecht: Kluwer.
- McQuaide, J., Leinhardt, G., & Stainton, C. (1999). Ethical reasoning : real and simulated. *Journal of Education Computer Research*, 21(4), 433-474.
- Murray, K. (1999). Bioethics in the laboratory: synthesis & interactivity. *The American Biology Teacher*, 61(9), 662-667.
- Meijer, J., Elshout-Mohr, M., & Hout-Wolters, B.H.A.M. van. (2001). An instrument for the assessment of cross-curricular skills. *Educational Research and Evaluation*, 7(1), 79-107.
- Noddings, N. (1995). A morally defensible mission for schools in the 21st century. *Phi Delta Kappan*, 76(5), 365-368.
- Oser, F. K. (1986). Moral education and values education: The discourse perspective. In M.C. Wittwoc (Ed.), *Handbook of Research on Teaching*. (pp 917-941). New York: Macmillan.
- Oser, F. K. (1994). Moral perspectives on teaching. *Review of Research in Education*, 20, 57-127.
- Parker, W. C. (1997a). Democracy and difference. *Theory and Research in Social Education*, 25(2), 220-234.
- Parker, W. C. (1997b). The art of deliberation. *Educational Leadership*, 54(5), 18-21.
- Parker, W. C. (2001). Educating democratic citizens: A broad view. *Theory into Practice*, 40(1), 6-13.

- Phillips, R. (2002). *Reflective Teaching of History, 11-18: Continuum Studies in Reflective Practice and Theory*. London & New York: Continuum.
- Power, F., Higgins, A., & Kohlberg, L. (1989). *Lawrence Kohlberg's approach to moral education*. New York: Columbia University Press.
- Preskill, S. (1997). Discussion, schooling, and the struggle for democracy. *Theory and Research in Social Education, 25*(3), 316-345.
- Rasbash, J., Browne, N., Cameron, B. & Charlton, C. (2005). *MLwiN Version 2.02*. Multilevel Models Project. London: Institute of Education.
- Roegholt, S. (1995) *Meerperspectiefisch onderwijs*. Dissertatie. Amsterdam: Vrije Universiteit.
- Rojas-Drummond, S. & Mercer, N. (2003). Scaffolding the development of effective collaboration and learning. *International Journal of Educational Research, 39*(1), 99-111.
- Saye, J. (1998). Creating time to develop student thinking: team-teaching with technology. *Social Education, 62*(4), 356-362.
- Sadler, T.D., & Zeidler, D.L. (2005). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of research in science teaching, 42*(1), 112-138.
- Schuitema, J. A., Dam, G. ten, & Veugelers, W. (in press). Teaching strategies for moral education: a review. *Journal of Curriculum Studies*.
- Schuitema, J. A., Riessen, M. van, Veugelers, W. & Dam, G. ten. (submitted). *Making history valuable: moral and social development through history teaching*.
- Schultz, L., Barr, D., & Selman, R. (2001). The value of a developmental approach to evaluating character development programmes: An outcome study of facing history and ourselves. *Journal of Moral Education, 30*(1), 3-27.
- Solomon, D., Watson, M., & Battistich, V. (2001). Teaching and schooling effects on moral/prosocial development. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 566-603). Washington, D.C.: AERA.
- Tappan, M. B. (1998). Moral education in the zone of proximal development. *Journal of Moral Education, 27*(2), 141-160.
- Tirri, K., & Pehkonen, L. (2002). The moral reasoning and scientific argumentation of gifted adolescents. *The Journal of Secondary Gifted Education, 8*(3), 120-129.
- Torney-Purta, J. Lehmann, R. Oswald, H. & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen*. Amsterdam: IEA.
- Tredway, L. (1995). Socratic seminars: engaging students in intellectual discourse. *Educational Leadership, 53*(1), 26-29.
- Veugelers, W. (2000). Different ways of teaching values. *Educational Review, 52*(1), 37-46.
- Veugelers, W. & Vedder, P. (2003). Values in teaching. *Teachers and Teaching: theory and practice, 9*(4), 89-101.
- Wilson, M. S. (2001). Research on history teaching. In V. Richardson (Ed.), *Handbook of Research on Teaching* (pp. 527-544). Washington, D.C.: AERA.
- Wrenn, A. (1999) Build it in, don't bolt it on: history's opportunity to support critical citizenship. *Teaching History, 96*, 6-13.
- Yore, L., Bisanz, G.L., & Hand, B. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25*(6), pp. 687-725.

Manuscript aanvaard: 28 januari 2007

Auteurs

Jaap Schuitema is als promovendus verbonden aan het Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam.

Wiel Veugelers is als docent en onderzoeker verbonden aan het Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam en als hoogleraar educatie aan de Universiteit voor Humanistiek in Utrecht.

Gert Rijlaarsdam is als hoogleraar onderwijsontwikkeling verbonden aan het Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam.

Geert ten Dam is als hoogleraar onderwijskunde verbonden aan het Instituut voor de Lerarenopleiding van de Universiteit van Amsterdam.

Correspondentieadres: Jaap Schuitema, Instituut voor de Lerarenopleiding, UvA, Spinozastraat 55, 1018 HJ Amsterdam.
E-mail: j.a.schuitema@uva.nl.

Abstract

Two instructional designs for citizenship education in the history class: an effect study

This study investigates the effectiveness of group work in the context of dialogical citizenship education in the history class. Two instructional designs were developed and implemented in the 8th grade of pre-university education. Both instructional designs aim to stimulate the ability of students to reflect on moral values and multiple perspectives. The two designs differed in the amount of group work and whole classroom teaching. The ability to reflect on values and multiple perspectives was assessed by means of short essays that students wrote about a moral issue. The results show that students who worked proportionally more in small groups reflect more on values and different perspectives compared to students who participated more in whole classroom teaching.

Appendix

De betrouwbaarheden van de beoordelaars werd geschat met behulp van LISREL-modellen beschreven door Van den Bergh en Eiting (1986). De modellen bleken allemaal goed te passen ($\chi^2/v < 1,76$; $p > ,12$). De geschatte betrouwbaarheden varieerde tussen ,51 en 1,05.

Elke steekproef werd door twee personen beoordeeld en de uiteindelijk score is bepaald door het gemiddelde te nemen van die twee scores. De betrouwbaarheid van deze gemiddelde scores is berekend op basis van de twee betrouwbaarheden van de beoordelaars van de betreffende steekproef met behulp van de Spearman-Brown formule voor testverlenging (zie Tabel 4).

Tabel 4

Betrouwbaarheden per steekproef

Steekproef	Leerstofgebonden			Transfer	
	Waarden	Multipers.	Hist. red.	Waarden	Multipers.
S1	,87	,92	,76	,83	,82
S2	,88	,84	,86	,93	,87
S3	,83	,90	,78	,86	,78
S4	,80	,87		,88	,85
S5	,85	,91		,89	,90