Zeker weten?

Leren de kwaliteit van biologie-onderzoek te bewaken in 5 vwo

Are you certain?

Learning to ensure the quality of biology research
in pre-university education

Tot stand gekomen in samenwerking met het Centrum voor Didactiek van Wiskunde en Natuurwetenschappen van de Universiteit Utrecht.

© 2006 Onderwijscentrum VU, Vrije Universiteit Amsterdam

ISBN-10: 90-74580-15-7
ISBN-13: 978-90-74580-15-1

Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt door middel van druk, fotokopie of op andere wijze dan ook, zonder voorafgaande schriftelijke toestemming van de uitgever.
All rights reserved. No part of this publication may be reproduced in any form, by print or photo print, microfilm or any other means, without written permission of the publisher.

Omslag: Audiovisueel Centrum, Vrije Universiteit Amsterdam
Druk: Ipskamp Print Partners, Enschede

Zeker weten?

Leren de kwaliteit van biologie-onderzoek te bewaken in 5 vwo

Are you certain?

Learning to ensure the quality of biology research in pre-university education

(with a summary in English)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. W.H. Gispen,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen
op vrijdag 8 december 2006 des ochtends te 10.30 uur

door

Herman Hendrik Schalk

geboren op 1 januari 1956 te Dordrecht

Promotor: prof. dr. K.Th. Boersma

Co-promotor: dr. J.A. van der Schee

Inhoudsopgave

Woord vooraf ***1***

Proloog **Het Grote Uien Wortelharen Onderzoek** ***5***

Hoofdstuk 1 **Waarom 'zeker weten?'** ***9***

1.1 Aanleiding, richting en kader van het onderzoek *11*
1.2 Onderzoeksvraag en aanpak van het onderzoek *13*

Hoofdstuk 2 **Kwaliteit van onderzoek** ***17***

2.1 Wetenschappelijk onderzoek *18*
2.2 Biologisch onderzoek *19*
2.3 Kwaliteit vraagt 'begrip van bewijs' *24*

Hoofdstuk 3 **'Onderzoek doen' in het biologieonderwijs** ***31***

3.1 Onderzoek doen volgens vakliteratuur en officiële leerdoelen *32*
3.2 Onderzoek doen volgens docenten *38*
3.3 Onderzoek doen volgens biologiemethodes *42*
3.4 Onderzoek doen in leerlingverslagen *46*
3.5 Keuzes *51*

Hoofdstuk 4 **Kwaliteit leren bewaken** ***53***

4.1 Denken, doen en reflecteren *56*
4.2 Uitgangspunten voor twee onderwijsleerstrategieën: expliciet en impliciet *64*

Hoofdstuk 5 **De eerste cyclus** ***67***

5.1 Het eerste ontwerp *68*
5.2 Momenten van expliciete reflectie 70
5.3 Onderzoekspopulatie en instrumenten *72*
5.4 Bevindingen *77*
5.4.1 Uitvoerbaarheid *77*
5.4.2 Effectiviteit *80*
5.5 Verbeterpunten voor de tweede cyclus *86*

Hoofdstuk 6 **Aanscherping** ***89***

6.1 Meer motief 90
6.2 Een didactische structuur *91*
6.3 Meer zicht op de ontwikkeling van begrip van bewijs *93*
6.4 Verwachte samenhang tussen onderwijsleerstrategie en patronen in de ontwikkeling van begrip van bewijs *99*

Hoofdstuk 7 **De tweede cyclus** ***103***

7.1 Het tweede ontwerp *104*
7.2 Reflectieopdrachten *109*
7.3 De nieuwe populatie en instrumenten *112*
7.4 Uitvoerbaarheid *117*
7.4.1 Gegaan zoals gepland? *117*
7.4.2 De startlessen *124*
7.4.3 Motief als drijvende kracht *127*
7.4.4 Het beoogde docentgedrag *129*
7.4.5 Een éigen onderzoek *134*
7.4.6 Het boekje *135*
7.4.7 Een artikel als eindproduct *138*
7.4.8 De reflectieopdrachten *139*
7.4.9 De impliciete onderwijsleerstrategie *148*
7.4.10 Het wetenschappelijke verhaal *151*
7.4.11 Samenvattend: in welke mate is de praktische opdracht uitvoerbaar? *152*
7.5 Effectiviteit *155*
7.5.1 De startlessen *156*
7.5.2 De weg naar de eindproducten *158*
7.5.3 De eindproducten *172*
7.5.4 Begin- en eindtoets *175*
7.5.5 Ervaren leereffect *179*
7.5.6 Samenvattend: in welke mate is de praktische opdracht effectief? *180*
7.6 Kan het beter? *182*

Hoofdstuk 8 **Terugblik en vooruitblik** ***187***

8.1 Antwoord op de onderzoeksvraag *188*
8.2 Hoe zeker is het antwoord? *191*
8.3 Reflectie op de theoretische uitgangspunten *194*
8.4 Lijnen naar betekenisvol leren *197*
8.5 Verder onderzoeken *202*
8.6 Hoog genoeg mikken *205*

Epiloog **Het korte-termijn geheugen** ***207***

Literatuur ***211***

Lijst van bijlagen ***223***

Samenvatting ***227***

Summary ***237***

Curriculum vitae ***247***

23 elementen van begrip van bewijs ***250***

Woord vooraf

Woord vooraf

Wat is goed onderzoek?

In mijn studentenjaren was goed onderzoek vooral onderzoek dat maatschappelijk relevant was. Ik heb veel discussies gevoerd over de illusie van 'waardevrije wetenschap'. Wetenschapsbeoefening kon en mocht niet in een ivoren toren plaatsvinden, maar moest in dienst staan van het welzijn van alle mensen en niet te vergeten de natuur. De ervaringen met de praktijk van de wetenschap sterkten me in die stellige overtuiging. Puzzelen met gegevens vond ik leuk, maar besefte ik wel welke belangen er uiteindelijk mee gediend waren? Competitie en publicatiedrang. Onderzoek om zeker te weten? Schijn bedriegt!

In mijn jaren voor de klas was goed onderzoek vooral onderzoek dat leerlingen enthousiast maakte voor de natuur en de biologie. Primair ging het om de ervaring. Al in de cursus didactiek – mijn opleiding tot leraar – kwam ik in aanraking met onderzoek door leerlingen. De werkweken van 4 havo en 2 vwo waar leerlingen eigen onderzoek opzetten en uitvoerden lieten die manier van 'wetenschap' bedrijven zien: leerlingen nieuwsgierig naar een antwoord op een zelfbedachte vraag. Het was het naïeve beeld dat daar naar voren kwam, maar het enthousiasme van de leerlingen was er niet minder om. Tegelijkertijd waren de leerlingen snel tevreden, ze problematiseerden de kwaliteit van hun antwoorden nauwelijks. Onderzoek om zeker te weten? Schijn bedriegt!

In mijn jaren bij het Cito kwam onderzoek van leerlingen ter sprake bij het maken van basisvormingstoetsen en vwo-examens. Wat moesten ze weten en kunnen? Wat kon je vragen en wat niet? Goed onderzoek werd dan uiteindelijk goed scoren op een practicumtoets of op vragen over deelvaardigheden in een schriftelijke toets. Onderzoek splitsen in deelvaardigheden is makkelijk als je een antwoordmodel en een scoringsvoorschrift moet opstellen. Maar kan een goedscorende leerling ook goed onderzoek doen? Schijn bedriegt!

Toen ik in de gelegenheid werd gesteld om binnen het kader van een onderzoeksprogramma rond onderzoeksvaardigheden van leerlingen een eigen onderwerp voor een promotieonderzoek te bepalen, was – gezien het bovenstaande – de keus niet moeilijk: de kwaliteit van leerlingonderzoek. Ik ben daarbij niet uitgekomen op de maatschappelijke relevantie ervan, niet op het enthousiasme waarmee het wordt uitgevoerd noch op het inzoomen op deelvaardigheden. Na een oriënterende periode, waarin ik gesprekken voerde met collega's en docenten, heb ik ingezet op een overkoepelend besef: een besef van kwaliteit, van 'weet ik het nu zeker?' en 'hoe weet ik of ik het zeker weet?' Sommigen noemen dat 'een wetenschappelijke houding', anderen 'procedurele kennis', maar hoe je het ook noemt, het is mij gaandeweg steeds duidelijker geworden dat dit de essentie van 'goed onderzoek' is.

En goed onderzoek doe je niet alleen. In dit proefschrift wordt op verschillende plaatsen het sociale karakter van wetenschapsbeoefening benadrukt. Ik wil dan ook de mensen die hebben bijgedragen aan de kwaliteit van míjn onderzoek op deze plaats hartelijk bedanken.

Waar zou dit onderzoek geëindigd zijn zonder de niet-aflatende positieve feedback die ik van mijn promotor en copromotor, Kerst Boersma en Joop van der Schee, heb gekregen? Mijn motor loopt op waardering, jullie zorgden dat mijn tank nooit leeg raakte. De ideeën die we samen genereerden, jullie steun en vertrouwen én de vrijheid om mijn eigen keuzes te maken, stelden me in staat de eindstreep te behalen. Zeer veel dank!

Een zeer waardevolle rol werd gespeeld door de klankbordgroep, Annelies van Goor, Lieke Kievits, Dirk van der Wulp, Jeroen den Hertog en Wim Bax. De ruim honderd jaar on-

derwijservaring (en niet de minste!) die rond de tafel zat als wij bijeenkwamen was zowel geruststellend als prikkelend en stimulerend. Ik kon mijn plannen en voorlopige conclusies blootstellen aan jullie positief kritische blik. Ook voor mij was dit – zoals Dirk het een keer uitdrukte – gratis nascholing!

De discussies met collega's, hierboven al even gemeld, waren eveneens waardevol voor mijn gedachtevorming. Ik voerde ze in vele gremia: de onderzoeksgroep en het bètacluster van het Onderwijscentrum VU, de Vereniging voor Didactiek van de Biologie, de redactie van *Niche* en niet te vergeten de wandelgangen. Met name wil ik Mieke Kapteijn, Lisette van Rens, Douwe Kok, Corinne van Velzen, Fred Korthagen, Mieke Lunenberg, Martijn Willemse, Herman Jonker, Anja Swennen, Alfred Schermer, Frans Wagenaar, Marcel Kamp, Michiel van Eijck en Els de Hullu bedanken voor de warme belangstelling, verwarrende én verhelderende discussies. Jullie hielden – en houden – me scherp!

Bijzonder veel dank ben ik ook verschuldigd aan de docenten bij wie ik in de klas onderzoek gedaan heb, Brigit Verbeek en Louis Siefkens, mijn koningskoppel, in de rest van dit proefschrift dan ook aangeduid met Bea en Lex. Jullie betrokkenheid, inzet en inspanningen, onontbeerlijk en essentieel voor het slagen van het project, kunnen niet genoeg benadrukt worden. Als ik jullie daden soms als 'niet optimaal' bestempel, is dat geenszins als afkeuring bedoeld. Het laat slechts zien dat ook jullie geen supermensen zijn. Maar ik vind het wel super wat jullie voor mijn onderzoek gedaan hebben. Uiteraard ook dank aan jullie scholen – de 'School in Amstelveen' en de 'Amsterdamse School' – voor de gelegenheid die ze boden om het onderzoek uit te voeren.

De leerlingen verdienen ook zeker een woord van dank. Je bloot stellen aan onderwijs – hoe doordacht vormgegeven ook – jullie zijn niet anders gewend. Opnameapparatuur op tafel, ook dat wende. En al spraken jullie daarin regelmatig je verbazing uit dat iemand dat allemaal zou naluisteren – laat staan uittypen – de opnames hebben me onschatbaar rijk materiaal opgeleverd. Behalve inzicht in jullie inzicht in onderzoek, gaven ze me een kijkje in de (school)cultuur van 17-jarigen. Ik heb gesmuld!

En dan de hulptroepen. Zonder het typewerk van Isabelle Sliepen, Hanneke Merten, Nicolien van Peursem, Ellen van der Woerd en Jan Maarten van Schalkwijk was dit onderzoek halverwege gestrand. Hulde aan jullie doorzettingsvermogen! De stageonderzoeken van (wederom) Isabelle Sliepen, Hans Ferket, Rick Appelman, Ahmet Yuksel en Rijco van Egdom hebben het onderzoek wezenlijk verder geholpen. Dank voor deze win-win producties. De analyses en verdere ondersteuning van Marthe Straatemeier, (wederom) Rijco van Egdom en Suzanne Langius hebben me veel werk uit handen genomen. Ik zou nu nog bezig zijn geweest! Speciale dank is er voor Frans Kleintjes, oud-collega van het Cito, voor zijn waardevolle adviezen in de slotfase. Die slotfase werd ook aanzienlijk verlicht door de inspanningen van Gerrie Buijze en Els Breedveld; ook jullie verdienen mijn dank.

Ik mag me verheugen in de aanwezigheid van Lisette van Rens en Frans Wagenaar als paranimfen. Lisette, jouw adviezen, vooral aan het begin en het eind van mijn onderzoekstraject, waren zeer waardevol en steunend. En Frans, met jouw niet aflatende, energieke inzet voor goed biologieonderwijs heb je me meer dan eens geïnspireerd. Ik reken me rijk dat jullie me vergezellen bij de promotie.

De discussies thuis aan tafel – redeneringen die die vuurproef doorstaan kun je met een gerust hart wereldkundig maken – hebben me ervoor behoed de dagelijkse werkelijkheid van het onderwijs niet uit het oog te verliezen. Lieverds, bedankt dat jullie mijn relativeringsvermogen zo gestimuleerd hebben. Wat een voorrecht, mazzel en luxe om in het belangrijkste onderzoeksjaar een 5-vwoër in huis te hebben. En Ada, mijn lief, wat zou ik zijn zonder jouw onvoorwaardelijk kritische steun en onmetelijke liefde? Zelfs de maan …

Proloog

Het Grote Uien Wortelharen Onderzoek

'In zo'n onderdeel kun je goed zien wat leerlingen zelfstandig kunnen.'
Daan van der Haar, docent biologie, over eigen experimenteel onderzoek.
(geciteerd in Van Goor & Den Hertog, 2001, p. 16)

Proloog Het Grote Uien Wortelharen Onderzoek

Leerlingen doen vaak met enthousiasme een eigen onderzoek. Dat enthousiasme is echter geen garantie voor kwaliteit. De onderstaande fragmenten uit het verslag, met bovenstaande titel, van een onderzoek door twee leerlingen uit 5 vwo zijn bedoeld om dat te illustreren. Op het eerste gezicht lijkt het een aardig onderzoek en verslag, maar wat wilden ze nu eigenlijk weten? Hoe zeker is het antwoord dat ze vinden en welke besef van (on)zekerheid laten deze leerlingen zien?

Een opmaat voor de probleemstelling in hoofdstuk 1.

(uit 'Inleiding')

Voor de praktische opdracht biologie 2003 hebben wij, Laura Mensrecht en Lidwien Sterling[1], besloten om de functie van wortelharen bij een plant te onderzoeken. De functie van wortelharen is het opnemen van voedingsionen. Wij kwamen op dit onderwerp omdat er ten eerste weinig scheikunde bij komt kijken. Wij wilden liever een onderzoekje zonder scheikunde omdat wij geen scheikunde in ons pakket hebben en er niet ontzettend in geïnteresseerd zijn. De tweede reden was dat het onderzoekje niet te ingewikkeld moest zijn en niet een half jaar onderzoekstijd moest innemen. De plant die wij voor ons onderzoek gaan gebruiken is de ui.
Onze onderzoeksvraag luidt:

Is de rol van de wortelharen van een ui bij de opname van water voor de ui van levensbelang?
(…)
Wij hebben natuurlijk wel enige ideeën waarop dit onderzoek zal uitlopen. Dus hier komt onze hypothese:

Wij denken dat wij de uien waarvan de wortelharen zijn gelakt en verwijderd tijdens ons onderzoek zullen gaan zien aftakelen. Dus het loof wordt geel en verdort en de wortels worden zacht en snotterig.

(uit 'Conclusie')

Onze conclusie is dat de rol van de wortelharen van een ui van levensbelang zijn voor de ui. Dit concluderen wij op grond van onze waarnemingen. Uit onze proef bleek namelijk dat de meeste uien uit opstelling 2, waarvan de wortelharen waren gelakt, de meeste aftakelingsverschijnselen vertoonden en geen vooruitgang boekten. De meeste uien uit opstelling 3, waarvan we de wortelharen hadden verwijderd, vertoonden ook aftakelingsverschijnselen en boekten ook geen vooruitgang. De uien uit opstelling 1, die onder normale omstandigheden verkeerden, boekten daarentegen wel vooruitgang.
(…)

[1] De namen van de leerlingen zijn veranderd, met handhaving van het geslacht.

Hieruit kunnen we het in onze inleiding genoemde feit over wortelharen bevestigen: wortelharen nemen voedingsionen op. Zonder wortelharen of met gelakte wortelharen nemen de uien dus geen of minder voedingsionen op waardoor de uien aftakelen.

(uit 'Foutendiscussie')

Toch hadden wij nog enkele onopgeloste vragen over ons onderzoekje. Was de schimmel die bij sommige uien aanwezig was gevaarlijk voor ons onderzoek? En hoe kwam het dat ui 4 en ui 7 loof schoten terwijl de andere uien van hun opstellingen aan het aftakelen waren?
De antwoorden konden we gelukkig vinden. De schimmel op de uien was niet gevaarlijk voor ons onderzoek. De schimmel zat namelijk op het dorre omhulsel van de ui dus dat konden wij verwijderen zonder de ui te beschadigen.
Het antwoord op onze tweede vraag was iets moeilijker te beantwoorden. Maar we kwamen er achter. Ui 4 hoorde bij opstelling 2, waarvan de wortelharen waren gelakt. Dus wij hadden verwacht dat deze ui zou aftakelen. Dit zelfde verwachtten wij bij ui 7 uit opstelling 3, waarvan de wortelharen waren verwijderd. Wat bleek? De wortels van deze uien waren na het lakken en het verwijderen van de wortelharen verder gaan groeien. Dus ze hadden een nieuw gedeelte wortel gekregen wat niet was gelakt en waarvan de wortelharen niet waren verwijderd. En dat gedeelte bevatte blijkbaar genoeg wortelharen om genoeg voedingsionen mee op te nemen.
Uit dit onderzoek konden wij niet opmaken of het verder groeien van de wortels van de behandelde uien een voorzorgsmaatregel van de ui was. Dit zouden we in een volgend onderzoek kunnen onderzoeken.

1 Waarom 'zeker weten?'

1.1 Aanleiding, richting en kader van het onderzoek
1.2 Onderzoeksvraag en aanpak van het onderzoek

'The real problem of intellectual education is the transformation of more or less casual curiosity and sporadic suggestion into attitudes of alert, cautious, and thorough inquiry.'
John Dewey (1933, p. 181)

Hoofdstuk 1 Waarom 'zeker weten?'

Onderzoek doen is gericht op het vinden van een antwoord op een vraag. De motivering van de onderzoeksvraag in het leerlingenverslag in de proloog geeft geen blijk van een grote gedrevenheid van de leerlingen om met hun onderzoek een zo zeker mogelijk antwoord op een prangende vraag te krijgen. Evenmin blinkt de vraag uit in helderheid. Willen ze de functie van wortelharen weten? Nee, die weten ze al, maar dat is blijkbaar onvoldoende om te concluderen dat de ui niet zonder kan. Zijn ze zeker van hun antwoord? Ja, met nog enkele vraagjes, die ze gelukkig weten te beantwoorden. Het tekent de situatie die rond leerlingonderzoek in het biologieonderwijs vaak aangetroffen wordt: *more or less causal curiosity and sporadic suggestion*. Maar wat zijn dan de *attitudes of alert, cautious and thourough inquiry* die Dewey in bovenstaand citaat als contrast daartegenover zet? Die houding omschrijven Smits, Lijnse en Bergen (2000, p. 17), ons inziens treffend, als volgt: 'Ieder dient zich bij iedere wetenschappelijke bewering, van zichzelf of anderen, voortdurend af te vragen of die, in het licht van de beschikbare evidentie, wel correct is.' Een houding die de vraag 'zeker weten?' voor ogen houdt, zo zou je de essentie ervan kunnen verwoorden.

Onderzoek doen is gericht op het vinden van een antwoord op een vraag en daarbij is het de bedoeling dat het antwoord voldoende zekerheid heeft. Om die zekerheid zo groot mogelijk te laten zijn, dient onderzoek zorgvuldig opgezet en uitgevoerd te worden. Als leerlingen in het voorbereidend wetenschappelijk onderwijs (vwo) iets moeten leren over het doen van onderzoek zal die zorgvuldigheid dus aan de orde moeten komen. Als ze moeten leren onderzoeken zullen ze die zorgvuldigheid ook moeten leren betrachten.

De westerse samenleving is doorspekt met resultaten van wetenschappelijk onderzoek en dagelijks komen nieuwe bevindingen in het nieuws. Dat varieert van 'Aarde sneller warmer dan verwacht' tot 'HIV-virus toch van chimpansees' en 'Huisarts te makkelijk met antibiotica'. Vaak is niet duidelijk hoe zorgvuldig het achterliggende onderzoek is uitgevoerd. Ook als leerlingen zelf geen onderzoeker worden, is het dus zinvol de zorgvuldigheid van onderzoek van anderen op waarde te kunnen schatten.

Het leren betrachten en beoordelen van zorgvuldigheid bij het doen van onderzoek is het onderwerp van dit proefschrift. De vraag van dit onderzoek komt voort uit de constatering dat het leren van zorgvuldigheid in het doen van onderzoek in het huidige vwo wel wat te wensen over laat. Dat is het onderwerp van paragraaf 1.1.

Onderzoek doen is gericht op het vinden van een antwoord op een vraag. Om een duidelijk antwoord te vinden, is een duidelijke vraag nodig. Daarom worden in paragraaf 1.2 de onderzoeksvraag en de daarvan afgeleide deelvragen geformuleerd en wordt uitgelegd op welke manier die aangepakt zijn. Daarmee is dan meteen de opbouw van dit proefschrift gegeven.

1.1 Aanleiding, richting en kader

- *Aanleiding*

Onderzoek door leerlingen is al jaren een bekend fenomeen in het Nederlandse biologieonderwijs en wordt ook al decennia gepropageerd in publicaties voor biologiedocenten (o.a. Falk, Keuchenius en Saaltink, 1974; Van Alderwegen-de Vries, Kapteijn en Thijssen, 1988; Griethuysen-Elfers, Kapteijn, Prop-van den Berg, Westra & Van Wijk, 1991). Het is niet pas ingevoerd met de invoering van de tweede fase, zoals onderzoek bij scheikunde (Van Rens, 2005). Eerder neemt de door docenten ervaren ruimte voor onderzoek bij biologie af (Morélis, 2001; Van Goor & Den Hertog, 2001). De lange traditie van onderzoek doen in het biologieonderwijs doet vermoeden dat er een doortimmerde didactiek voor ontwikkeld is en er een stevige praktijk is gegroeid. Dat is echter maar voor een beperkt deel van het biologieonderwijs het geval. Er zijn scholen waar er al jaren veel aan wordt gedaan (Griethuysen-Elfers e.a., 1991; Bax, Bos & Van Mourik, 1997; Van Goor & Den Hertog, 2002), maar ook scholen waar het nauwelijks ontwikkeld is (Kuiper, 1993). Sinds 1991 staan er in de examenprogramma's biologie voor havo en vwo onderzoeksvaardigheden genoemd en dat heeft geleid tot het opnemen van vragen daarover in de centrale examens (De Knecht-Van Eekelen & Pohlmann-Nepveu, 1994), maar niet tot een praktijk van onderzoek doen op alle scholen.

Het bewaken van de kwaliteit is een aspect van onderzoek doen dat bij leerlingen niet hoog scoort. Leerlingen beschouwen onderzoek doen vooral als 'iets te weten komen', 'een vermoeden controleren' of 'het vinden van een antwoord op een vraag' (Hubers, 2003). De kwaliteit van het antwoord houdt hen minder bezig: 'nagaan hoe het beter had gekund' zetten ze op een 13^{e} plaats van belangrijkheid in een rij van 19 onderzoeksvaardigheden, zowel voor henzelf als voor onderzoekers. Bovenaan zetten ze voor zichzelf 'conclusies trekken' en 'gegevens uitwerken'. In de praktijk blijkt dat ook: 'kennelijk wilden de leerlingen de onderzoeksvraag wel beantwoorden, maar hechtten ze er minder waarde aan om dat zo betrouwbaar en valide mogelijk te doen' (Van Rens & Dekkers, 2000, p. 86). Ook anderen constateren dat er meer aandacht voor validiteit en betrouwbaarheid in het leren onderzoeken door leerlingen zou moeten zijn (o.a. Smits, 1997; De Vos & Genseberger, 2000; Smits e.a., 2000).

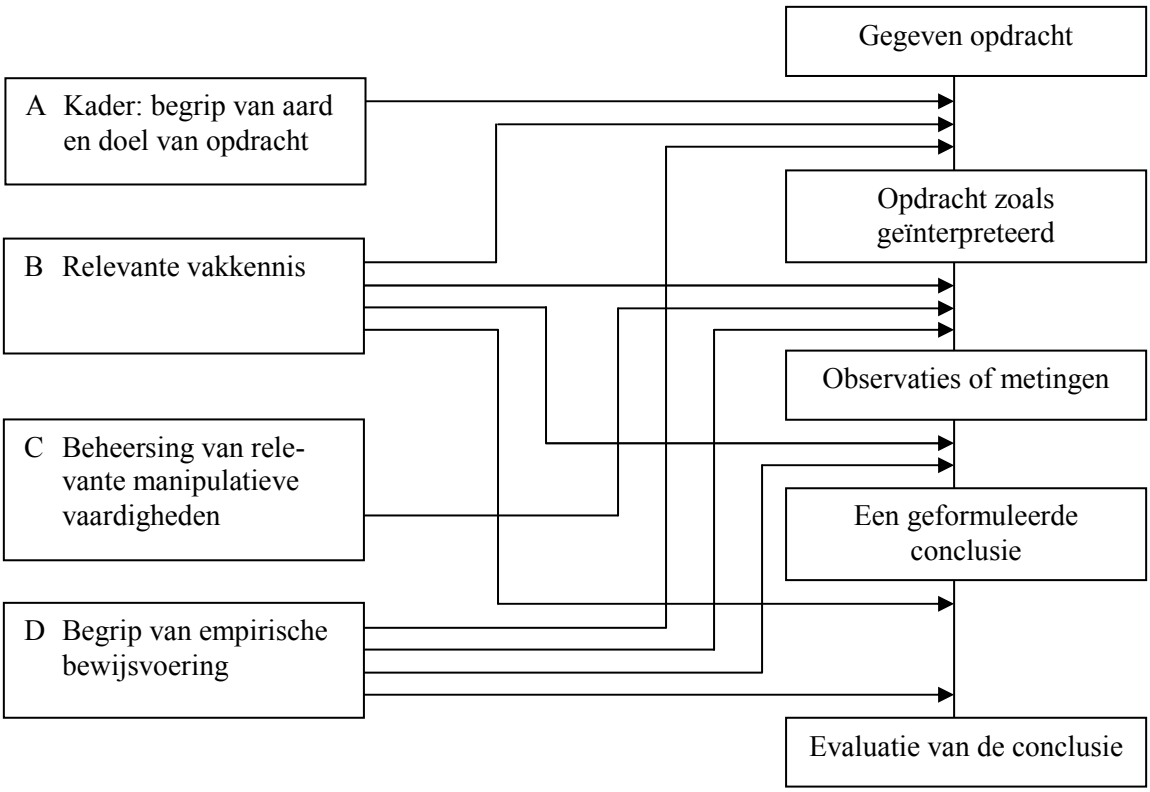
Niet alleen in Nederland is het bewaken van de kwaliteit in het doen van onderzoek door leerlingen een punt van zorg. In een vergelijkende studie van practica in de natuurwetenschappelijke vakken in zeven Europese landen (Tiberghien, Veillard, Le Marechal, Buty & Millar, 2001) blijkt dat daarin het opzetten van een onderzoek en het onderbouwen van een conclusie met gegevens een ondergeschikte rol spelen. In het Verenigd Koninkrijk blijkt begrip van bewijsvoering in het science-onderwijs erg weinig aanwezig (o.a. Gott & Duggan, 1995b; Gott & Johnson, 1999). Ook voor het biologieonderwijs wordt die constatering gedaan (Roberts & Gott, 1999). Millar, Lubben, Gott, en Duggan (1994, p. 245) betogen op basis van hun bevindingen in Engeland dat 'children's understanding of empirical evidence and of criteria for evaluating the quality of evidence needs to be addressed.' De redenen waarom dat belangrijk is, komen zowel uit de praktijk van wetenschappelijk onderzoek – 'evidence is the basis for testing competing theories' – als uit het beroepsperspectief – het kunnen nemen van beslissingen op basis van gegevens – als uit het perspectief van de kritische burger (Gott & Duggan, 1996, p.794). Ze voegen er elders nog aan toe dat het begrijpen van bewijsvoering leerlingen duidelijker maakt 'what it is to be a scientist' (Duggan & Gott, 1996, p. 30).

- *Richting*

Gegeven deze aanleiding is het dan de vraag op welke manier het gesignaleerde probleem benaderd kan worden. In een beschrijvend onderzoek kan gekeken worden hoe het in het algemeen aangepakt wordt en welke aanpak de beste resultaten geeft. Dat leidt dan tot curriculumonderzoek zoals bijvoorbeeld Kuiper (1993) dat doet. In een meer analytisch en

verklarend onderzoek kan ook gekeken worden welke ingrediënten *good practices* hebben, onderwijs dat al dicht in de buurt komt van het beoogde, zoals bijvoorbeeld Crawford, Krajcik en Marx (1999) en Crawford (2000) dat doen voor onderzoekend leren. Een andere manier om een effectieve onderwijsleerstrategie te ontdekken, is er een te ontwerpen in een ontwikkelingsonderzoek (Lijnse, 1995; Boersma, 1999; Van den Akker, 1999; Boersma, Knippels & Waarlo, 2005). Door de ontwikkeling van zo'n strategie, die liefst in samenwerking met direct betrokkenen plaatsvindt, nauwgezet te volgen en het ontwikkelde onderwijs uit te proberen en te evalueren, kan onderzocht worden of de onderwijsleerstrategie *goed genoeg* is om aan het gesignaleerde probleem tegemoet te komen. Uiteraard is het voor het ontwikkelen nodig om naar de bestaande praktijk te kijken, maar het biedt ook de mogelijkheid om iets nieuws aan die praktijk toe te voegen. Voor het onderwijs in de natuurwetenschappen zijn in het recente verleden enkele, vaak succesvolle ontwikkelingsonderzoeken uitgevoerd (o.a. Knippels, 2002; Kortland, 2001; Van Eijck, 2006; Van Rens, 2005; Verhoeff, 2003).

De keus voor een ontwikkelingsonderzoek zegt echter nog niets over de richting waarin je dan gaat ontwikkelen. Wat zijn bronnen voor nieuwe inspiratie en ingrediënten om een onderwijsleerstrategie te ontwerpen? Aanknopen bij ontwikkelingen elders lijkt dan een goede optie. Het model dat in het PACKS-project (*Procedural And Conceptual Knowledge in Science;* Millar e.a, 1994) is ontwikkeld leent zich daar goed voor (fig.1.1). Het geeft aan welke categorieën van factoren (A, B, C en D) de kwaliteit van onderzoek door leerlingen (rechts in het model) beïnvloeden. Begrijpen van het doel van de opdracht en beheersen van relevante manipulatieve vaardigheden zijn nodig bij respectievelijk het beginnen en het uitvoeren van een onderzoek. Vakkennis en begrip van bewijsvoering spelen in alle fasen van een onderzoeksopdracht een rol. Dat strookt met de opvatting die in het citaat eerder in dit hoofdstuk is weergegeven en die dezelfde auteurs later in het artikel opnieuw formuleren:



Figuur 1.1 Het PACKS-model (Millar e.a., 1994).

jezelf voortdurend de vraag stellen in hoeverre datgene wat je doet te verantwoorden is als het best mogelijke in de gegeven omstandigheden (Smits e.a., 2000, p. 23). Deze *understanding of empirical evidence* wordt verder omschreven als het begrijpen van criteria waarmee de kwaliteit van empirisch verkregen gegevens kan worden geëvalueerd. Nadruk komt daarbij te liggen op de overkoepelende begrippen van validiteit en betrouwbaarheid. Wat die criteria inhouden is door Gott en Duggan (1995a, 1996) eerst globaal uitgewerkt tot een lijstje van twaalf *concepts of evidence*. Op hun website (Gott, Duggan & Roberts, 2006) werken ze die regelmatig bij en is het inmiddels uitgebreid tot een lijst van 107 begrippen die het niveau van onderzoek doen in het voortgezet onderwijs verre overstijgt.

Het operationaliseren van de kwaliteit van onderzoek met behulp van *concepts of evidence* lijkt een begaanbare weg om leerlingen het bewaken ervan te laten leren. In hoofdstuk 2 wordt deze operationalisatie verder uitgewerkt.

- *Kader*

Een vraag in de huidige wetenschap staat nooit op zichzelf, maar komt voort uit eerdere vragen en antwoorden en vaak is hij onderdeel van een onderzoeksprogramma. Dat is niet anders met het in dit proefschrift beschreven onderzoek.

Uit de beschrijving van de aanleiding en de richting blijkt al dat het onderzoek zich richt op vragen uit de praktijk van het huidige biologieonderwijs en de didactiek daarvan. Daarmee is al een stevig kader gegeven.

Een tweede kader waarin dit onderzoek staat is in eerste instantie gevormd door het onderzoeksplan 2000-2003 'Didactische toerusting van docenten en hun opleiders' van het Instituut voor Didactiek en Onderzoekspraktijk (IDO/VU, in 2003 opgegaan in het Onderwijscentrum VU). Daarin wordt onder andere de aandacht gericht op onderzoek naar het verwerven van onderzoeksvaardigheden. Maar het onderhavige onderzoek is ook verankerd in het nieuwe onderzoeksprogramma (2004-2008) van het Onderwijscentrum VU 'Kennisontwikkeling tussen theorie en praktijk'[2]. Dan gaat het o.a. om de vraag hoe kennisontwikkeling van leerlingen plaatsvindt in wisselwerking tussen theorie en praktijk én hoe die kan worden bevorderd. Zoals uit het bovenstaande blijkt, omvat het doen van onderzoek meer dan het aanleren van een aantal vaardigheden, het heeft ook een duidelijke kennisbasis. De keus voor een ontwikkelingsonderzoek zet in op een sterke wisselwerking is tussen theorie en praktijk, zoals in de volgende paragraaf wordt beschreven.

1.2 Onderzoeksvraag en aanpak van het onderzoek

Op basis van de hierboven gesignaleerde onvrede rond de kwaliteit van leerlingonderzoek en de keus om meer aandacht te besteden aan *concepts of evidence* is het zinvol om te onderzoeken of gerichte aandacht voor begrip van bewijs de kwaliteit van leerlingonderzoek ten goede zou komen, maar met name of het ertoe zou leiden dat leerlingen zélf beter die kwaliteit in de gaten houden. Dat leidt tot de volgende onderzoeksvraag:

Hoe kunnen leerlingen in het biologieonderwijs in de tweede fase VO leren de kwaliteit van het door hen uitgevoerde eigen onderzoek beter te bewaken?

[2] Dit onderzoeksprogramma is te raadplegen op de website van het Onderwijscentrum VU: www.onderwijscentrum.vu.nl

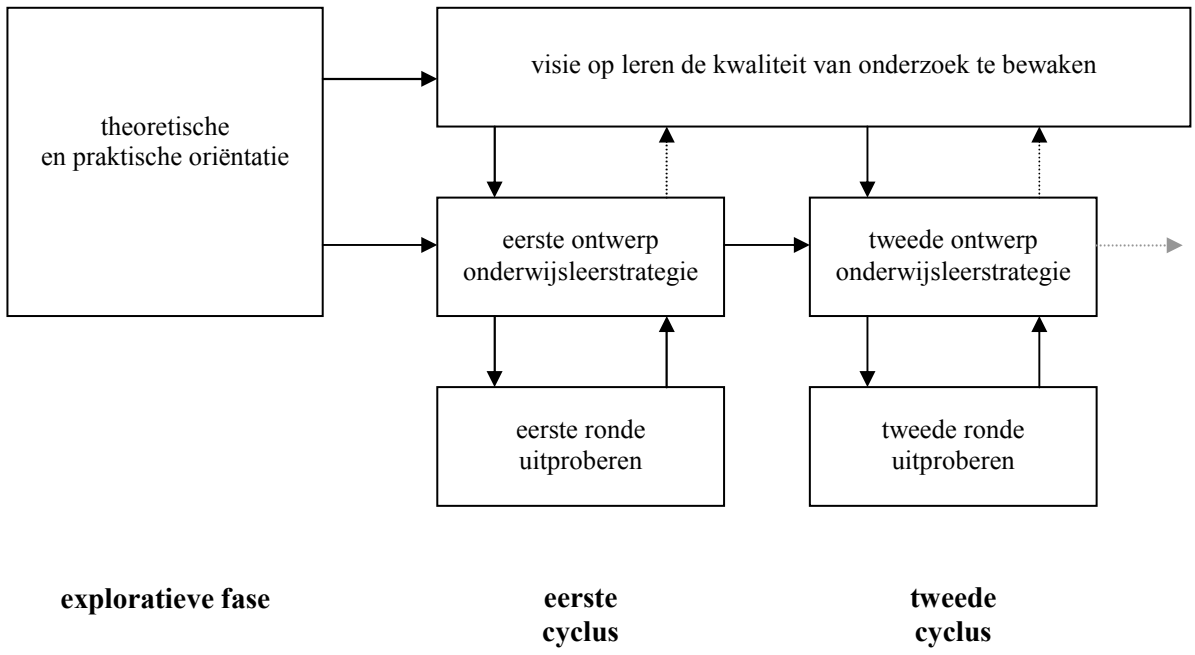
Waarbij het volgende verstaan wordt onder

- biologieonderwijs: het geheel van activiteiten van docent (organiseren, instrueren, begeleiden, beoordelen) en leerlingen (bestuderen van teksten, uitvoeren van opdrachten, reflecteren) binnen het schoolvak biologie;
- leren: het versterken en uitbreiden van kennis en inzicht van leerlingen met nieuwe begrippen en nieuwe relaties tussen begrippen (zie hoofdstuk 4);
- kwaliteit van het eigen onderzoek: de mate waarin het onderzoek voldoet aan de criteria zoals geoperationaliseerd in 'begrip van bewijs' (zie par. 2.3);
- beter bewaken: het actief betrekken van de elementen van 'begrip van bewijs' op het eigen onderzoek, en dat in sterkere mate dan nu gebeurt in het biologieonderwijs.

De tweede fase VO is in de praktijk van dit onderzoek klas 5 vwo.

Wij hebben dit onderzoek opgezet als een ontwikkelingsonderzoek (Van den Akker, 1999, Boersma e.a., 2005). De opzet van het onderzoek is weergegeven in figuur 1.2. Het onderzoek kent een exploratieve fase en twee ontwikkel- en onderzoekscycli. In de exploratieve fase vindt een theoretische en praktische oriëntatie plaats op kwaliteit van onderzoek, onderwijskundige theorieën en op de praktijk van het huidige biologieonderwijs. Die exploratie voedt zowel de visie op het leren bewaken van de kwaliteit van onderzoek als de ontwikkeling van het eerste concrete ontwerp voor een onderwijsleerstrategie. Het uitproberen van het eerste ontwerp levert feedback op zowel het ontwerp van een onderwijsleerstrategie als op de visie. De feedback leidt tot een bijgestelde strategie en een tweede ontwerp. Een tweede ronde uitproberen levert vervolgens aanwijzingen voor een eventuele derde versie van de onderwijsleerstrategie voor en visie op het leren bewaken van de kwaliteit van onderzoek.

Het onderzoek is begeleid door een klankbordgroep, bestaande uit vijf docenten met ruime ervaring met het begeleiden van leerlingen in het doen van onderzoek. De docenten zijn afkomstig uit het netwerk van het Onderwijscentrum VU en brengen met elkaar meer dan honderd jaar onderwijservaring in de bovenbouw van havo en vwo mee. In een tiental bijeen-



Figuur 1.2 Opzet van het onderzoek

komsten verspreid over alle fasen van het onderzoek hebben zij feedback gegeven en mede inhoud gegeven aan de visie op het leren bewaken van de kwaliteit van onderzoek.

- *Deelvragen*

Om te komen tot een antwoord op de centrale onderzoeksvraag is een aantal deelvragen gesteld die achtereenvolgens in de verschillende fasen van het onderzoek aan de orde komen. In de exploratieve fase is antwoord gezocht op een drietal deelvragen die van belang zijn om een effectief ontwerp van een onderwijsleerstrategie te kunnen maken.

(1) Wat is de biologische invulling van 'begrip van bewijs'?

Als je leerlingen wilt leren de kwaliteit van hun biologische onderzoek beter te bewaken, kun je niet zonder een beschrijving van wat er dan in dat kader onder kwaliteit van onderzoek wordt verstaan en hoe je die kunt bewaken. Die beschrijving staat in hoofdstuk 2: *Kwaliteit van onderzoek.* De *concepts of evidence* zoals Gott en Duggan (2003) en Gott e.a. (2006) die uitwerken, zijn niet zonder bewerking geschikt voor het biologieonderwijs. Een bewerking wordt in paragraaf 2.3 gegeven.

(2) Wat is de plaats van onderzoek doen in het biologieonderwijs?

Onderzoek doen is al decennia lang een deel van het biologieonderwijs, maar de plaats ervan in het curriculum is wel aan verandering onderhevig. Ook is er een verschil tussen het 'imaginaire curriculum' (Goodlad, Klein & Tye, 1979) zoals dat in de (internationale) literatuur geschetst wordt, het officiële curriculum in het examenprogramma en het uitgevoerde curriculum zoals dat in de praktijk gestalte krijgt. Bovendien is het zinvol de vraag te stellen welke plek biologiedocenten het bewaken van de kwaliteit van onderzoek zouden willen geven.

Een andere manier om te kijken naar de plaats van onderzoek doen in het biologieonderwijs is kijken naar welke stimulansen leerlingen daar nu al voor krijgen. Wat wordt er in het biologieonderwijs al aan gedaan? Drie bronnen werpen daar licht op: docenten, biologiemethoden en leerlingproducten. De exploratie van de plaats van onderzoek doen in het biologieonderwijs wordt beschreven in hoofdstuk 3. Dat leidt aan het eind daarvan tot argumenten voor de keus van elementen van begrip van bewijs voor de te ontwikkelen onderwijsleerstrategie.

(3) Hoe ontwikkelen leerlingen begrip van bewijs?

Uiteindelijk doel is dat leerlingen de kwaliteit van hun eigen onderzoek goed kunnen bewaken. Met andere woorden dat ze begrip van bewijs hebben en dat het in hun kennisstructuur verankerd wordt. Om lesmateriaal of richtlijnen voor instructie en begeleiding te ontwerpen, is het nodig een idee te hebben hoe dat proces van verankering verloopt. In hoofdstuk 4 leidt een verkenning van de onderwijskunde en vakdidactiek tot het begin van een visie op hoe ze het zouden kunnen leren.

Die visie behoeft echter toetsing aan de praktijk door het uitwerken en uitproberen van opdrachten voor leerlingen en begeleidingsstrategieën voor docenten. In het onderzoek zijn die opdrachten en strategieën in beide cycli in nauwe samenwerking met docenten ontwikkeld en uitgevoerd. Maar uiteindelijk gaat het niet alleen om het vormgeven van succesvolle lessen, we willen zicht krijgen op essentiële kenmerken van het ontwikkelen van begrip van bewijs. Daarvoor werden zowel de producten die de leerlingen produceren als de weg die naar die producten leidt nauwgezet onderzocht in de beide cycli. In de hoofdstukken 5 en 7 wordt dat beschreven. De eerste cyclus van ontwerpen en onderzoeken leidde tot aanscherping van de theoretische uitgangspunten en de manier om de ontwikkeling van begrip van bewijs te stimuleren en te volgen. Die aanscherping wordt beschreven in hoofdstuk 6.

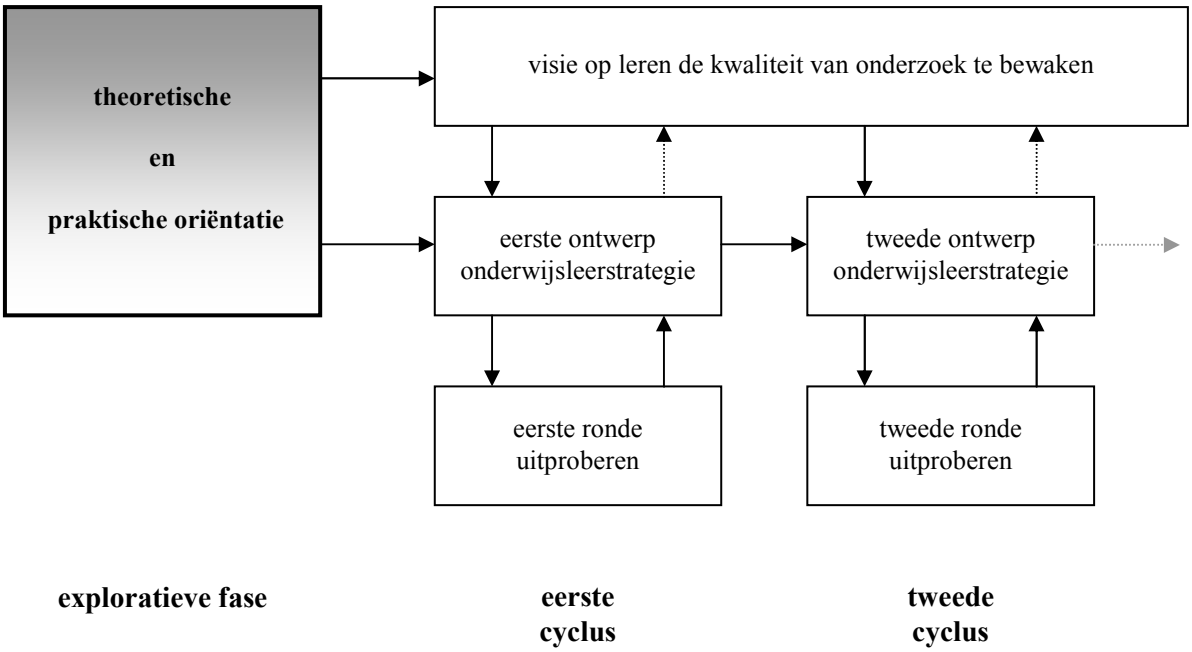
(4) Wat zijn kenmerken van een uitvoerbare en effectieve onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek?

De aanvankelijke visie op het ontwikkelen van begrip van bewijs leidt met inachtneming van een aantal praktische randvoorwaarden tot een concreet ontwerp voor een onderwijsleerstrategie, in de vorm van een praktische opdracht voor 5 vwo. Dat ontwerp omvat zowel materialen voor leerlingen en docenten als uitgewerkte scenario's waarin de beoogde gang van zaken geschetst wordt. Om tot een empirische gefundeerd antwoord op de vierde deelvraag te kunnen komen, moest dat ontwerp uitgeprobeerd en geëvalueerd worden. Kortom er moest bepaald worden

- *in welke mate de praktische opdracht uitvoerbaar is voor docent en leerlingen;*
- *in welke mate er leereffect te constateren is met betrekking tot begrip van bewijs*.

Aan het eind van hoofdstuk 5 wordt op een rijtje gezet waarin het tweede ontwerp beter moest zijn dan het eerste. Het bijgestelde ontwerp wordt beschreven in hoofdstuk 7. Een uitgebreide analyse van de uitvoerbaarheid en effectiviteit van het tweede ontwerp leidt enerzijds tot aanwijzingen voor een eventuele derde versie, maar anderzijds tot een meer algemeen antwoord op de vierde deelvraag: wat zijn kenmerken van een succesvolle onderwijsleerstrategie? In het laatste hoofdstuk wordt de balans opgemaakt met het geven van een antwoord op de centrale onderzoeksvraag: *Hoe kunnen ze het leren?*

Onderzoek doen is gericht op het vinden van een antwoord op een vraag. De aard van het onderwerp van dit onderzoek – de kwaliteit van onderzoek – maakt de vraag 'hoe zeker is dat antwoord eigenlijk?' onontkoombaar. En niet altijd is het gevonden antwoord afdoende en meestal roept zo'n antwoord ook nieuwe vragen op. Zo is het ook met dit onderzoek. In het laatste hoofdstuk wordt de hoofdvraag beantwoord en het antwoord gewogen. Ook reflecteren we op zowel de theoretische als de praktische uitgangspunten van dit onderzoek en doen we suggesties voor onderzoek om onderwijs rond van de kwaliteit van onderzoek verder te ontwikkelen. Dat resulteert in bespiegelingen over de vragen: 'wat is er nodig om het beter te doen?' en 'mikken we niet te hoog?'



2. Kwaliteit van onderzoek

2.1 Wetenschappelijk onderzoek
2.2 Biologisch onderzoek
2.3 Kwaliteit vraagt 'begrip van bewijs'

'Carrying out a scientific investigation, then, is primarily a display of understanding, and not of skill.'
Robin Millar, Fred Lubben, Richard Gott & Sandra Duggan (1994, p. 246)

Hoofdstuk 2 Kwaliteit van onderzoek

Een proefschrift met de vraag hoe leerlingen de kwaliteit van onderzoek kunnen leren bewaken, kan er niet omheen dat begrip inhoudelijk uit te werken. Dat gebeurt in dit hoofdstuk. Daarbij uitgaan van de vraag 'zeker weten?' zou kunnen leiden tot een kennistheoretisch verhaal over wat 'weten' is en wanneer je daar zeker van kunt zijn. Dat is niet de bedoeling, want dat voert veel te ver voor het kader waarin we nu over kwaliteit van onderzoek spreken: het vak biologie in het voortgezet onderwijs. Wat wel zinvol is, is om te kijken wat wetenschappelijk onderzoek inhoudt, in het bijzonder in de biologie, en hoe je de kwaliteit ervan kunt beoordelen en bewaken.

In paragraaf 2.1 schetsen we kort onze visie op wetenschap en wetenschappelijk onderzoek. Wat maakt een activiteit tot 'wetenschap' en is er een wetenschappelijke methode? In paragraaf 2.2 geven we een typering van biologisch onderzoek, ook bedoeld om handvatten te hebben om de kwaliteit ervan te kunnen beoordelen. In paragraaf 2.3 wordt die kwaliteit verder uitgewerkt in 23 criteria, 'elementen van begrip van bewijs'.

2.1 Wetenschappelijk onderzoek

Wat onder wetenschap verstaan wordt, is in de loop van de geschiedenis aan verandering onderhevig geweest. In zijn boek *De ontdekking van het weten* laat Kwa (2005) zien dat er sinds de Griekse oudheid verschillende wetenschappelijke 'stijlen' zijn ontstaan, zoals de deductieve, de experimentele, de hypothetische, de statistische en de evolutionaire stijl.[3] Deze stijlen sluiten elkaar niet uit, noch is er sprake is van een 'evolutie' waarbij een recentere stijl 'beter' is dan een voorgaande. Vaak is er zelfs sprake van combinaties van stijlen, bijvoorbeeld de experimentele stijl en de statistische. Het bestaan van verschillende wetenschappelijke stijlen laat zien dat er verschillende manieren zijn om kennis over de werkelijkheid te vergaren, elk met hun (voor een deel) eigen geaccepteerde methoden. Het onderstreept de door velen gedane constatering dat er niet één universele wetenschappelijke methode is (Chalmers, 1999). Bauer (1992) merkt terecht op dat als er een methode zou bestaan die onomstotelijk tot ware kennis zou leiden, discussie in de wetenschappelijke gemeenschap overbodig zou zijn. Het tegendeel is echter het geval en hij presenteert dan ook het beeld van het 'kennisfilter' waarin objectieve en betrouwbare kennis uitgefilterd wordt uit het geheel van geproduceerde kennis door onder andere peer review en gebruik door anderen. Dit laat ook goed zien dat wetenschap een menselijke onderneming is, een praktijk waarin veel activiteiten ondernomen worden om de wetenschap verder te helpen.

Vaak wordt het niet bestaan van een universele methode geïllustreerd met de uitspraak van Feyerabend (1975) 'anything goes' en wordt daaraan de conclusie verbonden dat elke aanpak even bruikbaar is. Chalmers (1999, p. 189) weerspreekt dit:

[3] Kwa leunt zich hierbij sterk op het werk *Styles of scientific thinking in the European tradition* van Crombie (1994).

> We hebben gezien dat de argumenten tegen de methode voor Feyerabend aanleiding zijn aan te nemen dat er geen methode bestaat, dat wetenschappelijk onderzoekers gehoor moeten geven aan hun eigen subjectieve wensen en dat alles mogelijk is. Maar de mogelijkheden zijn niet uitgeput met een universele methode of helemaal geen methode. Een tussenweg zou inhouden dat er in de wetenschap methoden en standaarden bestaan, maar dat deze van wetenschap tot wetenschap kunnen variëren en, binnen een wetenschap, kunnen worden veranderd en dat ook ten goede.

Hij noemt als mogelijkheid onder andere de stroming van het nieuwe experimentalisme (o.a. Ackermann, 1989; Mayo, 1996), die een hernieuwd pleidooi voeren voor de kracht van het experiment om theorieën te toetsen. Maar daar zijn wel eisen aan verbonden (Chalmers, 1999, p. 228):

> Een experiment[4] bevestigt een uitspraak alleen als mogelijke bronnen van fouten zijn uitgesloten en de uitspraak de toets dus waarschijnlijk niet zou doorstaan, tenzij ze juist is.

Niet één onveranderlijke, universele methode dus, maar wel onderscheid tussen wetenschap en niet-wetenschap. Maar wat maakt een activiteit dan tot wetenschap? Daar worden uiteraard ook verschillende antwoorden op gegeven, maar die gaan wel meestal in dezelfde richting. In een methodologieboek voor biologiestudenten geven Allen en Baker (2001) het volgende rijtje als karakteristiek voor (natuur)wetenschap: *(1) empirical knowledge, (2) rationality, (3) repeatability, (4) testability, (5) experimentation, (6) generality.* De American Association for the Advancement of Science (AAAS, 1990, p. 4-7) noemt: *(1) science demands evidence; (2) science is a blend of logic and imagination; (3) science explains and predicts; (4) scientists try to identify and avoid bias; (5) science is not authoritarian.* Wat hieruit naar voren komt is enerzijds de nadruk op empirie en logica en anderzijds het verklarende karakter. Natuurwetenschap is op zoek naar begrip van de materiële wereld en verzamelt daarvoor niet alleen waarnemingen, maar probeert daarmee ook logische, samenhangende verklaringen – theorieën – te geven voor verschijnselen. Dat geldt ook voor biologisch onderzoek, het onderwerp van de volgende paragraaf.

2.2 Biologisch onderzoek

> Biologisch onderzoek is (natuur-)wetenschappelijk onderzoek aan objecten uit de levende natuur, omvattende bouw en functie van organismen, hun ontstaanswijze en de relaties tussen organismen en hun levende (biotische) en dode (a-biotische) omgeving' (KNAW, 1997, p. 11).

Het beschrijven van diersoorten, het uitzoeken hoe hartslagfrequentie afhankelijk is van inspanning en het ophelderen van de ontstaansgeschiedenis van mensen, het is dus allemaal biologisch onderzoek. Biologische kennis komt tot stand door het verzamelen van gegevens over de levende natuur en het analyseren en interpreteren daarvan. Door observeren, experimenteren, interpreteren en theorievorming vindt groei van de biologie als 'body of knowledge' plaats.

[4] 'Experiment' moet hier trouwens in niet al te strikte zin opgevat worden. Het voorbeeld dat Mayo (1996) uitwerkt, betreft de waarneming van de afbuiging van passerende lichtstralen door de zon tijdens een zonsverduistering als toets voor Einsteins wet voor de zwaartekracht, die voortvloeit uit de algemene relativiteitstheorie. Geen gecontroleerd experiment in een laboratorium dus.

Biologisch onderzoek omvat alle activiteiten die bijdragen aan die groei. Die activiteiten kunnen op verschillende 'niveaus'[5] plaatsvinden. Van het nauwkeurig beschrijven van de ontwikkeling van een kiemplant uit een zaad, via het systematisch in kaart brengen van de aantallen uitgegroeide kiemplanten in verschillende bossen en het bepalen van de optimale groeicondities in een laboratorium tot het formuleren van een nieuwe theorie over de ontwikkeling van planten. Als een arts erachter probeert te komen wat haar patiënt mankeert of als een controleur wil bepalen of de concentratie bacteriën in het drinkwater binnen een vastgestelde norm valt, wordt ook vaak het werkwoord 'onderzoeken' gebruikt. Toch is dat ons inziens geen wetenschappelijk biologisch onderzoek. Gelukkig wordt er in die gevallen meestal wel op een systematische en wetenschappelijk verantwoorde manier gewerkt, maar het levert geen nieuwe kennis of inzicht op.

Biologisch onderzoek is divers. Niet alleen omdat de natuur zo divers is, maar ook omdat vraagstellingen en manieren om daarop een antwoord te vinden ontzettend van elkaar kunnen verschillen. Het is dan ook niet verwonderlijk dat er veel manieren bedacht zijn om biologie en biologisch onderzoek te categoriseren. De huidige indelingen van onderzoeksgroepen aan universiteiten laten sporen zien van indelingen naar studieobject (plantenecologie), naar organisatieniveau (celfysiologie, populatiegenetica) naar vraagstelling (ecotoxicologie) en werkwijze (theoretische biologie).

Ook onderzoek kan op veel manieren ingedeeld worden (Stokking & Van der Schaaf, 2000):

- aan de hand van de grondvorm, het design (bijv. experiment, survey, case study)
- aan de hand van de aard van de vraagstelling (bijv. beschrijven, toetsen van hypothese, zoeken naar relaties, bepalen van een constante, ontwikkelen van een methode)
- aan de hand van de situatie waarin het onderzoek wordt uitgevoerd (bijv. laboratorium, veld, archief)
- aan de hand van de methode van gegevensverzameling (bijv. meten, observeren, interviewen, literatuur).

Sommige auteurs (o.a. Mayr, 1997) delen onderzoek in aan de hand van vraag*woorden*: wat, hoe en waarom. Doordat je een wat-vraag echter vaak kan herformuleren in een hoe-vraag en andersom, is zo'n indeling ons inziens niet houdbaar.

Hoewel de laatste twee indelingen van Stokking en Van der Schaaf (2000) misschien meer aansluiten bij de praktische invalshoek van waaruit een docent tegen onderzoek van zijn leerlingen aankijkt, prefereren we toch het uitgangspunt dat onderzoek doen vooral te maken heeft met een bepaalde manier van denken en redeneren, wat de eerste twee indelingen kenmerkt. Onderzoek doen is gericht op het vinden van een antwoord op een vraag, stelden we in het eerste hoofdstuk al. Dat moet ons inziens richtinggevend zijn en daarom kiezen we voor een indeling aan de hand van de vraagstelling. Dat sluit ook aan bij didactische benaderingen om bij onderzoek vooral van vragen (van leerlingen of uit andere bron) uit te gaan (Barnard, Gilbert & McGregor 1993; Chin, 2004; Keys, 1998; Marbach-Ad & Claassen, 2001; Windschitl & Buttemer, 2000). We beperken ons daarbij tot drie categorieën:

- beschrijvend onderzoek, vanwege de constatering die Mayr (1997, p. 113) doet: 'Description is the first step in any branch of biology,' en de opvatting die hij daarover verwoordt: 'Description should never be maligned, because it the indispensable foundation of all explanatory and interpretive research in biology.'
- hypothesetoetsend onderzoek, want: 'The formulation and testing of hypotheses lie at the very heart of any scientific or rational inquiry' (Allen & Baker, 2001, p. 40).

[5] 'Niveau' verwijst hier niet naar de organisatieniveaus in de biologie (molecuul, cel, organisme, enz.), maar naar een verschillende mate van aggregatie van gegevens en interpretatie ervan.

- ontwerpend onderzoek, omdat: 'Anthony van Leeuwenhoek's drawings of the 'animalcules' his microscope revealed in a drop of pondwater, publicized in the 1670s, revolutionized biology to an even greater degree than Galileo achieved for cosmology when he turned his watching device, the telescope, to the heavens and observed the moons of Jupiter' (Rose, 1998, p. 58).

In een groot onderzoek zijn deze typen vaak alle vertegenwoordigd en lopen ze soms in elkaar over. Leerlingonderzoek kan zich beperken tot een van die fasen en daarom is het onderscheid zinvol. We werken de drie typen wat verder uit.

- *Beschrijvend onderzoek*

Beschrijvend onderzoek richt zich op het in kaart brengen een deel van de (biologische) werkelijkheid, zoals het uiterlijk van een plantensoort, de soortensamenstelling van een sloot of de basenvolgorde van het menselijk genoom. Een mooi voorbeeld van een beschrijvend leerlingonderzoek is dat van Piers (2000) waarin hij de broedvogelpopulaties in een polder in kaart brengt.

Bij beschrijvend onderzoek ligt het accent op goed waarnemen met de zintuigen of een instrument, zoals een microscoop of verrekijker en het zodanig weergeven van het waargenomene dat het voor anderen herkenbaar is. Dat weergeven kan gebeuren met o.a. woorden, tekeningen en foto's. In de biologie komt deze vorm van onderzoek veel voor in bijvoorbeeld de systematiek, de ethologie en de ecologie. Ook bepaalde vormen van meer maatschappelijk gericht onderzoek, bijvoorbeeld onderzoek naar de verspreiding van het HIV-virus onder de bevolking van Nederland, vallen in deze categorie. Leerlingonderzoek heeft soms ook deze vorm, zoals het inventariseren van broedvogelpopulaties in het park bij school of het enquêteren van docenten met betrekking tot hun rookgedrag.

Dat beschrijvend onderzoek niet van vooronderstellingen zou uitgaan, is een misverstand dat nog steeds bestaat. Elke waarneming is theoriegeladen, men neemt waar wat in het eigen conceptuele kader past en het is vaak heel moeilijk daarbuiten te treden. Bovendien is het aantal mogelijke waarnemingen zo groot dat men van tevoren een selectie zal moeten maken waarop men zal letten en waarop niet. Ook vereist beschrijving een systeem dat het geraamte vormt om de waarnemingen te ordenen. Als een onderzoeker de aantallen broedparen van zangvogels in een bepaald gebied in kaart wil brengen, zal hij moeten weten dan wel beslissen welke waarneming nodig is om te kunnen concluderen dat een broedpaar aanwezig is, wat zangvogels zijn e.d. Rose (1998, p.24-25) geeft hier een aardig voorbeeld van:

> Suppose I am interested in how animals behave, and how that behaviour changes as they grow from infancy to adulthood. I may be watching a family group of marmosets, a pride of lions, a nest of hatching blue tits and their parents. I want to record how any of these spend their time during a day, a week, a month, a year. But I can't watch them continuously over the entire period, even if I were to mount video cameras and record every aspect of their activity. There would simply be to much data to analyse. (…)
> My video records patterns of continuous changing activity. But to make sense of it I need to classify, to distinguish the different types of activity I observe. (…)
> Which of the interactions (…) counts as play – or is this not a meaningful category at all? If, during the first weeks of life, there is an increase in the proportion of the time spent playing, is this a 'real' change, or is it an artefact resulting from the fact that the amount of time the infant spends asleep decreases, so the other recorded activities simply expand to fill up the sample time? The problem of separating object from field, of determining which is the 'correct' interpretation of the ambiguous figure, is not confined to the psychologist's abstractions but is the everyday stuff from which science has to be built.

- *Hypothesetoetsend onderzoek*

Dit onderzoek wordt ook wel experimenteel onderzoek genoemd, maar dat vinden wij geen juiste benaming. Er bestaat onderzoek dat geen experimenteel onderzoek is, maar toch een hypothese toetst (veel biologisch veldonderzoek) en er bestaat experimenteel onderzoek dat geen hypothese toetst. Rose (1998) noemt interveniëren als kenmerk voor een experiment, terwijl dat best beschrijvend van aard kan zijn (wat gebeurt er als ik …). Wat hypothesetoetsend onderzoek kenmerkt is de hypothetisch-deductieve aanpak.

Er wordt een uitspraak gedaan over een mogelijk verband tussen verschijnselen en een mogelijke verklaring van een verschijnsel. Dat wordt een *hypothese* genoemd. Deze is meestal gebaseerd op eerdere waarnemingen en/of op logische consequenties van andere verklaringen c.q. theorieën. Bijvoorbeeld de hypothese dat kieviten in de herfst naar het zuiden trekken als gevolg van het opraken van hun voedselbronnen. Vervolgens wordt een manier bedacht om de hypothese te toetsen, om uit te zoeken of hij juist is of niet. Er worden logische consequenties uit de hypothese afgeleid (gededuceerd): als de hypothese waar is en ik dát doe (en aan een aantal randvoorwaarden is voldaan), dan zal ik die en die resultaten uit een experiment moeten krijgen of die en die waarnemingen moeten kunnen doen. Dit wordt vaak de *voorspelling* genoemd, bijvoorbeeld 'als ik zorg voor extra voedsel, zullen de kieviten niet naar het zuiden gaan.' Om de juistheid van de voorspelling te beoordelen worden er gegevens verzameld. Dat kan op veel verschillende manieren gebeuren: in ons voorbeeld door gericht te observeren, maar het kan ook door gecontroleerde experimenten uit te voeren, interviews te houden, literatuur door te vlooien, enzovoort.

Als de voorspelling niet door de gegevens gestaafd wordt, bijvoorbeeld doordat de kieviten wel vertrekken, wordt de hypothese verworpen (aannemende dat de methode van gegevensverzameling goed genoeg was: was het wel het juiste voedsel?). Als de voorspelling wel door de gegevens gestaafd wordt, wordt de hypothese niet verworpen. Onderzoekers hechten echter vaak aan hun hypothese. Het zomaar door waarnemingen of experimenten laten weerleggen is soms moeilijk te accepteren. De voorwaarde die hierboven tussen haakjes staat, wordt dan ook vaak betwist. Enerzijds is dat positief, want elk onderzoek moet kritisch bekeken worden, anderzijds worden er ook eindeloos ad hoc redeneringen verzonnen om hypotheses staande te houden.

Vaak is het mogelijk op grond van waarnemingen verschillende hypotheses op te stellen of, omgekeerd geformuleerd, blijven verschillende hypotheses overeind bij een bepaalde verzameling waarnemingen. Zo kunnen de hypotheses opgesteld worden dat de kieviten aangezet worden tot trekken door de kortere daglengte of de dalende temperatuur. Voor hypothesetoetsend onderzoek is het niet per definitie nodig *nieuwe* gegevens te verzamelen. Het kan ook gaan om herschikking en nieuwe interpretatie van oude gegevens, bijvoorbeeld vertrekdata van kieviten uit eerdere jaren.

- *Ontwerpend onderzoek*

De wetenschap zou niet zover gekomen zijn als er niet allerlei materialen, instrumenten en procedures waren ontworpen en ontwikkeld. Denk bijvoorbeeld aan de Skinnerbox (een experimenteeromgeving voor het doen van gedragsonderzoek bij ratten e.d.) of aan allerlei kleurtechnieken van microscopische preparaten. In essentie dus ondersteunend voor andere soorten onderzoek, maar zo belangrijk dat er zelfs Nobelprijzen voor worden uitgeloofd (K.B. Mullis kreeg in 1993 de Nobelprijs voor scheikunde voor de polymerase kettingreactie zonder welke de biotechnologie niet zo'n hoge vlucht had kunnen nemen).

In dit type onderzoek ligt dus het accent op het ontwikkelen of optimaliseren van materialen en methodes. Dan kan weer op een heleboel manieren, van het concretiseren van een volledig nieuw idee tot het systematisch uitproberen van concentraties werkzame stof, bij-

voorbeeld een medicijn. Of bijvoorbeeld het selecteren van de meest geschikte organismen voor een bepaald doel (voedselproductie, plantenveredeling).

Met ontwerpend onderzoek wordt hier niet gedoeld op toegepast onderzoek – dan wel 'strategisch' of 'marktgericht' onderzoek (KNAW, 2001) – hoewel er wel vaak wetenschappelijke kennis voor nodig is. Ontwerpend onderzoek kan wel toegepast zijn, maar net zo goed fundamenteel. Dat is het geval als de vraag naar een ontwerp niet voortkomt uit een maatschappelijke behoefte, maar uit een vraag uit de wetenschap (zoals bovengenoemde voorbeelden). De ontworpen producten stimuleren dan weer de verdere (fundamenteel) wetenschappelijke ontwikkeling.

- *Elk een eigen methode?*

Zoals in paragraaf 2.1 al is verwoord, één universele natuurwetenschappelijke methode bestaat niet. Evenmin is het zo dat voor elk van de drie hierboven onderscheiden typen onderzoek een vaste methode bestaat. Toch valt er wel degelijk iets te zeggen over wanneer zulk onderzoek gedegen is uitgevoerd en wanneer niet, wanneer je meer vertrouwen mag hebben in de resultaten en wanneer minder. De vijf kenmerken van wetenschappelijk onderzoek zoals genoemd aan het eind van paragraaf 2.1 (AAAS, 1990), vormen daarvoor een goed handvat, maar we zetten ze in een andere volgorde en voegen er een aan toe:

- *wetenschap is niet eigenmachtig*, dat wil zeggen dat er geen conclusies van tevoren vastliggen omdat dat door andere beoefenaren van de wetenschap zo is vastgelegd; onderzoek dat ongefundeerde gegevenheden als uitgangspunt neemt, verdient wantrouwen;
- *wetenschap is een mengsel van verbeelding en logica*, daarmee wordt bedoeld dat vooruitgang in de wetenschap nooit 'vanzelf' uit de gegevens of analyses daarvan tevoorschijn komt; maar het betekent tevens dat, hoe creatief een wetenschapper ook mogelijke verbanden bedenkt, hypotheses vroeg of laat toch aan de principes van de logica getoetst moeten worden;
- *wetenschap verklaart en voorspelt*, dat houdt in dat de geloofwaardigheid van wetenschappelijke theorieën berust op hun vermogen om verbanden tussen verschijnselen te verklaren en te voorspellen; het wezen van wetenschap is 'validation by observation' (AAAS, 1990, p. 6), bevestiging door waarneming, maar ook het kunnen voorspellen van nog niet gedane waarnemingen;
- *wetenschap probeert vooringenomenheid (bias) te identificeren en te voorkomen:* hoewel hierboven betoogd is dat waarnemingen en hypotheses onvermijdelijk in een bepaalde mate theoriegeladen zijn, is het van belang om dat te onderkennen en zoveel mogelijk te beperken en om het gehanteerde theoretisch kader te expliciteren;
- *wetenschap vraagt 'empirical evidence'*, geldig en betrouwbaar bewijsmateriaal, want vroeg of laat wordt de geldigheid van wetenschappelijke beweringen bevestigd of verworpen door empirische gegevens; daarom is het belangrijk om nauwkeurige en betrouwbare gegevens te verkrijgen;
- *wetenschap is een sociaal proces van kennisconstructie* (Bauer, 1992; Chinn & Malhotra, 2002): wetenschap komt in een gemeenschap van onderzoekers volgens de regels van die gemeenschap tot stand; dat betekent dat wetenschappelijk onderzoek voortbouwt op eerder onderzoek en dat de resultaten ook in die gemeenschap gewogen worden.

In een parafrase op het eerdere geciteerde fragment uit Chalmers (1999, p. 228) zou je de kwaliteit van een onderzoek als volgt kunnen omschrijven: een onderzoek bevestigt een uitspraak alleen als mogelijke bronnen van fouten zijn uitgesloten en de uitspraak het onder-

zoek dus waarschijnlijk niet zou doorstaan, tenzij ze juist is. Daarvoor is ons inziens 'begrip van bewijs' nodig, hetgeen we uitleggen in de volgende paragraaf.

2.3 Kwaliteit vraagt 'begrip van bewijs'

'Begrip van bewijs' is onze vertaling van *concepts of evidence.* Dat behoeft enige uitleg, want vertalen betekent interpreteren en keuzes maken. Het Van Dale woordenboek (Hannay, 1988) geeft als vertaling van *evidence* 'bewijs', in de betekenis van bewijsmateriaal of bewijssstuk, *concept* wordt vertaald met 'begrip'. 'Begrippen van bewijsvoering', is dus wellicht de meest adequate vertaling voor de lijst *concepts of evidence.* Maar behalve dat dat niet lekker in het gehoor ligt, mist het een connotatie die 'begrip van bewijs' wel heeft, die van 'begrijpen'. Want onderzoek doen is primair een kwestie van begrijpen, zoals ook het citaat van Millar e.a. (1994) aan het begin van dit hoofdstuk verwoordt. *Concepts of evidence* moeten leiden tot *understanding of empirical evidence,* is in het PACKS-model (fig. 1.1) weergegeven. 'Bewijs' heeft wel het risico in zich dat het opgevat wordt als *proof,* zoals een sluitend wiskundig bewijs. Dat dat in de natuurwetenschappen niet mogelijk is, vatten Allen en Baker (2001, p. 48) samen met 'Indeed, the word 'proof' should not be used in the context of science at all.' Met de term 'begrip van bewijs' bedoelen we dus het begrijpen van (c.q. inzicht in) de begrippen van empirische bewijsvoering.

In hoofdstuk 1 werd gesteld dat het uitwerken van de *concepts of evidence* een begaanbare weg lijkt ten behoeve van het leren bewaken van de kwaliteit van onderzoek. Daar werd ook gesteld dat een uitwerking ervan voor het Nederlandse biologieonderwijs nodig is. De eerste deelvraag van dit proefschrift is daarop gericht. Hieronder wordt die uitwerking gegeven. Die is grotendeels gebaseerd op de gedetailleerde uitwerkingen van Gott en medewerkers (Gott & Duggan, 1995a, 1996, 2003; Gott e.a.,2006; Roberts, 2001). Maar de 107 begrippen die in de meest recente versie staan, zijn enerzijds veel te uitgebreid voor ons doel, terwijl ze anderzijds te beperkt zijn. Zo omvatten ze wel de opzet van onderzoek en de kritische evaluatie van verzamelde gegevens, maar wordt er over vraagstelling niets anders gezegd dan dat je de juiste variabelen moet kiezen. Het woord 'hypothese' komt er niet in voor, noch hoe je het toetsen daarvan aanpakt! Dat lijkt een gevolg van de nogal natuurkundige insteek die gekozen is, al ontbreekt dit aspect ook in een artikel waarin Roberts (2001) zich speciaal op de biologie richt. Een aanvulling is dus nodig. Daarvoor wordt geput uit andere bronnen, o.a. *Algemene methodologie voor biologen* (Van der Steen, 1982) en *Biology, scientific process and social issues* (Allen & Baker, 2001).

Een aspect waar Gott en medewerkers uitgebreid aandacht aan besteden, het weergeven van waarnemingen en metingen in tabellen en grafieken, hebben wij niet opgenomen. Wij vinden dit namelijk geen primair aspect van de kwaliteit van onderzoek. Weliswaar helpt een overzichtelijke en inzichtelijke presentatie van gegevens bij de bewijsvoering, maar het raakt niet de essentie ervan. Allen en Baker (2001), bijvoorbeeld, vermelden het wel, maar in een appendix.

In het PACKS-model (fig. 1.1) is te zien dat begrip van bewijs een rol speelt in elke fase van een onderzoek. De hieronder volgende beschrijving van de elementen van begrip van bewijs is anders geordend, namelijk met betrekking tot achtereenvolgens de onderzoeksvraag, hypothese, onderzoeksopzet & variabelen, metingen & observaties, conclusie & verklaring en evaluatie. Dit lijkt een volgorde te suggereren waarin ze aan de orde (moeten) komen, maar zo is het niet bedoeld. Weliswaar komen ze in een onderzoeksverslag of artikel vaak in deze

volgorde voor, in de praktijk van onderzoek doen is er geen vaste volgorde en wordt bijvoorbeeld na het verzamelen van de gegevens de onderzoeksvraag nog aangepast.

Een conceptversie van de elementen van begrip van bewijs (Schalk, 2001) is voorgelegd aan een hoogleraar theoretische biologie, zes ervaren biologiedocenten en vier onderzoekers op het gebied van de didactiek van de natuurwetenschappen. De daaruit resulterende 23 elementen vormen ons inziens een overzichtelijke lijst van criteria voor kwaliteit van onderzoek, niet allemaal op elk onderzoek van toepassing, maar wel een voldoende compleet gereedschap om de kwaliteit van elk (biologisch) onderzoek ermee te beoordelen. De elementen, die hieronder geformuleerd en toegelicht worden, zijn ons inziens geschikt om te gebruiken in het biologieonderwijs, al is het taalgebruik is daar niet op afgestemd; het is geen leerlingtaal.

- *Elementen van begrip van bewijs m.b.t. onderzoeksvraag*

1. Een onderzoeksvraag:
   - bevat alleen ondubbelzinnige termen en formuleringen;
   - is voldoende specifiek en afgeperkt.
2. De beantwoording van de onderzoeksvraag vereist een beschrijving en/of de toetsing van een hypothese.

Elk onderzoek kent een onderzoeksvraag. Deze komt voort uit een waarneming, theoretische exercitie of uit eerder onderzoek, bijvoorbeeld uit de waarnemingen in bijgaand kader. De eerste vraag die bij menig bioloog zal opkomen na het lezen van zo'n serie waarnemingen is 'hoe doen ze dat?' Maar als je die verwondering over het precieze samenspel wilt omzetten in een onderzoeksvraag moet je duidelijker zijn over hetgeen je wilt uitzoeken. Dan krijg je zoiets als: 'Hoe bepalen koolmezen het tijdstip van eileg zo dat de rupsenpiek en de periode waarin hun jongen het meeste voedsel nodig hebben, samenvallen?' of een al wat meer toegespitste zoals: 'Welke omgevingsfactor bepaalt het begin van de eileg bij koolmezen?' Een andere verwondering is ook mogelijk: 'Hoe beïnvloeden tanninen de embryonale ontwikkeling van de wintervlinder(rups)?'

> Koolmezen in een bos met veel eiken voeren hun jongen met rupsen van de kleine wintervlinder. Deze rupsen komen massaal uit als de bladeren van de eiken nog erg jong zijn en nog geen tanninen (groeiremmende stoffen) bevatten. Omdat de wintervlinderrups snel groeit en verpopt, is er sprake van een rupsenpiek. Als het vroeg in het jaar warm is, komen de eiken eerder in blad en is de rupsenpiek ook vroeger.
>
> Koolmezen blijken hun eieren op een zodanig tijdstip te leggen dat de rupsenpiek gelijk valt met de periode wanneer hun jongen het meeste voedsel nodig hebben. Er kan soms wel een maand liggen tussen het leggen van het eerste ei en de periode waarin de jongen hun maximale behoefte aan voedsel hebben.
>
> *Bron: Tinbergen & Verhulst, 2000*

Zo'n vraagstelling is ondubbelzinnig en afgeperkt, maar ook nog enigszins open. Vaak wordt echter een onderzoeksvraag geformuleerd waarin eigenlijk al een hypothese verborgen zit, zoals: 'Is het tijdstip van eieren leggen van koolmezen afhankelijk van de temperatuur?' Ook deze vraag is ondubbelzinnig en afgeperkt, maar vraagt eigenlijk alleen nog maar om bevestiging of ontkenning. Daarmee doet hij afbreuk aan het proces van onderzoek doen, omdat het onderscheid van verschillende denkstappen verdwijnt. Een onderzoeksvraag beschrijft een probleem waarvoor een antwoord in de vorm van een beschrijving of verklaring gezocht wordt, het is meestal niet een vraag die alleen om bevestiging of ontkenning vraagt, dat is namelijk de functie van een hypothese.

Als validiteit van een onderzoek opgevat wordt als de mate waarin het bewijsmateriaal een antwoord verschaft op de vraag (Gott & Duggan, 2003), is het van belang een heldere, niet mis te verstane vraag te stellen. Daarom mogen gebruikte termen en formuleringen niet

voor meer dan één uitleg vatbaar zijn, ze moeten *eenduidig* zijn. Daarnaast is het van belang voldoende specifiek te zijn en de vraagstelling af te perken. De vraag hoe vogels het tijdstip van eileg bepalen is wel interessant, maar niet in z'n algemeenheid onderzoekbaar. Dat wordt het pas als het wordt gespecificeerd dat het om koolmezen gaat en wordt beperkt tot de situatie in eikenbossen waar wintervlinders voorkomen. Vaak verandert de precieze formulering van een onderzoeksvraag in de loop van het onderzoek, onder andere doordat termen nadere specificatie behoeven of verdere afperking nodig is.

Ook is het uiteraard van belang om te weten naar wat voor type antwoord je zoekt. Gaat het om de beschrijving van een verschijnsel of situatie (het aantal koolmeesjongen dat groot wordt en het aantal rupsen in de loop van de tijd) of om een verklaring (hoe die twee samenhangen), waartoe een hypothese getoetst wordt?

- *Elementen van begrip van bewijs m.b.t. hypothese*

3. Bij hypothesetoetsend onderzoek wordt er een hypothese (een mogelijke verklaring of voorlopig antwoord) opgesteld die bij de onderzoeksvraag aansluit.
4. De hypothese moet toetsbaar zijn.
5. (Bijna) altijd is er meer dan één hypothese mogelijk.
6. Op basis van de hypothese (en een aantal aannames) wordt een voorspelling geformuleerd over welke observatie(s) of meting(en) verwacht kunnen worden (als de hypothese waar is [en …], dan …); bij beschrijvend onderzoek kan een verwachting van de resultaten geformuleerd worden.

Als er voor de beantwoording van de onderzoeksvraag een beschrijving nodig is, is het geven van een hypothese niet nodig. Maar als een verklaring gezocht wordt is dat wel het geval. Een hypothese is een mogelijk antwoord op de onderzoeksvraag: het zou wel eens kunnen dat … Een slimme hypothese komt niet uit de lucht vallen. Meestal is op basis van wat al bekend is een aannemelijke hypothese op te stellen.

'Hypotheses are tentative explanations to account for observed phenomena' (Allen & Baker, 2001, p. 40), dus een uitspraak over hoe een verschijnsel (organisme, systeem) mogelijkerwijs in elkaar zit of werkt. In een onderbouwing van die hypothese kan aangegeven worden waardoor die hypothese aannemelijk is.[6]

Vaak zijn verschillende hypotheses mogelijk die elkaar al dan niet uitsluiten. Bijvoorbeeld de hypotheses dat koolmezen het tijdstip van hun eileg laten afhangen van de omgevingstemperatuur dan wel van de aanwezigheid van ander voedsel dan wintervlinderrupsen (die er dan als het ware een voorspeller van zijn). Maar misschien reageren koolmezen wel op beide factoren.

Een hypothese moet toetsbaar zijn, dat wil zeggen dat er een manier moet zijn om met empirische gegevens en/of logica de hypothese aan een test te onderwerpen. Maar de hypothese in een hypothesetoetsend onderzoek moet niet verward worden met de nul-hypothese of de alternatieve hypothese zoals die in de statistiek gebruikt worden (McPherson, 2001). Die laatste twee moeten altijd volledig complementair zijn om bij het verwerpen van de een de ander te kunnen accepteren, terwijl in biologische kwesties bijna altijd alternatieve, nog niet onder woorden gebrachte hypotheses denkbaar zijn.

[6] Daarmee nemen we een ander standpunt in dan anderen in het biologieonderwijs, die vinden dat een hypothese geen verklaring mag bevatten (zie de discussie in De Hullu, Van der Wulp en Sloep, 1999; Pohlmann, Dietvoorst, De Knecht & Lieverse,1999). Die discussie, waarvan de aanleiding gevormd werd door het antwoordmodel van een examenvraag, heeft er weliswaar toe geleid dat in examens de kwestie sindsdien omzeild is, maar de verwarring in het onderwijsveld is gebleven.

De verbindende schakel tussen de hypothese en de opzet van het onderzoek is de voorspelling: als de hypothese juist is dan moet ik de volgende waarnemingen kunnen doen; laat ik maar eens gaan kijken of dat klopt. Die voorspelling moet zo gedaan zijn, dat bij het *niet* optreden van de voorspelde uitkomst er logisch geconcludeerd kan worden dat de hypothese onjuist is. Daarentegen betekent het wel optreden van de voorspelde uitkomst niet dat de hypothese juist is, alleen dat hij niet verworpen kan worden. Vaak wordt een voorspelling als een overbodige stap gezien (ook door docenten, zie Schalk, 2002a). Dat is ook zo als de voorspelling neer komt op: ik voorspel dat de hypothese juist is. Als de voorspelling echter een logisch vervolg is op de hypothese, dan is hij richtinggevend: als ik denk dat de koolmezen hun eileg afstemmen aan de hand van omgevingsfactoren zoals de temperatuur, kan ik een voorspelling opstellen over moment van eileg in relatie met het temperatuurverloop in het voorjaar (in een warm voorjaar leggen ze eerder o.i.d.) en die met waarnemingen toetsen.

Merk op dat het ook voor een beschrijvend onderzoek van belang is een verwachting te hebben van de resultaten, al is dat iets anders dan een hypothese die al dan niet verworpen gaat worden.

- *Elementen van begrip van bewijs m.b.t. onderzoeksopzet en variabelen*

7. Voor beschrijvend onderzoek is het belangrijk de te beschrijven kenmerken (van organismen, ecosystemen e.d.) te benoemen.
8. Het is noodzakelijk de afhankelijke en de onafhankelijke variabele(n) te onderscheiden.
9. Het is noodzakelijk alle andere mogelijk beïnvloedende variabelen te expliciteren en zo mogelijk constant te houden.
10. Om de invloed van een variabele te kunnen beoordelen is een controle of een ‘blanco’ noodzakelijk.

De onderzoeksopzet (het design) betreft de daadwerkelijke uitvoering van het onderzoek. Hierin moeten, naast een lijst van benodigde materialen, een aantal zaken duidelijk gemaakt worden. In een beschrijvend onderzoek ten eerste de te observeren kenmerken: waar wordt naar gekeken en hoe worden eventuele categorieën gedefinieerd? (zie het citaat van Rose op pagina 19). Uiteraard moeten die zo gekozen worden dat er een antwoord mee kan worden gegeven op de onderzoeksvraag.

Gott en medewerkers (Gott & Duggan, 2003; Gott e.a., 2006) besteden heel veel aandacht aan de variabelen die in een onderzoek meespelen. Ze stellen zelfs dat alle onderzoeken deze onderliggende structuur hebben (Roberts, 2001). Met het onderscheid tussen categorische, geordende en continue variabelen delen ze onderzoek in in negen typen (zowel onafhankelijke als afhankelijke variabele categorisch; onafhankelijke variabele categorisch en afhankelijke geordend; enzovoort). Onderzoek waarbij beide variabelen continu zijn kennen ze dan de meeste waarde toe. Ook Allen en Baker (2001) voeren een pleidooi voor ‘experimenting’, omdat daarin variabelen onder controle kunnen worden gehouden. Een dergelijke voorkeur voor experimenteren en kwantificeren hebben wij voor het biologieonderwijs in Nederland niet. Vandaar dat we typen variabelen en een daarop gebaseerde indeling van typen onderzoek niet opgenomen hebben in onze elementen van begrip van bewijs.

Het constant houden van andere variabelen dan de onderzochte gaat in laboratoriumomstandigheden vaak goed, maar in veldstudies is het lastiger. Biologische verschijnselen zijn vaak erg variabel, dat wil zeggen dat veel factoren niet constant zijn, vaak ook niet constant te houden zijn, bijvoorbeeld de factoren die de eileg van koolmezen beïnvloeden. Er kan gezocht worden naar waarnemingssituaties en -tijdstippen die zo weinig mogelijk van elkaar verschillen of er kan dan gekozen worden voor ‘randomisering’, in de verwachting dat de verschillen elkaar uitmiddelen.

Om de invloed van een variabele te kunnen waarnemen is het vaak handig c.q. noodzakelijk om een vergelijking te maken met de situatie waarin die invloed afwezig is. Dat wordt een controle-experiment of blanco genoemd. Bijvoorbeeld bij het bepalen van de werking van een enzym een reageerbuis met substraat, maar zonder enzym. Niet altijd is dat mogelijk. In het koolmezenonderzoek is bijvoorbeeld geen situatie te realiseren zonder temperatuur(svariatie).

- *Elementen van begrip van bewijs m.b.t. metingen en observaties*

11. De observatie(s) of meting(en) zelf mogen geen invloed hebben op de uitkomst van die observatie(s)/meting(en) of die invloed moet expliciet gemaakt worden.
12. Het bereik en de intervallen van de gekozen waarden van de onafhankelijke variabele moeten passen bij de verwachte variatie van de afhankelijke variabele.
13. De gevoeligheid van het instrument moet passen bij de te verrichten meting(en).
14. De steekproef moet representatief zijn voor de populatie:
    - groot genoeg
    - aselect of gestratificeerd gekozen waarbij de manier van kiezen niets met het object te maken mag hebben.

    Het aantal metingen of observaties moet groot genoeg zijn.
15. Van uit de toon vallende resultaten ('uitbijters') moet beoordeeld worden of ze in de verwerking meegenomen moeten worden (zijn het artefacten?).
16. Wanneer uitkomsten van metingen gemiddeld worden, moet dit inhoudelijk verantwoord zijn.
17. Er mag alleen van een verschil tussen metingen gesproken worden als dat verschil statistisch significant is.

Observaties of metingen mogen de uitkomst ervan niet beïnvloeden. Een observator kan het gedrag van de door hem geobserveerde dieren beïnvloeden; een elektrode in een cel kan de werking van die cel verstoren. Maar ook een vraag in een interview kan sociaal wenselijke antwoorden uitlokken en zo de resultaten vertekenen. Daarom is het belangrijk alert te zijn op dergelijke effecten en ze zoveel mogelijk te vermijden.

Voor het doen van zinvolle waarnemingen is het uiteraard ook van belang om de juiste middelen te kiezen en intervallen die passen bij de vraag. De deling van bacteriën volgen met een stereomicroscoop gaat niet; voor het vinden van de optimumtemperatuur van de werking van een enzym zijn intervallen van 10 graden niet zinvol.

Statistische significantie komt in biologisch onderzoek op het vwo zelden aan de orde, maar voor het beoordelen van de waarde van een onderzoek is het vaak erg belangrijk, gezien de inherente variabiliteit van veel biologische verschijnselen. Daarmee direct samenhangend is de vraag hoe groot een steekproef moet zijn en hoe vaak metingen of waarnemingen moeten worden herhaald.

En de eerste stap bij het verwerken van gegevens is het op waarde schatten ervan: is er niet toevallig een vreemd effect opgetreden waardoor een sterk afwijkende waarde verklaard kan worden; mag ik de waarden zomaar middelen? Het idee van leerlingen dat één meting de 'ware' waarde van een variabele kan vaststellen, is erg hardnekkig (Lubben & Millar, 1996; Buffler, Allie, Lubben & Campbell, 2001).

- *Elementen van begrip van bewijs m.b.t. conclusie(s) en verklaring*

18. De conclusie(s) moet(en) aansluiten bij de onderzoeksvraag en de hypothese.
19. Een correlatie is nog geen oorzakelijk verband en mag dus niet zonder meer als zodanig geïnterpreteerd worden.

20. Er wordt aangegeven of een verklaring een oorzakelijke verklaring (verklaring vanuit de oorzaken) of een functionele verklaring (verklaring vanuit de gevolgen) is.

De interpretatie van de resultaten leidt achtereenvolgens tot een terugkoppeling naar de voorspelling (kwam die uit?), de hypothese (wordt die verworpen of niet?) en (of alleen, bij een beschrijvend onderzoek) de onderzoeksvraag (is er een antwoord gevonden?). Daarbij is het meestal nodig om een (logische) redenering op te zetten om van het een naar het ander te komen en om eventueel verschillende interpretaties tegen elkaar af te zetten.

Een onderzoeksvraag beschrijft een probleem waarvoor een antwoord in de vorm van een beschrijving of verklaring gezocht wordt, stelden we al eerder. En als er naar een verklaring van een biologisch verschijnsel gezocht wordt, is het van belang te onderscheiden of je op zoek bent naar de oorzaak (waardoor) of de functie ervan (waartoe). Hoewel sommigen beweren dat elke functionele verklaring via de omweg van de evolutionaire ontwikkeling te herleiden is tot een oorzakelijke, wordt door Wouters (1999; zie ook Zeiss, 2000) het belang van functionele verklaringen wetenschapsfilosofisch onderbouwd. Ook voor leerlingonderzoek is het een zinvol onderscheid om leerlingen meer inzicht te geven welk type antwoord ze eigenlijk geven op hun onderzoeksvraag.

Dat verbanden die tussen variabelen aangetoond worden, niet zonder meer iets verklaren, wordt uitgedrukt met het begrip dat een correlatie niet zonder meer als causaal verband geïnterpreteerd mag worden. Ten eerste is het bij een correlatie niet vanzelfsprekend dat welke variabele oorzaak zou kunnen zijn en welke gevolg, maar er is ook de mogelijkheid dat beide door een derde variabele veroorzaakt worden, dan wel dat er toeval in het spel is. Voorzichtigheid is dus geboden.

- *Elementen van begrip van bewijs m.b.t. evaluatie*

21. Als een onderzoek niet valide en betrouwbaar is uitgevoerd (met andere woorden: als niet aan de voorgaande criteria is voldaan), hebben de resultaten weinig waarde.
22. Indien mogelijk worden de resultaten van een onderzoek vergeleken met gegevens uit ander onderzoek.
23. De conclusie wordt vergeleken met geaccepteerde ideeën (theorieën), gezond verstand en ervaring.

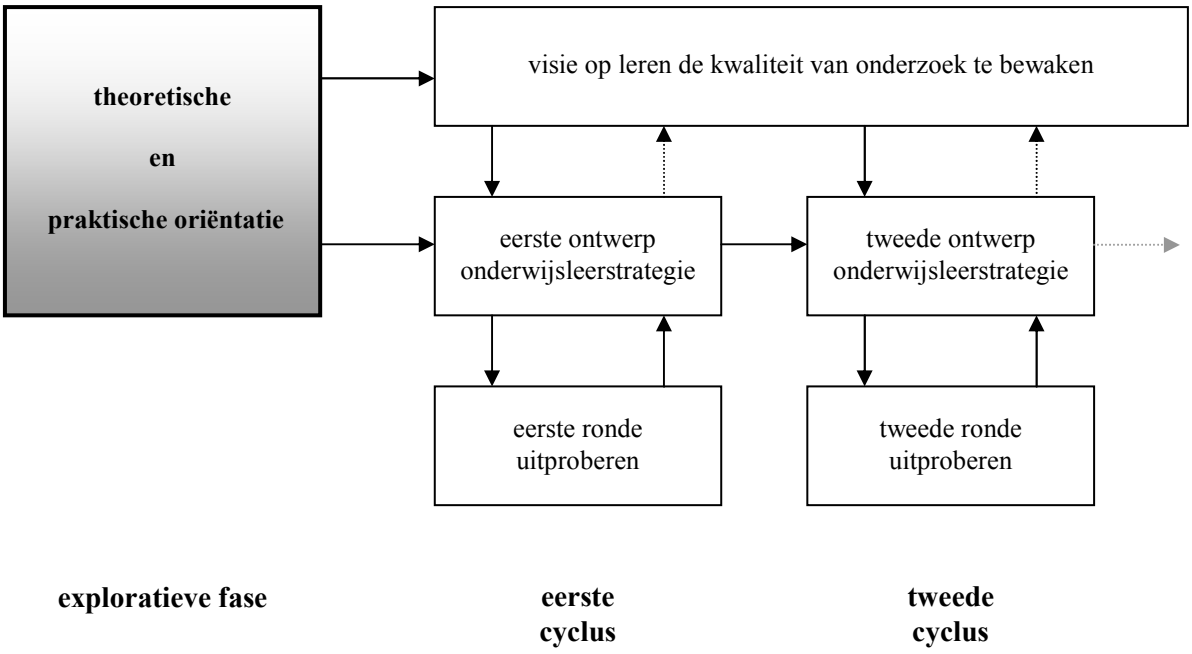
Bij de evaluatie draait het om de vraag ‘hoe zeker ben ik van het gevonden antwoord op de onderzoeksvraag?’ Dat kan eigenlijk alle aspecten van het onderzoek betreffen. Misschien blijkt achteraf het design toch niet geschikt om de hypothese te toetsen, blijkt de nauwkeurigheid van het instrument of de methode tegen te vallen, of blijkt de variatie van de onafhankelijke variabele verkeerd gekozen. Het kan zijn dat de hypothese niet verworpen kan worden, maar ook niet bevestigd.

Het gaat dus uiteindelijk om het inschatten van de validiteit en betrouwbaarheid van het onderzoek en de resultaten. Element 21 sluit daarmee alle voorgaande in, omdat die een uitwerking van de twee centrale begrippen zijn. Validiteit van een onderzoek is daarbij de mate waarin het bewijsmateriaal een antwoord verschaft op de vraag en de betrouwbaarheid de mate waarin de gegevens vertrouwd kunnen worden (Gott & Duggan, 2003). Daarbij wordt betrouwbaarheid vooral gezien als de mate waarin herhaling van het onderzoek tot dezelfde resultaten leidt. Vergelijking van de resultaten en conclusies met de bevindingen van anderen en gevestigde ideeën draagt daarom ook bij aan het vertrouwen dat je kunt hebben in de kwaliteit van je onderzoek. Dit verwijst naar het sociale karakter van wetenschapsbeoefening, dat in de elementen 22 en 23 tot uitdrukking komt.

- *Hoe belangrijk zijn de verschillende elementen?*

Niet alle elementen van begrip van bewijs zijn in het kader van het biologieonderwijs in de bovenbouw van het vwo even belangrijk. Na een verkenning van wat het examenprogramma voorschrijft en wat docenten in de praktijk het belangrijkst vinden, wordt in hoofdstuk 3 beargumenteerd een selectie uit de 23 elementen gemaakt.

Zoals gezegd zijn de als stellingen geformuleerde criteria ook niet in leerlingtaal geschreven. Om voor leerlingen toegankelijk te zijn is dus nog een vertaalslag nodig. Die wordt gemaakt bij het concretiseren van de onderwijsleerstrategie (hoofdstuk 5).



3.1 Onderzoek doen volgens vakliteratuur en officiële leerdoelen
3.2 Onderzoek doen volgens docenten
3.3 Onderzoek doen volgens biologiemethodes
3.4 Onderzoek doen in leerlingverslagen
3.5 Keuzes

3 'Onderzoek doen' in het biologieonderwijs

'Even in curricula that are largely content oriented, hands-on inquiry activities play an important role. The goal of these activities is to provide a context in which students can learn to reason scientifically.'
Clarck A. Chinn, Betina A. Malothra (2002, p. 175)

Hoofdstuk 3 'Onderzoek doen' in het biologieonderwijs

In hoofdstuk 1 is de aanleiding voor dit onderzoek beschreven: zorg om de kwaliteit van het onderzoek dat leerlingen in het kader van het biologieonderwijs doen en vooral om het inzicht in kwaliteit dat ze daarbij opdoen. In hoofdstuk 2 is kwaliteit van onderzoek uitgewerkt in een lijst van 23 elementen van 'begrip van bewijs' die geschikt geacht worden voor het biologieonderwijs op het vwo. De vraag is echter in hoeverre het beheersen van dat begrip van bewijs redelijkerwijs van vwo-leerlingen verwacht mag worden. Is het niet een te ambitieuze doelstelling, met andere woorden: moeten ze het wel leren? Om die vraag te kunnen beantwoorden is eerst gekeken naar de plaats die het doen van onderzoek in het biologieonderwijs heeft of zou moeten hebben, gebaseerd op literatuur vanuit de *science education research.* Daar zijn de officiële eindtermen uit het huidige examenprogramma biologie voor het vwo naast gelegd evenals een recente schets voor een nieuw programma (par. 3.1).

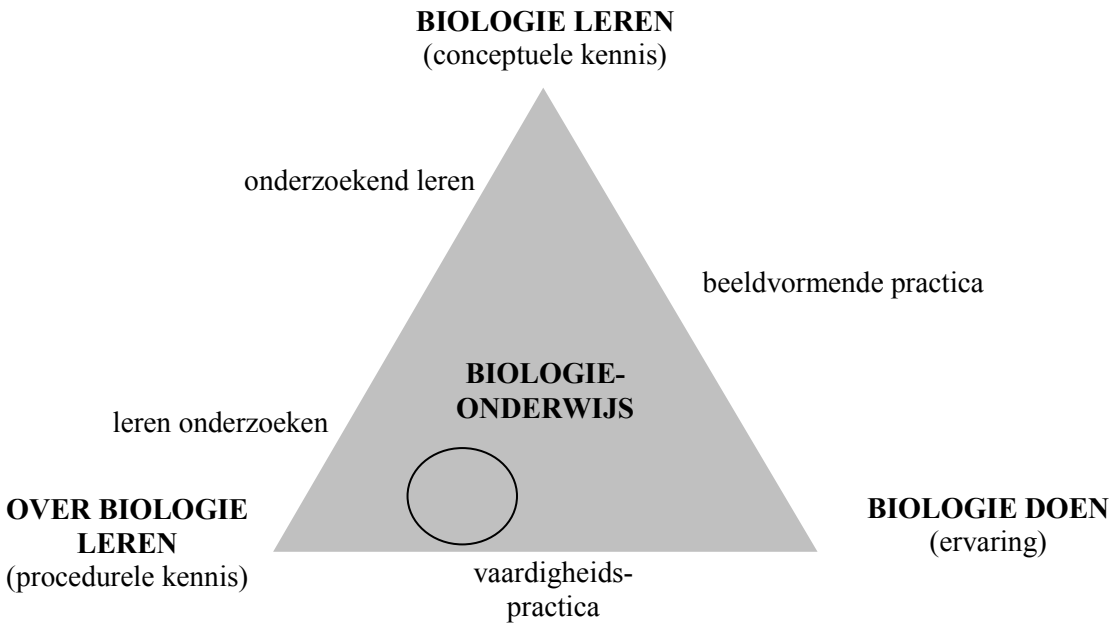
De gesignaleerde zorg om de kwaliteit van leerling-onderzoek zou zijn oorzaak kunnen vinden in een gebrek aan aandacht ervoor in het huidige biologieonderwijs in Nederland. Als er in lessen en schoolboeken weinig aandacht besteed wordt aan de kwaliteit van onderzoek, is het logisch dat leerlingen het niet leren. Aan de andere kant, als er in schoolboeken en het gegeven c.q. genoten onderwijs wel veel aandacht voor is, is er dan wel behoefte aan een (nieuwe) onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek? Leren ze het niet al of hebben ze tenminste de gelegenheid al om het te leren? Daarom is aan docenten gevraagd wat hun mening is over de 23 elementen van begrip van bewijs; welke vinden ze belangrijk voor hun leerlingen en welke niet, en aan welke besteden ze nu al aandacht (par. 3.2)? En daarom zijn ook schoolboeken en leerlingverslagen geanalyseerd (resp. par. 3.3 en 3.4).

Dat leidt tot de conclusie dat er nog wel wat te leren valt en dat een strategie daarvoor niet meteen voorhanden is. Het leidde ook tot het temperen van de ambitie om alle elementen van begrip van bewijs uit te werken in een onderwijsleerstrategie; een prioritering blijkt nodig. In paragraaf 3.5 wordt een selectie beargumenteerd.

3.1 Onderzoek doen volgens vakliteratuur en officiële leerdoelen

Onderwijs in de natuurwetenschappen, en dat geldt ook voor het biologieonderwijs, is op te vatten als een combinatie van *learning science, learning about science* en *doing science* (Hodson, 1992). Met het eerste wordt de vakinhoud, de kennis van biologische concepten en theorieën bedoeld, het tweede omvat kennis over de aard en methoden van de natuurwetenschap en over de historische en sociologische aspecten, *doing science* wordt door Hodson (1992, p. 549) omschreven als 'engaging in and developing expertise in scientific inquiry and problem solving.' Neergezet in een driehoek (fig. 3.1) geven ze het spanningsveld weer waar het accent van het biologieonderwijs zou moeten liggen.

Tussen 'biologie leren' en 'over biologie leren' ligt de spanning tussen wetenschap als geheel van concepten en theorieën en wetenschap als proces van kennisverwerving. De mees-



Figuur 3.1 Model voor het biologieonderwijs als spanningsveld tussen 'biologie leren', 'over biologie leren' en 'biologie doen'. De (beoogde) plaats van het leren bewaken van de kwaliteit van onderzoek is aangegeven met een cirkel.

te vernieuwingen in het science- of biologieonderwijs in de afgelopen eeuw hadden de bedoeling om het accent uit de hoek van 'biologie leren' te halen. Bijvoorbeeld de methoden van het BSCS-project (Biological Science Curriculum Study) van het American Institute of Biological Science (1959) waren gericht op een onderzoekende en kritische houding en de daarbij behorende vaardigheden en de methode *Science – A Process Approach* van de American Association for the Advancement of Science (1967) legde de nadruk op wetenschap als proces van kennisverwerving. Dientengevolge moesten leerlingen niet primair vakkennis leren, maar vooral algemeen inzetbare procesvaardigheden, 'generalisable process skills', zoals 'observing', 'measuring', 'formulating hypotheses', en 'experimenting'. Ook het veelgebruikte didactiekboek *De bioloog als leraar* (Falk e.a., 1974) pleit voor een evenwicht tussen de basisthema's uit de biologie en de natuurwetenschappelijke denk- en werkwijze in het Nederlandse biologieonderwijs. Op eenzijdige nadruk op procesvaardigheden is overigens ook weer kritiek geleverd, o.a. in het veel geciteerde artikel *Beyond processes* van Millar en Driver (1987): je hoeft kinderen niet te leren hypotheses op te stellen, dat kunnen ze al lang, het gaat erom dat ze *wetenschappelijke* hypotheses leren opstellen, en daarvoor is ook vakinhoud nodig.

'Onderzoeken' krijgt dan op twee manieren een plaats in het biologieonderwijs, als *onderzoekend leren* en als *leren onderzoeken.* Onderzoekend leren is gericht op het 'zelf ontdekken' van de vakkennis, want – zo is de veronderstelling – dat beklijft beter en zo wordt de ontdekkingscontext van de kennis duidelijk; het proces van onderzoeken is dan secundair. Bij leren onderzoeken staan onderzoeksvaardigheden centraal, daarbij is het minder van belang rond welk biologisch onderwerp er onderzocht wordt. In het model van figuur 3.1 staat onderzoekend leren dus dichter bij de hoek 'biologie leren' en leren onderzoeken dichter bij 'over biologie leren'.

De hoek 'biologie doen' tegenover de beide andere hoeken geeft de spanning weer tussen leren met je hoofd en leren met je handen en gaat dus over de plaats van practicum. In het standaardwerk *Practicum biologie* (Van Alderwegen-de Vries e.a., 1988) wordt het ver-

Tabel 3.1 Percentage docenten (natuurkunde, biologie, geschiedenis, economie) dat aangeeft in redelijke of sterke mate de volgende doelen na te streven met het laten doen van onderzoek door hun leerlingen (Stokking & Van der Schaaf, 2000, p. 107, 130).

doel		*survey 1998 (N=49)*	*survey 1999 (N=165)*
1	leerlingen op een motiverende manier kennis over het vak laten opdoen	91	88
2	leerlingen onderzoek laten doen als oefening in zelfstandig werken en leren	85	82
3	leerlingen inzicht in en overzicht over de onderzoekscyclus laten krijgen	54	69
4	leerlingen ervaring laten opdoen en laten oefenen met een aantal deelvaardigheden	91	89
5	leerlingen kennis laten opdoen van 'de gereedschapskist' van de onderzoeker (methoden en technieken)	55	71
6	leerlingen kennis laten maken met de werkwijze en houding van een onderzoeker (nieuwsgierig, systematisch, nauwkeurig, e.d.)	54	69
7	leerlingen laten inzien dat onderzoek een manier is om kennis te verkrijgen	80	90

band geëxpliciteerd tussen vorm en doel van practica. In 'beeldvormende practica' gaat het om het concretiseren van de leerstof en om de zintuiglijke gewaarwordingen die leerlingen daarbij krijgen. Het gaat erom te ervaren hoe de biologische objecten en verschijnselen er in het echt uitzien, hoe ze voelen en dergelijke. Daarom staat deze vorm van practicum langs de zijde tussen 'biologie doen' en 'biologie leren'. Als het accent geheel ligt op vaardigheden, van het bedienen van een microscoop tot het formuleren van hypotheses zijn deze practica te plaatsen langs de zijde tussen 'biologie doen' en 'over biologie leren'.

'Onderzoekspractica' hebben in de praktijk verschillende doelen en kunnen daarom eigenlijk op elke plek van de driehoek geplaatst worden, afhankelijk van hun doel en de invulling daarvan. In twee surveys, resp. in 1998 en 1999 onderscheidden Stokking en Van der Schaaf (2000) zeven doelen, zie tabel 3.1. De meest nagestreefde doelen betreffen zowel de vakkennis (doel 1) als het opdoen van ervaring (doel 4) en inzicht in de manier van kennis verkrijgen (doel 7). Opmerkelijk genoeg scoren de doelen die met procedurele kennis te maken hebben (doel 3, 5 en 6) het laagst.

Het doen van onderzoek door leerlingen kan dus verschillende doelen dienen, zowel kennis, vaardigheden als houdingen. Het is zinvol hier aandacht te besteden aan het onderscheid daartussen en de overlap ervan. Stokking en Van der Schaaf (2000, p. 24) geven in een 'principiële verkenning' een heldere uiteenzetting hoe kennis en vaardigheden verweven zijn en wat onder onderzoeksvaardigheden verstaan kan worden:

> Vaardigheden zijn in belangrijke mate gebaseerd op kennis van procedures. (..) Een procedure heeft betrekking op een aantal handelingen die in een bepaalde volgorde moet worden uitgevoerd om een bepaald doel te bereiken, resp. een probleem op te lossen.

Zij onderscheiden drie dimensies waarop je vaardigheden kunt indelen: smal versus breed, als vermogen versus als prestatie en deelvaardigheden versus een complexe totale vaardigheid. De smalle betekenis heeft daarbij de betekenis van het flexibel of routinematig

uitvoeren van een serie handelingen, terwijl de brede betekenis een vaardigheid gezien wordt als een competentie, te weten een geheel van kennis, vaardigheden en houdingen. Het onderscheid tussen vermogen ('hij kan het') en prestatie ('hij doet het') is vooral van belang voor het bepalen of iemand een vaardigheid beheerst; het (veronderstelde) vermogen kan eigenlijk alleen maar waargenomen worden via een handeling of product. Een veelomvattende vaardigheid als 'onderzoeken' of 'samenwerken' wordt vaak gezien (en onderwezen) als verzameling deelvaardigheden, terwijl het maar de vraag is of de som van de deelvaardigheden wel hetzelfde is als de totale vaardigheid.
De combinatie van deze drie dimensies geeft een matrix van 'vaardigheden in soorten en maten' (fig. 3.2). Van links naar rechts gaat het 'van oorzaak naar gevolg': cognitieve en affec-

<table>
<tr><td colspan="2" rowspan="2"></td><td rowspan="2">vaardigheid als (verondersteld) vermogen of proces, nodig voor het leveren van bepaalde prestaties</td><td colspan="2">vaardigheid als taakgerichte prestatie ('performance')</td></tr>
<tr><td>vaardigheid als het (snel, precies, flexibel) (kunnen) uitvoeren van (zichtbare) handelingen</td><td>vaardigheid, gedefinieerd in termen van de eisen waaraan het resultaat of product moet voldoen</td></tr>
<tr><td colspan="2">vaardigheid als complexe totale vaardigheid of taak
(op te vatten in de brede betekenis: als geheel van kennis, vaardigheid en houding)</td><td>'competentie'</td><td>dit betreft vooral het uitvoeren van meerdere handelingen in de juiste volgorde

(bijv.: uitvoeren van de stappen van een onderzoek</td><td>hierop worden mensen (jongeren, beroepsbeoefenaren) vaak uiteindelijk afgerekend

(bijv.: een gemaakte keuze, een uitgevoerd onderzoek</td></tr>
<tr><td rowspan="3">vaardigheid als deelvaardigheid of deeltaak</td><td>brede betekenis (kennis, vaardigheid, houding)</td><td>(vaardigheden op het vlak van monitoring, metacognitie, reflectie, zelfsturing?)</td><td>hierop wordt vaak gelet bij beoordeling van 'het proces'

(bijv.: samenwerken)</td><td>(bijvoorbeeld: geen keuze in overeenstemming met wat anderen willen dat jij kiest; geen planning waarvan je van te voren al weet dat die niet haalbaar is)</td></tr>
<tr><td rowspan="2">smalle betekenis (geroutiniseerde procedure)</td><td rowspan="2">het specificeren hiervan is een eindeloze zaak

het betreft echter wel 'wat in het hoofd van de leerling of student gebeurt'</td><td colspan="2">hierop kun je in onderwijs gericht instrueren, begeleiden en beoordelen
veel eindtermen zijn op deze manier geformuleerd</td></tr>
<tr><td>(bijvoorbeeld: gegevens analyseren, criteria toepassen)</td><td>(bijvoorbeeld: correcte conclusies trekken)</td></tr>
</table>

Figuur 3.2 Vaardighedenmatrix (uit: Stokking & Van der Schaaf, 2000, p. 27).

Tabel 3.2 Onderzoeksvaardigheden volgens de eindtermen voor de Tweede Fase, domein A-6, eindtermen 35 t/m 43 (SLO, 1996).

De kandidaat kan	
35	een natuurwetenschappelijk probleem herkennen en specificeren
36	verbanden leggen tussen probleemstelling, hypotheses, gegevens en aanwezige natuurwetenschappelijk voorkennis
37	een natuurwetenschappelijk probleem herleiden tot een onderzoeksvraag
38	hypotheses opstellen en verwachtingen formuleren
39	prioriteiten, mogelijkheden en randvoorwaarden vaststellen om een natuurwetenschappelijk onderzoek uit te voeren
40	een werkplan maken voor het uitvoeren van een natuurwetenschappelijk onderzoek ter beantwoording van een onderzoeksvraag
41	relevante waarnemingen verrichten en (meet)gegevens verzamelen
42	conclusies trekken op grond van verzamelde gegevens van uitgevoerd onderzoek
43	oplossing, onderzoeksgegevens, resultaat en conclusies evalueren

tieve vaardigheden → (zichtbare) handelingen → resultaten en producten. Van beneden naar boven gaat het van deel naar geheel: vaardigheden in de smalle betekenis → vaardigheden in brede betekenis → vaardigheden als complex geheel van kennis, vaardigheid en houding. De cel linksboven, de combinatie van 'vaardigheid als verondersteld vermogen' en 'vaardigheid als geheel van kennis, vaardigheden en houding', heet dan 'competentie'.

Zowel het model van figuur 3.1 als de vaardighedenmatrix laten zien dat er veel benaderingen van het doen van onderzoek door leerlingen mogelijk zijn. De praktijk van het Nederlandse biologieonderwijs is dan ook veelvormig (zie bijvoorbeeld Griethuysen-Elfers e.a., 1991; Kuiper, 1993). Tegelijkertijd wordt duidelijk dat het doen van onderzoek er niet uit weg te denken is. Dat roept de vraag op welke plaats en vorm eraan gegeven wordt door de officiële leerdoelen, zoals neergelegd in het examenprogramma voor het vwo.

Het examenprogramma voor de vakken biologie 1 en biologie 1,2 (SLO, 1996) is niet het eerste programma waarin onderzoeksvaardigheden zijn opgenomen. Ook het eraan voorafgaande examenprogramma vermeldde die. De vaardigheidsdoelen zijn voor biologie, natuurkunde en scheikunde op gelijke wijze geformuleerd. In tabel 3.2 zijn de eindtermen uit het examenprogramma biologie uit 1996 opgenomen, dus zoals ze vanaf de invoering van de tweede fase golden tot de aanpassing voor 2007 en verder. Dat lijstje oogt als stappen in het doen van een onderzoek dan wel als deelvaardigheden die daarbij van belang zijn. Dat is ook de insteek bij het maken van de centrale examens (Pohlmann-Nepveu, 1996), bijvoorbeeld het opstellen van een onderzoeksopzet of het trekken van een conclusie uit voorgeschotelde gegevens.

Bij de aanpassing voor 2007 en verder wordt in het officiële programma een globalere formulering gehanteerd: 'de kandidaat kan een natuurwetenschappelijk onderzoek voorbereiden, uitvoeren, de verzamelde onderzoeksresultaten verwerken en hieruit een conclusie trekken.' Maar de specificatie ervan in de *Handreiking schoolexamen biologie havo/vwo* (Legierse, 2006), is precies hetzelfde rijtje als in tabel 3.2.

In termen van de hierboven genoemde indelingen zijn de officiële leerdoelen dus vooral op te vatten als het (kunnen) *uitvoeren* van *deelvaardigheden*. In de geschetste drie-

hoek dicht bij de hoek 'biologie doen' en in de matrix op de kruising van enerzijds 'vaardigheid als taakgerichte prestatie' en anderzijds 'vaardigheden in de smalle betekenis'.

Er is echter ook een andere tendens waar te nemen. Boersma en Kamp (2001) pleiten voor competenties als doelstellingen in het biologieonderwijs, die zij omschrijven als een handelingsrepertoire dat functioneel is in een of meer praktijken. In een themanummer van het tijdschrift *Niche* over 'competenties in het biologieonderwijs' omschrijft de themaredactie een competentie als volgt (Schalk, Kapteijn, Van Gelderen, Neerincx & Wagenaar, 2002, p. 10):

> Essentieel voor een competentie is dat het geheel meer is dan de som der delen. Het is niet voldoende dat een leerling een aangeleerde 'truc' kan toepassen in de aangeleerde situatie. Er is pas sprake van een competentie als het aangeleerde toepasbaar is in diverse voorkomende situaties. De leerling moet dus kunnen inschatten hoe de aangeleerde vaardigheid voor een bepaalde situatie moet worden bijgesteld en welke kennis precies moet worden gebruikt.

Op de vraag welke competenties vervolgens voor het toekomstige biologieonderwijs zinvol worden geacht, noemt Boersma (2001, p. 314) onder andere 'oriënteren op de aard, inhoud en werkwijzen van natuurwetenschappelijk (biologisch) onderzoek' en voegt daaraan toe:

> Het is wat mij betreft niet de bedoeling dat leerlingen leren onderzoeken. Deze competentie is voor havo en vwo te hoog gegrepen en heeft te weinig algemeen vormende betekenis.

Wat dan wel beoogd wordt, blijkt uit een recent concept voor een nieuw examenprogramma (CVBO, 2006) en de daarbij behorende leerlijn (Boersma, Van Graft, Harteveld, De Hullu, Mazereeuw, Van den Oever & Van der Zande, 2006). Uitgaande van de 'concept-context-benadering' wordt onderzoeken beschreven als een activiteit én als twee verwante biologische denk- en werkwijzen. Onderzoeken wordt gezien als een kernactiviteit in wetenschappelijke

Tabel 3.3 De plaats en invulling van onderzoeken in het conceptexamenprogramma vwo van de Commissie Vernieuwing BiologieOnderwijs (CVBO, 2006) en de daarbij behorende leerlijn 4-18 (Boersma e.a., 2006).

Uit het conceptexamenprogramma vwo, onder 'denk- en werkwijzen'

De kandidaat kan (…)

5. Onderzoeken
- omschrijven welke fasen bij het uitvoeren van onderzoek veelal worden doorlopen
- omschrijven aan welke criteria valide uitgevoerd biologisch onderzoek moet voldoen

Uit de leerlijn biologie 4-18

A. Biologische denk- en werkwijzen

Beschrijvend en vergelijkend onderzoeken	Werkwijze waarbij biologische objecten worden beschreven, vergeleken worden met andere biologische objecten, en eventueel worden gecategoriseerd, op basis van vooraf vastgestelde kenmerken
Experimenteel onderzoeken	Denk- en werkwijze waarbij een onderzoeksvraag wordt geformuleerd en een daarop aansluitende hypothese experimenteel wordt getoetst; voor onderzoek naar herhaalbare processen

contexten (praktijken). Aan activiteiten ontlenen leerlingen de betekenis van concepten, in dit geval de criteria voor valide onderzoek (tabel 3.3). Het is echter het niet per se de bedoeling dat leerlingen de activiteit leren uitvoeren zoals die in de reële context wordt uitgevoerd. In school vindt een analoge gedidactiseerde activiteit, een leeractiviteit, plaats in een gedidactiseerde context. Denk- en werkwijzen worden gezien als een nadere operationalisering van activiteiten; de activiteit 'onderzoek' wordt geoperationaliseerd in 'beschrijvend en vergelijkend onderzoeken' en 'experimenteel onderzoeken'. Samen met de eindterm uit het conceptexamenprogramma vwo is de omschrijving van die werkwijzen weergegeven in tabel 3.3.

De benadering om onderzoeken te beschouwen als competentie, als handelingsrepertoire c.q. denk- en werkwijze toepasbaar in een of meer praktijken in het dagelijks leven of de wetenschap strookt met de holistische manier waarop Hodson (1992; zie ook Hodson & Hodson, 1998a, 1998b) tegen het onderwijs in de natuurwetenschappen aankijken: als *enculturation* in praktijken waarin natuurwetenschappen een rol spelen. Als we competentie dan zo breed zien als Stokking en Van der Schaaf (2000) dat doen, als geheel van kennis, vaardigheden en houdingen, kan het zowel de wetenschappelijke houding omvatten waar De Vos en Genseberger (2000) en Smits e.a. (2000) voor pleiten, als de procedurele kennis van 'begrip van bewijs' waar onder andere Millar e.a. (1994) op doelen en die in hoofdstuk 2 is omschreven.

De eerste vraag aan het begin van dit hoofdstuk: 'Moeten leerlingen eigenlijk wel leren de kwaliteit van onderzoek te bewaken?' kan dan ook genuanceerd beantwoord worden. Het antwoord is 'nee' als je strikt naar de eindtermen van het vigerende examenprogramma kijkt; dan is het in voldoende beheersen van deelvaardigheden genoeg. Het antwoord is echter 'ja' in het licht van de doelstelling dat leerlingen een begin van competentie ontwikkelen en zich daarvoor kennis, denk- en werkwijzen en houdingen van biologisch onderzoek eigen maken.

Hoe ver het biologieonderwijs daarin kan gaan, zal de praktijk moeten uitwijzen. Reden genoeg om aan docenten te vragen hoeveel belang zij hechten aan de elementen van begrip van bewijs voor hun leerlingen.

3.2 Onderzoek doen volgens docenten

Een conceptversie van de elementen van begrip van bewijs is met een klein aantal docenten besproken (Schalk, 2001). Op grond daarvan is de lijst bijgesteld en schriftelijk aan een grotere groep docenten voorgelegd.

Uit het ledenbestand van de Nederlandse Vereniging voor het Onderwijs in de Natuurwetenschappen (NVON) is een aselecte steekproef getrokken van 450 biologiedocenten met een eerstegraads bevoegdheid. Na schriftelijke en telefonische rapellering hebben 57 docenten de vragenlijst teruggestuurd, waarvan 48 ingevuld (een respons van 13 resp. 11 procent). Niet-reagerende docenten en de docenten die de vragenlijst oningevuld terugstuurden gaven daarvoor meestal als reden dat ze niet meer actief waren in het onderwijs. Er zijn twee argumenten om te veronderstellen dat de steekproef vertekend is in de richting van de meer actieve, bij het onderwerp betrokken docenten. Ten eerste omdat getrokken is uit de leden van de NVON en lid zijn van een vakorganisatie getuigt van betrokkenheid; een ander adressenbestand was echter niet voorhanden. Ten tweede zullen docenten die zich betrokken voelen op het onderwerp eerder geneigd zijn de vragenlijst in te vullen en te retourneren dan docenten die er geen interesse in hebben. De resultaten dienen dus met grote voorzichtigheid behandeld

te worden; het is waarschijnlijk dat ze een te hoge waardering opleveren van het belang en gebruik van de elementen van begrip van bewijs.

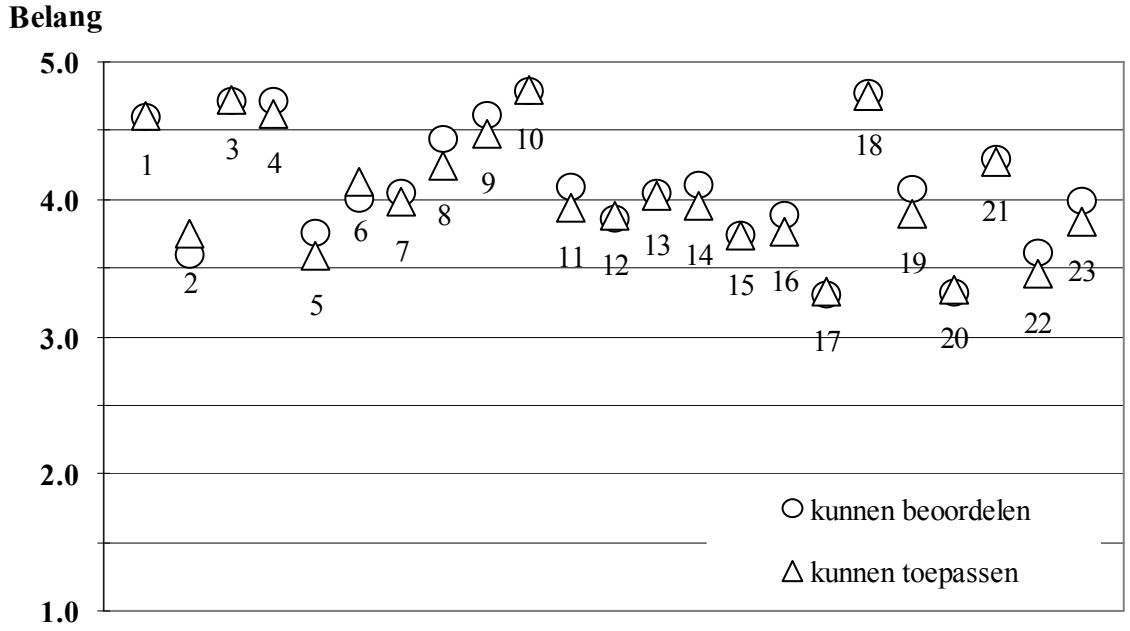
- *Toegekend belang aan de elementen*

In de vragenlijst (bijlage 1) zijn de 23 elementen van begrip van bewijs gepresenteerd en is de docenten gevraagd bij elk element op een vijfpuntsschaal hun mening te geven over twee stellingen met betrekking tot het belang dat ze er aan hechten:

- Ik vind het belangrijk dat mijn leerlingen de kwaliteit van (eigen of ander) onderzoek op dit element **kunnen beoordelen** [--: niet belangrijk / -: relatief onbelangrijk / ±: middelmatig belangrijk / +: relatief belangrijk / ++: zeer belangrijk (essentieel)]
- Ik vind het belangrijk dat mijn leerlingen dit aspect in eigen onderzoek **kunnen toepassen** [idem]

De resultaten zijn weergegeven in figuur 3.3. Alle elementen worden meer dan middelmatig belangrijk gevonden, want de gemiddelde waardering is overal hoger dan het midden van de schaal. Dat geldt zowel voor het kunnen beoordelen van onderzoek met betrekking tot het begrip (stelling 1) als het kunnen toepassen ervan in eigen onderzoek (stelling 2).

Toch worden niet alle elementen even belangrijk gevonden. Bijna neutraal (waardering 3) scoren de elementen 17 en 20[7] (resp. statistische significantie en het verschil tussen een oorzakelijke en een functionele verklaring), terwijl bij de eerste stelling ook 2, 5, 12, 15, 16 en 22 onder de waardering 'relatief belangrijk' (4) blijven en bij de tweede stelling de elementen 11, 14, 19 en 23 daar bijkomen. De twee elementen die het hoogst scoren zijn 10 en 18 (gemiddeld 4,8). Daarachter komen de elementen 3 en 4 (4,7), gevolgd door 1 en 9 (4,6), 8



Figuur 3.3 Belang dat docenten eraan hechten dat leerlingen de verschillende elementen van begrip van bewijs kunnen gebruiken bij het beoordelen van onderzoek van anderen (○) en kunnen toepassen in hun eigen onderzoek (Δ) . 1: niet belangrijk / 2: relatief onbelangrijk / 3: middelmatig belangrijk / 4: relatief belangrijk / 5: zeer belangrijk (essentieel). De cijfers verwijzen naar de elementen van begrip van bewijs.

[7] De elementen van begrip van bewijs staan omschreven in paragraaf 2.3. Een overzicht ervan staat op bladzijde 250 van dit proefschrift.

(4,4) en 21 (4,3). Daarbij is er weinig verschil in de toekenning van belang met betrekking tot toepassen in eigen onderzoek of beoordelen van onderzoek van anderen.

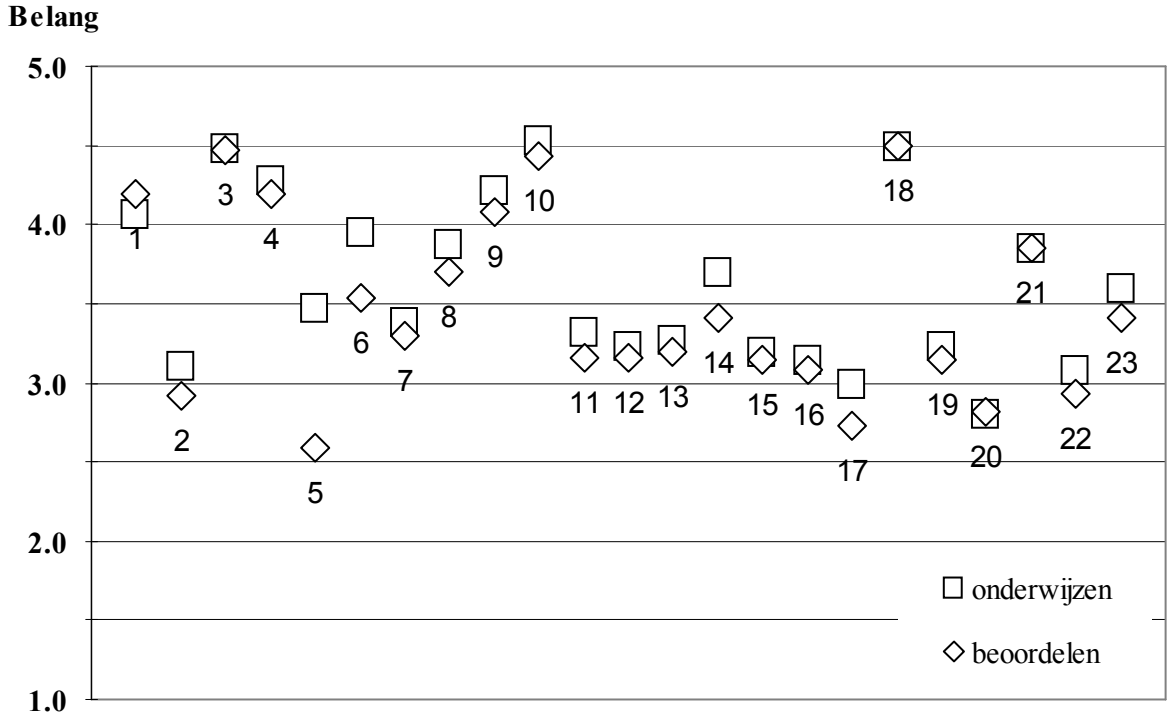
Uit de resultaten blijkt dat deze docenten aan de basale structuur van een hypothesetoetsend onderzoek het meeste belang toekennen: een duidelijke onderzoeksvraag (1), een daarop aansluitende (3) en toetsbare (4) hypothese, helderheid over variabelen (8, 9) en blanco (10), een conclusie die de onderzoeksvraag beantwoordt (18) en de notie van validiteit en betrouwbaarheid van het onderzoek (21).

- *De elementen in de onderwijspraktijk*

Docenten kunnen bepaalde onderwijsdoelen belangrijk vinden, maar dat wil nog niet meteen zeggen dat ze die in hun onderwijs daadwerkelijk nastreven. Andersom kan betoogd worden dat het belang dat docenten hechten aan bepaalde doelen afgelezen kan worden aan de mate waarin ze het in hun onderwijs vorm geven en aan de rol die ze spelen in de beoordeling. Daarom zijn aan de docenten nog twee andere stellingen voorgelegd met betrekking tot de 23 elementen van begrip van bewijs:

- Ik besteed er in mijn onderwijs **aandacht** aan [--: geen expliciete aandacht / -: weinig aandacht / ±: springt er qua aandacht niet uit / +: veel aandacht / ++: zeer veel aandacht]
- Dit aspect speelt een **rol in mijn beoordeling** van het onderzoek van leerlingen [--: geen rol / -: ondergeschikte rol / ±: gemiddelde rol / +: boven gemiddelde rol / ++: zware rol].

De gemiddelde oordelen van de docenten zijn weergegeven in figuur 3.4.

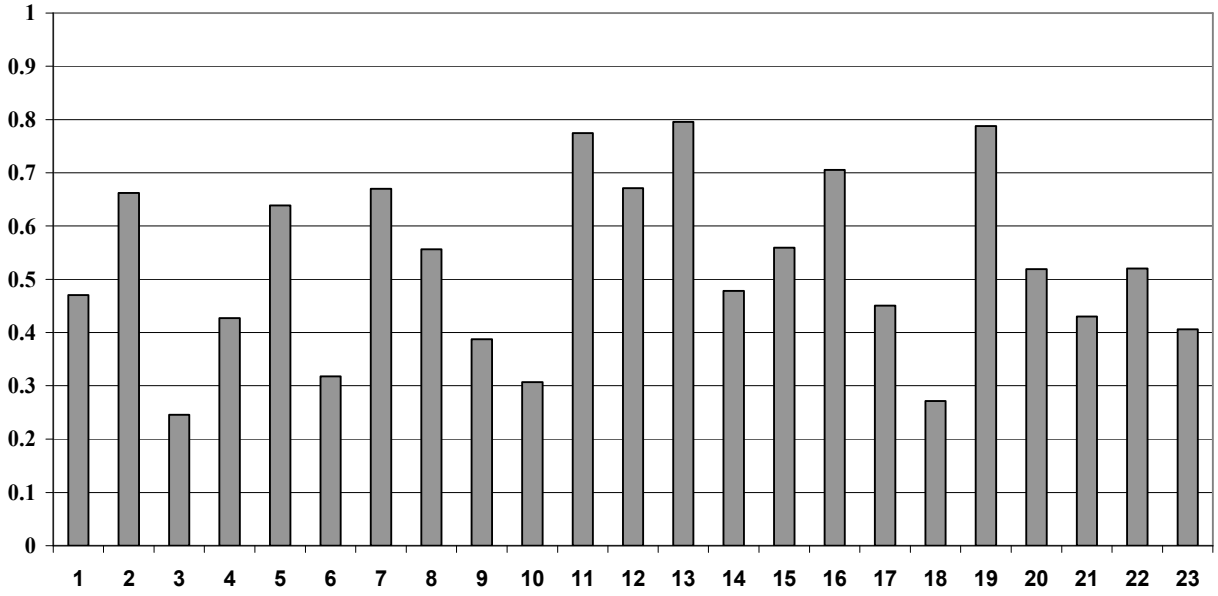


Figuur 3.4 Aandacht die docenten geven aan de verschillende elementen van begrip van bewijs in hun onderwijs en de rol die die elementen spelen in de beoordeling van leerlingonderzoek. □: aandacht in het onderwijs; 1: geen expliciete aandacht, 2: weinig aandacht, 3: springt er qua aandacht niet uit, 4: veel aandacht, 5: zeer veel aandacht; ◊: rol in de beoordeling; 1: geen rol, 2: ondergeschikte rol, 3: gemiddelde rol; 4: boven gemiddelde rol, 5: zware rol). De cijfers verwijzen naar de elementen van begrip van bewijs

Er zijn geen elementen waaraan de docenten geen aandacht geven, maar de meeste springen er qua aandacht niet uit. De 48 respondenten zeiden de meeste aandacht te geven aan onderzoeksvraag en een daarop aansluitende conclusie (elementen 1 en 18), het formuleren en toetsen van een hypothese (3, 4, 6) en het controleren van variabelen en een blanco (9, 10). Dezelfde elementen spelen de grootste rol in de beoordeling van het werk van hun leerlingen. Globaal volgen de scores met betrekking tot het gegeven onderwijs en de beoordeling dus hetzelfde patroon als de scores van het toegekende belang, en blijven ze bij elk element dicht bij elkaar. Alleen bij element 5 (onderkennen dat er meer hypotheses mogelijk zijn) is het verschil groot. Niet onbegrijpelijk dat daar bij de beoordeling minder gewicht aan wordt toegekend, een onderzoek wordt meestal beoordeeld op de hypothese die getoetst wordt, niet op de alternatieven.

De scores met betrekking tot de eigen praktijk liggen wel allemaal lager dan de scores voor het toegekende belang. In figuur 3.5 zijn de verschillen tussen de (gemiddelde) gegevens van figuur 3.3 en 3.4 weergegeven, waardoor het verschil tussen toegekend belang en eigen praktijk zichtbaar wordt. Dan blijkt dat de grootste discrepantie optreedt bij elementen die niet hoog scoren op toegekend belang: de elementen 11, 13 en 19 (meer dan 0,7 punt verschil), gevolgd door 2, 5, 7, 12 en 16 (meer dan 0,6 punt verschil). Dat is logisch, docenten doen meer moeite om doelen te realiseren die ze belangrijk vinden; doelen die ze minder belangrijk – ze noemen ze niet ónbelangrijk – komen minder aan bod. Andersom kan het echter ook betekenen dat ze de doelen waar ze de meeste aandacht aan besteden het meeste belang toekennen; een rechtvaardiging van de eigen praktijk.

Het beeld dat uit deze inventarisatie naar voren komt is beslist positief. Er wordt aandacht besteed aan begrip van bewijs in instructie en beoordeling. Er zijn echter twee kanttekeningen te maken die dit enigszins relativeren. Ten eerste – zoals aan het begin van deze paragraaf al genoemd – zijn er twijfels bij de representativiteit van de respondenten. Getrokken uit het bestand van NVON-leden en een kleine respons, is het niet onwaarschijnlijk dat vooral



Figuur 3.5 Verschil tussen toegekend belang (gemiddelde waarden uit figuur 3.3) en eigen praktijk (gemiddelde waarden uit figuur 3.4) met betrekking tot de elementen van begrip van bewijs.

actieve, bij het onderwerp betrokken docenten hebben gereageerd, die een bovengemiddeld belang hechten aan het doen van onderzoek door leerlingen en dientengevolge er relatief veel aandacht aan besteden. Ten tweede nodigt de vraagstelling vooral uit tot een *relatieve* inschatting van de aandacht die eraan gegeven wordt in instructie en beoordeling. Hoeveel aandacht dat betekent in uren of in vergelijking met bijvoorbeeld een onderwerp als genetica, blijft daardoor buiten zicht.

3.3 Onderzoek doen volgens biologiemethodes

Om een beeld te krijgen van de aandacht die in biologiemethoden aan begrip van bewijs wordt besteed, is een drietal methoden geanalyseerd.[8] Het betreft boeken van de eerste generatie voor de vernieuwde tweede fase in het vwo. Sommige uitgevers geven naast hun vakmethoden aparte uitgaven uit voor het aanleren van vaardigheden. Deze zijn bedoeld als aanvulling en ondersteuning van de vakmethoden en zijn meegenomen in de analyse om een volledig beeld te krijgen. Uit het totale aanbod van zes methodes voor vwo is gekozen voor de twee methoden met het grootste marktaandeel, *Biologie voor jou* en *Nectar* en voor de nieuwe methode *Biologie Actief*.

- *Aanpak van de analyse*

Van uitgeverij Nijgh Versluys is geanalyseerd de methode *Biologie Actief* vwo B1, B2A en B2B, van elk het informatieboek, het activiteitenboek, het antwoordenboek en voor vwo B1 ook het docentenboek (Walsarie Wolff-Cox & Krüger, 1999; Walsarie Wolff-Cox, Dertien & Krüger, 2000/2001/2002); daarnaast het vaardighedenboek *Zo doen we dat*, leerlingenversie en docentenversie (Weezenbeek, 1999). In het informatieboek van *Biologie Actief* vindt de leerling alle verplichte leerstof. Na elke paragraaf volgen enkele korte tekstvragen. De bedoeling van deze vragen is de leerling zelf te laten controleren of hij de inhoud van de tekst heeft begrepen. Daarna kan hij beginnen aan de activiteiten uit het activiteitenboek. Hierin wordt de leerling door verschillende werkvormen gestimuleerd iets met de aangeboden theorie te doen om zich zo de benodigde kennis eigen te maken. Het informatieboek bevat bovendien een hoofdstuk waarin alle vaardigheden kort worden samengevat. De leerling kan dit hoofdstuk raadplegen wanneer een bepaalde vaardigheid gevraagd wordt. *Zo doen we dat* is een handleiding bij praktische opdrachten voor de exacte vakken. Het beschrijft grote, veelomvattende vaardigheden als stappenplannen, bijvoorbeeld 'experimenteel natuurwetenschappelijk onderzoek' wordt verdeeld in de stappen 'keuze onderwerp', 'voorbereiding', 'keuring werkplan door docent', 'uitvoering' en 'verwerking'.

Van uitgeverij Malmberg is geanalyseerd de methode *Biologie voor Jou* vwo B1, B2 (deel 1, 2 en 3), steeds het leerlingenboek, het antwoordenboek, de docentenhandleiding en ook het boek B2 examenvragen (Smits & Waas, 1999/2000/2001), daarnaast de vaardighedenboeken *Algemene Vaardigheden Exact*, de delen '4 havo.vwo' en '5 havo.vwo en 6 vwo' (Edutekst, 1999/2000). De thema's in *Biologie voor jou* zijn steeds op dezelfde manier opgebouwd, namelijk inleiding, basisstof, samenvatting, diagnostische toets en verrijkingsstof. De basisstof met tekst en opgaven moet iedere leerling doen. Heeft een leerling tijd over, dan kan hij in overleg met de docent een keuze maken uit de 'verrijkingsstof'. De boeken *Algemene vaardigheden exact* zijn gemaakt om de leerlingen zelfstandig de vaardigheden uit het examenprogramma voor de natuurwetenschappen te laten verwerven. De aanpak lijkt op die van *Biologie voor jou* met geëxpliciteerde doelstellingen in termen van 'je kunt …' Bij de vaardigheid 'natuurwetenschappelijk onderzoeken' is de eerste daarvan 'je kunt de eisen noemen

[8] De analyse is uitgevoerd in het kader van een onderzoeksstage waarvan elders uitgebreider verslag is gedaan (Appelman, 2003).

waaraan een natuurwetenschappelijk onderzoek moet voldoen', waarvoor vervolgens naar de vakmethodes wordt verwezen.

Van uitgeverij Wolters-Noordhoff is geanalyseerd de methode *Nectar* vwo bovenbouw biologie 1, biologie 2 (deel 1 en 2), steeds het tekst- en opgavenboek, het uitwerkingenboek en de docentenhandleiding bij biologie 1 (Maier & Van Wijk, 1998/1999/2000). Basisteksten bevatten de kennis en vaardigheden uit het examenprogramma. Daarnaast biedt *Nectar* de leerling ondersteuning door leertaken, overzichten, 'test jezelf' en uitwerkingen. De leertaken oefenen biologische en algemene vaardigheden in opdrachten. Wanneer er geoefend wordt met de natuurwetenschappelijke methode wordt in de marge aangegeven welk deel van de natuurwetenschappelijke methode behandeld wordt. Wolters-Noordhoff geeft geen apart vaardighedenboek uit.

De uitgeverijen wilden of konden geen informatie geven over de mate waarin de vaardighedenboeken naast de biologiemethoden gebruikt werden. Daarom is een kleine peiling gehouden onder de deelnemers van een conferentie voor biologiedocenten. Van de dertig respondenten die les geven in de bovenbouw gebruiken elf docenten *Nectar*, tien docenten *Biologie voor jou* en één docent *Biologie Actief.* De vaardighedenboeken worden elk door één van de respondenten gebruikt. Daarom is besloten de resultaten van de analyse zowel met als zonder de bijdrage van de vaardighedenboeken weer te geven.

Er is ook een onderscheid gemaakt tussen verschillende typen tekst waarin begrip van bewijs aangetroffen kan worden. Het maakt immers verschil of een element van begrip van bewijs in een informatieve tekst staat, in een opgave gebruikt moet worden, in een uitwerking genoemd wordt of in de docentenhandleiding aan de orde komt. In het laatste geval is het via een beoogde invloed van de docent dat het begrip bij de leerlingen terecht komt.

Een derde onderscheid dat gemaakt is, is dat tussen het impliciet en het expliciet voorkomen van de elementen. Impliciet betekent dat het element wel aan de orde komt, maar dat de eisen die gesteld worden aan de uitvoering van het betreffende aspect van onderzoek doen, onbenoemd blijven; expliciet betekent dat de eisen wel benoemd worden. Een voorbeeld van impliciet is (*Biologie voor jou* vwo B2, deel2, thema 4, verrijkingsstof 2, p. 52, eerste opdracht):

> Kies een van de twee probleemstellingen of bedenk zelf een probleemstelling die te maken heeft met eigenschappen van planten die bescherming bieden tegen uitdroging. Formuleer vervolgens een onderzoeksvraag.

De leerlingen moeten hier wel een onderzoeksvraag formuleren aan de hand van een probleemstelling, maar er staat niet vermeld dat de onderzoeksvraag moet voldoen aan bepaalde eisen zoals ondubbelzinnigheid, specificiteit en afperking. Als de docent de opgave bespreekt, of als de leerlingen hun eigen werk nakijken, zal wellicht worden genoemd aan welke eisen een goede onderzoeksvraag moet voldoen.

Een voorbeeld van een expliciete manier waarop een element aan de orde komt is (*Nectar* vwo bovenbouw biologie 2, deel 1, praktische opdracht 1, p. 311, opgave 2):

> Om daadwerkelijk onderzoek uit te kunnen voeren is een nauwkeuriger formulering van de vraag nodig. Je zult de globale vraag moeten vertalen in een specifieke onderzoeksvraag. Lees de stappen door die daarvoor nodig zijn en bestudeer de bijbehorende uitwerking in het kader.

De leerlingen krijgen hier een algemene vraag die ze moeten omzetten in een specifieke onderzoeksvraag. Bij de stappen die ze moeten maken wordt er onder andere op gelet of de vraag niet te ruim is en of hij wel te beantwoorden is. Hier wordt dus uitdrukkelijk genoemd waar de leerlingen op moeten letten bij het formuleren van een onderzoeksvraag.

- *Resultaten van de analyse*

In tabel 3.4 zijn de resultaten weergegeven. Het vaakst komen de elementen van begrip van bewijs aan de orde in *Biologie Actief* in combinatie met *Zo doen we dat*, op de voet gevolgd door *Nectar*. *Biologie voor jou* samen met *Algemene vaardigheden exact* blijft daar enigszins bij achter. Zonder vaardighedenboeken scoren *Biologie Actief* en *Biologie voor jou* even hoog. Opmerkelijk is het verschil in verdeling over de verschillende typen tekst. In *Biologie Actief* staat begrip van bewijs het vaakst in de aangeboden informatie, terwijl dat bij *Nectar* vrijwel niet gebeurt. Evenals in *Biologie voor jou* komen de elementen in *Nectar*

Tabel 3.4 Aantallen keren dat de elementen van begrip van bewijs in verschillende biologiemethodes voor bovenbouw vwo aan de orde komen, impliciet dan wel expliciet, en percentuele verdeling daarvan over informatieve tekst voor leerlingen, opgaven, uitwerkingen en docentenhandleiding. AVE: *Algemene vaardigheden exact*; BA: *Biologie Actief*, BVJ: *Biologie voor jou*, ZDWD: *Zo doen we dat*.

element nr.	*BA*	*BA + ZDWD*	*BVJ*	*BVJ + AVE*	*Nectar*	*totaal*	*impliciet*	*expliciet*
1	16	34	11	13	23	70	56	14
2	0	0	0	1	1	2	2	0
3	30	44	17	19	31	94	85	9
4	5	5	2	3	8	16	12	4
5	3	3	0	2	6	11	11	0
6	2	4	17	19	7	30	17	13
7	1	1	0	0	1	2	1	1
8	2	6	1	1	11	18	9	9
9	4	6	10	10	6	22	15	7
10	13	15	12	13	15	43	25	18
11	3	3	2	7	5	15	13	2
12	0	0	1	2	2	4	4	0
13	1	1	0	6	4	11	10	1
14	4	4	6	8	3	15	10	5
15	0	0	0	1	0	1	1	0
16	0	0	0	0	0	0	0	0
17	0	0	0	1	0	1	1	0
18	2	17	7	11	7	35	25	10
19	0	0	0	0	1	1	1	0
20	0	0	0	0	0	0	0	0
21	0	4	0	0	4	8	5	3
22	1	1	1	1	2	4	4	0
23	0	0	0	0	0	0	0	0
totaal aantal	87	148	87	118	137	403	307	96
waarvan expliciet (%)	25,3	26,4	25,3	18,6	25,5	23,8		
verdeling (%)	*BA*	*BA + ZDWD*	*BVJ*	*BVJ + AVE*	*Nectar*	*totaal*	*impliciet*	*expliciet*
informatie	41,4	42,6	25,3	27,1	3,6	24,8	21,8	34,4
opgaven	35,6	20,9	50,6	39,0	63,5	40,7	45,3	26,0
uitwerkingen	11,5	6,8	20,7	31,4	10,9	15,4	17,3	9,4
handleiding	11,5	29,7	3,4	2,5	21,9	19,1	15,6	30,2

hoofdzakelijk in de opgaven aan de orde. De bijdrage van *Zo doen we dat* is vooral via de docentenhandleiding en die van *Algemene vaardigheden exact* via de uitwerkingen.

De elementen die het vaakst aan de orde komen zijn min of meer dezelfde als die de docenten noemen (par. 3.2): onderzoeksvraag (1) met aansluitende conclusie (18) en hypothese (3) met voorspeling (6). De blanco (10) scoort ook hoog, terwijl 'constant houden van variabelen' (9) iets vaker voorkomt dan het benoemen van de onafhankelijke en afhankelijke variabelen (8). De methodes verschillen enigszins in hun accenten. De notie dat een conclusie op de onderzoeksvraag moet aansluiten komt slechts twee keer voor in *Biologie Actief* maar juist vaak in *Zo doen we dat*. *Biologie voor jou* vindt de voorspelling erg belangrijk en *Nectar* is de methode die de onafhankelijke en afhankelijke variabele benadrukt. Element 21, met de begrippen validiteit en betrouwbaarheid, komt weinig aan de orde: vier keer in *Zo doen we dat* en vier keer in *Nectar*.

Alle methodes stellen de eisen voor goed onderzoek ongeveer drie keer zo vaak impliciet als expliciet aan de orde. Impliciet komen ze het vaakst in opgaven voor, expliciet hoofdzakelijk in informatieve teksten of in de docentenhandleidingen. Met name het lage aantal keren dat de elementen expliciet in de uitwerkingen staan, valt op (9,4% van 96 = 9 keer).

De mate waarin de elementen die het vaakst voorkomen geëxpliciteerd worden loopt nogal uiteen. Waar 1 en 18 dicht bij het gemiddelde van een kwart zitten, worden de eisen waar een hypothese aan moet voldoen slechts in minder dan een tiende van de keren dat een hypothese aan de orde is geëxpliciteerd. De elementen 8, 9 en 10 worden juist vaker dan gemiddeld expliciet aan de orde gesteld, respectievelijk in 50, 32 en 40 procent van de gevallen.

- *Betekenis van de resultaten*

Wat betekent deze analyse nu voor de tweede vraag die aan het begin van dit hoofdstuk is gesteld: leren ze het niet al? Daarvoor is een norm nodig om de resultaten tegen af te wegen. Zo'n norm is niet makkelijk te geven, maar kan bijvoorbeeld gevonden worden in een minimum aantal keren dat verschillende aspecten van onderzoek doen in de loop van de drie jaren al dan niet expliciet aan de orde komen in een methode. Appelman (2003) formuleert als norm dat per hoofdstuk van een methode minstens één vraagstuk zou moeten voorkomen waarbij gemiddeld een derde van de elementen aan de orde zou moeten zijn. Het aantal hoofdstukken is in *Biologie Actief, Biologie voor jou* en *Nectar* respectievelijk 23, 18 en 26. Dit lijkt een redelijke norm. In tabel 3.5 zijn de gevonden aantallen, gegroepeerd per fase van het onderzoek, naast die norm[9] gezet. Alleen voor de eerste fase van onderzoek doen – onderzoeksvraag, hypothese en voorspelling – blijken alle drie de methodes de norm te halen; vooral het slot – conclusie en evaluatie – komt er bekaaid af. Bovendien zijn deze cijfers geflatteerd want ook de getallen van de docentenhandleidingen zijn meegeteld. Als alleen gekeken wordt naar de keren dat de elementen expliciet aan de orde komen, blijven alle geanalyseerde methodes voor alle aspecten van onderzoek doen ver onder de maat.

Uitgerekend per element betekent de norm dat elk element in *Biologie Actief, Biologie voor jou* en *Nectar* respectievelijk acht, zes en negen keer aan de orde zou moeten komen. De elementen 1, 3 en 10 zijn de enige die dat in elke methode halen; in *Biologie voor jou* geldt het ook voor de elementen 6, 9 en 18 (tabel 3.4).

Ook al is de aangelegde norm een subjectieve keus en voor discussie vatbaar, het beeld dat uit de analyse van deze drie methodes naar voren komt, is niet florissant voor het leren bewaken van de kwaliteit van onderzoek. De ondersteuning van leren onderzoeken door schoolboeken blijkt dus niet sterk. Docenten die er serieus werk van maken, voorzien de leerlingen dan ook vaak van een speciaal op onderzoek doen gerichte instructie, al dan niet op

[9] De norm is als volgt berekend: (aantal hoofdstukken x aantal elementen) / 3

Tabel 3.5 Aantallen keren dat aspecten van onderzoek doen in termen van begrip van bewijs in verschillende biologiemethodes voor bovenbouw vwo aan de orde komen, impliciet dan wel expliciet, en een norm daarvoor. AVE: *Algemene vaardigheden exact*; BA: *Biologie Actief*, BVJ: *Biologie voor jou*, ZDWD: *Zo doen we dat*; im: impliciet; ex: expliciet.

	BA			*BA + ZDWD*		*BVJ*			*BVJ + AVE*		*Nectar*		
	im	*ex*	*norm*	*im*	*ex*	*im*	*ex*	*norm*	*im*	*ex*	*im*	*ex*	*norm*
Onderzoeksvraag, hypothese & voorspelling (1-6)	46	10	46	76	14	36	11	36	46	11	61	15	52
Variabelen, steekproef & metingen (7-17)	16	12	84	20	16	24	8	66	41	8	28	19	95
Conclusie & evaluatie (18-23)	3	0	46	13	9	5	3	36	9	3	13	1	52
Alle elementen	65	22	176	109	39	65	22	138	96	22	102	35	199

papier, zie bijvoorbeeld het boek *Practicum biologie 2* van Griethuysen-Elfers e.a. (1991). Dat de methodes niet alles vertellen over wat er op scholen gebeurt, wordt in dat boek duidelijk. Vandaar dat ook gekeken is naar wat de leerlingen ervan terecht brengen; dat is het onderwerp van de volgende paragraaf.

3.4 Onderzoek doen in leerlingverslagen

Om te zien in hoeverre leerlingen in hun onderzoeksverslagen blijk geven te beschikken over begrip van bewijs, zijn leerlingverslagen geanalyseerd.[10] Dat is in eerste instantie gebeurd met verslagen van een school in Uithoorn. Om het beeld breder te krijgen, zijn daar later verslagen van vijf andere scholen aan toegevoegd.

Omdat het een verkennend, exploratief onderzoek naar leerlingverslagen betrof zijn eerst scholen benaderd waarmee al contacten bestonden en waarvan bekend was dat er ervaring was met eigen onderzoek door leerlingen. Daarnaast is geprobeerd van meer scholen verslagen te verzamelen. Dat bleek heel moeilijk te zijn. Er werden vijftien scholen benaderd, waarvan slechts twee scholen konden en wilden meewerken. Veel van de andere scholen lieten hun leerlingen geen experimenteel onderzoek uitvoeren, maar literatuuronderzoek. Andere scholen wilden de leerlingverslagen niet ter beschikking stellen voor dit onderzoek.

Zoals in tabel 3.6 te zien is, is op de meeste van de scholen waarvan verslagen zijn geanalyseerd al vele jaren ervaring met eigen onderzoek door leerlingen. Dat betekent dat de steekproef van geanalyseerde verslagen niet de situatie weerspiegelt van het gehele biologieonderwijs in Nederland, maar wel een beeld geeft van wat er op sommige scholen bereikt wordt.

[10] De analyses zijn uitgevoerd in het kader van onderzoeksstages waarvan elders uitgebreid verslag is gedaan (Sliepen, 2002; Ferket, 2003).

Tabel 3.6 Scholen waarvan leerlingverslagen zijn geanalyseerd.

school in	*aantal jaren ervaring met leerlingonderzoek*	*klas waarin onderzoek plaatsvindt*	*aantal verslagen geanalyseerd*
Amstelveen	3	5-6	9
Castricum	>15	5 of 6	20
Oegstgeest	8	5	45
Schagen	10	5-6	9
Wassenaar	10	5-6	3
Uithoorn	20	6	83

Eerst worden de scholen en enkele kenmerken van het leerlingonderzoek kort beschreven, daarna wordt de methode van analyse uitgelegd en worden de resultaten gepresenteerd. We besluiten met het bespreken van de betekenis van de resultaten.

- *De scholen*

Op de school in Uithoorn bestaat een jarenlange traditie van eigen onderzoek door leerlingen. In de loop van de schooljaren doen de leerlingen al verschillende kleine onderzoeken. Het onderzoek waarvan de verslagen geanalyseerd zijn, vond plaats in het kader van een praktisch schoolonderzoek (voor de vernieuwing van de tweede fase) of een profielwerkstuk in klas 6 vwo. Zij deden dat meestal individueel. De onderwerpen waren zeer divers, van onderzoek in de natuur in de schoolomgeving tot meer psychologisch getinte onderwerpen. Sinds twintig jaar zijn alle verslagen daarvan bewaard, ten tijde van het onderzoek waren dat ongeveer 560 verslagen. Uit die verzameling zijn aselect 83 verslagen getrokken en geanalyseerd.

Op de school in Amstelveen heeft het leerlingonderzoek pas vorm gekregen door de invoering van de vernieuwde tweede fase. De negen onderzoeken beschrijven vooral hypothesetoetsend onderzoek, met onderwerpen als vissen en appels. De leerlingen voeren het eigen onderzoek uit in 5 vwo. Dit doen ze alleen of met een andere leerling samen.

De school in Castricum doet al vijftien tot twintig jaar aan leerlingonderzoek, voor de invoering van de vernieuwde tweede fase in de vorm van 'eigen experimenteel onderzoek', daarna is dat EXO kleiner in omvang geworden, omdat er ook nog een profielwerkstuk gemaakt moet worden. In 4 vwo gaan de biologiedocenten met de leerlingen een dag naar Artis in Amsterdam, waar alle leerlingen een gedragsonderzoek uitvoeren. Daarna kunnen de leerlingen kiezen of ze hun eigen onderzoek hierop laten volgen, of dat ze zelf een ander onderwerp uitkiezen. De onderwerpen van de verslagen die zijn beoordeeld zijn de dierentuinbewoners of geheel andere onderwerpen, zoals bacteriën en tuinkers of vegetatieonderzoek in de Castricumse duinen. De leerlingen schrijven de verslagen in vwo 5 of 6 en doen dit ofwel alleen of in groepjes met een grootte van maximaal vier leerlingen. Er waren twintig verslagen beschikbaar om te analyseren.

Op de school in Oegstgeest is al sinds ongeveer acht jaar sprake van eigen onderzoek door de leerlingen. Aan het eind van 5 vwo gaan de leerlingen drie dagen naar het nabij gelegen natuurgebied Meyendel, om daar veldwerk te doen. Als onderwerp hebben de meeste profielwerkstukken onderdelen van het ecosysteem, zoals mieren, meeuwen of de vegetatie in de duinen. De onderzoeken zijn uitgevoerd door groepjes tot drie personen of soms door een leerling alleen. Er waren 45 verslagen beschikbaar om te analyseren.

Op de school in Schagen vindt er sinds ongeveer tien jaar leerlingonderzoek plaats in het schoolvak biologie. De negen profielwerkstukken die zijn beoordeeld, zijn vooral hypothesetoetsende, experimentele onderzoeken naar bacteriën of voedselbederf.

De biologiesectie van de school in Wassenaar doet al sinds ongeveer tien jaar aan eigen onderzoek door de leerlingen. Het eigen onderzoek wordt gezien als een soort bekroning op het vwo. Het gaat meestal om hypothesetoetsende onderzoeken, met huis-tuin-en-keuken onderwerpen als kippen, kroos en aardappelen. De leerlingen werken in groepjes van twee en voeren het onderzoek uit in klas 5 en 6. Helaas zijn er slechts drie verslagen (het gaat hier om profielwerkstukken) beoordeeld, omdat de oudere verslagen verloren zijn gegaan bij de verbouwingen van de school.

- *Het leerlingonderzoek*

Op al deze scholen krijgen de leerlingen veel vrijheid in de keus van hun onderwerp en de opzet van hun onderzoek. Ze krijgen daarbij een meer of minder uitgebreide handleiding. De meest uitgebreide is die in Uithoorn (Dirks, 1988), waarin de eisen waaraan het onderzoek moet voldoen zijn omschreven. De docenten in Uithoorn beschouwen het praktisch schoolonderzoek c.q. profielwerkstuk als de eindfase van een traject van leren onderzoeken: 'Het geeft ook helemaal niet zozeer wat nou het onderwerp is, maar er zit gewoon een systematische manier van denken achter om problemen op te lossen. Onderzoek doen, of ze nou psychologisch onderzoek doen, geologisch onderzoek of wat dan ook, het zit hem in de manier van denken.' De docent in Oegstgeest laat het eigen onderzoek heel open: 'Ze mogen zelf heel erg bepalen welk onderzoek ze willen gaan doen, het is alleen beperkt tot het gebied. Dus het is in de duinen, in Meyendel, en ze krijgen daarvoor een heleboel boeken en tijdschriften en daaruit mogen ze kiezen en ze mogen ook zelf iets heel nieuws nemen, wat daar nog niet eerder onderzocht is.' De leerlingen krijgen wel een instructie om een proefopzet te maken volgens een bepaald model. Op de scholen uit Amstelveen en Castricum lijkt het eigen onderzoek van de leerlingen qua openheid vergelijkbaar te zijn. De docent in Wassenaar: 'Wij laten het leerling-onderzoek echt behoorlijk open. Ze moeten zelf het onderwerp bedenken, ze moeten zelf een vraagstelling bedenken. Wij proberen dus steeds te zeggen van 'in ieder geval moet er een hypothese in', (…) een eenvoudige hypothese van 'ik denk dat het zus of zo zal gaan', 'ik vermoed wel dat het zo en zo is', dat vind ik dan wel genoeg.' Iets minder open is het onderzoek in Schagen: 'Ze hebben vaak geen idee wat ze willen, of ze komen met iets waar ze niets mee kunnen. We begeleiden ze zó dat het onderzoek in ieder geval iets oplevert voor ze. Natuurlijk is niets opleveren ook een resultaat, maar dat is niet leuk voor ze. In zoverre sturen we ze behoorlijk.'

Alle docenten proberen hun leerlingen op het goede pad te houden in het onderzoek, zonder het voor te zeggen, het zijn immers de leerlingen die het onderzoek moeten uitvoeren en niet de docent. De docent in Schagen: 'Als het gebeurt, dan geef ik aan welke richting ze opgaan en welk risico ze lopen. Dit uit zich meestal in het veranderen of het laten vallen van een deelvraag, of het extra opnemen van een andere deelvraag.'

De docent in Wassenaar discussieert met de leerlingen als er niet voldaan is aan de eisen voor goed onderzoek: 'Met 6-vwo-leerlingen kan dat ook heel goed, ze zijn enthousiast genoeg om dat op te pakken.' Andere docenten (Amstelveen, Oegstgeest) geven aan dat er te weinig tijd is om alle punten die in het verslag horen aandacht te geven bij ieder groepje leerlingen. De eerste onderdelen, zoals de onderzoeksvraag en hypothese, komen nog wel aan bod bij het inleveren van het plan. Punten als resultaten en de discussie of de uiteindelijke conclusie kunnen niet uitgebreid worden behandeld.

- *Aanpak van de analyse*

De verslagen zijn geanalyseerd op de 23 elementen van begrip van bewijs. Dat wil zeggen dat voor elk element in één van de volgende vier categorieën gescoord is:
(1) *niet van toepassing*: niet elk element is op elk onderzoek van toepassing, zo zijn bijvoorbeeld in een beschrijvend onderzoek de eisen voor de hypothese niet van toepassing en wordt

niet in elk onderzoek een steekproef getrokken; sommige elementen zijn wel altijd van toepassing, daar kan dus niet in deze categorie gescoord worden;
(2) *aan voldaan*: uit het verslag blijkt dat het onderzoek voldoet aan de eis zoals verwoord in het betreffende element, bijvoorbeeld het eenduidig formuleren van een onderzoeksvraag;
(3) *beargumenteerd niet aan voldaan*: het onderzoek voldoet weliswaar niet aan de eis, maar in het verslag wordt uitgelegd waarom dat niet het geval is, bijvoorbeeld als door omstandigheden een kleine steekproef is getrokken;
(4) *niet beargumenteerd niet aan voldaan*: het onderzoek voldoet niet aan de eis zoals verwoord in het betreffende element en er wordt ook niet uitgelegd waarom dat zo is.

Score in categorie 2 of 3 wordt beschouwd als 'begrip van bewijs', omdat ook als niet aan een eis voldaan is, maar er beargumenteerd wordt waarom dat zo is, de leerlingen blijkbaar begrijpen dat er aan die eis voldaan zou moeten zijn. De mate van begrip van bewijs wordt dan bepaald door de breuk (score 2 + score 3)/(score 2 + score 3 + score 4), uitgedrukt in een percentage. Bij een aantal elementen is sprake van een meervoudige eis, bijvoorbeeld element 1: een onderzoeksvraag bevat alleen ondubbelzinnige termen en formuleringen en is voldoende specifiek en afgeperkt. Voor die drie elementen (1, 3 en 14) worden meer scores genoteerd en gemiddeld.

- *Resultaten van de analyse*

In tabel 3.7 staat per element het percentage vermeld dat in de verslagen van de verschillende scholen is gevonden. Dan blijken er opvallende overeenkomsten en verschillen te bestaan. Allereerst is het gemiddelde van alle elementen samen voor bijna alle scholen vrijwel gelijk. Alleen Uithoorn scoort 70 procent tegen ongeveer 60 procent op de andere scholen. Het verschil is met betrekking tot de elementen rond variabelen en metingen (7-17) zelfs nog iets groter, terwijl het met betrekking tot de elementen rond onderzoeksvraag en hypothese minder groot is. Daar vallen de scores met name in Oegstgeest en Schagen lager uit. De scores voor 'conclusie & evaluatie' liggen beduidend lager, maar ook daar is er een verschil tussen Uithoorn enerzijds en de andere scholen anderzijds. Een verklaring daarvoor ligt wellicht in het feit dat Uithoorn de enige school is waar het onderzoek van de leerlingen geheel in klas 6 plaatsvindt en het is bovendien de school met de langste traditie in het doen van onderzoek.

Voor een aantal elementen wordt er hoog gescoord. Dat zijn vooral de al eerder genoemde elementen rond onderzoeksvraag en hypothese (1, 3, 4), variabelen (7, 8) en conclusie (18). Het onderscheid tussen beschrijvend en hypothesetoetsend onderzoek (2), tussen een correlatie en een causaal verband (19) en tussen een oorzakelijke en een functionele verklaring (20) scoren erg laag, evenals het vergelijken met bestaande theorieën (23). De overige elementen zitten daar tussen in, ook het element met betrekking tot validiteit en betrouwbaarheid (21). Per element lopen de scores tussen de scholen soms ook sterk uiteen. Zo werd er in Castricum in alle verslagen een voorspelling aangetroffen (element 6), terwijl dat in Amstelveen, Schagen en Wassenaar nooit het geval was. Dit verschil is direct te verklaren uit de instructie die de leerlingen bij hun onderzoek kregen. De docent in Castricum vindt een voorspelling een essentiële stap in een hypothesetoetsend onderzoek, de docenten op de andere scholen hechten daar minder aan of vinden het zelfs een loze stap. De docent in Castricum vindt een blanco (10) eveneens erg belangrijk, maar waarom het onder controle houden van andere variabelen dan juist veel minder scoort, wordt niet duidelijk. Niet voor alle verschillen is dus een directe verklaring voorhanden; dit is ook niet in detail aan de docenten gevraagd. Het lijkt echter waarschijnlijk dat de invloed van de accenten die een school of docent legt in instructie of begeleiding, groot is.

Het gemiddelde van de resultaten is op twee manieren berekend, per verslag en per school. In de eerste manier is elk verslag even zwaar meegewogen en daardoor weegt Uithoorn ongeveer even zwaar als de andere samen (83 om 86 verslagen). In de tweede manier

Tabel 3.7 Percentages begrip van bewijs voor de 23 elementen in onderzoeksverslagen van leerlingen van zes scholen. 'Begrip van bewijs' wil zeggen dat aan een van toepassing zijnd criterium is voldaan of beargumenteerd is waarom er niet aan is voldaan.

element	*Amstelveen (N = 9)*	*Castricum (N = 20)*	*Oegstgeest (N = 45)*	*Schagen (N = 9)*	*Wassenaar (N = 3)*	*Uithoorn (N = 83)*	*gemiddelde (N=169)*	*gem. schoolpercentage*
1	89	77	64	88	89	80	77	81
2	0	0	0	0	0	1	1	0
3	89	81	70	87	92	91	84	85
4	100	84	100	100	100	100	98	97
5	44	32	52	14	67	29	35	40
6	0	100	12	0	0	60	45	29
1-6	*70*	*70*	*55*	*64*	*73*	*72*	*67*	*67*
7	56	50	82	100	100	100	80	81
8	100	100	80	83	100	87	89	92
9	22	39	50	63	100	82	65	59
10	86	100	67	57	0	40	52	58
11	78	55	71	78	33	94	80	68
12	75	67	68	50	67	86	77	69
13	56	80	78	78	0	83	78	62
14	31	60	63	67	80	62	61	60
15	100	33	58	0	0	83	56	46
16	100	13	44	0	0	100	75	43
17	0	33	43	0	0	92	78	28
7-17	*60*	*58*	*67*	*66*	*53*	*78*	*71*	*64*
18	100	95	93	78	100	96	95	94
19	50	71	0	17	67	19	26	37
20	0	0	0	0	0	2	1	0
21	33	35	31	67	67	74	55	51
22	11	20	27	22	33	79	51	32
23	33	15	27	0	0	33	27	18
18-23	*39*	*41*	*43*	*34*	*44*	*54*	*48*	*43*
1-23	*60*	*60*	*57*	*58*	*59*	*70*	*64*	*61*

zijn de scores van de scholen gemiddeld, waardoor Uithoorn (83 verslagen) evenveel meeweegt als Wassenaar (drie verslagen). Omdat we denken dat de invloed van de (instructie en begeleiding van de) docenten op de manier waarop leerlingen hun onderzoek inrichten, uitvoeren en verslaan, groot is, denken we dat de gemiddelden van de schoolscores een beter beeld geven van wat er op deze zes scholen aan begrip van bewijs bereikt wordt.

- *Betekenis van de resultaten*

De vraag is hoe deze resultaten te beoordelen. Daarvoor is eigenlijk een norm nodig voor een voldoende resultaat, een cesuur; bij welk percentage leerlingen is er voldoende ge-

Tabel 3.8 Aantallen elementen dan wel verslagen die voldoende scoren bij verschillende cesuren.

Grens onvoldoende – voldoende (%)	*Aantal elementen met voldoende score (N = 23)*	*Aantal verslagen met voldoende score (N = 169)*
49 / 50	13	146 (86%)
59 / 60	10	114 (67%)
66 / 67	8	78 (46%)
74 / 75	6	34 (20%)

scoord voor een element en bij welk percentage elementen heeft een leerling voldoende begrip van bewijs getoond? Een cesuur is in principe altijd arbitrair, het is een subjectieve keuze. Daarom zijn in tabel 3.8 de resultaten (gebaseerd op de schoolgemiddelden) weergegeven bij verschillende cesuren. De strenge grens waarbij 75 procent van de leerlingen een element moet beheersen, wordt door slechts zes elementen bereikt, te weten de elementen 1, 3, 4, 7, 8, en 18. Bij de soepele grens van 50 procent zijn het dertien van de 23 elementen. Voor leerlingen geldt vaak dat ze ruim de helft van de te behalen punten moeten halen om een voldoende te halen. Bij die grens haalt tweederde van de leerlingen dat.

Op het eerste gezicht lijkt de situatie redelijk positief. Daar zijn echter enkele kanttekeningen bij te plaatsen. Ten eerste betreft het eindproducten; bij alle scholen heeft min of meer intensieve begeleiding plaatsgevonden bij de totstandkoming ervan. Een slecht geformuleerde onderzoeksvraag, bijvoorbeeld, zal van commentaar voorzien zijn en zijn bijgesteld tot hij volgens de docent goed genoeg was. Ten tweede, zoals we in de inleiding van deze paragraaf al vermeldden, is er met deze exploratie geen representatief onderzoek gedaan naar wat leerlingen aan begrip van bewijs tonen in hun verslagen. De resultaten weerspiegelen niet de situatie van het gehele biologieonderwijs in Nederland, maar geven slechts een beeld van wat er op sommige scholen, die veel belang hechten aan onderzoek doen, bereikt wordt.

De conclusie lijkt gerechtvaardigd dat op scholen waar serieus gewerkt wordt aan onderzoek door leerlingen een beperkt aantal elementen van begrip van bewijs in een meerderheid van de leerlingverslagen aangetroffen kan worden. De vraag blijft echter of dat voldoende is om te kunnen stellen dat leerlingen de kwaliteit van hun onderzoek weten te bewaken. Als een onderzoek aan tweederde van de kwaliteitseisen voldoet, is het dan een goed onderzoek?

3.5 Keuzes

De verkenning van de literatuur laat zien dat aandacht schenken aan de manier waarop nieuwe kennis wordt verkregen via onderzoek wezenlijk is voor het biologieonderwijs. Even wezenlijk is het om daarbij de kwaliteitscriteria voor onderzoek aan de orde te stellen. Docenten onderschrijven de opvatting dat begrip van bewijs van belang is voor leerlingen en geven aan daar ook aandacht aan te besteden in hun onderwijs, al vinden ze niet alle elementen even belangrijk. De ondersteuning van leren onderzoeken door schoolboeken blijkt niet sterk en al geven leerlingverslagen een redelijk positief beeld, ook daar lijkt verbetering zinvol.

Een onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek zou ons inziens eigenlijk alle elementen van begrip van bewijs moeten omvatten. Als de uitwerking ervan echter aan randvoorwaarden gebonden is, zoals een beperkte tijdsspanne in één leerjaar, is het nodig om prioriteiten te stellen en een keus te maken welke elementen wel en welke niet aan de orde gesteld zullen worden.

Argumenten voor die keus kunnen ten eerste aan de officiële leerdoelen ontleend worden. Die volgen in grote lijnen de stappen van de empirische cyclus: vraagstelling, hypothese, voorspelling, waarnemingen, conclusie, evaluatie. Deze stappen zijn als deelvaardigheden geformuleerd. De huidige trend om minder vanuit deelvaardigheden uit te gaan, maar meer van kennis, vaardigheden en houdingen omvattende competenties is een argument om een samenhangende selectie uit de elementen te maken die de essentie van de kwaliteit van onderzoek weerspiegelt. Een tweede bron voor argumenten voor een selectie wordt gevormd door de mening en de praktijk van docenten. De elementen waar de docenten het meeste belang aan toekennen, sporen met de officiële leerdoelen: onderzoeksvraag, hypothese, variabelen, blanco, conclusie en de notie van validiteit en betrouwbaarheid. Dit zijn ook de elementen waar zij in hun eigen praktijk de meeste aandacht aan geven. Er zijn echter ook elementen die zij belangrijk vinden waar ze minder aandacht aan geven (zie fig. 3.5). Een mogelijkheid voor prioritering is om een te ontwikkelen onderwijsleerstrategie vooral te richten op de elementen waarbij het verschil tussen toegekend belang en gegeven aandacht het grootst is. Daar blijft de praktijk immers achter bij de wens, dus daar zou een nieuwe onderwijsleerstrategie het meeste zoden aan de dijk kunnen zetten.

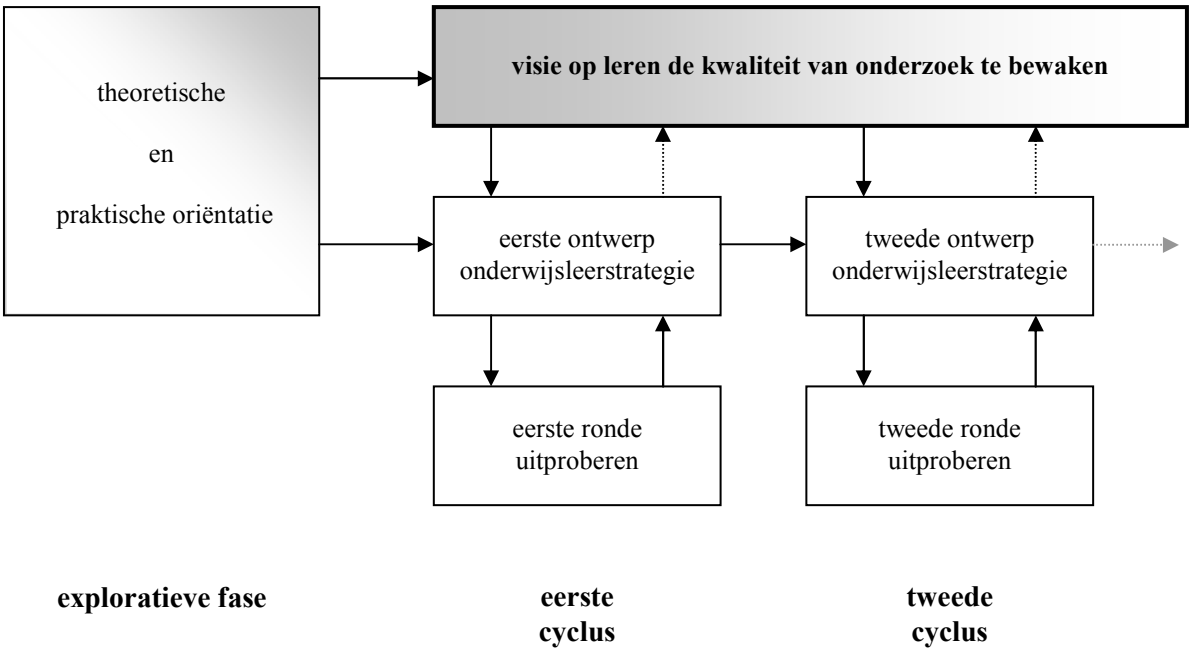
In de klankbordgroep van docenten is deze mogelijkheid overwogen, maar toch is er niet voor gekozen. De inschatting was dat docenten meer behoefte hebben aan een onderwijsleerstrategie voor de in hun ogen belangrijkste elementen, omdat zo'n strategie nog niet voorhanden is. Pas als de meeste docenten het onderwijzen van de belangrijkste elementen goed in de vingers hebben, zou een strategie voor de andere elementen in het onderwijs aanslaan.

Dat betekent dat de eerste bron van argumenten de doorslag heeft gegeven: een samenhangende selectie van elementen die de essentie van de kwaliteit van onderzoeken omvatten. Daarbij is gekozen voor hypothesetoetsend onderzoek. De selectie geeft daarvan de kern weer: de hypothetisch-deductieve aanpak.

Uiteindelijk is dus gekozen voor de volgende vier aspecten van de kwaliteit van onderzoek met de daarbij horende elementen van begrip van bewijs.

- *onderzoek geeft een duidelijk antwoord op een heldere vraag*: dat betekent dat enerzijds de onderzoeksvraag goed geformuleerd moet zijn, dus eenduidig, specifiek en afgeperkt (element 1) en anderzijds de conclusie er een antwoord op moet geven (element 18);
- *hypothesetoetsend onderzoek onderwerpt een mogelijke verklaring, een mogelijk antwoord aan een kritische toets*: er wordt dus een hypothese geformuleerd (element 3) zodanig dat hij toetsbaar is (4), die bovendien een van meer hypotheses is (5), waaruit een voorspelling wordt afgeleid (6);
- *in een onderzoek is duidelijk wat onderzocht wordt en wat niet*: primair gaat het dan om onafhankelijke en afhankelijke variabelen (8) en om het controleren van andere variabelen (9 en 10); vanwege de inherente variabiliteit van biologische verschijnselen en organismen is steekproeftrekking (14) hierbij eveneens belangrijk;
- *onderzoek is valide en betrouwbaar,* dat wil zeggen dat mogelijke bronnen van fouten worden uitgesloten en de hypothese het onderzoek niet zal doorstaan, tenzij hij juist is; dit overkoepelende idee van kwaliteit wordt verwoord in element 21.

De elf elementen van begrip van bewijs 1, 3, 4, 5, 6, 8, 9, 10, 14, 18 en 21 vormen daarmee de inhoud van de onderwijsleerstrategie en de concrete uitwerking daarvan in een serie lessen voor klas 5 vwo. Die uitwerking wordt beschreven in hoofdstuk 5. Daaraan voorafgaand wordt in hoofdstuk 4 een onderwijskundige en didactische basis gelegd voor de manier waarop leerlingen begrip van bewijs zouden kunnen leren.



4 Kwaliteit leren bewaken

4.1 Denken, doen en reflecteren
4.2 Uitgangspunten voor twee onderwijsleerstrategieën: expliciet en impliciet

'It is the adherence to criteria which is at the centre of thinking critically, and giving attention to explicating and applying the relevant criteria must be at the centre of attempts to foster critical thinking in science.'
Sharon Bailin (2002, p. 373)

Hoofdstuk 4 Kwaliteit leren bewaken

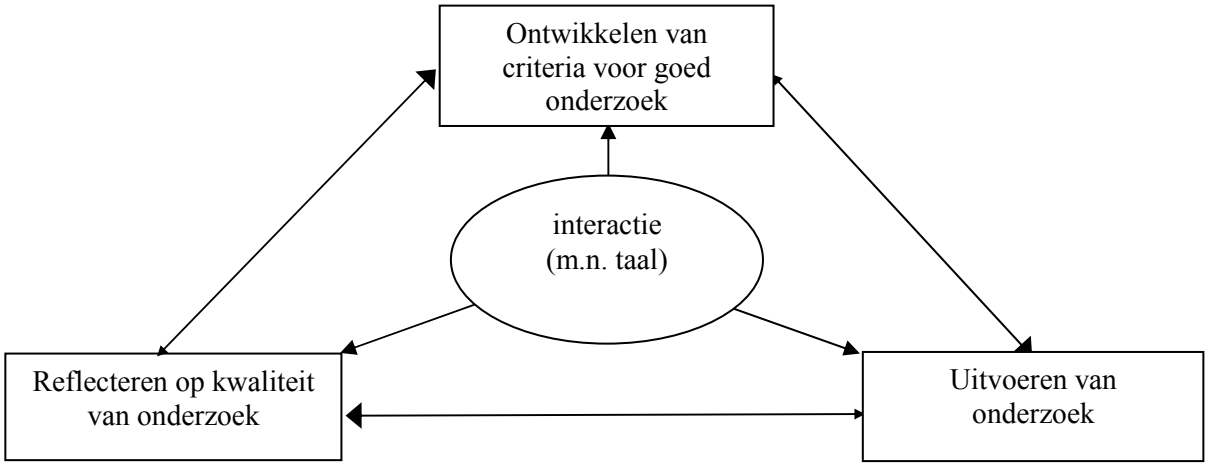
In hoofdstuk 1 is beargumenteerd dat leerlingen in het vwo zouden moeten leren onderzoek te doen, omdat dat een wezenlijk onderdeel vormt van de wetenschappelijke en maatschappelijke praktijk waar ze op voorbereid worden. In hoofdstuk 2 is die praktijk verder omschreven en is uitgelegd welke rol kwaliteitsbewaking daarin speelt. Aan het eind van hoofdstuk 3 hebben we vervolgens beargumenteerd welke aspecten van kwaliteitsbewaking leerlingen in het vwo zouden moeten leren beheersen.

Als leerlingen de kwaliteit van onderzoek moeten kunnen bewaken zullen daar naar verwachting de volgende componenten een rol bij spelen:

- ze moeten weten aan welke kwaliteitseisen onderzoek moet voldoen en hoe deze eisen bij een concreet onderzoek moeten worden toegepast; dat vraagt om begrip van bewijs, om daarmee te leren nadenken over kwaliteit van onderzoek;
- ze moeten de kwaliteitscriteria kunnen toepassen op hun eigen onderzoek; immers, ze moeten goed onderzoek leren doen en dat leer je volgens gangbare opvattingen vooral door onderzoek te doen;
- ze moeten, als zij onderzoek hebben uitgevoerd, kunnen beoordelen of de criteria wel goed zijn toegepast, met andere woorden: zij moeten kunnen reflecteren op hun onderzoek en de vraag kunnen beantwoorden of het goed genoeg was.

Het gaat dus om denken over de kwaliteit van onderzoek, het uitvoeren van (kwalitatief goed) onderzoek en het reflecteren op de kwaliteit van onderzoek.

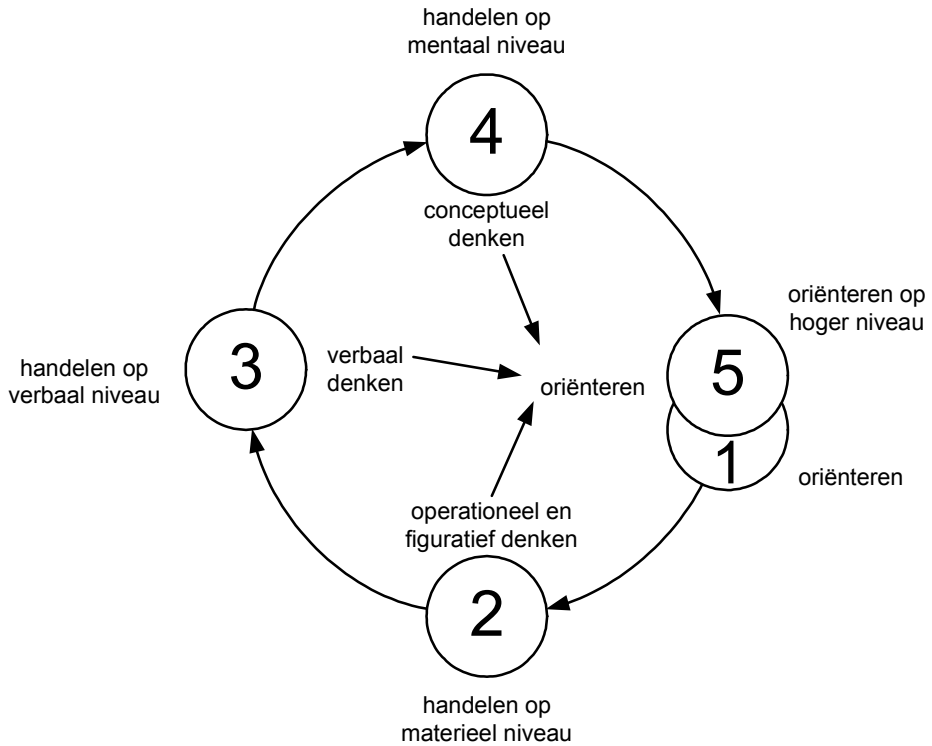
Dat is een complex geheel en als leerlingen dit moeten leren, kunnen we veronderstellen dat leerlingen dat niet in één keer zullen leren. De genoemde componenten zullen meer-



Figuur 4.1 Componenten van het leren bewaken van de kwaliteit van onderzoek. Hoewel de meest voor de hand liggende wijze van doorlopen met de klok mee is (aangegeven door grote pijlpunten), lijkt heen en weer gaan soms ook mogelijk (aangegeven met kleine pijlpunten).

malen moeten worden aangeboden. Bovendien, omdat leren een interactief proces is, zullen de leerlingen dat in interactie met anderen moeten doen. Daarbij worden betekenissen meestal verbaal uitgewisseld. De relatie tussen de componenten kan cyclisch worden afgebeeld, zie figuur 4.1. Hierbij is de gelijkenis met de driehoek 'biologie leren – over biologie leren – biologie doen' (figuur 3.1) niet helemaal toevallig. In beide driehoeken gaat het om inhoud (vakkennis resp. begrip van bewijs), ervaring (het doen van biologie resp. onderzoek) en procedure (methoden van de biologie resp. reflecteren op kwaliteit).

Het leren bewaken van de kwaliteit van onderzoek kan echter ook nog op een andere manier in een schema gezet worden, waarbij de nadruk ligt op de wijze waarop leerlingen het verwerven, op de wijze waarop het opgenomen wordt in hun mentale structuur (figuur 4.2). Dit schema is afgeleid van het schema waarmee Arievitch en Haenen (2005) de 'spiral formation of mental actions' volgens Galperin weergeven. Galperin werkte het concept van 'internalisatie' uit, een centraal begrip van de leertheorie van Vygotski (1978).[11] Het proces van internalisatie, zo toonde Galperin aan, gaat van het materiële niveau van de concrete objecten en handelingen, via het verbale niveau van (gesproken of gedachte) woorden voor objecten en handelingen naar het mentale niveau van beelden, associaties en begrippen. Het proces van het je eigen maken van begrippen vindt plaats in een spiraal die steeds weer via deze niveaus gaat. Na een oriëntatie, onder andere op wat je al weet en kan, gaat het leerproces vanuit het materiële niveau (doen van onderzoek) via het verbale niveau (praten over onderzoek) naar het mentale niveau (denken over onderzoek) en weer naar een hernieuwde oriëntatie, maar nu op hoger niveau. Daarna start de cyclus weer op het materiële niveau. De spiraal geeft de toenemende internalisatie en beheersing van een actie aan, waardoor de kwaliteit ervan ook toeneemt. Dat blijkt ook uit een gestaag toenemend vermogen van de lerende om zich te oriënteren, hetgeen leidt tot een steeds beter begrip en uitvoering van de actie. Stapsgewijs wordt de



Figuur 4.2 Spiraalsgewijs proces van verwerving van mentale acties volgens Galperin (naar Arievitch & Haenen, 2005).

[11] De leertheorie van Vygotsky wordt met verschillende namen aangeduid, 'sociocultural theory' of 'activity theory', in navolging van Van Oers (1998a) gebruiken we bij voorkeur 'cultuurhistorische theorie'.

leerling zich bewust van de 'ins and outs of an action', aldus Galperin volgens Arievitch en Haenen (2005). Dit proces wordt weergegeven met de pijlen in het midden van de figuur.

Het leerproces (figuur 4.2) begeeft zich dus op dezelfde niveaus als de beoogde niveaus waarop een leerling met de kwaliteitscriteria moet kunnen omgaan (figuur 4.1). Doel en middel gaan hand in hand. Voordat we dit kunnen vertalen in uitgangspunten voor een onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek, werken we deze niveaus – denken over onderzoek, onderzoek doen en reflecteren op onderzoek – verder uit.

4.1 Denken, doen en reflecteren

- *Denken*

'Denkt aleer gij doende zijt en doende denkt dan nog' hing boven het schoolbord in het lokaal waar ik als brugklasser de meeste lessen kreeg. Wie het had opgehangen weet ik niet, maar mijn mentor ('klassenleraar' heette het toen nog) vestigde er regelmatig de aandacht op. Het geeft het primaat aan zoals dat toen (in 1968) door velen gezien werd: het denken beheerst het doen en uiteindelijk gaat het erom dat de kennis in je hoofd komt.

De omschrijving in de inleiding bij dit hoofdstuk zou dat ook kunnen doen vermoeden: eerst moet je kennis nemen van de criteria voor goed onderzoek alvorens je ze kunt toepassen. Die invalshoek wordt ook bepleit door een aantal auteurs, waarvan Richard Gott de belangrijkste is (zie bijvoorbeeld Millar e.a., 1994; Gott & Duggan, 1995b; Gott & Roberts, 2003; Jones & Gott, 1998). Zij benadrukken dat procedurele kennis (het weten hoe) een aparte kennisbasis is die gelijkwaardig is aan vakinhoudelijke kennis. En omdat het kennis is, moet het volgens hen ook expliciet onderwezen worden. Dat betekent echter niet dat de criteria als reproduceerbare kennis aangeboden en overhoord moeten worden. Het gaat erom dat je ze kunt gebruiken. En tegelijkertijd worden de criteria ook juist tijdens het gebruik opnieuw geconstrueerd, zoals de cycli in de figuren 4.1 en 4.2 illustreren.

Het gebruik van criteria typeert 'kritisch' denken, tenminste volgens Bailin (2002). Zij problematiseert het veel gepropageerde idee (bijv. Adey & Shayer, 1990) dat kritisch denken uit vaardigheden en processen dan wel uit procedures zou bestaan. Bailin stelt dat er een norm mist: de norm die kritisch denken onderscheidt van onkritisch denken. Kritisch denken betekent het begrijpen en kunnen hanteren van criteria voor goed denken. Die kunnen verschillen in verschillende contexten, want kritisch denken heeft altijd betrekking op een bepaalde activiteit of situatie. Het interpreteren van een grafiek is bijvoorbeeld iets anders dan het interpreteren van een literaire tekst. Kritisch denken, het beheersen van criteria, houdt daarom ook in dat de denker herkent wanneer hij die criteria moet toepassen. Welke daarvan dan gebruikt worden, hangt erg van de specifieke activiteit of situatie af. Als belangrijkste in de natuurwetenschappelijke context noemt Bailin kennis van de criteria: begrippen als 'conclusie' en 'oorzaak en gevolg' en principes als 'testen van hypotheses'.

Hoe kunnen leerlingen dat leren? Bailin (2002) stelt dat het vasthouden aan criteria als de kern van kritisch denken betekent dat er expliciet aandacht geschonken moet worden aan het uitleggen en toepassen van de relevante criteria (zie ook het citaat aan het begin van dit hoofdstuk).

Er zijn in de loop van de tijd verschillende initiatieven geweest, geschoeid op verschillende leertheorieën, die zich specifiek richtten op het leren (kritisch) denken in en met de natuurwetenschappen, onder andere het project *Cognitive Acceleration through Science Education* (CASE; o.a. Adey & Shayer, 1990; Adey, Shayer & Yates, 1989), Lawson (1995) met zijn boek *Science teaching and the development of thinking* en het project *Thinking in Science*

Classrooms (Zohar, 1996, 2004). Doel van al deze initiatieven was en is het verhogen van de denkcapaciteiten van leerlingen, de 'hogere orde' denkvaardigheden of -strategieën. Daartoe zijn leeractiviteiten ontwikkeld die naast of in het normale onderwijsprogramma gebruikt worden en waarin verschillende strategieën aan de hand van natuurwetenschappelijke inhoud geoefend worden. De bedenkers van deze strategieën claimen dat ze behoorlijk effectief zijn (Adey & Shayer, 1990; Lawson, 2001; Zohar, 1996, 2004), maar aan de beweerde effectiviteit wordt ook getwijfeld (bijv. Leo & Galloway, 1996; Jones & Gott, 1998). Toch wordt nergens beweerd dat je géén activiteiten hoeft te ondernemen om kritisch denken bij leerlingen te stimuleren. Integendeel, hoewel er in de literatuur verschillende benaderingen en omschrijvingen van kritisch denken te vinden zijn, is de diversiteit kleiner als het gaat over de vraag hoe het geleerd zou kunnen worden, stellen Ten Dam en Volman (2002). Zij vatten het als volgt samen:

> Instructiekenmerken waarvan wordt gesteld dat ze kritisch denken bevorderen zijn: aandacht voor de (ontwikkeling van) epistemologische opvattingen van leerlingen, stimuleren van actief leren, gebruik van een probleemgestuurd curriculum en aanmoediging van interactie tussen leerlingen. Daarnaast benadrukken enkele auteurs het motiverende effect van het gebruik van praktijksituaties. Op kritisch denken wordt vooral een beroep gedaan bij heuristische, 'ill-defined' en complexe problemen. Tegenwoordig zijn de meeste onderzoekers het erover eens dat kritisch leren denken daarom moet plaatsvinden in een context van betekenisvolle, domeingebonden onderwerpen.
>
> In empirisch onderzoek naar instructievariabelen die bijdragen aan de ontwikkeling van kritisch denken, wordt het belang van een groot deel van deze kenmerken bevestigd. Dat geldt vooral voor de kenmerken die betrekking hebben op het stimuleren van een actieve betrokkenheid en inbreng van de leerlingen in het leerproces, zoals een intensieve interactie tussen leerlingen en docent en leerlingen onderling, leerlingen hun inzichten laten presenteren of verwoorden in een essaytentamen (in plaats van via multiple-choicetentamens). (p. 174)

Deze auteurs pleiten vervolgens voor een benadering waarin de lerende de beoogde kennis en vaardigheden actief moet verwerven en moet toepassen, wat betekent dat de leerlingen vooral zelf aan de slag moeten.

Dat brengt ons bij het werkwoord *doen*.

- *Doen*

'Je leert het pas echt door het veel te doen', was het adagium van de instructeur die mij leerde autorijden. Hij kon mij de theorie en de basisvaardigheden bijbrengen, maar om een goed chauffeur te worden, moest ik vooral veel kilometers maken. Daarmee was hij een exponent van de opvatting dat de praktijk de beste leermeester is, zeker waar het gaat om complexe vaardigheden.

Een metafoor die verwijst naar het beroepsonderwijs uit vervlogen tijden – dat archetypisch uitging van leren door doen – is die van de meester-gezel-relatie, in de onderwijskundige literatuur vaak aangeduid met *apprenticeship.* Eén van de grondleggers van deze benadering, Jean Lave (1997, p.19), formuleert het zo:

> Apprentices learn to think, argue, act, and interact in increasingly knowledgeable ways with people who do something well, by doing it with them as legitimate, peripheral participants.

Omdat het in het algemeen vormende onderwijs niet gaat om leren ten behoeve van een specifiek beroep, maar om manieren van denken en doen, wordt het meestal gespecificeerd tot *cognitive apprenticeship* (Collins, Brown & Newman, 1989). In tegenstelling tot traditioneel apprenticeship, gaat het er bij cognitive apprenticeship om de situatiegebonden kennis te verbreden tot meer situaties en zo de opgedane kennis te generaliseren voor een

range van toepassingscontexten (Hennessy, 1993). Natuurwetenschap is dan niet zozeer een kennisbestand, maar een activiteit, een praktijk (zie ook hoofdstuk 2).

Vanuit deze invalshoek bekeken wordt het leren bewaken van de kwaliteit van onderzoek een aspect van het ingeleid worden in de praktijk, de cultuur, van het natuurwetenschappelijk onderzoek. Hodson en Hodson (1998b) benoemen het onderwijs in de natuurwetenschappen dan ook expliciet als 'enculturation'. De drieslag die volgens Hodson (1992) bij science education hoort – learning science, learning about science, doing science – is in hoofdstuk 3 al uitgewerkt en spoort, zoals in de introductie van dit hoofdstuk opgemerkt, met zowel 'denken, reflecteren en doen' als met het mentale, het verbale en het materiële niveau.

Kennis wordt dus opgedaan in een (gesimuleerde) praktijk, waarbij er geen (of weinig) onderscheid is tussen weten en doen:

> The assumption that they can be separated ignores the ways in which situations structure cognition. Activities are not carried out with the explicit purpose of determining principles separate from acting in the world. (Hennessy, 1993, p. 21)

In een praktijk worden algemene principes dus nooit apart, buiten een context, geleerd.

Als we spreken van een praktijk van onderzoek doen, is het wel belangrijk een onderscheid te maken tussen de praktijk zoals die in de 'echte' wetenschap plaatsvindt, aan universiteiten, wetenschappelijke instituten en researchcentra, en de praktijk van onderzoek doen op school. In de eerste is de functie van onderzoek doen altijd het verkrijgen van kennis die nog niet bekend is, in de tweede gaat het meestal om kennis die wel al bekend is en kan bovendien het doel van de onderzoeksactiviteit zijn om bijvoorbeeld vaardigheden te leren. In de schoolpraktijk gaat het dus om een gedidactiseerde praktijk (zie par. 3.1) met voor een deel eigen gewoontes, regels en wijzen van spreken (Mortimer & Scott, 2003).

Zelf onderzoek doen lijkt echter wel een essentiële component van leren onderzoeken en practicum dus onontbeerlijk. In de natuurwetenschappelijke vakken op school bestaat een lange traditie van practicum doen, die echter gekenmerkt wordt door onhelderheid over doel en effectiviteit. Er bestaat dan ook een vrijwel even lange traditie van herdefiniëren en herorienteren van practicum (bijv. Hodson, 1993; zie voor een aardig overzicht het eerste hoofdstuk van Gott & Duggan, 1995a).

Concluderen dat zelf onderzoek doen een belangrijke component is bij het leren bewaken van de kwaliteit van onderzoek, betekent echter niet dat élke vorm van onderzoekspracticum daaraan bijdraagt. Als het gaat over manieren om de *concepts of evidence* te leren begrijpen en toepassen, bepleiten Gott en Duggan (1995a, 1995b, 1996) voor *investigations,* open onderzoek en projecten, omdat die de leerlingen in staat stellen de invloed te zien van verschillende elementen van begrip van bewijs: '… they will be required to synthesise all of these ideas into an overall solution' (Gott & Duggan, 1995b, p. 144) waarbij ze met 'ideas' doelen op de concepts of evidence. Om begrip van bewijs in de volle breedte aan de orde te laten komen, betogen zij, zijn grote, open opdrachten nodig.

De openheid van een opdracht aan leerlingen is echter een begrip dat tot verwarring kan leiden. Welke aspecten worden dan aan de leerlingen zelf overgelaten en welke worden door de docent bepaald? In *De bioloog als leraar* onderscheiden Falk e.a (1974) zes vrijheidsgraden van practicum, van vrijwel volledig gesloten, illustratieve kookboekpracticum tot een vrijwel geheel door de leerlingen te kiezen inrichting en uitvoering van een 'experimenteel practicum': 'Afgezien van een stukje beeldvorming met de klas en afbakening van het onderzoekveld zijn alle fasen open' (Falk e.a., 1974, p. 158). Lock (1990) onderscheidt zeven niveaus van openheid, aan de hand van een vijftal vragen:

- wie bepaalt het onderwerp, het terrein van onderzoek?

- wie formuleert het probleem, de onderzoeksvraag?
- wie doet de planning?
- wie beslist over de aanpak?
- wie interpreteert en evalueert de resultaten?

De meest open situatie wordt bereikt als het antwoord op al deze vragen 'de leerling' is (zie ook Schalk, 2005a).

Een grote, open opdracht om te leren de kwaliteit van onderzoek te bewaken vereist dat leerlingen zeggenschap hebben over veel aspecten van het onderzoek en daarin eigen keuzes kunnen maken. Het zelf uitvoeren is echter geen voldoende voorwaarde, zoals veel docenten weten. Behalve het kennis hebben van criteria en die toepassen in eigen onderzoek, is het nodig om terug te kijken of de criteria góéd zijn toegepast. En dat heet reflecteren.

- *Reflecteren*

'Ieder dient zich bij iedere wetenschappelijke bewering, van zichzelf of anderen, voortdurend af te vragen of die, in het licht van de beschikbare evidentie, wel correct is,' zo omschrijven Smits e.a. (2000, p. 17) de 'wetenschappelijke attitude' die zij noodzakelijk achten om onderzoek met kwaliteit te kunnen uitvoeren. Dat is de kern van het reflecteren waar het hier om gaat: beoordelen of (een stap in) het onderzoek correct is gedaan, of criteria goed zijn toegepast en of de goede criteria zijn toegepast.

Het kenmerk van reflecteren is volgens Korthagen (2001) dat het gericht is op het (her)structureren van een ervaring, een probleem of bestaande kennis of inzicht. Reflecteren (her)vormt de mentale structuren bij de lerende. Het verschil met *denken* is dat het bij denken in eerste instantie gaat om nadenken over de criteria en bij *reflecteren* om nadenken over het al dan niet juiste gebruik van de criteria. Daarin komt het cyclische karakter van het leren tot uitdrukking (figuur 4.2). Als leerlingen beginnen op het materiële niveau met het uitvoeren van een onderzoek en ze reflecteren vervolgens (op het verbale niveau) op de kwaliteit ervan, dan zouden daaruit criteria afgeleid kunnen worden die in hun kennisstructuur worden opgenomen (het mentale niveau). Als ze zich vervolgens op de volgende cyclus oriënteren, kunnen ze daarbij de criteria gebruiken en reflecteren op het gebruik ervan. Ze kunnen dan komen tot bijgestelde criteria of tot 'meta'criteria hoe en wanneer ze te gebruiken. Dit zou je een inductieve weg kunnen noemen. De deductieve weg start een stap eerder (of twee stappen later): eerst kennis nemen van de criteria en ze vervolgens toepassen; reflectie leidt dan weer tot bijstelling of kennisontwikkeling over het gebruik ervan. Deze 'criteria voor criteria' en dit 'nadenken over nadenken' weerspiegelen het metacognitieve karakter van het reflecteren.

Het 'verankeren' op het mentale niveau onderstreept ook het belang van de voorkennis, want dat is de kennisstructuur waar de nieuwe kennis in verankerd moet worden. De stap die in één cyclus van materieel, verbaal en mentaal niveau gemaakt wordt, mag daarom niet te groot zijn. Er kan anders geen verankering plaatsvinden. Reflectie blijkt ook een geschikte manier te zijn om de relatie met voorkennis te leggen (De Vries, 2004).

Reflecteren heeft ook te maken met transfer, want reflecteren op het juist gebruik van de criteria kan ook nadenken inhouden over in welke gevallen die criteria wel en niet moeten en kunnen worden toegepast. Simons (1990) beschrijft drie dilemma's als het gaat om het ontwerpen van instructie gericht op transfer.

Het eerste dilemma betreft de vraag of er gestreefd wordt naar nabije of verre transfer. Bij nabije transfer is er weinig (subjectieve) afstand tussen leersituatie en toepassing, bij verre transfer veel. Voor de inrichting van het onderwijs is dat een dilemma, want er wordt betwijfeld of voor deze twee manieren van transfer hetzelfde moet gebeuren: voor nabije transfer

zou je vooral moeten automatiseren en oefenen in een kleine *range* situaties, 'contextualiseren', voor verre transfer zou je moeten 'decontextualiseren'[12] in een veel grotere variatie aan situaties. De doelstelling is dat leerlingen de kwaliteit van onderzoek leren beoordelen en bewaken. En niet alleen moeten ze de kwaliteit van hun eigen onderzoek leren bewaken, het is ook de bedoeling dat ze de opgedane kennis kunnen veralgemeniseren en kunnen toepassen in nieuwe situaties. Met andere woorden: er wordt gestreefd naar verre transfer. Dat zou ervoor pleiten dat de criteria in een ruimer scala aan contexten aan de orde komen dan alleen die van het eigen onderzoek van de leerlingen.

Het tweede dilemma van Simons (1990) betreft de vraag of het beter is de aan te leren kennis en vaardigheden in een context aan te bieden of ze eerst contextvrij[13] aan te bieden en ze vervolgens in een context te laten toepassen. In het voorgaande is al betoogd is dat het leren denken met de criteria het best plaatsvindt in een context van betekenisvolle, domeingebonden onderwerpen en daarmee lijkt de discussie beslecht. Toch zou je dit kunnen vertalen in de vraag of er eerst in zijn algemeenheid over criteria voor de kwaliteit van onderzoek gesproken zou moeten worden, bij wijze van oriëntatie, voordat je leerlingen aan hun eigen onderzoek laat beginnen. Of zou het beter zijn om de criteria uit de eigen onderzoeken te 'destilleren'?

Het derde dilemma gaat over de taakverdeling tussen de instructeur of het materiaal en de lerende. Neemt de docent (eventueel via het materiaal) het initiatief tot reflectie of wordt dat aan de leerlingen overgelaten? Of, als je die laatste optie té open vindt, laat je de leerlingen bezig met hun eigen onderzoek en laat je het daarvan afhangen welke criteria relevant zijn om op te reflecteren of dwing je leerlingen op alle criteria te reflecteren, dus ook op criteria die wellicht (op dat moment) niet relevant zijn voor het eigen onderzoek?

Op grond van onderzoek (Hubers, 2003) is te verwachten dat de leerlingen vooral gefocust zullen zijn op het onderzoek zelf – wat gaan we onderzoeken, waar halen we de spullen vandaan, komt er wat uit? – of op een voldoende prestatie – vindt de docent het goed, voldoen we aan de eisen, lukt het om een mooi product (verslag, presentatie) te maken? En minder gefocust op de vraag of ze nu wel een valide en betrouwbaar onderzoek aan het uitvoeren zijn en of ze zeker kunnen zijn van hun conclusies, laat staan op de criteria die je daarbij zou kunnen hanteren. Met andere woorden: reflectie zal niet (altijd) spontaan plaatsvinden. Die aanzet tot reflectie zal dus van de docent of het materiaal moeten komen.

Een dilemma dat hierop aansluit, is of je aanstuurt op *reflection-on-action* of op *reflection-in-action* (Schön, 1983). In het eerste geval wordt na afloop van een activiteit erop teruggekeken en er zoveel mogelijk lering uit getrokken. Dus bijvoorbeeld aan het eind van een onderzoek bekijken of je de criteria correct hebt toegepast. *Reflection-in-action* is een proces dat tijdens een activiteit plaatsvindt en die activiteit mede stuurt. Schön (1983, p. 50) omschrijft het als volgt:

> There is some puzzling, or troubling, or interesting phenomenon with which the individual is trying to deal. As he tries to make sense of it, he also reflects on the understandings which have been implicit in his action, understandings which he surfaces, criticizes, restructures, and embodies in further action.

[12] Vanuit het idee dat begrippen hun betekenis ontlenen aan de context waarin ze gebruikt worden, is de term 'decontextualiseren' eigenlijk verkeerd (Van Oers, 1998b). Het gaat er in die visie om om de begrippen beschikbaar te maken voor gebruik in een groter aantal contexten. 'Recontextualiseren' is een betere term. Hier wordt echter omwille van de referentie aan Simons de term 'decontextualiseren' gebruikt.

[13] Vanuit het idee dat begrippen hun betekenis ontlenen aan de context waarin ze gebruikt worden, is de term 'contextvrij' een onmogelijkheid (Van Oers, 1998a). We vatten het hier op als 'gegeneraliseerd'.

Reflection-in-action wordt dus geïnitieerd en gestuurd door het onderzoek van de leerlingen zelf, het initiatief voor *reflection-on-action* kan ook van buiten het onderzoek komen.

De wens om de leerlingen door reflectie zich de criteria voor goed onderzoek eigen te laten maken (in hun mentale structuur te verankeren) door ze zelf onderzoek te laten doen, leidt dus tot dilemma's met betrekking tot de mate van koppeling van het aan de orde stellen van de criteria en het eigen onderzoek; koppeling qua inhoud, initiatief en tijdstip. In paragraaf 4.2 maken we een beargumenteerde keus in die dilemma's.

- *Interactie*

De interactie die tussen leerlingen onderling en tussen leerlingen en docent plaatsvindt is grotendeels verbaal, al komt er steeds meer aandacht voor de non-verbale elementen van communicatie in het natuurwetenschappelijke onderwijs (o.a. Kress, Jewitt, Ogborn & Tsatsarelis, 2001; Piccinini & Martins, 2003). Het ontwikkelen van en reflecteren op criteria zal in ieder geval een verbale component hebben.

Verschillende auteurs wijzen op het belang van de taal in het science-onderwijs. Sommigen zien het als deel van de *enculturation*, het lid worden van de wetenschappelijke gemeenschap, (o.a. Hodson & Hodson, 1998a). Anderen wijzen vooral op het belang van de taal zoals die in de klas wordt gesproken, als voertuig voor het construeren van kennis en betekenis (o.a. Mortimer & Scott, 2003; Sutton, 1996).

Wat het inwijden in de specifieke taal van de natuurwetenschappen betreft, gaat het niet alleen om het juist kunnen gebruiken van termen en begrippen (jargon, zo u wilt), maar ook om de manier waarop je waarnemingen onder woorden brengt of conclusies formuleert. Sommige elementen van begrip van bewijs zijn daar een goede illustratie van. Ze bevatten zowel terminologie ('eenduidig', 'significant', 'betrouwbaarheid' en dergelijke) als aanwijzingen voor de communicatie ('een correlatie is nog geen oorzakelijk verband en mag dus niet zonder meer als zodanig geïnterpreteerd worden'). Zohar (2004) merkt op dat complexe, technische taal niet altijd nodig is, omdat veel technische termen een eenvoudig alledaags equivalent hebben. Als voorbeeld noemt zij het aanduiden van 'onder controle houden van variabelen' met 'eerlijk meten'. De vraag is echter of het begrip 'eerlijk meten', ook in de context van onderzoek doen, wel voor iedereen die betekenis heeft. Het zou bijvoorbeeld ook 'nauwkeurig aflezen' kunnen betekenen.

Bij de rol van taal bij kennisontwikkeling wordt expliciet verwezen en voortgebouwd op de theorieën van Russische leerpsychologen als Vygotski en Bakhtin. Het opbouwen (construeren) van kennis is geen individuele zaak maar een interpersoonlijke, sociale gebeurtenis. Betekenisverlening vindt plaats in interactie en wordt daarna geïnternaliseerd. Mortimer en Scott (2003, hoofdstuk 2) geven een toegankelijke en heldere uiteenzetting hiervan. Kernpunt is de constatering dat kennisontwikkeling en betekenisverlening een *dialogisch* proces is, dat wil zeggen dat er uitwisseling, confrontatie en afstemming van betekenissen plaatsvindt tussen personen of tussen oude en nieuwe betekenissen, voordat een nieuwe (bijgestelde) betekenis in de kennisstructuur van de lerende wordt verankerd.

Over de rol van de docent hierin concludeert Scott (1998, p. 72) dat '… the analyses presented suggest that the way in which the teacher 'talks around the activity' is at least as important as the activity itself' (bijvoorbeeld over de betekenis van 'eerlijk meten'). De docent kan wetenschappelijk correcte kennis als zodanig – op een 'authoritative' manier – presenteren, maar dat is niet genoeg om het in de hoofden van de leerlingen te laten belanden. 'Authoritative discourse' is gericht op het verwoorden van de 'juiste' inhoud en daarmee van een gesloten karakter, terwijl 'dialogic discourse' open staat voor alternatieve visies. Uitein-

delijk zijn beide vormen nodig om tot het beoogde resultaat te komen: het verlenen van betekenis aan de inhoud en het verankeren daarvan in de eigen kennisstructuur.

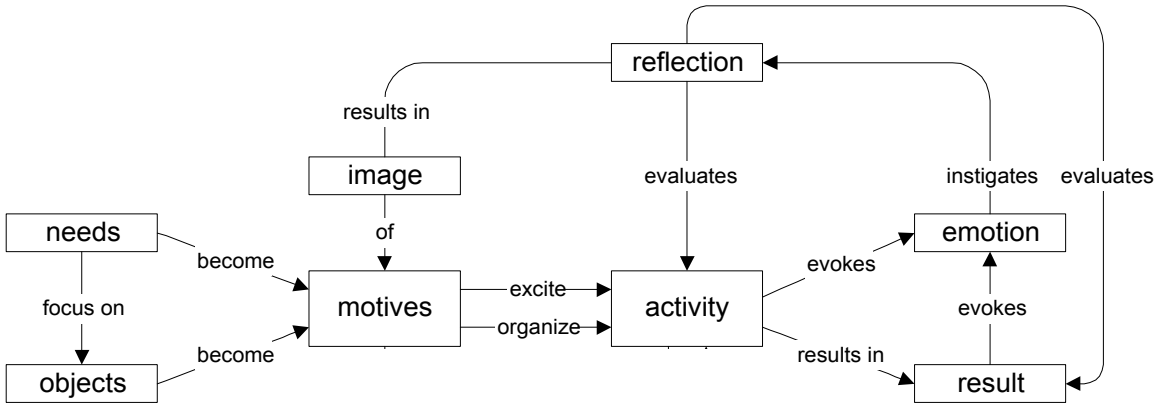
Een aanpak voor het leren bewaken van de kwaliteit van onderzoek moet dus voldoende ruimte bieden voor praten, tussen docent en leerling maar vooral tussen de leerlingen onderling. Het adagium zou dus kunnen luiden: 'Praat aleer gij doende zijt en doende praat dan nog', al zullen niet alle docenten dat graag in hun lokaal ophangen.

- *Denken, doen én reflecteren in interactie*

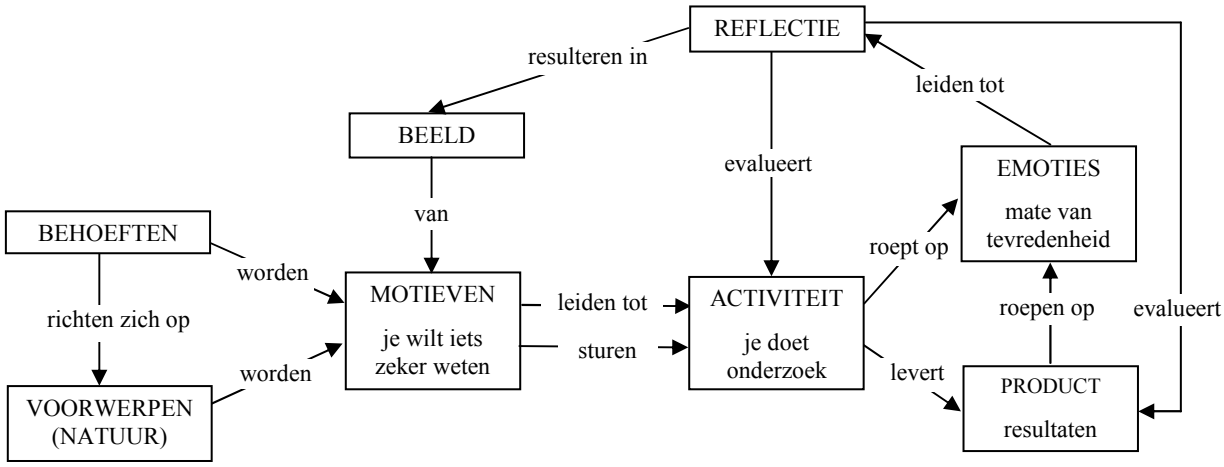
De relatie tussen de vier componenten is in figuur 4.1 schetsmatig aangegeven, maar het blijft natuurlijk van belang om deze relaties verder uit te werken en theoretisch in te bedden. Ook daarvoor kunnen we aanwijzingen ontlenen aan de cultuurhistorische theorie. Daarin staan inhoudelijk gedefinieerde en cultuurhistorisch bepaalde *activiteiten* centraal. Leren hoe activiteiten in een handelingspraktijk worden uitgevoerd gebeurt in interactie met een leraar of coach (meester-gezel relatie). Taal speelt daarbij een cruciale rol, doordat wat in interactie omgaat, wordt omgezet in 'internal speach' en denken (Arievitch & Haenen, 2005). Ook in de cultuurhistorische theorie speelt reflectie een belangrijke rol. Reflectie is echter niet alleen een individuele aangelegenheid, maar komt met name in interactie met anderen tot zijn recht.

Hoe deze elementen zich verhouden in het kader van activiteiten is weergegeven in figuur 4.3, ontleend aan het proefschrift van Van Aalsvoort (2000). Behoeftes worden motieven om activiteiten te ondernemen die tot resultaten leiden. De resultaten én de activiteit worden door middel van reflectie geëvalueerd.

In figuur 4.4 is dit schema geconcretiseerd voor het doen van onderzoek (als cultuurhistorisch bepaalde activiteit in de wetenschappelijke praktijk én als afspiegeling daarvan in het onderwijs). De behoefte om iets (zeker) te willen weten leidt tot onderzoek dat resultaten oplevert. De mate van tevredenheid leidt in meer of mindere mate tot reflectie op (de kwaliteit van) proces en product. Ook leidt reflectie tot een beeld van het motief: wat betekent dat, willen weten en wanneer weet je het goed genoeg, wanneer is de behoefte vervuld? Dan moet je een idee hebben van de mate van zekerheid waarmee je het weet.



Figuur 4.3 De ontwikkeling van een activiteit volgens de cultuurhistorische theorie (Van Aalsvoort, 2000, p. 189).



Figuur 4.4 De 'ontwikkeling van een activiteit' uitgewerkt voor de activiteit 'onderzoek doen'.

Figuur 4.4. voegt aan 'onderzoek doen' expliciet motieven toe. Motieven komen voort uit behoeften, de behoefte om iets te willen weten in dit geval. Evaluatie van die behoefte, met andere woorden reflecteren en beslissen of de behoefte in voldoende mate is vervuld, speelt zich op twee niveaus af: 'Heb ik een antwoord?' en 'Ben ik zeker (genoeg) van mijn antwoord?' Het blijkt leerlingen extra te motiveren als deze vragen ook niet op voorhand door de docent beantwoord kunnen worden, met andere woorden als de leerlingen echt 'nieuwe' kennis produceren (Hodson, 1996; Van Rens, 2005).

In het voorgaande is de nadruk gelegd op het leerproces zoals dat bij de leerlingen zou kunnen plaatsvinden en dat spoort met de trend in het hedendaagse onderwijs om het leren vóór het onderwijzen te plaatsen. De docent moet het leren echter wel mogelijk maken en ondersteunen. Mogelijk maken door de omgeving (onder andere qua ruimte, orde, werkzaamheden en materialen) te scheppen waarin geleerd kan worden. En ondersteunen – waarvoor in de Engelstalige literatuur vaak de term *scaffolding* gebruikt wordt, afgeleid van *scaffold*, steiger – door de lerenden in de gelegenheid te stellen een taak te volvoeren die zonder steun niet zou lukken. Zoals Wood, Bruner en Ross (1976, p. 90) het verwoorden:

> This scaffolding consists essentially of the adult 'controlling' those elements of the task that are initially beyond the learner's capacity, thus permitting him to concentrate upon and complete only those elements that are within his range of competence. The task thus proceeds to a successful conclusion. We assume, however, that the process can potentially achieve much more for the learner than an assisted completion of the task. It may result, eventually, in development of task competence by the learner at a pace that would far outstrip his unassisted efforts.

En daar gaat het uiteindelijk om: om het vermogen zo'n taak zelfstandig uit te voeren en er kritisch naar te kijken:

> The ultimate aim is to give learners control over their own learning processes and the confidence to engage in critical analysis. The apprenticeship process usually begins with a competent other person – the tutor – making explicit their tacit knowledge of modelling effective strategies through demonstrating desirable ways of problem solving in authentic activity. (Hennessy, 1993, p. 12)

Met betrekking tot de mate van sturing door de docent, valt een grote open onderzoeksopdracht in de indeling van Pieters en Verschaffel (2003) onder 'losse externe sturing',

maar *cognitive apprenticeship* onder 'gedeelde sturing'. Dat wil zeggen dat er op het terrein van de regulatieactiviteiten veel aan de leerlingen zelf overgelaten wordt, maar dat op het terrein van de cognitieve activiteiten de docent meer sturing geeft:

> Door de vaak verborgen denk- en leeractiviteiten te laten expliciteren, bijvoorbeeld door directe vragen tijdens de uitvoering van de activiteiten, en te reflecteren op het eigen resultaat van het uitvoeren van een handeling, krijgt de leerling inzicht in de eigen (denk)activiteiten. (Pieters & Verschaffel, 2003, p. 271)

Dit is de rol die de docent ook in de onderwijsleerstrategie is toebedeeld: bevragend en reflectie uitlokkend. Hoe expliciet dat uitlokken van reflectie moet zijn, is dan nog wel een open vraag. Die komt in de volgende paragraaf uitgebreider aan de orde.

4.2 Uitgangspunten voor twee onderwijsleerstrategieën: expliciet en impliciet

De inzichten die in het voorgaande verwoord zijn, leiden tot de volgende uitgangspunten voor het ontwerp van een onderwijsleerstrategie:

- het gaat om het opdoen van kennis over kwaliteit van onderzoek en niet om het opdoen van biologisch inhoudelijke kennis; de criteria voor goed onderzoek vormen de kennisinhoud;
- de motivatie om te leren over de kwaliteit van onderzoek wordt gezocht in de vragen van de leerlingen; ze mogen hun eigen onderwerp en onderzoeksvraag kiezen, wat willen ze echt (en zeker) weten?;
- de leerlingen doen een eigen onderzoek, waarin ze veel eigen keuzes mogen maken;
- er is veel ruimte voor interactie, zowel tussen leerlingen onderling als tussen leerlingen en docent; de leerlingen werken daarom in kleine onderzoeksteams (twee of drie personen) en de docent is beschikbaar voor vragen en feedback;
- het komen tot kennis van kwaliteit van onderzoek wordt gestimuleerd door reflectie op de kwaliteit van het eigen onderzoek; hierbij wordt een beroep gedaan op hetgeen de leerlingen al weten en kunnen.

Maar wat is de beste manier om leerlingen tot reflectie aan te zetten? De in 4.1 al genoemde dilemma's vragen om keuzes met betrekking tot inhoud van, initiatief tot en tijdstip van reflectie.

- *Inhoud van de reflectie*

Bouw je alle criteria structureel in en maak je ze zo voor alle leerlingen onontkoombaar? Veel auteurs bevelen een dergelijke expliciete aanpak van begrip van bewijs of reflectie aan, zonder dit echter empirisch te onderbouwen. Millar e.a. (1994) stellen op grond van de constatering dat het doen van onderzoek een specifieke kennisbasis heeft, dat het dus expliciet onderwezen moet worden. Dit uitgangspunt is omgezet in een aantal boeken en onderwijsmaterialen voor docenten voortgezet onderwijs, met name in Engeland (o.a. Goldsworthy, Watson & Wood Robinson, 2000; Gott, Foulds & Johnson, 1997; Gott, Foulds & Jones, 1998, Gott, Foulds & Roberts, 1999; Sang & Wood-Robinson, 2002).

Maar heeft het aan de orde stellen van bijvoorbeeld de begrippen afhankelijke en onafhankelijke variabele wel het gewenste leereffect als leerlingen niet zelf met die vraag in hun onderzoek bezig zijn? In een pleidooi voor onderwijs geschoeid op sociaal constructivistische leest, weliswaar voor heel jonge kinderen, pleiten Mallory & New (1994, p. 333) voor feedback op de eigen doelen van de leerlingen:

Children have a right to genuine and constructive feedback related to *their* goals of competence and community participation rather than adult goals associated with fragmented skills or classroom management. (*cursivering van hen*)

- *Initiatief tot reflectie*

Wacht je met het reflecteren op de kwaliteit van het onderzoek tot de leerlingen dat zelf problematiseren, of val je ze tussendoor lastig met vragen daarover? Het eerste heeft als risico in zich dat er dan misschien wel helemaal niet gereflecteerd wordt, zoals eerder al betoogd, maar het tweede lijkt te botsen met de beoogde zelfsturing van leerlingen, zoals in de betekenisgerichte en toepassingsgerichte leerstijlen van Vermunt (1992; zie ook Verloop & Lowyck, 2003). Omdat echter niet alle leerlingen die leerstijlen (al) hebben, wordt ervoor gekozen het initiatief ook van de docent laten komen.

- *Tijdstip van reflectie*

Leg je het eigen onderzoek stil om te reflecteren of doe je het tussen de bedrijven door? Met andere woorden: richt je je op *reflection-on-action* of op *reflection-in-action*? Ook hier is voor beide opties iets te zeggen. *Reflection-in-action* kan heel direct betrokken worden op het probleem onder handen, maar heeft weinig distantie en daardoor misschien ook weinig transfer. *Reflection-on-action* vereist wel een stap opzij en kan frustrerend en daardoor wellicht contraproductief werken. De vraag lijkt of je alleen op het eigen onderzoek reflecteert of de reflectie veralgemeniseert. Verschaffel (1995) betoogt dat expliciet op transfer gerichte instructie gegeven dient te worden om 'decontextualisatie'[14] te laten optreden.

- *Twee onderwijsleerstrategieën*

Binnen de klankbordgroep die dit onderzoek heeft begeleid, was alleen overeenstemming met betrekking tot het tweede dilemma: het initiatief tot reflectie moet niet alleen aan de leerling overgelaten worden, de docent moet het stimuleren. Met betrekking tot de andere twee dilemma's was er verschil van mening. Iedereen vond weliswaar dat het aan de orde stellen van begrip van bewijs op het moment dat het in het leerling-onderzoek aan de orde is de grootste kans op succesvol leren geeft. Maar dat het daarom ook alléén op dat moment aan de orde gesteld kan worden, vond geen algemene instemming. Het is immers mogelijk dat dan sommige aspecten van begrip van bewijs helemaal niet aan de orde komen, terwijl het allerminst vanzelfsprekend is dat de leerlingen het dan wel beheersen. Kortom, enerzijds werd de vrees uitgesproken dat een deel van de beoogde leerinhoud niet aan de orde zou komen, anderzijds werd er gesuggereerd dat het verspilde moeite zou zijn om te reflecteren op elementen van begrip van bewijs op momenten waarop dat voor het eigen onderzoek niet relevant is.

Er tekenden zich twee varianten af:

- een variant waarbij alle door de docent relevant geachte criteria in andere, gelijksoortige contexten aan de orde komen, waarbij dat niet afhankelijk is van de vraag van de leerling en waarbij de reflectie op momenten plaatsvindt dat het eigen onderzoek even onderbroken wordt;
- een variant waarbij alleen op de criteria gereflecteerd wordt die in het onderzoek van de leerlingen problematisch zijn, waarbij dit door de leerlingen óf de docent gesignaleerd wordt en de reflectie tijdens en alleen op het eigen onderzoek plaatsvindt.

Op dat moment was er onvoldoende grond om voor één van beide opties te kiezen. Besloten werd dat te onderzoeken. De onderzoeksvraag naar kenmerken van effectief materi-

[14] Zie voor een kanttekening bij dit begrip noot 12 op pagina 60.

aal en docentgedrag voor begrip van bewijs (par. 1.2) kan dus wat betreft de begeleiding geoperationaliseerd worden als:

Is er verschil in effectiviteit tussen een onderwijsleerstrategie waarin alle elementen van begrip van bewijs expliciet aan de orde worden gesteld en een onderwijsleerstrategie waarin elementen alleen aan de orde komen als het onderzoek van de leerlingen daar aanleiding toe geeft?

Dat is uitgewerkt in twee onderwijsleerstrategieën: een expliciete en een impliciete.

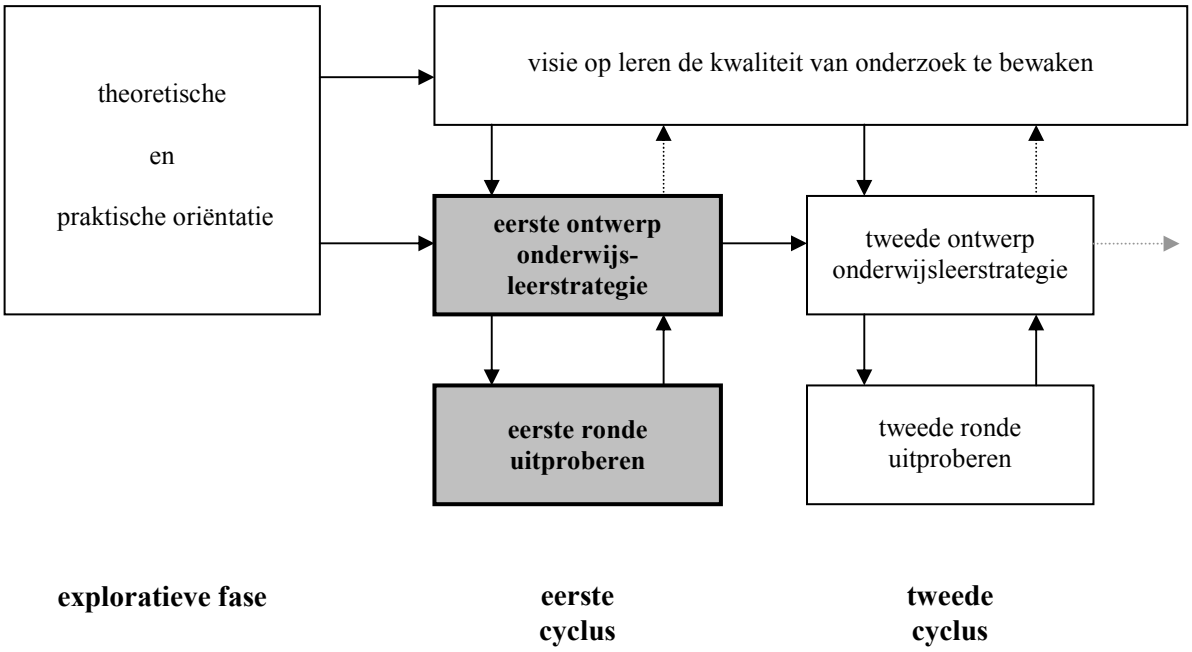
De expliciete onderwijsleerstrategie

De expliciete onderwijsleerstrategie kent een aantal 'momenten van expliciete reflectie'. Dat zijn momenten waarop de docent een of enkele elementen van begrip van bewijs aan de orde stelt die van belang zijn voor de fase waarin het onderzoek van de leerlingen geacht wordt zich volgens de planning te bevinden. De opdracht wordt centraal door de docent ingeleid, de leerlingen maken vervolgens met hun partners een opdracht die moet uitmonden in het onderkennen en formuleren van een of meer criteria voor goed onderzoek. In een onderwijsleergesprek worden de criteria geëxpliciteerd. Aan de hand van die criteria wordt op het eigen onderzoek gereflecteerd ('zijn ze goed toegepast?'). In paragraaf 5.2 worden deze momenten uitgebreider omschreven.

De impliciete onderwijsleerstrategie

Deze strategie kent geen centrale momenten van reflectie. In het begeleiden van de leerlingen is de docent alert op het signaleren van aanknopingspunten om een of meer criteria aan de orde te stellen. Dat zal vooral vragenderwijs gebeuren om ook in deze strategie de leerlingen tot eigen kennisconstructie te laten komen, bijvoorbeeld: 'Is jullie onderzoeksvraag duidelijk genoeg?'of 'Hebben jullie nagedacht over een blanco?' De term 'impliciet' betekent dus niet dat er vooral impliciet c.q. intuïtief geleerd wordt (Swaak, 1998) of dat de criteria impliciet blijven, want dat is juist niet de bedoeling. 'Impliciet' slaat op de aanleiding om te reflecteren (en daarbij criteria te expliciteren). Die aanleiding komt van het onderzoek van de leerlingen en is dus niet van te voren te plannen.

Beide strategieën kennen echter ook veel gemeenschappelijks, zoals in het begin van deze paragraaf in de uitgangspunten is verwoord. Het volgende hoofdstuk begint dan ook met het beschrijven van die strategieën, hun gezamenlijke kenmerken en hun verschillen.



5 De eerste cyclus

5.1 Het eerste ontwerp
5.2 Momenten van expliciete reflectie
5.3 Onderzoekspopulatie en instrumenten
5.4 Bevindingen
5.4.1 Uitvoerbaarheid
5.4.2 Effectiviteit
5.5 Verbeterpunten voor de tweede cyclus

'A very real challenge for researchers and practitioners to tackle together, then, is how to make it feasible to simultaneously scaffold students' conceptual understanding of complex topics and their metacognitive knowledge and regulation of how they are building that understanding together.'
Kathleen Hogan, (1999, p. 1105)

Hoofdstuk 5 De eerste cyclus

De uitgangspunten waar het vorige hoofdstuk mee geëindigd is, behoeven uitwerking om in praktijk omgezet te worden. Samen met de randvoorwaarden van de scholen waar de lessen uitgeprobeerd zijn, leidt dit tot een eerste opzet (par. 5.1). Speciale aandacht behoeven de momenten van expliciete reflectie. In paragraaf 5.2 wordt het ontwerp daarvan beschreven.

De zo ontwikkelde praktische opdracht is beproefd in de praktijk om in eerste instantie antwoord te krijgen op twee vragen:

- *in welke mate is de praktische opdracht uitvoerbaar voor docent en leerlingen, met speciale aandacht voor de momenten van expliciete reflectie?*
- *in welke mate is er leereffect te constateren met betrekking tot begrip van bewijs?*

De scholen, docenten en leerlingen die hebben meegewerkt aan de eerste cyclus alsmede de wijze waarop beoordeeld is of het eerste ontwerp aan de verwachtingen voldeed, vormen het onderwerp van paragraaf 5.3.

Hoe de eerste ronde uitproberen verliep, is beschreven in paragraaf 5.4. om uiteindelijk de vraag te kunnen beantwoorden in hoeverre aan de verwachtingen werd voldaan en welke aanpassingen nodig waren om de uitvoerbaarheid en de effectiviteit te vergroten (par. 5.5).

5.1 Het eerste ontwerp

Om tot een eerste ontwerp te komen van een lessenserie – op basis van de uitgangspunten uit paragraaf 4.2 – die in de praktijk beproefd kan worden, moest uiteraard ook rekening gehouden worden met de randvoorwaarden die vanuit de scholen en de docenten kwamen die hebben meegewerkt aan de eerste ontwerp- en onderzoekscyclus (zie voor omschrijving van de proefscholen en docenten paragraaf 5.3). Deze praktische randvoorwaarden waren de volgende:

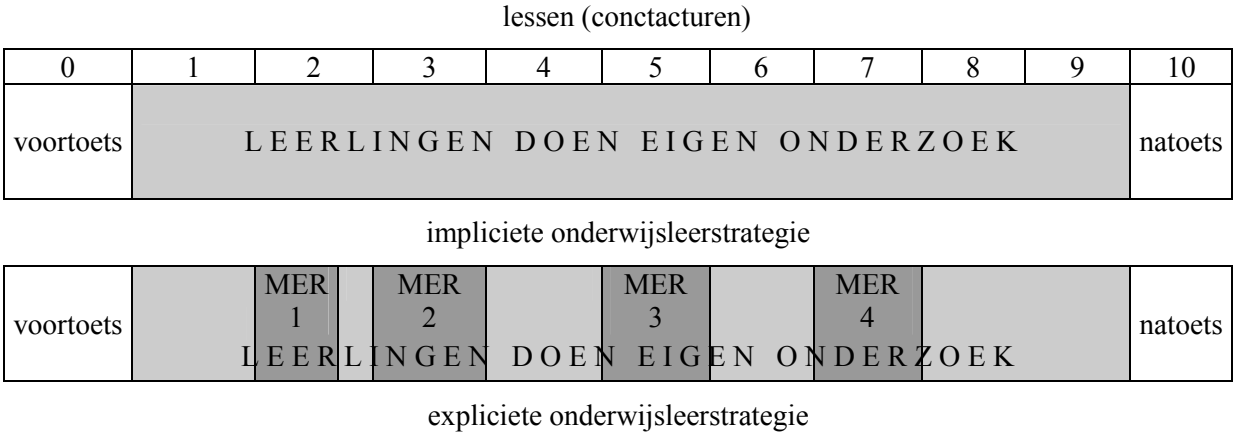
- de lessenserie beslaat maximaal elf lessen, inclusief voor- en natoets;
- de lessenserie moet kunnen meetellen als praktische opdracht;
- de technisch onderwijsassistent is tijdens de contacturen door de leerlingen te raadplegen voor advies en materiaal.

Het is moeilijk om op deze plaats recht te doen aan het creatieve ontwerpproces zoals dat heeft plaatsgevonden in wisselwerking tussen onderzoeker, begeleiders (promotor, copromotor, klankbordgroep) en docenten. We volstaan hier met de constatering dat het eerste ontwerp inhoud en vorm heeft gekregen in die wisselwerking, die ook nog tijdens de uitvoering doorging.

De uitgangspunten en randvoorwaarden leidden uiteindelijk tot de volgende kenmerken van het eerste ontwerp:

- de lessenserie omvat negen contacturen plus twee uren voor de voor- en natoets; tussen de contacturen zit een week; de leerlingen hebben die lesuren en de tijd ertussen om aan hun onderzoek te werken, maar er zijn ook lesuren biologie die andere onderwerpen behandelen;
- de leerlingen werken samen in groepjes van twee of drie[15] aan een eigen hypothesetoetsend onderzoek binnen het vak biologie; ze bedenken een eigen onderwerp, formuleren hun eigen onderzoeksvraag, maken een onderzoeksopzet, verzamelen en verwerken gegevens, beantwoorden de onderzoeksvraag en evalueren hun eigen onderzoek; dit is geen lineair proces, maar er wordt heen en weer gegaan tussen de verschillende 'stappen' van het onderzoek;
- het product telt mee als praktische opdracht; leerlingen maken een verslag als product;
- de docent heeft een ondersteunende rol, hij/zij[16] is vraagbaak voor de leerlingen wat betreft de opzet en uitvoering van het onderzoek; hij heeft echter ook een initiërende rol waar het gaat om het stimuleren van de reflectie op de kwaliteit van het onderzoek;
- in de wisselwerking tussen onderzoek doen en reflecteren op onderzoek komen de kwaliteitscriteria voor onderzoek aan de orde en hebben de leerlingen gelegenheid ze toe te passen.

Er zijn twee varianten van deze opzet ontwikkeld, een expliciete en een impliciete; in de expliciete variant werd de reflectie op de kwaliteit van onderzoek op vastgestelde tijdstippen expliciet aan de orde gesteld (de zogenaamde 'momenten van expliciete reflectie'); het eigen onderzoek werd daar dan even voor stilgelegd; in de impliciete variant vond die reflectie uitsluitend plaats als het eigen onderzoek van de leerlingen daartoe aanleiding gaf. Op beide scholen, dus door beide docenten, werd zowel de expliciete als de impliciete variant uitgevoerd.



Figuur 5.1 Schematische weergave van het eerste ontwerp van de lessenserie. In de impliciete onderwijsleerstrategie (boven) doen de leerlingen hun eigen onderzoek en vindt reflectie plaats als dat onderzoek er aanleiding tot geeft. In de expliciete onderwijsleerstrategie (onder) wordt het eigen onderzoek van de leerlingen onderbroken door vier momenten van expliciete reflectie (MER). De totale tijd is in beide strategieën hetzelfde.

[15] Omdat de meeste leerlingen in groepjes van twee werken, worden die vaak aangeduid als duo's, maar zo'n 'duo' kan dus ook uit drie leerlingen bestaan.

[16] In het onderzoek hebben een vrouwelijke en een mannelijke docent meegewerkt (zie par. 5.3). Als in algemene termen over hen gesproken wordt, wordt de mannelijke vorm gebruikt. Hetzelfde geldt voor de leerlingen.

Figuur 5.1 geeft de opzet van beide varianten in schema weer. De totale contacttijd was in beide varianten hetzelfde. In het eerste ('nulde') en het laatste uur van de lessenserie maakten de leerlingen een toets over de kwaliteit van onderzoek. Deze wordt in paragraaf 5.3 besproken.

Aan het begin van de eerste les werd de opzet van de hele praktische opdracht door de docent uiteengezet. Dat gebeurde door het uitdelen van een schriftelijke instructie die kort mondeling door de docent toegelicht werd. In die instructie werden enkele mededelingen gedaan over de onderwerpkeuze, onderdelen c.q. stappen van een onderzoek, wijze van verslaglegging en planning. Gedurende de gehele praktische opdracht konden de leerlingen de docent om advies vragen, terwijl de docent ook op eigen initiatief de voortgang ter sprake kon brengen en de leerlingen kon bevragen.

5.2 Momenten van expliciete reflectie

Het doel van de momenten van expliciete reflectie (MERs) was om de kwaliteit van onderzoek aan de orde te stellen met behulp van verschillende elementen van begrip van bewijs en die te laten toepassen op het eigen onderzoek van de leerlingen. In paragraaf 3.5 is beargumenteerd waarom er gekozen is om de leerlingen slechts met een selectie van de 23 elementen te confronteren, te weten de elementen 1, 3, 4, 5, 6, 8, 9, 10, 14, 18 en 21. Deze zijn over een aantal momenten van expliciete reflectie verdeeld, zodanig dat ze aansluiten bij verschillende fasen van het onderzoek. Om enerzijds voldoende tijd te hebben deze elementen aan de orde te stellen en anderzijds ook voldoende tijd over te laten voor het eigen onderzoek, is dit aantal momenten bepaald op vier, te weten bij het formuleren van de onderzoeksvraag, bij het maken van een onderzoeksopzet, bij het verzamelen en verwerken van de resultaten en bij het evalueren van het onderzoek.

De algemene opzet van de MERs was als volgt:

- oriëntatie: waarom deze opdracht, koppeling aan het bewaken van de kwaliteit van het eigen onderzoek (de vraag 'zeker weten?);
- opdracht met betrekking tot enkele elementen van begrip van bewijs;
- bespreking van de opdracht;
- toepassing op het eigen onderzoek.

In tabel 5.1 is weergegeven welke elementen van begrip van bewijs in welke MER aan de orde kwamen en hoe de MERs zijn opgebouwd. Alle opdrachten in de MERs werden mondeling geïntroduceerd. De leerlingen kregen vervolgens werkbladen die hen door het MER leidden. Ze bevatten ruimte voor het opschrijven van antwoorden op vragen bij de opdracht, de uitkomsten van de bespreking en de terugkoppeling op hun eigen onderzoek.

Voor elke MER werd in een scenario omschreven wat de beoogde doelen zijn en welke activiteiten er van de leerlingen en de docent werden verwacht. De scenario's zijn in samenwerking met de docenten ontwikkeld. Zo'n scenario diende als een steun voor de docenten bij de uitvoering van de MERs en tegelijkertijd als een 'hypothese'[17] om de waargenomen uitvoering aan te toetsen. We beschrijven kort de vier MERs; de volledige scenario's en werkbladen voor de leerlingen zijn opgenomen in bijlage 2.

[17] De term 'hypothese' om aan te geven dat een scenario een gewenst c.q. verwacht verloop van een (serie) leeractiviteit(en) beschrijft, komt van Knippels (2002). Op de keper beschouwd is er echter geen sprake van een hypothese als mogelijke verklaring (vgl. par. 2.2), maar slechts van een verwachting die al dan niet bewaarheid wordt. Uiteraard kan dan een hypothese geformuleerd worden om te verklaren *waarom* de waargenomen uitvoering al dan niet van de verwachte afwijkt.

Tabel 5.1 Elementen van begrip van bewijs en de leeractiviteiten in de momenten van expliciete reflectie (MER).

moment van expliciete reflectie (nr)	*les en verwachte fase van leerling-onderzoek*	*element van begrip van bewijs*	*leeractiviteiten*
1 Zeker weten? Wees scherp!	2 formuleren onderzoeksvraag	1, 18	- oriëntatie - practicum 'hoeveel voedingsstoffen bevat melk?' - bespreking - toepassing op het eigen onderzoek
2 Zeker weten? Stel het op de proef!	3 onderzoeksopzet	3, 4, 5, 6	- oriëntatie - opdracht 'de functie van stotten' - bespreking - toepassing op het eigen onderzoek
3 Zeker weten? Alles onder controle!	5 verzamelen en verwerken resultaten	8, 9, 10, 14, 18	- oriëntatie - onderzoeksverslag 'het effect van cafeïne' - bespreking - toepassing op het eigen onderzoek
4 Jullie eigen onderzoek: zeker weten?	7 evalueren onderzoek	21	- oriëntatie - commentaar geven op elkaars onderzoeksverslag - toepassing op het eigen onderzoeksverslag

- *MER 1: Zeker weten? Wees scherp!*

Elke MER werd ingeleid met het aangeven van de bedoeling van de opdracht. Voor de eerste luidde die als volgt:

> Vandaag doen we een activiteit die bedoeld is om jullie eigen onderzoek te ondersteunen. Het gaat dus vooral om het proces, de manier waarop je werkt en niet zozeer om de inhoud van het onderzoekje dat je nu gaat doen. Wat je in het eerste deel van de les leert, kun je daarna hopelijk op je eigen onderzoek toepassen. Jullie krijgen per onderzoeksgroepje (hetzelfde groepje als waarin je je eigen onderzoek doet) van mij een instructie op papier, de benodigde materialen staan hier klaar. Na een kwartier breek ik het af en gaan we het bespreken. Het is niet heel erg als je dan nog niet klaar bent, maar probeer wel zo ver mogelijk te komen.

In het lokaal stonden volle melk, hafvolle melk en karnemelk en indicatoren voor zetmeel, glucose, vet en eiwit. Ook waren reageerbuizen en andere practicumbenodigdheden aanwezig. De bedoeling was dat de leerlingen – voor of na de discussie – tot het inzicht kwamen dat de onderzoeksvraag 'hoeveel voedingsstoffen bevat melk?' verschillende onduidelijkheden bevat: 'hoeveel' is niet eenduidig, het kan zowel op het aantal voedingsstoffen als op de hoeveelheden van elk slaan, 'voedingsstoffen' is niet afgeperkt en 'melk' is niet gespecificeerd. Dit inzicht moest vervolgens toegepast worden op de eigen onderzoeksvraag: is die eenduidig, afgeperkt en specifiek?

- *MER 2: Zeker weten? Stel het op de proef!*

Het tweede moment van expliciete reflectie ging over het toetsen van hypotheses. Er werd een probleem voorgelegd waarin vier hypotheses worden gepresenteerd met betrekking tot het 'stotten' (het maken van een speciaal soort sprong) van Thomson gazellen (gebaseerd

op een opgave uit het vwo-examen van 2000, tweede tijdvak, die weer gebaseerd is op Alcock, 1998). Afzetten van de vier hypotheses tegen het beschreven gedrag van de gazellen zou tot het verwerpen van drie van de vier leiden. In de discussie werd de nadruk gelegd op de redenering die tot verwerpen leidt. Toepassing hiervan op het eigen onderzoek zou moeten leiden tot reflectie op de eigen onderzoeksopzet met betrekking tot het toetsen van hypothese(s).

- *MER 3: Zeker weten? Alles onder controle!*

  Het derde MER werd besteed aan het becommentariëren van een onderzoeksverslag van leerlingen van een andere school, getiteld 'Het effect van cafeïne'. Het verslag is samengevat op één bladzij (door de onderzoeker) en de studenten werd gevraagd drie punten van kritiek te noemen. Het onderhavige leerlingonderzoek beschrijft het effect van cafeïne op het uitvoeren van eenvoudige rekentaken. Het vertoont verschillende zwakheden, zoals het gebruik van zowel cafeïnevrije koffie als 'niets te drinken' als controles. In de resultaten zit een tijdseffect verborgen (alle onderzochte leerlingen deden de rekentaken beter op de derde dag, onafhankelijk of ze koffie met of zonder cafeïne of niets gedronken hadden). De reflectieopdracht zou er toe moeten leiden dat de leerlingen hun eigen onderzoek nogmaals kritisch bekijken op het juiste gebruik van controle-experimenten en onvoorziene invloed van andere variabelen.

- *MER 4: Jullie eigen onderzoek: zeker weten?*

  In het laatste moment van expliciete reflectie gaven de leerlingen commentaar op elkaars conceptverslag. Er werd hen gevraagd twee positieve en twee negatieve kanttekeningen te maken. Tevens moesten ze de vraag beantwoorden in hoeverre hun medeleerlingen zeker konden zijn van de conclusies uit hun onderzoek. Het commentaar dat ze ontvingen konden ze gebruiken om hun eigen conceptverslag te verbeteren.

5.3 Onderzoekspopulatie en instrumenten

De keuze van proefscholen is erg bepaald door de beschikbaarheid en bereidwilligheid van docenten en klassen die aan een aantal praktische randvoorwaarden moesten voldoen. De docent moest in twee opvolgende jaren twee 5 vwo-klassen (of één grote die voor de tijd van het project gesplitst kon worden) met het vak biologie 1,2 lesgeven en bereid zijn om die klassen een grote praktische opdracht met een eigen onderzoek te laten doen. Tevens moest de docent bereid zijn een aantal voorbereidende en evaluerende bijeenkomsten bij te wonen, extra werk te doen en het eigen gedrag in de klas te bespreken met de onderzoeker. Er stond een bescheiden financiële vergoeding tegenover. Twee docenten werden bereid gevonden, een van een school in Amsterdam en een van een school in Amstelveen. Ze worden aangeduid met *Lex* die werkt op de *Amsterdamse School* (AS) en *Bea* die lesgeeft op de *School in Amstelveen* (SA). Achtereenvolgens worden de scholen, de docenten en de proefgroepen beschreven. De beschrijvingen van de scholen zijn gebaseerd op rapporten van de Inspectie voor het Onderwijs (2003, 2004).

De *Amsterdamse School* (AS) is een school voor vwo en havo met een leerlingenaantal tussen de 800 en 850, waarvan zo'n 40% in klas 3 of hoger van het vwo. Het grootste deel van de leerlingen woont in Amsterdam Zuid en Amsterdam Centrum en behoort tot de kansrijkere sociale lagen van de maatschappij. Het aantal leerlingen van allochtone afkomst is miniem. Er werken ongeveer 65 leraren op de school en 13 personeelsleden met een onder-

wijsondersteunende taak. De leeftijd van de leraren ligt dicht bij het landelijk gemiddelde. Van de examenkandidaten in 2003 had 19% het profiel Natuur & Gezondheid gekozen en 13% het profiel Natuur & Techniek. Het gemiddeld cijfer voor het centraal examen in de exacte vakken was in 2003 een 6,1, dat volgens de inspectie rekening houdend met de leerlingkenmerken onder het landelijk gemiddelde ligt.

De Inspectie van het Onderwijs (2003) typeert de school als volgt:

> Op de 'Amsterdamse School' heerst een prettige sfeer onder leerlingen en leraren. Leerlingen en ouders voelen dat zij serieus genomen worden. Het team toont veel elan om iets extra's te realiseren voor de leerlingen. (…) De leraren werken met plezier op de school en tonen zich tevreden over het taakbeleid. Zij geven op een tamelijk traditionele manier les en willen de leerlingen veel structuur bieden. (p. 5)
>
> Het team beschikt over voldoende didactische basisvaardigheden. De lessen hebben een duidelijke structuur en de leraren leggen op heldere wijze uit. Zij zijn echter nogal leerstofgericht. Zij differentiëren het leerstofaanbod nauwelijks en controleren te weinig of alle leerlingen de stof begrijpen. Bovendien worden de leerlingen zelden aangemoedigd om hardop te redeneren. Sommige docenten verdiepen zich niet genoeg in de redenering van een leerling om het leerproces effectief te kunnen bijsturen. (p. 5)
>
> Tijdens de door ons bezochte lessen zagen wij voornamelijk frontaal onderwijs en situaties waarin leerlingen individueel aan opdrachten werken. Het repertoire van de leraren op het gebied van didactische werkvormen is nog beperkt. Het merendeel van de leraren weet de leerling geen geschikte structuur te bieden voor samenwerkend of zelfstandig leren, hoewel de school dit ambieert. De leerlingen worden nog onvoldoende in staat gesteld zelf verantwoordelijkheid te nemen voor hun leerproces. (…) De school heeft een begin gemaakt met het inventariseren van de wijze waarop de verschillende vakken aandacht besteden aan algemene vaardigheden. Het overleg per sector biedt hiervoor goede mogelijkheden en bij sommige clusters vindt daadwerkelijk afstemming plaats. In het algemeen verdient afstemming van inhoud van de verschillende vakken echter nog aandacht. De vakwerkplannen wijzen uit dat er nog spraakverwarring is over het begrip algemene vaardigheden. (p. 5, 6)

De *School in Amstelveen* (SA) is een school voor vwo, havo en vmbo-t met een leerlingenaantal rond de 1500, waarvan zo'n 25% in klas 3 of hoger van het vwo. Ongeveer tweederde van de leerlingen woont in Amstelveen, de overige leerlingen komen uit de omliggende gemeentes. De leerlingen zijn afkomstig uit alle sociale milieus. Het aantal leerlingen van allochtone afkomst ligt met drie procent iets onder het landelijk gemiddelde. Door de vele internationale bedrijven in de regio trekt de school leerlingen aan van meer dan twintig nationaliteiten. Er werken ongeveer 110 leraren op de school en 26 personeelsleden met een onderwijsondersteunende taak. De leeftijd van de leraren ligt onder het landelijk gemiddelde. Van de examenkandidaten in 2003 had 17% het profiel Natuur & Gezondheid gekozen en 10% het profiel Natuur & Techniek. Het gemiddeld cijfer voor het centraal examen in de exacte vakken was in 2003 een 6,5, dat volgens de inspectie rekening houdend met de leerlingkenmerken rond het landelijk gemiddelde ligt.

De Inspectie van het Onderwijs (2004) typeert de school als volgt:

> De 'School in Amstelveen' is een veilige school waar men op een respectvolle manier met elkaar omgaat. De leerlingen zijn tevreden over de docenten. Ze vinden hen aardig en voelen zich door hen serieus genomen. De sfeer is goed. Onder personeelsleden is er een grote mate van collegialiteit. (p. 17)
>
> Op het terrein van het leerstofaanbod kent de school een lange traditie van professionele zorgvuldigheid, waarbij de waarborgen vooral door de vaksecties en de methodeboeken worden

> geleverd. De inhoud ervan is er op gericht, de kerndoelen en de stof voor de examens in voldoende mate aan te bieden. (p. 15)
> Gemiddeld zijn de lessen op de school van voldoende kwaliteit. De docenten zijn veelal vriendelijk en tonen oprechte belangstelling voor hun leerlingen. De leerlingen uiten zich direct. De meeste docenten spelen daar goed op in. Daarbij speelt het geven van feedback op het leerproces dat de leerling doorloopt nog slechts een minimale rol. De leerlingen kunnen meer gestimuleerd worden tot nadenken over hun aanpak. De didactiek is meestal ordelijk en gestructureerd en in toenemende mate activerend. (p. 16)
> De leraren sluiten in de keuze en verlevendiging van hun lesstof vaak nog weinig aan bij de leefwereld van de leerlingen. Zij voelen zich doorgaans zo verantwoordelijk voor de prestaties van hun leerlingen dat zij vaak kiezen voor sterk docent gestuurde werkvormen waarbij er weinig ruimte is voor eigen inbreng van leerlingen. Dat maakt het ook moeilijk rekening te houden met niveauverschillen tussen leerlingen. (p. 16)

Beide docenten kunnen omschreven worden als redelijk ervaren.

Lex staat vanaf 1993 voor de klas, de AS is zijn eerste school. Hij geeft les in de vakken biologie en Algemene Natuurwetenschappen. Hij vindt het belangrijk om leerlingen aan het denken te zetten over de aard en de methode van het vak. Daarbij gaat het hem om de grote lijnen, leerlingen hoeven wat hem betreft geen jonge onderzoekers te worden. Hij bevestigt het beeld dat de inspectie schetst dat de leerlingen weinig zelf verantwoordelijkheid krijgen en nemen voor hun leerproces; ze nemen in zijn ogen een nogal afwachtende houding aan en zijn vooral gericht op het behalen van hun diploma als aanzet tot een verdere loopbaan.

> Mijn tweedeklassers vinden dat ik niet genoeg uitleg, maar als ik ze vraag wat ze uitgelegd zouden willen hebben weten ze niets te bedenken. Mijn 6V-ers vinden een onderwijsleergesprek over de fotosynthese te lang duren. Ze hebben niet het idee dat ze er iets van leren. Ik constateer echter dat ze tijdens zo'n les te weinig betrokken zijn op het moment dat een ander aan het woord is. Ze hebben het gevoel dat ik ze overhoor in plaats van ze prikkel. Als ik onjuiste beweringen op het bord schrijf met de bedoeling dat daar op geschoten wordt, gaan ze het overschrijven in hun schrift. Ik weet dus zeker dat onze 5V-ers (gemiddeld) in de basis niet geïnteresseerd zijn in een hogere kwaliteit van onderzoek als dat niet onmiddellijk in hun beoordeling tot uiting komt. Prestatiegerichtheid in plaats van procesgerichtheid, en taakgericht in plaats van leerdoelgericht!
> (…)
> Inderdaad weten veel docenten de definitie van 'algemene vaardigheden' niet. Iedere poging om de leerlingen (en leraren) iets van algemene vaardigheden bij te brengen wordt bij ons gegoten in de vorm van losse projectjes of activiteitenweken, maar zelden geïntegreerd in de lessen van alledag. Leraren spreken van lesgeven als ze een theorieles bedoelen. Andere werkvormen worden niet direct als lesgeven gezien. [1][18]

Samen met zijn sectiegenoten en met de leden van de secties natuurkunde en scheikunde streeft hij naar verbetering van hun onderwijs door onder andere vaardigheden beter op elkaar af te stemmen.

Met behulp van een vragenlijst bij de ‘Roos van Leary’ (ook wel *Model for Interpersonal Teacher Behaviour* genoemd, Brekelmans, Wubbels & Créton, 1989; Wubbels, Brekelmans, Den Brok & Van Tartwijk, 2006) typeert hij zichzelf als ‘leidend’, ‘actief helpend’ en ‘inlevend’; zijn leerlingen[19] typeren hem ook zo, zij het minder uitgesproken. Op ‘aanvallend’ geven ze hem een lagere score dan hij zichzelf geeft, op ‘afhankelijk/meegaand’ een

[18] De herkomst van de citaten is te vinden in bijlage 12.

[19] De vragenlijsten zijn ingevuld door de leerlingen van de groepen die aan de tweede ronde hebben deelgenomen.

hogere. In de typologie van Brekelmans e.a. (1989) is zijn gedrag als docent te kenschetsen als 'tolerant en gezaghebbend', waarbij zijn leerlingen hem wat meegaander vinden dan hijzelf.

Bea staat vanaf januari 1998 voor de klas, vanaf augustus 1998 op de SA. Zij geeft les in biologie en is daarnaast brugklascoördinator. Zij vindt het belangrijk om in de loop van de jaren dat een leerling op school zit spiraalsgewijs aandacht aan onderzoek doen te besteden. Onderdeel daarvan zou het aanleren van criteria voor goed onderzoek moeten zijn, maar daar is – voordat ze ging meedoen aan dit promotieonderzoek – nog niet expliciet aandacht aan besteed. Ze kenschetst de leerlingen als coöperatief, hoewel ze ook leerlingen heeft die volgens haar met minimale inspanning aan de eindstreep willen komen.

Met behulp van een vragenlijst bij de 'Roos van Leary' geven haar leerlingen[19] de hoogste scores voor 'inlevend' en 'actief helpend' en in iets mindere mate voor 'leidend'. 'tolerant'. Zelf geeft ze zich hogere scores bij inlevend en actief helpend gedrag dan de leerlingen haar geven. Bea zelf ziet haar gedrag in deze klas als meer leidend dan de leerlingen beoordelen en neigt daardoor naar de typering 'tolerant en gezaghebbend', waar haar leerlingen het houden op 'tolerant' (Brekelmans e.a., 1989).

In beide scholen zijn in winter 2003 de onderwijsleerstrategieën toegepast in twee 5 vwo groepen voor het vak biologie 1,2; één groep voor de impliciete onderwijsleerstrategie, één voor de expliciete. In tabel 5.2 zijn de aantallen leerlingen van de groepen weergegeven. Op de AS betrof het één klas die voor een deel van de praktische opdracht gesplitst werd: als er een MER plaatsvond, werkte de impliciete groep in een ander lokaal. Op de SA waren het twee kleine groepen die wel op verschillende uren biologieles kregen. De groepen waren niet geheel homogeen qua samenstelling. Op de AS is het voor leerlingen met een maatschappijprofiel niet mogelijk het vak biologie 1,2 te kiezen, op de SA wel.

- *Ervaringen met het doen van onderzoek*

Op beide scholen wordt er in een aantal vakken iets aan onderzoek doen gedaan. Zowel bij godsdienst als bij de gammavakken, maar het meest bij de exacte vakken. Toch zijn de docenten natuurkunde en scheikunde van beide scholen van mening dat het om niet meer dan een aanzet gaat. Er is weliswaar aandacht voor zaken als het stellen van een goede onderzoeksvraag, het opstellen van een hypothese en het terugkoppelen van de conclusie op de resultaten, maar vaak ligt bij practica nog de nadruk op het nauwkeurig uitvoeren van procedures en metingen en op illustratie van de theorie.

Binnen het vak biologie op de AS wordt vanaf de onderbouw aandacht besteed aan leren onderzoeken. Lex ziet daar in de bovenbouw echter weinig vruchten van. In klas 4 vwo doen de leerlingen een onderzoek naar diergedrag en in de eerste maanden van het vijfde leerjaar doen ze een onderzoek naar tolerantiegrenzen van tuinkerszaden voor abiotische factoren. Daarbij is de onderzoeksvraag gegeven, de opzet en uitvoering zijn aan de leerlingen. De leer-

Tabel 5.2 Omvang en samenstelling van de leerlinggroepen in de eerste ronde. AS: Amsterdamse School; SA: School in Amstelveen.

groep	*meisjes*	*jongens*	*natuurprofiel*	*maatschappijprofiel*	*totaal*
AS expliciet	9	5	14	-	14
AS impliciet	13	3	16	-	16
SA expliciet	8	4	1	11	12
SA impliciet	11	4	6	9	15

lingen komen tot aardige resultaten, maar van kritische reflectie op het gedane onderzoek is geen sprake. In de jaren voor de eerste cyclus deden leerlingen in 5 vwo een open eigen onderzoek. Dat werd aan het begin van het schooljaar geïntroduceerd en vond grotendeels plaats buiten de contacturen. Ook hierbij was het de ervaring van Lex dat leerlingen zelden of nooit tot een conclusie kwamen in de trant van 'dat hadden we anders moeten aanpakken.'

Op de SA is de situatie vergelijkbaar. Er wordt door de jaren heen wel aandacht besteed aan onderzoek doen, maar er wordt niet heel gericht gewerkt naar criteria voor goed onderzoek. In klas 5, in het najaar, wordt een praktische opdracht 'catalase' gedaan, waarbij de leerlingen op grond van waarnemingen een hypothese moeten opstellen en toetsen. Leerlingen vinden dit moeilijk en komen zonder hints van de docent zelden tot de juiste oplossing.

- *Instrumenten*

Het verloop van de lessen werd op verschillende manieren gevolgd en geëvalueerd. Tabel 5.3 geeft een overzicht van de instrumenten en geeft weer aan de beantwoording van welke onderzoeksvraag (zie de introductie van dit hoofdstuk) ze bijdragen.

Bijna alle lessen werden geobserveerd door de onderzoeker en gesprekken tussen de docent en leerlingen werden op geluidsband opgenomen. Docent en onderzoeker evalueerden elk 'moment van expliciete reflectie' direct na afloop. Ook werden uit elke groep twee leerlingen op vier momenten geïnterviewd, samenvallend met de MERs (dus ook uit de groep die geen MERs kreeg, op ongeveer hetzelfde moment in de lessenserie). Aan de leerlingen werd een opgave voorgelegd waarin dezelfde elementen van begrip van bewijs aan de orde kwamen als in de MER (zie bijlage 3). Deze opgave werd eerst gemaakt door de leerlingen (individueel) en daarna gezamenlijk besproken om de argumenten die de leerlingen erbij hadden gebruikt te vernemen. De argumentaties zijn geanalyseerd met behulp van het argumentatiemodel van Toulmin (1958). Tevens werd aan de leerlingen uit de expliciete groep gevraagd wat ze van de MER vonden.

De MER-werkbladen van de leerlingen en de uiteindelijke verslagen van de onderzoeken werden verzameld en geanalyseerd. Hierbij werd met name gekeken naar het gebruik van de verschillende elementen van begrip van bewijs.

De toets die voor en na de lessenserie werd afgenomen bestond uit dertien vragen met in totaal 27 items. Een aantal vragen in de toets zijn bewerkingen van vragen uit havo-examens biologie. Alle elementen van begrip van bewijs die in de MERs aan de orde kwa-

Tabel 5.3 Overzicht welke onderzoeksinstrumenten in de eerste cyclus bijdragen aan de beantwoording van de vragen naar uitvoerbaarheid en effectiviteit.

instrument	*vraag naar uitvoerbaarheid*	*vraag naar effectiviteit*
observatie door onderzoeker	*	
opname van gesprekken tussen leerling en docent	*	
evaluatie MERs met docent	*	
werkbladen van leerlingen bij de MERs	*	*
interviews met enkele leerlingen	*	*
onderzoeksverslagen van leerlingen		*
voor- en natoets		*

men, kwamen ook in de toets aan de orde. De toets, het antwoordblad voor de leerlingen en het antwoordmodel zijn opgenomen in bijlage 4.

Waarschijnlijk vanwege de grote diversiteit aan begrippen in de toets en vanwege het feit dat veel leerlingen niet alle opgaven hebben kunnen maken, is de betrouwbaarheid laag (Cronbach's alfa = 0,38). Om die reden is afgezien van een kwantitatieve verwerking van de resultaten. Wel is er op een kwalitatieve manier naar de antwoorden van de leerlingen gekeken, bijvoorbeeld naar welke argumenten ze gebruikten om een onderzoeksvraag of de opzet van een onderzoek te beoordelen.

5.4 Bevindingen

Zoals in de intro van dit hoofdstuk vermeld, gaat het in de eerste cyclus om het verkrijgen van een antwoord op de volgende vragen:

- *in welke mate is de praktische opdracht uitvoerbaar voor docent en leerlingen, met speciale aandacht voor de momenten van expliciete reflectie?* (par. 5.4.1)
- *in welke mate is er leereffect te constateren met betrekking tot begrip van bewijs?* (par. 5.4.2)

De antwoorden op deze vragen leiden in paragraaf 5.5 tot verbeterpunten voor het tweede ontwerp om de uitvoerbaarheid en de effectiviteit te vergroten.

5.4.1 Uitvoerbaarheid

De observaties, werkbladen, opnamen en evaluaties met docent en leerlingen leveren gezamenlijk een beeld op van de uitvoerbaarheid van de strategieën in de verschillende groepen. Eerst beschrijven we enkele algemene aspecten en daarna gaan we specifiek in op de momenten van expliciete reflectie.

- *Instructie en begeleiding*

Na de korte introductie van de opdracht hadden leerlingen vooral vragen met betrekking tot de planning (wanneer moet het af zijn?) en de eisen die aan het product gesteld worden (moet er ook literatuur bij?). Daarna gingen ze in hun groepjes overleggen over mogelijke onderwerpen en vragen. Sommigen raadpleegden daarbij hun biologieboek.

In alle groepen waren de docenten actief betrokken bij de onderzoeken van de leerlingen. Ze stimuleerden de leerlingen om hun kennis en creativiteit in te zetten door ze vragen te stellen die hen aan het denken zetten. Ze waren anderzijds soms zo betrokken dat ze bijna de onderzoeksvraag of onderzoeksopzet zelf formuleerden. Gevraagd naar de reden voor dit gedrag, zeiden de docenten dat ze wilden voorkomen dat de onderzoeken mislukten, want dat zou zo teleurstellend geweest zijn voor de leerlingen. Een tweede reden was dat sommige onderzoeken specifieke vakkennis vereisten, kennis waarover de leerlingen niet beschikten.

In de begeleiding in zowel de expliciete als de impliciete onderwijsleerstrategie werd veel tijd besteed aan het bespreken van de uitvoerbaarheid van de onderzoeken van de leerlingen, bijvoorbeeld hoe je aan voldoende proefdieren of proefpersonen kwam. Elementen van begrip van bewijs kwamen daarbij soms ter sprake, maar er werden ook kansen gemist. Uit observatie bleek dat meer elementen aan de orde gesteld hadden kunnen worden of explicieter gemaakt hadden kunnen worden. Bijvoorbeeld: de conceptonderzoeksvragen worden wel bekeken op eenduidigheid, specificiteit en afperking, maar die termen zelf werden niet gebruikt. Er werd dan wel gevraagd: 'Wat bedoel je met reactievermogen?' en, naar aanleiding van de term *hoeveelheid licht*: 'Bedoel je lichtsterkte of afwisseling van licht en donker?', maar er werd bijvoorbeeld niet gezegd: 'Hoeveelheid licht is voor tweeërlei uitleg vatbaar; dat is niet

goed, want de termen die je in je onderzoeksvraag gebruikt moeten eenduidig zijn.' Een ander voorbeeld is dat toen er werd gesignaleerd dat er teveel variabelen in een onderzoeksopzet betrokken waren, er werd gezegd: 'Dan wordt het een onderzoek in een onderzoek', maar er werd niet gezegd: 'Je moet maar één ding variëren en één ding meten,' of: 'Als je de invloed van één factor wil weten moet je die variëren en de rest constant houden' en al helemaal niet: 'Wat is je onafhankelijke variabele en wat is je afhankelijke variabele?' Relatief de meeste aandacht werd besteed aan de kwaliteit van de onderzoeksvraag (element 1) en de onderzoeksopzet (elementen 7 tot en met 10), maar de criteria bleven meestal onbenoemd.

Achteraf merkten de docenten op dat ze onzeker waren over het begeleidingsgedrag dat van hen verwacht werd. Dat betrof vooral hun gedrag in de impliciete strategie, maar ook hun gedrag in de begeleiding in de expliciete strategie tussen de MERs in. De instructie om zich te gedragen als ze gewend waren tijdens onderzoeksopdrachten van hun leerlingen, gaf onvoldoende duidelijkheid over de inbedding van de elementen van begrip van bewijs daarin.

- *Planning en tijdsbesteding*

Een eigen onderzoek doen in een periode van tien weken terwijl ook ander schoolwerk moet gebeuren, bleek een moeilijke taak voor de leerlingen. De meeste van hen lukte het niet om de gegeven planning te halen, hetgeen tot tijdsdruk leidde aan het eind van de beschikbare periode.

Vooral in de expliciete onderwijsleerstrategie leidde dat tot problemen, aangezien de MERs daardoor niet meer in de pas liepen met het onderzoek van de meeste leerlingen. De elementen van begrip van bewijs die in de MERs aan de orde kwamen, waren namelijk gericht op de fasen waarin het leerlingonderzoek geacht werd te zijn. De toepassing van die elementen op hun eigen onderzoek werd echter bemoeilijkt door het feit dat de leerlingen die fasen nog niet hadden bereikt op het moment dat de MERs plaatsvonden. Als je zelf nog geen onderzoeksvraag hebt geformuleerd kun je ook niet op de kwaliteit ervan reflecteren.

Dit leidde ook tot irritatie bij de leerlingen. Toen ze de klas binnenkwamen en constateerden dat het derde MER op het programma stond, riepen sommigen uit dat ze liever aan hun eigen onderzoek wilden werken, in plaats van aan 'nog zo'n MER'. Dat geeft aan dat de het belang van de MERs voor het eigen onderzoek niet voldoende duidelijk werd (gemaakt). Daarnaar gevraagd, zeiden sommigen dat elementen uit de MERs wel bruikbaar waren, maar het toepassen op hun eigen onderzoek vonden ze moeilijk, omdat naar hun mening dat onderzoek 'anders' was. Het verschil zat bijvoorbeeld in het feit dat zij niet verschillende alternatieve hypotheses naast elkaar hadden (zoals in MER 2) die ze op basis van waarnemingen al dan niet zouden kunnen verwerpen.

- *Zijn de momenten van expliciete reflectie uitvoerbaar?*

Tijdsdruk was niet alleen een probleem voor het onderzoek van de leerlingen, ook de meeste momenten van expliciete reflectie pasten krap in de toebemeten tijd, één lesuur. Hoewel ze zorgvuldig gepland waren samen met de docenten, inclusief een tijdsindeling, vroegen de opdrachten of de discussies meer tijd dan was voorzien. Gevolg daarvan was dat er weinig of geen tijd overbleef om de uitkomsten van de discussie op het eigen onderzoek te betrekken.

In het eerste MER liepen de leerlingen volledig over de onduidelijkheden in de vraagstelling 'hoeveel voedingsstoffen bevat melk?' heen. Ze begonnen meteen een van de aangeboden typen melk te testen met de aanwezige indicatoren. Niet verrassend, want leerlingen zijn gewend uit te voeren wat ze opgedragen krijgen en dat zijn maar al te vaak niet volledig duidelijke taken. Velen stellen dan geen vragen, maar gaan gewoon aan het werk, zeker als de spullen klaar staan. Sommige leerlingen aarzelden echter wel en vroegen bij zichzelf of bij medeleerlingen 'welke melk moet ik nemen?' Maar vervolgens kozen ze er gewoon één uit en

gingen door met de opdracht. In de discussie inventariseerde de docent de resultaten en trok daaruit de conclusie dat de leerlingen de onderzoeksvraag elk op hun eigen manier geïnterpreteerd hadden. Op hun werkblad werd de leerlingen gevraagd wat ze aan hun eigen onderzoeksvraag zouden gaan veranderen. Van degenen die al een vraag hadden geformuleerd vermeldden sommigen dat ze niets zouden veranderen, anderen probeerden hun formulering eenduidiger te maken. Bijvoorbeeld van 'verdwijnt na verloop van tijd vitamine C uit sinaasappelsap?' maakten ze 'vermindert de hoeveelheid vitamine C in een glas met 200 ml sinaasappelsap in een periode van twee dagen?'

Het probleem van de stottende gazellen in MER 2 leidde tot geanimeerde discussies in de groepen. Soms probeerden ze de verschillende hypotheses te beoordelen op basis van hun eigen inschatting van de situatie: 'Waarom zou een gazelle aan een predator willen meedelen dat hij gezien is? Hij zal zo snel mogelijk vluchten!' In de klassendiscussie volgden ze actief de redenering van de docent. Op het werkblad werd gevraagd in hoeverre de onderzoekers zeker konden zijn van hun conclusie. Sommige leerlingen zeiden van wel: 'De observaties komen alleen overeen met de communicatiehypothese'. Anderen zeiden van niet: 'Nee, want het is een hypothese, het moet nog uitgezocht worden.' Dat je een hypothese kan toetsen door de voorspelling met empirische gegevens te vergelijken, werd dus niet door iedereen begrepen. Toepassing van dit begrip op het eigen onderzoek werd ook bemoeilijkt doordat de meeste leerlingen – als ze er in hun eigen onderzoek al aan toe waren op dat moment – slechts één hypothese hadden geformuleerd. Bovendien was die hypothese meestal niet meer dan een bevestigend antwoord op de onderzoeksvraag (zie paragraaf 2.3; Schalk, 2002c; Schalk, Schermer & Bax, 2004). Toepassing van de essentie van deze reflectie op hun eigen onderzoek vonden de leerlingen daardoor moeilijk.

De relatie met het eigen onderzoek werd makkelijker gelegd in het derde MER. Het meeste commentaar van de leerlingen op het voorgelegde onderzoeksverslag betrof het onder controle houden van variabelen. Al tijdens de opdracht maakten sommige leerlingen de vergelijking met hun eigen onderzoek: 'In ons onderzoek moeten we zorgen dat we de temperatuur constant houden.' Het verborgen volgorde-effect werd door de meeste leerlingen wel opgemerkt, maar de docent moest het probleem van de dubbele blanco aanwijzen en uitleggen. De docenten merkten op dat de leerlingen makkelijker werk van anderen dan hun eigen werk van kritisch commentaar voorzien: 'Ze raken erg betrokken, het raakt ze, het wordt echt voor ze.' Dit derde moment van expliciete reflectie stimuleerde de leerlingen om met hernieuwde energie met hun eigen onderzoek verder te gaan, meer dan dat het stimuleerde om kritisch naar hun eigen onderzoeksopzet te kijken. Ernaar gevraagd, lukte het de leerlingen wel om onafhankelijke en afhankelijke variabele in hun eigen onderzoek te noemen (begrip van bewijs, element 8), maar hadden ze moeite om alle andere variabelen te onderkennen en te beheersen (element 9) en om een blanco te omschrijven (element 10).

In het laatste MER viel de opdracht en de toepassing op het eigen onderzoek samen, want de leerlingen moesten commentaar geven op elkaars concept-verslag. Bij het begin was er de nadruk op gelegd dat ze elkaar alleen op een constructieve manier kritiek mochten geven. Op een van de scholen waren de conceptverslagen nog niet beschikbaar op het moment van het vierde MER. De leerlingen hadden daarom korte beschrijvingen gemaakt van de beoogde verslagen. Veel commentaar dat ze elkaar gaven had de vorm van: 'Ik snap niet waarom je …' of: 'Dit aspect moet je wat duidelijker maken.' Dat in een van de concepten de onderzoeksvraag en de conclusie ontbraken werd niet als een groot probleem gezien. Erg weinig elementen van begrip van bewijs werden genoemd in de commentaren, ook al waren ze genoemd op het werkblad. Omdat er niet altijd een conclusie was, werd de vraag hoe zeker je

kon zijn van de juistheid van de conclusies helaas niet altijd beantwoord. Waar hij wel werd beantwoord, werd niet duidelijk op welke manier de leerlingen daarbij gebruikt maakten van 'begrip van bewijs'. Soms noemden ze wel aspecten ervan, maar bleef onduidelijk hoe die meespeelden in hun oordeel, bijvoorbeeld: 'Hun conclusie is juist, maar we snappen niet waarom ze het experiment niet herhaald hebben', en: 'Hun antwoord op hun onderzoeksvraag is juist; tijdens hun experiment waren alle omstandigheden constant, behalve één variabele.' Hoewel in termen van begrip van bewijs de commentaren niet vaak bruikbaar waren, waardeerden de leerlingen de feedback van hun medeleerlingen. Twee duo's kwamen zelfs overeen elkaars volgende versies weer van commentaar te voorzien.

De scenario's van de momenten van expliciete reflectie werden door de docenten als ondersteunend ervaren. Vooral de beschrijving van wat zijzelf geacht werden te doen vonden ze steunend. Anderzijds leken de docenten daardoor soms het doel van het MER – in termen van beoogde leerresultaten – te 'vergeten'. Ze waren vooral gericht op hetgeen gedaan 'moest' worden volgens het scenario en misten daardoor mogelijkheden om het doel via andere wegen te bereiken.

Samenvattend kan geconstateerd worden dat de momenten van expliciete reflectie niet geheel verliepen zoals gepland en de uitvoerbaarheid ervan tegenviel. Hoewel het wel lukte om de opdrachten en discussies uit te voeren, namen ze als geheel te veel tijd, werd het belang ervan onvoldoende duidelijk en pasten de leerlingen de betrokken elementen van begrip van bewijs slechts mondjesmaat toe op hun eigen onderzoek. Desondanks merkte een van de docenten op dat hij ze het volgende jaar opnieuw zou gebruiken, zelfs als ons gezamenlijke project (het in dit proefschrift beschreven onderzoek) niet voortgezet zou worden, want 'ze stellen een heleboel belangrijke aspecten van onderzoek doen aan de orde.'

5.4.2 Effectiviteit

Om leereffecten te kunnen constateren zijn er voor- en natoetsen afgenomen, zijn de werkbladen en verslagen van de leerlingonderzoeken geanalyseerd, en zijn aan enkele leerlingen in interviews nieuwe problemen voorgelegd.

- *Toetsresultaten*

De betrouwbaarheid van de toets viel dermate laag uit dat op basis daarvan geen uitspraken gedaan kunnen worden over eventuele toename van de beheersing van begrip van bewijs. Als er echter op een kwalitatieve manier naar de antwoorden op de toets gekeken wordt, kunnen wel enkele opmerkingen gemaakt worden.

Zo bevatten bijvoorbeeld de antwoorden op de vragen over de kwaliteit van onderzoeksvragen (element 1) niet alleen argumenten met betrekking tot eenduidigheid, specificiteit en afperking, maar ook met betrekking tot de inhoud – bijvoorbeeld: 'Onderzoeksvraag II vind ik minder goed, omdat het per mens verschillend is' – en hoe interessant een leerling de vraag vindt. In de natoets was die laatste categorie met name in de expliciete groep kleiner in omvang. Met een correcte onderzoeksopzet hadden de leerlingen problemen, het herkennen of benoemen van de afhankelijke en onafhankelijke variabele (element 8) lukte zelden. Het onderkennen van de variabelen die onder controle moeten worden gehouden (9) ging beter, maar dan nog deden de meesten het niet correct. Het begrip 'blanco' (10) was vrijwel onbekend. Bij het beoordelen van de kwaliteit van een beschreven onderzoek (21) kunnen leerlingen zwakke punten aanwijzen, met name wat het aantal waarnemingen (14) en het onder controle houden van variabelen betreft (9). Dit geldt voor de antwoorden op zowel de voortoets als de natoets.

Tabel 5.4 Begrip van bewijs in de verslagen van de onderzoeksteams (N = 28). 'Begrip van bewijs' wil zeggen dat aan een van toepassing zijnd criterium is voldaan of beargumenteerd is waarom er niet aan is voldaan.

element	*omschrijving*	*aantal keer van toepassing*	*aantal keer 'begrip van bewijs'*	*percentage 'begrip van bewijs'*
1	formuleren onderzoeksvraag	79	53	67
2	onderscheid beschrijvend / hypothesetoetsend onderzoek	28	0	0
3	formuleren hypothese	97	72	74
4	toetsbaarheid hypothese en uitvoerbaarheid	23	19	83
5	meer dan één hypothese	28	0	0
6	voorspelling	28	9	32
1-6	*onderzoeksvraag, hypothese en voorspelling*	*283*	*153*	*54*
7	kenmerken bij beschrijvend onderzoek	4	3	75
8	onafhankelijke en afhankelijke variabele	24	19	79
9	controlevariabelen	25	15	60
10	blanco	16	8	50
11	invloed metingen op uitkomst	27	20	74
12	intervallen van variabelen	13	8	62
13	gevoeligheid meetinstrument	28	16	57
14	steekproef aselect en groot genoeg	59	32	54
15	omgaan met uit de toon vallende resultaten	14	5	36
16	verantwoord middelen	9	5	56
17	onderkennen significantie	12	1	8
7-17	*variabelen, steekproef en metingen*	*231*	*132*	*57*
18	conclusie sluit aan op onderzoeksvraag	28	21	75
19	onderscheid correlatie en causaal verband	5	1	20
20	verschil oorzakelijke en functionele verklaring	9	2	22
21	uitspraken over betrouwbaarheid en validiteit	28	7	25
22	vergelijking met ander onderzoek	28	1	4
23	vergelijking met theorie	28	2	7
18-23	*conclusie en evaluatie*	*126*	*34*	*27*
1-23	*alle elementen*	*640*	*318*	*50*

- *Leerlingverslagen*

De verslagen van de leerlingen zijn op dezelfde manier geanalyseerd als de leerlingverslagen in de voorstudie. Voor alle elementen van begrip van bewijs is beoordeeld of ze van toepassing waren in het betreffende onderzoek, zo ja, of eraan voldaan was en als er niet aan voldaan was, of dat dan beargumenteerd was. De resultaten van de analyse staan in tabel 5.4. Relatief vaak werd een toetsbare hypothese (element 3 en 4) geformuleerd, maar nooit werd meer dan één hypothese vermeld (5). De kenmerken en variabelen die onderzocht werden,

Tabel 5.5 Percentages begrip van bewijs voor verschillende combinaties van elementen (gerealiseerd ten opzichte van mogelijk) in de artikelen in de verschillende groepen en in de voorstudie (par. 3.4). AS: Amsterdamse School; SA: School in Amstelveen.

	expliciet (N=13)	*impliciet (N=15)*	*AS (N=13)*	*SA (N=15)*	*Totaal (N=28)*	*voor-studie*
Onderzoeksvraag, hypothese & voorspelling (1-6)	58	50	53	55	54	67
Variabelen, steekproef & metingen (7-17)	58	56	56	58	57	64
Conclusie & evaluatie (18-23)	28	26	34	21	27	43
Alle elementen	52	48	51	49	51	61

werden vaak duidelijk vermeld (7 en 8), het controleren van de andere variabelen (9) en de blanco kwamen minder vaak uit de verf. Het onderkennen van mogelijke invloed van de metingen op de uitkomst (11) werd door veel leerlingen gedaan, aan de andere criteria met betrekking tot steekproef en metingen werd in redelijke mate voldaan, met uitzondering van het omgaan met uit de toon vallende resultaten (15) en statistische significantie (17). Het formuleren van een bij de onderzoeksvraag aansluitende conclusie (18) ging de meeste leerlingen goed af, maar de andere elementen met betrekking tot conclusie en evaluatie (19-23) werden minder goed toegepast. Aan de laatste drie elementen, van toepassing op elk onderzoek, werd weinig tot nauwelijks voldaan. Zelden werden de resultaten of de conclusie met ander onderzoek of theorie vergeleken (22 en 23) en slechts zeven keer werd het onderzoek wat betreft validiteit en betrouwbaarheid geëvalueerd (21).

Met betrekking tot al deze elementen van begrip van bewijs werden weinig verschillen tussen de expliciete en impliciete groepen en tussen de scholen waargenomen (tabel 5.5). Het grootste verschil is 13 procent, tussen de scholen wat betreft de elementen 18 tot 23. Nadere analyse wijst dat dit veroorzaakt wordt door meer bij de onderzoeksvraag passende conclusies (element 18, 11 van de 13 tegenover 10 van de 15) en doordat de vergelijking met ander onderzoek en theorie alleen in Amsterdam heeft plaatsgevonden (elementen 22 en 23).

Vergelijking met de getallen uit de voorstudie (tabel 5.5, meest rechtse kolom) laat zien dat de leerlingen in deze eerste cyclus over de hele linie minder presteren. De praktische opdracht, al dan niet in combinatie met de MERs, heeft geen groot effect gehad op het bewaken van de kwaliteit van het door de leerlingen uitgevoerde onderzoek; dat komt althans niet tot uitdrukking in de verslagen.

Wat verder opviel aan de verslagen was de informele toon die gebezigd werd. Veel leerlingen vermeldden allerlei irrelevante, persoonlijke opmerkingen in hun verslag zoals: ‘Onze samenwerking liep perfect. Nadat we hadden besloten samen te gaan werken omdat onze ‘partners’ met lustrum waren gegaan is alles eigenlijk heel geregeld gegaan. We hebben als eerste de donderdag erop een werkplan gemaakt …’. Dit illustreert het bewaken van de kwaliteit van hun onderzoek voor de leerlingen geen hoofdzaak was.

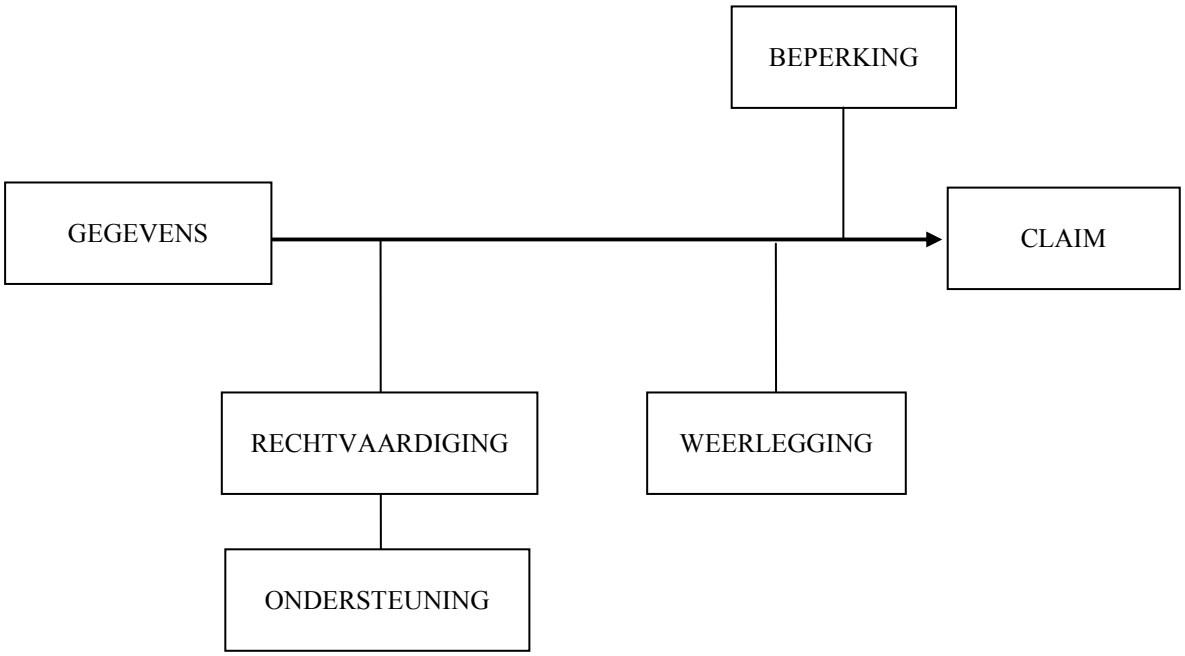
- *Interviews*

Uit zowel de expliciete als de impliciete groep van elke school is één duo op verschillende momenten tijdens de praktische opdracht geïnterviewd. In die interviews werd de leerlingen steeds een nieuwe opgave voorgelegd dat verband hield met de elementen van begrip

van bewijs die in het MER aan de orde waren geweest en werden het MER nabesproken (alleen met de leerlingen van de expliciete groep). De argumentaties van de leerlingen bij de opgaven zijn geanalyseerd.[20] Primair doel hierbij was om te kijken of en op welke manier leerlingen elementen van begrip van bewijs gebruikten bij het argumenteren over onderzoek. Doel was niet om uitspraken te doen over argumentatieve vaardigheden van de leerlingen noch over geldigheid van de argumentaties.

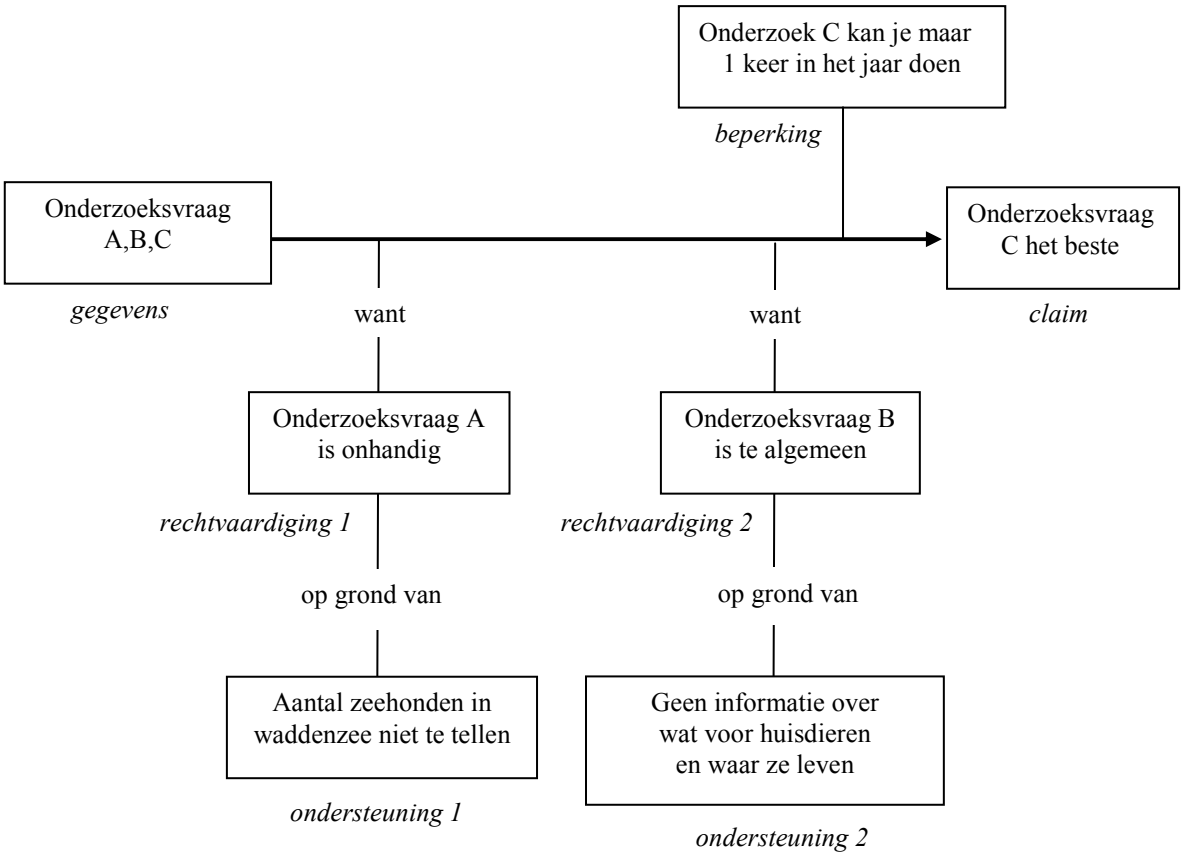
Voor de analyse hebben we gebruik gemaakt van de methode van Toulmin (1958). Die is uitermate geschikt om de structuur van argumentaties weer te geven zonder uitspraken te doen over de juistheid ervan. Het is een model dat de essentiële elementen van een argumentatie en hun functionele relaties weergeeft (figuur 5.2). De structuur van het argument ziet er dan als volgt uit: Aangezien (*gegevens*) ... en omdat (*rechtvaardiging*) ... op grond van (*ondersteuning*).... dus (*claim*). Als aanvulling onderscheidt Toulmin twee andere kenmerken in complexe argumentaties: een *beperking* geeft de mate van zekerheid van de claim aan en een *weerlegging* specificeert de situaties waar de claim niet geldig is. In figuur 5.3 is een argumentatie uit een interview weergegeven volgens het model, waarin overigens een weerlegging ontbreekt.

Uit de analyse bleek dat leerlingen inderdaad in hun argumentaties elementen van begrip van bewijs gebruikten, maar ook andere overwegingen, zoals ook blijkt uit de inhoudelijke analyse van de antwoorden op de toets. De leerlingen bleken in hun argumentaties meer elementen te gebruiken die niet met begrip van bewijs te maken hadden dan wel (figuur 5.4). De claim wat de beste onderzoeksvraag is, werd bijvoorbeeld ook gerechtvaardigd met interesse of belang van zo'n onderzoek: 'Ik vind die vraag veel interessanter' of: 'Daar is al redelijk veel onderzoek naar gedaan, dacht ik.' Of er werden overwegingen aangevoerd die inhoudelijk op de vragen ingingen: 'Dan is er ook nog het verschil dat je tussen katten en leeuwen hebt en zo …' Hoewel het maar weinig uitspraken betreft (17), komen de beperkingen wel in meerderheid vanuit begrip van bewijs (63 procent).

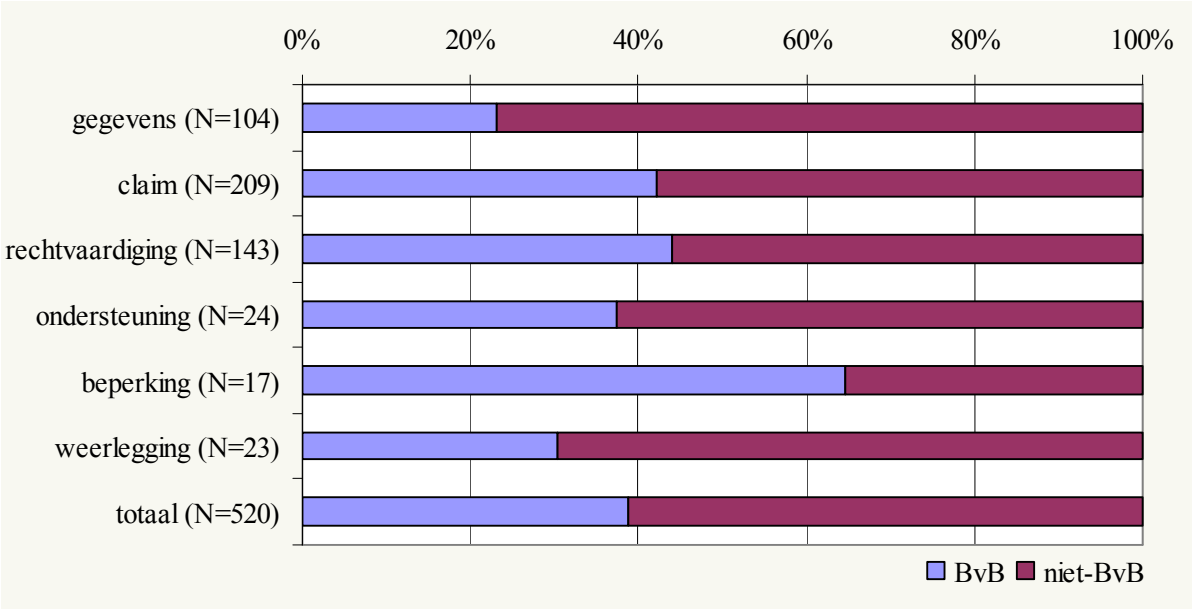


Figuur 5.2 Argumentatiemodel volgens Toulmin (1958) in schemavorm.

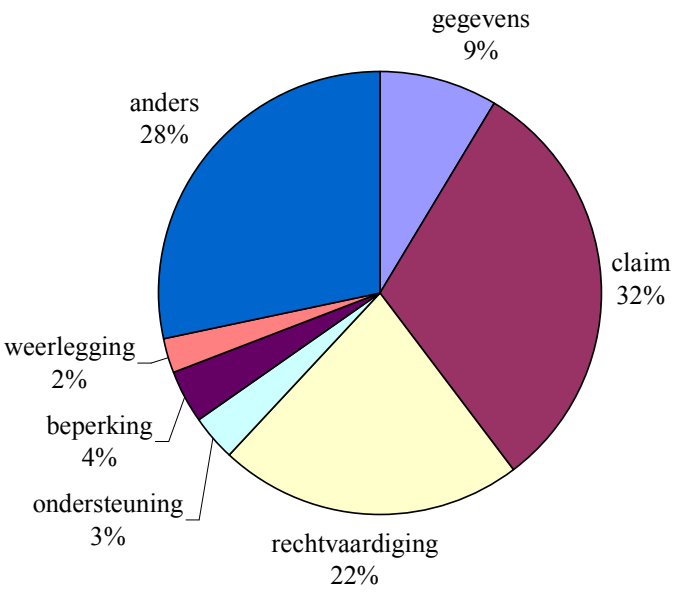
[20] De analyse is uitgevoerd in het kader van een onderzoeksstage waarvan elders uitgebreid verslag is gedaan (Yuksel, 2004; Schalk & Yuksel, 2004).



Figuur 5.3 Argumentatieanalyse volgens het model van Toulmin in schemavorm. Argumentatie uit interview 1, met een beperking en zonder weerlegging. De opgave waarop dit interview betrekking had is te vinden in bijlage 3.



Figuur 5.4 Percentage uitspraken in de leerlinginterviews waarin een element van begrip van bewijs werd gebruikt in verschillende delen van een argumentatie. Linker deel van elke balk: met begrip van bewijs; rechter deel: anders dan met begrip van bewijs.



Figuur 5.5 Verdeling van de uitspraken die een element van begrip van bewijs bevatten over de delen van een argumentatie. Anders: uitspraken die niet tot een van de delen van een argumentatie gerekend kunnen worden. Het totaal aantal uitspraken waarin een element van begrip van bewijs is opgenomen is 282.

De argumentaties van de leerlingen waren vaak eenvoudig van opbouw: een claim met een rechtvaardiging, waarbij de gegevens meestal impliciet bleven. Figuur 5.5 laat dit ook zien: begrip van bewijs werd het vaakst gebruikt in de claims en de rechtvaardigingen.[21] Een uitspraak die dit illustreert over de betrouwbaarheid van onderzoek (element 21) is de volgende. De uitspraak gaat over de vraag welke gegevens, uit eigen onderzoek dan wel uit onderzoek van anderen, een hypothese het best ondersteunen: 'Kijk, dit is een eigen observatie en dit is een eigen experiment *(gegevens)*, dan zou ik zeggen, dat je dat goed kan vertrouwen *(claim)*.' Daarnaast werden elementen van begrip van bewijs ook vaak gebruikt buiten de argumentaties, dus zonder dat ze bijdroegen aan een redenering om een claim te ondersteunen.

Al met al kan er weinig leereffect waargenomen worden in toets, verslagen en argumentaties. Dat kan twee oorzaken hebben, namelijk dat er te weinig geleerd is of dat de meetinstrumenten niet adequaat meten. Voor beide is wat te zeggen. De eerder in deze paragraaf gedane constatering dat de uitvoerbaarheid te wensen over laat, heeft natuurlijk zijn weerslag in de resultaten. Het onderwijs was niet optimaal, dus verwacht dat ook niet van de resultaten. Dat ook de meetinstrumenten tegen het licht gehouden moeten worden, is meer dan een theoretische opmerking. Bij nadere beschouwing van de toets en de toetsresultaten blijkt dat de toets te lang was (veel leerlingen hadden de laatste opgaven niet gemaakt), dat de betrouwbaarheid te laag is en dat sommige vragen toch het betreffende begrip van bewijs minder goed operationaliseren dan gedacht. De analyse van de argumentaties blijkt een erg arbeidsintensieve aangelegenheid, zowel voor het onderkennen van de verschillende delen van argumentaties als voor het scoren van de verschillende elementen van begrip van bewijs. Bovendien maakt het wel duidelijk hoe de leerlingen de criteria gebruiken, maar niet hoe de ontwikkeling

[21] De relatie tussen figuur 5.4 en 5.5 is de volgende: de uitspraken uit de linker delen van de balken van figuur 5.4 bepalen de overeenkomstige segmenten van het cirkeldiagram van figuur 5.5; daaraan zijn als categorie 'anders' de uitspraken toegevoegd die niet als een deel van een argumentatie gescoord konden worden, maar wel een element van begrip van bewijs bevatten.

van begrip van bewijs plaatsvindt. Vergelijking van het gebruik in de loop van de tijd kan wel verschillen zichtbaar maken, maar nog steeds niet in hoeverre de leerlingen zich de criteria eigen maken op de manier zoals in hoofdstuk 4 beschreven is. Daarmee draagt het te weinig bij aan de beantwoording van de derde onderzoeksvraag (par. 1.2). Het lijkt daarom beter in de tweede cyclus een andere manier van analyseren van de gesprekken van leerlingen te gebruiken.

Dat er in de toets, de verslagen en de interviews weinig of geen onderscheid tussen de resultaten van de twee strategieën te constateren is, zou kunnen betekenen dat beide even geschikt of ongeschikt zijn voor het aanleren van begrip van bewijs, maar ook dat ze beide nog niet 'uitgesproken' genoeg zijn. Een heroverweging en verdere articulatie van het onderscheid blijkt nodig. In beide strategieën is nog meer aandacht nodig voor het toepassen van de elementen van begrip van bewijs, zowel in de directe begeleiding als in de momenten van expliciete reflectie. In de volgende paragraaf wordt omschreven op welke punten het ontwerp verbeterd kan worden.

5.5 Verbeterpunten voor de tweede cyclus

Wat de eerste ronde ons geleerd heeft voor de revisie van de onderwijsleerstrategieën is samen te vatten in de volgende punten.

- *Aanscherping van de uitgangspunten*

Om een verbetering van de beide onderwijsleerstrategieën te kunnen bewerkstelligen, is een heroverweging en aanscherping van de uitgangspunten zoals onderbouwd en verwoord in hoofdstuk 4 nodig. Met name de gerichtheid van de leerlingen op het bewaken van de kwaliteit van hun eigen onderzoek is onvoldoende. Leerlingen blijken onvoldoende motief te hebben ontwikkeld om hun kennis daarover uit te breiden. De rol van hun motief en de wijze om de ontwikkeling daarvan te stimuleren verdienden nieuwe aandacht.

De invloed van de expliciete reflectie op de ontwikkeling van begrip van bewijs bij de leerlingen is in de eerste ronde niet duidelijk geworden. Dat is gedeeltelijk veroorzaakt door de manier waarop deze in de MERs is vormgegeven, maar een andere manier van analyseren kan wellicht ook meer zichtbaar maken. Dat betekent dat de manier van analyseren van de (gesproken) interactie tussen de leerlingen en tussen docent en leerlingen beter de ontwikkeling van begrip van bewijs zichtbaar moet maken. Daarvoor is een extra theoretische exercitie nodig voordat de opzet voor de tweede ronde gemaakt kan worden.

- *Duidelijker omschrijving van het beoogde docentgedrag bij de begeleiding*

De beide strategieën moeten duidelijker ingevuld en omschreven worden. De docenten waren te vaak te onzeker over wat van hen verwacht werd. Beide strategieën hebben als doel om te leren de kwaliteit van onderzoek te bewaken door een aantal elementen van begrip van bewijs te leren. Het verschil zit in de manier waarop die elementen aangeboden worden. In de tweede ronde wordt het begeleidingsgedrag van de docenten meer hierop gericht. Ze moeten meer dan in de eerste ronde alert zijn op mogelijkheden om begrip van bewijs aan de orde te stellen. In de aansturing van de docenten is dat tot uitdrukking gebracht.

- *Betere oriëntatie van de leerlingen op het doel van de opdracht*

Dat het doel van de praktische opdracht beter voor ogen moet worden gehouden, geldt niet alleen voor de docenten, maar eveneens voor de leerlingen. Zij waren erg gefocust op het

doen van hun onderzoek en veel minder op het bewaken van de kwaliteit ervan met behulp van de elementen van begrip van bewijs. De momenten van expliciete reflectie werden door hen dan ook soms als lastige onderbreking gezien, in plaats van als nuttige pas op de plaats. Vanaf het begin van de opdracht af aan moet het doel 'het leren bewaken van de kwaliteit' meer benadrukt worden.

De geringe gerichtheid op het bewaken van de kwaliteit kwam ook tot uitdrukking in de toon van de verslagen. De verslaggeving zou meer op die hoofdzaak gericht moeten zijn. De eisen die aan het eindproduct gesteld worden moeten dat weerspiegelen en moeten ook duidelijker onder de aandacht van de leerlingen gebracht worden.

- *Meer structuur en overzicht*

In de eerste ronde kregen de leerlingen een erg open opdracht waarmee ze aan de slag moesten. Daarbij was het vaak onduidelijk wat ze precies moesten doen, hoe verschillende fasen in het onderzoek zich tot elkaar verhielden en hoe de planning in elkaar zat. Dat leidde tot vertraging en een gebrek aan overzicht bij de leerlingen. Het lijkt raadzaam de leerlingen meer structuur aan te bieden in de vorm van een boekje waarin de informatie is opgenomen over de bedoeling, planning, eisen en randvoorwaarden van de praktische opdracht en informatie over het doen van onderzoek, zoals criteria, fasering en een voorbeeld van een beoogd eindproduct.

Het tijdsbestek waarin de opdracht moest worden uitgevoerd was erg lang, gedurende tien weken één lesuur per week. Dit maakt concentratie en het houden van overzicht lastig. Hetzelfde aantal lessen in een kortere periode lijkt om die reden te verkiezen.

- *Decentralisatie en herziening van de MERs*

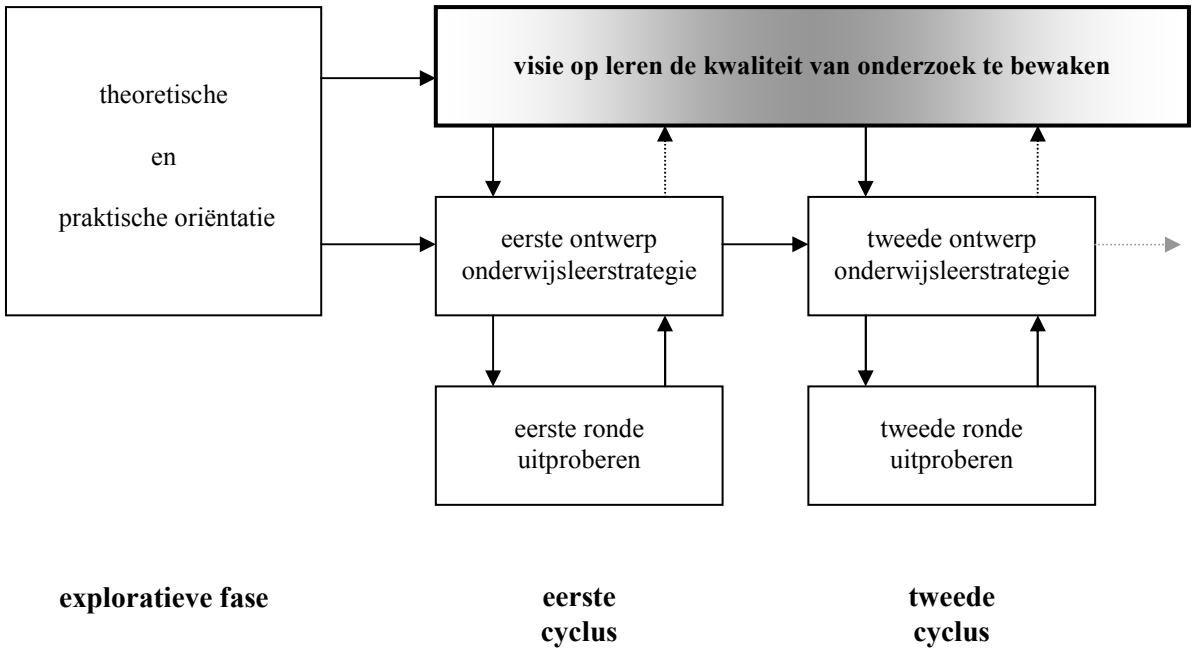
De momenten van expliciete reflectie vielen inhoudelijk vaak niet samen met de fase van het onderzoek van de leerlingen. Hen werd bijvoorbeeld gevraagd na te denken over de criteria voor een goede hypothese op het moment dat ze in hun eigen onderzoek daar nog helemaal niet mee bezig waren. Omdat de leerlingen in verschillend tempo en in een verschillende volgorde door de fasen van hun onderzoek lopen, lijkt decentralisatie geboden. Dan kunnen de leerlingen namelijk zo'n opdracht maken op het moment dat ze eraan toe zijn in hun onderzoek. Alleen voor de vierde reflectieopdracht, waarin wederzijdse feedback op de conceptartikelen plaatsvindt, blijft afstemming tussen de leerlingen noodzakelijk.

Ook de inhoud van de opdrachten behoeft aanpassing. Met name dient de terugkoppeling op het eigen onderzoek beter ingebouwd te worden, aangezien die in de eerste ronde vaak tussen wal en schip terecht kwam. Bovendien bleken de opdrachten te veel tijd te kosten. Inkorten is dus eveneens noodzakelijk.

- *Herziening meetinstrumenten*

Het niet kunnen constateren van leereffect en van verschillen daarin tussen beide strategieën zou kunnen liggen aan de gebruikte instrumenten om dat leereffect in kaart te brengen. Er moet kritisch naar de toets gekeken worden. Analyse van interviews met de methode van Toulmin maakt het in principe mogelijk om ontwikkeling waar te nemen in het gebruik van elementen van begrip van bewijs, maar laat niet zien hoe die in de uitvoering van de praktische opdracht tot stand komt. Met andere woorden: het laat niet zien of en hoe het proces van de 'spiral formation of mental actions' (fig. 4.2) plaatsvindt. De analyse zou die zichtbaar moeten kunnen maken, evenals de invloed die reflectie – zowel '*in action*' (tijdens het werken aan het onderzoek) als '*on action*' (in de reflectiemomenten) – daarop heeft in de interactie tussen leerlingen onderling en tussen leerlingen en docent. De manier waarop deze ontwikkeling het best gevolgd kan worden, moet dus eveneens heroverwogen worden.

Het volgende hoofdstuk is gewijd aan het aanscherpen van de theoretische uitgangspunten voor de onderwijsleerstrategieën en de wijze van volgen van de ontwikkeling van begrip van bewijs bij de leerlingen. In de eerste paragrafen van hoofdstuk 7 wordt de uitwerking daarvan in een verbeterd ontwerp en instrumenten beschreven.



6 Aanscherping

6.1 Meer motief
6.2 Een didactische structuur
6.3 De ontwikkeling van begrip van bewijs
6.4 Verwachte samenhang tussen onderwijsleerstrategie en patronen in de ontwikkeling van begrip van bewijs

'Needed are the voices of the teacher and students, which are vital to developing understanding of the nature of an inquiry-based classroom.'
Barbara Crawford (2000, p. 933)

Hoofdstuk 6 Aanscherping

Een van de belangrijke tekortkomingen van de eerste ronde is de geringe gerichtheid van de leerlingen op het bewaken van de kwaliteit van hun eigen onderzoek. Leerlingen blijken onvoldoende gemotiveerd om hun kennis daarover uit te breiden. De rol van motieven en de wijze van motiveren krijgen daarom in paragraaf 6.1 nieuwe aandacht.

Dit leidt tot een herbezinning op de algehele structuur van de praktische opdracht. Hoe hangen de uitvoering van het onderzoek en de reflectie op de kwaliteit ervan samen en hoe leiden die uiteindelijk tot veralgemeniseerde criteria voor goed onderzoek? In paragraaf 6.2 wordt een didactische structuur uiteengezet die daar een mogelijk antwoord op geeft.

Tegelijkertijd is het voor de beantwoording van de derde onderzoeksvraag (zie par. 1.2) nodig om meer zicht te krijgen op hoe leerlingen begrip van bewijs ontwikkelen tijdens de uitvoering van hun onderzoek. Zoals in paragraaf 4.1 betoogd, is de (talige) interactie daarbij wezenlijk. In paragraaf 6.3 wordt een manier om daarnaar te kijken geïntroduceerd die daarvoor bruikbaar lijkt, zowel in de expliciete als in de impliciete onderwijsleerstrategie.

De heroverwegingen met betrekking tot de structuur en de manier om de ontwikkeling van begrip van bewijs te volgen, leiden tenslotte in paragraaf 6.4 tot het formuleren van verwachtingen met betrekking tot de beide onderwijsleerstrategieën. In hoofdstuk 7 wordt dat uitgewerkt in een aangescherpte opzet voor de praktische opdracht en de beoordeling ervan.

6.1 Meer motief

In paragraaf 4.1 werd de ontwikkeling van een activiteit uit de cultuurhistorische theorie – ten tonele gevoerd om de rol van motieven in het leren te verduidelijken: behoeften leiden tot motieven.[22] Motieven hebben zowel een emotionele als een cognitieve functie: ze zetten aan tot een activiteit en richten en organiseren die (Van Aalsvoort, 2004). In de context van het doen van onderzoek kan dat betekenen dat de behoefte iets te willen weten (het antwoord op een vraag te vinden) aanzet tot het doen van onderzoek, terwijl de behoefte iets *zeker* te willen weten dat onderzoek organiseert. Dat is althans de theorie. De praktijk blijkt weerbarstiger, want de behoefte iets te willen weten – de door de leerlingen gestelde onderzoeksvragen – bleek in de eerste ronde weinig motief op te roepen om het onderzoek te organiseren volgens criteria van goed onderzoek. Het belang van de reflectie op de kwaliteit van het onderzoek – of die nu naar aanleiding van het eigen onderzoek of aan de hand van de momenten van reflectie aan de orde kwam – werd te weinig opgeroepen en daardoor kwam de terugkoppeling van (tussen)product naar motief en activiteit (zie figuur 4.3 en 4.4) onvoldoende tot stand. De vraag 'weet ik het nu zeker genoeg?' werd door de leerlingen niet gesteld.

Dat doet de vraag rijzen hoe het tweede deel van het motief, het *zéker* willen weten, beter opgeroepen kan worden. In dit kader doen Lijnse en Klaassen (2004) zinvolle uitspra-

[22] Het is van belang onderscheid te maken tussen *motief* en *motivatie*: met motivatie wordt een algemene stimulans c.q. gerichtheid bedoeld, bijvoorbeeld motivatie om te leren, terwijl met motief een inhoudelijke reden om een bepaalde activiteit uit te voeren bedoeld wordt.

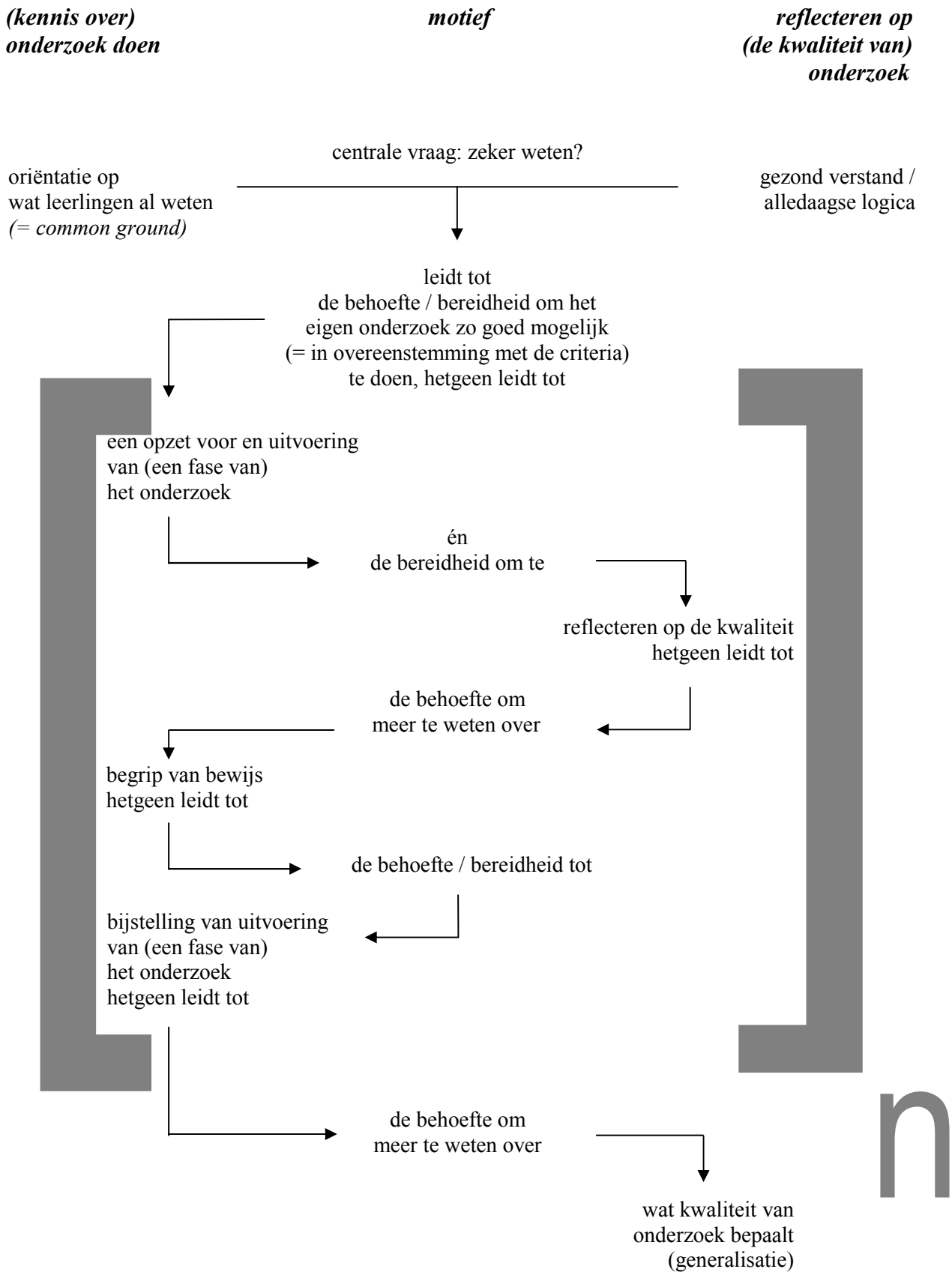
ken. Zij zeggen dat om leerlingen kennis aan hun conceptuele structuur te laten toevoegen, m.a.w. ze te laten leren, het nodig is om ze te laten zien wat de betekenis ervan voor hen is. De nadruk van wat zij de *probleemstellende benadering* noemen ligt er dan op de leerlingen in een dusdanige positie te brengen dat zij zelf de zin zien van – d.w.z. een inhoudelijk motief hebben voor – het uitbreiden van hun kennis, ervaringen en overtuigingen. Deze opgeroepen kennisbehoefte motiveert en stuurt dan weer de te ondernemen activiteiten. Dat betekent dat de zekerheid van het eventuele gevonden antwoord geproblematiseerd moet worden, met andere woorden: de vraag 'Zeker weten?' moet centraal staan.

De uitdaging is dan, volgens Lijnse en Klaassen (2004), om dit in een effectieve onderwijsleerstrategie om te zetten. De richting die deze auteurs wijzen, begint bij het vinden van *common ground*, de – bewust – gedeelde manier om naar de werkelijkheid te kijken. Daarbij is het belangrijk een evenwicht te vinden tussen *freedom from below* en *guidance from above*, waarmee wordt bedoeld dat je moet uitgaan van wat de leerlingen al weten en dat moet uitbreiden in de gewenste richting. Hoewel dit constructivistische uitgangspunt door velen gedeeld wordt, stellen zij dat zij er anders dan anderen vanuit gaan dat wat de leerlingen al weten en kunnen grotendeels correct is en meestal voldoende basis is om de *common ground* te vinden. Het gaat dan om het expliciteren van die gezamenlijke basis: wat weten we al van onderzoek doen? De kennis die al in de groep – bij de verschillende deelnemers – aanwezig is, kan gemeenschappelijk gemaakt worden en zonodig gecorrigeerd en geactualiseerd en vervolgens dienen als gezamenlijk referentiepunt. Dan kan ook gezamenlijk geconstateerd worden welke kennis en vaardigheden ontbreken, maar wel nodig lijken. Met andere woorden: er kan een gemeenschappelijke kennisbehoefte geformuleerd worden.

6.2 Een didactische structuur

Hoe leidt het bovenstaande naar een aanpassing van de praktische opdracht? Het betekent ten eerste dat er meer aandacht en ruimte moet zijn om een gezamenlijk vertrekpunt te vinden. Door samen op een rijtje te zetten wat je al weet van de kwaliteit van goed onderzoek, kunnen de activiteiten die volgen (het eigen onderzoek) daar meer op gericht worden. Bovendien kan bij de MERs of bij de impliciete begeleiding gerefereerd worden aan die gezamenlijke start. En daarnaast is het belangrijk dat de gezamenlijke start zo ingericht is dat die de behoefte oproept om er achter te komen hoe goed onderzoek eruit ziet, hoe je iets met zoveel mogelijk zekerheid te weten kunt komen.

Analoog aan de schema's die Lijnse en Klaassen (2004) in hun artikel presenteren, is in figuur 6.1 een hernieuwde didactische structuur van de praktische opdracht geschetst, waarbij ook de centrale vraag is geformuleerd, analoog aan andere uitwerkingen van de probleemstellende benadering (Knippels, 2002; Kortland, 2001). De figuur bestaat uit drie kolommen. De eerste kolom, '(kennis over) onderzoek doen', betreft de uitvoering van het onderzoek met alle daarbij behorende kennis en vaardigheden, dus ook de elementen van begrip van bewijs. De derde kolom, 'reflecteren op (de kwaliteit van) onderzoek', betreft het bewaken van de kwaliteit van het onderzoek, het nadenken over 'waar zijn we nu helemaal mee bezig en waarom'. De verbindende schakel wordt gevormd door het motief dat voortkomt uit de centrale vraag 'zeker weten?' Omdat een leerling zijn eigen onderzoek zo goed mogelijk wil doen – het antwoord zo zeker mogelijk wil weten – is hij bereid om erop te reflecteren en



Figuur 6.1 Didactische structuur van de lessenserie / voor het leren van begrip van bewijs. $[..]_n$ geeft aan dat het deel tussen de haken een aantal maal herhaald kan worden.

kan een kennisbehoefte naar kwaliteitscriteria ontstaan die tot bijstelling leidt. Deze serie 'gebeurtenissen' (uitvoering, reflectie en criteriumontwikkeling en bijstelling) kan een aantal keren plaatsvinden, steeds ten aanzien van een ander aspect of fase van het onderzoek, in de figuur aangeduid met $[\ldots]_n$. Uiteindelijk zal het uitvoeren van het onderzoek en de reflectie erop kunnen c.q. moeten leiden tot de vraag (en de beantwoording daarvan): 'Hoe weet ik of ik het zeker weet?' Het lijkt er dus op dat de bereidheid om te reflecteren op de kwaliteit staat of valt met een goede gezamenlijke oriëntatie op de centrale vraag 'zeker weten?'

Waar in de eerste ronde de leerlingen met hoofdzakelijk organisatorische informatie aan het werk gestuurd werden, is er voor de tweede ronde een meer inhoudelijke start voorzien: waar gaat het om, wat weten we daar al van en hoe breiden we onze kennis uit?; een stevige *common ground* waar de rest van de praktische opdracht op kan worden gefundeerd.

Of de stappen naar de reflectie en terug naar de bijstelling gemaakt worden, hangt af van de motieven daarvoor. In de impliciete onderwijsleerstrategie moet dat motief van het eigen onderzoek zelf komen, al dan niet verwoord door de docent ('weet je dit nu zeker?'). In de expliciete strategie worden de stappen naar reflectie en bijstelling ook extern aangestuurd door de momenten van reflectie. Ze moeten dan wel op het juiste 'tijdstippen' aangestuurd worden, dat wil zeggen als de criteria aan de orde zijn in het eigen onderzoek. Daarom lijkt het beter de momenten van expliciete reflectie te decentraliseren, dat wil zeggen om het ieder duo op zijn eigen tijdstippen te laten doen in plaats van op voor iedereen dezelfde zoals in de eerste cyclus..

Na een aantal cycli van uitvoering, reflectie en bijstelling zal de stap naar generalisatie ook aandacht moeten krijgen. Dat is geen vanzelfsprekende stap, de leerlingen hebben niet uit zichzelf een motief om tot veralgemenisering over te gaan. Het initiatief zal van het materiaal of de docent moeten komen.

In hoofdstuk 7 wordt deze didactische structuur verder uitgewerkt in de bijstelling van de praktische opdracht.

6.3 Meer zicht op de ontwikkeling van begrip van bewijs

In paragraaf 5.5 is aangegeven dat een extra theoretische exercitie nodig is om meer zicht te krijgen op de manier waarop leerlingen begrip van bewijs ontwikkelen. Het blijkt dat de in de eerste ronde gebruikte wijze van analyseren – kijken naar de argumentaties van leerlingen aan de hand van het model van Toulmin (1958) – niet genoeg informatie levert om zinvolle uitspraken te kunnen doen over de effectiviteit van de beide onderwijsleerstrategieën op dat terrein. Met name de invloed van de interactie tussen leerlingen onderling en tussen de leerlingen en de docent op het ontwikkelen van begrip van bewijs blijft in het model van Toulmin buiten beeld, terwijl die in het proces van (impliciet of expliciet aangestuurd) reflecteren wel belangrijk zijn.

Daarom is een ander analyse-instrument ontwikkeld dat wél de ontwikkeling van begrip van bewijs in beeld brengt, zowel de inhoudelijke ontwikkeling als de wijze waarop het zich in de interacties tussen docent en leerlingen en tussen leerlingen onderling ontwikkelt.

- *Uitspraken en episodes*

Alvorens te beschrijven welke categorieën het ontwikkelde analyse-instrument kent, is het van belang om aan te geven op welke eenheden die categorieën betrekking hebben. Wat geanalyseerd is, zijn de gesprekken die tijdens de lessen plaatsvonden; klassengesprekken als die plaatsvonden en tevens de gesprekken van in totaal acht onderzoeksduo's, zowel gesprekken onderling als met de docent.

De klassengesprekken en de gesprekken van de duo's over het onderzoek zijn opgenomen op geluidsband en volledig uitgeschreven. De transcripten zijn in episodes verdeeld waarbij de grenzen van episodes worden gevormd door verandering van onderwerp of gesprekspartners. De episodes zijn onderverdeeld in uitspraken waarbij uitspraken begrensd worden door verandering van spreker. Een uitspraak is dus hetgeen één persoon achter elkaar uitspreekt tot een ander het woord neemt. Uiteraard is er ook regelmatig meer dan één persoon tegelijk aan het woord. Gepoogd is dat zo getrouw mogelijk op te schrijven, ook al kent tekst op papier maar één dimensie. Als iemand onderbroken wordt door een ander, maar daarna zijn betoog voortzet, is dat in de analyses beschouwd als één uitspraak.

- *Categorieën van uitspraken*

Uiteindelijk gaat het erom dat de leerlingen criteria voor goed onderzoek kunnen gebruiken in meer dan één context. Met andere woorden dat ze de gegeneraliseerde criteria kunnen hanteren. En dan gaat het niet alleen om het kennen en kunnen toepassen van de algemene regel, maar ook om het begrijpen van de achterliggende gedachte. En uiteraard gaat het om het kunnen herkennen en benoemen van concrete verschijningsvormen ervan. In een voorbeeld: leerlingen kennen niet alleen de regel dat alle andere variabelen dan de onafhankelijke en de afhankelijke onder controle moeten worden gehouden, ze begrijpen ook dat als aan die regel niet voldaan is er geen sprake is van een eerlijk onderzoek én ze kunnen onderkennen dat bij onderzoek naar de invloed van de hoeveelheid licht op de groei van kiemplantjes de temperatuur constant gehouden moet worden.

Deze drie niveaus – algemene regel, achterliggende gedachte, concreet voorbeeld – komen overeen met een driedeling die vaak gebruikt wordt met betrekking tot het praten over natuur en natuurwetenschap, zij het in omgekeerde volgorde: beschrijving, verklaring en generalisatie (o.a. Mortimer & Scott, 2000, 2003). Bovendien ligt daarin ook een parallel met het model van Galperin (fig. 4.2). Op het materiële niveau bevinden zich concrete objecten en gebeurtenissen (beschrijving), op het verbale niveau worden die onder woorden gebracht (explicitering) en op het mentale niveau gaat het om innerlijke weergave ervan (generalisatie).

Toegepast op de criteria voor goed onderzoek wordt dat dan de volgende driedeling.

- *Beschrijving* betreft uitspraken over specifieke voorwerpen, systemen of fenomenen in termen van hun samenstelling, verplaatsingen en dergelijke, bijvoorbeeld: 'Deze onderzoeksvraag is te vaag.'
- *Explicitering*[23] betreft uitspraken die een of ander theoretisch model of mechanisme inbrengen om een specifiek fenomeen duidelijk te maken, bijvoorbeeld: 'De vraag is voor meer dan een uitleg vatbaar, want het begrip 'hoeveel' kan op verschillende manieren worden opgevat.'
- *Generalisatie* betreft beschrijvende of expliciterende uitspraken die niet meer gekoppeld zijn aan één specifieke context, maar een algemene regel of wetmatigheid aangeven, bijvoorbeeld: 'Onderzoeksvragen mogen alleen eenduidige begrippen bevatten.'

Deze manier van categoriseren van uitspraken van leerlingen en docenten maakt het mogelijk om de ontwikkeling van begrip van bewijs in beeld te brengen. Ten eerste door te kijken of de criteria alleen op een concreet, beschrijvend niveau aan de orde komen of dat ze ook expliciet gemaakt en gegeneraliseerd worden. Met andere woorden óf er wel sprake is van een ontwikkeling.

[23] Mortimer en Scott (2000, 2003) gebruiken *explanation*, dat vertaald kan worden met o.a. 'verklaring', en 'uitleg'. De term *verklaring* geeft een oorzakelijk verband tussen gebeurtenissen aan: het geeft aan *waardoor* een verschijnsel plaatsvindt; bij het gebruik van de criteria voor goed onderzoek lijkt de term uitleg of *explicitering* beter, omdat het er eerder om gaat aan te geven *waarom* iets wel of niet zou moeten gebeuren.

Ten tweede kunnen verschillende volgordes waarin beschreven, geëxpliciteerd en/of gegeneraliseerd wordt iets zeggen over de manier van ontwikkelen. Bijvoorbeeld of het van beschrijven via expliciteren naar generaliseren gaat, de inductieve weg, of juist het tegenovergestelde, de deductieve weg.

In een gesprek over onderzoek kunnen leerlingen 'spontaan' verwoorden hoe of waarom ze voor een bepaalde aanpak kiezen, met andere woorden ze kunnen beschrijven dat of expliciteren waarom ze bepaalde intervallen voor hun onafhankelijke variabele kiezen. Maar het kan ook als een probleem opgeworpen worden, dus in een vragende vorm verwoord worden. Die problematisering kan door de docent gebeuren, maar ook door een leerling, in de richting van zijn of haar onderzoekspartner of in de richting van de docent of technisch onderwijsassistent. Naast de categorieën beschrijving, explicitering of generalisatie is het daarom ook zinvol te registreren of er sprake is van het problematiseren van het onderzoek met betrekking tot een bepaald kwaliteitsaspect, en op wiens initiatief er geproblematiseerd wordt.

Met betrekking tot de vraag of de criteria door de leerlingen ook buiten het eigen onderzoek kunnen worden toegepast, of ze kunnen recontextualiseren, is het zinvol om te registreren of dat plaatsvindt. Met andere woorden of de criteria in een andere context – een ander onderzoek – worden toegepast. Dat is dus nog net een stap verder dan het formuleren van een criterium als algemene regel.

Samengevat kunnen de volgende categorieën van uitspraken onderscheiden worden.

- *Problematiseren*: de docent of een leerling stelt een vraag of signaleert een probleem waardoor er een probleem ontstaat waar een antwoord / oplossing voor gezocht moet worden.
- *Beschrijven*: de docent of een leerling omschrijft het antwoord of de oplossing op een specifiek probleem: hoe zit dit in elkaar of hoe los je dit op?
- *Expliciteren:* de docent of een leerling legt uit wat de achterliggende gedachte achter het antwoord of de oplossing is. Dit kan eventueel incorrect zijn.
- *Generaliseren*: de docent of een leerling formuleert een algemene regel, een algemeen geldend criterium.
- *Toepassen*: een specifiek antwoord of een algemeen (gegeneraliseerd) antwoord wordt toegepast in een nieuwe context.

• *Patronen, volledigheid en volgorde*

Zoals al opgemerkt is het voor het volgen van de ontwikkeling van begrip van bewijs ook van belang te kijken naar de volgordes waarin de verschillende uitspraken voorkomen. Worden de beschrijvingen, expliciteringen en generalisaties los van elkaar gedaan of in samenhang? Met andere woorden: zijn er *patronen* te onderkennen in de gesprekken en hoe *volledig* zijn die dan? Een driedeling lijkt zinvol.

- een *volledig patroon* doorloopt – in welke volgorde dan ook – zowel het materiële als het verbale als het mentale niveau en bevat dus in ieder geval de drie categorieën beschrijven, expliciteren en generaliseren, al dan niet in combinatie met problematiseren en/of toepassen;
- een *onvolledig patroon* mist één van die drie en bestaat dus uit een combinatie van beschrijven en expliciteren, expliciteren en generaliseren of beschrijven en generaliseren;
- in een *rudimentair patroon* ontbreken twee van de drie en is eigenlijk nauwelijks een patroon te noemen.

Behalve de aanwezigheid van beschrijven, expliciteren en generaliseren geeft ook de volgorde waarin deze voorkomen iets aan van de ontwikkeling van een criterium. Begint die op het concrete, beschrijvende niveau of met het poneren van de algemene regel? Uitgaande van de veronderstelde cyclus *oriëntatie – materieel – verbaal – mentaal – nieuwe oriëntatie* (fig. 4.2), zou een patroon *problematiseren – beschrijven – expliciteren – generaliseren – toepassen* verwacht kunnen worden. Maar vindt het ook plaats?

- *Voorbeeld van scoren van uitspraken en patronen*

Aan de hand van een fragment van het klassengesprek in een van de startlessen kan de indeling toegelicht worden (fig. 6.2). Er wordt gesproken over de vraag of en hoe je de onderzoeksvraag 'zijn blonde meisjes dommer dan andere meisjes?' zou kunnen onderzoeken. Het fragment begint met een vraag van de docent naar nog meer beïnvloedende factoren – die je dus constant of onder controle moet houden, al is dat criterium (element van begrip van bewijs 9, zie p.250) nog niet expliciet genoemd – behalve de leeftijd van de meisjes, die daarvoor net aan de orde geweest is. Op deze problematisering van de docent volgt de beschrijving van enkele concrete factoren, genoemd door leerlingen (uitspraken 48, 49, 56) en bevestigd door de docent (50, 57). Leerling 3 problematiseert dat op zijn beurt weer (51). Dan komt er opeens een andere kwestie tussendoor. De vraag van leerling 5 (54) wordt eerst verkeerd begrepen door de docent (55), maar nadat hij het probleem opnieuw formuleert (58), snapt de docent waar het over gaat, namelijk of IQ leeftijdafhankelijk is. Dat grijpt terug op de discussie die voorafgaand aan dit fragment heeft plaatsgevonden. De docent verwijst dan ook naar iets wat eerder (door Martin[24]) gezegd is (59); als leerling 5 aanhoudt, expliciteert de docent de achterliggende gedachte (62, eerste deel). Dat dit eigenlijk niet correct is (IQ kan wel veranderen), doet voor het scoren niet terzake. De docent pakt daarna de eerdere lijn weer op (62, tweede deel) en breidt dat uit naar de steekproefsamenstelling (element 14, uitspraak 64). Leerling 6 komt vervolgens met een suggestie om de invloed van andere factoren onder controle te houden: neem tweelingen waarvan de een blond is en de ander niet (66, ze praat nogal zacht en wat ze zegt is op de opname niet precies te horen; uit het vervolg kan echter afgeleid worden wat ze waarschijnlijk gezegd heeft). Als leerling 5 deze suggestie problematiseert (67) weerspreekt de docent dit (68) en legt leerling 6 uit wat ze bedoelt. De docent pakt de suggestie op, maar problematiseert of je aan voldoende van die tweelingen zou kunnen komen (71 en 73). Leerling 6 zet haar idee nogmaals uitgebreider uiteen (72), waarna de docent het generaliseert (74).

In het fragment komen zowel een volledig, een onvolledig als een rudimentair patroon voor. De hoofdlijn van het gesprek gaat over de beïnvloedende factoren en kent de volgorde PBBBPBPBPBPBPBEEG met betrekking tot element 9, hetgeen verkort kan worden geschreven als $[PB]_6E_2G9$ (zes keer PB, twee keer E en vervolgens een G; het betreft element

(op de volgende bladzij)

Figuur 6.2 Deel van het transcript van episode 4 van een startles met toegekende scores voor categorie/element van begrip van bewijs (cat/nr bvb) en patronen. Om de scores die de patronen vormen goed te kunnen onderscheiden zijn ze afwisselend met normale, cursieve en vette tekens weergegeven. Ter illustratie is van twee patronen aangegeven uit welke uitspraken ze opgebouwd zijn. P9 betekent problematiseren m.b.t. element 9, B9 betekent beschrijven m.b.t. element 9, enz.; $[PB]_6$ betekent zes keer de volgorde PB.

[24] De namen van de leerlingen zijn veranderd, met handhaving van het geslacht.

epi	uit	spreker	tekst	cat/ nr bvb	patroon
04	47	Docent	Zijn er misschien nog andere dingen van invloed, die … behalve leeftijd?	P9	
	48	Leerling 1	Ja, bijvoorbeeld hangt ervan af of je allemaal mensen op het VWO gaat bekijken, of …	B9	
	49	Leerling 2	Criminele achtergrond.	B9	
	50	Docent	Ah, opleiding!	B9	
	51	Leerling 3	Maar opleiding zou eigenlijk ook niet van invloed moeten zijn	P9	
	52	Leerling 4	[onverstaanbaar]		
	53	Docent	Ik zag daar een vinger …		
	54	Leerling 5	Ja. Hoe lang is, is je IQ …?	*P9*	
	55	Docent	Hoe lang ze zijn die meisjes? Nee.	B9	
	56	Leerling 4	Je hersenomvang?	P9	
	57	Docent	Dat soort dingen. Maar jij wou wat vragen?	B9	
	58	Leerling 5	Maar is een IQ-test, als die eenmaal is vastgesteld, is die voor de rest van je leven, zeg maar …?	*P9*	
	59	Docent	Martin zegt van wel.	*B9*	
	60	Leerling 2	Ja het is niet …		
	61	Leerling 5	Ik bedoel … blijft het altijd hetzelfde, verandert het niet in de loop van de tijd, in de loop van de jaren?	*P9*	
	62	Docent	Een IQ-test test eigenlijk waar je toe in staat bent hebt, en niet hoeveel je bijvoorbeeld weet of zo, zoiets.	*E9*	*PBPE9*
			Ehm, nou dan zitten we wel een beetje met dat probleem, hè? Dan ehm, welke opleiding moeten die meisjes hebben of maakt dat niks uit?	P9	
	63	Leerling 2	Nee, want dat merk je toch dat, nee. Nee.		
	64	Docent	Ik weet niet, zeg het maar. Hoe …		
	65	Leerling 6	Nee.		
	(64)	Docent	… hoe ga jij je die groepen samenstellen, hoe zien ze d'r uit? Even Veronica.	**P14**	
	66	Leerling 6 (Veronica)	Het beste is als [onverstaanbaar] [waarschijnlijk zegt Veronica dat het het beste is om tweelingen te nemen waarvan de een blond is en de ander niet]	B9	
	67	Leerling 5	Dat kan dus niet.	P9	
	68	Docent	Nou, het kan wel. Maar, dat komt dan juist zelden voor inderdaad.	B9	
	69	Leerling 2	Muteren!		
	70	Leerling 6	Dan een tweeling, maar dan twee-eiig. Dat zijn dezelfde [onverstaanbaar]	E9	
	71	Docent	Goed, maar jij zegt dus. Dus, zeg jij, het best zou zijn dat je eh, want een groep van één is een beetje weinig. Als jij eeneiige tweelingen hebt, splits je in tweeën, en daar doe je onderzoek mee. Dat kan, maar het zou natuurlijk prettiger zijn als je d'r dan eh, honderd had.	**B14**	
	72	Leerling 6	Maar twee …		
	73	Docent	Maar, hoe zien die groepen er dan uit? Helaas, een eeneiige tweeling wordt volgens mij lastig.	**P14**	**PBP14**
	(72)	Leerling 6	… twee-eiige tweeling hebben toch ook gewoon dezelfde achtergrond, zelfde … opvoeding	E9	
	74	Docent	Dus jij zegt eh, goed de leden van beide groepen, de, blonde en de andere, moeten eh, overeen komen, in alles. leeftijd, achtergrond.	G9	$[PB]_6E_2G9$

9). Omdat hierin zowel beschrijving als explicitering als generalisatie zit, is dit een *volledig* patroon. Daarnaast zijn de uitspraken 54 + 58 t/m 62 samen te vatten als PBPE9. Het betreft een zijpad van het gesprek en wordt daarom als een apart patroon beschouwd. Het bevat slechts twee van de drie B, E en G en is dus een *onvolledig* patroon. Tenslotte is er het patroon PBP14, dat naast problematisering alleen een beschrijving bevat en daarom *rudimentair* genoemd wordt.

Hier blijkt dus ook meteen dat patronen vervlochten kunnen zijn: de kortere patronen PBPE9 en PB14 slingeren zich door het lange patroon $[PB]_6E_2G9$. Tot slot valt bij dit fragment nog te noteren wat de bijdragen van docent en leerlingen zijn aan de discussie. De problematiseringen komen vier keer van de docent en vijf keer van een leerling, de expliciteringen zijn ook vrijwel gelijk verdeeld (één keer docent, twee keer leerling) en de generalisatie gebeurt door de docent.

- *Eenheid van analyse en score*

De eenheid van analyse is de uitspraak. Elke uitspraak van de docent of de leerling is bekeken op twee aspecten: of er een verwijzing naar een van de elementen van begrip van bewijs voorkomt en zo ja, in welke van de categorieën van analyse de uitspraak valt. Het scoren van uitspraken leverde dus aantallen uitspraken op per element van begrip van bewijs in de categorieën P, B, E, G en T.

Daarnaast is op patroonniveau gescoord (tabel 6.1). Om goed het optreden en de volgorde van beschrijven, expliciteren en generaliseren in volledige, onvolledige en rudimentaire patronen in beeld te brengen (zie p. 92/93) is voor het scoren van de patronen niet gelet op het voorkomen van *problematiseren*, wel op het voorkomen van *toepassen*. Maar het al dan niet voorkomen van *toepassen* had geen invloed op de indeling in volledig – onvolledig – rudimentair, want die is uitsluitend gebaseerd op het voorkomen van respectievelijk drie, twee of één van de drie categorieën *beschrijven, expliciteren* en *generaliseren.* Herhaalde delen van patronen zijn als één deel beschouwd. De patronen uit figuur 6.2 – PBPE9, PB14 en $[PB]_6E_2G9$ – zijn dus respectievelijk gescoord als BE, B en BEG.

Zowel bij de volledige als de onvolledige patronen is er een categorie 'anders': *V anders* en *O anders*. Hierin vallen de weinig verwachte patronen zoals BGE, BEB of BEBG. *R anders* komt niet voor, want dat kan theoretisch alleen B, E of G in combinatie met toepassen betreffen. Bij een beschrijving valt er echter geen regel of idee toe te passen en als een explicitering of generalisatie concreet gemaakt wordt kan er alleen sprake zijn van toepassing als het om een ánder onderzoek gaat dan eerder in hetzelfde patroon (in een beschrijving) concreet aan de orde geweest is. Een explicitering of een generalisatie met een concretisering zijn dus als EB of GB gescoord.

Tabel 6.1 Scorecategorieën van volledige, onvolledige en rudimentaire patronen. B: beschrijven, E: expliciteren, G: generaliseren, T: toepassen.

Volledig	*Onvolledig*		*Rudimentair*
BEGT	BET	GET	B
BEG	BE	GE	E
GEBT	BGT	GBT	G
GEB	BG	GB	
V anders	EGT	EBT	
	EG	EB	
		O anders	

- *Interpretatie*

Het scoren van uitspraken en patronen maakt het mogelijk om de ontwikkeling van begrip van bewijs te volgen in de gesprekken tussen leerlingen en docent. Komen de criteria aan de orde en zo ja, op welk niveau? Blijft het bij het concrete beschrijven of wordt er ook geëxpliciteerd en gegeneraliseerd? En als er gegeneraliseerd wordt, via welke wegen komt die generalisatie dan tot stand? Gaat dat via volledige patronen en zo ja, zijn die vooral inductief of juist niet? Waarbij het ook nog interessant is om te zien in hoeverre dit verschilt in de beide onderwijsleerstrategieën. In de volgende paragraaf wordt verder uitgewerkt welke verwachtingen geformuleerd kunnen worden ten aanzien van de ontwikkeling van begrip van bewijs in die verschillende strategieën.

6.4 Verwachte samenhang tussen onderwijsleerstrategie en patronen in de ontwikkeling van begrip van bewijs

Door de verschillen in aanpak tussen de impliciete en expliciete onderwijsleerstrategie (binnen het kader van de probleemstellende benadering) zijn er andere dominante patronen in de beide strategieën te verwachten.

Allereerst lijkt het logisch te veronderstellen dat het werken aan een eigen onderzoek vooral tot een inductieve wijze van ontwikkelen zal leiden. Er wordt immers vanuit een concreet probleem gewerkt en via de reflectie daarop verwachten we dat de leerlingen tot veralgemenisering komen (zie fig. 6.1). Dat betekent dat er tijdens het werken aan het eigen onderzoek – in de impliciete strategie gebeurt niets anders – vooral BE, BG, EG of BEG patronen zullen optreden en minder EB, GB, GE of GEB.

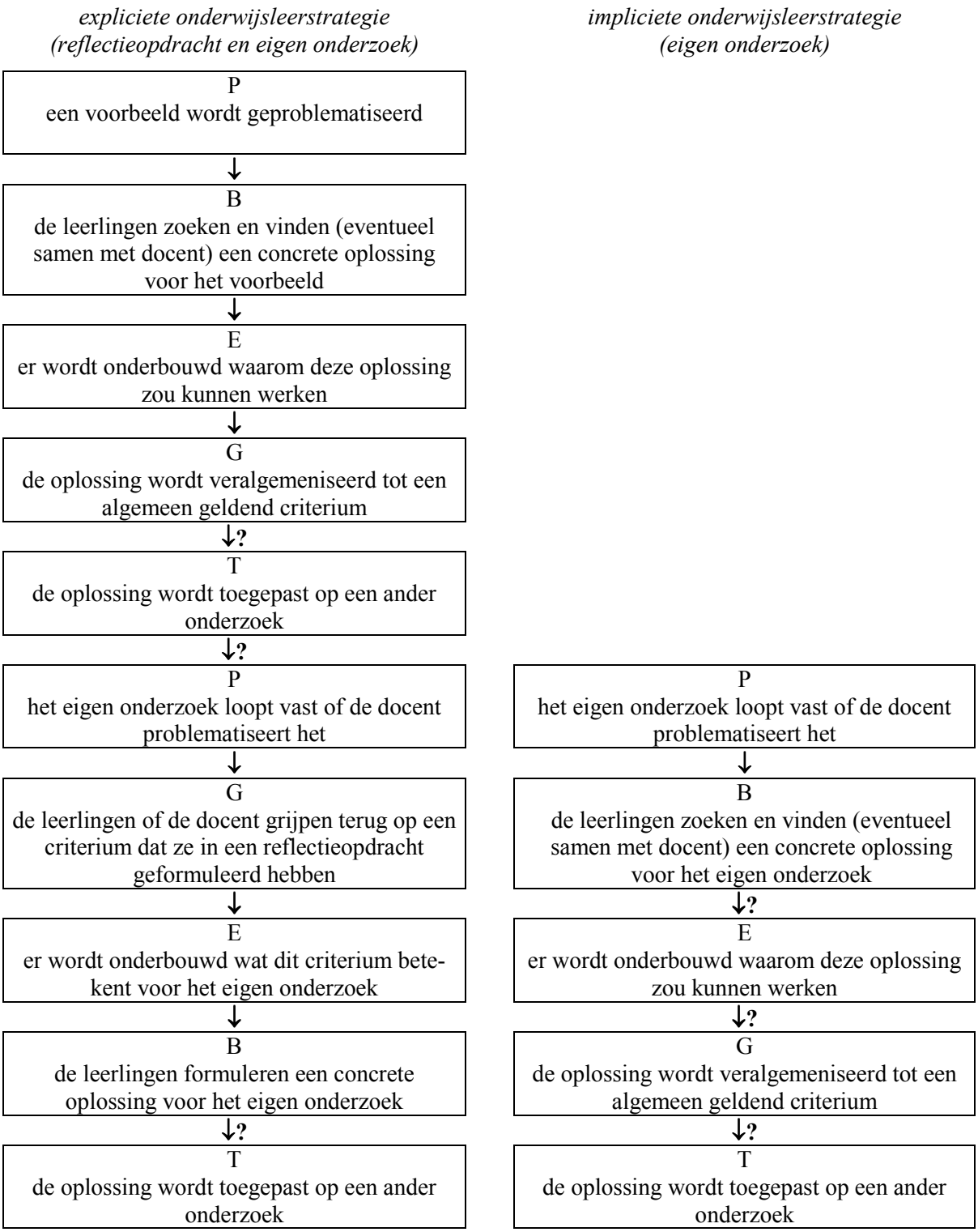
In de expliciete aanpak worden echter ook los van het eigen onderzoek een of meer criteria aan de orde gesteld in de reflectieopdrachten door middel van een concreet voorbeeld. Er wordt daarbij expliciet gevraagd de gevonden criteria op het eigen onderzoek toe te passen. Dat betekent dat in de expliciete strategie er mogelijk vaker een deductief patroon zal optreden, van algemeen naar specifiek. Dus de genoemde EB, GB, GE of GEB en ook vaker een T.

Dat leidt tot de eerste, tweeledige verwachting:

(1a) *in de impliciete strategie komen vooral inductieve patronen voor, in de expliciete onderwijsleerstrategie vaker deductieve patronen;*

(1b) *in de expliciete onderwijsleerstrategie wordt vaker een veralgemeniseerd criterium in een nieuwe context toegepast.*

In figuur 6.3 is dat schematisch weergegeven. In de expliciete strategie worden leerlingen ‘gedwongen’ een criterium te veralgemeniseren en het toe te passen op hun eigen onderzoek. Dat werkt echter alleen als het voorbeeld en het eigen onderzoek voldoende op elkaar lijken en voldoende met elkaar in de pas lopen. In de impliciete strategie is het gevaar van ‘uit de pas lopen’ er niet, maar worden explicitering en generalisatie niet afgedwongen, behalve als de docent er expliciet naar vraagt. Generaliseren is daar moeilijker omdat er niet een tweede onderzoek voorhanden is om het criterium in een andere context te gebruiken. Interessant is dan de vraag of in de expliciete strategie de leerlingen in hun eigen onderzoek teruggrijpen op de in de reflectieopdrachten gegeven voorbeelden en de daar ontwikkelde criteria. De vraag is echter of er überhaupt wel veralgemenisering optreedt, of de gesprekken wel het concrete niveau ontstijgen. Met andere woorden: leidt het doen van een onderzoek wel tot reflectie op de kwaliteit ervan? Zeker bij het eigen onderzoek, als de leerlingen een concrete oplossing hebben gevonden voor hun probleem, is het maar de vraag of ze



Figuur 6.3 Mogelijke ontwikkeling van begrip van bewijs in de expliciete onderwijsleerstrategie (links) en de impliciete onderwijsleerstrategie (rechts). Een vraagteken betekent dat het de vraag is of die stap plaatsvindt.

dat dan nog onderbouwen en veralgemeniseren. Dat betekent dat er in de impliciete strategie wellicht vooral onvolledige of rudimentaire patronen optreden, die dan ook nog eens op het beschrijvende niveau blijven. In de expliciete strategie worden leerlingen in de reflectieopdrachten gedwongen om criteria te formuleren en die toe te passen op hun eigen onderzoek. Het is de verwachting dat ze in het praten over hun eigen onderzoek – los van de reflectieopdrachten – dan ook vaker criteria of de gedachten erachter gebruiken. Als dat in de impliciete strategie gebeurt, zal dat vaker op initiatief van de docent zijn.

De tweede, eveneens tweeledige verwachting luidt daarom:

(2a) *in de expliciete strategie komen meer volledige patronen voor en minder rudimentaire dan in de impliciete strategie;*

(2b) *in de impliciete strategie zijn relatief meer uitspraken op het beschrijvende niveau.*

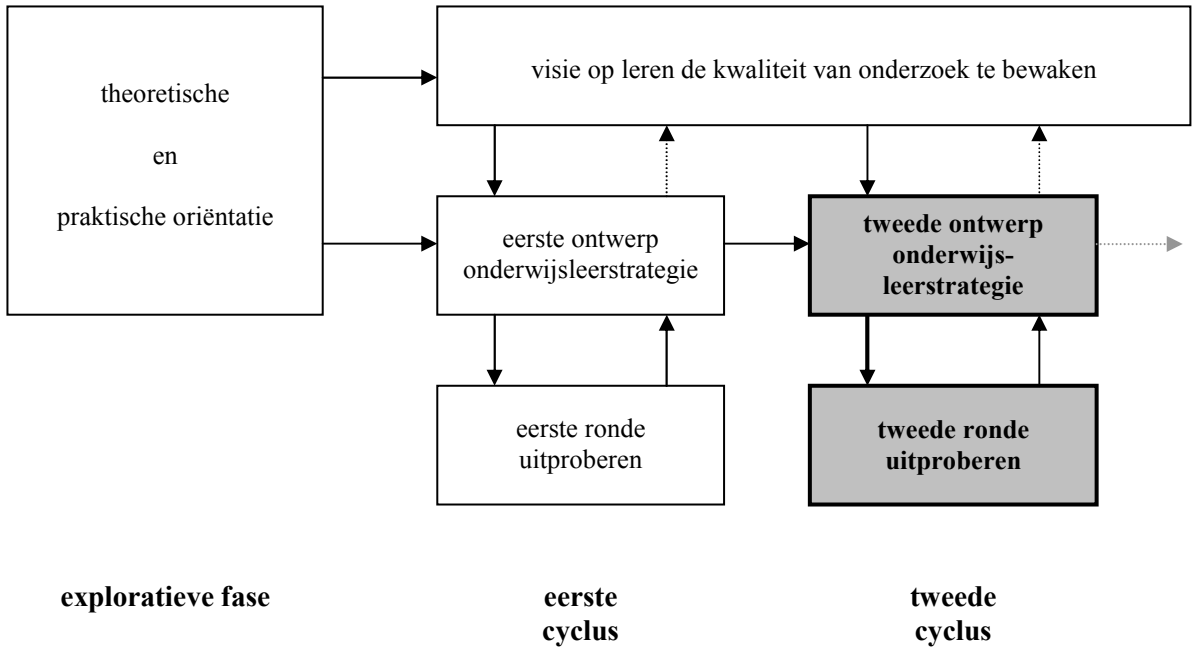
Dat de docent een initiërende rol is toebedacht, zou te zien kunnen zijn in het aandeel van de docent in categorieën en patronen. Niet alleen zou dat te zien zijn in meer problematiseringen van zijn kant, maar wellicht ook in het aantal expliciteringen en generaliseringen die in zijn of haar bijzijn gedaan worden. De docent wordt geacht door te vragen en criteria boven tafel te krijgen, al heeft het de voorkeur ze door de leerlingen te laten verwoorden. Doordat de leerlingen in de expliciete groepen in de reflectieopdrachten zeker met de criteria in aanraking komen, explicitern of generaliseren zij wellicht vaker zelf bij hun eigen onderzoek.

Dat leidt tot een derde, drieledige verwachting:

(3a) *de problematiseringen komen vaker van de docent dan van leerlingen;*

(3b) *in de gesprekken tussen docent en leerlingen komen vaker expliciteringen en generaliseringen voor – en dientengevolge meer volledige en onvolledige patronen – dan in gesprekken tussen leerlingen onderling;*

(3c) *deze verschillen zijn groter in de impliciete strategie dan in de expliciete.*



7.1 Het tweede ontwerp
7.2 Reflectieopdrachten
7.3 De nieuwe populatie en instrumenten
7.4 Uitvoerbaarheid
7.4.1 Gegaan zoals gepland?
7.4.2 De startlessen
7.4.3 Motief als drijvende kracht
7.4.4 Het beoogde docentgedrag
7.4.5 Een éigen onderzoek
7.4.6 Het boekje
7.4.7 Een artikel als eindproduct
7.4.8 De reflectieopdrachten
7.4.9 De impliciete onderwijsleerstrategie
7.4.10 Het wetenschappelijke verhaal
7.4.11 Samenvattend: in welke mate is de praktische opdracht uitvoerbaar?
7.5 Effectiviteit
7.5.1 De startlessen
7.5.2 De weg naar de eindproducten
7.5.3 De eindproducten
7.5.4 Begin- en eindtoets
7.5.5 Ervaren leereffect
7.5.6 Samenvattend: in welke mate is de praktische opdracht effectief?
7.6 Kan het beter?

7 De tweede cyclus

'Each part constitutes a different way of looking at this lived experience of learning in the science laboratories. (...) It is this polyphony which constitutes my understanding.'
Wolff-Michael Roth (1995, p. 265)

Hoofdstuk 7 De tweede cyclus

Dit hoofdstuk beschrijft de opzet, uitvoeringen en bevindingen van het tweede ontwerp van de praktische opdracht. Eerst worden de bijgestelde uitgangspunten op een rij gezet en de belangrijkste veranderingen ten opzichte van de eerste ronde verwoord (7.1). Dan volgt de beschrijving van de nieuwe reflectieopdrachten (7.2) en de onderzoekspopulatie en het bijgestelde onderzoeksinstrumentarium (7.3).

In paragraaf 7.4 en 7.5 worden de bevindingen beschreven die moeten leiden tot de beantwoording van respectievelijk de volgende vragen:

- *in welke mate is de praktische opdracht uitvoerbaar voor docent en leerlingen, met speciale aandacht voor de reflectieopdrachten en de ontwikkeling van begrip van bewijs?*
- *in welke mate is er leereffect te constateren met betrekking tot begrip van bewijs, met speciale aandacht voor de manier waarop dat zich ontwikkelt?*

Aan het slot van dit hoofdstuk (7.6) wordt onder woorden gebracht welke aspecten van de praktische opdracht nog verbetering behoeven.

7.1 Het tweede ontwerp

Na ervaringen uit de eerste ronde, zoals verwoord in paragraaf 5.5 en de aanscherping van het theoretische kader in het vorige hoofdstuk, zijn de uitgangspunten voor de tweede versie de volgende, de veranderingen ten opzichte van het eerste ontwerp zijn gecursiveerd:

- de lessenserie omvat *tien* contacturen *plus twee* uren voor de voor- en natoets; er zijn *twee aaneengesloten* contacturen per week (m.u.v. de eerste en de laatste les); de leerlingen hebben die lesuren en de tijd ertussen om aan hun onderzoek te werken;
- *de opdracht wordt gestart met een gezamenlijke oriëntatie die gericht is op het vinden en expliciteren van een gezamenlijk startkapitaal aan kennis, gebaseerd op hetgeen de leerlingen al weten over onderzoek doen;*
- de leerlingen werken samen in groepjes van twee of drie[25] aan een eigen hypothesetoetsend onderzoek binnen het vak biologie; ze bedenken een eigen onderwerp, formuleren hun eigen onderzoeksvraag, maken een onderzoeksopzet, verzamelen en verwerken gegevens, beantwoorden de onderzoeksvraag en evalueren hun eigen onderzoek; dit is geen lineair proces, maar er wordt heen en weer gegaan tussen de verschillende delen van het onderzoek;
- *de leerlingen ontvangen een boekje waarin de opdracht nader omschreven wordt; daarin houden ze ook de voortgang van hun onderzoek bij;*
- het product telt mee als praktische opdracht; *de leerlingen rapporteren over hun onderzoek in de vorm van een wetenschappelijk artikel; zij krijgen daarvoor richtlijnen en een voorbeeld;*

[25] Omdat de meeste leerlingen in groepjes van twee werken, worden die vaak aangeduid als duo's, maar zo'n 'duo' kan dus ook uit drie leerlingen bestaan.

- de docent heeft een ondersteunende rol, hij/zij is vraagbaak voor de leerlingen wat betreft de opzet en uitvoering van het onderzoek; hij heeft echter ook een initiërende rol waar het gaat om het stimuleren van de reflectie op de kwaliteit van het onderzoek;
- in de wisselwerking tussen onderzoek doen en reflecteren op onderzoek wordt het *'wetenschappelijke verhaal'*[26] ontwikkeld van de kwaliteitscriteria voor onderzoek en hebben de leerlingen gelegenheid het toe te passen *en uit te breiden*.

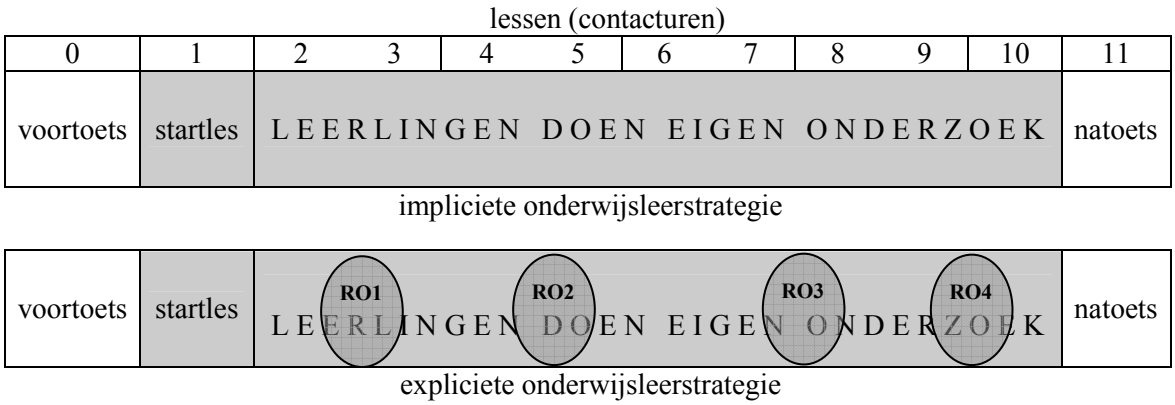
Net als in de eerste cyclus zijn twee varianten van deze opzet ontwikkeld, een expliciete en een impliciete. In de expliciete variant wordt de reflectie op de kwaliteit van onderzoek expliciet aan de orde gesteld in verplichte *reflectieopdrachten* die uitgevoerd worden als het eigen onderzoek een bepaald stadium heedt bereikt. Het eigen onderzoek wordt daar dan even voor stilgelegd en de opdrachten worden nabesproken met de docent.

In de impliciete variant vindt die reflectie uitsluitend plaats als het eigen onderzoek van de leerlingen daartoe aanleiding geeft. Wel zijn twee *momenten van geschreven feedback* ingebouwd: de leerlingen leveren aan het eind van les 3 hun onderzoeksvraag, hypothese en onderzoeksopzet in en op een later tijdstip het concept van hun artikel; de docent geeft hier schriftelijk commentaar op.

Op beide scholen, dus door beide docenten, is zowel de expliciete als de impliciete variant uitgevoerd.

Figuur 7.1 geeft de opzet van beide varianten in schema weer. De totale contacttijd is in beide varianten hetzelfde. Voor beide varianten is een scenario geschreven waarin voor elke les en de tussenliggende periodes is aangegeven welke activiteiten leerlingen en docent verwacht worden te ondernemen en wat de verwachte (leer)opbrengst is. Deze scenario's zijn uitgebreid met de docenten besproken en op grond van deze bespreking aangepast; de uiteindelijke scenario's zijn te vinden in bijlage 5.

In het eerste en het laatste uur van de lessenserie maken de leerlingen een toets over de kwaliteit van onderzoek. Deze aangepaste versie van de toets uit de eerste cyclus wordt in paragraaf 7.3 besproken.



Figuur 7.1 Schematische weergave van het tweede ontwerp van de lessenserie. In de impliciete onderwijsleerstrategie (boven) doen de leerlingen hun eigen onderzoek en vindt reflectie plaats als dat onderzoek er aanleiding toe geeft. In de expliciete onderwijsleerstrategie (onder) wordt het eigen onderzoek van de leerlingen onderbroken door vier reflectieopdrachten (RO1 t/m RO4) die voor verschillende duo's op verschillende momenten kunnen plaatsvinden.

[26] Deze term is afkomstig van Mortimer en Scott (2003). Zij bedoelen er de 'uitleg' van een wetenschappelijk verschijnsel mee, zoals die stap voor stap gestalte krijgt in een onderwijsleergesprek of andere leeractiviteiten. Wat er in deze context mee bedoeld wordt, wordt uitgelegd op pagina 107.

De aanpassingen ten opzichte van het eerste ontwerp – wat er in de tweede ronde beter moet gaan, zie paragraaf 5.5 – worden hieronder verder uitgewerkt, de reflectieopdrachten komen in paragraaf 7.2 aan de orde. Of en hoe het gewerkt heeft, wordt in de paragrafen 7.4 en 7.5 beschreven.

- *Startles*

Zoals betoogd in paragraaf 6.2 begint de tweede cyclus met een gezamenlijke oriëntatie. Deze 'startles' is voor alle leerlingen hetzelfde, omdat de doelstelling van het project voor alle leerlingen hetzelfde is: leren wat goed onderzoek doen is en welke criteria daarbij een rol spelen. De doelen van de startles zijn:

- een globaal motief ontwikkelen, de leerlingen richten op het doel van de praktische opdracht: leren wat goed onderzoek is door als centrale vraag 'Zeker weten?' op te werpen;
- het vinden en expliciteren van een *common ground*, een gezamenlijk startkapitaal aan kennis, gebaseerd op hetgeen de leerlingen al weten over onderzoek doen;
- de ontwikkeling van het wetenschappelijke verhaal een aanvang laten nemen, d.w.z. liefst gezamenlijk tot de formulering van criteria voor goed onderzoek komen (de elementen van begrip van bewijs 1, 3, 4, 5, 8, 9, 10, 14, 18 en 21);
- de leerlingen inlichten over enkele praktische randvoorwaarden.

Met de docenten is de opzet van de startles uitvoerig doorgesproken, aan de hand van een door de onderzoeker geschreven scenario, een gedetailleerd verloop van het klassengesprek (figuur 7.2, zie bijlage 6 voor het volledige scenario).[27] De docenten waren zich ervan bewust dat het scenario slechts één mogelijk verloop van de les schetste, maar naar hun inschatting bood het voldoende zicht op het beoogde verloop en de beoogde uitkomst. Het scenario is ook gebruikt om het daadwerkelijke verloop van de (vier keer uitgevoerde) les te vergelijken met het van te voren bedachte.

Aan de hand van een prikkelende onderzoeksvraag, 'zijn blonde meisjes dommer dan andere meisjes?', worden de leerlingen uitgedaagd om aan te geven hoe je dit *goed* zou kunnen onderzoeken, hoe je het antwoord op die vraag zeker zou kunnen weten.

Ter ondersteuning van de docent tijdens de uitvoering is het beoogde lesverloop in steekwoorden op kaartjes (A6-formaat) gezet die ze bij de hand kunnen houden. Ook zijn de criteria op twee overheadsheets en in de boekjes voor de leerlingen gezet, zodat de docent ze niet op het bord hoeft te schrijven en de leerlingen ze niet hoeven over te nemen, hetgeen tijdwinst betekende. Aan het eind van de startles wordt de opzet en de planning van de hele praktische opdracht door de docent kort uiteengezet aan de hand van een boekje dat aan iedere leerling individueel wordt uitgereikt.

- *Boekje*

De leerlingen wordt meer structuur dan in de eerste cyclus geboden in de vorm van een boekje (bijlage 7) met daarin de volgende informatie:

(op de volgende bladzij)

Figuur 7.2 Eerste deel van het scenario van de startles. D: docent; LL: leerling.

[27] Het scenario voor de startles heeft een iets ander karakter en een iets andere functie dan de scenario's voor de beide onderwijsleerstrategieën (zie noot 14 op p. 68). Dit scenario heeft de vorm van een volledig uitgeschreven klassengesprek, als één mogelijk en gewenst verloop van het gesprek. Het heeft vooral de functie van voorbeeld voor de docenten van een manier om vanuit leerlingreacties naar de criteria voor goed onderzoek te komen.

D: Vandaag gaan we beginnen met de praktische opdracht 'onderzoek doen'.
Zoals ik al heb verteld, gaat het er niet alleen om een leuk en goed onderzoek te doen, maar ook, vooral, om te leren – en te laten zien dat je het geleerd hebt – wat goed onderzoek doen is.
Voordat jullie de boekjes die ik uitgedeeld heb gaan doorlezen, wil ik met jullie van gedachten wisselen over wat dat is: goed onderzoek.
Jullie hebben al een beetje ervaring met het doen van onderzoek, bij biologie en bij andere vakken, bijvoorbeeld het onderzoek dat jullie voor de kerst hebben gedaan met catalase.
Jullie weten er dus al het een en ander van.
Ik doe een voorzet, dat staat ook al op de kaft van het boekje: onderzoek doe je als je iets wilt weten, goed onderzoek is nodig als je iets ZEKER wilt weten.
Als ik een vraag heb, bijvoorbeeld (*schrijft op bord*): 'zijn blonde meisjes dommer dan andere meisjes?' hoe moet ik het dan aanpakken om er zeker van te zijn dat ik een goed antwoord krijg?
LL1: Nou, alle meisjes een IQ-test laten maken en dan kijken of de blonde meisjes lager scoren.
D: Oké, dank je wel voor deze suggestie, laten we hem eens bekijken. Je zegt (*schrijft op bord*):
- alle meisjes
- een IQ-test
- verschil in score

Laten we eens bij het eind beginnen: hoe groot moet het verschil zijn om te kunnen zeggen dat blonde meisjes dommer zijn? Bij een IQ-test is het gemiddelde altijd 100, onder de 80 is zwakbegaafd en boven de 140 is briljant.
Dus … Hoeveel verschil?
LL1: 10 of zo?
D: Waar hangt het van af of je 10 een groot genoeg verschil vindt?
LL2: Van wat je dom vindt.
D: Precies, het gaat erom dat je van te voren goed omschrijft wat je 'dom' vindt.
Kijk nog eens naar de vraagstelling: 'zijn blonde meisjes echt dommer dan andere meisjes?' Zijn er nog meer begrippen die je eerst moet omschrijven?
LL3: blond
D: Inderdaad. Want ik noem mezelf wel blond, maar noemen jullie bijvoorbeeld hem (wijst een jongen aan met donkerblond/licht bruin haar) blond?
LLN Ja/nee
D: Ik wil die discussie hier nu niet beslechten, maar het eerste criterium voor goed onderzoek is nu wel duidelijk: de vraag bevat eenduidige begrippen, of je moet goed omschrijven wat je ermee bedoelt.
Dat geldt dus ook voor 'dom' en de vraag of je dat met een IQ-test kunt meten.
Dan naar een ander aspect van de suggestie: alle meisjes. Is het nodig alle meisjes te testen?
LL4: Nee, dat lukt nooit.
D: Wie dan wel en hoe zoek je die uit?
LL4: Bijvoorbeeld alle meisjes hier op school of in deze klas.
D: Dat zijn twee suggesties; deze klas, zou dat genoeg zijn om het zeker te weten?
LL4: Nee,want dat zijn er maar acht.
D: En acht is te weinig?
LL4: Ja, want dan kan het toeval een te grote rol spelen.
D: Oké: voldoende grote aantallen. Dan maar de hele school, weet je het dan zeker?
LL5: Wel zekerder.
D: Oké, maar zou het een eerlijke steekproef zijn van alle blonde meisjes?
Welk criterium kunnen we daar uit halen voor goed onderzoek?
LL6: Dat je een steekproef eerlijk moet trekken.
D: Oké, mee eens, je moet een steekproef eerlijk trekken. Maar wat is 'eerlijk'?
(…)

- een inleidende tekst over de bedoeling van de praktische opdracht en het doen van onderzoek;
- de criteria voor goed onderzoek;
- de planning van de lessen en de deadlines;
- randvoorwaarden bij de uitvoering, zoals inperking van de onderwerpkeus en praktische aanwijzingen;
- de 'stappen' van het onderzoek, met ruimte voor aantekeningen bij elke stap;
- (in de expliciete groep) de reflectieopdrachten;
- de eisen aan het eindproduct, een artikel; 'auteursinformatie' en een voorbeeldartikel.

Het risico dat door het aanbrengen van meer structuur de leerlingen het als een serie achter elkaar uit te voeren opdrachten gaan beschouwen, is onderkend. In plaats van zelf hun eigen aanpak van een onderzoek te volgen, zouden leerlingen zich aan de hand van het boekje door een 'recept' heen kunnen werken, van onderzoeksvraag naar hypothese naar opzet enzovoort, zonder het geheel te overzien. Om dat te voorkomen is op een aantal plaatsen de zinsnede 'pas zonodig eerdere stappen aan' opgenomen en in de inleiding de volgende tekst opgenomen:

> Dit boekje is opgezet volgens de globale structuur van een onderzoek. Het is geen invuloefening, het is meer een logboek waarin je alle relevante gegevens tijdens het hele onderzoek noteert.
>
> Het kan dienen als leidraad, maar onderzoek is een heen-en-weer proces. Als je bezig bent met het bedenken van de uitvoering, kan het zijn dat je je onderzoeksvraag toch eigenlijk nét iets anders wilt formuleren. Het komt zelfs voor dat je aan het eind blijkt gegevens verzameld te hebben die antwoord geven op een andere vraag. In de wetenschap komt dat ook regelmatig voor, maar in wetenschappelijke artikelen zie je daar weinig van terug.
> Dit boekje begeleidt je bij de verschillende onderdelen van het onderzoek. Om het heen-en-weer karakter van onderzoek te benadrukken wordt regelmatig opgemerkt 'pas zonodig eerdere stappen aan.' Je schrijft dat dan op in dit boekje, maar in het artikel schrijf je alleen de uiteindelijke vraag, aanpak en conclusie (enz.) op. Uiteindelijk lever je zowel het artikel als het boekje in. Alleen het artikel wordt beoordeeld.

- *Artikel als product*

De aanleiding om een artikel als eindproduct te vragen in plaats van een verslag is gelegen in de constatering dat in de eerste cyclus veel leerlingen allerlei irrelevante, persoonlijke opmerkingen in hun verslag vermeldden zoals: 'Onze samenwerking liep perfect. Nadat we hadden besloten samen te gaan werken omdat onze 'partners' met lustrum waren gegaan is alles eigenlijk heel geregeld gegaan. We hebben als eerste de donderdag erop een werkplan gemaakt …'

Door te vragen volgens de opzet van een wetenschappelijk artikel te schrijven zou dit voorkomen kunnen worden of in ieder geval als niet juist gekwalificeerd kunnen worden. Tevens bevordert schrijven in deze vorm wellicht de aandacht voor de kwaliteit van het onderzoek in de zin van 'zeker weten?' In het boekje is een lijstje met eisen opgenomen.

Om de leerlingen een beeld te geven hoe het artikel eruit zou moeten zien, is een voorbeeldartikel gezocht van een onderzoek dat leerlingen zou aanspreken. Het moest Nederlandstalig zijn en van voldoende wetenschappelijk gehalte, maar tegelijkertijd voldoende toegankelijk voor de leerlingen. Uiteindelijk is een artikel uit *Lutra* gevonden en enigszins bewerkt (Lammertsma, Broekhuizen & Muskens, 1994). Ook is de informatie voor auteurs van dit tijdschrift in het boekje voor de leerlingen opgenomen (VZZ, 2006).

- *Het wetenschappelijke verhaal*

In hun boek *Meaning making in the secondary science classroom* gebruiken Mortimer en Scott (2003) de term 'staging the scientific story'.[28] Zij bedoelen daarmee dat in de gesprekken tussen docent en leerlingen een opbouw gevolgd kan worden naar de beoogde wetenschappelijke uitleg van verschijnselen e.d., waarbij voortgebouwd wordt op hetgeen de leerlingen al weten en kunnen.

> In staging the scientific story, the teacher needs to be aware of the existing and developing understandings of the students, being sensitive to the kinds of things students are saying in class, and drawing upon their knowledge of students' everyday views in the topic area. The challenge for the teacher is then to develop convincing lines of argument to interact and engage dialogically with those existing understandings. (p. 19)

In deze praktische opdracht wordt met het 'wetenschappelijke verhaal' de ontwikkeling van begrip van bewijs bedoeld. Het doel van de opdracht is immers om te leren de kwaliteit van onderzoek te bewaken met daarvoor geschikte criteria, volgens de opvattingen binnen de wetenschappelijke gemeenschap. In paragraaf 2.3 zijn die criteria, de elementen van begrip van bewijs, omschreven en in paragraaf 3.5 is beargumenteerd welke daarvan prioriteit in de praktische opdracht verdienen. De ontwikkeling en toepassing van het wetenschappelijke verhaal houdt dus de ontwikkeling en toepassing van (die selectie van elementen van) begrip van bewijs bij de leerlingen in.

Het is de bedoeling dat het wetenschappelijke verhaal op verschillende momenten in de praktische opdracht ten tonele gevoerd wordt. In de startles wordt een begin gemaakt, dat dan verder uitgewerkt kan worden in de begeleiding van het eigen onderzoek en in reflectieopdrachten. Gaandeweg het eigen onderzoek – al dan niet gelardeerd met de reflectieopdrachten – worden de verschillende criteria aan de orde gesteld en toegepast, zodat een samenhangend verhaal ontstaat van wat de wetenschap onder kwaliteit van onderzoek verstaat.

- *Duidelijker omschrijving van het beoogde docentgedrag bij de begeleiding*

In de eerste cyclus waren de docenten soms in verwarring over hetgeen van ze verwacht werd bij de begeleiding van leerlingen in de beide strategieën; mochten ze nu wel of niet de criteria voor goed onderzoek aan de orde stellen bij de impliciete strategie? Daarom is opnieuw verwoord wat van hen verwacht wordt, namelijk dat de docent ook zelf het initiatief neemt om te vragen hoe het onderzoek van de leerlingen vordert. Waar het leerlingonderzoek erom vraagt, stelt de docent de elementen van begrip van bewijs expliciet aan de orde, zowel in de expliciete als in de impliciete onderwijsleerstrategie.

In principe is er dus geen onderscheid tussen de strategieën waar het de directe begeleiding van de docent betreft. Het onderscheid zit erin dat bij de expliciete strategie de leerlingen reflectieopdrachten maken en die nabespreken met de docent. In die nabesprekingen komen de criteria – ook degene die niet opgeroepen worden door het eigen onderzoek van de leerlingen – expliciet aan de orde. In de impliciete strategie zijn twee momenten ingebouwd waarop de docenten schriftelijk feedback geven op tussenproducten.

7.2 Reflectieopdrachten

Algemene klacht over de centrale momenten van expliciete reflectie (MERs) uit de eerste cyclus was dat ze teveel tijd kostten: eerst een opdracht uitvoeren, dan erover napraten

[28] Mortimer en Scott (2003) wijzen zelf als bron het boek *Explaining science in the classroom* (Ogborn e.a., 1996) aan. Daarin wordt wel betoogd dat *explanation* een verhaal is, maar de metafoor van het theater (*staging)* is van Mortimer en Scott zelf.

en vervolgens (als de tijd het nog toeliet) toepassen op het eigen onderzoek. Een tweede manco betrof de relatie met het eigen onderzoek. Die werd nauwelijks gelegd, doordat ze in de tijd niet spoorden met het eigen onderzoek, maar ook doordat er vaak de tijd voor ontbrak. Aan beide problemen is gepoogd tegemoet te komen.

Ten eerste zijn de opdrachten gedecentraliseerd, dat betekent dat ze niet meer door alle leerlingen op hetzelfde tijdstip gemaakt worden, maar op het moment dat de leerlingen eraan toe zijn in hun eigen onderzoek. De naam is overeenkomstig gewijzigd van '*moment* van expliciete reflectie' in 'reflectie*opdracht*'.

Ten tweede is – in het verlengde van de eerste aanpassing – de structuur aangepast:

- onder de kop 'Zeker weten?' wordt de opdracht schriftelijk uiteengezet;
- nadat de leerlingen het eerste deel hebben gedaan en de bijbehorende vragen hebben beantwoord, wordt gevraagd welk criterium of welke criteria voor goed onderzoek uit deze opdracht zijn af te leiden; daarbij wordt ook verwezen naar het overzicht van de criteria in het boekje;
- vervolgens wordt gevraagd om deze criteria toe te passen op het eigen onderzoek en zonodig dat (deel van het) eigen onderzoek aan te passen;
- daarna wordt de opdracht nabesproken met de docent, die daarbij expliciet vraagt naar de toepassing van de criteria op het eigen onderzoek.

Door deze nieuwe structuur zouden de opdrachten in kortere tijd uit te voeren moeten zijn, ongeveer in twintig minuten. In figuur 7.3 is de eerste reflectieopdracht weergegeven; de andere opdrachten zijn te vinden in het boekje voor de leerlingen van de expliciete onderwijsleerstrategie (bijlage 7).

Voor het nabespreken van de opdrachten zijn scenario's opgesteld waarvan de essentie is om vooral op de criteria in te gaan en minder op het al dan niet juist beantwoorden van de opdrachten. De scenario's zijn te vinden in bijlage 8.

De inhoud van de opdrachten is grotendeels dezelfde gebleven als die van de MERs, aangezien die wel voldeed. Een uitzondering hierop is de eerste. Uit MER 1 is het practicum weggelaten (indertijd begroot op 20 minuten, de praktijk bleek nog iets langer). Het practicum bleek ook niet goed te functioneren in het naar boven krijgen van de criteria. Daarom is in de nieuwe reflectieopdracht 1 gekozen voor een opzet die in de leerling-interviews in de eerste cyclus gevolgd is. Leerlingen worden geconfronteerd met drie onderzoeksvragen waarvan vermeld wordt dat ze niet goed zijn. De eerste is dezelfde die in het klassengesprek aan de orde is geweest. Daarna wordt ze gevraagd er drie criteria uit te destilleren. Ze kunnen die eventueel uit hun aantekeningen van het klassengesprek halen of uit het boekje.

De opdrachten uit MER 2 en 3 (respectievelijk de springende gazelle en het koffieonderzoek, zie par. 5.2) zijn grotendeels gehandhaafd, omdat die wel voldeden in het duidelijk maken van criteria. Ze zijn wel aangepast aan de nieuwe structuur.

In de laatste reflectieopdracht is het centrale karakter wel grotendeels gehandhaafd, omdat in die opdracht de leerlingen commentaar op elkaars conceptartikel moeten geven. Dat vereist ten eerste dat die concepten op hetzelfde tijdstip beschikbaar zijn en ten tweede dat groepjes met elkaar in gesprek moeten, hetgeen uiteraard ook op hetzelfde tijdstip moet gebeuren. Er is wel sprake van iets meer flexibiliteit aangezien niet alle groepjes op hetzelfde tijdstip hoeven te overleggen. Wel is in de laatste opdracht de aandacht van de leerlingen meer gericht op de algemene kwaliteit (validiteit en betrouwbaarheid) en de centrale vraag 'hoe zeker ben je ervan dat de beschreven conclusies kloppen?'

Zeker weten?

Reflectieopdracht 1

Zoek de fout!

Weet je zeker dat je onderzoeksvraag goed geformuleerd is?

Hieronder staan drie onderzoeksvragen die geen van alle geheel goed zijn geformuleerd.
Elke onderzoeksvraag voldoet NIET aan een van de criteria voor een goede onderzoeksvraag.

- Onderstreep het gedeelte dat niet goed is en schrijf erbij waarom.
  Herformuleer daarna de onderzoeksvraag zodat hij wel goed is.

1 Zijn blonde meisjes dommer dan andere meisjes?

Niet goed omdat

Herformulering:

2 Waardoor leven huisdieren langer dan dieren in het wild?

Niet goed omdat

Herformulering:

3 Hoe komt het dat het aantal grijze zeehonden is toegenomen?

Niet goed omdat

Herformulering:

- Vat het bovenstaande samen in drie criteria voor een goede onderzoeksvraag:
  1
  2
  3

- Kijk nog eens naar jullie eigen onderzoeksvraag. Voldoet die aan deze criteria?
  Zo nee, pas hem dan aan zodat hij er wel aan voldoet.

BESPREEK DEZE OPDRACHT NA MET JE DOCENT

Figuur 7.3 Reflectieopdracht 1; de antwoordruimtes zijn in deze afbeelding enigszins verkleind.

7.3 De nieuwe populatie en instrumenten

Het tweede ontwerp is in winter en voorjaar 2004 op dezelfde scholen en met dezelfde docenten uitgevoerd als de eerste cyclus (zie par. 5.3). In beide scholen zijn de onderwijsleerstrategieën weer toegepast in twee 5 vwo groepen voor het vak biologie 1,2; één groep voor de expliciete onderwijsleerstrategie, één voor de impliciete. In tabel 7.1 zijn de aantallen leerlingen van de groepen weergegeven.

Op beide scholen betrof het één klas die voor de gehele duur van de praktische opdracht gesplitst werd. Op de AS was er voor elke groep een blokuur (90 minuten) ingeroosterd, daarnaast was er nog een uur biologie voor de hele klas; daarin werden andere onderwerpen behandeld. Op de SA was er voor elke groep eveneens een blokuur (80 minuten) ingeroosterd; daarnaast waren er voor de klas geen uren biologie.

De groepen zijn qua samenstelling homogener dan in de eerste ronde. Op twee leerlingen na hebben ze allen een natuurprofiel gekozen. Wel ligt in de impliciete groep op de SA de verhouding tussen jongens en meisjes erg scheef.

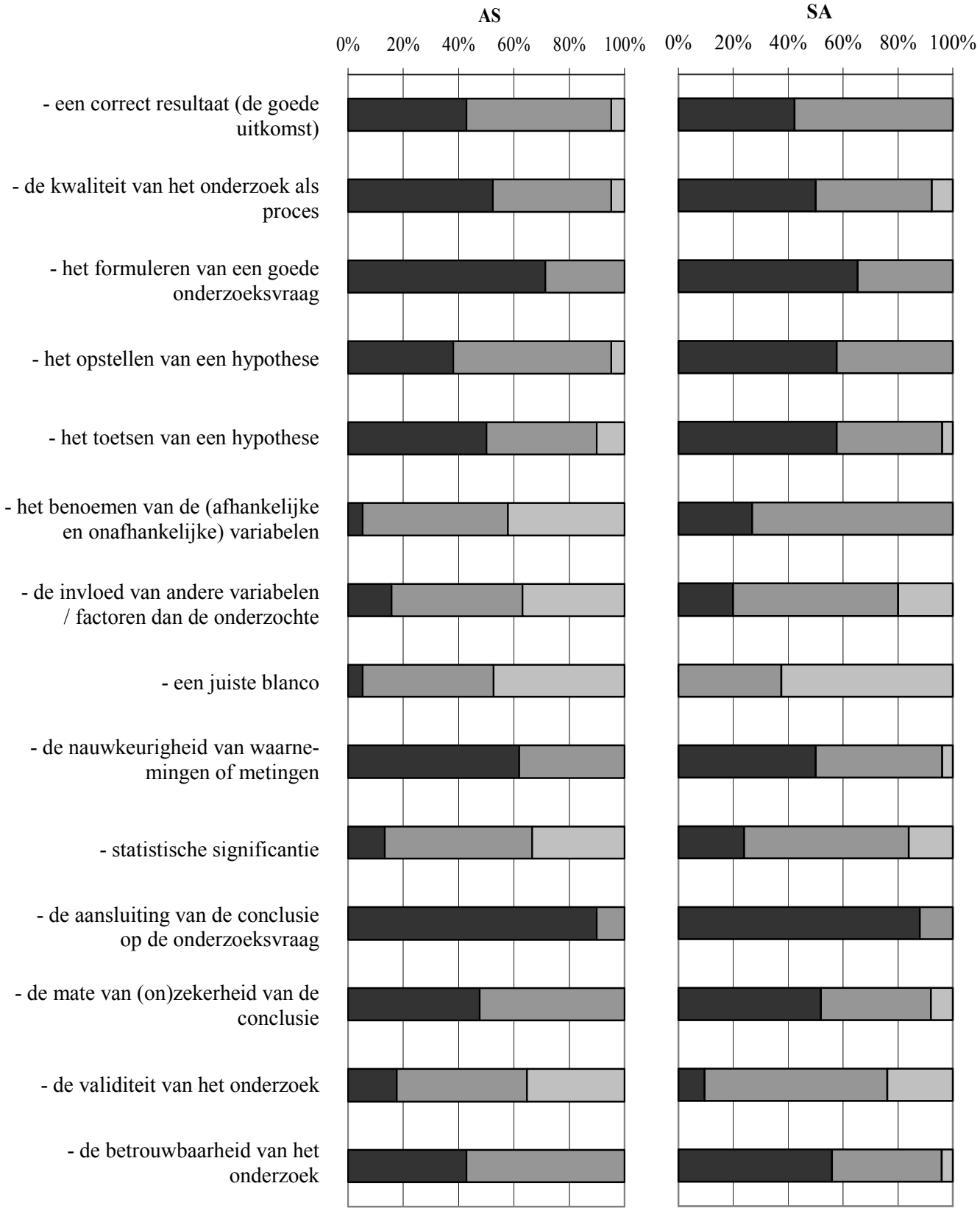
- *Ervaringen in het doen van onderzoek*

Bij deze groepen is met behulp van een vragenlijst geïnventariseerd welke ervaringen ze met onderzoek doen op school hadden opgedaan; die vragenlijst is afgenomen in hetzelfde lesuur als de voortoets. Er is gevraagd in welke vakken ze ervaringen met onderzoek doen hebben opgedaan en in hoeverre er aandacht voor verschillende elementen van begrip van bewijs is geweest (voor de vragenlijst, zie bijlage 9).

De leerlingen op beide scholen gaven aan dat ze vooral bij de natuurwetenschappelijke vakken onderzoek gedaan hebben. Ook de gammavakken en godsdienst werden genoemd, zij het in mindere mate. Van veel aspecten van onderzoek doen gaven de leerlingen aan dat daar soms of altijd aandacht aan wordt besteed (fig. 7.4); op de Amsterdamse School vooral bij scheikunde, op de School in Amstelveen vooral bij natuurkunde. Bij navraag bij de docenten natuurkunde en scheikunde blijkt dit beeld behoorlijk genuanceerd te moeten worden. Weliswaar wordt er bijvoorbeeld aandacht besteed aan het formuleren van een goede onderzoeksvraag, maar niet aan de criteria daarvoor. De nadruk ligt volgens de docenten hoofdzakelijk op een correct resultaat en daarmee op zaken als de aansluiting van de conclusie op de onderzoeksvraag en de nauwkeurigheid van procedures en metingen. Het duidelijkst wordt de overschatting van de leerlingen geïllustreerd met de antwoorden bij ‘de validiteit van het onderzoek’; op beide scholen zei meer dan de helft van de leerlingen dat het soms of altijd aan de orde komt, volgens de docenten is die term nooit gevallen. De verschillende aspecten komen niet systematisch aan de orde, wel af en toe, aldus de Amstelveense docent scheikunde. Het lijkt er dus op dat veel leerlingen wat deze aspecten betreft wel de klok hebben horen luiden, maar dat de docenten niet weten waar de klepel hangt. Zowel in Amsterdam als in Amstelveen komt open onderzoek, waarbij de leerlingen zelf onderwerp en onderzoeksvraag mogen bepalen, wat natuur- en scheikunde betreft niet eerder dan bij het profielwerkstuk (eind 5 vwo, begin 6 vwo) aan de orde.

Tabel 7.1 Omvang en samenstelling van de leerlinggroepen in de tweede ronde.

Groep	*meisjes*	*jongens*	*natuurprofiel*	*maatschappijprofiel*	*totaal*
AS expliciet	4	6	10	-	10
AS impliciet	6	5	11	-	11
SA expliciet	9	6	14	1	15
SA impliciet	13	1	13	1	14
totaal	32	18	48	2	50



Figuur 7.4 Onderzoekservaringen zoals door de leerlingen waargenomen. De vraag luidde: 'In hoeverre is bij onderzoek doen aandacht besteed aan de volgende aspecten?' Het linker deel van elke balk geeft het percentage leerlingen aan dat 'altijd' antwoordde, het middelste deel 'soms', het rechter (lichte) deel 'nooit'. Links staan de ervaringen van de leerlingen van de Amsterdamse School, rechts die van de School in Amstelveen.

Op beide scholen wordt door de docenten onderkend dat meer afstemming tussen de natuurwetenschappelijke vakken wenselijk is. Er worden daartoe initiatieven ondernomen, maar voor de groepen uit de tweede ronde hebben die geen effect gehad. De ervaringen binnen het vak biologie zijn voor deze groepen vergelijkbaar met die uit de eerste ronde (zie par. 5.3).

- *Instrumenten*

Ook in de tweede ronde is de bijgestelde praktische opdracht beoordeeld op uitvoerbaarheid en effectiviteit. In tabel 7.2 wordt een overzicht gegeven van de instrumenten die zijn gebruikt om de onderzoeksvragen te beantwoorden. De instrumenten zijn aangepast op de manier zoals in paragraaf 6.3 is beschreven: er zijn geen leerlingen meer geïnterviewd, maar er zijn veel meer gesprekken tussen leerlingen onderling en tussen docent en leerlingen opgenomen. Door uit elke groep twee duo's te volgen is geprobeerd meer zicht te krijgen op de ontwikkeling van begrip van bewijs. Van vrijwel alle leerlingen zijn de aantekeningen, conceptartikelen en definitieve artikelen verzameld. Het was ook de opzet om een jaar later (in klas 6) de profielwerkstukken van de leerlingen te verzamelen om te zien in hoeverre daarin kwaliteitscriteria gehanteerd waren. Dat is slechts ten dele gelukt. Bovendien hebben sommige leerlingen een profielwerkstuk gemaakt in samenwerking met een leerling die niet in de onderzoeksklas zat. Om deze redenen is afgezien van het analyseren van de profielwerkstukken, ook van de werkstukken die wel verzameld waren.

- *Observaties en evaluaties met de docenten*

Tijdens de lessen is door de onderzoeker geobserveerd welke de activiteiten docent en leerlingen ondernamen om daarna te kunnen beoordelen in hoeverre die overeenkwamen met de beoogde activiteiten zoals omschreven in de scenario's.

Na afloop van vrijwel elke les is er nagepraat met de docent, waarbij ook de observaties betrokken werden. Zonodig werd de docent gevraagd naar redenen waarom hij of zij van het scenario afgeweken was. Na afloop van de gehele praktische opdracht op beide scholen is met beide docenten samen het geheel geëvalueerd.

Tabel 7.2 Overzicht welke onderzoeksinstrumenten in de tweede cyclus bijdragen aan de beantwoording van de vragen naar uitvoerbaarheid en effectiviteit.

instrument	*vraag naar uitvoerbaarheid*	*vraag naar effectiviteit*
observatie door onderzoeker	*	
opname van gesprekken tussen leerling en docent en leerlingen onderling (inclusief de reflectieopdrachten): acht onderzoeksgroepjes	*	*
evaluatie lessen en opdrachten met docenten	*	
aantekeningen van leerlingen in boekjes	*	*
conceptartikelen		*
definitieve artikelen		*
voor- en natoets		*
evaluatie door de leerlingen	*	*

- *Analyseren van de gesprekken*

De complete startlessen en alle gesprekken van de geselecteerde onderzoeksgroepjes met betrekking tot hun onderzoek en gelieerde activiteiten tijdens de lesuren zijn uitgetypt. De transcripten zijn onderverdeeld in episodes op basis van het gespreksonderwerp en de gesprekspartners. De episodes zijn onderverdeeld in uitspraken per spreker. Alle uitspraken zijn bekeken om te zien of er een of meer elementen van begrip van bewijs aan de orde komen. Zo ja, dan is de uitspraak ingedeeld in de categorieën van het analysekader: problematiseren, beschrijven, expliciteren, generaliseren, toepassen. De gecategoriseerde uitspraken zijn geordend en de erin voorkomende patronen gescoord (zie par. 6.3). De indeling in categorieën en toewijzing aan elementen van begrip van bewijs zijn onafhankelijk uitgevoerd door twee beoordelaars, waarna bespreking volgde tot consensus was bereikt. De interbeoordelaarsovereenstemming is berekend voor beide indelingen. De proportie overeenstemming tussen de scores voor wat element van begrip van bewijs betreft bedroeg 0,79, voor wat de indeling in categorieën betreft 0,84. Cohen's kappa was respectievelijk 0,58 en 0,64. Dat is niet erg hoog, maar wordt wel gezien als acceptabel, 'moderate' resp. 'substantial' volgens Landis en Koch (1977, geciteerd uit Heuvelmans & Sanders, 1993).

- *Artikelen, conceptartikelen en aantekeningen van leerlingen*

Bij het analyseren van de aantekeningen die de leerlingen in de boekjes hebben gemaakt en de conceptartikelen, is gekeken naar het gebruik van de verschillende elementen van begrip van bewijs. De conceptartikelen en de definitieve artikelen zijn geanalyseerd op dezelfde manier als de leerlingverslagen in de exploratieve fase (zie par. 3.4). Voor alle elementen van begrip van bewijs is beoordeeld of eraan voldaan wordt en of als er niet aan voldaan wordt, of dat dan beargumenteerd is. De analyse is uitgevoerd door twee onafhankelijke beoordelaars (proportie overeenstemming 0,81, Cohen's kappa 0,71), waarna bespreking volgde tot overeenstemming was bereikt.

- *Toets en evaluatieformulier*

Aan het begin en aan het eind van de praktische opdracht hebben de leerlingen weer een toets gemaakt. Op basis van de bevindingen in de eerste ronde is de toets bijgesteld. De belangrijkste tekortkomingen waren de omvang (te groot), de moeilijkheidsgraad (sommige vragen te moeilijk, andere te makkelijk) en de betrouwbaarheid (te laag).

Alle vragen zijn heroverwogen en zonodig bijgesteld of vervangen, waarbij gelet is op voldoende dekking van de elementen van begrip van bewijs en begripsvaliditeit. De nieuwe versie omvat 14 opgaven met in totaal 31 items; de maximumscore bedraagt 34 punten. Dat lijkt niet minder dan de toets uit de eerste ronde, maar de formulering van de vragen is zo aangepast dat de beantwoording minder tijd kost. Er is daardoor ruimte ontstaan voor het invullen van een vragenlijst over ervaringen met onderzoek doen (bij de voortoets, zie eerder in deze paragraaf) of een evaluatieformulier (bij de natoets, zie onder). De toets is verdeeld in drie deeltoetsen op basis van de elementen van begrip van bewijs die (als belangrijkste) in de vragen aan de orde komen (resp. element 1-6, 7-17 en 18-23). Toets en formulieren zijn opgenomen in bijlage 9, 10 en 11.

In eerste instantie zijn de antwoorden gescoord met een code die aangeeft of en zo ja, welk element van begrip van bewijs gebruikt is. Als er sprake is van een of meer voor de vraag relevante)elementen, is één of meer punten toegekend aan het antwoord. Dit levert zowel informatie op over de mate van begrip van bewijs die leerlingen ten toon spreiden als informatie over welke elementen ze in stelling brengen om aspecten van onderzoek te beoordelen.

Alle toetsen zijn door twee onafhankelijke beoordelaars geanalyseerd (proportie overeenstemming 0,82), waarna bespreking volgde tot overeenstemming was bereikt. De scores

Tabel 7.3 Betrouwbaarheid van de toets en drie deeltoetsen op verschillende manieren berekend.

	aantal leerlingen	*aantal items*	*max. score*	*coëfficiënt alpha*	*GLB*	*alpha bij 40 items*
Hele toets	41	24	27	0,59	0,91	0,70
deeltoets 1	41	10	12	0,53	0,84	0,82
deeltoets 2	41	7	8	0,46	0,67	0,83
deeltoets 3	41	7	7	0,33	0,54	0,74

zijn geanalyseerd met het programma TiaPlus® (build nr. 300), een toets- en itemanalyseprogramma van het Cito.[29] Op basis van die analyse zijn zeven items verwijderd. Het aantal items komt daarmee op 24, de maximumscore op 27 punten. Een aantal leerlingen scoorde op de natoets beduidend lager dan op de voortoets. Waarschijnlijk is dat veroorzaakt doordat deze leerlingen de natoets veel minder serieus hebben gemaakt dan de voortoets. De scores van deze leerlingen zijn uit de analyses verwijderd, waarmee het aantal leerlingen in de analyses op 41 komt.

De betrouwbaarheid (coëfficiënt alpha) van de bijgestelde toets is 0,59. Dat is laag, maar wordt voor een deel veroorzaakt door het gering aantal opgaven, de grote diversiteit van de inhoud en de kleine afnamepopulatie. Coëfficiënt alpha is immers vooral een maat voor de interne consistentie van een toets (Veldhuijzen, Godebeld & Sanders, 1993). Verhelst (2000) brengt onder de aandacht dat alpha (net als andere manieren om de betrouwbaarheid te berekenen) slechts een schatting van de ondergrens van de betrouwbaarheid is. Hij gaat in op verschillende manieren om aan de hand van één toetsafname de betrouwbaarheid te bepalen. De grootste ondergrens van die schattingen van de betrouwbaarheid is de 'greatest lower bound' (GLB)[30], die echter bij kleine groepen leerlingen een te hoge schatting geeft van de betrouwbaarheid. Hij laat zien dat bij grote groepen (minimaal 200 personen) de GLB de betrouwbaarheid asymptotisch nadert. De GLB berekend voor de hele toets is 0,91, wat dus een overschatting is.

Coëfficiënt alfa is ook erg afhankelijk van de lengte van de toets. Een manier om iets meer te kunnen zeggen over de interne consistentie van een korte toets is het omrekenen naar een standaardlengte van 40 items met de formule van Spearman-Brown.

In tabel 7.3 zijn die verschillende schattingen – coëfficiënt alpha, GLB en alpha bij 40 items – weergegeven. Coëfficiënt alpha en de GLB zijn afhankelijk van het aantal items en dus het hoogst voor de hele toets. De werkelijke betrouwbaarheid ligt ergens tussen deze waarden, waarschijnlijk wat dichter bij alpha dan bij de GLB. De formule van Spearman-Brown is eigenlijk bedoeld om te bepalen hoeveel van hetzelfde type items toegevoegd zou moeten worden om een bepaalde betrouwbaarheid te bereiken. Het laat zien dat de interne consistentie van de deeltoetsen 1 en 2 groter is dan die van de hele toets.

In de eerste cyclus was de betrouwbaarheid van de toets dusdanig laag dat aan de scores geen betekenis kon worden toegekend. De schattingen van de betrouwbaarheid van de herziene toets zijn nog steeds niet erg hoog, maar wel hoger. Daarom kennen we met de nodige voorzichtigheid toch enige betekenis aan de toetsscores toe.

[29] Het programma TiaPlus® is voor niet-commercieel, wetenschappelijk gebruik te verkrijgen bij het Cito in Arnhem.

[30] In de handleiding van TiaPlus® (Cito, 2006, p.18) staat de GLB als volgt omschreven: Where Cronbach's Alpha is suitable for homogeneous tests (all items mapping the 'same' trait), the GLB is a better estimate for the reliability if the test is heterogeneous (items mapping two or more different traits or abilities). The GLB coefficient is closely related tot the Lambda family of reliability coefficients and was proposed by ten Berge, Snijders and Zegers (*Psychometrika,* 1981, p. 201-213).

Bij de natoets is de leerlingen ook een evaluatieformulier voorgelegd met negen vragen, variërend van hoeveel uur ze buiten de lessen aan hun onderzoek hadden besteed, via hun waardering van een aantal aspecten van de opdracht tot de ervaren leeropbrengst. De vragen naar waardering bevatten elk drie of vier vijfpuntssschalen tussen twee tegengestelde begrippen, bijvoorbeeld:

2. Wat vond je van het werken aan het onderzoek?

moeilijk	0	0	0	0	0	makkelijk
nuttig	0	0	0	0	0	overbodig
te lang duren	0	0	0	0	0	te weinig tijd
motiverend	0	0	0	0	0	demotiverend

Aan eerste bolletje is de waarde 1 toegekend, aan het laatste de waarde 5. Vervolgens zijn gemiddeldes en standaarddeviaties berekend. De antwoorden op de verschillende vragen zijn apart verwerkt.

7.4 Uitvoerbaarheid

Na de eerste ronde is geconstateerd dat er nog het nodige schortte aan de uitvoerbaarheid van de praktisch opdracht, zowel voor de docenten als voor de leerlingen (par. 5.4). Om de uitvoerbaarheid te verbeteren zijn de beoogde docent- en leerlingactiviteiten in de verschillende lessen opgenomen in scenario's voor beide varianten van de praktische opdracht. Ook zijn er enkele aanpassingen gedaan die zouden moeten leiden tot een betere uitvoerbaarheid (par. 7.1).

Aan de hand van observaties van de onderzoeker, opnames van de gang van zaken tijdens de lessen, en evaluaties van docenten en leerlingen wordt de invloed van deze aanpassingen op de uitvoerbaarheid besproken. Allereerst wordt beschreven in hoeverre de praktische opdracht is verlopen zoals gepland, dus of de beoogde docent- en leerlingactiviteiten daadwerkelijk zijn uitgevoerd op de geplande tijdstippen (par. 7.4.1). Daarna komen de aangepaste uitgangspunten – zoals beschreven in paragraaf 7.1 – aan de orde, zij het in een iets andere volgorde:

- de startlessen (par. 7.4.2);
- motief als drijvende kracht (7.4.3);
- het beoogde docentgedrag (7.4.4);
- het doen van een éigen onderzoek (7.4.5);
- het boekje (7.4.6);
- een artikel als eindproduct (7.4.7);
- de reflectieopdrachten (7.4.8);
- de impliciete onderwijsleerstrategie (7.4.9);
- de ontwikkeling van het wetenschappelijke verhaal van begrip van bewijs (7.4.10).

De paragraaf wordt besloten met samenvattende conclusies met betrekking tot de uitvoerbaarheid van de praktische opdracht (7.4.11).

7.4.1 Gegaan zoals gepland?

De gang van zaken tijdens de lessen is vergeleken met de geplande activiteiten zoals beschreven in de scenario's. Een samenvatting daarvan staat in de tabellen 7.4 en 7.5.

De praktische opdracht is door de docenten in grote lijnen uitgevoerd zoals gepland. Dat wil zeggen, de startlessen zijn gegeven, de reflectieopdrachten besproken, in de impliciete groep is er tussentijds schriftelijk feedback gegeven en zijn de conceptartikelen van commen-

taar voorzien (tabel 7.4). Ondanks deze constateringen zijn er wel enkele knelpunten met betrekking tot de algemene gang van zaken te signaleren.

Ten eerste de omvang en de planning. In Amsterdam zijn er tien plus twee lessen aan besteed, in Amstelveen was een les meer nodig. De toets aan het eind is niet meteen na de laatste les afgenomen, maar later. Op beide scholen blijkt het moeilijk zo'n omvangrijke serie lessen te laten plaatsvinden op de van te voren geplande tijdstippen. De algemene planning van activiteiten op een school blijkt vaak zwaarder te (moeten) wegen dan de planning van docenten, die dan vervolgens moeten schuiven.

Ten tweede was het voor de docenten soms moeilijk om overzicht te houden over de gang van zaken, onder andere door de twee verschillende groepen, maar ook doordat elk duo leerlingen een ander onderzoek deed. Het vereiste een goede voorbereiding en planning, die in sommige gevallen echter ook door de onderzoeker bewaakt moest worden. Met name het aangeven wanneer de impliciete groep wat moest inleveren of wanneer de expliciete groep een reflectieopdracht moest maken of bespreken, werd een aantal keren door de onderzoeker in herinnering gebracht. Hij was daardoor degene die de scenario's bewaakte, want die hadden voor de docenten niet altijd de beoogde sturende werking. Niet omdat de scenario's te onduidelijk waren, maar omdat de docenten ze niet altijd raadpleegden bij het voorbereiden. Want naast het begeleiden van de praktische opdracht in 5 vwo hadden de docenten uiteraard andere klassen en andere werkzaamheden, hetgeen het houden van overzicht en de voorbereiding niet positief beïnvloedde.

In de eindevaluatie zeiden beide docenten de meer geconcentreerde periode – ten opzichte van de eerste ronde – als positief ervaren te hebben. Bea merkte echter wel op dat ze nu meer druk voelde: '… mensen hebben me nodig, willen me spreken en ik ben er niet voor ze.' In de eerste ronde was er vaker de gelegenheid om tussen de lessen door, voor of na de lessen die niet aan het onderzoek werden besteed, even met leerlingen over hun onderzoek te spreken. Tijd blijkt daarmee een sterk bepalende factor voor de uitvoering.

Ook de leerlingen hebben globaal gezien de opdracht uitgevoerd zoals bedoeld. Dat wil zeggen, ze hebben een onderwerp bedacht, een onderzoeksvraag, hypothese en onderzoeksopzet geformuleerd, een onderzoek uitgevoerd en er een artikel of verslag over geschreven. Maar meer in detail bekeken, zijn er veel verschillen met de scenario's te constateren (tabel 7.5). De belangrijkste afwijking is de planning. Al in les 2+3 begonnen ze achter te lopen, omdat de meeste leerlingen nog niet aan het formuleren van een hypothese en een onderzoeksopzet waren toegekomen. Vooral in de expliciete groep trad die vertraging op, onder andere doordat ze reflectieopdracht 1 moesten maken en bespreken. De aanloop naar de uitvoering, waarin veel praktische zaken geregeld moesten worden, bleek de leerlingen meer tijd te kosten dan voorzien was. Waar de uitvoering in les 5 had moeten plaatsvinden, werd dat vrijwel altijd later, tussen 5 en 6, in les 6 of soms zelfs pas na les 9! Niet alleen het bedenken van onderwerp, onderzoeksvraag, hypothese en voorspelling leverde vertraging op, ook de beschikbaarheid van de technisch onderwijsassistent of materiaal (bacteriën, pH-meter, broedstoof op de juiste temperatuur) deed dat. Daardoor werd dan ook later dan gepland aan het schrijven begonnen, hoewel de docenten de leerlingen wel stimuleerden alvast gegevens op te schrijven als ze om praktische redenen nog niet met de uitvoering konden beginnen, hetgeen

(op de volgende bladzijde)

Tabel 7.4 Vergelijking van de gerealiseerde docentactiviteiten met de in de scenario's geplande; +: geheel gerealiseerd; ±: gedeeltelijk gerealiseerd; -: niet gerealiseerd; *: niet op dat moment, maar wel later gerealiseerd; nvt: niet van toepassing; ?: onduidelijk of gerealiseerd; RO: reflectieopdracht.

Lesuur	Docentactiviteit zoals beschreven in scenario	AS expliciet	AS impliciet	SA expliciet	SA impliciet
1	▪ Start de serie lessen met een gesprek over de vraag 'wat is goed onderzoek?'	+	+	+	+
	▪ Deelt boekjes uit	+	+	+	+
	▪ Geeft praktische informatie met betrekking tot werkwijze, randvoorwaarden en planning.	-	-	+	+
	▪ Geeft 'huiswerk' op: inleiding lezen, overleggen over onderwerp en onderzoeksvraag, info zoeken.	-	+	+	+
	▪ Beantwoordt eventuele vragen	-	-	+	+
2 + 3	▪ Geeft aan wat er aan het eind van dit blokuur gedaan moet zijn en dat boekje ingeleverd moet worden.	+	+	+	+
	▪ Stelt vragen over voortgang keuzeproces en formulering van vraagstelling.	+	+	+	+
	▪ Helpt leerlingen op weg door vragen te stellen, maar doet zelf slechts mondjesmaat suggesties.	+	+	+	+
	▪ Expliciteert zonodig criteria voor goed onderzoek (met name bvb 1).	+	±	±	-
	alleen expliciet				
	▪ Bespreekt RO 1.	+		+	
	▪ Controleert of de criteria juist zijn opgeschreven. Geeft zonodig uitleg.	+		+	
	▪ Bespreekt eventueel RO 2. Geeft zonodig uitleg.	-		-	
tussen 3 en 4	*alleen impliciet*				
	▪ Geeft schriftelijk commentaar op de boekjes.		+		+
	▪ Expliciteert zonodig criteria voor goed onderzoek (met name bvb 1, 3, 4, 5, 6, en 18).		-		±
4 + 5	*alleen impliciet*				
	▪ Geeft zonodig verduidelijking bij het commentaar.		+		+
	▪ Ondersteunt leerlingen door vragen te stellen, maar doet zelf slechts mondjesmaat suggesties.	+	+	±	±
	▪ Expliciteert zonodig criteria voor goed onderzoek (met name bvb 8, 9 en 10).	+	±	+	±
	alleen expliciet				
	▪ Bespreekt RO 2.	+		+	
tussen 5 en 6	▪ Geeft zonodig feedback	+	nvt	nvt	+
6 + 7	▪ Ondersteunt leerlingen door vragen te stellen, maar doet zelf slechts mondjesmaat suggesties.	+	+	±	-
	▪ Expliciteert zonodig criteria voor goed onderzoek (met name bvb 8, 9, 10, 18 en 21).	+	±	+	-
	alleen expliciet				
	▪ Bespreekt RO 3.	*		*	
tussen 7 en 8	▪ Geeft zonodig feedback.	nvt	nvt	+	nvt
	▪ Expliciteert zonodig criteria voor goed onderzoek.			-	
8 (+ 9)	▪ Geeft zonodig feedback.	+	+	+	+
	▪ Expliciteert zonodig criteria voor goed onderzoek.	+	±	-	-
9	*alleen expliciet*				
	▪ Begeleidt RO 4. Geeft zonodig uitleg.	*		+	
tussen 9 en 10	▪ Geeft zonodig feedback.	+	+	+	nvt
	▪ Expliciteert zonodig criteria voor goed onderzoek.	?	?	-	
	alleen impliciet				
	▪ Geeft commentaar op de concept artikelen.		4/5		+
	▪ Expliciteert zonodig criteria voor goed onderzoek.		±		-
10 (+ 11)	▪ Expliciteert zonodig criteria voor goed onderzoek.	-	±	-	+
	alleen impliciet				
	▪ Geeft zonodig verduidelijking bij commentaar		+		+
na 10 (11)	▪ Beoordeelt artikelen, geeft commentaar.	+	+	+	+

Lesuur	Leerlingactiviteit zoals beschreven in scenario	AS expliciet	AS impliciet	SA expliciet	SA impliciet
1	▪ Nemen deel aan gesprek 'wat is goed onderzoek?'	+	+	+	+
	▪ Kijken boekjes in.	+	+	+	+
	▪ Luisteren naar praktische informatie en noteren huiswerk.	-	-	+	+
	▪ Overleggen over onderwerp en onderzoeksvraag.	-	-	±	+
tussen 1 en 2	▪ Lezen inleiding van het boekje.	?	?	?	?
	▪ Overleggen over onderwerp en onderzoeksvraag.	±	1/5	5/7	4/7
	▪ Zoeken informatie over een of meer onderwerpen.	?	-	?	3/7
2 + 3	▪ Zoeken informatie over een of meer onderwerpen.	-	-	+	+
	▪ Vragen zonodig advies van docent.	+	+	+	+
	▪ Maken keuze met betrekking tot onderwerp en onderzoeksvraag.	±	+	6/7	+
	▪ Formuleren een geschikte hypothese.	-	±	-	4/7
	▪ Bedenken onderzoeksopzet en voorspelling.	-	±	-	3/7
	alleen impliciet				
	▪ Leveren boekje in.		+		6/7
	alleen expliciet				
	▪ Maken RO 1.	+		+	
	▪ Bespreken de uitkomst met de docent.	+		+	
	▪ Passen zonodig onderzoeksvraag aan.	±		+	
	▪ Maken eventueel al RO 2.	-		-	
tussen 3 en 4	*alleen expliciet*				
	▪ Formuleren een geschikte hypothese.	*		±	
	▪ Bedenken onderzoeksopzet en voorspelling.	*		±	
4 + 5	*alleen impliciet*				
	▪ Lezen commentaar op hun opzet. Vragen zonodig verduidelijking.		4/5		+
	alleen expliciet				
	▪ Maken (indien nog niet gedaan) RO 2.	+		5/7	
	▪ Bespreken de uitkomst met de docent.	+		+	
	▪ Passen zonodig hypothese, opzet en voorspelling aan.	+		+	
	▪ Werken precieze opzet uit, bereiden uitvoering voor. Vragen zonodig feedback.	4/5	+	+	+
	▪ Voeren onderzoek uit. Vragen zonodig feedback.	-	-	2/7	-
tussen 5 en 6	▪ Voeren onderzoek uit. Vragen zonodig feedback.	±	±	3/7	1/7
6 + 7	▪ Voeren onderzoek uit. Vragen zonodig feedback	±	±	6/7	+
	▪ Trekken een conclusie en evalueren hun onderzoek.	*	*	*	*
	▪ Maken een begin met het schrijven van een artikel.	±	±	±	3/7
	alleen expliciet				
	▪ Maken RO 3.	1/5		+	
	▪ Bespreken de uitkomst met de docent.	-		3/7	
	▪ Passen zonodig onderzoeksvraag aan.	-		?	
tussen 7 en 8	▪ Schrijven aan hun artikel. Vragen zonodig feedback.	±	±	±	±
8 (+ 9)	▪ Schrijven aan hun artikel. Vragen zonodig feedback.	+	+	±	4/7
	▪ Leveren concept artikel in.	*	*	3/7	*
9	*alleen expliciet*				
	▪ Maken RO 4.	*		3/7	
tussen 9 en 10	*alleen expliciet*				
	▪ Stellen hun concept bij.	-		-	
10 (+ 11)	*alleen impliciet*				
	▪ Lezen commentaar op hun concept artikel, vragen zonodig verduidelijking.		+		+
	▪ Stellen hun concept bij.	-	-	-	±
na 10 (11)	▪ Stellen hun concept bij.	+	+	+	+
	▪ Leveren definitieve versie in.	*	*	+	+

(op de vorige bladzijde)
Tabel 7.5 Vergelijking van de gerealiseerde leerlingactiviteiten met de in de scenario's geplande; +: geheel gerealiseerd; ±: gedeeltelijk gerealiseerd.; ./. aantal duo's van totaal waarbij gerealiseerd; -: niet gerealiseerd; *: niet op dat moment, maar wel later gerealiseerd; ?: onduidelijk of gerealiseerd; RO: reflectieopdracht.

de leerlingen ook deden. Vandaar dat bijvoorbeeld in tabel 7.5 bij les 6+7 te lezen valt dat de leerlingen (nog) geen conclusie trekken, maar wel met schrijven beginnen.

Opmerkelijk genoeg liepen de leerlingen op de Amsterdamse School meer achter op hun planning dan de leerlingen van de School in Amstelveen, terwijl de lessen in Amstelveen in een korter tijdsbestek – steeds een week tussen de blokuren – plaatsvonden dan die in Amsterdam, waar de reeks onderbroken werd door toetsweken, vakantie en schoolreis. De leerlingen in Amsterdam benutten de extra tussentijd dus niet. Integendeel, dit leek contraproductief te werken. Na een periode van drie weken tussen les 6+7 en les 8+9 – waarin vakantie en een schoolreis vielen – waren veel leerlingen nauwelijks opgeschoten en waren veel leerlingen hun boekje vergeten mee te nemen; na die datum zijn dan ook weinig aantekeningen in de boekjes te vinden. Een korte periode waarin het onderzoek in zijn geheel moet plaatsvinden blijkt hier dus ook beter te zijn.

De hoeveelheid tijd die leerlingen zeiden te hebben besteed naast de lessen liep sterk uiteen; van 4 tot 15 uur op de Amsterdamse School en van 3 tot maar liefst 23 uur op de School in Amstelveen. Leerlingen hebben – zeker in hun eigen beleving – ook veel andere dingen te doen, zowel in het kader van hun schoolwerk als daarbuiten. Zeker in een toetsweek maken leerlingen geen tijd vrij om aan het onderzoek te werken.

In de les gebruikten de leerlingen hun tijd echter niet efficiënt, zo blijkt uit de observaties: ze deden veel andere dingen tussendoor, zoals kletsen over andere zaken (de reis, het weekend), werken voor andere vakken of wachten tot ze de docent konden spreken. Op de vraag in de schriftelijke evaluatie bij de natoets 'Waren de lessen die aan het onderzoek besteed zijn voor jou de moeite waard?' antwoordde 72 procent van de leerlingen 'ja', in Amsterdam was de score lager dan in Amstelveen: 65 tegen 78 procent. Toch vonden ze achteraf dat ze genoeg tijd hadden gehad om aan het onderzoek te werken (figuur 7.5).

Op de vraag in hoeverre ze het werken aan het onderzoek motiverend dan wel demotiverend vonden antwoordden de meeste leerlingen in het midden waarbij in Amstelveen meer uitersten werden gekozen. De meeste leerlingen vonden het vaker moeilijk dan makkelijk en vaker nuttig dan overbodig.

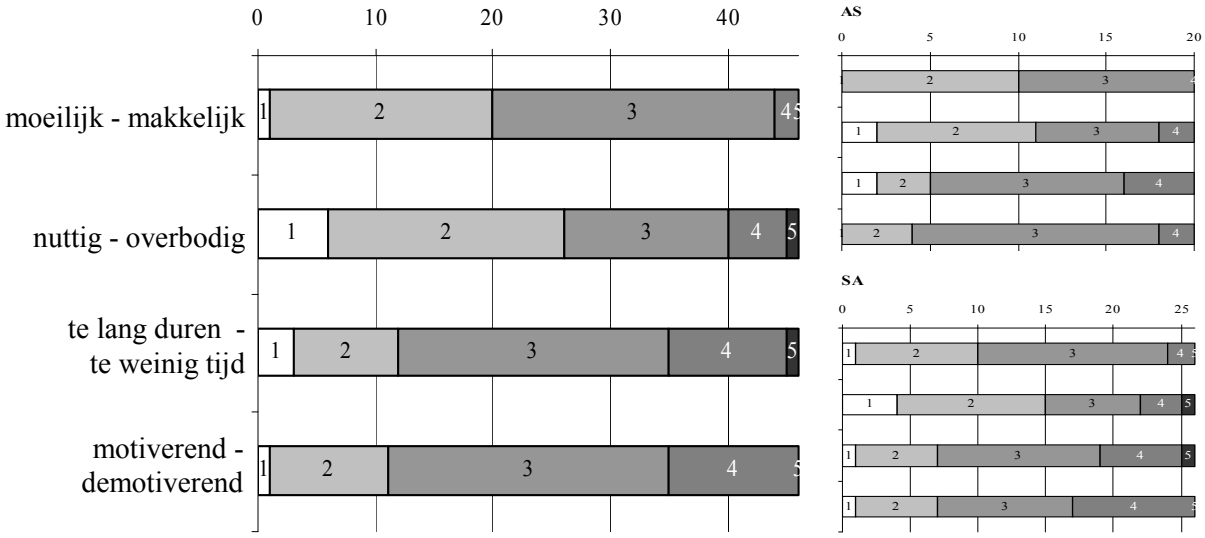
In de evaluatie, maar ook tijdens de lessen uitten leerlingen regelmatig hun frustratie over het feit dat ze op dat moment niet verder konden. Was dat in Amstelveen vooral vanwege praktische, technische complicaties of de afwezigheid van de technisch onderwijsassistent, in Amsterdam waren leerlingen ook vaak aan het wachten op een moment dat ze Lex konden spreken. De tijd die Lex met een duo in gesprek was, soms wel een kwartier, was dan ook aanmerkelijk langer dan bij Bea. Maar waar leerlingen in Amstelveen vaak – maar zeker niet altijd – doorwerkten als ze even niet bij Bea terecht konden, deden leerlingen in Amsterdam dat meestal niet, ze werden onrustig; ze gingen dan rondjes lopen, met hun mobiele telefoon spelen, hun lippen stiften, ballen, op het bord schrijven of consumpties halen. Geconfronteerd met die observaties, verklaart Lex dat met een verwijzing naar de schoolcultuur:

> In het algemeen is het zo, een kenmerk van leerlingen in het algemeen of speciaal van onze leerlingen, maar … en daarmee wil ik dus niet zeggen dat ze wat lopen te luiwammesen, dat nadrukkelijk niet, maar onze leerlingen zijn wel eens niet goed in het zonder de hand van de

> docent zelf verder gaan. (…) Het initiatief om het zelf te doen, een knoop door te hakken, dat wordt niet snel genomen. Het zijn op-de-docent-wachters hier. [1][31]

In de eindevaluatie herkende Bea het ook, met name in de expliciete onderwijsleerstrategie. Zij schrijft het echter niet zozeer toe aan de schoolcultuur, maar aan de onzekerheid van de leerlingen:

> Als zij een reflectieopdracht af hebben en ze moeten die met mij bespreken en ze wachten er toch mee voordat ze met hun onderzoek verder gaan en daarnaast lopen ze in bepaalde aspecten vast, dus ze willen jou ook spreken omdat ze het niet weten of ze nu op de goede weg zijn en je bent alleen maar reflectieopdrachten aan het nabespreken met leerlingen, en ze zitten echt te wachten tot ze aan de beurt zijn en het duurt een half uur, dan zijn ze ook dat half uur niet met een leerproces bezig. En ze kunnen voor mijn gevoel ook niet met een leerproces bezig zijn, want ze willen een soort fiat hebben dat ze weer een stap verder mogen zetten of kunnen zetten of eh … al is het alleen maar van dat ik het heb gezien en niet volledig heb afgekraakt en eh, als ik zeg van 'Nou ja, jij bent op de goede weg', 'Oké,bedankt', eigenlijk heb ik ze dan niks gegeven, maar dan gaan ze toch weer verder en dat is … daar zit hem dan de beg…; het is niet zozeer de onderzoekstijd, het is de begeleiding …
> (…)



Schaal	*Totaal (N =46)*	*AS (N=20)*	*SA (N=26)*
Moeilijk - makkelijk	2.6 (0,6)	2.5 (0,5)	2.7 (0,7)
Nuttig - overbodig	2.5 (0,9)	2.5 (0,8)	2.5 (1,0)
Te lang duren - te weinig tijd	2.9 (0,9)	2.9 (0,9)	3.0 (0,9)
Motiverend - demotiverend	3.0 (0,8)	2.9 (0,6)	3.0 (0,9)

Figuur 7.5 Evaluatie van het werken aan het onderzoek. Antwoorden van leerlingen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op vier 5-puntsschalen; hoe lager het cijfer, hoe moeilijker, nuttiger, langer, motiverender; links voor alle leerlingen, rechtsboven voor de leerlingen van de Amsterdamse School (AS), rechtsonder voor de leerlingen van de School in Amstelveen (SA).

[31] De herkomst van de citaten is te vinden in bijlage 12.

> Als je zegt: 'Ik wil in ieder geval dat in dit uur je reflectieopdracht is gemaakt en besproken', dan hebben ze toch altijd de neiging om niet verder te gaan tot dat die twee dingen zijn gebeurd. En dan zeg je wel: 'Daarnaast kun je met je eigen onderzoek verder', nou, voor hun gevoel niet dus. [2]

Voornemens van Lex om zijn tijd evenwichtiger te verdelen en de gesprekken korter te houden, lukten echter vaak niet. Hij dook dan zo diep in het gesprek over het onderzoek van een groepje leerlingen dat hij de tijd niet meer in de gaten hield. Die intensiteit waarmee hij de gesprekken voerde, leidde volgens hem er ook toe dat hij niet waarnam wat de andere leerlingen op dat moment aan het doen waren.

Ook Bea verklaarde na afloop van de les waarin ze de hele tijd bezig was geweest met het bespreken van reflectieopdracht 1, dat ze geen idee had waar de leerlingen mee bezig waren geweest in de les (buiten de besprekingen). Zij ervoer de lessen als intensief, vooral de lessen waarin reflectieopdrachten nabesproken moesten worden. Toen de leerlingen eenmaal aan de uitvoering begonnen waren, werd het in haar ogen wat relaxter. Bij beide docenten merkte de onderzoeker een keer na afloop op dat zij het waren die dat lesuur het hardst gewerkt hadden en niet de leerlingen.

Lex realiseerde zich terdege dat de planning voor de leerlingen krap was, hij sprak een aantal keren tegen de onderzoeker uit dat hij zich zorgen maakte of ze niet achter gingen lopen. Aan het begin van elk blokuur spoorde hij zijn leerlingen ook steeds aan en vertelde hen wat ze volgens hem allemaal zouden moeten doen in dat blokuur en in de tijd tot de volgende les. Achteraf was hij van mening dat de leerlingen de lessen niet goed genoeg benut hadden. In zijn opinie hoort het uittypen van hun inleiding eigenlijk niet in de les, de lessen zijn ervoor om de kwaliteit van hun onderzoek te verhogen. Waarmee hij onderling overleg en overleg met hem bedoelde. Hij stond een aantal keer leerlingen dan ook niet toe naar de mediatheek te gaan, maar hield hen in het lokaal.

Gevolg van de uitlopende planning was dat de conceptartikelen – die nodig waren voor reflectieopdracht 4 dan wel door de docent becommentarieerd moesten worden – later klaar waren dan voorzien. Slechts drie van de in totaal 24 duo's deden dat op tijd. De leerlingen uit de impliciete onderwijsleerstrategie leverden hun concepten uiteindelijk wel bijna allemaal bij de docent in, ergens tussen les 9 en 10 (waar op de AS veel tijd tussen zat), zodat ze nog wel commentaar konden krijgen. In Amstelveen werd voor een deel van de leerlingen uit de expliciete groep een apart tijdstip gezocht om alsnog reflectieopdracht 4 te kunnen doen. In Amsterdam vond reflectieopdracht 4 plaats in les 10. Daarna was er in de les geen tijd meer om het concept bij te stellen, dat hebben de leerlingen buiten de les moeten doen. In de impliciete groep op de AS was die tijd er wel, de leerlingen wachtten echter bijna allemaal eerst de gelegenheid af om Lex om commentaar te vragen, en gingen er – op één uitzondering na – daarna in de les evenmin mee aan de slag.

Opvallend was dat Lex, ondanks zijn zorgen over de vertraging, na afloop van les 8+9 van de impliciete groep verklaarde dat het hem honderd procent meeviel waar de leerlingen waren, terwijl slechts anderhalf van de vijf onderzoeken uitgevoerd waren. Die positieve waardering blijkt beïnvloed te zijn door eerdere ervaringen:

> Het gaat oneindig veel gestructureerder dan in het verleden, toen kreeg ik het ene verslag in maart en het andere eind mei. [3]

De vraag lijkt gerechtvaardigd hoe realistisch de planning was en of Lex er eigenlijk zelf wel in geloofde dat het dat was. Met andere woorden of het eerste uitgangspunt zoals verwoord in paragraaf 7.1 geen bijstelling behoeft. In paragraaf 7.6 komt die vraag terug.

Met betrekking tot de planning spraken beide docenten ook uit dat hun voorkeur uitgaat naar de periode tussen kerst- en krokusvakantie (zoals gerealiseerd in Amstelveen) en minder naar de periode tussen krokus- en meivakantie (zoals gerealiseerd in Amsterdam). De eerste periode was weliswaar iets korter dan de tweede, maar de tweede periode werd veel vaker onderbroken door andere schoolbrede activiteiten. Ook de leerlingen in Amsterdam spraken uit dat aan het eind van het jaar de werkbelasting groter is dan halverwege het jaar en dat zo'n grote praktische opdracht wat hen betreft beter eerder in het jaar kan.

Uit de gang van zaken kan de conclusie getrokken worden dat tijd een bepalende factor is in de uitvoering, zowel voor de docenten als de leerlingen. De docenten hebben niet altijd genoeg tijd (genomen) voor de voorbereiding en kwamen tijdens de lessen tijd te kort om alle leerlingen naar hun idee voldoende aandacht te geven en de reflectieopdrachten na te bespreken, vooral in de beginfase van de onderzoeken. De leerlingen hebben meer tijd nodig gehad voor het opzetten van hun onderzoek dan gepland, maar ze gingen nogal inefficiënt met de beschikbare tijd om. Ook bij hen moest het onderzoek plaatsvinden naast andere tijdrovende activiteiten.

7.4.2 De startlessen

Doel van de startlessen was het ontwikkelen van een globaal motief naast het vinden en expliciteren van een gezamenlijk startkapitaal aan kennis, gebaseerd op hetgeen de leerlingen al weten over onderzoek doen, en een start maken met het wetenschappelijke verhaal omtrent het doen van onderzoek. Tevens zouden er enkele praktische zaken aan de orde komen.

Het gedetailleerde scenario (figuur 7.2 en bijlage 6), waarin aan de hand van de onderzoeksvraag 'zijn blonde meisjes dommer dan andere meisjes?' de beoogde criteria voor goed onderzoek aan de orde komen, blijkt goed uitvoerbaar (zie tabel 7.4). De docenten hadden de grote lijn goed voor ogen, lieten de leerlingen meedenken en vatten discussies samen met de algemene criteria. Bea hield ook de tijd goed in de gaten, zodat ze de gelegenheid had de boekjes uit te delen en enige praktische zaken mee te delen (tabel 7.4). Lex schakelde halverwege in tempo terug, omdat hij toen goed op schema lag, waardoor hij aan het eind echter alsnog in tijdnood kwam.

Dat het scenario slechts één mogelijk gewenst verloop van het gesprek schetst, bleek wel duidelijk. Maar hoewel Bea zich dat terdege realiseerde ('Het gaat er niet om of ze IQ noemen, als ze 'dom' maar definiëren'), voelde het toch als een keurslijf ('O, ik mag er niet van afwijken'). Daarbij bleek het lastig om het verhaal zoals opgeschreven (de verwachte antwoorden) te onderscheiden van de onderliggende gedachte:

> Het frame is dan zo eigenlijk strak in je hoofd dat je vergeet dat je dan natuurlijk nog steeds mag schuiven, het onderliggende dat zit er heus wel. [4]

Lex had daar minder moeite mee; hij had bijvoorbeeld veel meer tekst nodig voor de elementen 9 en 14 – in leerlingtekst: de invloed van andere variabelen dan de onderzochte wordt constant gehouden of evenwichtig gespreid – met name voor het tweede deel. Pas na heel veel heen-en-weer gepraat kon hij het criterium formuleren; hij haalde er een onderzoek bij dat de leerlingen het jaar ervoor hebben gedaan naar de invloed van licht op de groei van waterpest. Hieronder het slot van de discussie over andere factoren:

> Docent: Ja, maar dan ben je het toch met me eens dat je, eh … dus met andere woorden, de andere dan de factor die je wil onderzoeken, donker of licht, dat een andere factor, verschillende lengte, dat mag geen rol spelen, tussen die twee. Dus moet je ze even

lang maken. [*Hij verwijst hier naar stukjes waterpest die even lang zouden moeten zijn.*]

Peter[32]: Dus je moet …

Maaike: Ja, dus daarom neem je toch ook alleen maar meisjes, ik bedoel, ja, lengte en weet ik veel wat …

Docent: Maar tussen meisjes kunnen meer verschillen zitten dan alleen blond en donker.

Zander: Maar je moet ook kijken wat voor school ze hebben gehad, welke boeken …

Docent: Toch?

Peter: [*terzijde*] Boeken, ja …

Maaike: Maar ik denk sowieso, dat …

Docent: Dus wat moet je met die andere verschillen doen?

Zander: Beperken.

Maaike: Je kan niet twintig verschillende meisjes neerzetten, want dat kan niet.

Docent: Nee? Oh oké. Oké. Zander zegt beperken. Die andere verschillen, die er tussen die meisjes in de ene groep en in de andere groep zitten, die moet je beperken.

Zander: Maar je kan ze nooit op nul brengen.

Docent: Zoveel mogelijk, liefst honderd procent.

Zander: Ja, dat kan niet.

Maaike: Dan heb je dus sowieso al geen betrouwbaar onderzoek.

Docent: Da's moeilijk. Een andere, er is nog een andere oplossing, die je daarvoor kan bedenken. Hè, want dalijk een eh, een eh … Ja, nou heb ik een heleboel stukjes waterpest van allemaal verschillende lengtes, en je hebt geen schaar.

Zander: Scheuren.

Docent: En je hebt ook geen nagels aan je vingers.

Zander: Oh.

Docent: Dus je hebt een hele berg waterpest. En toch die twee bakken – donker en licht – wat kan je nou nog meer doen, behalve al die stukjes op gelijke lengte?

Zander: Dezelfde lengte zoeken.

Peter: Nou.

Docent: Dat dezelfde lengte. Ja oké, dat is waar.

Peter: Ja!

Docent: Ja, dat dat …

Peter: Een paar lange en een paar korte in de lichte en de donkere gooien.

Zander: Ja!

Peter: Je gooit een paar lange en een paar korte in de donkere en een paar lange en korte in de lichte

Docent: Met andere woorden je verdeelt of verspreidt …

Maaike: … het verschil …

Docent: … de verschillen, ja. Oké, goed. Dus met andere woorden, ofwel je houdt de andere factoren constant, hè geen enkel ander verschil dan alleen die haarkleur van de meisjes. Ofwel je zorgt dat in beide … meisjes uit Australië … zitten; dat in beide groepen meisjes van twee … en meisjes van tien en twintig en tachtig …

Zander: Ja dan is nog steeds m'n vraag wanneer …

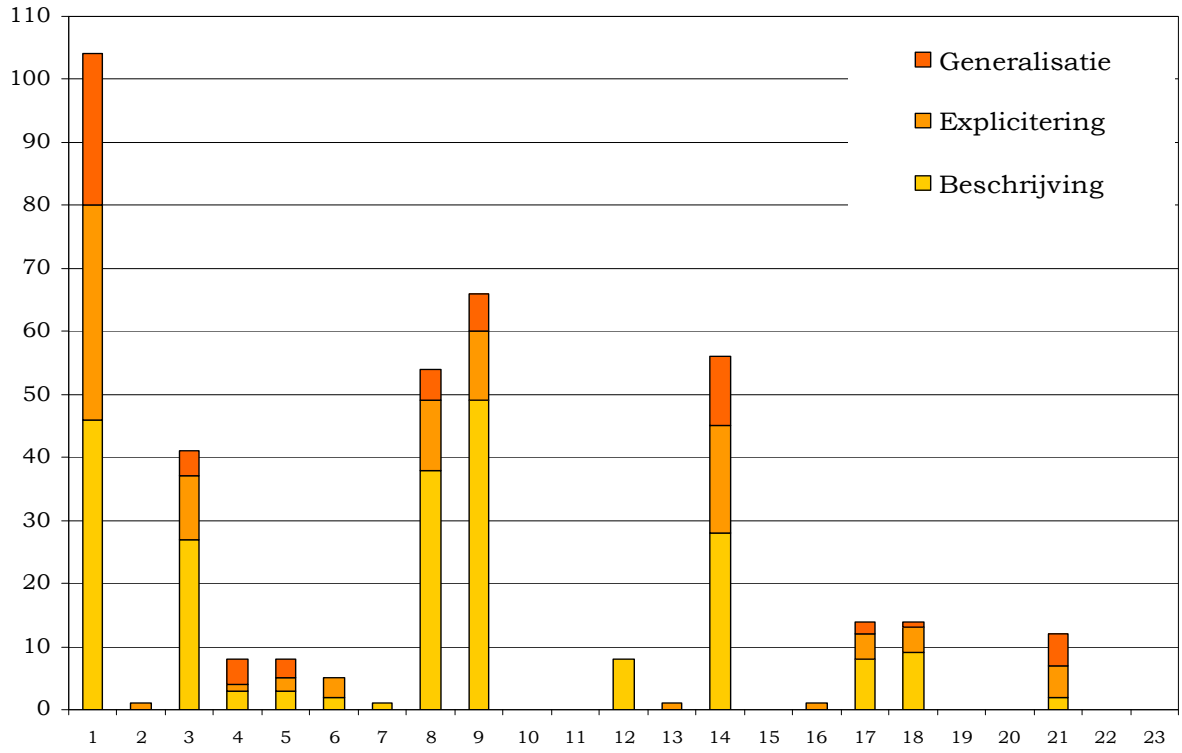
Docent: Of andere … evenwichtig, noemen we dat maar even. [5]

Uiteindelijk lukte het Lex dus wel om het criterium helder te maken. Bij het recapituleren, later in de les, refereerde hij ook nog eens aan dit deel van het gesprek. Zo ging het bij een aantal criteria: de docent werpt het probleem op en aan de hand van reacties van leerlingen gaat hij stapje voor stapje naar het criterium. Op deze manier kreeg dus het *wetenschap-*

[32] De namen van de leerlingen zijn veranderd, met handhaving van het geslacht.

pelijke verhaal van de criteria vorm. Maar niet bij alle criteria lukte het op die manier. Sommige criteria werden sneller door de docent genoemd, zonder een gezamenlijke weg erheen.

In het scenario is voorzien dat de criteria voor goed onderzoek aan de orde zouden komen die ook genoemd worden in het boekje voor de leerlingen (de elementen 1, 3, 4, 5, 6, 8, 9, 14, 17, 18 en 21, zie bladzijde 250). Element 10, over een blanco, is niet in het scenario opgenomen, omdat dat niet past bij de onderzoeksvraag over domme blondjes. De elementen zijn – in dezelfde formulering als in de boekjes – ook op twee overheadsheets gezet die de docenten ter beschikking hadden. Deze sheets fungeerden goed als (tussentijdse) samenvatting van het gesprek; alle criteria passeerden zo (nogmaals) de revue, ook als ze niet in het gesprek aan de orde waren geweest. Figuur 7.6 laat zien welke criteria in de startlessen aan de orde zijn gekomen en of dat in de vorm van een concrete beschrijving (‘wat is dom?’), een explicitering (‘het is niet duidelijk wat je met dom bedoelt’) of een generalisatie ('er mogen geen onduidelijke begrippen in de onderzoeksvraag staan’) gebeurde. Te zien is dat de verschillende elementen van begrip van bewijs niet in gelijke mate aan de orde kwamen; element 1 (over de onderzoeksvraag), element 3 (over hypothese), de elementen 8 en 9 (over variabelen) en element 14 (over steekproeftrekking) kwamen veel vaker ter sprake dan de andere. Dat is niet onlogisch gezien het onderwerp (de onderzoeksvraag ‘Zijn blonde meisjes dommer dan andere meisjes?’). Maar element 18 (over antwoord op de onderzoeksvraag) en 21 (over valide en betrouwbaar onderzoeken) werden relatief weinig genoemd, terwijl de vraag die daarmee samenhangt – ‘heb ik een antwoord en hoe zeker ben ik ervan dat het klopt?’ – wel de centrale vraag van de les moest zijn. De criteria kwamen dus wel over het voetlicht, maar niet geheel in de bedoelde verhouding. De criteria die aan de orde kwamen, zijn ook vrijwel allemaal geëxpliciteerd en gegeneraliseerd. Dat is logisch gezien het scenario en het gebruik van de overheadsheets, maar het illustreert dat het uitvoerbaar is het wetenschappelijke verhaal een aanvang te laten nemen.



Figuur 7.6 Aantal uitspraken (van docent en leerlingen) waarin de verschillende elementen van begrip van bewijs in de startlessen aan de orde zijn geweest in een beschrijving, een explicitering of een generalisatie.

De docenten hadden soms moeite met de inhoud van de criteria; Lex vergiste zich met de assen van een te schetsen diagram en Bea ging de fout in met het begrip statistische significantie. Daarop gewezen zei ze dat, omdat er zoveel informatie in de startles aan de orde moest komen, ze bij dat begrip geen prioriteit had gelegd. Blijkbaar is in de gesprekken over het scenario die van te voren met de docenten gezamenlijk én met de docenten afzonderlijk gevoerd waren, onvoldoende ingegaan op de inhoud van de criteria.

De leerlingen waren in wisselende mate betrokken bij het gesprek, maar de verwachte reacties kwamen wel uit de klas (tabel 7.5). In Amstelveen was de ene groep meer betrokken bij de discussie dan in de andere. Bea zei daar na afloop over dat ze vond dat ze met name in de tweede groep soms de antwoorden eruit moest trekken, maar dat volgens haar de leerlingen wel aan het meedenken waren. Lex was niet ontevreden over de mate waarin de leerlingen meededen, ook al leek het erop dat sommigen soms met hun aandacht elders zaten. 'Het is een positieve groep,' aldus Lex.

Al met al blijkt het scenario voor de startlessen dus redelijk goed uitvoerbaar. In paragraaf 7.5 wordt verder ingegaan op de effectiviteit van de startles in het ontwikkelen van een globaal motief en het leggen van de gemeenschappelijke basis.

7.4.3 Motief als drijvende kracht

De titel van paragraaf 6.1 is 'Meer motief'. In die paragraaf wordt betoogd dat het nodig is om een gemeenschappelijke kennisbehoefte te genereren als drijvende kracht achter het leerproces. De leerlingen zouden voorzien moeten worden van een inhoudelijk motief voor het uitbreiden van hun kennis, in dit geval de kennis met betrekking tot het doen van goed onderzoek. In paragraaf 6.2 is dat uitgewerkt in een didactische structuur (zie figuur 6.1).

De vraag die hier beantwoord moet worden is of het gelukt is: zijn er aanwijzingen te vinden voor een motief om het onderzoek *goed* uit te voeren en zo dus de criteria beter te gaan beheersen? Daarbij moet het verschil tussen motief en motivatie niet uit het oog verloren worden (zie noot 10 op p. 86). Ook al zijn de leerlingen met motivatie met hun onderzoek bezig – ze hebben er zin in, ze vinden het leuk, het houdt ze bezig, ze zijn leergierig – dan betekent dat nog niet dat het inhoudelijke motief met betrekking tot de criteria voor goed onderzoek aanwezig is.

De docenten waren niet ontevreden over de motivatie van hun leerlingen. Ze vonden die hoger dan tijdens andere lessen. De onderzoeker registreerde daarentegen dat leerlingen vaak met andere zaken dan hun onderzoek bezig waren, vooral als ze aan het wachten waren op een mogelijkheid de docent te spreken. In de eindevaluatie scoren de leerlingen vooral in het midden van de 5-puntsschaal motiverend-demotiverend (figuur 7.5), met een gemiddelde score van 3,0 (AS: 2,9; SA: 3,0; expliciet: 3,1; impliciet: 2,9).

In de geregistreerde gesprekken van leerlingen zijn aanwijzingen te vinden dat leerlingen een goed onderzoek wilden afleveren, zoals in het volgende citaat

Johan: Alleen het enige is, ja dan heb je, nee dan, ja dan hebben we wel de proef met nog een proefpersoon. We doen hem drie keer, we doen hem één keer bij blinden, één keer bij zienden en dan moeten we dus goeie blinddoeken hebben bij de zienden met een blinddoek.

Nina: En we moeten absoluut mensen …

Johan: Van die zwarte slaapmaskers of zo.

Nina: En we moeten heel duidelijk zeggen dat de mensen die zien niet echt … hun ogen dichtdoen, maar juist de ogen moeten openhouden. Dat is wel belangrijk.

Johan: Dat is wel belangrijk, ja. Dus niet je ogen dicht doen om je meer te concentreren, want dat is juist niet de bedoeling. Dat is wel moeilijk, want dan moet je mensen dus psychologisch zo ver krijgen dat ze iets doen, maar ze mogen zich niet inspannen om het te doen. Dat is wel moeilijk. We kunnen ze afleiden op de een of andere manier.

Nina: Nee, nee, dat zou ik niet doen; gewoon open.

Johan: Maar je kan natuurlijk ook terwijl je gewoon recht naar voren kijkt, kan je ook afgeleid worden, dus je zal nooit een honderd procent betrouwbaar onderzoek krijgen op die manier. Je zal ze toch echt iets moeten laten zien, denk ik. Ik denk dat je een placebo onderzoek ernaast moet doen. Nina, wat vind je daarvan?

Nina: Waarvan?

Johan: Dat je een nep onderzoekje voor de zienden erbij doet. Dat je ze dingen laat zien of zo: kijk hiernaar terwijl je dingen hoort. Zij weten niet dat daar geen vragen over gevraagd zullen worden. Je doet in feite een placebo onderzoek ernaast.

Nina: Maar hoe doe je dat nou precies? Dan zijn er vier onderzoeken.

Johan: Nee, je zegt 'ga op een stoel zitten.' En dan ga je iets laten horen. Terwijl ze dit doen moeten ze ook naar een paar bladen kijken die je geeft voor… Zij weten niet dat dat geen enkel belang heeft, dat weten ze niet, zij denken dat echt 'oké'. Dat zou toch kunnen, dat je een nep-onderzoek …? Daar hoef je geen resultaten voor te verwerken.

Nina: Dat is waar, maar mag ik nog heel even iets anders zeggen? Maar moeten we niet eigenlijk de mensen met en zonder blinddoek dezelfde mensen laten zijn? Oh, maar dan weten ze natuurlijk al precies wat ze kunnen verwachten. Ja, dat kunnen we op zich wel doen.

Johan: En als je er nou gewoon voor zorgt dat je mensen heb die geen gehoorafwijking hebben? Dan zal je een redelijk betrouwbaar onderzoek krijgen; je zal niet super, maar tsja, het is ook geen onderzoek waarvoor we misschien op de universiteit .. [6]

Slechts één keer werd expliciet fraude voorgesteld, de partner ging er echter niet op in:

Anna: Oké, maar dan zeggen we van oké, één eenheid cafeïne, dus één blikje cola, één kop koffie, één blikje Red Bull.

Maartje: Dat kan wel, want dan is alles gelijk voor iedereen. We kunnen ook gewoon liegen in ons verslag.

Anna: Ja, maar daarbij … als ze het 's ochtends nemen is het echt wel 's middags uitgewerkt, weet je. Het maakt op zich niet uit.

Maartje: Ja, maar we zouden gewoon in ons verslag……

Anna: Dat vragen we gewoon, dan. En dan kijken we of het meevalt. Als iedereen nou één eenheid heeft gehad, dan kunnen we gewoon zeggen alle deelnemers hebben één eenheid gehad. [7]

Waarbij overigens opgemerkt kan worden dat het voor het gemak gelijkschakelen van de hoeveelheid cafeïne in een kopje koffie met die in een blikje Red Bull niet getuigt van zorgvuldig willen werken.

Leerlingen deden moeite om hun onderzoeksvraag netjes te formuleren, om genoeg proefpersonen te krijgen en om variabelen onder controle te houden, ze puzzelden hoe ze eerlijk konden tellen, ze waren teleurgesteld toen verschillen kleiner uitvielen dan verwacht, ze waren trots toen er mooie resultaten uitkwamen en ze vonden het dom van zichzelf toen ze erop gewezen werden dat ze een voor de hand liggend gegeven niet opgezocht hadden. Geen enkele keer is er echter een opmerking geregistreerd in de trant van 'ik wil weten hoe dat zit met die criteria', 'wat is nou precies een blanco'? of 'wanneer zijn resultaten nou betrouwbaar?'

Wel zijn er opmerkingen te vinden die hun eigen inspanningen om het goed te doen relativeren: 'Als het niet goed is, zegt ze dat wel,' en 'Hij zei de vorige keer dat het zo goed was.' Dat lijkt erop te duiden dat de leerlingen het toch vooral als een schooltaak zagen, één van vele en die dus volgens de eisen van zo'n taak uitvoerden. De ene leerling deed dat met meer ambitie en motivatie dan een ander. Dus leerlingen waren wel degelijk tevreden toen ze een onverwacht verband hadden aangetoond tussen rookgedrag van ouders en het beginnen met roken van scholieren en wel degelijk teleurgesteld toen ze moesten constateren dat de emotionele lading van kleuren anders is dan ze dachten.

De conclusie lijkt daarom gerechtvaardigd dat de leerlingen in behoorlijke mate gemotiveerd waren om een goed onderzoek te doen, maar dat er niet eenduidig gesproken kan worden van een inhoudelijk motief, een leerbehoefte om meer te weten te komen van criteria voor goed onderzoek. Wat overigens niet de mogelijkheid wegneemt dat ze daar toch iets over geleerd hebben, maar dat komt in de paragraaf 7.5 aan de orde.

7.4.4 Het beoogde docentgedrag

Het voorgaande onderstreept al het belang van de rol van de docent in de gesprekken, reden om na te gaan of hen duidelijk was wat er van hen verwacht werd. Uitgangspunt is dat de docent een ondersteunende rol heeft, vraagbaak is voor de leerlingen wat betreft de opzet en uitvoering van het onderzoek, maar ook een initiërende rol vervult waar het gaat om het stimuleren van de reflectie op de kwaliteit van het onderzoek.

Behalve dat in beide strategieën zonodig en zo mogelijk de criteria aan de orde gesteld zouden moeten worden, was het de bedoeling het denkproces van de leerlingen te stimuleren door vooral wedervragen te stellen en slechts mondjesmaat zelf suggesties te doen voor onderzoeksvragen of praktische aanpak. De docenten verwachtten dat met name dat laatste hen moeite zou kosten. Leerlingen willen graag aanwijzingen en zelf zijn zij naar eigen zeggen ook geneigd die te geven, onder andere vanwege het feit dat vragende feedback meer tijd kost dan sturende. In tabel 7.4 (p. 119) is te zien dat het gaandeweg de praktische opdracht inderdaad minder lukte, vooral in de lessen in Amstelveen. Lex lukte het beter dan Bea, hij was zich er vaker bewust van dat hij het de leerlingen niet moet voorzeggen:

> Dat vind ik moeilijk. Als ik nog verder ga, ga ik ze toch dingen in de mond leggen en ik wil dat ze met wat ze tot nu tot weten, met behulp van de reflectieopdrachten, toch zelf die …, kijk, als ik zeg 'je moet het zo stellen', dan … en dat vinden ze vet irritant. [8]

Het viel de leerlingen inderdaad op dat de docenten niet zomaar antwoorden gaven, zowel in Amsterdam als in Amstelveen, zoals blijkt uit de volgende twee fragmenten:

> Interviewer: Maar vind je überhaupt dat als je een vraag stelt dat de docent het antwoord moet geven? Of mag hij ook een tegenvraag stellen?
> Peter: Hij mag van mij best een tegenvraag stellen.
> Zander: Maar eerst antwoord geven.
> Peter: Maar hij moet wel, zeg maar, gewoon het antwoord geven. Zeg maar, hij moet – ik bedoel het maakt me niet uit – maar het moet niet tien minuten duren voordat het antwoord komt. Ik bedoel als hij er een uitleg omheen wil doen, ik bedoel best, want dat doet hij altijd, dat weet ik, maar geef dan ook het antwoord en doe het in twee minuten. [9]

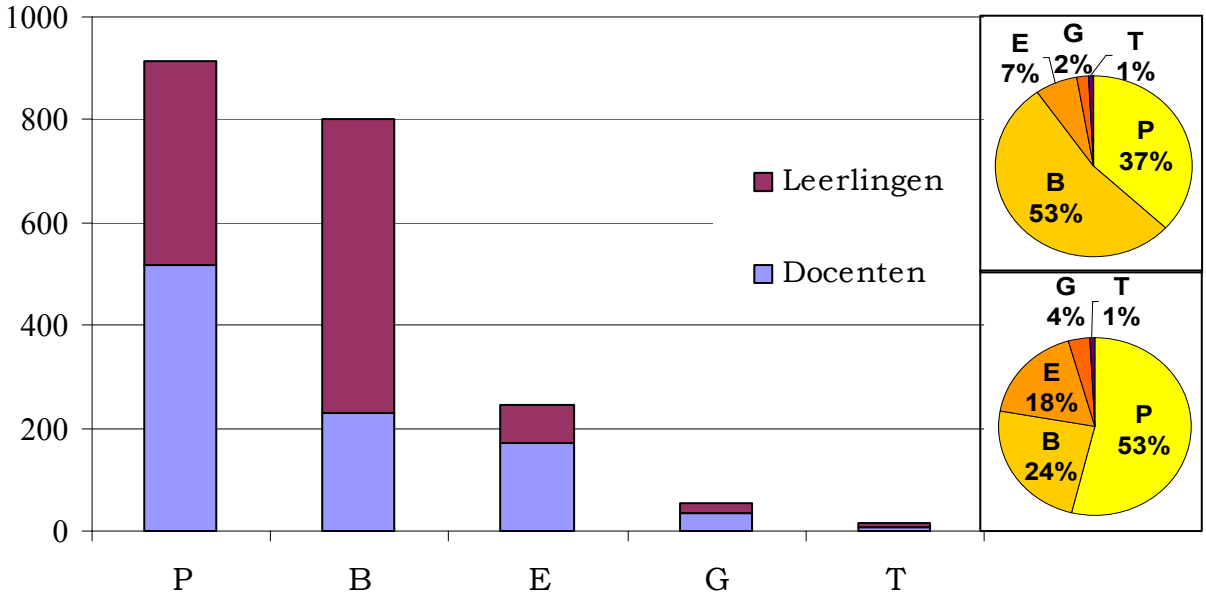
> Kees: Maar ze, ze liet wel veel aan onszelf over. Dan gingen we dus vragen van 'wat lijkt u beter, als we dit doen of dit?', dan zei ze 'ja, dat mag je zelf eh, dat is eh, aan jullie de keus', dus dan ...
> Int.: Jaja, dan moest je d'r zelf over gaan nadenken.
> Kees: Wij hebben eh …

Giorgio: Zo'n psychiater: 'Wat vind jij?'
Kees: Wat vind jij? Dat moet je voor jezelf bedenken!
Int.: Ja, dat zijn jullie niet gewend?
Kees: Ja, nou, de meeste leraren die zijn zo enthousiast, dan ga je 't vragen en dan ja…
Giorgio: … dan vertellen ze de hele opdracht al.
Kees: Heb je natuurlijk wel dát probleem, maar toch zou ik voor dit gaan. Zo kan je dus ook, als je daar slim daarmee omgaat, kan je dus echt heel veel, heel veel d'r uit putten, maar daar wou zij niet aan meedoen, zeg maar. Ik vind het ook wel logisch dat zij het niet doet, als leraar zou ik ook niet iedereen gaan helpen. Maar ja, de meeste leraren doen 't gewoon, ik bedoel, we hebben een economieleraar …
Int.: Jaja.
Kees: en dan, stellen we iets op en dan zo 'ja, maar kun je het niet beter zo doen?' (…). Nou dan maakt ie, zeg maar, je hele onderzoeksopzet voor je en eh, denk jij 'nou mooi'. [10]

Uit deze fragmenten blijkt dat het de docenten gelukt is, meer dan de leerlingen gewend zijn, om tegenvragen te stellen en het denkwerk bij hen te laten. Wat niet wegneemt dat het niet altijd gebeurde.

Ook het scoren van de uitspraken van docenten in de categorieën problematiseren, beschrijven, expliciteren, generaliseren en toepassen laat zien dat de docenten het merendeel van het problematiseren voor hun rekening namen en de leerlingen de meeste beschrijvingen op hun conto schreven (fig. 7.7). Daarmee is verwachting 3a uit paragraaf 6.4 uitgekomen.

Het expliciteren van de criteria voor goed onderzoek, het tweede doel van de begeleiding van de docenten lukte minder vaak dan het stellen van wedervragen (zie tabel 7.4). Lex deed het wel in de expliciete groep, maar lang niet altijd in de impliciete groep. Bea deed het in minder dan de helft van de lessen, waarvan drie keer dan ook nog maar ten dele.



Figuur 7.7 Aantal uitspraken gedaan door de docenten en de leerlingen in gesprekken over het eigen onderzoek en de reflectieopdrachten waarin elementen van begrip van bewijs aan de orde zijn geweest in de vorm van een problematisering (P), een beschrijving (B), een explicitering (E), een generalisatie (G) of een toepassing (T). Rechts de procentuele verdeling van de uitspraken van leerlingen (boven) en docenten (onder).

Uit gedetailleerdere analyse van de gesprekken komt een positiever beeld naar voren (fig. 7.7). De docentuitspraken met betrekking tot de criteria zijn voor meer dan de helft problematiseringen en betreffen bijna evenveel expliciteringen, generalisaties en toepassingen als beschrijvingen.

Toch is te constateren dat er, net zoals in de eerste ronde, mogelijkheden gemist werden om de criteria te expliciteren, bijvoorbeeld in het volgende fragment. De leerlingen willen de werking van tandpasta onderzoeken.

Docent: Ja?
Ineke: We zijn er nu zeg maar vanuit gegaan dat het bacteriedodend is.
Docent: Dat moet je nog wel even checken inderdaad in de literatuur.
Ineke: Want anders kunnen we dit niet doen. Dan hadden we zeg maar 'welke bacteriën worden gedood in de bestanddelen', dan moeten we daar een paar van uitzoeken uit tandpasta. Of 'in hoeverre zijn ze bacterie dodend?' Maar we vroegen ons af van als je dat gaat onderzoeken onder de microscoop of zo, hoe kun je dan zien wat voor bacteriën het zijn?
Docent: Welke variabele heb je? Welke bacteriën er worden gedood, hè?
Ineke: En welke bestanddelen, verschillende bestanddelen uit tandpasta. En verschillende bacteriën.
Bonnie: Want welke bacteriën worden gedood door tandpasta? Toch elke keer dezelfde?
Docent: Dus wat…
Ineke: Misschien wel dezelfde bacteriën.
Docent: Welke aanname doe jij hier?
Ineke: Bacteriën en verschillende soorten bestanddelen.
Bonnie: Nee, kijk als fluor nou niet bacteriëndodend is, misschien is het dan wel chloride, zeg maar.
Docent: Maar wil je dat weten? Wil je…?
Ineke: Wat?
Docent: Nou, welke bestanddelen.
Ineke: Nou ja, anders heb je..
Bonnie: Maar als je zeg maar alleen de vraag hebt 'is tandpasta bacteriedodend'?
Ineke: Of 'welke bacteriën worden gedood door tandpasta?'
Docent: Nou, in dit onderzoeksvraagje, daar zitten eigenlijk twee verschillende onderzoeken in. Namelijk aan de ene kant wil je weten dat, even aangenomen dat tandpasta bacteriën doodt, en heb je in deze onderzoeksvraag zitten 'welke'.
Bonnie: Welke en door wat.
Docent: Ja, eerst dan welke. Dat betekent, dat is al een onderzoek op zich. Want dan ga je kijken van, in eerste instantie, hoeveel verschillende zijn er en dan ga je kijken welke…
Bonnie: Dat hebben wij altijd! Wij hebben altijd …
Docent: Jullie willen altijd even alles onderzoeken. En daarna, ik weet niet hoe je tandpasta wilt gaan … hoe je onderdelen eruit gaat halen, maar daarnaast moet je ook nog weten wélke onderdelen. Ik denk dat…
Ineke: Dat is apart.
Docent: Ja, maar dan krijg je natuurlijk gewoon, dan moet je een ander onderzoek gaan doen, dat is … ja, weet je wat jullie dan doen, jullie zijn dan bezig met twee, drie vervolgonderzoeken, wat je dus eigenlijk doet van… Stel dat het doodt, dan ga je kijken welke doodt het en dan ga je ook nog eens kijken, als je dan aanneemt als hypothese fluor, dan ga je dan weer iets opzetten met alleen maar fluor. Wat je weer uit scheikunde kan … Je moet keuzes maken.

Ineke: Maar kunnen we wel zien dat … onderscheid maken tussen bacteriën, dus als we doen 'welke bacteriën worden gedood door tandpasta?' [11]

De docent laat in het begin wel de term 'variabele' vallen, maar gaat daar niet mee verder, terwijl het heel verhelderend zou zijn de leerlingen de vraag te stellen welke onafhankelijke en afhankelijke variabelen ze nu willen gaan onderzoeken.

Maar soms ging het ook wél goed, zoals in het volgende fragment, waar de docent het criterium 'een hypothese moet toetsbaar zijn' (begrip van bewijs, element 4) aan de orde stelt aan de hand van de eigen onderzoeksvraag van de leerling (of jongens sneller kunnen typen dan meisjes en of dat komt doordat jongens meer tijd achter de computer doorbrengen) met verwijzing naar reflectieopdracht 2, die over gazellen gaat.

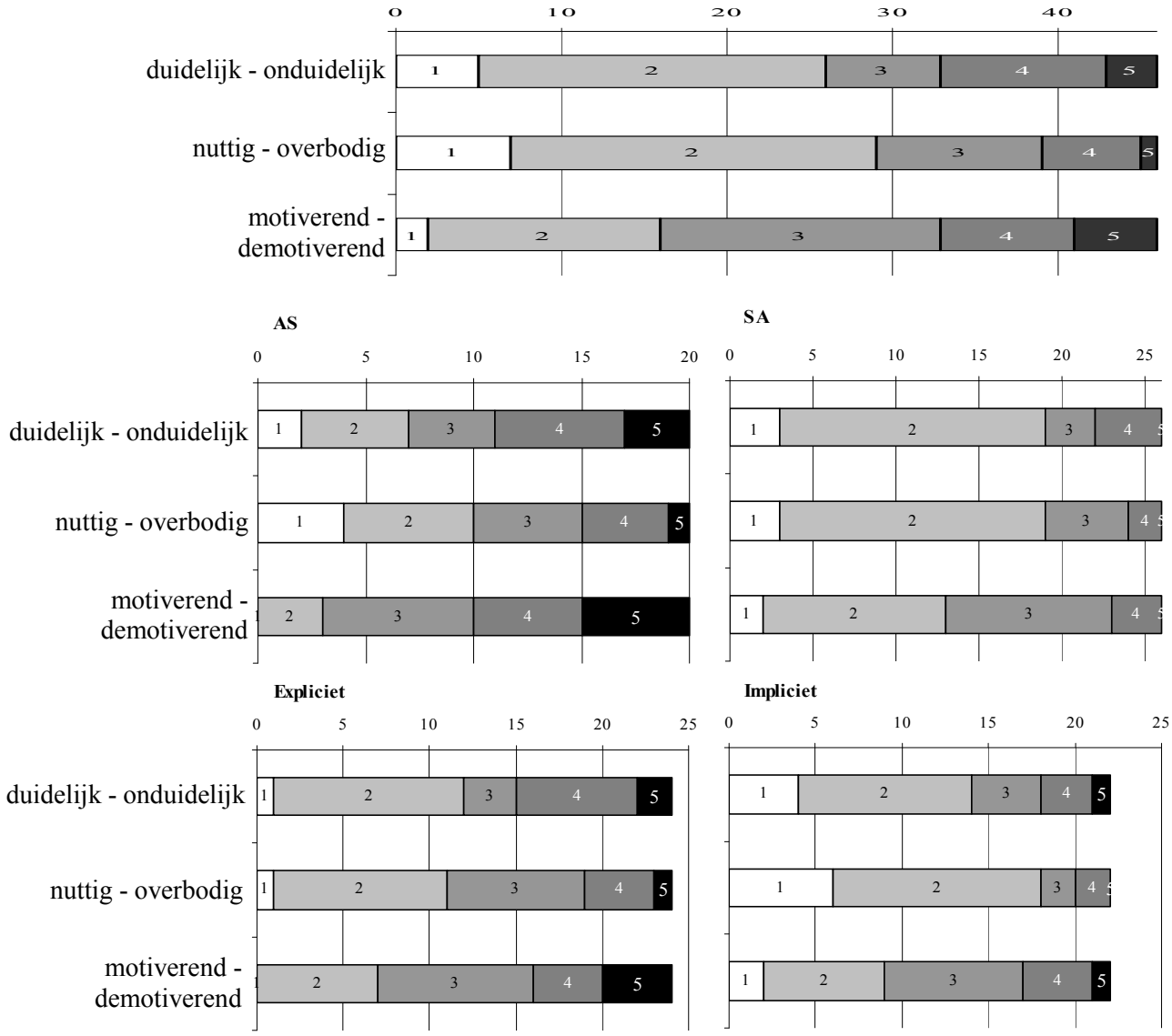
Docent: Mijn vraag, waar ik mee begon, was 'is hij toetsbaar?'.
Theo: Ja.
Docent: Hoe weet je dat?
Theo: Wat bedoelt u precies met toetsbaar?
Docent: Of je de hypothese kan toetsen.
Theo: Dat hij uitvoerbaar is?
Docent: Dat wilde ik inderdaad zeggen … nou, een onderzoek is uitvoerbaar, een hypothese toets je.
Theo: Ja ik denk het wel, want je neemt gewoon de tijd op van hoe snel het allemaal gaat en je laat ze invullen hoeveel uur ze per dag of hoeveel tijd ze per dag achter de computer zitten. En dan ga je daar aan werken en als dat dan, als er dan uit komt dat jongens veel sneller zijn. Je kijkt eerst naar alleen de tijd en dan kijk je daarna nog naar het aantal uur dat ze dan achter de computer zitten. Dus is denk dat hij wel toetsbaar is.
Docent: Ja, want wat maakte in dat voorbeeld van die gazelle, wat maakt die hypothese uiteindelijk toetsbaar? Op welk moment weet die onderzoeker, ik kan die hypotheses toetsen?
Theo: Hier. [*wijst op plaats in reflectieopdracht 2 waar waarneming beschreven staat*]
Docent: Dat was er al na, ditte? Dit is al een waarneming.
Theo: Bedoel je voordat, als hij die hypotheses maakt?
Docent: Ja. Voordat hij in dat veld gaat zitten.
Theo: Hij heeft drie duidelijk, drie duidelijke verschillen, die zeg maar waar te nemen zijn als je dat gaat onderzoeken.
Docent: Ja.
Theo: Dus dan kan je denken dat is toetsbaar.
Docent: Ja, dat is het. Hij heeft, hè, dit zijn dus eigenlijk voorspellingen, die hij hier doet in het geval dat de hypothese juist is.
Theo: Ja.
Docent: Dus dat, op het moment dat je dat hebt – en dat er duidelijk onderscheid in is natuurlijk, want ja – dan weet je in ieder geval dat je hypothese aan het criterium voldoet, die is toetsbaar. [12]

Na het problematiseren en expliciteren van de achterliggende gedachte, het laten beschrijven van de concrete aanpak, wordt het toegepast op het voorbeeld uit de reflectieopdracht en tot slot wordt het criterium in zijn algemeenheid herhaald.

Hieruit blijkt dat het de docenten in ieder geval wel duidelijker was wat er van ze verwacht werd. Desalniettemin is het niet altijd volgens de bedoeling uitgevoerd.

Het oordeel van de leerlingen over de mondelinge feedback van hun docent in de schriftelijke evaluatie na afloop is weergegeven in figuur 7.8. Het blijkt dat meer leerlingen de

feedback duidelijk, nuttig en motiverend vonden dan het tegendeel, een positief oordeel dus. In Amstelveen oordeelden ze veel positiever dan in Amsterdam. Alle extreem negatieve antwoorden ('onduidelijk', 'overbodig', 'demotiverend') werden in Amsterdam gegeven, hoofdzakelijk in de expliciete groep. Waarschijnlijk heeft dat laatste te maken met de manier waarop Lex de reflectieopdrachten nabesprak, waar sommige leerlingen na afloop nogal negatief over oordeelden.



Schaal	*Totaal (N =46)*	*AS (N=20)*	*SA (N=26)*	*Expliciet (N=24)*	*Impliciet (N=22)*
Duidelijk - onduidelijk	2,7 (1,1)	3,2 (1,3)	2,3 (0,9)	2,9 (1,1)	2,4 (1,1)
Nuttig - overbodig	2,4 (1,0)	2,6 (1,2)	2,2 (0,8)	2,8 (0,9)	2,0 (0,9)
Motiverend - demotiverend	3,0 (1,1)	3,6 (1,1)	2,5 (0,8)	3,2 (1,0)	2,7 (1,0)

Figuur 7.8 Evaluatie van de mondelinge feedback van de docent. Antwoorden van leerlingen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op drie 5-punts-schalen duidelijk-onduidelijk, nuttig-overbodig, en motiverend–demotiverend; hoe lager het cijfer, hoe duidelijker, nuttiger, motiverender; boven voor alle leerlingen, midden links voor de leerlingen van de Amsterdamse School (AS), midden rechts voor de leerlingen van de School in Amstelveen (SA), linksonder voor de leerlingen van de expliciete groepen en rechtsonder voor de leerlingen van de impliciete groepen.

Al met al lijkt de conclusie gerechtvaardigd dat de docenten beter dan in de eerste ronde wisten wat er van hen verwacht werd en ze door het stellen van wedervragen en het zonodig expliciteren van criteria voor goed onderzoek dat in de praktijk probeerden te brengen. Door de hectiek van het begeleiden van de verschillende leerling-onderzoeken lukte dat niet altijd optimaal. Het lijkt veel gevraagd van de docenten om zowel adequaat inhoudelijk te reageren op de diversiteit van onderzoeksvoorstellen, als het denken van de leerlingen zelf te stimuleren, als alert alle gelegenheden voor het aan de orde stellen van begrip van bewijs te benutten.

7.4.5 Een éigen onderzoek

Eén van de uitgangspunten van de praktische opdracht is dat leerlingen een eigen onderzoek doen binnen het vak biologie, een eigen onderwerp bedenken, hun eigen onderzoeksvraag formuleren, een onderzoeksopzet maken, gegevens verzamelen en verwerken, de onderzoeksvraag beantwoorden en hun eigen onderzoek evalueren. Dat is uiteindelijk alle leerlingen gelukt, maar er zaten wel de nodige haken en ogen aan.

De planning kwam vanaf het begin onder druk te staan omdat de leerlingen meer tijd dan gepland nodig bleken te hebben voor de start van hun onderzoek: het kiezen van een onderwerp, het formuleren van onderzoeksvraag, hypothese, voorspelling en het verzinnen van een onderzoeksopzet. Dat geldt zowel voor Amsterdam als voor Amstelveen, maar doordat de onderwerpkeuze in Amsterdam minder ingeperkt was dan in Amstelveen, speelde het daar sterker. Bea heeft om praktische redenen de keuze beperkt tot bacteriën/schimmels, enzymwerking, voedingsstofbepaling of gisting; van Lex mocht elk onderwerp binnen de biologie. Die vrijheid werd erg gewaardeerd:

Zander: Ik vond het op zich leuk om te doen. Alleen, ja, van tevoren werd gezegd, ja, je mag een onderzoek doen, je mag zelf weten waarover. En dat is echt heel erg ruim. Je hebt gewoon geen idee waar je het over moet doen. Nou we hadden dan ook wel gelijk door, dat we een heleboel dingen niet wilden doen. Ja, toen kwamen we meer op dit onderzoek.

Int.: Ja, het roken.

Zander: Ja, het is wel erg psychologisch gezocht misschien.

Int.: Ja, jullie wilden niet biologisch, zeg maar met....

Zander: Ja... ja dat zeiden we ook in de inleiding, van we willen niet met bakjes aarde wroeten, met pissebedden.

Int.: O, dat waren jullie ja, dat heb ik ergens gelezen, maar ik wist even niet meer dat....

Zander: Ja. Daar hadden we totaal geen zin in. Krijg je maar vieze handen, nee hoor.

Int.: Nee, maar dat psychologische sprak jullie meer aan?

Zander: Ja, alleen hebben we ook wel gelijk, kwamen we erachter dat het toch wel vrij lastig was om daar conclusies uit te trekken. Omdat het echt heel erg … bij sommige mensen ligt het zo diep, dat je daar niet met een enquêtetje achter kan komen.

Int.: Dus dat is dan wel de beperking van zo'n onderzoek?

Zander: Ja.

Int.: Peter, wat vond jij?

Peter: Nou, ik vond het ook wel leuk om te doen, maar het is inderdaad zo lastig, weet je. Als je dat niet wil, ja dan blijven er een paar dingen over, maar die soms lastiger zijn dan toch. Zeg maar, dan met die bakjes en zo.

Int.: Maar wat had je zeg maar prettiger gevonden? Een afgeperkt gebied, bijvoorbeeld het gaat over planten of het gaat over diergedrag? Of deze grote vrijheid van het mag over alles gaan, als het maar enigszins met biologie te maken heeft?

Peter: Ja, qua keuze was het misschien makkelijker om een onderwerp te geven, maar aan de andere kant vond ik het ook wel leuk, dat je zelf mocht nadenken over wat je

> ging kiezen, of... dus wel je eigen idee zeg maar kan geven. Dat je niet gebonden zit aan iets. [13]

Bij het zoeken naar een onderwerp en een onderzoeksvraag werd weinig gebruik gemaakt van het biologieboek, hoewel beide docenten dat wel suggereerden. Het voorbeeld uit de startles – *Zijn blonde meisjes dommer dan andere meisjes?* – bleek daarentegen wel inspirerend. Zoveel zelfs dat één duo lang bij Bea bleef aandringen om dat toch te mogen doen. Uiteindelijk berustten ze in Bea's weigering. Lex schreef de keuze van bijna alle leerlingen voor een psychologisch getint onderwerp ook toe aan het voorbeeld van de blonde meisjes.

Het zoeken naar achtergrondinformatie bij een onderwerp om daarmee een relevante vraagstelling en een plausibele hypothese te formuleren – wel voorzien in de scenario's en aangegeven in de boekjes – werd in Amsterdam nauwelijks uitgevoerd, in Amstelveen door minder dan de helft (tabel 7.5, p. 120). Slechts enkele leerlingen keken hun biologieboek in om achterliggende theorie op te zoeken. Meestal pas bij het schrijven van het verslag, als onderzoeksvraag en hypothese allang waren bedacht en het onderzoek vaak al was uitgevoerd, werd informatie gezocht om in de inleiding te vermelden. De docenten eisten het nauwelijks bij de bespreking van onderzoeksvraag en hypothese. Soms zeiden ze wel 'zoek maar even op in je boek', maar met name Lex zei een aantal keren op een vraag van leerlingen dat je niet de hele theorie hoefde te vermelden. Toen een leerlinge hem echter in les 9, toen ze bezig was haar artikel te schrijven, vroeg welke achtergrondinformatie nodig was bij haar onderzoek, zei hij daar achteraf over:

> Gek eigenlijk; die achtergrondinformatie zou onmisbaar moeten zijn bij het bedenken van de vraagstelling en de hypothese en de onderzoeksopzet. Dat is grappig, dat komt er bij hen apart bij.
> Dat zijn ze [*hij bedoelt hier zijn leerlingen in het algemeen*] niet gewend. (…) Ze doen weinig met de kennis die ze al hebben opgedaan, dit onderzoek staat heel apart voor ze. [14]

Voor de beide docenten betekende de grote diversiteit van onderwerpen een uitdaging, zowel wat de theoretische inhoud als de praktische uitvoering betrof. Bea verwees voor de praktische zaken regelmatig naar haar technisch onderwijsassistent (toa) en gaf ruiterlijk toe dat ze sommige praktische aspecten niet beheerst. Dat blijkt af en toe ook in haar feedback naar leerlingen, bijvoorbeeld toen ze stelde dat je de groeisnelheid van bacteriën kunt bepalen aan de grootte van kolonies op een voedingsbodem, in plaats van aan het aantal. Voor een aantal onderzoeken werd een dusdanig beroep gedaan op de toa van scheikunde, dat het Bea wrijving met haar collega scheikunde opleverde. Lex verwees eveneens naar de toa voor praktische zaken met betrekking tot pH en dergelijke, maar aangezien de meeste onderzoeken meer psychologisch getint waren, begeleidde hij die geheel zelf. Achteraf zei hij van één onderzoek (naar beweegredenen om Frankrijk als vakantiebestemming te kiezen) dat hij dat had moeten afschieten als zijnde te weinig biologisch.

De moeite met het vinden van een goed onderwerp en de bijbehorende theoretische basis, gevoegd bij de moeite die de docenten hadden om de uiteenlopende onderzoeken te overzien en inhoudelijk te begeleiden, nopen tot een heroverweging van het uitgangspunt dat leerlingen een volledig eigen onderzoek doen. Die heroverweging vindt plaats in paragraaf 7.6.

7.4.6 Het boekje

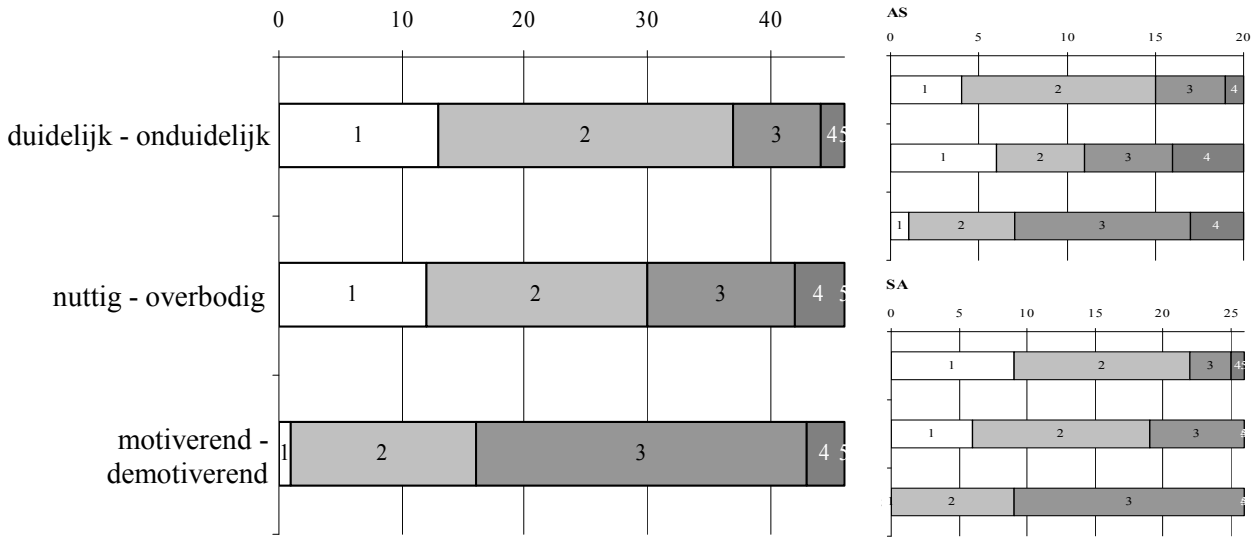
Het boekje dat de leerlingen in les 1 ontvingen van hun docent had in het begin van de praktische opdracht duidelijk een sturende werking. Lex drukte zijn leerlingen stevig op het hart dat zij dit boekje de komende tijd steeds bij zich moesten hebben, dat het een gedetailleerde studiewijzer, hun 'spoorboekje', was. Aan het begin van de eerste blokuren verwees hij

er ook vaak naar, bijvoorbeeld met bladzijdenverwijzing bij het aangeven welke reflectieopdracht dat blokuur gemaakt moest worden. Hij legde daarbij weinig accent op de beoogde heen-en-weer-benadering (zie par. 7.1). De leerlingen hebben de boekjes vooral in de aanloop naar hun onderzoek gebruikt: op de eerste bladzijden zijn duidelijk meer notities te vinden dan op de latere. Als daar al iets staat, dan is dat meestal een verwijzing naar hun verslag c.q. artikel. Slechts enkele leerlingen noteerden hun meetresultaten in hun boekje, de meesten gebruikten daar kladblaadjes voor. Twee boekjes per duo vonden de meeste leerlingen ook overbodig; alleen in het begin schreven ze allebei in hun boekje, al gauw werd bij de meeste duo's slechts één boekje gebruikt. Het kwam regelmatig voor dat leerlingen hun boekje niet bij zich hadden in de les of dat het exemplaar met de aantekeningen in het kluisje van de afwezige onderzoekspartner lag.

Toch oordeelden de leerlingen in de schriftelijke evaluatie opvallend positief over de structurering in het boekje (figuur 7.9). Ze vonden het duidelijk en nuttig (al is dat in Amsterdam minder dan in Amstelveen) en meer motiverend dan demotiverend.

Behalve ondersteuning voor de verschillende stappen in het onderzoek, bevatte het boekje een planning, de eisen voor het artikel en een voorbeeldartikel. Gevraagd naar de redenen dat de leerlingen zich niet aan die planning hielden, merkte Lex op:

> De planning in het boekje is niet hún planning en bovendien kijken ze niet in het boekje. [15]



Schaal	*Totaal (N =46)*	*AS (N=20)*	*SA (N=26)*
Duidelijk - onduidelijk	2,0 (0,8)	2,1 (0,8)	1,9 (0,8)
Nuttig - overbodig	2,2 (0,9)	2,3 (1,1)	2,0 (0,7)
Motiverend - demotiverend	2,7 (0,6)	2,7 (0,7)	2,7 (0,5)

Figuur 7.9 Evaluatie van de structurering van het boekje (stappen in het onderzoek). Antwoorden van leerlingen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op 5-puntsschalen duidelijk-onduidelijk, nuttig-overbodig en motiverend-demotiverend; hoe lager het cijfer, hoe duidelijker, nuttiger, motiverender; links alle leerlingen, rechtsboven de leerlingen van de Amsterdamse School, rechtsonder de leerlingen van de School in Amstelveen.

Dat er in het boekje een voorbeeldartikel te vinden was, is ook niet tot elke leerling doorgedrongen, zo bleek tenminste uit opmerkingen van leerlingen, die elkaar er nog tegen het eind van de praktische opdracht op wezen. Het voorbeeldartikel was niet voor alle leerlingen een steun, sommige vonden het te ver afstaan van wat er – volgens hen – van ze verwacht werd:

Ineke: Maar ik vond ja, en dit was zeg maar we, wij eh, dit was echt een heel ander onderzoek. Want, zoals dit [*ze wijst in het boekje*], weet je wel herkomst en zo, dat konden wij allemaal niet gebruiken dus we zaten echt zo van ja wat moeten we daar nou mee. Van eh …

Bonnie: En misschien dat er eerder een beetje onderzoek …

Ineke: … hoe kunnen wij dat vervangen met onze eigen onderwerpen?

Bonnie: Ja, precies. Want we hadden bacteriën en we hadden echt geen idee hoe we, ja eigenlijk, het hiermee konden vergelijken. En heel veel mensen hadden toch niet zoiets, maar meer iets met eh bacteriën of enzymen en zo.

Ineke: Ja.

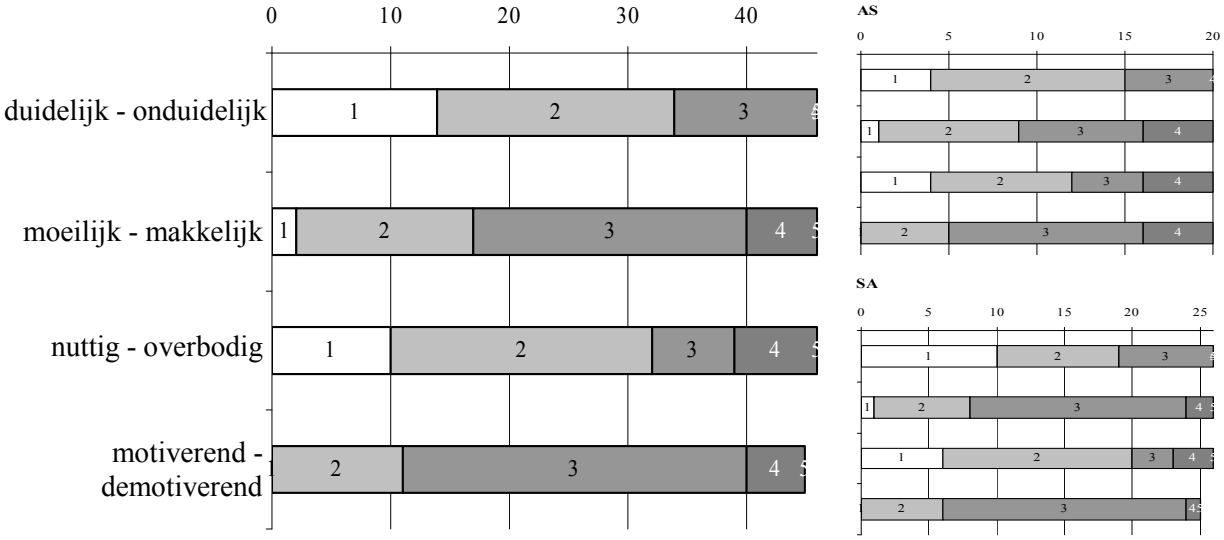
Int.: Dus het gaf te weinig voorbeeld hoe je eigen artikel eruit zou moeten zien?

Ineke: Ja.

Bonnie: Omdat was zo'n compleet ander onderwerp was.

Ineke: Ja.

Bonnie: Misschien dat er iets meer in de richting van eh…



Schaal	*Totaal (N =46)*	*AS (N=20)*	*SA (N=26)*
Duidelijk - onduidelijk	2,0 (0,8)	2,0 (0,7)	1,9 (0,8)
Moeilijk - makkelijk	2,7 (0,7)	2,7 (0,8)	2,7 (0,7)
Nuttig - overbodig	2,2 (1,0)	2,4 (1,0)	2,1 (0,9)
Motiverend - demotiverend	2,9 (0,6)	3,0 (0,7)	2,8 (0,5)

Figuur 7.10 Evaluatie van de informatie in het boekje (inleiding, eisen aan rapportage, voorbeeldartikel). Antwoorden van leerlingen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op 5-puntsschalen duidelijk-onduidelijk, makkelijk-moeilijk, nuttig-overbodig en motiverend-demotiverend; hoe lager het cijfer, hoe duidelijker, moeilijker, nuttiger, motiverender; links voor alle leerlingen, rechtsboven voor de leerlingen van de Amsterdamse School, rechtsonder voor de leerlingen van de School in Amstelveen.

Int.: Ja, bacteriën.
Ineke: Het is ook een heel groot, lang en zo, weet je wel, dus het is een heel andere opzet. [16]

Maar ook wat dit aspect van het boekje betreft oordelen de leerlingen positief bij de evaluatie (figuur 7.10). Duidelijk, nuttig, meer makkelijk dan moeilijk en even vaak motiverend als demotiverend.

Ten opzichte van de situatie in de eerste ronde is er duidelijk een structurerende werking van de boekjes uitgegaan en de leerlingen oordelen positief, maar toch moet de conclusie luiden dat de boekjes niet zo ondersteunend gefunctioneerd hebben als bedoeld en de leerlingen de ondersteunende werking van de boekjes niet ten volle hebben benut.

7.4.7 Een artikel als eindproduct

Nagenoeg alle leerlingen leverden aan het eind van de praktische opdracht een artikel in, zij het dat in Amsterdam de oorspronkelijke einddatum verschoven werd vanwege de vertraging die leerlingen opgelopen hadden (tabel 7.5). Ook het concept daarvan werd door de meeste leerlingen later dan gepland ingeleverd voor commentaar van de docent (impliciete groep) of meegenomen om commentaar van medeleerlingen te krijgen (expliciete groep, reflectieopdracht 4).

De gedachte achter het laten schrijven van een artikel in plaats van een verslag, komt voort uit de nogal losse stijl die in de eerste ronde door de leerlingen gebruikt werd, met name de vele irrelevante opmerkingen en de te weinig doelgerichte opbouw. Dit is met de docenten samen besproken en besloten en Bea communiceerde dat ook zo naar de leerlingen. Lex daarentegen bagatelliseerde het verschil: ‘Wij noemen dat een verslag, hier wordt dat een artikel genoemd’. Hij noemde het dan ook meestal ‘verslag’, evenals veel leerlingen; ook in Amstelveen deden de meeste leerlingen dat trouwens.

Deze houding van de docent wordt weerspiegeld in de geleverde producten, maar niet alleen de docent had er invloed op, zo lijkt het althans. In Amsterdam zijn de artikelen / verslagen beduidend korter, gemiddeld een kleine 1300 woorden tegen gemiddeld ruim 2300 in Amstelveen, terwijl ± 2500 de voorgeschreven omvang was (tabel 7.6). De opbouw van de producten in Amstelveen volgt op een enkele uitzondering na het voorgeschreven stramien (inleiding – materiaal en methoden – resultaten – conclusie en discussie), terwijl de leerlingen in Amsterdam bijna allemaal de opbouw van het boekje volgen (inleiding – onderzoeksvraag – hypothese – onderzoeksopzet – resultaten – conclusie – discussie). De Amsterdamse leerlingen hebben dus duidelijk te weinig gekeken naar de eisen die aan het product gesteld werden, de docent heeft hier ook weinig op gewezen.

In Amstelveen zijn alle producten ingeleverd in een snelhechter, met een kleurig geïllustreerd voorblad, waarop vaak naast de titel en de namen van de leerlingen ook de naam van de begeleidend docent en de datum vermeld zijn. Sommige zijn voorzien van een inhoudsopgave, slechts enkele hebben een soort artikelkop op de tweede bladzij. In Amsterdam voorzagen de meeste duo’s hun verslag ook van een voorblad, al dan niet gekleurd, maar iedereen vond een nietje erdoor voldoende om de bladzijden bijeen te houden. Slechts één van de tien duo’s leverde een artikel in zonder voorblad en met een echte artikelkop. Opmerkelijk genoeg zijn er bij de conceptverslagen meer die zo’n kop hadden.

De beoogde invloed op de stijl van schrijven is waarneembaar in de producten. De helft van de verslagen (11) bevat geen irrelevante procesinformatie, een kwart (6) in beperkte vorm of mate (zoals het vermelden van de data en tijden van het experiment) en een kwart (6) bevat wel opmerkingen zoals ‘Aangezien wij biologie in ons pakket hebben, zijn wij verplicht dit jaar een eigen onderzoek uit te voeren.’ (tabel 7.6). De meeste verslagen beginnen met een meer of minder gedegen inhoudelijke inleiding op het onderwerp. In Amsterdam hebben

Tabel 7.6 Kenmerken van de eindproducten (N = 23); AS: Amsterdamse School; SA: School in Amstelveen; *: met uitzondering van verwijzingen naar de docent, toa of het begeleidende boekje.

groep	*aantal producten in artikelvorm (ja/nee)*	*aantal woorden (gemiddelde, minimum en maximum)*	*aantal producten zonder/ met een beetje / met veel irrelevante proces-opmerkingen* (- / ± / +)	*aantal literatuur-verwijzingen (gemiddelde)**
AS	1 / 9	± 1290 (400 – 1600)	5 / 2 / 3	0,6
SA	0 / 13	± 2315 (600 – 4800)	6 / 4 / 3	3,1
expliciet	1 / 10	± 1875 (400 – 4800)	3 / 5 / 3	2,8
impliciet	0 / 12	± 1940 (1300 – 2600)	8 / 1 / 3	1,3

slechts twee van de tien verslagen een literatuurlijstje, in Amstelveen zijn er slechts vier van de dertien die het níét hebben.

In de geschreven commentaren van de docenten op de concepten van de leerlingen van de impliciete groep zijn zeker opmerkingen te vinden over de stijl van schrijven en de indeling, al zijn die niet overal gemaakt waar het mogelijk was. Bea schreef er een paar keer bij 'meer to-the-point' en 'is dit relevant?'; Lex een enkele keer 'het mag wat zakelijker, dit is geen dagboek!' Hij leverde echter nergens commentaar op kopjes als 'hypothese' en 'voorspelling', die volgens de eisen toch in de inleiding verwerkt horen te zijn. In de herschreven verslagen is door de leerlingen steeds iets met het commentaar van de docent gedaan. Vandaar dat in de uiteindelijke verslagen van de impliciete groep ook veel minder irrelevante opmerkingen meer worden aangetroffen dan in die van de expliciete groep. In de expliciete groep hebben de duo's elkaars concepten van gesproken commentaar voorzien, maar daar ging het nooit over vorm of stijl.

Al met al lijkt een artikel als eindproduct uitvoerbaar, ook al is het niet volledig uit de verf gekomen. Het voorbeeld moet wel dichter bij de te verwachten producten gezocht worden en duidelijker onder de aandacht gebracht worden. Nadruk van de docent op de vormeisen die aan het product gesteld worden, lijkt te werken, dus nog wat meer nadruk kan het resultaat wellicht verbeteren. Ook de mate waarin gegevens uit andere bronnen gezocht en verwerkt worden, lijkt door de docent te beïnvloeden te zijn.

7.4.8 De reflectieopdrachten

De Momenten van Expliciete Reflectie uit de eerste cyclus zijn vervangen door kortere, flexibeler in te zetten reflectieopdrachten. Geen centrale momenten meer die door de hele groep tegelijk gemaakt en besproken worden, maar opdrachten die decentraal te maken en te bespreken zijn op het moment dat ze voor het eigen onderzoek van de leerlingen het meest relevant zijn. Dit zou met name de toepassing bevorderen van de criteria die in de opdrachten aan de orde gesteld worden.

De leerlingen zouden de opdrachten moeten uitvoeren op momenten die bij de voortgang van hun eigen onderzoek paste. Die voortgang was echter ook wel enigszins gepland in de planning in de boekjes. En hoewel de boekjes niet als spoorboekje, maar als ondersteuning bedoeld waren, werkte de planning toch sturend. Aan het begin van de les zeiden de docenten namelijk wat er die les van de leerlingen verwacht werd, bijvoorbeeld:

> Goed mensen, luister even goed. We hebben vandaag een blokuur en dat blokuur gaan we heel goed besteden, hè zodat je na dit blokuur, eh, ja eigenlijk moet je zien dat je na dit blokuur een

flink eind op weg bent. Realiseer je ook dat dit de tweede van de zes lessen is die je aan je eigen onderzoek besteedt enne, je ziet hoe hard het gaat. Oké, eh, als het goed is, en dat hoop ik, heb je een onderzoeksvraag bedacht, voor jezelf, misschien wel meer dan één, je twijfelt nog. Hè, dat mag, is zelfs wel goed. Kan ook nog zijn, dat je d'r nog geen één, dat je nog geen knoop hebt doorgehakt, dat je d'r nog niet toe gekomen bent en dat het noodzakelijk is, dat je die eerst formuleert. Hè, hij mag best een beetje rammelen, dat is niet erg, maar d'r moet wel iets op papier, een onderzoeksvraag. Oké. Dan heb je hopelijk ook al gezien dat in het groene boekje op bladzijde 10 een oefening staat. Die je gedaan hebt, of moet doen of gaat doen of, in ieder geval, eh, op heel korte termijn, dus dit eerste uur al van eh, van deze les. Die je helpt om die onderzoeksvraag, bij te schaven, te verbeteren, misschien wel eh, te verwerpen. Eh, omdat je er hopelijk achter komt, of je onderzoeksvraag aan die lijst van criteria voldoet of niet. Nou, die criteria hebben we in die eerste les al genoemd. In die oefening, worden ze weer gevraagd eigenlijk. Kijken of je ze, eh kan toepassen op een aantal onderzoeksvragen die daar genoemd staan. In de tussentijd, kom ik wel een keer langs bij je. Je hoeft niet bij mij te komen, maar ik kom bij jou langs. Enne, ik kijk hoe je dat er van hebt afgebracht. Eh, van die oefening. Vervolgens, ga je die criteria natuurlijk toepassen op je eigen onderzoeksvraag, want daar gaat het om. Dus, je onderzoeksvraag moet aan die criteria voldoen, honderd procent, na vandaag. Dan kom ik nog wel een keer langs, om te kijken of dat lukt. Hè of je die criteria kan toepassen op je eigen onderzoeksvraag, en dan is er eh, vast nog wel wat tijd over. Zeg ik optimistisch, en dan werk je gewoon verder, eigenlijk, met dat boekje … de hypothese die we moeten gaan formuleren, misschien heb je dat ook wel gedaan. De proefopzet, hoe die d'r uit zou moeten zien, voorspelling van een resultaat, hoe dat zou zijn. Nou, als je helemaal zo ver komt dan heb je vandaag eh, vrij veel gedaan. Dus, ga aan de slag, controleer of je zelf een onderzoeksvraag hebt geformuleerd inmiddels, niet dat je hem in je achterhoofd hebt zitten, maar op papier. En maak als je dat nog niet gedaan hebt die oefening van bladzijde 10. Simpel, hè? [17]

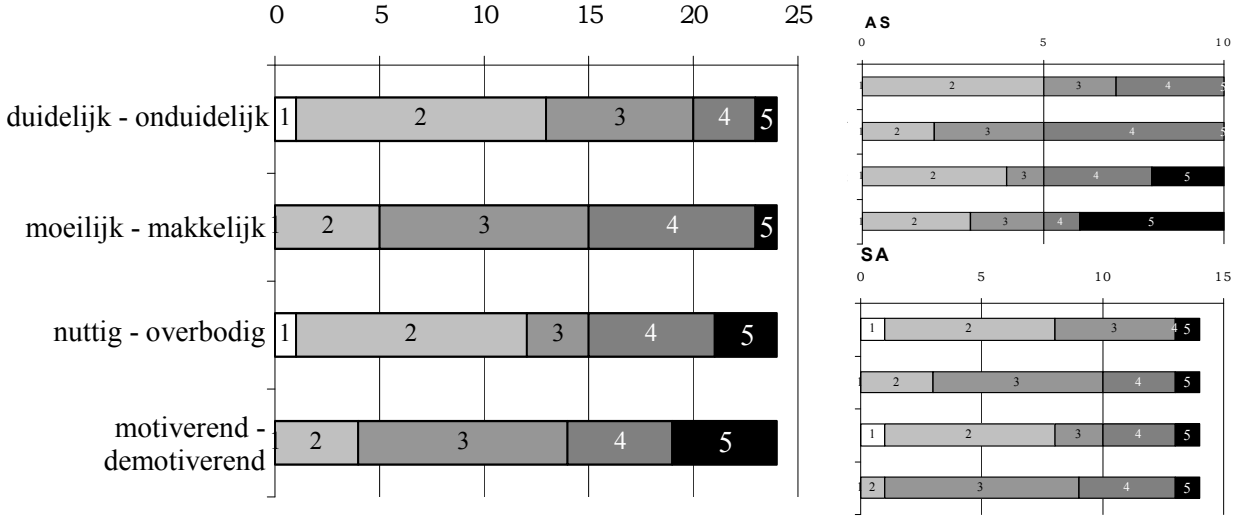
In dit citaat wordt wel de koppeling gemaakt tussen de reflectieopdracht en de eigen onderzoeksvraag, maar tegelijk wordt gesteld dat die opdracht wél die dag gedaan moet worden. Het blijkt ook zo bij de leerlingen te werken. Als ze even niet verder konden met hun eigen onderzoek, bijvoorbeeld omdat ze even niet bij de docent terecht konden of er geen materialen voorhanden waren, zeiden ze tegen elkaar: 'Zullen we dan nu die opdracht even doen?' In de praktijk werd het dus soms een opdracht náást het eigen onderzoek, in plaats van ín het eigen onderzoek. Weliswaar wordt er in de opdrachten expliciet naar de toepassing op het eigen onderzoek gevraagd, maar het maken van de opdracht werd niet altijd als 'check' in het eigen onderzoek ingebouwd. Niet altijd, maar soms dus wel, zoals in fragment 12 te zien was, waar de reflectieopdracht gebruikt werd om over het eigen onderzoek na te denken. De koppeling werd daar echter door de docent gemaakt en niet door de opdracht zelf.

Gevolg van deze instructie was wel dat alle leerlingen vrijwel alle opdrachten daadwerkelijk maakten en nabespraken met hun docent. In die nabesprekingen kwamen de beoogde criteria ook daadwerkelijk op tafel.

De meeste leerlingen waardeerden de opdrachten niet uitgesproken positief of negatief (figuur 7.11). Ze vonden ze behoorlijk duidelijk, aan de moeilijke kant, even nuttig als overbodig en minder vaak motiverend dan demotiverend. Uit de interviews achteraf bleek dat de vierde opdracht het meest gewaardeerd werd: direct commentaar van medeleerlingen op je eigen onderzoek werd als zeer zinvol ervaren.

De leerlingen in Amsterdam waren negatiever in hun oordeel dan die in Amstelveen, met name waar het het nut en het motiverende aspect betrof. Sommige leerlingen waren erg uitgesproken in hun mening en één keer leidde dat zelfs tot een heftige confrontatie tussen leerling en docent:

Docent: Maartje en Anna, hebben jullie de oefening gedaan, die ik gevraagd heb?
Anna: Welke oefening?
Docent: Dat onderzoek naar die cafeïne?
Anna: Ja, volgens mij wel.
Docent: Heb je die gedaan?
Anna: Waar staat die? [*bladert*]
Docent: Ja.
Anna: We hebben wel gelezen.
Docent: Gelezen, zou je hem dan even kunnen maken, kom ik over een half uur.
Anna: Nou, ik vind het zo stom, die andere groep hoeft het niet te doen en het scheelt alleen maar tijd en we hebben er echt geen ass aan …
Docent: Nou dat …
Anna: … dus je kan het nu ook al aan die meneer vertellen, je hebt niks aan die reflectieopdrachten. Die hebben we allemaal in de tweede en de derde … en bij ANW hebben we alles al geleerd.
Docent: Nou hoe weet je zo zeker dat je er niks aan hebt.?
Anna: Nou omdat ik al die andere opdrachten al heb gedaan, ik vind ze heel stom, het kost alleen maar tijd. Er zitten, elke les dat we hier hebben gezeten, hebben we alleen maar aan die reflectieopdrachten gezeten. En we komen gewoon niet verder met ons onderzoek. Dat is toch veel belangrijker?
Docent: Mag ik dan vragen of je het tegen je zin even wil doen?



Schaal	*Totaal (N =24)*	*AS (N=10)*	*SA (N=14)*
Duidelijk - onduidelijk	2,6 (0,9)	3,0 (0,9)	2,3 (0,9)
Moeilijk - makkelijk	3,2 (0,8)	3,5 (0,8)	3,0 (0,8)
Nuttig - overbodig	3,0 (1,2)	3,5 (1,2)	2,5 (1,1)
Motiverend - demotiverend	3,5 (1,0)	3,7 (1,3)	3,2 (0,7)

Figuur 7.11 Evaluatie van de reflectieopdrachten. Antwoorden van leerlingen van de expliciete groepen (in absolute aantallen) op 5-puntsschalen duidelijk-onduidelijk, makkelijk-moeilijk, nuttig-overbodig, en motiverend-demotiverend; hoe lager het cijfer, hoe duidelijker, moeilijker, nuttiger, motiverender; links voor alle leerlingen, rechtsboven voor de leerlingen van de Amsterdamse School, rechtsonder voor de leerlingen van de School in Amstelveen.

Anna: euh.
Docent: Ik wil het over een half uur even met jullie bespreken.
Anna: Nou we gaan tegen onze zin die stomme opdrachten maken.
Docent: Ja?
Anna: Oké..
Docent: Oké, dank je wel. Ik bespreek het over een half uurtje even met jullie. [18]

Helaas motiveert de docent niet waarom hij de opdracht wél zinvol vindt, terwijl het nota bene hier om reflectieopdracht 3 gaat, het koffieonderzoek, en Anna en Maartje zelf ook een onderzoek deden naar het effect van cafeïne.

In een interview achteraf bevestigde Anna haar opvatting. Ze zei dat ze voldoende had gehad aan:

> gewoon een boekje (…) met een soort van stappenplan voor een werkstuk, voor een onderzoek en dan bij elke stap gewoon uitgelegd van wat is belangrijk bij het formuleren van je onderzoeksvraag, wat is belangrijk bij het bedenken van je onderzoek en dat soort dingen. [19].

Zij zegt daarmee dat ze liever de criteria gewoon op een rijtje had gehad, zonder ze zelf uit een opdracht te destilleren. Over de vergelijking van het onderzoek uit reflectieopdracht 3 en hun eigen onderzoek, zeiden Maartje en Anna:

Int.: En toen je het gezien had, heeft dat nog invloed gehad op jullie eigen onderzoek?
Anna: Nee, eigenlijk niet.
Maartje: Nee, we zijn toch echt gewoon onze weg gegaan. En we hebben er niet echt….
Anna: We hebben er niet echt op gelet verder, we hebben die opdracht wel gedaan, maar we hebben het wel echt op onze eigen manier gedaan, zoals we al van tevoren hadden bedacht. Maar hadden sowieso, onze test en zo, onze uitvoering hebben we helemaal zelf bedacht. Dat hebben we ook niet van internet of zo. Alleen het idee om het over cafeïne te doen hadden we wel van internet, maar verder onze hele uitvoer en zo hadden we allemaal zelf bedacht. [20]

Eigenlijk vinden ze het dus juist goed van zichzelf dat ze hun eigen weg zijn gegaan en hun onderzoek niet door de reflectieopdracht hebben laten beïnvloeden. Terwijl de bedoeling juist andersom was!

Niet alleen de leerlingen kostte het tijd, ook de docenten waren een flink deel van hun tijd bezig met het nabespreken van de reflectieopdrachten. Bea besteedde een volledig blokuur (80 minuten) aan het nabespreken van reflectieopdracht 1. In overleg met de onderzoeker besloot Bea daarom om reflectieopdracht 2 klassikaal te bespreken aan het begin van het daarop volgende blokuur. Dat kostte toen inderdaad maar 20 minuten, maar de animo van de leerlingen was gering; er werd wel meegedaan, maar niet door iedereen. De derde opdracht besprak ze met twee duo's tegelijk, wat de efficiëntie verhoogde.

Lex besteedde ook veel tijd aan het nabespreken, terwijl hij maar vijf duo's in de expliciete groep had. Reflectieopdracht 2 kostte ook ruim een blokuur, al besteedde hij tussendoor ook tijd aan vragen over het eigen onderzoek van de leerlingen. Net als Bea combineerde hij bij reflectieopdracht 3 vaak de besprekingen van verschillende duo's.

Voor het nabespreken van de reflectieopdrachten zijn door de onderzoeker scenario's geschreven (zie bijlage 7). Essentie daarvan was om te beginnen met het vragen naar de criteria, om daarmee niet de vragen maar de criteria centraal te stellen. Daarbij zouden de docenten het gedetailleerd nabespreken van de opdracht eventueel kunnen overslaan. De docenten volgden die opzet bijna altijd, maar sloegen toch bijna nooit de bespreking van de vragen uit de opdracht zelf over:

Docent: Welke criteria heb je gevonden?
Yves: Goed omschreven begrippen, …
Docent: Mmhm. [*instemmend*]
Yves: … vraag een beetje afgeperkt.
Docent: Mmhm. Beetje?
Yves: Nou, gewoon doen …
Docent: Nou goed een beetje, voor mijn part, best ja.
Yves: … en dan een duidelijke vraag.
Docent: Dat is je derde: 'duidelijk', wat bedoel je dan met duidelijk?
Yves: Nou. Ehm, ja. Nou gewoon dat …
Docent: Als dit en dit, als je aan die twee voldoet, dan is ie als het goed is duidelijk. Denk ik.
Yves: Ja.
Docent: Je schrijft eerst, goed omschreven begrippen. Dus …
Anton: Niet …,
Docent: … bijvoorbeeld in het geval van …
Anton: … het moet niet, moet niet te moeilijk uitvoerbaar zijn.
Docent: Ehm, ja, okee. Ja ja ja ja.
Anton: … financiën, wij kunnen niet bepaalde onderzoeken financieren.
Docent: Precies. Dan denk je aan materiaal dat je nodig hebt en aan de tijd die het duurt, en ook de omvang van je, pfff, van je van je steekproef en zo. Oké, dat is ook waar, hoewel dat op zichzelf niet maakt dat een onderzoeksvraag slecht is. Dan kan het nog een goed geformuleerde onderzoeksvraag zijn, die helaas in de praktijk voor ons niet uitvoerbaar is, dat kan.
Yves: Ja.
Docent: Maar ehm, welke begrippen heb je ontdekt, die niet ehm … goed omschreven zijn? [*en gaat vervolgens op de opgave in*] [21]

Blijkbaar wilden beide docenten toch nagaan of de opdrachten goed gemaakt waren en dat lieten ze niet liggen als de uitkomsten bevredigend waren. Het scenario is daarin dus ook niet duidelijk genoeg.

Beide docenten ervoeren het nabespreken van de opdrachten als intensief. Lex zei in een nabespreking dat hij het vooral zo zwaar vond, omdat hij tegelijkertijd zowel de onderzoeksvragen, hypotheses en proefopzetten van de leerlingen paraat moest hebben, als moest nadenken over welk criterium uit de opdracht daarvoor het meest van belang was, als ervoor moest opletten het ze zelf te laten formuleren, het ze niet in de mond te leggen.

Al zijn vrijwel alle reflectieopdrachten door de leerlingen gedaan en al zijn bij de nabesprekingen de beoogde criteria aan de orde geweest, toch is er in z'n algemeenheid dus nog wel wat aan te merken op de uitvoerbaarheid ervan. Ze functioneren teveel als onderbreking en doordat er niet altijd een relatie met het eigen onderzoek werd gelegd, functioneren ze te weinig als *reflection-on-action.* Ze vragen veel tijd en aandacht van de docenten die daardoor niet aan vragen van leerlingen over het eigen onderzoek besteed kunnen worden, hetgeen frustratie bij de leerlingen oplevert.

In paragraaf 7.5 wordt op de effectiviteit van de reflectieopdrachten ingegaan worden (of ze hun rol in de ontwikkeling van begrip van bewijs naar verwachting vervullen), hieronder wordt meer in detail ingegaan op de uitvoerbaarheid van elk van de opdrachten apart.

- *Reflectieopdracht 1 'Zoek de fout'*

De eerste opdracht (zie figuur 7.3) betreft de criteria voor de onderzoeksvraag. Geconfronteerd met drie 'foute' onderzoeksvragen, hadden de meeste leerlingen geen moeite om de

criteria 'eenduidigheid' en 'afperking' te vinden (meestal op bladzij 5 van het boekje) en aan te wijzen waar er tegen gezondigd werd in de opgave. Soms werd het rudimentair aangegeven met 'wat', 'wie' en 'waar'. Omdat er drie criteria gevraagd werden en er in het boekje maar twee stonden die direct de onderzoeksvraag betroffen, vroeg het derde om eigen inzicht. Sommigen noemden geen derde criterium, sommigen noemden uitvoerbaarheid, anderen het benoemen van de onafhankelijke en afhankelijke variabelen of het controleren van andere factoren.

De terugkoppeling naar de eigen onderzoeksvraag vond zeker plaats, vooral wat de afperking betrof. Veel leerlingen perkten hun doelgroep in ('brugklassers van de School in Amstelveen') of hun onderzoekslocatie, soms zover dat daarmee de onderzoeksvraag teveel details van de werkwijze ging bevatten:

> Op welke van de deurklinken [aan de kant van het lesgebouw] van de lokalen 017, 117, 217 en 317 en van het noordelijk trappenhuis waar veel mensen komen, zijn op dinsdag 10-2-2004 [het 7e lesuur] meer bacteriën te vinden?

Maar niet alle leerlingen vonden dat de opdracht aansloot bij hun behoefte:

> Int.: En je hebt er ook zo'n opdracht over gemaakt. Was dat eh, zinvol om dat te doen?
> Bonnie: Ik vond die opdracht eh, was zeg maar al die voorbeelden toch, was dat?
> Int.: Mhm. Ja.
> Bonnie: Die waren eigenlijk zo slecht zeg maar, dat je d'r weer niet eh, eh, heel veel mee kon, zeg maar, omdat …'
> Int.: Mhm. Het was te duidelijk slecht, je hoefde er te weinig aan te puzzelen?
> Bonnie: Ja, zoiets. Het was, je kon moeilijk herformuleren omdat het toch allemaal al fout was. Huisdieren, ja wat voor huisdieren, dan moest je een totaal andere onderzoeksvraag maken dus, ik denk ja, daar hadden wij niet heel veel aan.
> Ineke: Nee, dan had ik meer gehad aan dat koffieonderzoek.
> Bonnie: Ja.
> Ineke: Maar dat is dan weer verderop hoor, maar, dat was dan toch wat, ja hoe zeg je dat?
> Bonnie: Als het lijkt dat het goed is, en dat je dan moet kijken van oh ja hè, het is toch fout, is dat bij onze onderzoeksvraag dan ook niet zo? Maar, dat nu was het zo overduidelijk dat …
> Int.: Mhm.
> Ineke: … want in eerste instantie leek ons, als we een nieuwe onderzoeksvraag hadden bedacht, leek het ons ook goed maar ja, dan, als we weer wat verder gingen kijken was ie opeens toch weer fout. Terwijl dat …
> Bonnie: … iets minder duidelijk.
> Ineke: dat het in eerste, en bij die domme blondjes bijvoorbeeld heb, ten minste, had ik meteen van dat kan nooit. [22]

Bij de nabespreking boog Bea de criteria steeds om tot eenduidigheid, afperking en specificiteit, terwijl deze drieslag na de eerste ronde was teruggebracht tot een tweeslag, omdat afperking en specificiteit teveel overlapten. Helaas honoreerde ze zo niet de inventiviteit van de leerlingen. Achteraf zei ze dat ze inderdaad nog teveel de drieslag van het jaar ervoor voor ogen had.

Deze reflectieopdracht blijkt goed uitvoerbaar, maar – zoals Bonnie in het citaat hierboven opmerkt – de toepassing op het eigen onderzoek zou waarschijnlijk makkelijker zijn als de onderzoeksvragen in de opdracht dichter bij het type onderzoeksvragen zouden staan zoals de leerlingen die formuleren.

- *Reflectieopdracht 2 'Stel het op de proef'*

De tweede opdracht is een aangepaste versie van MER 2 uit de eerste ronde, het toetsen van hypotheses door voorspellingen met waarnemingen te vergelijken. Veel leerlingen hadden moeite om de redenering die in de opdracht opgebouwd wordt te volgen. Ze waren van mening dat om een voorspelling van de uitkomsten van een onderzoek te doen je moet weten hoe het zit:

Peter: Maar er staat hier een vraag over de hypothese, dus 'als de hypothese van de alarmfunctie juist is, verwacht je dat de gazelle zijn witte achterste naar soortgenoten zal keren?'
(…)
Peter: Maar hoe kan je er nou een verwachting over uitspreken als je er nog geen informatie over hebt? Kijk, die man heeft het zitten observeren.
Docent: Ja?
Peter: Maar wij niet.
Docent: Nee, maar dat klopt, maar hij moet ook nog een keer gaan observeren.
Peter: Ja, dat snap ik ook wel, maar voordat hij een hypothese formuleert, weet hij al waar hij over praat.
Docent: Jij bedoelt; hij heeft vast wel gezien of … welke die doet?
Peter: Ja, hij heeft daar al ideeën over, hij ziet dat, maar wij zien dat niet voor ons en dan wordt een vraag gesteld waar allebei de antwoorden goed zijn.
Docent: Maar ga er maar vanuit dat hij dat ook niet weet. Hij heeft ze wel zien springen, en zit vervolgens thuis en dan denkt hij van; ja, maar hij heeft, heeft niet….
Peter: Ja, maar hoe…
Docent: Gewoon aannemen dat….
Peter: Het gaat om het woord verwachten, hoe kan je een verwachting uitspreken over iets waar je geen idee van hebt, waar het over gaat? [23]

Uit de antwoorden die de leerlingen bij de opdracht in de boekjes geschreven hebben blijkt dat, op enkele leerlingen na, de leerlingen de essentie niet begrepen hebben. Bij de leerlingen uit Amstelveen zijn veel correcties te zien in de boekjes, waarschijnlijk naar aanleiding van de klassikale bespreking. In de boekjes van de Amsterdamse leerlingen staat bij de vraag of het eigen onderzoek aan de criteria met betrekking tot hypotheses voldoet, vaak netjes 'ja', maar er is alle reden daaraan te twijfelen, gezien het weinige begrip dat de leerlingen in de gesprekken toonden. De transcripten van de gesprekken tussen de leerlingen onderling bij het maken van reflectieopdracht 2 ondersteunen die conclusie. De strategie die de meeste leerlingen wilden hanteren om te juistheid van hun hypothese te beoordelen, valt samen te vatten met 'kijken of het zo is', getuige de uitspraak:

Nou, wat verwacht je als ie niet klopt? Ja, dat ie niet klopt, wat moet ik anders verwachten? [24]

Lex had een gesprek van bijna een half uur met Peter en Zander over de opdracht en hun eigen hypothese en voorspelling (waarvan citaat 23 een deel is), waarna Peter aan het eind verzuchtte: 'Ik begrijp jouw antwoord wel, maar dat is geen antwoord op mijn vraag.' Het is Lex al die tijd niet gelukt Peter de essentie van het toetsen van een hypothese duidelijk te maken. In dezelfde les sprak hij er ook met Theo over (citaat 12), waar het wél lukte om het idee over te brengen.

Doordat Bea deze opdracht klassikaal besprak zonder de geëxpliciteerde criteria toe te passen op een of meer van de leerlingonderzoeken, is de terugkoppeling naar het eigen onderzoek aan de leerlingen overgelaten en niet meer besproken met Bea. Ineke en Bonnie hebben naast hun artikel ook de uitwerking van de reflectieopdrachten ingeleverd en uit de beschrij-

ving van reflectieopdracht 2 lijkt te volgen dat ook zij de essentie wel hebben opgepikt (hypotheses verwerpen door verwachting met waarneming te vergelijken), maar wel moeite hadden met de inhoud van de hypotheses uit de opgave.

Deze reflectieopdracht blijkt dus moeilijk uitvoerbaar vanwege de complexe inhoud en de geringe aansluiting met de manier waarop leerlingen gewend zijn met hypotheses om te gaan (Carey, Evans, Honda, Jay & Unger, 1989; De Hullu e.a., 1999; Lawson, 2001).

- *Reflectieopdracht 3 'Alles onder controle?'*

De derde opdracht is niet door alle leerlingen gemaakt. Vooral in Amsterdam stopten de leerlingen er weinig tijd en energie in, getuige citaat 18. Sowieso deden ze het later dan in de planning stond (tabel 7.5), in les 8+9 in plaats van in les 6+7, maar Lex had moeite om ze ertoe te zetten de opdracht na te bespreken. Hijzelf vond het ook vervelend het eigen onderzoek te onderbreken. Het kostte hem moeite zich afwisselend goed te concentreren op de opdracht en de eigen onderzoeken, zo zei hij na afloop. Toen was hij ook van mening dat hij het beter aan het eind van het blokuur centraal had kunnen doen, bij wijze van afsluiting.

De halfslachtige uitvoering van de leerlingen in Amsterdam uitte zich ook in het resultaat; geen enkel duo merkte het tijdseffect in de resultaten op (elke groep in het onderzoek in de reflectieopdracht scoort de tweede dag beter dan de eerste en de derde beter dan de tweede, ongeacht de behandeling). De leerlingen in Amstelveen én alle leerlingen in de eerste ronde hebben dat wel opgemerkt.

In Amstelveen ging het beduidend beter. Vrijwel alle leerlingen hebben de opdracht gemaakt en de besprekingen verliepen volgens plan. Het criterium dat andere variabelen dan de onderzochte onder controle gehouden moeten worden, kwam er – ook in Amsterdam – duidelijk uit, al was dat soms in zeer bedekte vorm: 'onderzoek moet eerlijk zijn voor en na'. Dat conclusies moeten aansluiten bij de onderzoeksvraag en hypothese en dat er een duidelijke blanco hoort te zijn, werd minder vaak genoemd.

Op zich lijkt er met deze opdracht weinig mis, de redenen waarom hij in Amsterdam niet optimaal werd uitgevoerd zijn vooral van externe aard.

- *Reflectieopdracht 4 'Zeker weten?'*

In tegenstelling tot de weerstand die de derde reflectieopdracht soms opriep, schatten de meeste leerlingen de vierde opdracht, commentaar geven op elkaars conceptartikelen, van te voren als zinvol in.

Bea leidde hem perfect in, ze legde de link met de praktijk van de wetenschap en ze gaf aanwijzingen over de manier van feedback geven:

> Docent: Oké, wat in het onderzoek altijd heel gebruikelijk is, is dat je, als je het onderzoek hebt gedaan, en je hebt … die ga je publiceren hè – je resultaten komen altijd in een *Nature* of *Science* eh, of eh, wordt gepubliceerd – en eh, voordat je dat doet wil je wel zeker, ook zeker weten dat dat wat je publiceert en de hele wereld over gaat toch niet met de grond gelijk wordt gemaakt. Trouwens ook, als je het opstuurt, zit daar ook een hele commissie die eerst maar eens gaat kijken van, gaan kijken van, hoe zeker is het wat je hebt gevonden misschien wel een aidsremmer is, maar hoe zeker weet je dat? Dus, wat ze altijd ook doen in eh, met je collega's als je aan het werk bent in het onderzoek, is dat je je artikel laat reviewen. Dat wil zeggen, een ander kijkt ook naar je onderzoek, die verdiept zich daarin en die geeft feedback op, hoe zeker is het dat wat je gevonden hebt inderdaad antwoord is op je vraag. En al die andere criteria. En jullie gaan elkaar van elkaar, eh het onderzoek eh, bekijken zoals je eigenlijk ook het koffieonderzoek hebt bekeken. En gekeken op grond van welke criteria ehm, eh het onderzoek goed is, of nog niet volledig genoeg. En waarom is dat? Met die feedback kun je dan zorgen dat je een goede discussie

schrijft. En in de discussie kan je jezelf, eh, zo indekken dat je zegt van ja ik heb al wel door dat dit misschien …

Ineke: … niet helemaal …

Docent: … maar eh, ik heb wel door wat ik aan het doen was. En ik weet wel dat het misschien niet honderd procent zeker is maar wel zo zeker dat… Ja?

(…)

Docent: Eh, nou is het wel heel belangrijk, kijk je kan wel zeggen van, dit is echt slecht, dit is toch veel te weinig, zo kan je eh, feedback geven. Maar je kan het ook onderbouwen. En het kan ook door middel van vragen stellen. Waarom heb je gekozen voor maar drie petrischalen, waarom heb je gekozen om eh, eh ...

(…)

Docent: Ehm, dus het gaat erom dat je niet keiharde feedback geeft eh, soms.

Maggie: Afkraken.

Docent: Dus niet afkraken inderdaad maar feedback geven, op die manier dat de ander ook denkt van, echt even goed over na kan denken en het kan gebruiken. Ja? [25]

Deze inleiding leek effect te hebben, want de leerlingen gingen serieus lezen en commentaar opschrijven en wisselden dat vervolgens serieus uit. In het geval dat de bedoeling minder duidelijk werd neergezet (later, bij de duo's in Amstelveen die eerst nog niet ver genoeg waren en in Amsterdam), was de uitvoering ook minder diepgaand: er werd meer oppervlakkig naar de concepten gekeken, de criteria werden er niet expliciet bijgehaald.

De vierde opdracht is degene die het meest werd gewaardeerd door de leerlingen, afgaande op de reacties in de klas en het commentaar van de leerlingen in de interviews na afloop. Wel was de waardering uiteraard afhankelijk van de kwaliteit van het commentaar dat de leerlingen van hun medeleerlingen ontvingen. Want sommigen vonden dat ze eigenlijk alleen maar iets te horen hadden gekregen wat ze zelf wel wisten. Anderen beschouwden commentaar blijkbaar toch als kritiek, die hen dan eigenlijk nog wel meeviel. Anna en Maartje schrijven het volgende in hun boekje bij de vraag wat ze van het commentaar leren en wat ze ermee doen:

> We hebben eigenlijk vrij weinig kritiek gehad, hoewel onze klasgenoten vonden dat de resultaten van ons onderzoek niet heel erg geloofwaardig waren. Wel hebben ze ons op nuttige punten kunnen wijzen die wij waren vergeten, zoals de ingrediënten van Golden Power en de mogelijkheid dat onze testpersonen dáchten dat ze meer energie en concentratie hadden na het drankje en daardoor de test sneller maakten.
> Deze punten hebben we in ons verslag verwerkt. [26]

Dus ondanks het zo te horen fundamentele commentaar ('niet geloofwaardig') ervoeren ze de kritiek als mild.

Maar ook als het commentaar van je medeleerlingen tegenvalt, kan deze opdracht nog zinvol zijn, getuige het volgende fragment uit het interview met Bonnie en Ineke:

Int.: En eh, die laatste reflectieopdracht, het bespreken van je conceptartikel? Vond je dat zinvol?

Bonnie: Ja.

Ineke: Ja, zeker weten.

Bonnie: Ja.

Int.: Je kreeg zinvol commentaar van eh, …

Bonnie: Eh, eh.

Ineke: Ja, nou het was een meer vooral eh, eh dat andere onderzoek, daar hadden wij …

Bonnie: Wij gaven commentaar en dat was ook weer, dat het ons weer aan het denken zette over ons eigen onderzoek.

Int.: Oh, op die manier!

Ineke: We hadden echt, ja, ze hadden zeven hypotheses. Maar ze hadden dan een vraag, met ehm, eh …
Bonnie: Eh, welke etage heeft de meeste bacteriën?
Ineke: Ja, en daar hadden ze in hun hypotheses hadden ze allerlei: 'de derde etage omdat…', 'de derde etage omdat …' …
Bonnie [*synchroon*]: De derde etage omdat, maar …
Ineke: Dus toen hadden ze, toen kwamen we weer op die discussie van die hypothese, want toen dachten we van moet dat nou wel en eh. En, ja ook bij een andere, aantal andere eh, onderdelen hadden ze ook percentages en zo en nou, dan kwam je ook weer bij dat koffieonderzoek dus het, hing wel een beetje met elkaar samen eigenlijk.
Int.: Maar vooral door het commentaar geven ging je over je eigen onderzoek nadenken?
Ineke: Ja, wij wel.
Bonnie: Ja, we kregen niet heel erg veel commentaar en, zeg maar qua, ja …
Ineke: Ja we hadden niet echt, heel veel commentaar gekregen waar we echt iets mee konden …
Bonnie: Nee.
Ineke: … dat we dachten van nou, dat moet echt nog worden veranderd. [27]

Deze leerlingen hebben dus ervaren dat het geven van commentaar minstens net zo leerzaam is als het krijgen ervan. Die ervaring is niet bij alle leerlingen opgetreden.

Goed ingeleid en ingekaderd blijkt deze opdracht goed uitvoerbaar voor docent en leerlingen. De kwaliteit van de door de leerlingen gegeven feedback wisselde echter sterk en daarmee de waardering.

7.4.9 De impliciete onderwijsleerstrategie

De reflectieopdrachten zijn het onderscheidende kenmerk tussen de expliciete en de impliciete onderwijsleerstrategie. Over de impliciete strategie, die daarvoor in de plaats twee momenten van geschreven feedback kent, vallen ook enkele algemene opmerkingen te maken met betrekking tot de uitvoering.

De waardering van de leerlingen voor het geschreven commentaar van de docent blijkt groter te zijn dan die voor de reflectieopdrachten (vergelijk figuur 7.11 met 7.12). Leerlingen ervoeren het commentaar meestal als duidelijk, nooit als overbodig en vaker als motiverend dan als demotiverend. Hoewel sommige leerlingen geïrriteerd en niet-begrijpend reageerden toen ze het onder ogen kregen ('wat bedoelt ze nou met …?'), zagen ze er meestal na de toelichting van de docent de waarde wel van in. In de interviews werd dit bevestigd: schriftelijk commentaar is handig, zeker als je er ook nog toelichting bij kan vragen. Leerlingen uit de expliciete groep die gehoord hadden van het schriftelijke commentaar dat leerlingen uit de andere groep hadden gekregen, voelden zich soms benadeeld: 'Ik vind het vals, ik vind het echt vals dat die andere groep gewoon wel eh, gewoon op- en aanmerkingen erbij krijgt.'

Het geven van schriftelijk commentaar geeft de docent de mogelijkheid de elementen van begrip van bewijs expliciet onder de aandacht te brengen en dat eventueel nog mondeling toe te lichten. Toch is dat moeilijk:

Els: [*leest wat de docent heeft opgeschreven*] De afhankelijke variabele? Dat is toch gewoon de bacterie? Sorry hoor, het spijt me zeer. Maar, de hoeveelheid bacteriën hangt toch af van het pH-optimum, dat is toch een afhankelijk, dan is die toch afhankelijk van de pH?
Minke: Volgens mij wel. [*even stil, geeuwt*] We gaan het vragen.
(…)

Els [*tegen docent*] Ik had een vraag, want … mijn afhankelijke variabele, dat is toch gewoon de bacterie, want die bacterie is afhankelijk, of tenminste die voort.. de hoeveelheid …

Docent: Als ik deze lees hè, weet je wat ik dan voor me zie? Als ik, dan zie ik jullie één bacterie isoleren en ergens neerleggen en dan 'plant voort, plant voort!'

Els: Oké.

Minke: Kijk van die bacterie moeten we nog even, want we weten nog niet welke bacteriën we gaan gebruiken want dat moeten we nog even met Alfons [*de toa*] regelen dan, welke bacterie we krijgen.

Docent: Ja, en waar dachten jullie aan?

Els: Een twintig-minuten-voortplant bacterie. Daar had die man het over.

Docent: Oh, oké.

Els: Maar ik weet niet hoe die heet, die bacterie.

Minke: Maar in ieder geval zo'n bacterie.

Docent: Maar wat ga je meten? Want voortplanting meten, wat ….

Els: Nou, hoeveel dat beestje, dat dingetje is vermenigvuldigd.

Docent: Hoe kun je dat meten?

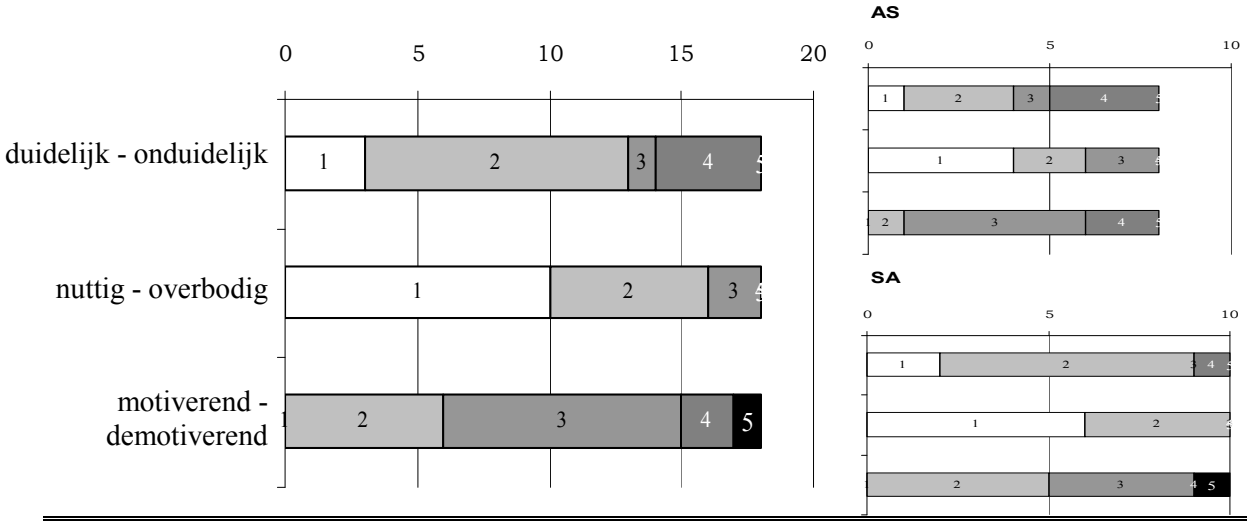
Els: Tellen, hij had het erover, dat je moet, we moesten iets doen. Oh ja, we moesten zo weinig mogelijk bacteriën in het begin nemen zodat het nog telbaar was.

Minke: Ja.

Docent: En wat ga je dan, je gaat zo'n voedingsbodem uitzetten hè?

Minke: Ja.

Docent: En dan krijg je van die kleine klontjes die overal zitten hè? Herken je dat niet, die kleine kolonies? Die moet je tellen. [28]



Schaal	*Totaal (N =18)*	*AS (N=8)*	*SA (N=10)*
Duidelijk - onduidelijk	2,3 (1,0)	2,8 (1,1)	2,0 (0,8)
Nuttig - overbodig	1,6 (0,7)	1,8 (0,8)	1,4 (0,5)
Motiverend - demotiverend	2,9 (0,8)	3,1 (0,6)	2,7 (0,9)

Figuur 7.12 Evaluatie van het geschreven commentaar van de docent. Antwoorden van leerlingen van de impliciete groepen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op 5-puntsschalen duidelijk-onduidelijk, nuttig-overbodig, en motiverend-demotiverend; hoe lager het cijfer, hoe duidelijker, nuttiger, motiverender; links voor alle leerlingen, rechtsboven voor de leerlingen van de Amsterdamse School, rechtsonder voor de leerlingen van de School in Amstelveen.

De docent pakt het criterium 'afhankelijke variabele benoemen' wel op ('wat ga je meten?'), maar expliciteert niet dat dát nu de afhankelijke variabele is.

Vaak betrof het commentaar het concrete niveau van het onderzoek van de leerlingen. Soms stond er 'afperken', maar soms ook 'wat bedoel je hiermee?' bij een onduidelijke term, maar niet 'dat is niet eenduidig'. In het commentaar dat de docenten gaven op de conceptartikelen betroffen de meeste opmerkingen de onderzoeksvraag, de variabelen of de opbouw en stijl van het conceptartikel, meestal op het concrete niveau van het betreffende onderzoek. Twee keer is een opmerking gemaakt over betrouwbaarheid, validiteit of 'zeker weten?'

Het geven van commentaar werd bemoeilijkt doordat niet alle leerlingen hun boekje of conceptartikel op de afgesproken tijdstippen inleverden. De docenten getroostten zich moeite om de leerlingen dan later toch nog van commentaar te voorzien.

Het commentaar kan dus de aanleiding zijn om over de elementen van begrip van bewijs te praten. Maar ook zonder dat lukte het, soms zelfs zodanig dat het sterk ging lijken op de nabespreking van een reflectieopdracht:

Rita: We hebben twee dingen bedacht.
Docent: [*leest wat leerlingen hebben opgeschreven*] 'Is er een samenhang tussen de leeftijd en het functioneren van het korte- en lange-termijn geheugen?' Bedoel je van het korte- en het lange-termijn geheugen?
Rita: Ja, daar zaten we dus net over te denken, …
Meinte: Ja.
Rita: … want als je alleen het korte termijn geheugen neemt is het wel makkelijk uit te voeren, als je het lange neemt dan …
Meinte: Ja inderdaad, dat is iets lastiger.
Docent: Staan er onduidelijke begrippen in deze... maar daar hadden we het vorige week over gehad, hè?
Meinte: Oh, over het algemeen of die erin staan? Ja, leeftijd, moet je zeg maar een bepaalde leeftijd nemen dan.
Rita: Ja, maar tussen leeftijd …
Docent: Ja, maar wacht even, voordat je daar antwoord op geeft, denk eerst even na. Want het is maar een vraag van mij 'staan er onduidelijke begrippen in?'
Rita: Ja, er staat … ja, leeftijd.
Docent: Wat is daar onduidelijk aan?
Rita: Nou, aangezien je niet over welke leeftijd.
Meinte: Maar de vraag is 'Is er samenhang tussen leeftijd …?'
Rita: Oh ja.
Meinte: Dus dan neem je een jonge groep en dan neem je zeg maar mensen van 40 jaar en mensen van 70 jaar, dus eigenlijk is het wel duidelijk.
Docent: Dat vind ik ook, dat vind ik geen onduidelijk ... iets anders misschien? [29]

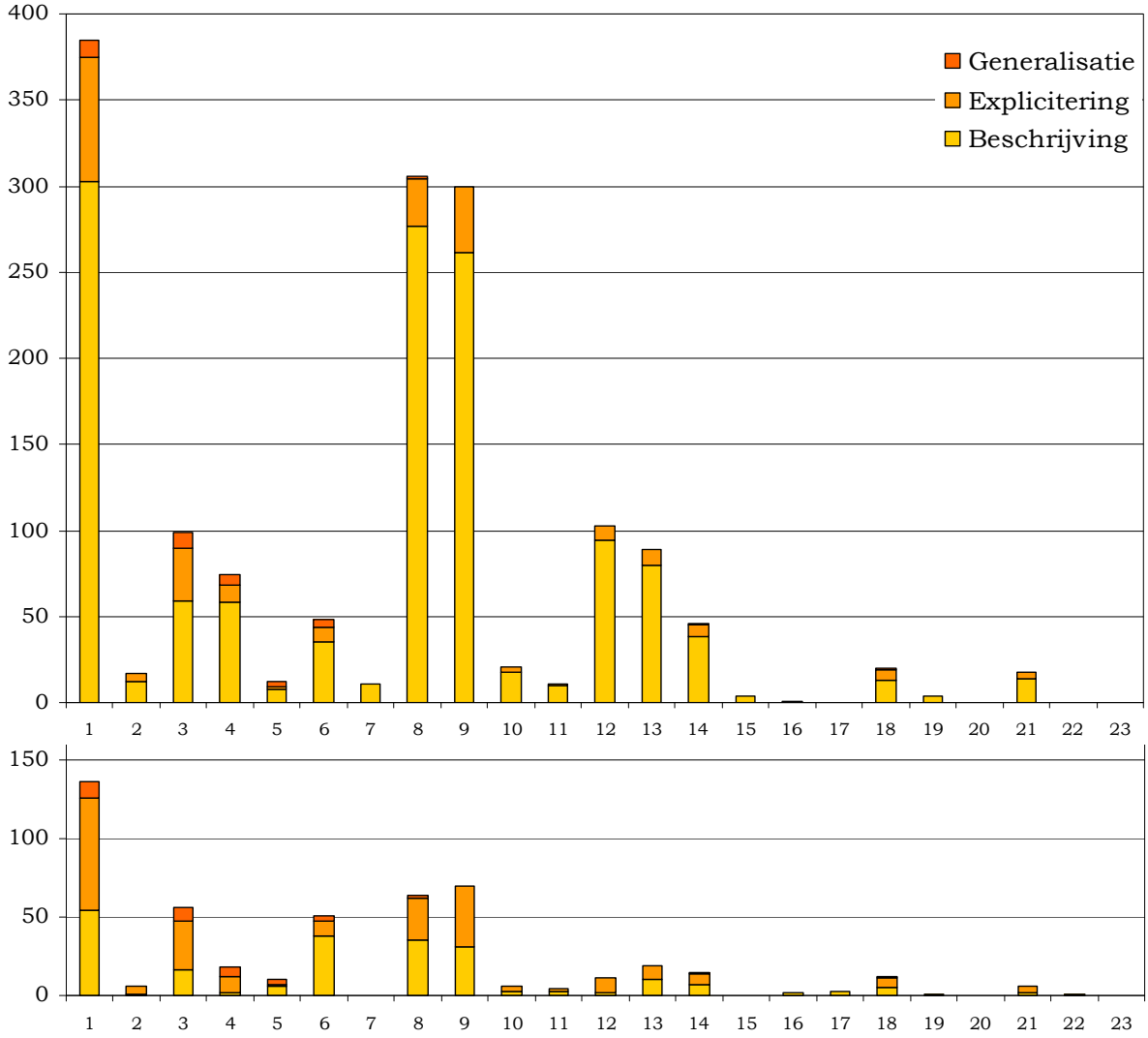
Het is dus zeker mogelijk om ook zonder de expliciete aanleiding in de vorm van een reflectieopdracht de criteria aan de orde te stellen. Hoe effectief dat gebeurde, komt in paragraaf 7.5 aan de orde.

De docenten waren genuanceerd in hun mening. In de eindevaluatie antwoordden ze op de vraag welke strategie uiteindelijk hun voorkeur heeft dat ze wellicht een combinatie zouden maken: de leerlingen wel confronteren met reflectieopdrachten en ze elkaar commentaar laten geven, maar ook zelf schriftelijk commentaar geven. Lex' idee om dan vaker commentaar op schrift te geven, wees Bea af, omdat ze dat teveel werk vond. Maar de waarde van schriftelijk commentaar stond voor beiden buiten kijf.

7.4.10 Het wetenschappelijke verhaal

Als uitgangspunt is ook geformuleerd: 'in de wisselwerking tussen onderzoek doen en reflecteren op onderzoek wordt het 'wetenschappelijke verhaal' ontwikkeld van de kwaliteitscriteria voor onderzoek en hebben de leerlingen gelegenheid het toe te passen en uit te breiden.' Daarmee wordt bedoeld dat gaandeweg de praktische opdracht de verschillende criteria voor het doen van goed onderzoek aan de orde zouden komen, niet in een vastgestelde volgorde, maar wel zodanig dat een compleet beeld ontstaat. Of dat beeld werkelijk ontstaan is bij de leerlingen komt in de volgende paragraaf aan de orde. Hier gaat het om de vraag of het gelukt is om het aan te bieden. Mortimer en Scott (2003) gebruiken er de terminologie van het theater voor; zij hebben het over de *staging,* het ten tonele voeren van het verhaal. De vraag die nu aan de orde is, kan verwoord worden als: 'Is de opvoering gelukt?'; in de volgende paragraaf gaat het er om of het publiek het verhaal begrepen heeft.

Het beeld dat naar voren komt, is dat het ten tonele voeren van begrip van bewijs zeker gelukt is. Alle beoogde criteria aan de orde zijn geweest, zij het niet in dezelfde mate. De vrees dat zonder expliciete aanleiding niet alle beoogde elementen van begrip van bewijs aan de orde zouden komen, is niet bewaarheid. Globaal gezien is er ongeveer dezelfde verhouding in de lessen rond eigen onderzoek en reflectieopdrachten als in de startlessen (vergelijk fig.



Figuur 7.13 Aantal uitspraken (van docent en leerlingen) waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest als beschrijving, als explicitering of als generalisatie. Boven: alle uitspraken rond het eigen onderzoek; onder: uitspraken rond de reflectieopdrachten.

7.6 met fig. 7.13). Een hoofdrol was weggelegd voor het eerste element van begrip van bewijs, dat aangeeft waaraan een onderzoeksvraag moet voldoen, en voor de elementen 8 en 9 die over variabelen gaan. De steekproeftrekking, die in de startlessen ook een voorname rol speelde, kwam in het eigen onderzoek beduidend minder aan bod. In het eigen onderzoek was wel meer aandacht voor criteria rond observeren en meten (elementen 12 en 13). Teleurstellend is echter de bijrol die de elementen 18 (antwoord op onderzoeksvraag) en 21 (hoe zeker ben je van je conclusie?) speelden. Hoewel niet helemaal afwezig, werd die elementen slechts een marginale plek toegekend.

Om nog even in toneeltermen te blijven spreken: de diepgang waarmee de karakters zijn uitgewerkt verschilde sterk. Het lijkt erop dat het drama voor een aanzienlijk deel aan de oppervlakte is gebleven. Er is te zien dat de meeste uitspraken zich op het concrete niveau van het onderzoek en de concrete aanpak bevonden en veel minder uitspraken gedaan werden op het generaliserende niveau van de criteria en de kwaliteit. De onderste delen van de kolommen in beide figuren – de beschrijvingen – betreffen uitspraken die zich op het concrete niveau begeven: hoe moet mijn onderzoeksvraag geformuleerd worden, welke variabelen betrek ik in mijn onderzoek, welke onderzoekspopulatie neem ik, enzovoort. De middelste delen – de expliciteringen – betreffen de uitspraken die onder woorden brengen waaróm het de ene en niet de andere variabele zou moeten zijn of waaróm het beter is meer proefpersonen te nemen. De bovenste delen – de generaliseringen – tenslotte, betreffen het formuleren van de criteria zelf. Voor de volledige ontwikkeling van het wetenschappelijke verhaal zijn alle drie niveaus nodig.

In het werken aan het eigen onderzoek (fig. 7.13, boven) was de overgrote meerderheid van de uitspraken op beschrijvend niveau (83% van alle uitspraken in de categorieën beschrijven, expliciteren en generaliseren), terwijl dat in de startlessen minder het geval was (52%, fig.7.6). De situatie bij het maken en bespreken van de reflectieopdrachten ligt daar tussenin (65%, fig. 7.13, onder). Deze scheve verdeling hoeft echter nog niet te betekenen dat het niet gelukt is het wetenschappelijke verhaal met voldoende diepgang over het voetlicht te brengen. Eén opmerking kan een karakter raak typeren en zo de interpretatie van de gehele voorstelling bepalen. Eén generalisering of explicitering ging dan ook meestal gepaard met een veelvoud aan beschrijvingen. Dat in de startlessen meer op expliciterend en generaliserend niveau gesproken werd, is inherent aan de opzet van die les, waar de docent erop gespitst was de algemene regels uit de besprekingen naar voren te halen.

De combinatie van uitspraken in patronen is dus ook erg belangrijk om te bezien. Maar omdat de opvoering van het wetenschappelijk verhaal bij uitstek als interactief toneel bedoeld is, kan dat niet bekeken worden zonder er de vraag bij te betrekken of het publiek de bedoeling van de voorstelling heeft opgepikt. Dan gaat het dus om de effectiviteit en die komt in paragraaf 7.5 aan de orde.

7.4.11 Samenvattend: in welke mate is de praktische opdracht uitvoerbaar?

Het antwoord op deze vraag is om twee redenen belangrijk. Ten eerste om te weten of de opdracht in de klas realiseerbaar is, of de docenten en leerlingen niet voor een onmogelijke opgave gesteld zijn. Ten tweede is het ook belangrijk als voorwaarde voor het onderzoek naar de effectiviteit van de opdracht. Als namelijk geconstateerd wordt dat er – bijvoorbeeld op één van de scholen – dusdanige knelpunten zijn dat hij niet uitgevoerd kan worden volgens de intentie van de makers, dan hebben uitspraken over de effectiviteit eigenlijk geen zin meer.

De opdracht is in grote lijnen door zowel docenten als leerlingen uitgevoerd zoals de bedoeling was. In de startlessen kwamen de beoogde elementen van begrip van bewijs over het voetlicht en werd een basis gelegd waarop in de rest van de praktische opdracht teruggegrepen werd. De docenten waren zich beter bewust van hun rol als initiator en stimulator van het denken van de leerlingen; ook waren ze alerter op het aan de orde stellen van de criteria

voor goed onderzoek. De reflectieopdrachten fungeerden zoals beoogd, in de zin dat ze aanleiding waren om over de criteria te praten. De praktische opdracht werd daarmee een uitvoerbare aanpak om leerlingen met de kwaliteit van onderzoek bezig te laten zijn.

Daarmee is ten opzichte van de eerste ronde ontegenzeglijk vooruitgang geboekt. Maar er blijken zeker ook knelpunten te zijn. Die hebben enerzijds betrekking op factoren die buiten het ontwerp liggen zoals de schoolomgeving en anderzijds op kenmerken van de praktische opdracht zelf.

- *Externe factoren*

Docent zijn is een drukke baan en leerling zijn in 5 vwo omvat meer dan biologie. Dit zijn gegevenheden die vanzelfsprekend zijn, maar toch hun invloed op de uitvoering van de praktische opdracht deden gelden. Docenten hebben ook andere klassen en werkzaamheden; roosterwijzigingen en binnenvallende conrectoren vragen (letterlijk!) aandacht en zijn niet bevorderlijk voor de concentratie. Toetsweken en buitenlandse reizen doorbreken het proces bij de leerlingen. Daarnaast waren er nog andere, evenmin bijzondere, externe factoren zoals ziekte van de technisch onderwijsassistent en falende apparatuur.

Dit zijn de 'normale' randvoorwaarden en een ervaren docent weet hoe hiermee om te gaan. Maar toch zet het druk op de docenten en de uitvoerbaarheid, en de docenten pleiten dan ook voor een geconcentreerde planning in een relatief rustige periode van het jaar.

- *Kenmerken van het ontwerp*

Sommige kenmerken van de praktische opdracht zelf vragen aandacht met betrekking tot de uitvoerbaarheid. Dat betreft de planning en tijdsbesteding, de invloed van de vrije onderwerpkeuze, de plaats van reflectie en feedback en de ontwikkeling van het verhaal rond begrip van bewijs.

Planning en tijdsbesteding

Hoewel de opdracht in beide strategieën in grote lijnen gegaan is zoals gepland, lijkt ook binnen de opdracht tijd een allesoverheersende factor te zijn. Van meet af aan is er onder tijdsdruk gewerkt, zowel door de leerlingen als door de docent. De leerlingen bleken veel tijd te stoppen in de aanloop: het bepalen van een onderwerp en het formuleren van een onderzoeksvraag. Niet omdat ze veel tijd stopten in het zoeken van achtergrondinformatie, want dat deden ze niet. Tegen het eind van de opdracht kwamen de leerlingen dan ook vrijwel allemaal in tijdnood en werd de planning bijgesteld. Er werd hard gewerkt door leerlingen, maar er werd ook tijd aan andere zaken besteed dan aan de opdracht.

Door de inzet van de boekjes was er voor de leerlingen duidelijk meer structuur en meer overzicht voor de leerlingen, al buitten ze dat ons inziens niet ten volle uit. Met andere woorden: leerlingen kunnen meer steun kunnen ontlenen aan het boekje en hoeven daardoor wellicht ook minder in tijdnood te komen.

Vrije onderwerpkeuze

Hoewel de leerlingen moeite hadden om een eigen onderzoek op te zetten, is het ze uiteindelijk allemaal gelukt. De grote vrijheid om te kiezen werd gewaardeerd, ondanks het lange nadenken over een onderwerp. Deze vrijheid van de leerlingen leidt wel tot een grote druk op de docenten om veel uiteenlopende onderzoeken te begeleiden, die ze niet alle qua inhoud of technieken even goed beheersen. Het komt ons inziens de uitvoerbaarheid voor docenten ten goede als de onderwerpskeus enigszins ingeperkt wordt. Bea deed dit al, maar met name om praktische redenen als beschikbaarheid van materiaal; het was geen keus op basis van haar grootste expertise. Het is logisch dat ervaring van een aantal jaren in het begeleiden van onderzoeken het docenten makkelijker zal maken veel uiteenlopende onderwerpen tegelijk te begeleiden.

De leerlingen hebben het als hun eigen onderzoek ervaren en wilden graag goede resultaten. Ze hebben allemaal een artikel c.q. verslag geschreven en ingeleverd, al is de vormeis van een wetenschappelijk artikel niet volledig uit de verf gekomen. Dat lijkt echter met meer nadruk van de kant van de docent wel uitvoerbaar.

Reflectie en feedback

De reflectieopdrachten blijken redelijk tot goed uitvoerbaar voor de leerlingen, maar kosten toch nog veel tijd, vooral door het nabespreken. Ondanks de gedecentraliseerde opzet, werden ze toch vaak als lastige onderbreking van het eigen onderzoek ervaren, meer dan als steun erbij. De relatie met het eigen onderzoek zou daarom nog versterkt moeten worden. Dat de reflectieopdracht waarin commentaar gegeven moest worden op elkaars conceptartikel, het meest gewaardeerd werd, geeft aan dat er wellicht meer momenten van peer review ingebouwd zouden kunnen worden. Ervaringen bij onderzoek in het scheikundeonderwijs daarmee zijn positief (Van Rens, 2005).

Directe feedback, zowel mondeling als schriftelijk werd door de leerlingen zeer gewaardeerd, ook al was die vaak in vragende vorm. Knelpunt hierbij is wederom de tijd, leerlingen vonden dat ze vaak lang moesten wachten tot ze de docent konden spreken. In die tijd werkten ze soms wel en soms niet door aan hun onderzoek. Meer momenten van schriftelijke feedback dan de twee in de impliciete strategie is door de docenten gesuggereerd, maar ook afgewezen vanwege de tijd die dat weer zou kosten.

De ontwikkeling van begrip van bewijs

Ondanks het gemotiveerde werken aan het eigen onderzoek, hebben de leerlingen het niet opgepikt als een motief om meer te weten te komen over begrip van bewijs. Ze bleven gefocust op het concrete niveau van hun eigen onderzoek.

Gezien de zware wissel die op de docenten werd getrokken, is het niet verwonderlijk dat ze niet alle kansen benutten voor het expliciteren en generaliseren van de criteria. Desondanks kwamen alle beoogde elementen van begrip van bewijs wel aan de orde, zij het niet in dezelfde mate. Veel nadruk viel op de startfase van het onderzoek: onderzoeksvraag, variabelen en hypothese. Ook was er aandacht voor eerlijk tellen en meten. Het overkoepelende element 21 (over validiteit en betrouwbaarheid) kreeg slechts weinig aandacht.

Het lijkt dus wel degelijk mogelijk het verhaal van begrip van bewijs over het voetlicht te brengen, maar meer aandacht voor de centrale vraag ‘zeker weten?’ lijkt nodig.

- *De last van extern én intern*

De praktische opdracht trekt een zware wissel op de docenten. Ze moeten deze arbeidsintensieve opdracht combineren met hun andere werkzaamheden. Binnen de opdracht moeten ze als een jongleur veel ballen tegelijk in de lucht houden. Ze moeten zowel het kader scheppen waarin de leerlingen aan de slag kunnen gaan als alle onderzoeksteams bijstaan bij het opzetten en uitvoeren van hun eigen onderzoek. Tegelijkertijd wordt van ze verwacht dat ze de ontwikkeling van begrip van bewijs bij de leerlingen in de gaten houden en stimuleren. Dat wil zeggen dat ze continu alert moeten zijn op mogelijke aanknopingspunten voor het aan de orde stellen van kwaliteitscriteria voor onderzoek. In de expliciete groep hebben ze daarvoor weliswaar de reflectieopdrachten, maar het bespreken daarvan vraagt veel van hun tijd en aandacht die ze daardoor niet in het begeleiden van het eigen onderzoek van de leerlingen kunnen steken.

In de tweede ronde wisten de docenten beter dan in de eerste ronde wat er van hen verwacht werd en wat ze te wachten stond, waarschijnlijk alleen al doordat het een tweede ronde betrof. Dat onderstreept het belang van een goede voorbereiding, zowel inhoudelijk met betrekking tot begrip van bewijs als wat begeleidingsgedrag betreft.

Aan het begin van dit hoofdstuk stelden we de vraag *in welke mate is de praktische opdracht uitvoerbaar voor docent en leerlingen, met speciale aandacht voor de reflectieopdrachten en de ontwikkeling van begrip van bewijs?* Die vraag beantwoorden we met: het vraagt veel van de docenten en de leerlingen, maar ons inziens niet té veel. De uitvoerbaarheid zou ons inziens echter wel verbeterd kunnen worden door een aantal aanpassingen.

Aanpassingen zouden moeten plaatsvinden met betrekking tot de planning (nog geconcentreerder), de onderwerpskeus (meer inperken), de nadruk op het vereiste eindproduct, maar vooral met betrekking tot de focus op begrip van bewijs. Het zou zowel de docenten als de leerlingen nog duidelijker voor ogen moeten staan dat het primair gaat om het leren over de kwaliteit van onderzoek. Daarnaast zal ervaring met het uitvoeren van deze opdracht het voor docenten ongetwijfeld makkelijker maken hem uit te voeren.

En de vraag of de uitvoering dusdanige problemen heeft gegeven dat die het kijken naar de effectiviteit beïnvloedt, beantwoorden we negatief. De uitvoering heeft ons inziens een leerproces niet onmogelijk gemaakt, noch op een van beide scholen, noch in een van beide onderwijsleerstrategieën.

7.5 Effectiviteit

In deze paragraaf worden de leereffecten van de praktische opdracht beschreven en worden daar verklaringen voor gezocht. Effectiviteit heeft in het kader van dit onderzoek enkele – gerelateerde, maar toch te onderscheiden – aspecten. Ten eerste gaat het om de effectiviteit in het ontwikkelen van een motief om te leren onderzoeken en een *common ground,* een gezamenlijk startkapitaal aan kennis over onderzoek doen. Vooral de startles had die intentie, maar ook het uitgangspunt dat de leerlingen een éigen onderzoek zouden opzetten en uitvoeren had dat doel.

Ten tweede betreft de effectiviteit het ontwikkelen van de criteria voor goed onderzoek, de elementen van begrip van bewijs. Dat moest plaatsvinden in de gesprekken rond het eigen onderzoek én in het maken en bespreken van de reflectieopdrachten. Behalve in de gesprekken kreeg het ook zijn beslag in de aantekeningen van de leerlingen en de feedback van de docent daarop.

Ten derde wordt met de effectiviteit van de praktische opdracht de mate bedoeld waarin de leerlingen naderhand blijk gaven de criteria voor de kwaliteit van onderzoek te beheersen. Dat wil zeggen, de mate waarin ze in hun eindproducten blijk geven de criteria juist te gebruiken én de mate waarin ze die criteria juist toepassen bij de beoordeling van ander onderzoek.

Tenslotte heeft effectiviteit ook betrekking op de ervaren leeropbrengst. In hoeverre hebben de leerlingen zélf het idee dat ze hebben leren onderzoeken?

In deze paragraaf worden deze vier aspecten achtereenvolgens besproken aan de hand van de in tabel 7.3 genoemde instrumenten. Waar mogelijk zijn er verbindingen gelegd tussen de bevindingen die deze verschillende manieren om naar effectiviteit te kijken opleveren, want uiteindelijk is het vooral de samenhang van deze aspecten die het oordeel over de effectiviteit bepaalt (de *polyphony* waar Roth het over heeft in het citaat aan het begin van dit hoofdstuk). Daar wordt dan ook deze paragraaf mee afgesloten. Achtereenvolgens komen de vier bovengenoemde aspecten van effectiviteit aan de orde:

- de effecten van de startlessen voor wat betreft motief en *common ground* (par. 7.5.1);
- de weg naar de eindproducten, dat wil zeggen de mate waarin en de manier waarop begrip van bewijs aan de orde kwam rond het eigen onderzoek en in de reflectieopdrachten, zoals op te maken is uit de gesprekken en de aantekeningen in de boekjes en de concepten van de artikelen; daarin worden ook de verwachtingen omtrent de induc-

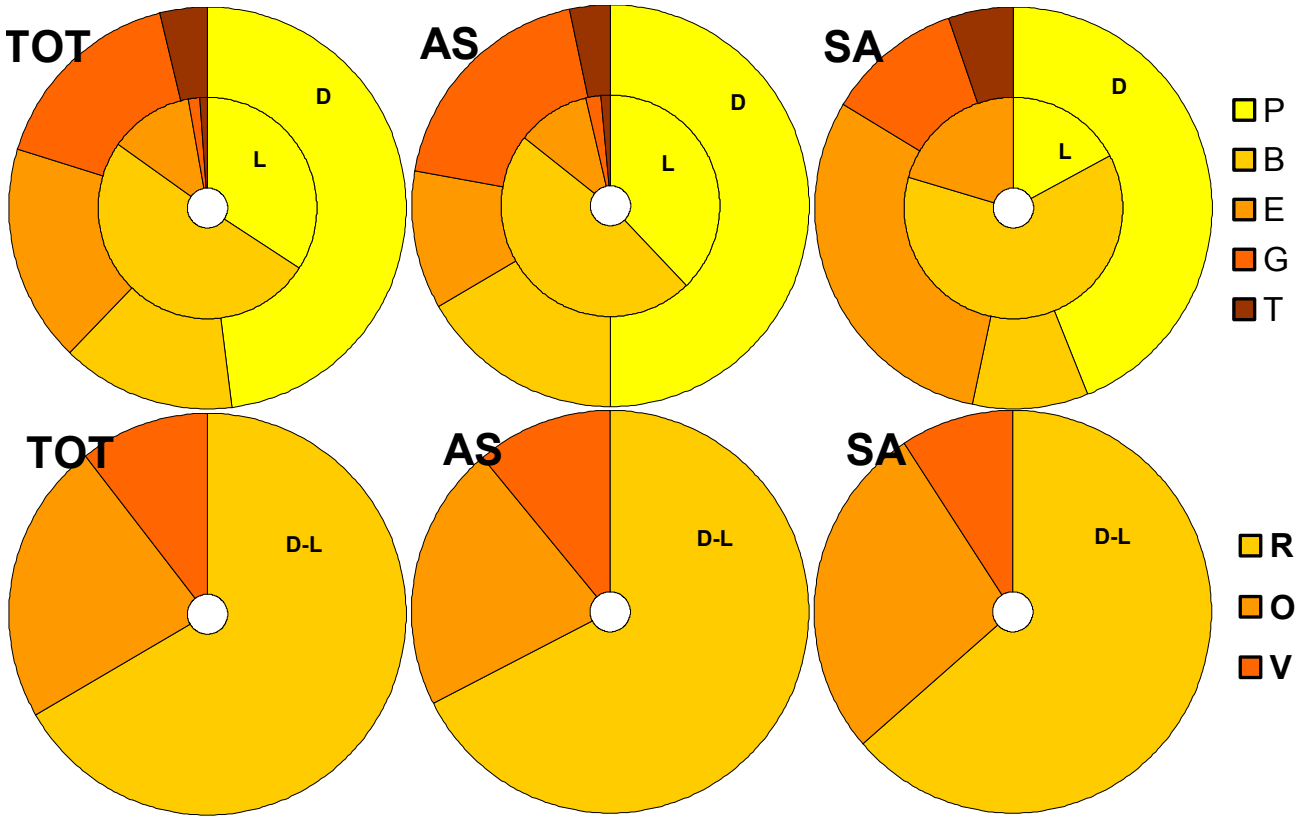
tieve dan wel deductieve manier van ontwikkelen van begrip van bewijs betrokken (7.5.2);
- de mate waarin de elementen van begrip van bewijs te vinden zijn in de uiteindelijke artikelen (7.5.3) en de resultaten op de voor- en natoets, waarin de leerlingen begrip van bewijs moesten toepassen op ander onderzoek (7.5.4);
- de mening van de leerlingen over het leereffect (7.5.5).

De paragraaf wordt besloten met samenvattende conclusies over de effectiviteit van de praktische opdracht met betrekking tot het ontwikkelen van begrip van bewijs (7.5.6).

7.5.1 De startlessen

In paragraaf 7.4.2 is beschreven hoe de startlessen verliepen en in welke mate de verschillende elementen van begrip van bewijs aan de orde zijn geweest als beschrijving, explicitering of generalisatie (fig. 7.6). Om de vraag te beantwoorden of daarmee een voldoende gezamenlijke kennisbasis gelegd is voor het uitvoeren van het eigen onderzoek, is het zinvol er meer in detail naar te kijken. Figuur 7.14 laat dat zien.

In de bovenste drie cirkeldiagrammen is weergegeven hoe de uitspraken van leerlingen (de binnenste cirkel) en docenten (de buitenste cirkel) verdeeld waren over de categorieën *problematiseren – beschrijven – expliciteren – generaliseren – toepassen*. Het aandeel pro-



Figuur 7.14 Verdeling van uitspraken en patronen in de startlessen waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest.
Boven: verdeling van uitspraken in de categorieën problematisering (P), beschrijving (B), explicitering (E), generalisatie (G) en toepassing (T). D: uitspraken van docent; L: uitspraken van leerlingen.
Onder: mate waarin de uitspraken van leerlingen en docent volledige (V), onvolledige (O) of rudimentaire (R) patronen vormen.
TOT: totaal; AS: Amsterdamse School; SA: School in Amstelveen.

blematiseringen was bij de docenten beduidend groter dan bij de leerlingen (49 tegen 34 procent), terwijl het voor de beschrijvingen andersom was (14 tegen 50). Dat is geheel volgens de opzet van de startlessen: de docent legt een probleem neer en daagt de leerlingen uit met oplossingen te komen. Maar de oplossingen van de leerlingen bleven grotendeels op het beschrijvende, concrete niveau. De expliciteringen waren redelijk gelijk verdeeld over leerlingen en docent (17 tegen 13), de generalisaties en de toepassingen kwamen echter grotendeels voor rekening van de docent (respectievelijk 16 en 4 tegen 2 en 1 procent). De lessen in Amsterdam en Amstelveen droegen in verschillende mate bij aan dit resultaat. Ten eerste was het aantal uitspraken in Amsterdam groter dan in Amstelveen (528 tegen 174), omdat er bij Lex veel meer interactie via veel korte uitspraken plaatsvond (met ongeveer evenveel uitspraken van leerlingen als van Lex) en omdat hij meer tijd voor het gesprek nam (en daardoor in tijdnood kwam). Het totaalresultaat lijkt daardoor meer op het resultaat uit Amsterdam dan op dat uit Amstelveen. In Bea's lessen problematiseerden leerlingen minder, bleven meer op het concrete niveau en generaliseerden in het geheel niet, evenmin pasten ze criteria toe op ander onderzoek. Wel expliciteerden ze relatief meer (21 procent), maar in absolute termen blijft het weinig, omdat er maar weinig (12) leerlinguitspraken waren. Bea zelf compenseerde dat door meer aan het woord te zijn en vooral door meer te expliciteren (30 procent).

De mate waarin er volledige, onvolledige of rudimentaire patronen in de gesprekken te onderkennen waren in Amsterdam dan wel Amstelveen (de onderste diagrammen van figuur 7.14) verschilt veel minder. Ongeveer een twaalfde van de patronen bevatte zowel beschrijvende, expliciterende als generaliserende uitspraken over criteria voor goed onderzoek (en was daarmee volledig), ongeveer een kwart bevatte twee van die drie (en was daarmee onvolledig) en tweederde bevatte maar één van die drie (rudimentair). In Amstelveen waren het wat minder volledige en wat meer onvolledige patronen.

Het is moeilijk te bepalen of de mate waarin in de startlessen de elementen van begrip van bewijs aan de orde zijn geweest *voldoende* is, omdat een hard criterium daarvoor ontbreekt. Maar omdat alle beoogde elementen ter sprake zijn gekomen én zowel docenten als leerlingen aan de discussie erover hebben bijgedragen – zij het in Amsterdam meer evenwichtig dan in Amstelveen – lijken de startlessen brede, gezamenlijke basis gelegd te hebben in de zin dat leerlingen zich actief met de vraag 'wat is goed onderzoek?' hebben beziggehouden. Bovendien is het 'domme blondjes-onderzoek' een soort ijkpunt geworden, waar later regelmatig terug naar verwezen werd, zowel door docenten als leerlingen. Dat gebeurde vooral met betrekking tot eenduidigheid en afperking van de onderzoeksvraag, maar ook met betrekking tot andere elementen van begrip van bewijs, zoals het onder controle houden van variabelen.

Toch is enige reserve ten aanzien van de kwaliteit van die gezamenlijke kennisbasis op zijn plaats. In het nagesprek over de startlessen keek Lex terug op het aan de orde komen van de criteria met betrekking tot variabelen en blanco. Hij baseerde er een algemenere verwachting op over het omgaan met criteria door de leerlingen in hun eigen onderzoek:

> Ze zaten er dichtbij, maar vonden het uiteindelijk niet zo belangrijk. Dat is heel grappig, want dát is zoals hun eigen onderzoek ook gaat: ze doen het wel, ze doen bijvoorbeeld …, ze zijn in staat om een blanco proef in te zetten maar daar geen consequenties aan te verbinden. Dus dan krijg je bijvoorbeeld van 'ja, maar het kan ook wel liggen aan dit of dat, dat die dingen sneller gegroeid zijn'. En dan vergeten ze dus gewoon naar hun blanco te kijken die ze om die reden hebben ingezet. Dus ze doen het wel, ze weten wel dat het moet, maar of dat nou zo belangrijk is voor een goed onderzoek, ja …, dat idee, dat belang daarvan… Dat geldt een beetje algemeen voor criteria waaraan hun .. ; als ze straks bezig gaan, zul je zien – of tenminste, we hopen van niet, maar, eh we hopen dan dat er verschil in zit, tussen de twee groepen – maar, dat zul je, denk ik, wel zien, dat er nonchalant mee omgesprongen wordt in de praktijk, met de criteria. [30]

En dat spoort met de al in paragraaf 7.4.3 geformuleerde constatering dat het de leerlingen vooral ging om het eigen onderzoek zo goed mogelijk (= in overeenstemming met de criteria) uit te voeren en niet om zoveel mogelijk te leren over wat goed onderzoek is (de criteria zelf). De startlessen zijn dus wel effectief geweest in het ontwikkelen van een motief voor een zo goed mogelijk eigen onderzoek (zie fig. 6.1), maar niet in het ontwikkelen van een motief voor het leren bewaken van kwaliteit van onderzoek meer in het algemeen.

7.5.2 De weg naar de eindproducten

Na de startlessen begonnen de leerlingen met hun eigen onderzoek. Ze praatten erover met elkaar en met de docent en maakten aantekeningen in hun boekjes. De expliciete groep maakte ook de reflectieopdrachten. De docenten begeleidden het proces door met de leerlingen in gesprek te gaan en in de impliciete groep schriftelijk commentaar te geven op de onderzoeksopzet en het conceptartikel.

Van acht onderzoeksgroepjes zijn alle in de lessen gevoerde gesprekken opgenomen en van de meeste leerlingen zijn de gebruikte boekjes verzameld, en daarmee het geschreven commentaar van de docenten en de uitwerking van de reflectieopdrachten. Ook zijn de conceptartikelen van de meeste leerlingen ingenomen. Met behulp van die gegevens wordt de rol die de criteria voor goed onderzoek (begrip van bewijs) hebben gespeeld bij het verloop van de eigen onderzoeken van de leerlingen geïllustreerd. Eerst wordt een beeld geschetst van de gesprekken, daarna van de aantekeningen van leerlingen en docenten in de boekjes, inclusief de reflectieopdrachten, en tot slot van de voorlopige eindproducten en het commentaar van de docenten daarop.

- *Begrip van bewijs in de gesprekken*

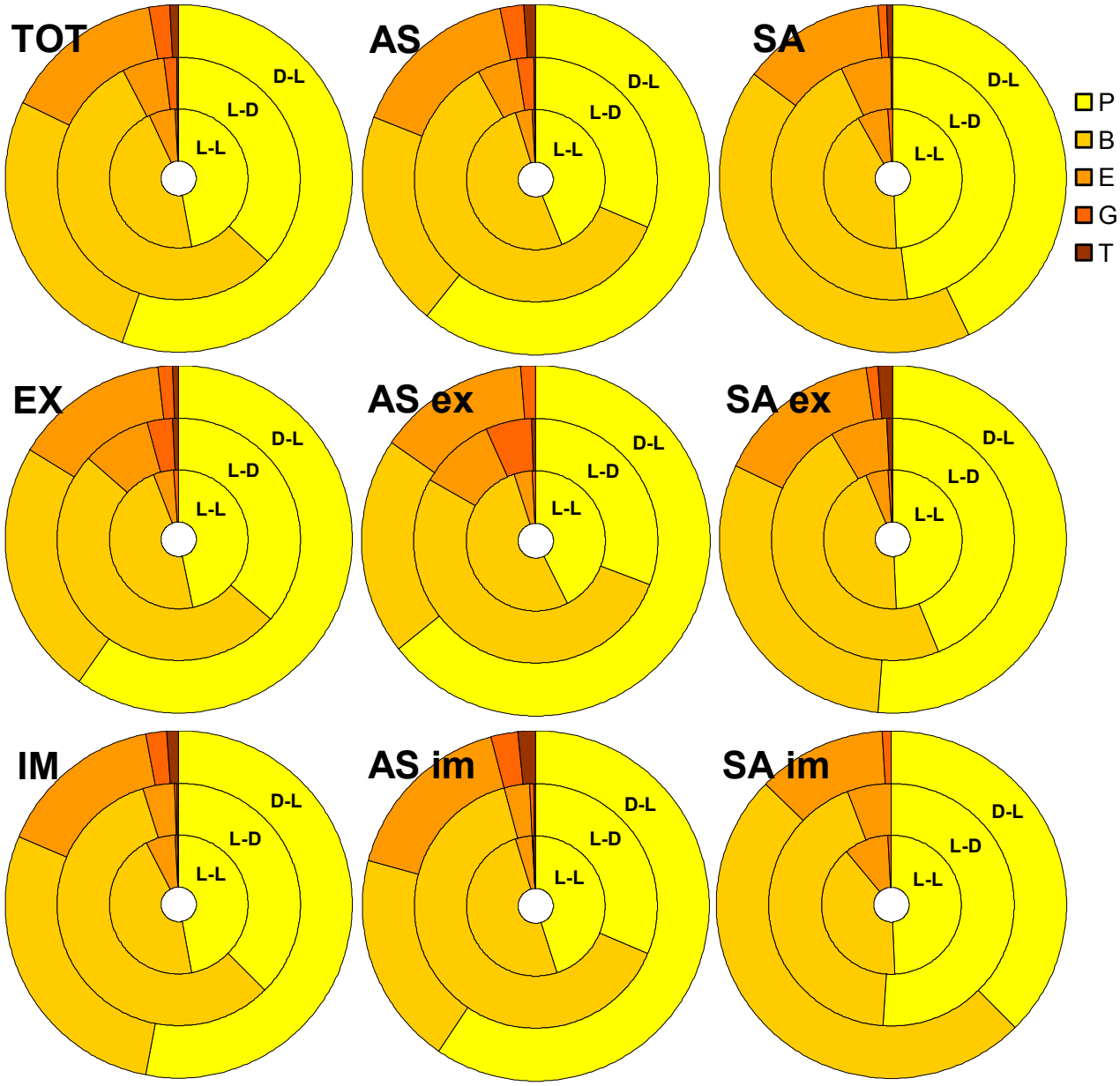
In paragraaf 7.4.10 is al beschreven welke elementen van begrip van bewijs in de opgenomen gesprekken rond de onderzoeken van leerlingen aan de orde zijn gekomen (fig. 7.13). Ook is weergegeven in welke mate er in de uitspraken van docenten en leerlingen sprake was van problematisering, beschrijving, explicitering, generalisatie of toepassing van de elementen (fig. 7.7). Dat werd gedaan om te laten zien dat het gelukt is het wetenschappelijke verhaal aan de orde te stellen. Om de vraag te beantwoorden of dat ook in voldoende mate is gebeurd om effectief te zijn, is een nadere analyse daarvan nodig.

De gegevens die in figuur 7.7 al gegeven zijn, zijn uitgesplitst in docent-leerlinggesprekken en leerling-leerlinggesprekken in de figuren 7.15 (gesprekken over eigen onderzoek) en 7.16 (gesprekken over de reflectieopdrachten). Ze worden gegeven voor alle leerlingen tezamen en voor beide scholen en beide groepen. Te zien is dat voor de docent-leerlinggesprekken rond het eigen onderzoek (fig. 7.15, de twee buitenste ringen van elke cirkel) globaal hetzelfde beeld ontstaat als in de startlessen (fig. 7.14): de docent problematiseerde, de leerlingen bleven hoofdzakelijk op het beschrijvende niveau. Maar niet alleen de generalisaties kwamen voor rekening van de docent, ook de expliciteringen, die ook nog eens een kleiner deel van de gesprekken vormden. Dat laatste is niet onlogisch, want in de startlessen vond een nogal 'geconcentreerd' gesprek plaats over de kwaliteit van een onderzoek, terwijl in de gesprekken over het eigen onderzoek veel meer aandacht was voor de concrete gang van zaken. In hun onderlinge gesprekken (binnenste ring van elke cirkel) expliciteerden en generaliseerden de leerlingen niet meer dan ze deden in hun gesprekken met de docent. Ook daar vond het dus maar erg weinig plaats.

Net als in de startlessen is er een verschil tussen de scholen (AS en SA), dus tussen de docenten. Lex problematiseerde vaker dan Bea (61 tegen 44 procent); hij speelde de bal van de leerlingen veel vaker terug. Dat wordt dan ook weerspiegeld in de verdeling van de uitspraken die leerlingen doen in gesprek met hun docent. In Amsterdam gaven ze vaker een beschrijving dan in Amstelveen (62 tegen 45 procent). Voor het aandeel van explicitering,

generalisatie en toepassing maakt het echter weinig uit. Dat kwam in beide scholen vaker voor rekening van de docent (19 en 14 tegen 7 procent). De verschillen in de gesprekken van leerlingen onderling zijn minder groot, al werd er in Amstelveen iets vaker geëxpliciteerd en gegeneraliseerd.

In paragraaf 6.4 zijn verwachtingen geformuleerd ten aanzien van het voorkomen van uitspraken uit bepaalde categorieën in de impliciete en de expliciete groep. Het tweede deel van de eerste verwachting (1b), luidt: *in de expliciete onderwijsleerstrategie wordt vaker een veralgemeniseerd criterium in een nieuwe context toegepast.* In figuur 7.15 is te zien dat deze verwachting niet is uitgekomen. Sowieso werden de criteria nauwelijks toegepast op ander onderzoek.



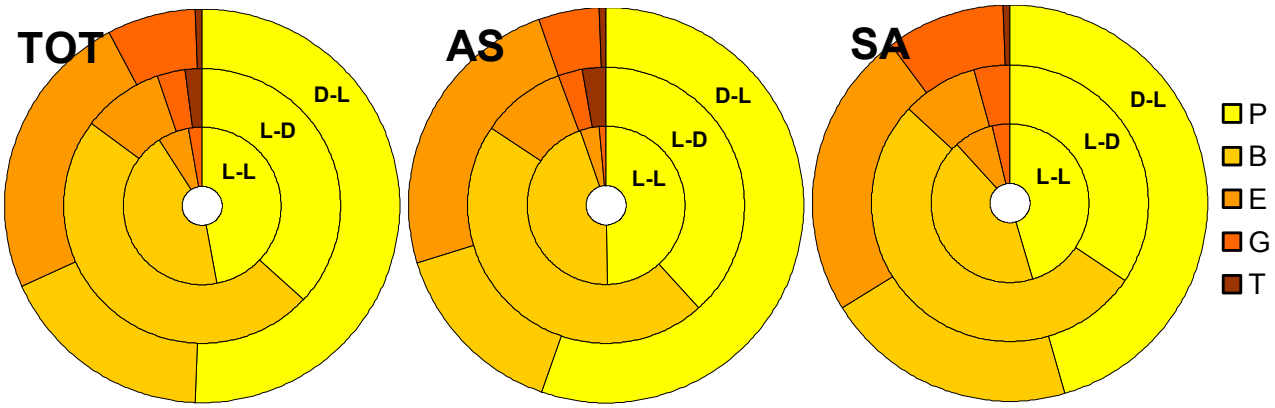
Figuur 7.15 Verdeling van uitspraken rond het eigen onderzoek waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest als problematisering (P), beschrijving (B), explicitering (E), generalisatie (G) of toepassing (T). D-L: uitspraken van docenten in gesprek met leerlingen; L-D: uitspraken van leerlingen in gesprek met docenten; L-L: uitspraken van leerlingen in gesprek met leerlingen; TOT: totaal; AS: Amsterdamse School; SA: School in Amstelveen; EX: expliciete groep; IM: impliciete groep.

Verwachting 2b luidt: *in de impliciete strategie zijn relatief meer uitspraken op het beschrijvende niveau.* Die kwam wel uit (59 tegen 50 procent), maar alleen voor de uitspraken van leerlingen in gesprek met hun docent. Leerlingen in de expliciete groep deden in die situatie meer expliciteringen en generalisaties (14 tegen 4 procent). In de onderlinge gesprekken was er nauwelijks verschil.

Dat strookt met verwachting 3b, die ook bewaarheid werd: *in de gesprekken tussen docent en leerlingen komen vaker expliciteringen en generaliseringen voor dan in gesprekken tussen leerlingen onderling.* Dat verwachting 3a uitkwam, is al eerder geconstateerd: *de problematiseringen komen vaker van de docent dan van leerlingen.* Verwachting 3c – *deze verschillen zijn groter in de impliciete strategie dan in de expliciete* – kwam een klein beetje uit wat het eerste deel betreft: in de expliciete groep werd iets vaker door leerlingen geproblematiseerd. Wat het tweede deel betreft kwam hij niet uit. In de impliciete groep werden in docent-leerlinggesprekken elf procent expliciteringen en generalisaties gedaan (door docent en leerlingen samen) en acht procent in onderlinge gesprekken; in de expliciete groep respectievelijk veertien en zes procent. In de impliciete groep kwamen ze echter wel veel minder vaak van de leerlingen dan in de expliciete groep (vier tegen dertien procent).

De vraag is of dat verschil geïnduceerd is door de docent of dat het uit de leerlingen zelf kwam. In de expliciete groep zijn namelijk de criteria voor goed onderzoek in de reflectieopdrachten ter sprake geweest. Het kan zijn dat de docent daar in die groep dan makkelijker een beroep op kon doen – en het dus door de leerlingen uitgesproken werd – dan in de impliciete groep. Dat lijkt aannemelijk, want de onderlinge gesprekken tussen leerlingen vertoonden namelijk nauwelijks verschil wat het percentage expliciteringen en generalisaties betreft (6 tegen 8).

In de leerlinguitspraken in de gesprekken met de docent over de reflectieopdrachten (fig. 7.16) is dan ook ongeveer hetzelfde percentage expliciteringen en generalisaties te zien als in die gesprekken over het eigen onderzoek in de expliciete groep, namelijk ongeveer vijftien. Ook de verdelingen in de gesprekken tussen leerlingen onderling in de verschillende situaties (eigen onderzoek en reflectieopdracht) komen redelijk met elkaar overeen. Het verschil tussen de gesprekken in de verschillende situaties zit echter vooral in de mate waarin de docenten expliciteerden en generaliseerden (31 tegen 17 procent), wat grotendeels in de plaats kwam van beschrijven (18 tegen 27 procent). Het betekent dat de docenten bij het bespreken van de reflectieopdrachten de criteria onder woorden hebben gebracht, en dat de leerlingen



Figuur 7.16 Verdeling van uitspraken rond de reflectieopdrachten waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest als problematisering (P), beschrijving (B), explicitering (E), generalisatie (G) of toepassing (T). D-L: uitspraken van docenten in gesprek met leerlingen; L-D: uitspraken van leerlingen in gesprek met docenten; L-L: uitspraken van leerlingen in gesprek met leerlingen; TOT: totaal; AS: Amsterdamse School; SA: School in Amstelveen.

dat nauwelijks gedaan hebben. Het volgende citaat uit de nabespreking van reflectieopdracht 1 is wat dat betreft illustratief:

36	Docent:	Oké, dan deze [*ze bedoelt de tweede onderzoeksvraag in reflectieopdracht 1*]. En hier zit eigenlijk ook nog een ander, ehm, criterium aan die onderzoeksvraag waar die aan moet voldoen. Wat is hier niet goed aan?
37	Giorgio:	Nou, eh, huisdieren; zoveel dieren zijn er. Ik denk dat je toch wel een dier moet kiezen, dus bijvoorbeeld een kat of vogels, kanarietje of zo.
38	Docent:	Ja?
39	Giorgio:	En eh […]
40	Docent:	Hoe zou je dat, hoe zou je dat noemen?
41	Giorgio:	Eh, wij hebben …
42	Docent:	Dus, het is niet afperken in tijd en ruimte of de doelgroep, beetje doelgroep, maar het is eigenlijk niet …
43	Giorgio:	Ja, wat je onderzoekt gewoon, totaal, je object gaat onderzoeken.
44	Docent:	Ja, het is niet specifiek genoeg.
45	Giorgio:	Nee.
46	Docent:	Dus je moet specificeren …
47	Giorgio:	Ja.
(46)	Docent:	… om welk dier het gaat. Als ik zou zeggen van eh, welke voedingsmiddelen zitten in een gezonde lunch? Ja dan, dan dat alle voedingsmiddelen, maar dat je … welke voedingsmiddelen? Je moet toch wel specifiek zijn over welk voedingsmiddel heb je het dan? Als ik zeg nou, hoeveelheid koolhydraten, in de lunch, nou gezonde lunch is natuurlijk niet eenduidig, want wat is gezond?
48	Giorgio:	Nou, gevarieerde voeding toch! Is toch gezond?
49	Docent:	Nou goed, dat moet je dan schrijven, dus het moet naast dat het afgeperkt is, naast dus dat het eenduidig is, goed omschreven, moet het specifiek genoeg zijn. [31]

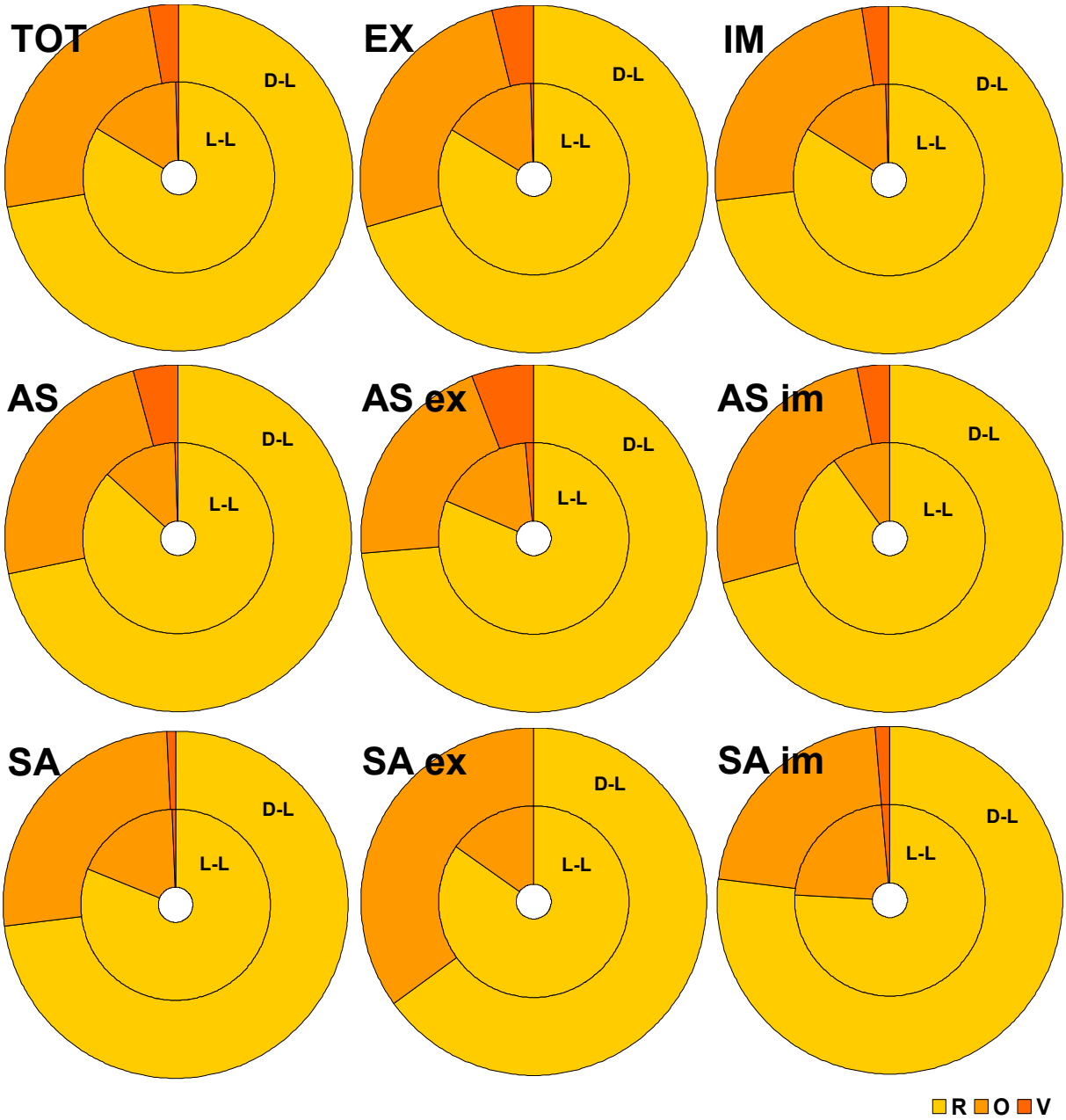
Hoewel de docent in eerste instantie vooral problematiserende vragen stelt (uitspraak 36), sluit zíj de bespreking van de onderzoeksvraag uit de opdracht af met een uitleg van het criterium (44, 46) en recapituleert zíj de criteria in gegeneraliseerde vorm (49).

Deze rolverdeling komt ook tot uiting in de patronen die in de gesprekken zijn waargenomen. De volledige patronen (met zowel *beschrijving, explicitering* als *generalisatie*) in de gesprekken rond het eigen onderzoek (fig.7.17) waren veruit in de minderheid (slechts 11 van de in totaal 672 patronen) en traden op in interactie tussen docent en leerlingen en in veel mindere mate tussen leerlingen onderling (9 tegen 2). Daarbij valt op dat het aandeel in Amsterdam beduidend groter is dan in Amstelveen. Lex was dus niet alleen breedsprakiger (617 gecategoriseerde uitspraken tegen 348 van Bea), maar maakte redeneringen ook vaker volledig (8 tegen 1 van de volledige patronen!). Hij deed dat overigens in gelijke mate in de expliciete en de impliciete groep (4 om 4). Dat het in zijn impliciete groep minder lijkt, komt doordat daar meer patronen geteld zijn (228 tegen 138 in de expliciete groep in Amsterdam).

De onvolledige patronen, die slechts twee van de drie categorieën *beschrijving, explicitering en generalisatie* bevatten, traden in de gesprekken rond het eigen onderzoek in Amstelveen even vaak op als in Amsterdam en in de expliciete groep even vaak als in de impliciete, allemaal rond de 25 procent. In Amstelveen in de expliciete groep vaker dan in de impliciete (35 tegen 22), maar in absolute aantallen is het verschil minder spectaculair (17 tegen 13). Het aandeel volledige en onvolledige patronen in de gesprekken tussen leerlingen onderling was meestal een stuk kleiner dan in de gesprekken met de docenten. Alleen in de impliciete groep in Amstelveen was dat in de gesprekken over het eigen onderzoek ongeveer even groot.

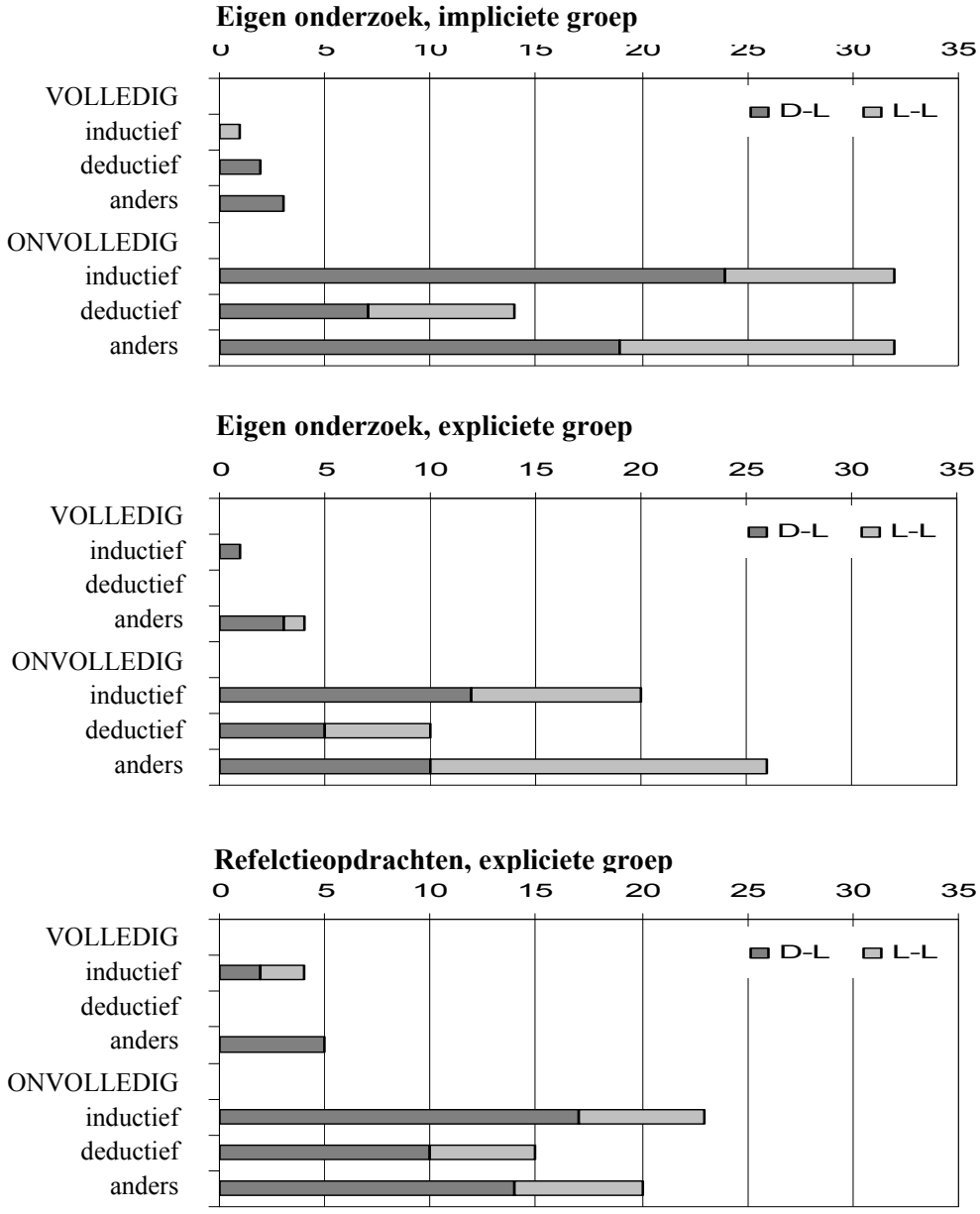
In paragraaf 6.4 is ook een verwachting geformuleerd ten aanzien van het voorkomen van volledige, onvolledige dan wel rudimentaire patronen in de impliciete en de expliciete groep (2a): *in de expliciete strategie komen meer volledige patronen voor en minder rudimentaire dan in de impliciete strategie.* In figuur 7.17 is te zien dat het aantal volledige patronen in beide groepen en in beide situaties erg klein c.q. nihil was. Het aandeel rudimentaire patronen in de expliciete strategie was niet veel kleiner dan in de impliciete: 70 tegen 74 procent in de gesprekken tussen docent en leerlingen; in de onderlinge gesprekken tussen leerlingen is er helemaal geen verschil te zien: beide 83 procent.

Er is ook een verwachting geformuleerd ten aanzien van het voorkomen van inductie-

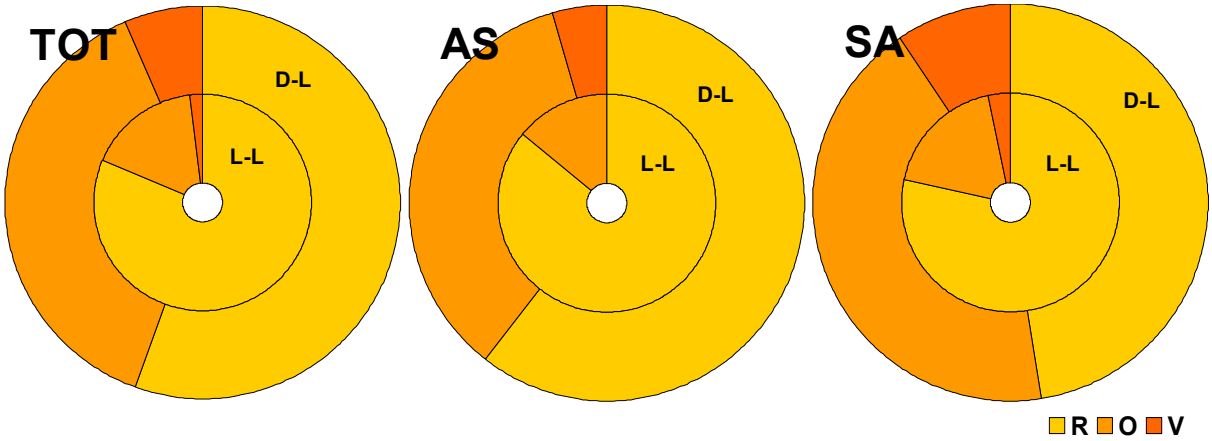


Figuur 7.17 Voorkomen van rudimentaire (R), onvolledige (O) en volledige (V) patronen in de uitspraken rond het eigen onderzoek waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest. D-L: in gesprekken tussen docent en leerlingen; L-L: in gesprekken tussen leerlingen onderling; AS: Amsterdamse School; SA: School in Amstelveen; EX: expliciete groep; IM: impliciete groep.

ve dan wel deductieve patronen (1a): *in de impliciete strategie komen vooral inductieve patronen voor, in de expliciete onderwijsleerstrategie vaker deductieve patronen.* Figuur 7.18 laat zien dat die verwachting maar ten dele is uitgekomen. Weliswaar kwamen er meer dan twee keer zoveel inductieve als deductieve patronen voor in de gesprekken in de impliciete groep, maar iets meer nog dan inductieve patronen kwamen andere patronen voor. Deze hebben meestal de vorm *beschrijving – explicitering – beschrijving* en zijn dus inductief noch deductief. En ook in de expliciete groep was het aantal inductieve patronen groter dan deductieve. Bij de reflectieopdrachten is het verschil het kleinst, in de onderlinge gesprekken van leerlingen daarover is het aantal zelfs vrijwel even groot.



Figuur 7.18 Aantallen inductieve, deductieve en andere patronen in de uitspraken over het eigen onderzoek en de reflectieopdrachten in de impliciete en expliciete groep.
D-L: in gesprekken tussen docent en leerlingen; L-L: in gesprekken tussen leerlingen onderling.



Figuur 7.19 Voorkomen van rudimentaire (R), onvolledige (O) en volledige (V) patronen in de uitspraken rond de reflectieopdrachten waarin de verschillende elementen van begrip van bewijs in de lessen aan de orde zijn geweest. D-L: in gesprekken tussen docent en leerlingen; L-L: in gesprekken tussen leerlingen onderling; TOT: totaal; AS: Amsterdamse School; SA: School in Amstelveen.

In de gesprekken tussen leerlingen en docent over de reflectieopdrachten was het aantal waargenomen onvolledige patronen kleiner dan bij het eigen onderzoek (fig. 7.18), maar het relatieve aantal was groter (38 tegen 2.5 procent, fig. 7.17 en 7.19), hetgeen gezien de aard en de duur van die gesprekken niet verwonderlijk is. Het aandeel volledige patronen was wat groter dan in de gesprekken over het eigen onderzoek (9 van de 210), en ook daar traden ze vooral op tussen docent en leerlingen (7 tegen 2). In deze situatie was Bea wel ongeveer even vaak volledig (4 van de 42 patronen) als Lex (3 van de 66). Maar in de gesprekken tussen leerlingen onderling was nauwelijks een groter aandeel waar te nemen (19 tegen 17 procent). De opdrachten op zich stimuleren de leerlingen onderling blijkbaar niet tot volledige redeneringen, in de nabesprekingen met de docent komen die redeneringen wel tot stand.

Ook hier kan de vraag of de elementen van begrip van bewijs in de lessen in *voldoende mate en volledig genoeg* aan de orde geweest voor het leren bewaken van de kwaliteit van onderzoek, niet beantwoord worden zonder criterium daarvoor. Want uiteraard is het niet alleen van belang dat de criteria ter sprake komen, maar ook dat de leerlingen ze oppikken, dat wil zeggen dat ze de criteria begrijpen en toepassen. Dat dat ze niet altijd makkelijk gaat, blijkt uit het volgende fragment, waar Lex het verschil tussen hypothese en voorspelling uitlegt naar aanleiding van zijn geschreven commentaar bij een conceptartikel van een onderzoeksgroepje van drie leerlingen. Lex expliciteert, generaliseert en past het toe op andere voorbeelden (uitspraak 88, 96, 100, 102, 104), maar vraagt niet aan de leerlingen het zelf te verwoorden voor hun eigen onderzoek. Als ze zeggen dat ze het snappen (103) spoort hij ze aan het te gaan toepassen, maar de leerlingen gaan er niet mee aan de slag, althans niet op dat moment.

86 Docent: Kijk, het gaat er eerst om, dat geef ik aan hè, een hypothese, er moet een hypothese bij.

87 Martin: Maar ik snap het verschil tussen een hypothese en een voorspelling echt niet.

88 Docent: Nou een hypothese is een, zeg maar, een vooronderstelling op basis van een theorie. Dus je zegt, zo zit het. Kijk, bijvoorbeeld als bij de alvleesklier weghalen, heeft dat invloed op de vertering van vet? Nou, mijn hypothese is ja, ik denk het wel, want de alvleesklier maakt lipase, dat is een enzym en dat breekt vetten af in de darm. Dat geeft dus de theoretische onderbouwing van jouw veronderstelde antwoord. Een

voorspelling is, als je dat nou vervolgens gaat onderzoeken, je, dat als jouw hypothese klopt, als inderdaad die alvleesklier daar verantwoordelijk voor is, dan zal ik dít zien in mijn resultaat. Namelijk de groep met alvleesklier gaat het goed, en zonder alvleesklier zie ik dat, dat er nog een hoop vetten in de ontlasting zitten of zo. Weet je wel. Dus als een hypothese waar is, dan komt dit uit mijn onderzoek. Dat is voorspelling.

89 Martin: Wat is dan de voorspelling?
90 Docent: Dat is de voorspelling.
91 Martin: Ik snap het verschil niet.
92 Docent: Nou, Jochem weet het.
93 Jochem: Ik snap echt nog steeds het verschil niet, want volgens mij…
94 Martin: Ja, ik snap wel dat het net iets anders is, maar …
95 Carla: Ik denk toch, bij allebei hetzelfde antwoord.
96 Docent: De hypothese is het verwachte antwoord op de onderzoeksvraag.
97 Carla: Ja en de voorspelling toch ook?
98 Docent: Op basis van de theorie, die je kent. Wat je daar van weet.
99 Carla: Ja en de voorspelling?
100 Docent: En de voorspelling is de verwachtte uitkomst van je onderzoek.
101 Martin: Maar moet je dan als hypothese …?
102 Docent: En dat is iets heel anders, dat is heel simpel. Kijk, je zegt van mijn voorspelling is als planten inderdaad, als het klopt, dat mijn hypothese, als mijn hypothese is planten hebben zonlicht nodig om te groeien. Dus ik denk inderdaad dat ze, dan heb je nog een hele uitleg over fotosynthese en weet ik veel wat, dus ik denk inderdaad dat het antwoord op mijn onderzoeksvraag is, ja. Licht stimuleert de groei van planten. Maar je voorspelling is dan, als dat zo is, dan zullen dus die tien potjes, daar zullen de planten langer in zijn, dan in die.
103 Carla: Oh, oké, hèhè.
104 Docent: Nou zo moeten jullie het ook doen. Als jouw hypothese klopt, want komt er en dan in die tabel te staan? Wat ga je dan zien? En dan kun je heel makkelijk, dan is de conclusie ook veel simpeler te trekken. Want dan hoef je maar naar je tabel te kijken en zeggen: ja.
105 Carla: Oké.
106 Docent: Gaan jullie dat doen?
107 [*geen hoorbare reactie, als de docent weg is, praten ze over andere zaken dan hun onderzoek.*] [32]

Wat Lex zegt over hypothese en voorspelling is overigens op zich niet onjuist, maar laat de voorspelling als uitkomst van de toets van de hypothese niet expliciet naar voren komen. Het blijft, ook voor de docenten, een lastig onderscheid. In de (klassikale) bespreking van reflectieopdracht 2 door Bea bleef dat ook onderbelicht. De opvatting dat een hypothese niet meer is dan een idee over het antwoord op je onderzoeksvraag blijkt hardnekkig ('we denken dat het zo is', zie paragraaf 2.3 en Schalk 2002a; 2002c).

Het geheel van de gesprekken overziend komt het beeld naar voren dat de docenten het leeuwendeel van de expliciteringen en generalisaties én de meer volledige redeneringen voor hun rekening komen, in de bespreking van de reflectieopdrachten nog meer dan in de gesprekken over het eigen onderzoek. Deze expliciete aandacht voor de criteria in de bespreking van de opdrachten lijkt door te sijpelen in de docent-leerlinggesprekken rond het eigen onderzoek, maar niet in de gesprekken van de leerlingen onderling. Anders geformuleerd: de nabesprekingen van de reflectieopdrachten geven de docenten, net als de startlessen, de gelegenheid om criteria te expliciteren en te generaliseren.

- *Begrip van bewijs in de boekjes*

Hoe de leerlingen de criteria gebruikt hebben in hun eigen onderzoek, is ook voor een deel te volgen in de aantekeningen van de leerlingen in de boekjes. De leerlingen hebben vooral in het begin van hun onderzoek de boekjes gebruikt: voorlopige onderzoeksvragen, definitieve onderzoeksvraag, hypothese, voorspelling en globale onderzoeksopzet; slechts een paar keer is de precieze onderzoeksopzet erin opgeschreven of zijn de resultaten, conclusie en discussie vermeld.

Te zien is dat leerlingen zeer wel in staat waren om hun *onderzoeksvraag* beter te maken qua eenduidigheid en afperking (begrip van bewijs, element 1), bijvoorbeeld: 'Wat voor invloed …' wordt 'In hoeverre bevordert …' wordt 'Bevordert …' wordt 'In welke mate bevordert ….'. En 'de gisting' wordt eerst veranderd in 'de werking van gist' en daarna in 'de productie van koolzuurgas bij het gistingsproces'. De week erna wordt de vraag uiteindelijk 'Wat is de optimumtemperatuur voor de productie van koolzuurgas bij het gistingsproces in brooddeeg?'

De invloed van de geschreven feedback van de docent is ook te zien: bij de onderzoeksvraag 'Bij welk pH-optimum plant een bacterie zich het snelst voort?' staan haakjes om 'optimum' en 'plant … voort' zijn met een gegolfd lijntje onderstreept; erbij staat 'wat is je afhankelijke variabele?'; er staan pijlen bij 'een' en 'bacterie' en daarbij 'niet specifiek,' erbij staat geschreven: 'welke omstandigheden? de vraag is niet afgeperkt'. Hiermee gingen de leerlingen aan de slag, ze vroegen nog wat verduidelijking aan de docent, maar uiteindelijk sloegen deze leerlingen te ver door in specificiteit en vermeldden erg veel details:

> Bij welke pH groeit de colibacterie het snelst na 19 uur (onder voorwaarde dat de verschillende voedingsbodems geplaatst zijn in een stoof) bij een constante temperatuur van 37 ºC? [33]

Twee leerlingen, Ineke en Bonnie, rapporteerden naast hun artikel uitgebreid over hun proces van onderzoek doen, met name de lange weg van het formuleren van de onderzoeksvraag.

> De voorlopige onderzoeksvraag luidde: 'Welke bacteriën uit de mond worden door bepaalde (bv. fluor – chloride) bestanddelen uit tandpasta gedood?'. Bij deze onderzoeksvraag zouden we bacteriën uit de mond op voedingsbodems kweken en hierbij een aantal verschillende bestanddelen uit tandpasta toevoegen. We wilden eerst de bacteriekolonies identificeren en na toevoeging zouden we bepalen of de bacteriën dood waren. Tijdens overleg met mevrouw B. bleek onze onderzoeksvraag eigenlijk meerdere vragen te bevatten. Met deze vraag bedoel je 'welke bacteriën worden gedood' en 'door welke bestanddelen die bacteriën worden gedood'. Daarom moesten we de vraag dus specificeren tot één duidelijke vraag en eerst aantonen of tandpasta wel echt bacteriedodend is. [34]

De structuur die het boekje biedt, heeft dus effect gehad: eerst ruimte geven voor voorlopige onderzoeksvragen en daarna voor een (voorlopig) definitieve onderzoeksvraag. Leerlingen schaafden en herformuleerden hun vraag een paar keer. Wel bleven leerlingen vaak hangen in beperkte vraagstellingen die in wezen beschrijvend waren en niet hypothesetoetsend, zoals de voorwaarde luidde ('wat is de optimumtemperatuur…' of iets dergelijks). De docenten braken dat niet of met veel moeite open door de leerlingen een verklarende hypothese te laten formuleren.

Doordat de eerste vraag van *reflectieopdracht 1* terugverwijst naar de startles wisten alle leerlingen wel aan te wijzen dat de gepresenteerde onderzoeksvragen rammelen qua eenduidigheid en afperking. Die criteria zijn dan ook bijna altijd opgeschreven. Sommigen maakten ook nog een opmerking over de variabelen (begrip van bewijs, element 8) of de uitvoer-

baarheid (element 4). Slechts weinig leerlingen vulden iets in bij de toepassing op hun eigen onderzoeksvraag; soms staat er 'z.o.z.'of '√'. Ineke en Bonnie bespraken in hun bijlage drie voorlopige versies van hun onderzoeksvraag en gaven ze aan waarom die niet aan de criteria voldeden. Over hun uiteindelijke onderzoeksvraag merkten ze onder meer op:

> Wij vinden dat deze vraag wel aan alle criteria voldoet. Het is eenduidig en duidelijk geformuleerd. Er zitten geen begrippen in waar men verschillende betekenissen aan kan geven. De onderzoeksvraag is specifiek, want het is duidelijk wat we willen onderzoeken. Ook denken we dat de proef uitvoerbaar is en dat we met de beschikbare middelen een goed antwoord op de onderzoeksvraag kunnen geven.
> Verder hebben wij ook maar één afhankelijke en één onafhankelijke variabele. [35]

Reflectieopdracht 1 lijkt dus in ieder geval een positief effect gehad te hebben op de kwaliteit van hun onderzoeksvragen. In hoeverre de leerlingen daarna de criteria voor een onderzoeksvraag beter beheersen, blijft onduidelijk.

De in de gesprekken gesignaleerde onduidelijkheid rond *hypothese en voorspelling* (begrip van bewijs, elementen 3-6) is ook terug te vinden in de boekjes. Soms is er een correct onderscheid, vaak is de hypothese als voorspelling geformuleerd en de voorspelling heeft dan weinig zin meer:

> Bij hypothese:
> *Als je de temperatuur van 10 mL Spa&Fruit Appel-guave invriest of kookt, zal het vitamine C gehalte in de Spa dalen.*
> Bij voorspelling:
> *Wij denken dat onze hypothese klopt.* [36]

Soms leek het de docent te lukken om een beschrijvend onderzoek om te buigen in de richting van een hypothesetoetsend onderzoek door een mogelijke verklaring voor een verwacht verschijnsel te laten formuleren. Maar dat beklijfde toch niet altijd, getuige de voorspelling in volgend fragment:

> Bij hypothese:
> *~~1: ja, jongens typen sneller~~*
> *~~2: nee, meisjes typen sneller~~*
> *~~3: gelijk~~*
> *Ja, jongens kunnen sneller (ch/s) typen omdat ze ~~langer~~ meer uren per dag typend achter de pc zitten.*
> bij voorspelling:
> *'dat er duidelijk te zien is dat de jongens sneller typen.'* [37]

Soms werd de voorspelling ook verward met het verwachte antwoord op de onderzoeksvraag of zelfs de interpretatie ervan, bijvoorbeeld: 'Als de hypothese blijkt te kloppen zitten er op onze school een hoop na-apers' of '[als de hypothese niet klopt] dan klopt de algemene veronderstelling over dat ouderdom gepaard gaat met vergeetachtigheid, niet'. De docenten gaven vrijwel geen geschreven feedback op dit punt.

Reflectieopdracht 2 is bedoeld om meer duidelijkheid te scheppen over het toetsen van hypotheses en het doen van voorspellingen. Het blijkt moeilijk voor de leerlingen, zoals al in paragraaf 7.4 werd vermeld. Het idee dat je verwachte waarnemingen moet vergelijken met daadwerkelijk gedane waarnemingen om zo hypothese te kunnen verwerpen of niet, is voor leerlingen vooral 'kijken of het werkt' (Carey e.a., 1989) en veel minder het ontwerpen van een waarnemingssituatie c.q. experiment om de verwachte waarnemingen al dan niet te kunnen doen. Ondanks dat de meeste leerlingen aan het eind van de opdracht wel correct op-

schreven aan welke criteria een hypothese moet voldoen en sommigen ook vermeldden dat hun eigen hypothese eraan voldeed, is twijfel gerechtvaardigd of ze de essentie van de hypothetisch-deductieve aanpak hebben opgepikt. Een citaat uit de eerder genoemde toelichting van Ineke en Bonnie illustreert dat:

> Wij denken dat onze twee hypothesen voldoen aan de gestelde criteria. De hypothesen zijn toetsbaar en geven een mogelijk antwoord op de onderzoeksvraag. De hypothesen zijn weliswaar tegengesteld, maar toch achtten wij beide mogelijk als antwoord op de onderzoeksvraag. Voor beide redeneringen is wat te zeggen. Verder zijn de hypothesen eenduidig en het blijkt dus dat er meerdere hypothesen mogelijk zijn. [38]

Hun hypotheses luidden:

> Hypothese 1: Na eenmalig gebruik van Listerine zal de bacteriedodende werking van dit mondwater op de tanden in een uur zijn afgenomen. Gedurende het uur zal de hoeveelheid bacteriën op de tanden toenemen. (…)
> Hypothese 2: Na eenmalig gebruik van Listerine zal de bacteriedodende werking van dit mondwater op de tanden in een uur niet zijn afgenomen. Gedurende een uur zal de hoeveelheid bacteriën ongeveer gelijk blijven. [39]

Dat zij met hun tegengestelde hypotheses dachten aan het criterium ‘vaak zijn er meer hypotheses mogelijk’ voldaan te hebben, is tekenend voor het onbegrip. In feite waren ze één hypothese aan het toetsen en lijkt het formuleren van elkaar complementerende hypotheses op toetsing van een ‘nul-hypothese’ uit de statistiek. Die verwarring is ook onder onderwijsgevenden wijd verspreid (McPherson, 2001; Schalk, 2002a, 2002c).

Eén team van leerlingen formuleerde bij de onderzoeksvraag ‘Op welke van de deurgrepen in het Noorderzijtrappenhuis zijn de meeste bacteriën te vinden?’ maar liefst zeven naast elkaar staande hypotheses (in verkorte vorm weergegeven):

> Hypothese 1: op de derde etage, omdat het er rustig is en daar iedereen zelf de deur open doet, op de andere etages is het drukker en blijft de deur openstaan.
> Hypothese 2: op de derde etage, omdat het er rustig is en de bacteriën daar meer gelegenheid hebben zich te vermenigvuldigen.
> Hypothese 3: op de etages waar veel mensen langskomen, zullen meer bacteriën te vinden zijn.
> Hypothese 4: op de derde etage, omdat daar meer practica plaatsvinden en mensen dus viezere handen hebben.
> Hypothese 5: op de tweede etage, omdat daar de mediatheek is.
> Hypothese 6: op de tweede etage omdat dat de lokalen voor Engels zijn en iedereen Engels in zijn pakket heeft.
> Hypothese 7: op de begane grond, want daar zijn meer lokalen en dus meer mensen. [40]

Helaas zetten ze dat onderzoek niet zo op het dat uitsluitsel kon brengen over welk van de hypotheses het meest waarschijnlijk is. Ook hen lijkt de essentie te zijn ontgaan.

De aantekeningen bij *globale onderzoeksopzet* laten zien hoe de leerlingen omgingen met de afhankelijke en onafhankelijke variabelen (begrip van bewijs, element 8), dat bleef namelijk vrijwel altijd impliciet, bijvoorbeeld: ‘Met een banaan controleren of hij sneller verrot in een steriele / niet-steriele omgeving’. Een enkele keer, of als de docent er in het commentaar naar vroeg, werd het expliciet benoemd (‘We nemen waarschijnlijk pH = 1, 3, 5, 7, 9 → onafhankelijke’), maar dan nog niet altijd correct:

> onafh. → de gevoelens bij kleuren volgens eerdere psychologische tests
> afh. → gekozen kleur en gevoel [41]

In het conceptartikel is dit wel gecorrigeerd door deze leerlingen.

Het onder controle houden of randomiseren van andere variabelen (begrip van bewijs, elementen 9 en 14), blijft ook meestal impliciet (bijvoorbeeld 'van elke buis een even grote hoeveelheid'), maar soms wordt het explicieter, zoals het vermelden dat het om gelijke hoeveelheden brooddeeg gaat en dat de temperatuur waarschijnlijk de enige beperkende factor is of zoals in het volgende fragment:

> Als je de gem. snelheid voor de inname van cafeïne vergelijkt met de gem. snelheid na de inname van cafeïne. Zo weten we zeker dat we geen last hebben van individuele verschillen in gewenning aan cafeïne → waardoor bij sommigen de concentratie minder wordt beïnvloed dan andere omdat ze al meer gewend zijn aan cafeïne [42]

In de boekjes wordt bij de globale onderzoeksopzet geen enkele keer aan een controle of blanco gerefereerd (begrip van bewijs, element 10).

De opzet van het boekje om eerst globaal te beschrijven wat je wilt gaan onderzoeken, blijkt dus redelijk effectief. De leerlingen benoemden welke variabelen ze in hun onderzoek betrekken, al gebruikten ze de term *(on)afhankelijke variabele* hoogst zelden expliciet. Slechts af en toe somden ze gedetailleerd de benodigdheden op.

In *reflectieopdracht 3* werd het niet onder controle houden van variabelen of een verkeerde steekproef (element 9 en 14) in het voorbeeldonderzoek het vaakst als tekortkoming gesignaleerd, naast het niet beantwoorden van de onderzoeksvraag (element 18). Minder vaak werd opgemerkt dat het geen betrouwbaar of valide onderzoek was (element 21). Dat er eigenlijk sprake was van twee onafhankelijke variabelen (wel of geen cafeïne, wel of geen koffie, element 8) werd ook een paar keer genoemd, evenals het onterecht middelen van de resultaten (element 16). Opmerkelijk is dat bij het expliciet noemen van criteria enkele leerlingen terugvielen op eerdere criteria als het eenduidig formuleren en afperken van de vraag (element 1).

Het lijkt erop dat leerlingen de criteria rond 'eerlijk meten' (element 7-17) redelijk makkelijk in hun onderzoek toepasten. Ze beseften dat het nodig was, op de concrete toepassing ervan moesten ze soms gewezen worden.

Zoals al opgemerkt, staan er weinig aantekeningen in de boekjes met betrekking tot *de resultaten, de conclusie en de discussie*. Alleen al op basis van die constatering kan gesteld worden dat de boekjes in die fase van het onderzoek hun functie grotendeels verloren hadden (zie ook par. 7.4). Toch waren er enkele leerlingen die wel iets opschreven over het slot van hun onderzoek. Onder 'discussie' werden soms factoren opgesomd die de kwaliteit van de resultaten beïnvloed hebben:

> - Onderzoek moest 2 keer gedaan worden doordat het verdunnen van de bacterie is onderschat.
> - Bij tweede onderzoek waren 3 v.d. 6 petrischaaltjes telbaar, dat is beter dan het eerste onderzoek, maar niet het gewenste resultaat
> - De precisie v.d. resultaten is niet nauwkeurig. Want we hebben 3 schaaltjes moeten schatten, en dat is geen exact resultaat, niet precies.
> - Er is helaas niet steriel genoeg gewerkt, dat kwam door slecht materiaal.
> - We hadden tijdnood. Omdat er zoveel bacteriën waren en te weinig TOA's werd er te weinig aandacht besteed aan hoeveel keer verdunnen in het eerste onderzoek en ons 2e onderzoek liep hierdoor uit. [43]

De evaluatiefase van het eigen onderzoek tijdens de uitvoering, een belangrijk deel van de *reflection-in-action*, is dus in de boekjes niet goed te volgen. Alleen in het onderling bespreken van de concepten van de eindproducten, de vierde reflectieopdracht voor de expli-

ciete groep, werd het zichtbaar. Maar alvorens de effecten van die opdracht te bespreken, gaan we eerst in op de concepten zelf.

- *Begrip van bewijs in de voorlopige producten*

De conceptartikelen zijn door de leerlingen gemaakt op basis van hun aantekeningen tijdens het doen van hun onderzoek en de gesprekken die ze daarbij met elkaar en met de docent gevoerd hebben. Deze voorlopige producten laten zien tot welk begrip van bewijs dat de leerlingen gebracht heeft.

De conceptartikelen van de leerlingen van de impliciete groep uit Amstelveen en van alle leerlingen uit Amsterdam zijn geanalyseerd op het voorkomen van elementen van begrip van bewijs, op dezelfde manier als de verslagen van leerling-onderzoek (par. 3.4). Daarbij wordt van elk element, als het van toepassing is, beoordeeld of aan dat criterium is voldaan, of er beargumenteerd niet aan is voldaan of dat er niet-beargumenteerd niet aan is voldaan. De eerste twee gevallen worden beschouwd als de aanwezigheid van 'begrip van bewijs'.

En dan blijkt dat over alle elementen samen genomen in 63 procent van de gevallen dat een element van toepassing is in de conceptartikelen van 'begrip van bewijs' gesproken kan worden (tabel 7.7). Dat is het gemiddelde over 23 elementen. Een gedetailleerdere blik laat grote verschillen zien. Het eenduidig en specifiek formuleren van een onderzoeksvraag en een toetsbare hypothese ging de leerlingen goed af. Daar is ook veel aandacht aan besteed in de opdracht. Het naast elkaar zetten van alternatieve hypotheses en het doen van een voorspelling ging echter veel minder goed. In het voorgaande werd al beschreven dat dat de leerlingen ook in de gesprekken niet altijd duidelijk werd.

De elementen met betrekking tot variabelen, steekproef en metingen laten ook een gevarieerd beeld zien. Wat opvalt zijn de lage scores voor de elementen 10 en 15. Element 10, het inzetten van een controle-experiment of blanco, is wel opgenomen in de selectie van elementen die in de praktische opdracht aan de orde zouden moeten komen (zie par. 3.5), maar is niet speciaal uitgewerkt in het boekje of in de reflectieopdrachten (hoewel het in reflectieopdracht 3 aan de orde zou kunnen komen). Bij het bespreken van de aantekeningen in de boekjes is ook al opgemerkt dat er nergens een opmerking over is gemaakt. Element 15, omgaan met uit de toon vallende resultaten, is evenmin expliciet aan de orde geweest.

Begrip van bewijs in het trekken van de conclusie en het evalueren van het onderzoek vond slechts in 40 procent van de gevallen plaats. Het trekken van een bij de onderzoeksvraag aansluitende conclusie ging meestal wel (element 18, 83 procent), maar het beschouwen van de vraag 'zeker weten', met uitspraken over de betrouwbaarheid en validiteit van het onderzoek (element 21) lukte minder (54 procent), terwijl het vergelijken van de resultaten met ander onderzoek of theorie (element 22 en 23) slechts sporadisch voorkwam. Ten aanzien van element 21 blijkt echter wel een opmerkelijk verschil tussen de expliciete en de impliciete groep op te treden: in de expliciete groep werd in 75 procent van de gevallen iets over betrouwbaarheid, validiteit en nauwkeurigheid opgemerkt, in de impliciete groep in maar 44 procent. Een voorbeeld:

> Als je kritisch naar het onderzoek kijkt, kun je nog heel wat kritiekpuntjes noemen. Deze kritiekpuntjes maken het onderzoek toch niet zo betrouwbaar als vooraf de bedoeling was.
> Dit is bijvoorbeeld het geval met de enquête die we hebben gehouden [zie bijlage]. Hierin staat bijvoorbeeld de vraag: 'Denk je dat dit van invloed is geweest op het beginnen met roken?' In deze vraag komt 'denk je' voor. Je vraagt dus naar de gedachte én de mening van de geënquêteerde.[…]
> Achteraf kunnen wij zeggen dat dit onderzoek misschien iets te hoog gegrepen was voor ons.
> [44]

Wellicht heeft het vaker dan wel explicieter aan de orde stellen van criteria voor goed onderzoek, onder andere in de reflectieopdrachten, hierop een positieve invloed gehad.

Tabel 7.7 Begrip van bewijs in de conceptartikelen van de onderzoeksteams (N = 15). 'Begrip van bewijs' wil zeggen dat aan een van toepassing zijnd criterium is voldaan of beargumenteerd is waarom er niet aan is voldaan.

element	omschrijving	aantal keer van toepassing	aantal keer 'begrip van bewijs'	percentage 'begrip van bewijs'
1	formuleren onderzoeksvraag	43	36	84
2	onderscheid beschrijvend / hypothesetoetsend onderzoek	15	0	0
3	formuleren hypothese	54	44	81
4	toetsbaarheid hypothese en uitvoerbaarheid	13	11	85
5	meer dan één hypothese	14	2	14
6	voorspelling	15	5	33
1-6	*onderzoeksvraag, hypothese en voorspelling*	*154*	*98*	*64*
7	kenmerken bij beschrijvend onderzoek	1	1	100
8	onafhankelijke en afhankelijke variabele	14	11	79
9	controlevariabelen	15	11	73
10	blanco	10	2	20
11	invloed metingen op uitkomst	12	12	100
12	intervallen van variabelen	12	7	58
13	gevoeligheid meetinstrument	15	13	87
14	steekproef aselect en groot genoeg	33	23	70
15	omgaan met uit de toon vallende resultaten	4	1	25
16	verantwoord middelen	5	5	100
17	onderkennen significantie	12	8	67
7-17	*variabelen, steekproef en metingen*	*133*	*94*	*71*
18	conclusie sluit aan op onderzoeksvraag	12	10	83
19	onderscheid correlatie en causaal verband	4	2	50
20	verschil oorzakelijke en functionele verklaring	2	0	0
21	uitspraken over betrouwbaarheid en validiteit	13	7	54
22	vergelijking met ander onderzoek	11	1	9
23	vergelijking met theorie	10	1	10
18-23	*conclusie en evaluatie*	*52*	*21*	*40*
1-23	*alle elementen*	*339*	*213*	*63*

Zoals in 7.4.9. al gemeld, lag in het commentaar dat de docenten gaven op de conceptartikelen (in de impliciete groep) ook niet de nadruk op betrouwbaarheid en validiteit of op 'zeker weten?' Welgeteld twee keer werd er een opmerking in die richting gemaakt. De meeste opmerkingen betroffen de onderzoeksvraag, de variabelen of de opbouw en stijl van het conceptartikel.

In de tekst van *reflectieopdracht 4* is expliciet verwoord hoe de leerlingen naar het conceptartikel van een ander onderzoeksteam geacht werden te kijken: 'Concentreer je op de vraag: hoe zeker is het dat het antwoord op de onderzoeksvraag juist is? Denk daarbij aan de criteria voor goed onderzoek.' Ook in de introducties van de opdracht werd dat accent gelegd. Buiten die introducties is die vraag echter nergens in de transcripten terug te vinden. Element 21 kwam in de transcripten maar in twee patronen voor, beide in een gesprek van leerlingen met de docent.

De meeste leerlingen in Amsterdam schreven in hun boekjes dat ze vonden dat ze niet erg goed of bruikbaar commentaar kregen, één team noemde het commentaar zelfs, 'niet behoorlijk'. Er werd volgens hen teveel op futiliteiten, punten en komma's gelet. Toch concludeerde één leerling:

> Van het positieve commentaar hebben we geleerd wat we in het vervolg weer moeten. Van het negatieve commentaar heb ik geleerd dat voor ik het inlever er nog iemand anders naar laat kijken. [45]

In Amstelveen schreven de leerlingen helemaal niet op wat ze van de opdracht geleerd hebben, alleen Ineke en Bonnie, in de bijlage bij hun artikel. Ze beschreven welke criteria ze hebben toegepast op het door hen bekeken onderzoek en dat ze naar aanleiding daarvan, maar nog voor de bespreking, al gereflecteerd hadden op hun eigen onderzoek. Ook beschreven ze op welke punten ze commentaar hebben gekregen en wat ze ermee gedaan hebben. Ze concludeerden met:

> Deze reflectieopdracht vonden we erg nuttig. We hebben er kritischer door naar ons eigen onderzoek gekeken. [46]

Ook in de interviews met de gevolgde duo's achteraf gaven leerlingen aan dat ze wederzijds commentaar leerzaam en nuttig vonden. Een direct effect op het bewaken van de kwaliteit van hun onderzoek is echter alleen waar te nemen door vergelijking met de eindproducten.

7.5.3 De eindproducten

De definitieve artikelen zijn op dezelfde manier geanalyseerd als de concepten (tabel 7.8). De totale score ligt iets hoger dan in de concepten (68 tegen 63 procent), waarbij het grootste verschil optreedt met betrekking tot de conclusie en evaluatie (52 tegen 40 procent). Vergelijking van tabel 7.8 met tabel 7.7 in meer detail geeft de grootste verschillen bij de elementen 8, 15 en 21 te zien. Ook als alleen gekeken wordt naar de artikelen waarvan zowel het concept als de definitieve versie ingeleverd is (gegevens niet weergegeven), zijn dit de elementen waar de meeste vooruitgang plaatsvindt. Bespreking van de conceptartikelen met medeleerlingen (de expliciete aanpak) of het ontvangen van commentaar van de docent (de impliciete aanpak) leidt dus wel tot verbetering en daarmee blijken beide strategieën in deze fase effectief.

Als we naar de onderlinge verhouding van de scores voor de verschillende elementen voor de definitieve artikelen kijken – en met name naar de selectie ervan: 1, 3, 4, 5, 6, 8, 9, 14, 18 en 21 – vallen enkele getallen op. De scores bij element 1, 4, 8, 9 en 18 zijn relatief hoog (boven de 75 procent). Voor de elementen 1 (onderzoeksvraag), 4 (toetsbaarheid en uitvoerbaarheid), 8 en 9 (variabelen) is dat in overeenstemming met de verwachtingen en het beeld uit de gesprekken. Dat ook 18 (aansluiting van conclusie bij onderzoeksvraag) hoog scoort, wekt geen verbazing, eerder wellicht dat het geen honderd procent is.

De scores op element 5, 6, 10 en 14 zijn laag, net als bij de conceptartikelen. Zoals in het voorgaande echter duidelijk werd, kwam het begrip van bewijs met betrekking tot de relatie tussen hypothese en voorspelling (6) niet uit de verf in de gesprekken en daarmee werd ook het idee van naast elkaar bestaande hypotheses (5) gemist. Dat er wel op element 3 goed gescoord is, komt omdat het daarbij gaat om aansluiting van hypothese op onderzoeksvraag en de precieze formulering ervan. In het voorgaande bleek ook dat nergens gerichte aandacht is besteed aan de functie en opzet van een blanco (10).

Tabel 7.8 Begrip van bewijs in de artikelen van de onderzoeksteams (N = 23). 'Begrip van bewijs' wil zeggen dat aan een van toepassing zijnd criterium is voldaan of beargumenteerd is waarom er niet aan is voldaan.

element	*omschrijving*	*aantal keer van toepassing*	*aantal keer 'begrip van bewijs'*	*percentage 'begrip van bewijs'*
1	formuleren onderzoeksvraag	67	61	91
2	onderscheid beschrijvend / hypothesetoetsend onderzoek	23	1	4
3	formuleren hypothese	89	77	87
4	toetsbaarheid hypothese en uitvoerbaarheid	22	19	86
5	meer dan één hypothese	21	7	33
6	voorspelling	23	9	39
1-6	*onderzoeksvraag, hypothese en voorspelling*	*245*	*174*	*71*
7	kenmerken bij beschrijvend onderzoek	1	1	100
8	onafhankelijke en afhankelijke variabele	22	22	100
9	controlevariabelen	23	18	78
10	blanco	16	4	25
11	invloed metingen op uitkomst	23	20	87
12	intervallen van variabelen	19	13	68
13	gevoeligheid meetinstrument	23	21	91
14	steekproef aselect en groot genoeg	52	28	54
15	omgaan met uit de toon vallende resultaten	12	8	67
16	verantwoord middelen	11	10	91
17	onderkennen significantie	19	12	63
7-17	*variabelen, steekproef en metingen*	*220*	*157*	*71*
18	conclusie sluit aan op onderzoeksvraag	23	21	91
19	onderscheid correlatie en causaal verband	10	7	70
20	verschil oorzakelijke en functionele verklaring	4	0	0
21	uitspraken over betrouwbaarheid en validiteit	23	17	74
22	vergelijking met ander onderzoek	23	5	22
23	vergelijking met theorie	22	5	23
18-23	*conclusie en evaluatie*	*105*	*55*	*52*
1-23	*alle elementen*	*571*	*386*	*68*

Wat opvallend is, is de aanzienlijke hogere score op element 21 (validiteit en betrouwbaarheid, net onder de 75 procent) in vergelijking met de conceptartikelen. De aandacht voor 'zeker weten' in de praktische opdracht en met name in de laatste feedback – van de docent of van medeleerlingen – heeft vruchten afgeworpen, zo blijkt.

Twee elementen buiten de selectie vallen op, 13 (de gevoeligheid van het meetinstrument) en 16 (verantwoord middelen). Het zijn voorbeelden van de problemen waar de leerlingen tegenaan lopen (hoe tel je bacteriekolonies als de voedingsbodem bijna vol zit?) of waar de docent toch een opmerking over maakt ('waarom neem je daar percentages?').

Tabel 7.9 Percentages begrip van bewijs voor verschillende combinaties van elementen (gerealiseerd ten opzichte van mogelijk) in de artikelen in de verschillende groepen en in de voorstudie (par. 3.4). *Boven*: voor alle elementen; *onder*: voor de geselecteerde elementen. AS: Amsterdamse School; SA: School in Amstelveen.

	expliciet (N=12)	*impliciet (N=11)*	*AS (N=10)*	*SA (N=13)*	*Totaal (N=23)*	*voor-studie*
Onderzoeksvraag, hypothese & voorspelling (1-6)	66	76	71	71	71	67
Variabelen, steekproef & metingen (7-17)	67	76	75	67	71	64
Conclusie & evaluatie (18-23)	49	56	46	58	52	43
Alle elementen	63	72	68	67	68	61
Onderzoeksvraag, hypothese & voorspelling (1, 3, 4, 5, 6)	74	80	78	77	77	74
Variabelen, steekproef & metingen (8, 9, 10, 14)	57	72	65	63	64	64
Conclusie & evaluatie (18, 21)	79	86	80	85	83	72
Alle onderwezen elementen	69	78	73	73	74	70

De verschillen tussen de groepen en de scholen zijn niet heel erg groot, waarbij ook aangetekend moet worden dat door de kleine aantallen de score van één artikel flinke invloed heeft op de getoonde percentages. Dat in het achterhoofd houdend, kan geconcludeerd worden dat de grootste verschillen tussen de expliciete en de impliciete groep zitten in het voordeel van de laatste (tabel 7.9). Over de hele linie deden de leerlingen in de impliciete groep het iets beter in de artikelen (72 tegen 63 procent begrip van bewijs), en dat uitte zich zowel in de fase van onderzoeksvraag, hypothese en voorspelling (76 tegen 66 procent) en de fase van de onderzoeksopzet (variabelen en metingen, 76 tegen 67) als in die van conclusie en evaluatie (56 tegen 49). Zou dit betekenen dat de leerlingen in die groep, doordat ze alle tijd aan hun eigen onderzoek konden besteden en niet aan reflectieopdrachten, de criteria beter opgepikt hebben? Of zat het vooral in de geschreven feedback van de docent?

Wat de scholen betreft is de situatie divers. In Amsterdam is beter gescoord met betrekking tot variabelen en metingen (75 tegen 67), in Amstelveen met betrekking tot de conclusie en evaluatie (58 tegen 46). In totaal ontlopen de scores elkaar nauwelijks (68 tegen 67). Als alleen gekeken word naar de 'onderwezen' elementen (onderste deel van tabel 7.9), is de situatie niet anders.

Net als bij de beoordeling van de gesprekken over de elementen van begrip van bewijs, is de vraag wanneer er gesproken kan worden van een *voldoende* resultaat; bij welk percentage leerlingen is er voldoende gescoord voor een element en bij welk percentage elementen heeft een leerling voldoende gescoord? Elke grens is in wezen arbitrair, daarom is in tabel 7.10 een overzicht gegeven van aantallen elementen en leerlingen met een voldoende resultaat behaald bij verschillende cesuren. Als we de strenge cesuur hanteren dat 75 procent van de

Tabel 7.10 Aantallen elementen en artikelen die voldoende scoren bij verschillende cesuren.

Grens onvoldoende – voldoende (%)	*Aantal elementen met voldoende score (N = 23)*	*Aantal artikelen met voldoende score (N = 23)*	*Aantal artikelen in expliciete groep met voldoende score (N = 12)*	*Aantal artikelen in impliciete groep met voldoende score (N = 11)*
49 / 50	16	21 (96%)	10 (83%)	11 (100%)
59 / 60	15	19 (87%)	9 (75%)	10 (91%)
66 / 67	14	13 (52%)	4 (33%)	9 (82%)
74 / 75	10	5 (22%)	2 (17%)	3 (27%)

leerlingen begrip van bewijs moet vertonen met betrekking tot een element, dan komen slechts tien elementen daarboven (tabel 7.8, de nummers 1, 3, 4, 7, 8, 9, 11, 13, 16 en 18). Van de beoogde elementen benaderen 5, 6, 10 en 14 die grens niet; element 21 haalt het net niet.

Voor leerlingen wordt meestal een minder strenge grens gehanteerd, soms zelfs de coulante 50 procent (Van Rens, 2005). Dan zijn er slechts twee artikelen waarin te weinig begrip van bewijs getoond wordt. Bij een iets hogere eis worden dat er uiteraard meer; bij een eis van tweederde van het mogelijk begrip van bewijs scoort bijna de helft van de artikelen onvoldoende. Ook uit deze analyse blijkt dat de impliciete groep het beter deed dan de expliciete.

Behalve vergelijking van de beoordeling van de definitieve artikelen met de concepten, is het zinvol te kijken naar de resultaten van de voorstudie van dit onderzoek (zie par. 3.4 en tabel 7.9, rechter kolom). Allereerst moet dan worden opgemerkt dat het in de voorstudie ging om verslagen van praktische schoolonderzoeken en profielwerkstukken van leerlingen uit 6 vwo. Helaas konden de profielwerkstukken van de leerlingen die de praktische opdracht 'Zeker weten?' hebben gemaakt, niet verzameld worden, dus moeten we het doen met deze 'scheve' vergelijking. En dan blijkt dat de leerlingen uit dit onderzoek het beter doen dan die 6-vwo-leerlingen (68 tegen 61 procent). Dat geldt het sterkst voor de conclusie en evaluatie: een score van 52 procent tegenover 43 in de voorstudie. En dat verschil wordt nog groter als alleen element 18 en 21 gekeken wordt. Dan is het percentage begrip van bewijs in alle groepen hoger dan de 72 procent uit de voorstudie. Dat wijst erop dat het leggen van accent op de vraag naar betrouwbaarheid en validiteit vruchten afwerpt, in ieder geval daar waar het dit type evaluerende opmerkingen over het onderzoek betreft.

Ook als we kijken naar de aantallen artikelen met voldoende score, doet de groep als geheel het beter dan de leerlingen in de voorstudie (vgl. tabel 7.10 met tabel 3.8), evenals de impliciete groep. De getallen voor de expliciete groep liggen bijna allemaal iets lager dan in de voorstudie.

7.5.4 Begin- en eindtoets

Voor en na de praktische opdracht hebben de leerlingen een toets gemaakt waarin dezelfde elementen van begrip van bewijs aan de orde kwamen die ook in de praktische opdracht een belangrijke rol spelen. In paragraaf 7.3 is de betrouwbaarheid van de toets besproken. Die laat zeker te wensen over, waardoor er voorzichtigheid betracht moet worden bij het toekennen van betekenis aan de scores. De toets is te vinden in bijlage 10.

Sommige opgaven in de toets waren gefocust op één element, andere opgaven konden vanuit meer elementen beantwoord worden. De antwoorden zijn gecodeerd en als er een relevant element van begrip van bewijs op een juiste manier in het antwoord betrokken is, is er daarvoor een scorepunt toegekend. Uit de resultaten is dus zowel iets af te lezen over het al dan niet correct beantwoorden van de opgaven als over de inzet van de verschillende elementen van begrip van bewijs daarbij. De toetsresultaten zijn zowel voor de gehele toets bepaald als voor delen ervan. Opgaven waar het accent lag op het opstellen of beoordelen van onderzoeksvraag, hypothese of voorspelling vormden deeltoets 1, opgaven over variabelen, blanco, steekproef en het correct uitvoeren van metingen vormden deeltoets 2 en opgaven waarbij de kwaliteit van een compleet onderzoek moest worden beoordeeld vormden deeltoets 3.

- *Begrip van bewijs*

In tabel 7.11 wordt een overzicht gegeven van de mate waarin de leerlingen elementen van begrip van bewijs hebben gebruikt in hun antwoorden. In alle groepen is dan sprake van een lichte vooruitgang, in de expliciete iets meer dan in de impliciete en in Amstelveen iets meer dan in Amsterdam. De percentages voor de elementen 1 tot en met 6 liggen gemiddeld hoger dan voor de elementen die met latere fasen van onderzoek te maken hebben, met uitzondering van de toetsbaarheid (element 4). Het betekent dat in hun beoordeling van onderzoek van anderen leerlingen eerder iets opmerkten over de vraagstelling dan over het design. Algemene beoordeling in de trant van 'niet valide' of 'niet betrouwbaar' (element 21) werden zelden gemaakt.

Tabel 7.11 Percentage keren dat een bepaald element van begrip van bewijs gebruikt in de beantwoording werd (van het aantal keren dat dat element gebruikt had kunnen worden) in voor- en nameting voor alle leerlingen, de expliciete en de impliciete groep en de Amsterdamse School (AS) en de School in Amstelveen (SA).

	totaal		*expliciet*		*impliciet*		*AS*		*SA*	
	(N = 41)		*(N = 20)*		*(N = 21)*		*(N = 15)*		*(N = 26)*	
element	*voor (%)*	*na (%)*	*voor (%)*	*na (%)*	*voor (%)*	*na (%)*	*voor (%)*	*na (%)*	*voor (%)*	*na (%)*
1	57	69	59	74	54	65	44	64	64	72
3	85	93	75	90	95	95	100	87	77	96
4	33	36	28	32	37	39	41	44	28	31
6	62	60	57	59	67	60	62	60	62	59
1-6	*62*	*72*	*57*	*68*	*67*	*76*	*62*	*73*	*62*	*72*
8	38	46	31	46	45	45	40	42	38	48
9	20	34	25	43	15	25	15	15	23	44
10	22	22	27	28	17	16	27	12	19	28
14	38	44	40	47	37	41	44	51	35	40
7-17	*29*	*31*	*30*	*32*	*27*	*30*	*30*	*29*	*28*	*32*
18	25	31	25	32	25	30	22	29	27	32
21	5	7	4	5	6	8	2	7	7	7
18-23	*14*	*17*	*13*	*16*	*14*	*18*	*10*	*16*	*15*	*18*
totaal	*34*	*41*	*34*	*43*	*35*	*39*	*34*	*38*	*34*	*43*

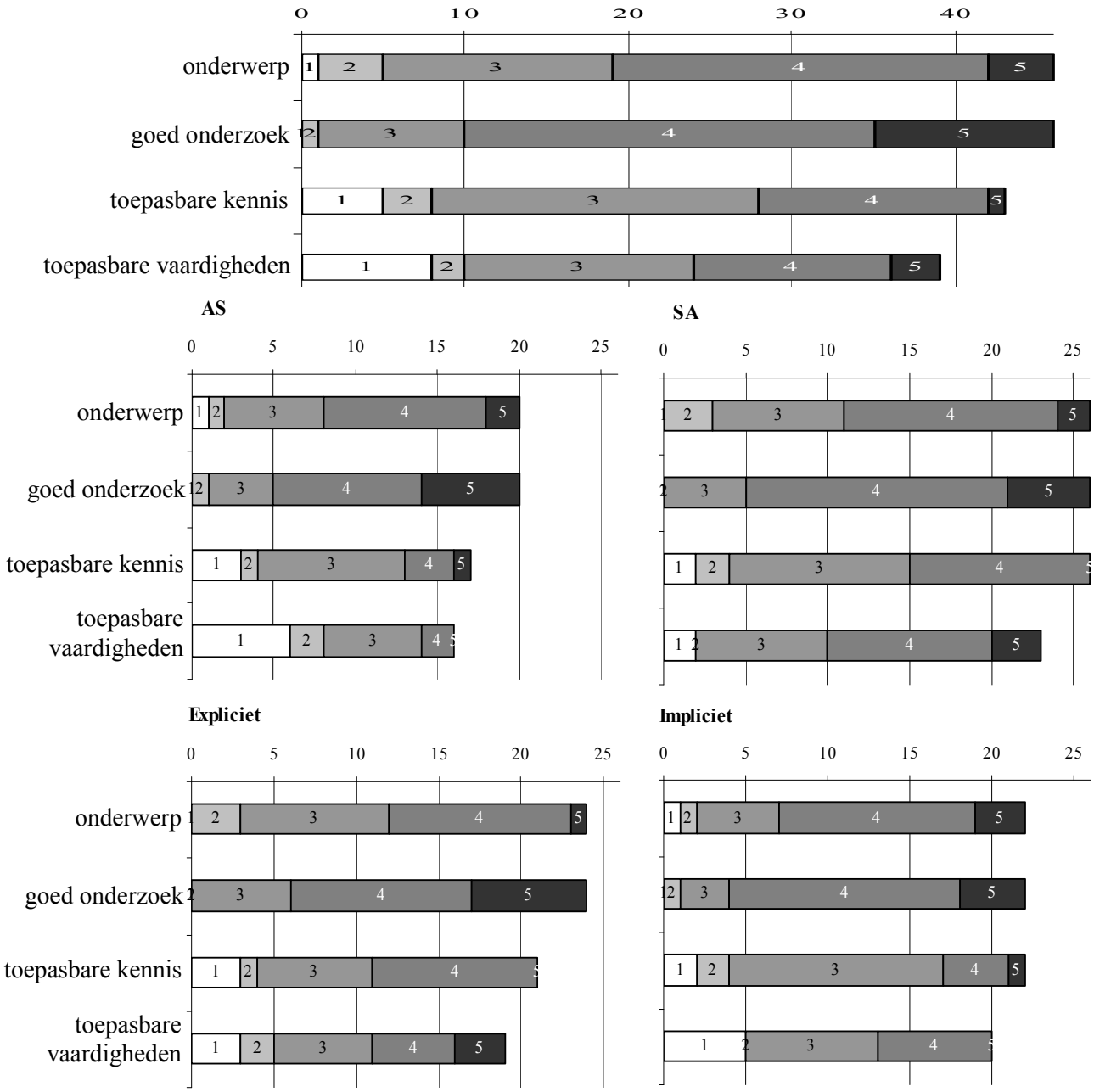
Tabel 7.12 Gemiddelde scores en p'-waarden (fractie van maximumscore) op de gehele toets en de drie deeltoetsen in voor- en nameting voor alle leerlingen, de expliciete en de impliciete groep en de Amsterdamse School (AS) en de School in Amstelveen (SA). *: significant beter dan de voortoets ($p < 0,05$).

VOOR		*Totaal* *(N = 41)*	*Expliciet* *(N = 20)*	*Impliciet* *(N = 21)*	*AS* *(N = 15)*	*SA* *(N = 26)*
Hele toets	gem. score	14,07	13,90	14,24	14,33	13,92
(max. 27)	p'	0,52	0,51	0,53	0,53	0,52
deeltoets 1	gem. score	8,32	7,95	8,67	8,40	8,27
(max. 12)	p'	0,69	0,66	0,72	0,70	0,69
deeltoets 2	gem. score	2,80	2,85	2,76	2,93	2,73
(max. 8)	p'	0,35	0,36	0,35	0,37	0,34
deeltoets 3	gem. score	2,95	3,10	2,81	3,00	2,92
(max. 7)	p'	0,42	0,44	0,40	0,43	0,42
NA						
Hele toets	gem. score	17,05*	17,70*	16,43*	15,87	17,73*
(max. 27)	p'	0,63	0,66	0,61	0,59	0,66
deeltoets 1	gem. score	9,51*	9,40*	9,62*	9,53	9,50*
(max. 12)	p'	0,79	0,78	0,80	0,79	0,79
deeltoets 2	gem. score	4,05*	4,60*	3,52	2,73	4,81*
(max. 8)	p'	0,51	0,58	0,44	0,34	0,60
deeltoets 3	gem. score	3,49*	3,70	3,29	3,60	3,42*
(max. 7)	p'	0,50	0,53	0,47	0,51	0,49

- *Scores*

Voor het toekennen van een scorepunt was echter niet altijd van belang wélk element van begrip van bewijs in de beantwoording gebruikt werd, maar óf er een (relevant) element gebruikt werd. Daarom geven de scores een iets ander beeld, hoewel de grote lijn uiteraard overeenkomt. In tabel 7.12 staan de berekende gemiddelde scores en p'-waarden (gemiddelde behaalde score gedeeld door maximumscore). Te zien is dat de verschillen tussen de voor- en de nameting klein zijn, maar met uitzondering van de Amsterdamse School is de vooruitgang ten opzichte van de voortoets significant (t-test voor gepaarde waarnemingen, $p<0,05$), ook in de meeste gevallen voor de deeltoetsen. Er is een aanwijsbare, maar kleine leerwinst opgetreden ten aanzien van de criteria voor goed onderzoek. Om het in schooltermen uit te drukken: ze zijn van een vijf naar een zesje gegaan. Nu zegt de absolute hoogte van de score niets over voldoende of onvoldoende zonder dat er inhoudelijk een grens daarvoor is aangegeven, maar duidelijk is dat er niet spectaculair veel vooruitgang is geboekt. Het zelf uitvoeren van een onderzoek en het reflecteren op de kwaliteit ervan hebben de leerlingen blijkbaar niet heel veel extra's meegegeven om opgaven te maken waarin ze criteria met betrekking tot die kwaliteit moesten toepassen.

In Amstelveen lijkt op alle fronten dus meer geleerd te zijn dan in Amsterdam, al is het onderlinge verschil tussen de scores net niet significant (t-test voor onafhankelijke steekproeven voor de gehele natoets, $t(39)=-1,93$, $p=0,06$), alleen wel voor deeltoets 2 in de nameting ($t(39)=-4,67$, $p<0,05$). Het is opvallend dat bij de opgaven met betrekking tot variabelen, steekproef en metingen (deeltoets 2) de leerlingen in Amsterdam in de nameting slechter scoorden dan in de voormeting, terwijl in Amstelveen juist op dat onderdeel de meeste



Schaal	*Totaal*	*AS*	*SA*	*Expliciet*	*Impliciet*
het onderwerp van onderzoek	3,5 (0,8) *(N =46)*	3,6 (0,9) *(N=20)*	3,5 (0,8) *(N=26)*	3,4 (0,8) *(N=24)*	3,7 (0,9) *(N=22)*
wat goed onderzoek doen is	4,0 (0,7) *(N =46)*	4,0 (0,8) *(N=20)*	4,0 (0,6) *(N=26)*	4,0 (0,7) *(N=24)*	4,0 (0,7) *(N=22)*
elders toepasbare kennis	3,1 (1,0) *(N =43)*	2,9 (1,1) *(N=17)*	3,2 (0,9) *(N=26)*	3,1 (1,0) *(N=21)*	3,0 (0,9) *(N=22)*
elders toepasbare vaardigheden	3,0 (1,2) *(N =39)*	2,3 (1,1) *(N=16)*	3,5 (1,0) *(N=23)*	3,2 (1,3) *(N=19)*	2,9 (1,2) *(N=20)*

Figuur 7.20 'Hoeveel heb je van deze praktische opdracht geleerd ten aanzien van …?' Antwoorden van leerlingen (boven in absolute aantallen, daaronder gemiddeldes en standaarddeviaties) op de vraag hoeveel ze geleerd hebben tijdens deze praktische opdracht ten aanzien van de in de linker kolom genoemde aspecten op 5-puntsschalen van 'niets' tot 'veel'; boven voor alle leerlingen, midden links voor de leerlingen van de Amsterdamse School (AS), midden rechts de leerlingen van de School in Amstelveen (SA), linksonder voor de leerlingen van de expliciete groepen en rechtsonder voor de leerlingen van de impliciete groepen.

vooruitgang is geboekt. Een verklaring voor deze verschillen is niet meteen te vinden, maar kan gezocht worden in docent- of leerlingkenmerken of een combinatie daarvan. In de figuren 7.8 en 7.11 is te zien dat de leerlingen in Amsterdam negatiever oordeelden over motiverende effect van de mondelinge feedback en de reflectieopdrachten.

De verschillen tussen de expliciete en de impliciete groep zijn klein, en alleen significant voor deeltoets 2 ($t(39)= -2{,}13$, $p<0{,}05$), maar enkele opmerkingen willen we er toch bij maken. Ten eerste dat de verschillen groter zijn tussen de scholen dan tussen de expliciete en impliciete groepen.werd. Nadere analyse leert dan ook nog eens dat het grootste verschil zit tussen de beide expliciete groepen (AS exp: -0,06; SA exp: 0,33 verschil in p'-waarde op deeltoets 2). Dat roept de vraag op wat er gebeurd is.

Een ander verschil dat opvalt, is dat voor alle leerlingen samen bij de scores de grootste vooruitgang in deeltoets 2 geboekt is (p'-waarde van 0,35 naar 0,51), terwijl bij begrip van bewijs daar juist de minste vooruitgang te constateren is (van 29 naar 31 procent). De grootste vooruitgang bij begrip van bewijs lag, zoals gezegd, bij de elementen met betrekking tot de beginfase van het onderzoek. Bij de vragen waar leerlingen de mogelijkheid hadden om onderzoek op verschillende criteria te beoordelen heeft dus een verschuiving plaatsgevonden naar criteria met betrekking tot onderzoeksvraag, hypothese of voorspelling ten koste van criteria met betrekking tot variabelen, steekproef en het doen van metingen. Dit effect was weer het sterkst in de expliciete groep en in Amstelveen.

7.5.5 Ervaren leereffect

Bij de evaluatie is aan de leerlingen gevraagd hoeveel ze van de praktische opdracht geleerd hebben over het onderwerp van onderzoek, over wat goed onderzoek doen is, over elders toepasbare kennis en over elders toepasbare vaardigheden. Zoals in figuur 7.22 te zien is, vonden vrijwel alle leerlingen dat ze matig tot zeer veel geleerd hebben over het doen van goed onderzoek, meer dan ten aanzien van het onderwerp van onderzoek. Dat is opvallend want meestal zijn de leerlingen op dat laatste veel meer gericht dan op het eerste (Hubers, 2003). In dit opzicht is de praktische opdracht dus effectief, de leerlingen vinden het nuttig om aan het onderzoek te werken (fig. 7.5) en het komt over dat het erom gaat om over het doen van goed onderzoek te leren. Dat ze geleerde kennis of vaardigheden (met betrekking tot onderzoek doen) elders kunnen toepassen, daar waren ze minder van overtuigd, al sloeg de balans nog naar de positieve kant door. Bij die laatste twee delen van deze vraag was er de mogelijkheid om toe te lichten wát ze dan geleerd hadden (…., namelijk: ……). De meerderheid van de toelichtingen (20 van de 37) betrof dan het doen van goed onderzoek in het algemeen en nog eens vier noemden een specifieke onderzoeksvaardigheid (zoals 'enquête afnemen'). Negen toelichtingen betroffen inhoudelijke kennis of vaardigheden ('sinaasappels persen') en vier waren moeilijk te duiden ('?' of 'bij scheikunde').

Is er in de posities op de vijfpuntsschalen nauwelijks verschil tussen de groepen (fig. 7.20), in de gegeven toelichtingen lijkt die er wel (tabel 7.13). In de expliciete groep werd veel vaker een antwoord gegeven in de categorie 'onderzoek doen, algemeen' dan in de impliciete groep, al is het wel zo dat die antwoorden vrijwel allemaal uit Amstelveen kwamen.

Tabel 7.13 Aantallen antwoorden van leerlingen op de vraag wat ze geleerd hebben ten aanzien van elders toepasbare kennis en vaardigheden.

	Totaal	*AS*	*SA*	*expliciet*	*impliciet*
onderzoek doen, algemeen	20	2	18	14	6
onderzoek doen, specifiek	4	2	2	3	1
inhoudelijk	9	4	5	3	6
onduidelijk	4	3	1	2	2

Het zijn dus met name de leerlingen uit de expliciete groep uit Amstelveen die verwoorden dat ze kennis en vaardigheden ten aanzien van onderzoek doen elders kunnen toepassen.

7.5.6 Samenvattend: in welke mate is de praktische opdracht effectief?

Aan het begin van deze paragraaf zijn vier aspecten van effectiviteit in het leren bewaken van de kwaliteit van onderzoek doen genoemd: (1) effectiviteit in het ontwikkelen van een motief en een gezamenlijk startkapitaal aan kennis, (2) effectiviteit in het ontwikkelen van de elementen van begrip van bewijs, (3) effectiviteit in het blijk geven de criteria voor de kwaliteit van onderzoek te beheersen en toepassen in hun eindproducten én bij de beoordeling van ander onderzoek en tenslotte (4) effectiviteit als ervaren leeropbrengst.

Op basis van deze vier aspecten en de samenhang ertussen kan de vraag beantwoord worden of de praktische opdracht effectief genoemd mag worden en of de expliciete en de impliciete onderwijsleerstrategie daarin verschillen.

- *Het ontwikkelen van een inhoudelijk motief en een gezamenlijke kennisbasis*

Op basis van de transcripten van de startlessen kan de conclusie getrokken worden dat de beoogde criteria voor goed onderzoek in ruime mate aan de orde zijn geweest, en ook dat zowel docenten als leerlingen daar beide aan hebben bijgedragen, zij het in Amsterdam meer evenwichtig dan in Amstelveen. De vraag 'zeker weten?' hebben de leerlingen wel meegenomen als motief om het eigen onderzoek zo goed mogelijk – in overeenstemming met de criteria – uit te voeren, maar minder om zoveel mogelijk te leren over wat goed onderzoek is – de criteria zelf. De vraag 'hoe weet ik of ik het zeker weet?' hield de leerlingen niet bezig. De gezamenlijke basis die de startlessen gelegd wordt, houdt dus vooral in dat leerlingen zich actief met de vraag 'wat is goed onderzoek?' bezighouden.

- *Ontwikkeling van begrip van bewijs*

Bij de weg naar het eindproduct, de ontwikkeling van begrip van bewijs gedurende de praktische opdracht, blijkt de rol van de docent cruciaal bij het expliciteren en generaliseren van de criteria voor goed onderzoek. Dat expliciteren en generaliseren vindt namelijk hoofdzakelijk in docent-leerlinggesprekken plaats. In hun onderlinge gesprekken blijven de leerlingen vrijwel geheel op het concrete niveau van beschrijven en problematiseren. De gesprekken over de reflectieopdrachten blijken zich beter te lenen voor het expliciteren en generaliseren van criteria dan de gesprekken over het eigen onderzoek. Begrip van bewijs ontwikkelt zich vaker langs de inductieve weg dan langs de deductieve, maar ongeveer net zo vaak via een heen-en-weer weg tussen concreet en algemeen. In de gesprekken over de reflectieopdrachten zijn wat meer deductieve patronen te onderkennen dan in de gesprekken rond het eigen onderzoek.

De structuur van het opdrachtenboekje – met name in de beginfase – en het geschreven commentaar van de docent op de opzet dragen bij aan het bewaken van de kwaliteit van het onderzoek door de leerlingen.

In de conceptartikelen is al een redelijke hoeveelheid 'begrip van bewijs' waar te nemen, de bespreking ervan met medeleerlingen of geschreven feedback erop van de docent verhogen de kwaliteit vooral waar het de evaluatie betreft. In de definitieve artikelen zijn meer opmerkingen over validiteit en betrouwbaarheid te vinden.

De gekozen aanpak om leerlingen aan een eigen onderzoek te laten werken en ze tijdens dat proces te bevragen over de kwaliteit ervan leidt er toe dat ze nadenken over de kwaliteit op het concrete niveau van hun eigen onderzoek. Veel minder breiden ze dat uit naar het algemene niveau van de criteria voor goed onderzoek. In de expliciete onderwijsleerstrategie, bij het maken en bespreken van de reflectieopdrachten, blijkt dat iets meer plaats te vinden.

- *De leerwinst zoals waargenomen*

De definitieve artikelen bevatten, zoals gezegd, al een meer dan redelijke hoeveelheid begrip van bewijs. In ongeveer tweederde van de voorkomende gevallen, wordt een criterium juist gehanteerd. De impliciete onderwijsleerstrategie doet het hierin iets beter. Er is voor de groep leerlingen uit 5 vwo als geheel een klein, maar positief verschil waar te nemen ten opzichte van in de voorstudie beoordeelde leerlingverslagen uit 6 vwo, waarbij het grootste verschil betrekking heeft op de evaluatie van het eigen onderzoek.

De leerlingen hebben een kleine, maar significante stap vooruit gedaan in het beoordelen van ander onderzoek, zoals in toetsvragen aan hen voorgelegd. Ze hebben dus zeker iets geleerd ten aanzien van het beoordelen van onderzoek. De criteria met betrekking tot het opzetten van een onderzoek (onderzoeksvraag, hypothese en variabelen) beheersten ze beter dan criteria met betrekking tot latere fasen. Maar ook al maken de leerlingen in hun artikelen wel in redelijke mate opmerkingen over validiteit en betrouwbaarheid, als ze ander onderzoek beoordelen komen die criteria niet of nauwelijks op tafel. De leerwinst van de leerlingen uit de expliciete onderwijsleerstrategie is iets groter dan die van de leerlingen uit de impliciete groep, maar de waargenomen leerwinst in Amstelveen is ook groter dan in Amsterdam. Een al te stellige uitspraak dat de expliciete groep het beter doet is daardoor niet verantwoord.

- *De leerwinst zoals ervaren*

De leerwinst die de leerlingen hebben ervaren, ligt duidelijk meer op het terrein van 'wat goed onderzoek doen is', dan op leren over het onderwerp van hun onderzoek. Hierbij waren de leerlingen van de expliciete groep in Amstelveen het meest uitgesproken. In dit opzicht is de praktische opdracht dus zeker effectief, het komt over dat het erom gaat om te leren over het doen van goed onderzoek.

- *Expliciet of impliciet effectief?*

Is met de iets betere score van de impliciete groep op de artikelen het pleit beslecht voor wat betreft de aanleiding van het praten over criteria? Die conclusie lijkt voorbarig, gezien hetgeen met betrekking tot de ontwikkeling van begrip van bewijs is geconcludeerd. Het weliswaar kleine verschil is wellicht te verklaren doordat de docenten in de impliciete groepen meer tijd hadden om over de leerlingonderzoeken te praten, omdat ze geen reflectieopdrachten hoefden na te bespreken. Bovendien gaven ze op het eigen onderzoek toegespitst schriftelijk commentaar, terwijl in de expliciete groep meer algemene aandacht aan de criteria werd gegeven in de reflectieopdrachten. Maar de wat betere score van de artikelen van de impliciete groep is gebaseerd op wat die leerlingen zeggen over hun éigen onderzoek. De toets, waar de leerwinst van de expliciete groep wat groter lijkt, is juist gebaseerd op het beoordelen van de kwaliteit van ánder onderzoek. Beide benaderingen blijken hun sterke kant te hebben.

Al met al is een genuanceerde conclusie over de effectiviteit van de praktische opdracht op zijn plaats. De leerlingen leren over goed onderzoek doen en ervaren dat ook als zodanig. De opbrengst ten opzichte van de verslagen uit de voorstudie en ten opzichte van de voortoets is positief en dat terwijl de voorstudie is gebaseerd op producten uit 6 vwo. De praktische opdracht lijkt daarmee *goed genoeg* om de leerlingen te laten leren over onderzoek doen met kwaliteit.

Maar er is ook twijfel, omdat de uitvoering tegen grenzen aan lijkt te lopen, zowel qua tijdsinvestering als qua eisen aan de docent. De elementen van begrip van bewijs zijn aan de orde geweest, maar niet alle even goed en vaak niet in gegeneraliseerde vorm. De in hoofdstuk 4 en 6 geschetste cyclus van *uitvoeren – reflecteren – ontwikkelen* van criteria (fig. 4.1 en 6.1) blijkt niet of nauwelijks doorlopen te worden. Zowel het motief van de leerlingen als de

focus van de gesprekken over de kwaliteit van onderzoek blijven op het concrete niveau van het eigen onderzoek. De vraag of de opdracht dan goed genoeg is om de leerlingen te leren de kwaliteit van onderzoek te bewaken kan helaas niet onvoorwaardelijk bevestigend beantwoord worden.

Dit eindoordeel over de praktische opdracht leidt ertoe de vraag te stellen of díe doelstelling – het leren bewaken van de kwaliteit van onderzoek – wel door zoiets als één praktische opdracht te bereiken is. Die discussie vindt plaats in hoofdstuk 8. Het leidt er ook toe de praktische opdracht zelf nogmaals onder de loep te nemen. Dat mondt dan uit in een overzicht van wat er goed ging en wat er in een eventuele nieuwe versie wellicht beter zou kunnen. Dat is het onderwerp van de volgende paragraaf.

7.6 Kan het beter?

In de vorige twee paragrafen zijn de uitvoering en de effecten van de praktische opdracht beschreven. Zowel de vraag naar de uitvoerbaarheid van de praktische opdracht als die naar de effectiviteit zijn aan het eind van paragraaf 7.4 respectievelijk 7.5 positief beantwoord, maar zowel de uitvoerbaarheid als de effectiviteit zouden beter kunnen, is het oordeel. In deze paragraaf bespreken we daarom nogmaals de uitgangspunten, zoals verwoord in paragraaf 7.1, om te bezien waar en hoe wellicht verbetering te bereiken is.

- *Omvang en planning*

Eerste uitgangspunt is een praktische opdracht van tien of elf lessen verspreid over vijf of zes weken, waarbij de leerlingen geacht worden tussendoor hun onderzoek uit te voeren. Dat het in die tijd vrijwel alle leerlingen gelukt is, valt te beschouwen als iets dat goed ging. Er werd echter door zowel leerlingen als docenten een forse tijdsdruk ervaren en allen zouden meer tijd zeker op prijs gesteld hebben. De vraag is echter of uitbreiding van de contacttijd deze druk verlicht. Wat ons inziens dan ook beter kan, is het tijdmanagement; docenten én leerlingen kunnen hun tijd efficiënter gebruiken. En dat is nog belangrijker wanneer de praktische opdracht door een klas van dertig leerlingen tegelijk wordt uitgevoerd. Met name het wachten van leerlingen tot ze met de docent kunnen overleggen en dus de verdeling van de tijd van de docent over de leerlingen vragen om een oplossing. Het afspreken van overleg buiten de lesuren is een mogelijkheid, evenals het concentreren van het hele onderzoek in een werkweek. Er is dan meer tijd te verdelen en de afleiding door andere beslommeringen is minder (Van Goor & Den Hertog, 2002).

Wat ook beter zou kunnen is de verdeling van de tijd over de fasen van het onderzoek. Er werd relatief veel tijd besteed aan de beginfase, het vinden van een onderwerp en een onderzoeksvraag. Hoewel er wel sturing was in de vorm van deadlines voor het inleveren van een vraag voor commentaar of het bespreken van een reflectieopdracht, werden deze niet strak gehanteerd. Dat zou beter kunnen, bijvoorbeeld door met onderzoeksteams vaste tijdstippen af te spreken om fases in het onderzoek (vraag, opzet, gegevensverzameling, conclusie en evaluatie) te bespreken en elkaar daaraan te houden. De leerlingen hebben dan een belang om op die tijdstippen voortgang te hebben geboekt, omdat ze anders kostbare begeleidingstijd verspillen.

- *De startlessen*

In de startlessen zijn de docenten er in behoorlijke mate in geslaagd een gezamenlijk startkapitaal aan kennis boven tafel te krijgen. Aanwezige kennis over onderzoek doen werd geactiveerd en er ontstonden levendige discussies over een goede aanpak van het voorbeeldonderzoek. De keus voor een voorbeeld op de grens van biologisch en psychologisch onder-

zoek leidde ertoe dat veel leerlingen ook in die richting hun onderwerp zochten. Een meer biologisch voorbeeld kan de leerlingen wellicht meer in die richting doen zoeken. In de gesprekken rond het eigen onderzoek werd af en toe door leerlingen of docent aan de startlessen en de daar ter sprake gebrachte criteria gerefereerd. Deze lessen fungeren dus als een goede start.

In de startlessen werd bij de leerlingen een motief opgeroepen voor het (leren) bewaken van de kwaliteit van onderzoek. Ze gingen na de startles gemotiveerd aan de slag, maar vooral gemotiveerd om een leuk en goed onderzoek te doen en minder om te leren hoe je kwaliteit van onderzoek bewaakt. De vraag is hoe dat beter kan. Wellicht kan dat door de leerlingen een onderzoek voor te leggen waar – ook in hun ogen – wel een en ander aan te verbeteren valt en ze uit te dagen dat beter te doen. Maar ook in een situatie waarin dat gebeurde, spanden de leerlingen zich vooral in om zelf iets beters op te zetten en niet om kwaliteitscriteria te ontwikkelen (Van Rens, 2005). Het blijft dan nodig om bij het presenteren van een voorbeeldonderzoek met de vraag 'hoe kun je beoordelen of dit onderzoek goed is uitgevoerd?' de stap naar veralgemenisering expliciet te maken: 'hoe weet je of je iets zeker weet?'

- *Een éigen onderzoek*

De leerlingen hebben (vrijwel) binnen de gestelde tijd allemaal een onderzoek opgezet en uitgevoerd. Ze waardeerden daarbij de keuzevrijheid en waren betrokken op hún onderzoek. Aan de andere kant vonden ze het erg moeilijk om binnen zoveel ruimte een geschikt onderwerp en een geschikte vraag te bedenken. Ondanks de suggestie in hun boekje, gebruikten ze vrijwel nooit hun biologieboeken of andere inhoudelijke informatie als inspiratiebron. Wel werd er gekeken naar onderzoeken van andere leerlingen die via internet te vinden waren.

De vraag is of inperking van de keuzevrijheid het bewaken van de kwaliteit ten goede komt. Door de keuzevrijheid in te perken, bijvoorbeeld tot een onderzoek naar diergedrag, komen de leerlingen wellicht makkelijker tot een geschikte onderzoeksvraag, maar kan ook de motivatie voor een 'eigen' onderzoek afnemen. In een situatie met een veel verder ingeperkte vrijheid met betrekking tot de onderzoeksvraag zijn in 5 vwo (bij scheikunde) echter positieve ervaringen opgedaan voor wat betreft de ervaren *ownership* van het onderzoek door de leerlingen (Van Rens, 2005). Door de leerlingen uit te dagen een gepresenteerd, door anderen gedaan onderzoek op een betere manier opnieuw te doen, ze in een licht competitieve omgeving te brengen (meedingen naar een prijs) en onderling commentaar in te bouwen (via een internetsymposium) 'gingen veel leerlingen ervoor'. Zo'n opzet kan ook stimuleren dat leerlingen meer vakinhoudelijke achtergrondinformatie betrekken in de evaluatie van hun onderzoek. Zeker als leerlingen later (in 6 vwo) een profielwerkstuk over een eigen onderwerp maken, is een beperktere keuzevrijheid voor deze praktische opdracht wellicht te verkiezen.

Voor de docenten en technisch onderwijsassistenten betekende de grote diversiteit aan onderwerpen een groot beroep op hun expertise, zowel inhoudelijk als qua technieken. Die expertise bleek niet voor alle onderzoeken in dezelfde mate aanwezig. Het beperken van de keuzevrijheid lijkt daardoor ook meer waarborgen te kunnen geven voor de kwaliteit van de begeleiding van de docenten en toa's, zowel wat inhoud als wat gebruikte onderzoeksmethoden betreft. Dus hoe meer een docent zich bekwaam weet op veel terreinen, hoe meer vrijheid er voor de leerlingen zou kunnen zijn wat keuze betreft.

- *Structuur van de opdracht en een artikel als eindproduct*

Vooral in de beginfase van hun onderzoek hadden de leerlingen houvast aan de redelijk open structuur van de opdracht zoals neergelegd in hun boekjes. Met name de stap om eerst voorlopige onderzoeksvragen te formuleren en die daarna te verbeteren leverde goede resultaten op. Toch gebruikten ze het niet optimaal, zeker niet in de latere fasen van het onderzoek. Dat zou beter kunnen, enerzijds door als docenten het gebruik te benadrukken en

wellicht zelfs te verplichten, anderzijds door het nog ondersteunender te maken. Bijvoorbeeld door bij de verschillende fasen (resultaten, conclusie, discussie) nog meer prikkelende vragen over de kwaliteit op te nemen (kun je de resultaten ook anders verklaren?; hoe zeker ben je van je conclusie?; denk je dat de redactie van een wetenschappelijk tijdschrift je artikel goed genoeg vindt om het te publiceren?).

De eisen aan het eindproduct en het voorbeeldartikel werden door de leerlingen beduidend minder vaak gebruikt en als minder bruikbaar gewaardeerd. In een volgende versie kan dan ook zeker een beter voorbeeld opgenomen worden. Daarbij hoeft geen (bijna) authentiek artikel te zijn, er kan ook een fictief artikel worden geschreven.

De leerlingen hebben in hun artikelen blijk gegeven over een redelijke mate van begrip van bewijs te beschikken. Met nog wat meer nadruk op de kenmerken van een 'wetenschappelijk' artikel is de eis zo'n eindproduct te leveren reëel. De meeste leerlingen is het namelijk gelukt daarbij in de buurt te komen. De feedback die zij daarbij kregen op hun voorlopige product bleek positief te werken. De feedback kan zich echter meer richten op de elementen van begrip van bewijs en de vraag 'hoe zeker weet je het nu?' Bijvoorbeeld door in 'rubrics' aan te geven welke eisen je stelt en welke niveaus daarin aangebracht kunnen worden.[33]

- *De rol van de docent*

De docenten hebben in de tweede ronde hun stimulerende rol duidelijk beter vervuld dan in de eerste ronde. Ze zetten de leerlingen aan het denken door vragen meestal niet direct te beantwoorden, maar een tegenvraag te stellen of door het onderzoek van leerlingen te problematiseren. De leerlingen waren tevreden met de schriftelijke feedback van hun docent en konden er ook lering uit trekken.

De docenten hebben ook regelmatig de criteria voor goed onderzoek aan de orde gesteld, op concreet niveau dan wel geëxpliciteerd of veralgemeniseerd. Dat expliciteren of generaliseren in meer volledige patronen vond in de gesprekken over de onderzoeken van de leerlingen echter minder plaats dan in de gesprekken over de reflectieopdrachten, waar expliciet naar de criteria gevraagd werd. In de gesprekken tussen de leerlingen onderling werd weinig geëxpliciteerd of gegeneraliseerd. Het blijkt dat deze aandacht voor de criteria van de docent moet komen, want de leerlingen doen dit niet uit zichzelf.

Er was minder sprake van gemiste kansen om begrip van bewijs aan de orde te stellen dan in de eerste ronde, maar het is gezien de druk waaronder de docenten begeleiden begrijpelijk dat niet alle kansen gegrepen werden. Wat echter wel beter zou kunnen, is het aangrijpen van de concrete vragen uit de onderzoeken van de leerlingen om de criteria te expliciteren en te generaliseren in volledige redeneringen.

De docenten hebben de leerlingen veel uitgelegd over de criteria, maar helaas bleek ook dat ze niet alle criteria even goed beheersten. Teveel is ervan uitgegaan dat een eerstegraads docent – die dus een wetenschappelijke opleiding heeft gehad – de criteria wel beheerst of snel oppikt. Dit betekent dat in elke volgende versie van de praktische opdracht de inhoudelijke voorbereiding van de docenten op de elementen van begrip van bewijs beter moet zijn. Het verdient dan ook serieuze overweging om nascholing voor docenten op te zetten waarin zowel de criteria voor goed onderzoek aan de orde komen als de manier om ze te benoemen, te expliciteren en te generaliseren. Ervaringen met nascholing van docenten op het terrein van de kwaliteit van leerlingonderzoek wijzen er echter op dat die alleen kans op suc-

[33] Er zijn verschillende rubrics voor het uitvoeren van natuurwetenschappelijk onderzoek op school gemaakt, bijvoorbeeld door Brouwers (2006) en door Lunsford en Melear (2004). Daarin is veel aandacht voor de proceskant, en slechts weinig voor validiteit en betrouwbaarheid van het onderzoek. Ook in beoordelingsschema's voor bijvoorbeeld profielwerkstukken krijgen deze begrippen weinig plaats (Sinkeldam, 1998; Weezenbeek, 1999).

ces heeft als de betreffende docenten die kwaliteit als een probleem zien waar ze aan willen werken. (Smits, 2003). Ook in een project waar docenten enthousiast en gedreven werkten aan het ontwikkelen van een onderzoekende houding kwamen criteria voor goed onderzoek niet expliciet aan de orde (Van der Valk e.a., 2005).

- *Het wetenschappelijke verhaal*

De conclusie dat de tweede versie van de praktische opdracht niet onvoorwaardelijk goed genoeg is om te leren de kwaliteit van onderzoek te bewaken, kan je vertalen met de stelling dat het wetenschappelijke verhaal van de criteria voor goed onderzoek onvoldoende gestalte heeft gekregen. Zo zijn in het eerste deel van het verhaal de eisen aan de onderzoeksvraag en het onderscheiden van afhankelijke en onafhankelijke variabelen goed over het voetlicht gekomen en goed opgepikt. Maar de hypothetisch-deductieve manier van redeneren blijkt moeilijk voor zowel docenten als leerlingen.

In de fase van metingen was er ruim aandacht voor allerlei beïnvloedende factoren en het trekken van een acceptabele steekproef. Het inbouwen van vereiste blanco's werd echter vaak over het hoofd gezien.

Wat de evaluatiefase van onderzoek betreft is het beeld genuanceerd. Element 21, over validiteit en betrouwbaarheid, kwam in de gesprekken over het onderzoek weinig aan de orde, maar de meerderheid van de leerlingen maakte er wel opmerkingen over in hun artikel. Hadden de leerlingen hier aan een half woord genoeg of werkte hier toch het motto van de praktische opdracht 'zeker weten?' door? Daarentegen scoorden ze op dit onderdeel in de toets (deeltoets 3) het laagst, lager dan op de andere onderdelen.

Als het gaat om het gefundeerd kunnen beoordelen en bewaken van de kwaliteit van een onderzoek blijft er ons inziens daarom nog wel wat te wensen over. De vraag naar de kwaliteit en de waarborgen voor een zo zeker mogelijk antwoord zou vaker aan de orde moeten komen, niet alleen aan het begin, maar ook gedurende de opzet en uitvoering van het onderzoek. De docenten kunnen die vaker stellen: 'Als je dit zo aanpakt, ben je dan straks zeker van je antwoord?' En, misschien nog wel belangrijker, de vraag kan aan het eind prominent gesteld worden: hoe zeker ben je, terugkijkend, van je antwoord en waarom wel of niet? Bij zo'n afsluitende *reflection-on-action* kunnen de criteria en de begrippen betrouwbaar en valide nogmaals geëxpliciteerd worden. Daarmee wordt het een meer compleet en 'rond' verhaal. Behalve een startles waarin de gezamenlijke kennisbasis en de criteria expliciet gemaakt worden, kan zo'n 'slotles' ons inziens zinvol zijn. Zaak is dan wel om het niet als 'mosterd na de maaltijd' op te dienen, dus het is alleen zinvol als het vóór het inleveren van de definitieve producten plaatsvindt.

- *Expliciete of impliciete aanleiding tot reflectie?*

Voor sommigen wellicht de hamvraag: zou in een derde versie van de praktische opdracht de expliciete of de impliciete aanpak gevolgd moeten worden? Beide onderwijsleerstrategieën lijken goed genoeg. Leerlingen in de impliciete groep doen het iets beter in de artikelen, leerlingen uit de expliciete groep in de toets. Beide strategieën hebben dus hun sterke kant en een derde versie van de praktische opdracht zou daarom moeten zoeken naar een manier om beide sterke kanten te verenigen.

De reflectieopdrachten blijken in het algemeen goed uitvoerbaar en worden door de meeste leerlingen ook volgens plan uitgevoerd. Ze worden echter niet erg gewaardeerd en vaak als hinderlijke onderbreking van het eigen onderzoek gezien. Over de vierde opdracht, het geven van onderlinge feedback, zijn leerlingen positief, maar zeker niet alle leerlingen. Er waren er ook die het commentaar dat ze kregen beneden de (hun) maat vonden. Daarentegen zijn de meeste leerlingen wel tevreden met de schriftelijke feedback van hun docent.

De leeropbrengst van de reflectieopdrachten is niet heel hard te maken. Van de eerste opdracht was dat misschien nog het duidelijkst, want sommige leerlingen pasten hun onder-

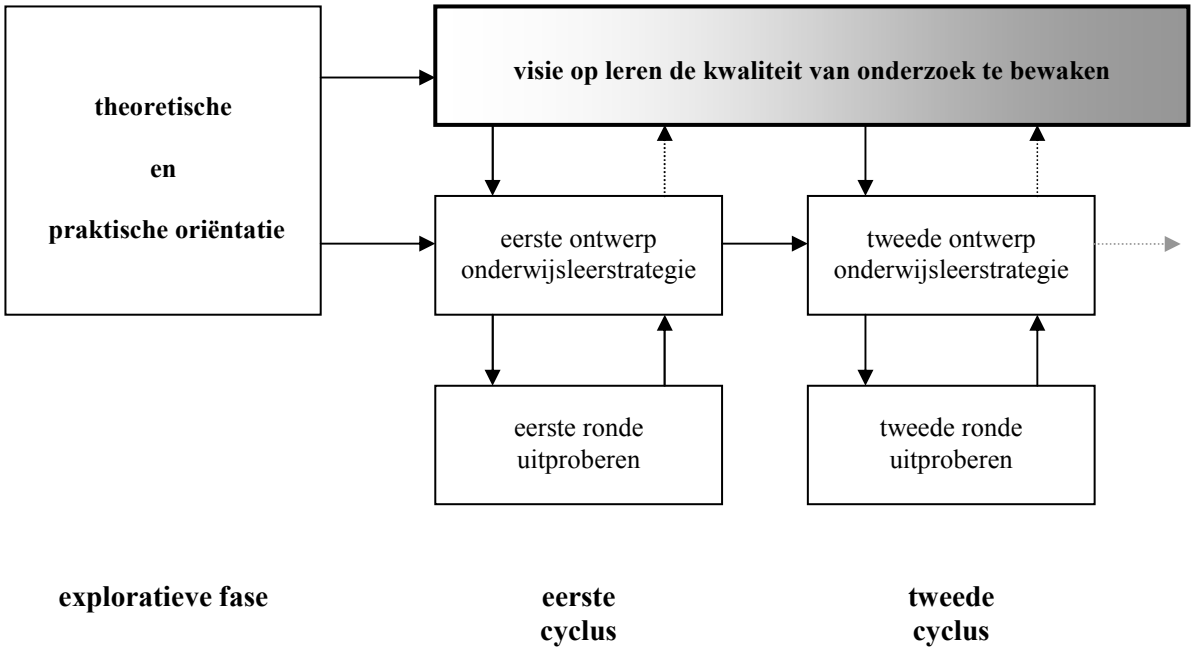
zoeksvraag naderhand daadwerkelijk aan. En ook van de vierde opdracht gaven sommige leerlingen aan dat hij heel leerzaam was, niet in het minst doordat kritisch kijken naar een ander onderzoek aanzette nogmaals kritisch te kijken naar het eigen onderzoek. Wel is waargenomen dat in de gesprekken rond de reflectieopdrachten tussen docent en leerlingen vaker in expliciete en algemene zin over de criteria voor goed onderzoek gesproken werd.

In de gesprekken over het eigen onderzoek heeft weinig explicitering en generalisatie van de criteria plaatsgevonden. In onze ogen schiet de impliciete strategie daarom tekort als de criteria alleen maar op het concrete niveau van het eigen onderzoek aan de orde komen. De aanleiding die het eigen onderzoek geeft om over begrip van bewijs te praten is blijkbaar wel voldoende – in die zin zijn de reflectieopdrachten dus niet nodig – maar de diepgang c.q. de mate van explicitering en veralgemenisering niet.

In een volgende versie van de praktische opdracht zou er naar gestreefd moeten worden om meer *reflection-in-action* in de gesprekken rond het eigen onderzoek te laten plaatsvinden en die reflectie naar een hoger plan te tillen. De docent kan de opdrachten als facultatief in te zetten reflectie achter de hand houden om die door leerlingen te laten maken die moeite hebben met bepaalde elementen van begrip van bewijs. Voor de problematiek rond hypothese en voorspelling zou dan een nieuwe opdracht ontwikkeld moeten worden. De opdrachten mogen ook nog wat kleiner in omvang zijn.

De vierde opdracht, waarbij het gaat over het eigen onderzoek én over het onderzoek van anderen, kan gehandhaafd worden, waarbij de docent beter in de gaten zou kunnen houden of de gegeven feedback van voldoende inhoudelijke kwaliteit is. Of er kan op een andere manier ruimte voor peer review gemaakt moeten worden, bijvoorbeeld door leerlingen hun voorlopige resultaten aan de andere leerlingen te laten presenteren, waarna er een discussie over de kwaliteit ervan kan plaatsvinden.

Behalve die *reflection-in-action* kunnen in een hierboven al gesuggereerde ‘slotles’, terugkijkend op volledig afgesloten onderzoeken, de criteria nogmaals expliciet als *reflection-on-action* aan de orde komen, eventueel aan de hand van een nieuw voorbeeldonderzoek.



8.1 Antwoord op de onderzoeksvraag
8.2 Hoe zeker is het antwoord?
8.3 Reflectie op de theoretische uitgangspunten
8.4 Lijnen naar betekenisvol leren
8.5 Verder onderzoeken
8.6 Hoog genoeg mikken

8 Terugblik en vooruitblik

'Our science should be adult enough to rejoice in complexity.'
Steven Rose (1998, p. 133)

Hoofdstuk 8 Terugblik en vooruitblik

In het laatste hoofdstuk van dit proefschrift maken we de balans op. Want het onderzoek heeft weliswaar een voorzichtig positief resultaat opgeleverd, er zijn ook aanwijzingen gevonden dat de randvoorwaarden van het huidige voorbereidend wetenschappelijk onderwijs niet altijd voldoende ruimte bieden voor het bereiken van deze doelstelling, zowel wat betreft de praktische mogelijkheden voor leerling en docent als de inhoudelijke inrichting.

Maar voordat geschetst wordt wat er ons inziens nodig is om dichter bij die doelstelling te komen (in par. 8.4 en 8.5), wordt teruggekeken op de theoretische uitgangspunten (par. 8.3) en de methodologische aanpak (par. 8.2) van dit onderzoek. In een laatste paragraaf wordt de vraag opgeworpen of we eigenlijk niet een te ambitieuze doelstelling nastreven. Allereerst wordt echter de onderzoeksvraag beantwoord.

8.1 Antwoord op de onderzoeksvraag

In hoofdstuk 1 is de onderzoeksvraag van dit onderzoek beschreven: *Hoe kunnen leerlingen in het biologieonderwijs in de tweede fase VO leren de kwaliteit van het door hen uitgevoerde eigen onderzoek beter te bewaken?* Daar werden ook vier deelvragen geformuleerd die moeten leiden tot de beantwoording daarvan.

- *De eerste en tweede deelvraag*

De deelvraag *Wat is de biologische invulling van 'begrip van bewijs'?* leidde in hoofdstuk 2 tot de omschrijving van kwaliteit van onderzoek in 23 elementen van begrip van bewijs. Op de deelvraag *Wat is de plaats van onderzoek doen in het biologieonderwijs?* is in hoofdstuk 3 een antwoord gegeven. Vergelijking van wat volgens de literatuur, de officiële leerdoelen en docenten gewenst is met wat er aan begrip van bewijs te vinden is in schoolboeken en verslagen van leerlingen levert een beeld op van dat wat ons inziens verbetering behoeft. Als basis voor een te ontwikkelen onderwijsleerstrategie die binnen het bestek van dit onderzoek ontwikkeld zou kunnen worden, is een beargumenteerde selectie uit de 23 elementen gemaakt.

- *De derde deelvraag*

Om zicht te krijgen op de derde deelvraag *Hoe ontwikkelen leerlingen begrip van bewijs?* is in hoofdstuk 4 een theoretische verkenning gedaan en in hoofdstuk 6 is die verkenning op basis van de eerste onderzoekscyclus bijgesteld. Uitgaande van de 'spiral formation of mental actions' van Galperin (Arievitch & Haenen, 2005) en de ontwikkeling van activiteiten zoals beschreven in de cultuurhistorische theorie (Van Aalsvoort, 2000) is een didactische structuur (Lijnse & Klaassen, 2004) geschetst waarin leerlingen van het concrete handelen in hun eigen onderzoek via het expliciet verwoorden van de kwaliteit ervan komen tot het generaliseren van criteria voor goed onderzoek. In eerste instantie met een analyse van de argumentaties (Toulmin, 1958) en later met een analytisch kader afgeleid van Mortimer en Scott (2003) is die ontwikkeling in beeld gebracht.

In hoofdstuk 5 en 7 is beschreven wat er van die ontwikkeling is waargenomen. Leerlingen bleken zich vooral op het concrete, beschrijvende niveau met begrip van bewijs bezig

te houden en uit zichzelf weinig redeneringen op te zetten waarin kwaliteit geëxpliciteerd of criteria gegeneraliseerd worden. In gesprek met de docent kwam het vaker tot expliciteringen en generalisaties, die echter toch nog vaker door de docent dan door de leerlingen onder woorden gebracht werden.

- *De vierde deelvraag*

De aard van dit ontwikkelingsonderzoek – met in essentie een ontwerpvraag – komt het best tot uitdrukking in de laatste deelvraag: *Wat zijn kenmerken van een uitvoerbare en effectieve onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek?* In twee onderzoeksrondes is gewerkt aan de vormgeving van zo'n onderwijsleerstrategie en in hoofdstuk 5 en 7 is daar uitgebreid verslag van gedaan. Daarbij is – zij het voorzichtig – geconcludeerd dat de tweede versie van de ontwikkelde praktische opdracht zowel uitvoerbaar als voldoende effectief is wat betreft het leren doen van een onderzoek met voldoende kwaliteit. Maar dat hij goed genoeg is wat betreft het generaliseren van de criteria voor goed onderzoek is niet zeker. In paragraaf 7.6 is verwoord wat er goed ging in de tweede versie en wat er ons inziens beter zou kunnen in een eventuele derde.

Maar een ontwikkelingsonderzoek gaat niet alleen om het produceren van een succesvol ontwerp. Het beoogt ook inzichten op te leveren die verder gaan dan het concrete product, i.c. de praktische opdracht. De vraag 'hoe kunnen ze het leren?' beantwoorden met 'met deze opdracht' zou immers het biologieonderwijs en de biologiedidactiek in Nederland niet veel verder helpen. We proberen daarom de ons inziens essentiële kenmerken van de onderwijsleerstrategie in meer algemene zin te verwoorden.

Leren de kwaliteit te bewaken kan niet zonder kwaliteitscriteria. Daarom zal altijd duidelijk moeten zijn om welke criteria het gaat. Daarmee krijgen niet alleen Millar e.a. (1994) gelijk wat betreft hun pleidooi voor een aparte *body of knowledge* voor onderzoek doen, maar wordt ook duidelijk dat iedere onderwijsleerstrategie daarop gericht zal moeten zijn. Daarmee bedoelen we dat het niet volstaat om leerlingen een onderzoek te laten doen en in de marge ervan criteria aan de orde te stellen, maar dat de criteria het doel en de kern van de leeractiviteiten moeten zijn. Gebleken is dat gerichte aandacht voor elementen van begrip van bewijs zich terugbetaalt in de producten van leerlingen. Tegelijkertijd is gebleken dat het niet eenvoudig is; leerlingen hebben er moeite mee zich de kern van de in de biologie zo belangrijke hypothetisch-deductieve wijze van redeneren eigen te maken. Kwaliteit bewaken is meer dan het afvinken van criteria, het gaat om het geheel, om begrip van bewijs in de betekenis van het begrijpen van wetenschappelijke bewijsvoering, onafhankelijk van de precieze uitwerking in elementen (Schalk, 2005b; Van Eijck, 2005).

En dat 'wetenschappelijke verhaal' over de kwaliteit van goed onderzoek moet ontwikkeld worden c.q. over het voetlicht gebracht worden. Mortimer en Scott (2003) spreken steeds van *staging the scientific story,* het ten tonele voeren ervan. In dat spel zijn zowel leerlingen als docent actief. Maar al is het dan bij voorkeur interactief toneel, duidelijk is dat de docent de regie moet voeren. Gebleken is dat begrip van bewijs vrijwel alleen in algemene zin of geëxpliciteerd aan de orde komt als de docent in het gesprek betrokken is. Bij leerlingen blijft het gesprek meestal op het concrete niveau van het eigen onderzoek. Niet alleen per element van begrip van bewijs geldt dat, ook wat het verhaal als geheel betreft, blijkt de docent een onmisbare rol te vervullen. Het leggen van een gemeenschappelijke basis waaraan later gerefereerd kan worden blijkt te lukken. Een afsluitend gesprek over wat er geleerd is lijkt op een zelfde manier nodig om het gezamenlijke verhaal compleet te maken. Begrip van bewijs moet dus niet alleen de kern van de leerinhoud zijn, de ontwikkeling ervan moet ook de didactische uitwerking bepalen.

Daarbij is reflecteren nodig. De onderwijskundige literatuur is daar stellig over: reflecteren is nodig bij het leren van hogere orde vaardigheden (Korthagen, Kessels, Koster, La-

gerwerf & Wubbels, 2001; Lagerwerf en Korthagen, 2006). In de ontwikkelde praktische opdracht is gevarieerd met de aanleiding om op de kwaliteit van onderzoek te reflecteren in een impliciete en expliciete onderwijsleerstrategie. Het eigen onderzoek van de leerlingen kan een 'impliciete' aanleiding zijn als de problematiek ervan het aan de orde stellen van een kwaliteitscriterium nodig of mogelijk maakt. Een 'expliciete' aanleiding wordt gevormd door reflectieopdrachten, die weliswaar qua tijdstip bij het eigen onderzoek moeten aansluiten, maar qua vraag daar onafhankelijk van zijn. Daarmee wordt gewaarborgd dat alle beoogde elementen van begrip van bewijs aan de orde komen. Het is gebleken dat bij het bespreken van de reflectieopdrachten (met de docent) de criteria vaker van het concrete beschrijvende niveau getild worden door ze in meer algemene zin te verwoorden dan in de gesprekken over het eigen onderzoek. De leerlingen in de expliciete groep verwoordden ze ook in de gesprekken over het eigen onderzoek iets vaker dan leerlingen uit de impliciete groep, waar het verwoorden vooral voor rekening van de docent kwam. Voor het overige was er echter weinig verschil te zien tussen de groepen. In de producten, waarbij het over het eigen onderzoek ging, paste de impliciete groep iets vaker de criteria toe, in de toetsen, waar het om ander onderzoek ging, scoorde de expliciete groep iets beter. Beide benaderingen blijken vruchten af te werpen, zowel de ' impliciete' *reflection-in-action* – 'waar ben ik mee bezig?' – als de 'expliciete' *reflection-on-action* – 'wat heb ik gedaan?' (Schön, 1983). Een succesvolle aanpak voor het leren bewaken van de kwaliteit van onderzoek zou ons inziens dan ook beide ingrediënten moeten bevatten.

Het ervaren van het onderzoek als 'eigen', lijkt ook een essentieel kenmerk. In het biologieonderwijs bestaat al een lange traditie van 'eigen experimenteel onderzoek', waarbij leerlingen zelf hun onderwerp kiezen en hun eigen onderzoeksvraag formuleren. Ook in dit onderzoek bleek weer hoe eigen onderzoek de leerlingen motiveert om aan de slag te gaan. Dat ze dan ook volledige vrijheid moeten krijgen met betrekking tot onderwerp en onderzoeksvraag is daarmee echter niet gezegd. Inperking van die vrijheid kan om andere redenen nodig zijn, zoals de expertise van de docent. Dat er met minder vrijheid toch voldoende *ownership* bereikt kan worden, blijkt bijvoorbeeld waar leerlingen een reëel probleem uit de samenleving onderzochten (Crawford, 2000; Roth & Lee, 2004), of de situatie waar leerlingen een (gefingeerd) onderzoek van andere onderzoekers beter probeerden uit te voeren (Van Rens, 2005). En leerlingen ervoeren het ook als 'eigen' in een nogal gesloten opzet waarbij ze werkten aan een vraag en met de technieken van een universitair onderzoeker, doordat ze stoffen produceerden die niemand anders ooit nog gemaakt had (het project 'LigandIG' van De Vrolijke School, zie *www.devrolijkeschool.nl*). Het voordeel van een minder grote variatie aan onderwerpen en technieken is dat de docent meer gelegenheid heeft om aandacht te besteden aan de kern van de zaak: de ontwikkeling van begrip van bewijs.

Van de docent wordt namelijk veel gevraagd. Hij moet een expert zijn, zowel wat betreft begrip van bewijs, het ontwikkelen ervan via *reflection-in-action* en *reflection-on-action*, vakinhoud en onderzoekstechnieken, als wat betreft het regisseren van het onderwijsleerproces. Maar hij moet ook een kei in tijdmanagement zijn. Dat die expertise nodig is, is gebleken in de voorbereiding op en de uitvoering van de ontwikkelde praktische opdracht. Niet altijd bleek die expertise in voldoende mate aanwezig of in voldoende mate uit de verf te komen. De onbekendheid met sommige onderwerpen van onderzoek maakte het soms onmogelijk om fundamenteler op de kwaliteit van het onderzoek in te gaan. Maar ook de beheersing van begrip van bewijs was niet optimaal, al is dat iets waarin deze docenten geen uitzondering vormden (Brinkman & Van Rens, 1999; Lawson, 2002; Schalk 2002a, 2002b; Yip, 1999). De rol die docenten (moeten) spelen bij het begeleiden van onderzoek van leerlingen is niet eenvoudig, zowel wat het bewaken van de kwaliteit van de onderzoeken betreft (Brinkman & Van Rens, 1999; Smits, 2003; Van der Schee & Rijborz, 2003) als het begeleiden van het hele proces (bijv. Crawford, 2000; Van Rens & Dekkers, 2001). Crawford (2000) noemt wel tien rollen die de docent moet vervullen, onder andere die van motivator, diagnosticus, gids en

medewerker. Zij concludeert dan ook (p. 932):

> The point I am stressing is that teachers' work in an inquiry-based classroom requires taking a myriad of roles – roles that demand a high level of expertise.

Om in het onderwijs het leren bewaken van de kwaliteit van onderzoek te stimuleren lijkt aandacht voor en versterking van de expertise van zittende en nieuwe docenten dan ook zeker noodzakelijk. Eigen ervaring met het doen van onderzoeksopdrachten in de opleiding is daarbij van grote invloed (Windschitl, 2002, 2003).

Kortom, terugkijkend op de ontwikkelde en uitgeprobeerde opdracht beschouwen wij de volgende kenmerken essentieel:

- duidelijkheid over wat bewaken van kwaliteit van onderzoek inhoudt, zowel voor docent als voor leerlingen;
- stevige inbedding van het doen van onderzoek in het wetenschappelijke verhaal over onderzoek doen, niet als slagroom op de taart, maar als gist in het beslag;
- gestalte geven aan de wisselwerking tussen onderzoek doen en leren over onderzoek doen via *reflection-in-action* en *reflection-on-action*;
- genoeg 'eigenheid' voor de leerlingen in het doen van onderzoek, er moet een belang zijn om een zo zeker mogelijk antwoord te krijgen;
- voldoende expertise van de docent in het doen en begeleiden van onderzoek én in het bewaken van de kwaliteit ervan.

- *Kunnen ze het leren en zo ja hoe?*

Is met het beantwoorden van de deelvragen – vooral met het aangeven van essentiële elementen van een onderwijsleerstrategie – ook de hoofdvraag *Hoe kunnen leerlingen in het biologieonderwijs in de tweede fase VO leren de kwaliteit van het door hen uitgevoerde eigen onderzoek beter te bewaken?* beantwoord? Naar onze mening niet helemaal. Want het in dit proefschrift beschreven onderzoek laat zien dat het niet eenvoudig is om in het voortgezet onderwijs aan deze voorwaarden te voldoen. Misschien moet de conclusie wel luiden dát er in de huidige tweede fase niet in voldoende mate aan is te voldoen. De leeropbrengst van het ontwikkelde materiaal was gering en de inspanning groot. In paragraaf 8.4 wordt daarom de vraag aan de orde gesteld onder welke omstandigheden er wellicht beter aan de voorwaarden voldaan kan worden. Maar eerst wordt ingegaan op de kwaliteit van dít onderzoek.

8.2 Hoe zeker is het antwoord?

De aard van het onderwerp van dit onderzoek – de kwaliteit van onderzoek – maakt de vraag 'hoe zeker is dat antwoord eigenlijk?' onontkoombaar. Daarom wordt in deze paragraaf de aanpak van dit onderzoek nog eens kritisch beschouwd. Achtereenvolgens staan we stil bij:

- de gebruikte begrippen;
- de keus voor scholen, docenten en leerlingen;
- het ontwikkelde materiaal en de suggesties voor verbeteringen;
- de onderzoeksinstrumenten en de zekerheid van de gegevens;
- de geldigheid van de conclusies.

- *Begrippenkader*

De eerste stap in het onderzoek was het uitwerken van 'biologisch begrip van bewijs' op basis van de *concepts of evidence* van Gott en Duggan (2003) en Gott e.a. (2006). De hieruit resulterende 23 elementen zijn vervolgens gebruikt om docenten te bevragen, schoolboeken en leerlingproducten te analyseren, een toets te ontwerpen, een praktische opdracht mee te

ontwikkelen en uit te voeren en gesprekken mee te analyseren. Dat tekent wel het belang van de keuzes die in het begin van dit onderzoek gemaakt zijn ten aanzien van het begrippenkader. Twijfel aan de validiteit van deze uitwerking van 'kwaliteit van onderzoek' legt daarom de bijl aan de wortel ervan. Die twijfel hebben we zoveel mogelijk proberen weg te nemen door ze voor beoordeling op consistentie en bruikbaarheid voor te leggen aan een wetenschapsfilosoof, vakdidactici en ervaren docenten. Ongetwijfeld is een andere operationalisering van de kwaliteit van onderzoek mogelijk. We denken echter met de 23 elementen van begrip van bewijs 'kwaliteit van onderzoek' – met daarin de kern van de in de biologie zo belangrijke hypothetisch-deductieve werkwijze – goed geoperationaliseerd te hebben. Ook met de gebruikte selectie ervan is die kern in beeld gebleven.

Met de begrippen *problematiseren, beschrijven, expliciteren, generaliseren* en *toepassen* hebben we het proces van ontwikkeling van begrip van bewijs in beeld proberen te brengen. En hoewel dit werkwoorden met een min of meer bekende betekenis in het onderwijs en erbuiten zijn én het gebaseerd is op een door Mortimer en Scott (2003) ontwikkeld analytisch kader, kan de vraag gesteld worden of het daarmee gelukt is. Bevordert hardop uitgesproken explicitering en generalisatie het opnemen van kwaliteitscriteria in je mentale structuur? Op die vraag kunnen we helaas geen robuust antwoord geven. Wat we gedaan hebben, is constateren dat het mondjesmaat gebeurde en dat de leerlingen in wiens aanwezigheid het iets vaker gebeurde, iets beter scoorden op een toets die om toepassing van de criteria vroeg. Maar ook dat leerlingen die vrijwel alleen op het concrete, problematiserende en beschrijvende, niveau bleven producten van voldoende kwaliteit afleverden. Vooralsnog lijkt het echter een bruikbaar begrippenkader.

- *Scholen, docenten en leerlingen*

Een ontwikkelingsonderzoek kiest er bewust voor met een beperkt aantal scholen en docenten te werken. De te ontwikkelen onderwijsleerstrategie moet namelijk in een concrete schoolpraktijk passen. Maar tegelijkertijd moet het onderzoek iets opleveren waar meer dan alleen de deelnemende scholen en docenten iets aan hebben. De vraag is dus of de 'gekozen' scholen en docenten voldoende vergelijkbaar zijn met andere Nederlandse scholen voor voorbereidend wetenschappelijk onderwijs. *Gekozen* staat niet voor niets tussen aanhalingstekens, want er was niet veel keus. We hebben veel docenten benaderd voor medewerking, maar vooral praktische bezwaren weerhield velen van hen. Toch zijn we van mening dat de Amsterdamse School en de School in Amstelveen aan het criterium 'gemiddelde school' voldoen en dat Bea en Lex 'gemiddelde docenten' zijn – en zijn we zeer tevreden over de inzet en samenwerking. Dat komt ons inziens de ecologische validiteit van het onderzoek ten goede. Wat deze 'gemiddelde docenten' gepresteerd hebben, ligt ons inziens binnen het bereik van velen.

Het onderzoek is uitgevoerd met weinig leerlingen. Die zijn niet zomaar representatief voor alle leerlingen in het Nederlandse vwo. In Amstelveen komen de leerlingen weliswaar zowel uit de stad als van het platteland, maar in Amsterdam zijn het vooral goedgebekte grotestadskinderen. Weinig leerlingen leveren natuurlijk geen bergen mooie, statistisch te bewerken gegevens op voor klinkende resultaten. Dat is een inherente zwakte van ontwikkelingsonderzoek. Desalniettemin hebben we kwantitatieve gegevens verzameld en bewerkt, omdat het het beeld van de uitvoering en de opbrengst verrijkt. Dat de onzekerheidsmarges dan groot zijn, nemen we voor lief.

Voor het volgen van de gesprekken is zelfs slechts een deel van deze kleine populatie leerlingen gevolgd, slechts zeventien van de vijftig leerlingen. Dat is puur om praktische redenen gedaan, nog meer opnames waren binnen de gestelde tijd niet uit te werken geweest. Maar dan is het wel van belang dat er een goede selectie van leerlingen plaatsvindt. In het onderzoek is de selectie aan de docenten overgelaten op basis van de uitgesproken wens dat

het min of meer 'gemiddelde' leerlingen moesten zijn. En hoewel ze samen waarschijnlijk wel een gemiddelde vormen, zijn de verschillen tussen leerlingen en hun manier van werken groot. De kans bestaat dus dat met een andere keus voor te volgen leerlingteams de gegevens van de gesprekken er anders uitgezien hadden.

Tien à veertien leerlingen in de klas of dertig maakt veel verschil. Als met tien tot veertien leerlingen al een grote (tijds)druk ervaren wordt door de docenten, dan laat het zich raden hoe dat in een grotere groep gaat. Dan is wellicht een enigszins andere opzet nodig, bijvoorbeeld standaard groepjes van drie leerlingen. Ongetwijfeld zal ook ervaring het uiteindelijk soepeler laten verlopen. Maar het onderstreept wel nog eens de invloed van de grenzen van de huidige inrichting van het onderwijs.

Ondanks deze slagen om de arm ten aanzien van de selectie van scholen, docenten en leerlingen is naar onze mening een ecologisch valide, herkenbaar beeld neergezet van de uitvoering van de ontwikkelde praktische opdracht. En zoals Roth (1995, p. 279) het uitdrukt:

> Once all aspects of the local situation and the knowledge construction process are made explicit (that is, if teachers have provided thick descriptions of the process and products of their inquiries), readers of research reports can decide for themselves how much of the local knowledge can be transported into other settings.

- *Het materiaal*

Op basis van het werken met (twee versies van) een praktische opdracht is een antwoord gegeven op de vraag hoe leerlingen kunnen leren de kwaliteit van onderzoek te bewaken. Zou dat antwoord anders zijn geweest als de opdracht er anders uitgezien had? Misschien in detail, maar we denken niet in grote lijn. De opdracht is enerzijds dicht bij de traditie in het biologieonderwijs gehouden (een eigen open onderzoek) en anderzijds van elementen voorzien waarvan werd verondersteld dat ze dienstbaar waren aan het bereiken van de doelstelling (het vaak aan de orde stellen van criteria in de begeleiding, de reflectieopdrachten). Ook was er de beperking in tijd, het ging om een opdracht voor een aantal weken, niet gespreid over jaren. Met de uitgangspunten zoals die samen met de docenten zijn opgesteld (zie hoofdstuk 5), lijkt het antwoord op de onderzoeksvraag niet op z'n kop gezet te worden door een andere concrete operationalisering van de opdracht.

Dat betekent dat we ook van mening zijn dat de suggesties voor verbetering die we in paragraaf 7.6 doen een stevig fundament hebben. Uiteraard hebben ze een tentatief karakter – we weten niet zeker of ze zullen werken – maar ons inziens zijn ze voldoende onderbouwd.

- *De onderzoeksinstrumenten*

Steven Rose (1998) doelt in het citaat aan het begin van dit hoofdstuk op de biologie, maar het geldt ons inziens ook voor de onderwijskunde en de didactiek. In dit onderzoek is op een complexe manier naar een complexe materie gekeken. Hoe leerlingen een begrippenkader over kwaliteit van onderzoek ontwikkelen gedurende een weken durend onderzoek aan een niet van te voren vastgesteld onderwerp, hoe ontwikkelde onderwijsmaterialen en het doen en laten van docenten daarop van invloed zijn. Daarbij zijn vragenlijsten en toetsen afgenomen, onderlinge gesprekken, aantekeningen en (concept)artikelen geanalyseerd, observaties gedaan en interviews gehouden die verschillende aspecten belichten van het proces dat heeft plaatsgevonden.

Vanwege de beperkte omvang van de onderzoekpopulatie zijn kwantitatieve gegevens bij ontwikkelingsonderzoek nooit hard. Generaliseren op zo'n smalle basis is een hachelijke zaak. Dit type onderzoek toont slechts aan dat iets (on)mogelijk is. Ons inziens is Lemke's (1998, p. 1185) opvatting hier (m.m.) van toepassing:

> Discourse analysis will not tell us a lot about how all classrooms or all science writing is alike (though it will tell us something), but it provides us with the tools to analyse and understand more exactly what is going on in any particular discourse or text that we wish to analyse. That is as much as any theory does in practice.

Toch had een en ander beter gekund. De toets die in de eerste ronde een erg lage betrouwbaarheid vertoonde is wel bijgesteld, maar de betrouwbaarheid is daardoor eigenlijk te weinig toegenomen. Aangezien we van mening zijn dat validiteit hier voor betrouwbaarheid moet gaan, zijn we niet zover gegaan om de hele toets aan de kant te schuiven. De gegevens over het gebruik van de verschillende elementen van begrip van bewijs leveren namelijk wel degelijk zinvolle informatie. Maar met name de 'leereffect'-gegevens van de toets moeten desalniettemin met slagen om de arm geïnterpreteerd worden.

De interbeoordelaarsbetrouwbaarheid van de beoordelingen van transcripten van gesprekken en leerlingproducten blijkt acceptabel. Dat is een hele geruststelling aangezien ze gebaseerd zijn op een nieuw ontwikkeld begrippenkader. Wat echter minder zeker is, is de overeenstemming tussen de beoordelingen van de verslagen uit de voorstudie en die van de producten van de leerlingen uit Amsterdam en Amstelveen. Weliswaar is hetzelfde beoordelingsinstrument gebruikt en is een aantal verslagen in de voorstudie ook beoordeeld door een van de beoordelaars van de latere producten. Maar bij de voorstudie is geen overeenstemming berekend en bovendien zat er een tijdsspanne van twee of drie jaren tussen de ene en de andere beoordeling. Ook hier is dus voorzichtigheid op z'n plaats bij de interpretatie van de gegevens. We zijn van mening dat we die voorzichtigheid betracht hebben bij het trekken van conclusies.

- *Zeker weten?*

Ondanks de hierboven beschreven onzekerheden hebben we in paragraaf 8.1 een antwoord op de onderzoeksvraag gegeven. Het betekent echter ook dat we de vraag 'zeker weten?' ten aanzien van dat antwoord niet ronduit bevestigend kunnen beantwoorden. Toch zijn we van mening dat er voldoende grond is om de conclusies te trekken die we trekken, zij het dat we alert moeten blijven op signalen die in een andere richting wijzen. En dat geldt ons inziens ook voor de suggesties die we in paragraaf 8.4 doen op basis van het antwoord op de onderzoeksvraag. Voordat we echter in de laatste paragrafen lijnen uitzetten voor verdere onderzoek, reflecteren we in 8.3 op de theoretische uitgangspunten die aan dit onderzoek ten grondslag liggen.

8.3 Reflectie op de theoretische uitgangspunten

In hoofdstuk 1 en 2 is beschreven op welke theoretische basis het onderzoek is gefundeerd: er is een geheel van kennis waarmee de kwaliteit van onderzoek beschreven kan worden en het is zinvol die als een te onderscheiden kennisbestand in het onderwijs een plaats te geven. In hoofdstuk 4 zijn de theoretische uitgangspunten beschreven die ten grondslag liggen aan de inrichting van de praktische opdracht zoals ontwikkeld. In hoofdstuk 6 zijn die uitgangspunten aangescherpt.

Wat zeggen de uitkomsten van dit onderzoek over die theoretische uitgangspunten? Is er reden om daar kritische kanttekeningen bij te zetten? Hier willen we de belangrijkste uitgangspunten van enig commentaar voorzien.

- *Kwaliteit van onderzoek als kennisbestand*

De elementen van begrip van bewijs zoals beschreven in paragraaf 2.3 zijn een verkorte versie van de *concepts of evidence* zoals Gott en Duggan (2003) die beschrijven en op hun

website uitwerken (Gott e.a., 2006). Ook anderen hebben veel geschreven over criteria voor wetenschappelijke methoden (voor biologie zie o.a. Giere, Sloep, Reddingius & Voorzanger, 1990; Van der Steen, 1993; Allen & Baker, 2001; Van der Steen & Ho, 2001). Alleen daaruit al blijkt dat er veel te weten valt over methoden en kwaliteit van wetenschappelijk onderzoek. Dat het een geheel van kennis ís, staat hier dan ook niet ter discussie, wel de vraag of die 'apart' onderwezen moet worden.

In de ontwikkeling van de onderwijsleerstrategie is ervan uitgegaan dat het uitvoeren van een onderzoek voldoende ruimte biedt om gedegen kennis over de kwaliteit van onderzoek op te doen. De startlessen van de tweede versie waren er expliciet op gericht om al opgedane kennis op dit terrein te activeren en zo een basis te leggen voor verdere ontwikkeling ervan. Geconcludeerd is dat de startlessen er in redelijke mate in geslaagd zijn dat te doen. Tegelijkertijd is geconcludeerd dat ten aanzien van enkele elementen van begrip van bewijs er na afloop van de praktische opdracht nog steeds kennislacunes bestonden, met name met betrekking tot het opstellen van een voorspelling ter toetsing van een hypothese en het principe van een blanco. De gesprekken van de docenten over het eigen onderzoek en de reflectieopdrachten bleken dat niet voldoende te hebben gestimuleerd. Anderzijds gaven de leerlingen wel aan dat ze meer over het doen van goed onderzoek geleerd hadden dan over het onderwerp van hun eigen onderzoek. Hun onderzoeken gingen uiteraard wél over vakinhoud, maar het is dus niet zo dat het leren over de kwaliteit van onderzoek daardoor is ondergesneeuwd.

Betekent dit dat er in de praktische opdracht nóg meer accent gelegd moet worden op kennis over bewaken van kwaliteit of betekent het dat er iets anders nodig is? De vraag komt op of het uitvoeren van één compleet, weken durend eigen onderzoek genoeg is om voldoende over het bewaken van de kwaliteit te leren. In de startles is de voorkennis wel geactiveerd, maar ten aanzien van bijvoorbeeld de blanco was die voorkennis gering. Ook bleek uit de analyse van schoolboeken al dat de aandacht voor kwaliteitscriteria voor onderzoek klein is.

Het uitgangspunt van de onderwijsleerstrategie dat de kwaliteitscriteria voor onderzoek dan wel in voldoende mate aanwezig zijn, dan wel in voldoende mate bij het doen van een groot eigen onderzoek geleerd kunnen worden, lijkt bijgesteld te moeten worden. Er is meer voor nodig dan één grote onderzoeksopdracht. Hoe dat vorm zou kunnen krijgen, wordt geschetst in paragraaf 8.4.

- *Denken, doen en reflecteren*

Uitgangspunt voor het leren bewaken van de kwaliteit van onderzoek is gezocht in de samenhang van denken, doen en reflecteren. In figuur 4.1 is weergegeven hoe die samenhang in talige interactie zou kunnen gestalte krijgt tussen het concrete uitvoeren (doen), het expliciete nadenken over de kwaliteit (denken) en het ontwikkelen van algemene criteria voor goed onderzoek (reflecteren). Figuur 4.2 laat min of meer analoog daaraan zien hoe Galperin (volgens Arievitch en Haenen, 2005) zich het leerproces, de 'spiral formation of mental actions', voorstelde. Na een oriëntatie op basaal niveau gaat het via het concrete niveau van handelen, het verbale niveau van verwoorden naar het niveau van de mentale structuur, om vervolgens te leiden tot een hernieuwde oriëntatie op hoger niveau, waarna een nieuwe cyclus kan beginnen. Die cirkelgang is in het onderzoek echter maar zelden waargenomen. Meestal leken de leerlingen onderin de cirkel heen en weer te schommelen, tussen concrete problematiseringen en dito beschrijvingen met af en toe een uitstapje naar een explicitering. De beweging leek noch vooruit te gaan door een algemene regel af te leiden uit de explicitering ('oh, dus zo zit het'), noch achteruit te gaan en terug te grijpen op een eerdere generalisatie ('o ja, dat zat zo'). Was de impuls om de schommel 'over de kop' te laten gaan afwezig of niet groot genoeg? Of bleef de stap van het verankeren in de mentale structuur meestal onuitgesproken? De vraag is dan ook of die cirkelgang niet via gerichte docentinstructie gestimuleerd moet worden, als hij blijkbaar zelden spontaan optreedt.

Veralgemeniseren is op te vatten als recontextualiseren (Van Oers, 1998b; zie noot 12 op p. 58), dat wil zeggen dat om te kunnen generaliseren de begrippen minstens in een tweede context aan de orde moeten komen. Als dat zo is, zou eventuele leerwinst wellicht waar te nemen zijn in de kwaliteit van een hernieuwde oriëntatie, bijvoorbeeld als de leerlingen opnieuw een onderzoek zouden moeten opzetten en uitvoeren.

Idee achter het laten uitvoeren van een eigen onderzoek door de leerlingen was 'leren door doen', waarbij gerefereerd werd aan het concept van *cognitive apprenticeship* (Collins e.a., 1989): in een meester-gezel relatie leert de novice de kneepjes van het vak van de expert. In paragraaf 4.1 is al opgemerkt dat de praktijk van onderzoek doen op scholen niet dezelfde is als de 'echte' onderzoekspraktijk van de wetenschap. In de uitvoering van een praktische opdracht lijkt zich dat te wreken. De docenten zijn geen onderzoekers en missen expertise, zowel wat betreft de uiteenlopende inhoud en de technieken van de onderzoeken als de bewaking van de kwaliteit. Bovendien onderzoeken ze niet mee. In een situatie waar de docent ook als onderzoeker betrokken is in een gezamenlijk onderzoeksproject, kan hij bovendien als rolmodel c.q. 'meester' fungeren (bijv. Crawford, 2000). Als het leren bewaken van de kwaliteit van onderzoek beschouwd wordt als *enculturation* in de praktijk van de wetenschapsbeoefening (Hodson & Hodson, 1998b), dringt zich de vraag op hoe dat moet als op scholen die cultuur niet aanwezig is. En als het doen van natuurwetenschap op school eigenlijk als een eigen praktijk met eigen gewoontes, regels en wijzen van spreken gezien zou moeten worden (Mortimer & Scott, 2003), lijkt het erop of die op de onderzoeksscholen nog niet voldoende gestalte heeft gekregen.

Dat noopt ons ook de ontwikkeling van een activiteit van figuur 4.4 te bekijken. Het geeft de (mogelijke) relaties weer tussen behoeften, motieven en activiteiten in de praktijk van onderzoek doen. De behoefte om iets te weten te komen vormt het motief om onderzoeksactiviteiten te ontplooien. De resultaten leiden tot reflectie op de kwaliteit ervan en tot de vraag of de behoefte in voldoende mate vervuld is: weten we het nu en weten we het zeker genoeg? Het eerste deel van het motief bleek bij de leerlingen in voldoende mate aanwezig, het tweede deel minder en een mogelijk derde deel: *hoe weten we of we het zeker genoeg weten?* – de vraag naar criteria – bleek vrijwel afwezig. Dit lijkt een cruciaal verschil tussen de praktijk van de 'echte' wetenschap en wetenschap op school. In de 'echte' wetenschap word je beoordeeld op die zekerheid, op school speelt dat veel minder. Als je aan kan geven waar zwakke plekken zitten, is het vaak al voldoende.

Het gelijkstellen van het leren over onderzoek doen in de praktijk op school met die van de wetenschap is dus niet probleemloos. Bij het gebruiken van wetenschappelijke contexten in de schoolsituatie moet daar terdege rekening mee worden gehouden.

- *Motief als drijvende kracht*

Na het uitproberen van de eerste versie van de praktische opdracht is al geconcludeerd dat er meer nadruk gelegd moest worden op de ontwikkeling van het motief *Zeker weten?* De opdracht is aangepast, er is een didactische structuur ontworpen (fig. 6.1), gebaseerd op de probleemstellende benadering (Lijnse & Klaassen, 2004). Deze laat een beweging zien tussen uitvoeren van een onderzoek en reflecteren op de kwaliteit ervan. Na een aantal keren tussen het concrete, beschrijvende en het reflectieve, expliciterende niveau heen en weer gegaan te zijn, zou de behoefte c.q. de bereidheid ontstaan om te veralgemeniseren. Drijvende kracht om deze beweging in gang te zetten en te houden zou het motief om zeker te weten zijn.

In de tweede ronde was er wel een motief bij de leerlingen aanwezig om een goed onderzoek te doen, maar nog steeds geen motief om te leren over de criteria in het algemeen; de stap naar het gegeneraliseerde niveau werd niet spontaan genomen. De leerlingen bleken geen expliciete behoefte, geen direct belang te hebben om uit te stijgen boven het concrete niveau van hun eigen onderzoek. Eigenlijk is dat wel logisch, omdat de leerlingen beoordeeld werden

op een goed product en niet op het kunnen generaliseren. Om de didactische structuur in de praktijk te laten functioneren, is consistentie tussen doelstellingen, motief en beoordeling uiteraard nodig. Dat kan in de vorm van een schoolse beoordeling, maar misschien ook op een andere manier, daarop wordt in de volgende paragraaf verder ingegaan.

8.4 Lijnen naar betekenisvol leren

In paragraaf 7.6 is beschreven wat er beter zou kunnen in een eventuele derde versie van de praktische opdracht. Daarbij aansluitend doen we hier suggesties die dat toch beperkte kader overstijgen. Het leren bewaken van de kwaliteit van onderzoek lijkt meer nodig te hebben dan één eigen onderzoek en meer dan de schoolpraktijk als context.

Wat nodig lijkt, is het vinden van meer gelegenheden om de spiraal van beschrijven, expliciteren en generaliseren te laten draaien. We zien twee mogelijkheden om dat te stimuleren. Ten eerste een doorgaande lijn, waarbij spiraalsgewijs op concreet, verbaal en mentaal niveau begrip van kwaliteit van onderzoek opgebouwd kan worden. 'Steeds beter' wordt dan een drijvende kracht. Ten tweede door het directe belang, de directe betekenis van het onderzoek te vergroten, bijvoorbeeld door een grotere authenticiteit van de (grotere) onderzoekscontext of door beoordeling door anderen. 'Zeker weten?' krijgt dan extra belang en betekenis.

- *Leerlijnen*

Onderzoek doen leer je niet in één keer. Daarvoor is het een veel te complex geheel van kennis en vaardigheden. Dat wordt algemeen onderkend en in de schoolboeken voor biologie zijn dan ook in alle jaren wel onderzoeksopdrachten te vinden. Het leren bewaken van de kwaliteit van onderzoek en het expliciteren van criteria daarvoor is ook iets dat niet in één stap gerealiseerd kan worden. Naast voldoende aanwezigheid van begrip van bewijs in methoden en leeractiviteiten, zowel door de jaren heen (steeds een beetje meer) als in de breedte van het vak (hoe hebben ze dat eigenlijk uitgezocht?) en zelfs over vakken heen (hoe doen ze dat bij natuurkunde?), is het nodig dat er een doelgerichte opbouw wordt gecreëerd. Nodig lijkt een doorgaande lijn in het onderwijs. Het opdoen van kennis over kwaliteitscriteria zou in de loop van jaren moeten plaatsvinden.

Met het accent dat de tweede fase legt op vaardigheden zijn er door veel uitgevers boeken ter ondersteuning van het leren daarvan op de markt gebracht (bijvoorbeeld De Boer, De Jong & Van Warmerdam, 1996; Wijgerinck, 1998), soms specifiek gericht op de bètavakken of onderzoek doen (bijvoorbeeld de vaardighedenboeken die geanalyseerd zijn of Claessens & Van Griethuysen, 2001; Geene, 2004). Deze boeken beschrijven wel een scala aan vaardigheden en hoe je die goed kunt uitvoeren, maar kennen geen opbouw in didactische zin. Daarvoor is een leerlijn nodig.

Wat onder een leerlijn wordt verstaan lijkt overigens niet altijd hetzelfde; soms ligt de nadruk op de samenhang voor de leerlingen, soms op de planning van leerstofonderdelen of activiteiten. Hubers (2004, p. 10) omschrijft een leerlijn als:

> Een opzet van het onderwijs waarin de leerling zo min mogelijk overlap, breuken of lacunes in zijn/haar leerproces ervaart.

De omschrijving die de Biologische Raad (KNAW, 2003, p. 20) geeft, is daar niet mee in tegenspraak, maar legt wel een heel ander accent:

> Onder een leerlijn wordt verstaan een opeenvolging van leeractiviteiten die zich richt op de systematische ontwikkeling van een of meer specifieke doelstellingen (begrippen, vaardighe-

den, competenties). Soms wordt alleen de inhoudelijke component (leerstof) als leerlijn weergegeven.

Van der Valk en Van Soest (2004, p. 7) merken op dat er eigenlijk een onderscheid gemaakt moet worden tussen een *onderwijslijn* (in het onderwijs dat aangeboden wordt) en een *leerlijn* (in wat er door de leerlingen geleerd wordt) en benadrukken 'dat er een *voor leerlingen inzichtelijke* voortgang zou moeten zijn in de eisen die aan de kwaliteit van vragen en antwoorden gesteld worden' (onze cursivering). Terecht benadrukken ze dat het niet alleen gaat om steeds vaardiger worden, maar dat er ook sprake moet zijn van een opbouw in complexiteit en kwaliteit en dat die ontwikkeling ook voor de leerlingen waarneembaar moet zijn.

In aanzetten voor leerlijnen met betrekking tot onderzoek doen is toenemende complexiteit te herkennen. Bax e.a. (1997), bijvoorbeeld, hebben activiteiten die in de drie natuurwetenschappelijke vakken met onderzoeksvaardigheden te maken hebben in tijd en moeilijkheid op een rij gezet. Vaak wordt bij het doen van onderzoek min of meer van achteren naar voren gewerkt: de onderdelen aan het einde van een onderzoeksproces, de uitvoering en de conclusie, worden meestal het eerst aan de leerlingen overgelaten, de onderzoeksvraag meestal pas het laatst. Falk e.a. (1974) spreken over 'vrijheidsgraad' als mate van openheid voor de leerlingen; ze onderscheiden zes vrijheidsgraden van een practicum, van 'illustratief' tot 'experimenteel'. Pas bij de grootste openheid kiezen groepjes leerlingen verschillende problemen. Ook Van Rens en Dekkers (2000, p. 90) achten het formuleren van een onderzoeksvraag het sluitstuk:

> Wie geheel zelfstandig een onderzoeksvraag wil formuleren waarop experimenteel onderzoek een antwoord kan geven, dient het gehele onderzoekstraject al in zekere mate te overzien. De onderzoeker dient dus alle relevante handelingen al in voldoende mate te beheersen. Het vermogen om *geheel zelfstandig* onderzoeksvragen te formuleren, komt dus in het traject pas laat aan bod. (cursivering van hen)

Anderen (o.a Chin, 2004; Van Eijck, 2005) zijn van mening dat je juist bij de vragen van leerlingen moet beginnen om een onderzoekende houding te stimuleren, voordat je ze in het keurslijf van een experimentele methode dwingt, als je dat al doet. Ook Van der Valk en Van Soest (2004) betwijfelen of de mate van openheid wel een factor voor de ordening van een leerlijn moet zijn. In hun onderzoek bleek dat leerlingen van een school waar het accent lag op het aanleren van onderzoeksvaardigheden een door de docenten opgezette leerlijn minder goed herkenden en minder waardeerden dan leerlingen van een school waar het accent lag op het aanbrengen van een onderzoekende houding. Als we in een leerlijn vooral het motief 'steeds beter?' de kans willen geven zich te ontwikkelen, lijkt deze laatste insteek meer kansen te bieden dan het stapsgewijze 'loslaten' van de leerlingen.

Hubers (2004) schetst een leerlijn van basisonderwijs tot hoger onderwijs, beginnend bij vragen stellen over natuurwetenschappelijke verschijnselen in het dagelijks leven in het basisonderwijs, via het vertalen van problemen in vragen die met onderzoek te beantwoorden zijn in de basisvorming en het formuleren van toetsbare hypotheses in de tweede fase naar het opzetten en uitvoeren van een wetenschappelijk onderzoek in het hoger onderwijs. De houding die daarbij hoort gaat van 'zelf op zoek gaan naar antwoorden op hun vragen' (basisonderwijs) via 'een idee ontwikkelen van wat goed onderzoek inhoudt' (basisvorming) en 'interesse en gevoel voor onderzoek in realistische situaties buiten hun *eigen huis en tuin* context'(tweede fase) naar 'in staat zijn een eigen positie in te nemen in de wetenschappelijke wereld en met betrekking tot de rol van wetenschappelijk onderzoek in de praktijk' (hoger onderwijs). Het is uitdrukkelijk een schets, bedoeld voor debat, waar echter duidelijk aankno-

pingspunten zijn om met kwaliteit(scriteria) van onderzoek bezig te zijn op verschillende niveaus. Van der Valk en Van Soest (2004, p. 8) verwoorden het zo:

> Vooral het aspect kritische houding zal een vooruitgang moeten tonen op punten als validiteit van begrippen en werkwijzen en betrouwbaarheid van methoden en resultaten, maar ook interne consistentie en het aansluiten van resultaten bij die van anderen.

Naast de bovengenoemde initiatieven in Nederland, kunnen bouwstenen voor een leerlijn in het buitenland gezocht worden. In Engeland is relatief veel ontwikkeld, zowel voor de onderbouw als voor de hogere leerjaren (onder andere Goldsworthy e.a., 2000; Gott e.a., 1997, 1998, 1999; Sang & Wood-Robinson, 2002). Deze boeken bevatten allerlei leeractiviteiten waarmee het begrip 'valide onderzoek' in stapjes kan worden opgebouwd. Hoewel veel van die activiteiten de natuur- en scheikundige kant van science als onderwerp hebben, is er ook veel voor biologie te vinden.

Losse elementen zijn echter wel noodzakelijke, maar zeker nog geen voldoende voorwaarde voor het ontwikkelen van een solide leerlijn voor het leren bewaken van de kwaliteit van onderzoek. Waar het om gaat is dat er een samenhang gecreëerd wordt van leeractiviteiten gericht op het centrale concept 'onderzoek doen' waarbij dat vanuit een kiem – zoals de kern van het hypothetisch-deductief redeneren – langzaam verder uitgebouwd en ontwikkeld wordt.

Van der Valk en Van Soest (2004) breiden het begrip leerlijn nog wat verder uit; naast een opbouw in de tijd (wat zij de verticale leerlijn noemen) is er de noodzaak van afstemming en samenhang over vakken heen (wat zij de horizontale leerlijn noemen). Dat is niet eenvoudig te realiseren (Bax e.a., 1997; Millenaar, Mooldijk & Den Boer, 1997), maar Van der Valk e.a. (2005) laten zien dat er veel te bereiken is als docenten, eventueel samen met begeleiders/onderzoekers, een *discourse community* vormen, waarin overeenstemming bereikt wordt met betrekking tot de leerinhoud – in dit geval onderwijzen tot een onderzoekende houding in de bètavakken – en uitvoering.

De voorwaarden lijken gunstig om in Nederland tot leerlijnen te komen. De Commissie Vernieuwing Biologieonderwijs, die de opdracht heeft nieuwe examenprogramma's voor havo en vwo te ontwikkelen, zet die programma's ook uitdrukkelijk in het kader van een leerlijn voor biologie voor 4- tot 18-jarigen. In een conceptversie van die leerlijn (Boersma e.a., 2006) worden de denk- en werkwijzen 'beschrijvend en vergelijkend onderzoeken' en 'experimenteel onderzoeken' gekoppeld aan de activiteit 'onderzoeken', een activiteit die in alle delen van de leerlijn gepropageerd wordt en voor havo en vwo wordt omschreven. In de conceptexamenprogramma's voor havo en vwo (Commissie Vernieuwing Biologieonderwijs, 2006a, 2006b, p. 4) wordt bij die activiteit o.a. vermeld: 'omschrijven aan welke criteria valide uitgevoerd biologisch onderzoek moet voldoen.' Er is intensief overleg met de commissies die voor natuurkunde, scheikunde en wiskunde nieuwe examenprogramma's ontwikkelen. Maar veel zal moeten gebeuren op schoolniveau. Een aanzet is te vinden bij Brouwers, Halfwerk, Spek, Visser, Legierse en Oosterbaan (2003), die een 'praktijkhulp' hebben ontwikkeld waarmee docenten in vijf bijeenkomsten een leerlijn onderzoeksvaardigheden voor vmbo-BBL-leerlingen kunnen maken of beter gezegd kunnen implementeren. Want die publicatie bevat al een aantal uitgewerkte (deel)leerlijnen die de basis vormen voor een schoolleerlijn. Zoiets zou ook ontwikkeld kunnen worden voor een leerlijn 'kwaliteit van onderzoek bewaken'. De bevindingen die in dit proefschrift beschreven zijn, kunnen daar hopelijk een bijdrage aan leveren.

- *Betekenisvol leren*

De praktijk van wetenschap op school wijkt op cruciale punten af van die van de 'echte' wetenschap, onder andere op het punt van het belang dat gehecht wordt aan een zeker ant-

woord. Een motief om de kwaliteit te bewaken zou dan ook kunnen liggen in een verbinding met de praktijk van echte wetenschap. We willen hier twee – elkaar niet uitsluitende – mogelijkheden schetsen om daar meer gestalte aan te geven: meer authenticiteit van de onderzoekscontext en beoordeling door anderen.

Authenticiteit van de onderzoekscontext

Met 'authentiek leren' wordt enerzijds bedoeld dat onderwijs aansluit bij de behoeften van leerlingen en anderzijds dat je het best kunt leren in een levensechte of realistische omgeving (Volman, 2006). Maar onder authenticiteit van een onderzoeksopdracht wordt niet door iedereen hetzelfde verstaan. Wij willen hier aansluiten bij de omschrijving die Crawford e.a. (1999, p. 703) geven:

> Authentic instruction involves creating environments similar to those of practitioners (Resnick, 1987) and related to students' own interests. In the case of the science classroom, authentic instruction engages students in pursuing answers to significant questions that are relevant to students and similar to activities of professional scientists (Brown, Collins & Duguid, 1989). Solving problems and doing 'real' science in the classroom stems from Dewey (1938) who argued that the school community is a place where a child gains authentic experiences by doing projects rather than solving isolated problems.

Authenticiteit hoeft dus niet per se te betekenen dat het om een reële vraag uit de maatschappij gaat, maar wel om een realistische en voor de leerlingen betekenisvolle vraag. Het 'similar to activities of professional scientists' uit bovenstaand citaat sluit aan bij de concept-context benadering die de basis vormt van de ontwikkeling van nieuwe examenprogramma's voor de exacte vakken (o.a. Boersma e.a., 2006, p. 12, zie ook par. 3.1), waarbij onder een context het volgende wordt verstaan:

> Onder een context verstaan we een 'praktijk' waarbinnen een aantal cultuurhistorisch bepaalde activiteiten worden uitgevoerd, die gericht zijn op het realiseren van de doelen die in die context worden nagestreefd. Bij de uitvoering van voor de context kenmerkende activiteiten worden veelal specifieke materialen, gereedschappen, symbolen en kennis gebruikt. Een context heeft eigen spelregels, die bepalen hoe mensen met elkaar omgaan, welke dingen belangrijk zijn en hoe zaken moeten worden opgevat.

Hoe kan de praktijk van onderzoek doen op school meer gaan lijken op die 'authentieke' onderzoekscontext, ook al gaat het om een gedidactiseerde versie ervan? Waarschijnlijk op vele manieren. We noemen er hier twee: door het te koppelen aan een direct maatschappelijk belang en door het onderzoek van de leerlingen te plaatsen in het kader van een groter onderzoeksprogramma.

Als met het onderzoek een direct maatschappelijk belang gediend wordt, is het duidelijk dat aan de kwaliteit van de resultaten belang wordt gehecht. Een bekend voorbeeld is het onderzoek in Oceanside (Roth & Lee, 2004), waar leerlingen met onderzoek aan de waterkwaliteit van een beek bijdroegen aan een lokale discussie over watervervuiling door bedrijven. De resultaten dienden dus overtuigingskracht te hebben voor de lokale bestuurders. Opmerkelijk is wel dat de auteurs het 'koloniaal' vinden als er aan de resultaten van de leerlingen dezelfde eisen worden gesteld als in de wetenschappelijke wereld. Als de gemeenschap het maar genoeg vindt, is het genoeg. Nadeel van koppeling aan een concreet lokaal maatschappelijk belang is de geringe herhaalbaarheid; als het maatschappelijke belang is gediend, is er geen aanleiding het onderzoek te herhalen met een volgende klas.

In de hedendaagse (natuur)wetenschap werkt vrijwel niemand meer in zijn eentje aan een onderwerp, maar vindt onderzoek plaats binnen het onderzoeksprogramma van vakgroe-

pen en onderzoeksscholen. Er zijn verschillende initiatieven genomen en uitgewerkt om leerlingen in zo'n meer realistische onderzoeksituatie te plaatsen.

In de biologiemethode *Synaps* (Molenaar & Schermer, 1999) worden leerlingen regelmatig in de situatie van een vakgroep gebracht. Leerlingen zoeken niet allemaal hetzelfde uit, maar leveren elk – vaak in het verband van een 'expertgroep' – een bijdrage aan de beantwoording van de centrale vraag. Daarvoor is het uiteraard van belang dat de kwaliteit van de ingebrachte deelantwoorden voldoende is. Deze *onderlinge afhankelijkheid* wordt geacht de kwaliteitsbewaking te bevorderen. Dat blijkt ook te gebeuren in andere van dit soort *communities of learners* (Crawford e.a., 1999). Crawford (2000) beschrijft een situatie waarin de klas als geheel een onderzoeksproject aanpakt en waarin iedereen het belang van kwaliteit inziet. De docent is hier echter wel de motor die veel functies moet combineren (o.a. coachen en model staan als onderzoeker).

In het project *De Vrolijke School* van het Onderwijscentrum VU en het Sint Ignatiusgymnasium in Amsterdam worden projecten opgezet waarin leerlingen werken aan een onderzoeksvraag die voortkomt uit het onderzoeksprogramma van een universitaire vakgroep (Van der Schee, Van Rens & Beishuizen, 2006). Het *werken aan vragen waarop het antwoord nog niet bekend is* blijkt motiverend te werken en maakt leerlingen nieuwsgierig naar het antwoord en de zekerheid ervan. Het kost de leerlingen wel enige tijd om te beseffen dat ze niet aan het zoeken zijn naar 'paaseieren'[34], waarvan de docent al weet waar ze liggen. Voordeel van de samenwerking met universitair onderzoekers is dat het model staan van de docent als onderzoeker dan minder noodzakelijk is; er is immers een 'echte' onderzoeker aanwezig. Expliciteren van die rol lijkt echter wel noodzakelijk. Om een positieve verandering teweeg te brengen in kennis en attitude blijkt het ook nodig dat de opdracht van begin tot eind uitdagend is door te werken aan zelfgestelde vragen die met de vakinhoud te maken hebben.

Er zijn ook verschillende landelijke en internationale projecten waarin leerlingen kunnen participeren door het verzamelen van gegevens in samenwerking met onderzoekers van researchinstituten. De landelijk of internationaal verzamelde gegevens zijn dan vaak beschikbaar om eigen onderzoeksvragen op los te laten. Voorbeelden zijn *De grote griepmeting* (www.degrotegriepmeting.nl), *Koetnet* (www.nioo.knaw.nl/koetnet), *Science across the world* (www.scienceacross.org) en *Coastwatch Online* (www.coastwatch.nl).

Een andere manier om te participeren in een onderzoeksprogramma wordt mogelijk als de school of sectie zélf zo'n programma heeft, bijvoorbeeld rond de ecologie van een natuurgebied in de buurt of rond het gedrag van de eenden in de sloten naast de school. Onderzoek levert altijd nieuwe vragen op en dus kunnen leerlingen voortbouwen op de resultaten van anderen. De docenten bouwen in de loop der jaren expertise op in het veld van onderzoek en worden daarmee goede rolmodellen voor de leerlingen als onderzoekers.

Ook is er de mogelijkheid dat leerlingen meedoen in het onderzoeksprogramma van een (natuur)vereniging en daar hun motief aan ontlenen om de kwaliteit van het onderzoek en de resultaten te bewaken (zie bijv. Piers, 2000).

Beoordeling door anderen

Beoordeling van de kwaliteit van onderzoek door anderen dan de onderzoekers zelf speelt in het wetenschapsbedrijf een grote rol. Visitatie door collega-onderzoekers en het reviewen van ter publicatie aangeboden artikelen zijn niet weg te denken en worden als garantie voor kwaliteit beschouwd (Bauer, 1992). Het benadrukt de sociale kant van wetenschapsbeoefening. Als gezocht wordt naar mogelijkheden om de kwaliteit van leerlingonderzoek op school door anderen – ook anderen dan de docent – te laten beoordelen, kan dat op verschillende manieren. We bespreken hier de rol van peer review en van jurybeoordelingen.

[34] Deze prachtige metafoor wordt aan Hans Freudenthal toegeschreven

Met peer review (of peer assessment) wordt de feedback bedoeld die onderzoekers krijgen van medeonderzoekers. In een schoolcontext betekent dat feedback van medeleerlingen uit de eigen klas, school of andere scholen. Net als in de onderhavige studie is in verschillende onderzoeken aangetoond dat het leerlingen of studenten aanzet hun voorlopige producten bij te stellen.

Feedback via internet van leerlingen die aan hetzelfde project werkten (Van Rens, 2005) gaf met name kwaliteitsverbetering met betrekking tot betrouwbaarheid en nauwkeurigheid (Hofman, 2006). Lui, Pysarchik en Taylor (2002), lieten studenten biologie zowel de onderzoeksvoorstellen als de onderzoeksverslagen van medestudenten beoordelen aan de hand van gegeven criteria. De studenten waren grotendeels positief over het geven en ontvangen van commentaar op elkaars werk, en beschouwden het als een stimulans tot kritisch nadenken én als een bijdrage aan de kwaliteit van hun product. Ze vonden het een wezenlijk deel van hun wetenschappelijke opleiding.

Van den Berg (2003) beschrijft op basis van een onderzoek bij geschiedenisstudenten de kenmerken van een optimaal ontwerp voor peer assessment: het moet een conceptversie van het product betreffen, aanvullend zijn aan de docentbeoordeling, wederzijds plaatsvinden in groepen van twee of drie personen die at random zijn samengesteld rond een verwant onderwerp, zowel schriftelijk (buiten contacttijd) als mondeling (binnen contacttijd) gegeven worden en niet apart beloond. Reflectieopdracht 4 heeft bijna al deze kenmerken, alleen wordt daarin, voorafgaand aan de bespreking, geen schriftelijke feedback opgesteld buiten de les.

In hoeverre het vooruitzicht dat het product door medeleerlingen becommentarieerd gaat worden een motief genereert om tijdens de uitvoering de kwaliteit beter te bewaken én om daarbij de criteria voor die kwaliteit te generaliseren is in de geciteerde publicaties echter niet beschreven.

Jurybeoordelingen van prijzen voor leerlingonderzoek wegen uiteraard niet alleen de methodologische kwaliteit, maar ook zaken als creativiteit en relevantie (zie bijv. Marijnissen, 2005). Toch wordt er wel degelijk gelet op de degelijkheid van het ontwerp en de uitvoering van het onderzoek. De vraag of leerlingen die prijzen voor profielwerkstukken hebben gewonnen extra op de criteria voor kwaliteit hebben gelet wordt daarmee echter niet beantwoord. Het vooruitzicht dat met een onderzoek een prijs gewonnen kan worden en het winnende artikel in een bekend tijdschrift gepubliceerd wordt, blijkt leerlingen in het scheikundeonderwijs echter zeker te motiveren tot kwaliteit (Van Rens, 2005).

8.5 Verder onderzoeken

Onderzoek doen is gericht op het vinden van een antwoord op een vraag. Maar niet altijd is het gevonden antwoord afdoende en meestal roept zo'n antwoord ook weer nieuwe vragen op. Zo is het ook met dit onderzoek. Zoals in paragraaf 8.2 al aangegeven is, is het antwoord op de centrale vraag niet geheel zeker en de suggesties uit 7.6 en 8.4 vragen ook om verdere onderbouwing. Daarom doen we in deze paragraaf enkele suggesties voor nader onderzoek.

- *Het begrippenkader*

In paragraaf 8.2 hebben we betoogd dat onze uitwerking van de *concepts of evidence* en het begrippenkader om de uitspraken te categoriseren robuust genoeg zijn. Toch zou verder onderzoek daarnaar nuttig zijn. Met name zou gezocht kunnen worden naar een Nederlands draagvlak voor de elementen van begrip van bewijs, wellicht uitgebreid naar de andere natuurwetenschappelijke vakken. Ons inziens zou daarbij vooral de hypothetisch-deductieve

manier van redeneren bij het doen van onderzoek centraal moeten staan, niet als een keurslijf (Van Eijck, 2005), maar als een basisrichting voor het kritisch denken over kwaliteit van bewijsvoering: 'het lijkt wel zo, maar is het wel echt zo?' dan wel 'kan ik een mogelijke verklaring aan een dusdanige test onderwerpen, dat hij alleen overeind blijft als de verklaring klopt?' Initiatieven die nu in Nederland bestaan voor leren onderzoek doen met kwaliteit (o.a tot uitdrukking komend in Hubers, 2004; Smits, 2003; Van Rens, 2005; Van der Valk e.a., 2005) bieden wellicht mogelijkheden om tot een gezamenlijk begrippenkader van kwaliteit van onderzoek te komen.

De koppeling van de reeks *problematiseren – beschrijven – expliciteren – generaliseren – toepassen* aan het model van 'spiral formation of mental actions' van Galperin is ook een onderwerp dat nadere studie behoeft. Is de verbinding niet te makkelijk gelegd? Onderzoek hiernaar zou zowel vanuit een theoretische (vanuit de cultuurhistorische theorie) als vanuit een praktische invalshoek gedaan kunnen worden om meer zicht te krijgen op hoe criteria voor goed onderzoek in de mentale structuur verankerd worden. Een secundaire analyse van de geregistreerde gesprekken zou hier wellicht van nut kunnen zijn.

- *Evalueren, toetsen en beoordelen*

Dit onderzoek was primair gericht op het ontwikkelen van een onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek. Om de ontwikkelde onderwijsleerstrategie te evalueren zijn verschillende instrumenten ingezet. Er is uitgebreid beschreven hoe het onderwijsleerproces heeft plaatsgevonden en dat is volgens sommige onderzoekers ook de meest aangewezen weg (cf. de *thick descriptions* in het citaat van Roth op p. 189). Ook heeft een analyse van leerlingproducten plaatsgevonden, waarbij bekeken is of in de leerlingonderzoeken voldaan dan wel beargumenteerd niet voldaan was aan de elementen van begrip van bewijs. Daarnaast is een toets ontwikkeld en afgenomen. We zijn van mening dat door die triangulatie een goed beeld ontstaan is van de uitvoerbaarheid en de effectiviteit van de onderwijsleerstrategie. Desalniettemin zou verder onderzoek met betrekking tot evaluatie-instrumenten zeker zinvol zijn, bijvoorbeeld met betrekking tot de operationalisering van de elementen van begrip van bewijs in toetsopgaven. Veenhoven (2004) heeft iets dergelijks gedaan voor het ontwerpen van een onderzoek bij het vak aardrijkskunde, als was daarbij voor een toets over het formuleren van een onderzoeksvraag met deelvragen en het maken van een onderzoeksopzet al 90 minuten nodig. Ook Roberts en Gott (2004a, 2004b) hebben een toets voor de *concepts of evidence* gemaakt en zij propageren zelfs om zo'n test landelijk in te voeren als alternatief voor praktische toetsen. Zij stellen in hun (60 minuten) toets veel meer elementen van begrip van bewijs aan de orde door leerlingen 19 verschillende problemen voor te leggen. De resultaten van hun toets correleren met die van andere beoordelingen van het onderzoek van leerlingen.

De manier van beoordelen in welke mate begrip van bewijs in de artikelen of verslagen van leerling aanwezig is, verdient eveneens verder onderzoek. Dergelijke analyses kunnen heel ver gaan, door veel epistemologische niveaus van argumenteren te onderscheiden (Takao & Kelly, 2003), maar het kan ook eenvoudiger, bijvoorbeeld door te analyseren of er een sluitende hypothetisch-deductieve redenering is gevolgd (Lawson, 2002).

De cijfers die de leerlingen die met het onderhavige onderzoek meededen voor hun artikelen hebben gekregen (van hun docent), zijn gebaseerd op de eisen zoals die in de boekjes beschreven zijn. Daarin is weliswaar aandacht voor de kwaliteit van hun onderzoek, maar toch liep het niet geheel synchroon met het toepassen van de beoogde elementen van begrip van bewijs. Als het doel van het doen van onderzoek door leerlingen (ook) het leren over kwaliteit van onderzoek is, moet dat ook in de beoordeling tot uitdrukking komen. Een dergelijke discrepantie van doelen en beoordelingscriteria is niet onbekend (Stokking & Van der Schaaf, 2000), maar moet uiteraard zoveel mogelijk vermeden worden. Ook Giddings, Hof-

stein en Lunetta (1991) benadrukken dat de validiteit van veel beoordelingsinstrumenten voor leerlingonderzoek betwijfeld moet worden, omdat ze niet meten wat er nagestreefd wordt of zou moeten worden. Een manier om dat tegen te gaan is het expliciteren van de beoordelingscriteria, ook voor leerlingen. Dat kan in de vorm van eisen, zoals in het onderhavige onderzoek gedaan is, maar daarbij wordt nog niet duidelijk wanneer je aan een eis voldaan hebt. Ook als de eisen de vorm hebben van *checklists,* bijvoorbeeld voor het beoordelen van onderzoek in profielwerkstukken (Sinkeldam, 1998; Weezenbeek, 1999), worden de criteria wel duidelijk, maar blijft de norm vaak voor veel interpretatie vatbaar. Een vorm die zowel de docent als de leerlingen inzicht geeft in de criteria én de normen waaraan een product of proces moet voldoen, is die van rubrieken of *rubrics.* Hierin worden per criterium verschillende niveaus van beheersing c.q. prestatie beschreven. Voorbeelden van rubrics voor het beoordelen van onderzoek leggen vaak nog veel nadruk op vormeisen (Brouwers, 2006), maar ze kunnen ook inhoudelijk meer toegesneden worden op de kwaliteit, inclusief een weging van verschillende delen daarvan (Lunsford & Melear, 2004). De uitdaging is om rubrics te ontwikkelen en te valideren die het bewaken van de kwaliteit van het eigen onderzoek de plaats geven die het toekomt. Daarmee kunnen dan de leerlingen worden geïnformeerd en beoordeeld en kan de consistentie tussen doelen en beoordeling worden bewaakt. Tevens kan het leerproces ermee worden geëvalueerd.

- *Verdere ontwikkeling van de onderwijsleerstrategie in leerlijn en authentieke context*

In het vorige hoofdstuk is op een rijtje gezet wat er allemaal verbeterd zou kunnen worden aan de ontwikkelde praktische opdracht. Dat kan leiden tot een derde versie waarvan de uitvoerbaarheid en effectiviteit onderzocht kan worden. Maar dat zou dan meer moeten zijn dan nog een derde onderzoekscyclus aan dit onderzoek toevoegen. Gebruik op meer dan de twee proefscholen en uitwisseling van ervaringen van docenten die het elk in hun eigen praktijk beproeven, kan informatie toevoegen die in de beperkte opzet van dit promotieonderzoek niet naar boven kwam. Dan kunnen bijvoorbeeld vragen aan de orde gesteld worden met betrekking tot schoolcontext, docentdeskundigheid en inbedding in een leerlijn.

Dat leren (over de kwaliteit van) onderzoek doen in een leerlijn geplaatst zou moeten worden is in de vorige paragraaf betoogd en daar zijn verschillende initiatieven genoemd (Brouwers e.a., 2003; Hubers, 2004; Van der Valk & Van Soest, 2002). Ook de ontwikkeling van een nieuwe examenprogramma's biologie voor havo en vwo is nadrukkelijk in het kader van een doorlopende leerlijn gezet (Boersma e.a., 2006). Weliswaar is die niet exclusief gericht op leren onderzoeken, maar onderzoek en de criteria ervoor maken deel uit van het concept. Het klimaat lijkt gunstig, maar het ontwikkelen van een leerlijn in de praktijk is een lange weg met valkuilen waarvan het bestaan alleen vermoed wordt. Zo'n ontwikkeling kan niet zonder onderzoeksmatige ondersteuning. We hopen met dit onderzoek daar een bijdrage aan te leveren, maar het doortrekken van de benadering uit de ontwikkelde onderwijsleerstrategie tot delen van een leerlijn behoeft meer onderzoek. Zo zou onderzocht kunnen worden welke fasering van 'begrip van bewijs' te verkiezen is en in hoeverre die vorm te geven is als een spiraal van beschrijven, expliciteren en generaliseren.

Het pleit met betrekking tot impliciete dan wel expliciete aanleiding voor reflectie is niet beslecht in het voordeel van de een of de ander. Gebleken is dat beide een positieve bijdrage leveren aan het beoordelen en bewaken van de kwaliteit van onderzoek. Maar is 'het ene doen en het andere niet laten' wel een bevredigende oplossing voor dit dilemma? De vraag wanneer en op welke manier beide benaderingen het best kunnen worden ingezet behoeft nog nadere studie. Zo kan bijvoorbeeld ook communiceren met anderen over je onderzoek – per e-mail of via internet (De Vries, 2004; Van Rens, 2005) – reflectie stimuleren.

In de vorige paragraaf zijn ook verschillende projecten genoemd waar in meer of minder authentieke contexten aan onderzoek gedaan wordt. Vaak is daar ook onderzoek aan gekoppeld (o.a. Van der Schee e.a., 2006). Of de in de vorige paragraaf gedane suggestie dat werken aan authentieke vragen ook (meer) bewaken van de kwaliteit van dat onderzoek met zich meebrengt, zou onderzocht moeten worden. Anderzijds zou meer aandacht voor het leren bewaken van de kwaliteit van het leerlingonderzoek ons inziens een waardevolle bijdrage kunnen leveren aan die projecten, als die authentieke contexten een plaats zouden kunnen krijgen in een brede leerlijn.

Een vraag die in dit proefschrift nauwelijks aan de orde is gekomen, is die naar de relatie met de vakinhoud. Niet alleen kan de invulling van begrip van bewijs verschillend zijn in verschillende vakken of leerstofdomeinen (Brown, Moore, Silkstone & Botton, 1996; Perkins & Salomon, 1989), maar ook kan het (in par. 3.1) gemaakte onderscheid tussen *leren onderzoeken* en *onderzoekend leren* geproblematiseerd worden. Gebeurt het leren bewaken van de kwaliteit niet gewoon al doende, tijdens het werken aan een inhoudelijk bepaald onderzoek? In beschrijvingen van zogenaamde *inquiry classrooms* (bijv. Crawford, 2000; Roth & Roychoudhury, 1993; Roth, 1995) wordt betoogd dat leerlingen door al werkende aan de inhoud van het vak op een onderzoeksmatige manier een onderzoekende denk- en werkwijze leren. Om onderzoek een plaats te geven in een leerlijn zou dan onderzocht kunnen worden of de keus voor een bepaalde vakinhoud invloed heeft en zo ja, welke vakinhouden zich dan goed lenen voor het leren bewaken van de kwaliteit van onderzoek.

8.6 Hoog genoeg mikken

In de klankbordgroep die dit onderzoek heeft begeleid is meer dan eens de vraag gesteld of er niet te veel of te hoog gegrepen doelstellingen werden nagestreefd. Dat betrof zowel de ambitie om leerlingen in het bestek van tien lessen zowel een eigen onderzoek te laten uitvoeren als te laten leren over de bewaking van de kwaliteit ervan, maar ook de vraag of het bewaken van de kwaliteit van onderzoek niet een te abstract begrip is. Die zorg is begrijpelijk. Regelmatig is gerapporteerd dat het beheersen van zoiets als begrip van bewijs niet alleen voor leerlingen, maar ook voor biologiestudenten en biologiedocenten – al dan niet in opleiding – moeilijk is (bijv. Almekinders, Thijs & Lubben, 1998; Ergazaki & Zogza, 2005; Lawson, 2002; Windschitl, 2003; Yip, 1999). Ook nieuwe inzichten uit de neurologie en neuropsychologie doen vermoeden dat het brein van adolescenten nog niet uitgerust is om doordachte beslissingen en afwegingen te maken; zie bijvoorbeeld de uitspraken van Jolles (in Mols, 2002) en het boek van Strauch (2003). Als pubers nog niet in staat zijn om verantwoorde afwegingen te maken met betrekking tot hun eigen gedrag, kunnen ze dan de keuzes met betrekking tot de kwaliteit van een onderzoek kunnen wegen?

Het antwoord op die vraag hebben we niet, maar dat leerlingen in de bovenbouw een hypothetisch-deductieve redenering aankunnen is wel aangetoond (Lawson, 2003). Daarom zijn wij van mening dat onze ambitie niet té groot is als we in het voorbereidend wetenschappelijk onderwijs (vwo) leerlingen willen uitdagen begrip te ontwikkelen voor kwaliteit van onderzoek. Maar inderdaad – zoals in de klankbordgroep met algemene instemming werd verwoord – leren onderzoeken vergt meer dan één goed opgezette praktische opdracht en binnen de kaders in het huidige voortgezet onderwijs zit niet heel veel ruimte. Er zal samenhang en samenwerking gecreëerd moeten worden tussen vakken, tussen jaren en ook tussen mensen. Leerlingen laten leren om kwaliteit van onderzoek te bewaken vereist serieuze, langdurige inspanning. Dus als een docent, een sectie, een school, een schoolbestuur, een examenprogrammacommissie of een minister serieus werk willen maken van het leren bewaken van de

kwaliteit van onderzoek in de hierboven beschreven zin, zal de eerste vraag aan hen moeten zijn: 'Zeker weten?'

Om te laten zien dat er wel degelijk iets bereikt kan worden, eindigt dit proefschrift met het (vrijwel gehele) artikel van twee leerlingen uit de tweede ronde. Dat het niet volledig conform de kwaliteitscriteria is, wil nog niet zeggen dat we te hoog mikken.

Epiloog

Het korte-termijn geheugen

'Omdat de pijl zakt bij het vliegen, moeten we hoog genoeg mikken.'
Frederik Daman (geciteerd in Vervaecke, 2002, p. 153)

Epiloog Het korte-termijn geheugen

Hieronder volgt vrijwel de gehele tekst van een artikel van twee leerlingen – Rita en Meinte – uit de tweede ronde.

Inleiding
(…)
In het verleden werd geheugen vaak beschouwd als een enkelvoudige geestelijke 'gave'. Tegenwoordig wordt het menselijke geheugen echter gezien als een hele reeks processen, inclusief de processen die zorgen voor de kortdurende opslag van zintuiglijke informatie tijdens de waarneming, de processen die ten grondslag liggen aan het vasthouden van kennis of taal, en de processen die het weer verzamelen van persoonlijke ervaringen en gebeurtenissen ondersteunen.

Om een moeilijk verhaal makkelijker te maken kan men bij het woord 'geheugen' denken aan de mate waarin de informatie die we via onze zintuigen ontvangen, gecodeerd wordt binnen het zenuwstelsel. In dit onderzoek hebben wij onderzocht of de leeftijd van invloed is op de mate waarin personen net gehoorde informatie ontvangen en coderen. We hebben dus gekeken naar het functioneren van het korte-termijn geheugen. Dit is een deel van het geheugen met een beperkte capaciteit, welke primair verbaal is en waarvan de inhoud snel vergeten wordt. Korte-termijn geheugen is het systeem dat gebruikt wordt om zich informatie te herinneren die 'in gebruik' is, zoals een telefoonnummer, terwijl men dit intoetst.

Als men iets vergeten is, gebruikt men vaak het excuus "och ja, ik word oud." Maar is dit wel correct? Mogen oudere mensen zeggen dat ze minder onthouden, dus dat het geheugen minder goed werkt, omdat ze al langer leven dan jongere mensen?
Dit onderzoek gaat in het speciaal over het korte-termijn geheugen. Om antwoord op deze vraag te krijgen bedachten wij een onderzoek naar het functioneren van het korte-termijn geheugen, met als duidelijke onderzoeksvraag:
Is er samenhang tussen de leeftijd en het functioneren van het korte-termijn geheugen?
Aan de hand van deze onderzoeksvraag hebben wij ook een hypothese kunnen stellen. Wij denken dat jonge mensen een beter korte-termijn geheugen hebben dan oudere mensen. Dit omdat oudere mensen langer leven en van wie het geheugen dus ook al langer 'aan het werk' is. Wij denken dat het geheugen op een één of andere manier 'versleten' raakt.

Proefopzet
De primaire capaciteit van het korte-termijn geheugen is verbaal en daarom hebben wij bij de proefpersonen een bandje met ingesproken woorden laten afspelen. Tussen de ingesproken woorden op het bandje zit een tussenpauze van enkele seconden. De woorden bestaan niet in de Nederlandse taal, dit omdat men anders linken gaat leggen met het lange-termijn geheugen. Bijvoorbeeld het woord 'taart', dat onthouden zou kunnen worden met het woord 'verjaardag', dat dan uit ervaringen uit het lange-termijn geheugen gehaald wordt. Zo is het onderzoek puur gericht op het functioneren van het korte-termijn geheugen.
Ook hebben de proefpersonen tijdens het luisteren naar de woorden een blad voor de neus met de woorden erop. De woorden zijn dus gezien en gehoord en aan de hand hiervan kunnen we-

ten wij in welke leeftijdsgroep er de meeste woorden onthouden zijn; en dus bij welke leeftijd het korte-termijn geheugen het best functioneert.

Het twee dagen durende onderzoek werd uitgevoerd op dertig proefpersonen. We hebben de personen naar leeftijd ingedeeld.
Groep 1 > 10-20 jaar
Groep 2 > 40-50 jaar
Groep 3 > 70 jaar en ouder.
Met deze groepsindeling hebben we drie verschillende generaties onderzocht op hoeveel onzinwoorden er onthouden worden. Door deze groepsindeling hebben we niet te maken met grensgevallen die het onderzoek lastiger zouden maken.
De onzinwoorden waren ingesproken op een cassetterecorder en getypt op een wit A4-tje.
We hebben elk proefpersoon thuis bezocht en zijn daar aan een tafel gaan zitten, met de proefpersoon tegenover de twee onderzoekers.
Direct na het beluisteren en tegelijkertijd lezen van de woorden, vroegen wij de personen om zoveel mogelijk woorden op te noemen.

Voorspelling
Als de hypothese klopt, dan zullen er in groep 1 de meeste woorden onthouden worden. Wij schatten dat er per persoon in de eerste groep ongeveer 8 van de 10 woorden onthouden zullen worden. Dit aantal zal dan teruglopen naarmate de leeftijd hoger ligt. In groep 2 schatten wij dat er gemiddeld 6 woorden onthouden zullen worden. Tenslotte zal er bij de derde generatie het minst onthouden worden.

Resultaten
(…)
Het verschil tussen de eerste groep en de tweede groep is niet erg groot, de eerste groep heeft in totaal achtenvijftig woorden van de honderd woorden onthouden en de tweede groep heeft in totaal vijftig van de honderd woorden te onthouden. Maar de derde groep, van zeventig jaar en ouder, heeft maar in totaal achtentwintig woorden van de honderd onthouden.
Hieruit kunnen we afleiden dat mensen tot ongeveer vijftig jaar nog een goed korte-termijn geheugen hebben, hoewel het wel al wat minder goed werkt dan dat van de eerste generatie. Vanaf ongeveer zeventig jaar is er duidelijk sprake van een vermindering van de capaciteit van het korte-termijn geheugen.

Conclusie
Antwoord op onze onderzoeksvraag:
Is er samenhang tussen de leeftijd en het functioneren van het korte-termijn geheugen?
Wij hebben uit ons onderzoek afgeleid dat er wel degelijk een samenhang is tussen de leeftijd en het functioneren van het korte-termijn geheugen. Tot een leeftijd van vijftig jaar functioneert het korte-termijn geheugen nog goed, maar na ongeveer vijftig jaar begint het korte-termijn geheugen steeds slechter te functioneren. En vanaf zeventig jaar en ouder functioneert het korte-termijn geheugen nogal slecht. En aan deze cijfers te zien wordt dan nog maar ongeveer éénderde van de woorden onthouden.

Onze hypothese klopt, want hoe jonger onze proefpersonen zijn, hoe beter ze de woorden konden onthouden en hoe ouder, hoe slechter de woorden onthouden werden.
Mensen kunnen dus wel als excuses gebruiken “och ja, ik word oud”, als ze dingen vergeten, maar bij de gemiddelde mens gebeurd dat pas rond je zeventigste levensjaar. Dus een erg goed excuus is het niet, want de mensen die dat excuus gebruiken zijn over het algemeen veel jonger dan deze derde generatie.

Discussie

Ons onderzoek is erg goed gegaan. De meetfouten die we in het onderzoek zouden kunnen hebben, zijn dan veroorzaakt doordat we, relatief gezien, bij weinig proefpersonen de test hadden afgenomen. De uitslag van onze proef zou nog betrouwbaarder zijn als we in plaats van tien personen per generatie, bijvoorbeeld vijftig of honderd personen per generatie genomen hadden. Alleen was het wel wat te moeilijk om aan zoveel mensen te komen.

Dat was het enige wat nog verbeterd zou kunnen worden. Wij denken dat onze uitslag toch wel redelijk betrouwbaar is, omdat we de test bij een toch wel redelijk aantal mensen afgenomen hebben.

In dit onderzoek hebben we slechts één factor onderzocht die het geheugen zou kunnen aantasten, de leeftijd. Maar in praktijk zijn er meerdere factoren waardoor het geheugen minder goed zou kunnen functioneren. Hierbij denken wij aan drugs-, alcohol- en medicijngebruik gedurende het leven, maar bijvoorbeeld ook verschillen in de capaciteit van het geheugen die een rol spelen. Hoe komt het dat de één meer onthoudt dan de ander? Hiervoor is geen leeftijdsverschil noodzakelijk.

Hiermee willen wij duidelijk maken dat andere factoren ook een rol spelen bij het functioneren van het korte-termijn geheugen, maar dat wij ons beperkt hebben tot de factor leeftijd.

Literatuur

Literatuur

Aalsvoort, J.G.M. van (2000). *Chemistry in Products*. Proefschrift Universiteit Utrecht.

Aalsvoort, J.G.M. van (2004). Activity theory as a tool tot address the problem of chemistry's lack of relevance in secondary school chemical education. *International Journal of Science Education* 26 (13), 1635-1651.

Ackermann, R.J. (1989). The new experimentalism. *British Journal for the Philosophy of Science* 40, 185-190.

Adey, P.S., & Shayer, M. (1990). Accelerating the development of formal thinking in middle and high school pupils. *Journal of Research in Science Teaching* 27 (3), 267-285.

Adey, P.S., Shayer, M., & Yates, C. (1989). *Thinking science: The materials of the CASE Project*. London: Nelson.

Akker, J. van den (1999). Principles and methods of development research. In J. van den Akker, R.M. Branch, K. Gustafson, N. Nieveen, & T. Plomp (Eds.), *Design Approaches and Tools in Education and Training* (1-14). Dordrecht: Kluwer Academic Publishers.

Alcock, J. (1998). *Animal behaviour: an evolutionary approach*. Sunderland. Mass.: Sinauer Associates.

Alderwegen-de Vries, A. P. van, Kapteijn, J.M., & Thijssen, G. (1988). *Practicum biologie*. Groningen: Wolters-Noordhoff.

Allen, G.E., & Baker, J.J.W. (2001). *Biology. Scientific process and social issue.* Bethesda, Md., [etc.]: Fitzgerald Science Press.

Almekinders, R., Thijs, G., & Lubben, F. (1998). Development of procedural understanding among South African science students at pre-tertiary level. *Journal of Biological Education* 33 (1), 33-38.

American Association for the Advancement of Science (1967). *Science: a process approach*. New York: Oxford University Press.

American Association for the Advancement of Science (1990). *Science for all Americans*. New York: Oxford University Press.

Appelman, R. (2003). *Begrip van bewijs in biologiemethoden*, Stageverslag. Amsterdam: IDO/VU.

Arievitch, I.M., & Haenen, J.P.P. (2005). Connecting sociocultural theory and educational practice: Galperin's approach. *Educational Psychologist* 40 (3), 155-165.

Bailin, S. (2002). Critical Thinking and Science Education. *Science & Education* 11, 361-375.

Barnard, C.J., Gilbert, F.S., & McGregor, P.K. (1993). *Asking Questions in Biology*. Essex: Longman.

Bauer, H.H. (1992). *Scientific literacy and the myth of the scientific method*. Urbana / Chicago: University of Illinois Press.

Bax, W., Bos, K., & Mourik, F. van (1997). Ontwikkeling van een gemeenschappelijke vaardigheden-lijn binnen een toekomstig profiel. *NVOX* 22 (9), 458-460.

Berg, I. van den (2003). *Peer assessment in universitair onderwijs: een onderzoek naar bruikbare ontwerpen*. Proefschrift Universiteit Utrecht.

Boer, E. de, Jong, J. de, & Warmerdam, P. van (1996). *Handboek vaardigheden*. Loenen aan de Vecht: Edumedia.

Boersma, K.Th. (1999). The Janus face of developmental research. Review of some designs of biology-didactical research. In O. de Jong, K. Kortland, A.J. Waarlo, & J. Buddingh' (Eds.), *Bridging the gap between theory and practice: What research says to the science teacher. Proceedings of the 1998 International Summer Symposium* (31-47). Hongkong: International Council of Associations for Science Education.

Boersma, K.Th. (2001). Het biologieprogramma in de 21e eeuw. *NVOX* 26 (6), 312-317.

Boersma, K.Th., Graft, M. van, Harteveld, A., Hullu, E. de, Mazereeuw, M., Oever, L. van den, & Zande, P.A.M. van der (2006). *Leerlijn biologie van 4 tot 18 jaar vanuit de concept-contextbenadering*. Commissie Vernieuwing Biologie Onderwijs.

Boersma, K., & Kamp, M. (2001). Competenties als doelstellingen voor het biologieonderwijs. *Niche* 32 (4), 28-31.

Boersma, K.Th., Knippels, M.C.P.J, & Waarlo, A.J. (2005). Developmental Research: the improvement of learning and teaching of science topics. In J. Bennett, J.Holman, R. Millar, & D. Waddington (Eds*.), Making a difference. Evaluation as a tool for improving science education* (5-98). Münster / New York: Waxmann.

Brekelmans, J.M.G., Wubbels, Th., & Créton, H.A. (1989). Een typologie van leerlingenpercepties van leraarsgedrag. *Pedagogische Studieën* 66, 315-326.

Brinkman, F.G., & Rens, E.M.M. van (1999). Student teachers' research skills as experienced in their educational training. *European Journal of Teacher Education*, 22, (1), 115-125.

Brouwers, T. (2006). Rubrics of Rubrieken. Bron: http://digimap.slo.nl/aanpakindeklas/beoordeling/Rubrics_of_Rubrieken.doc/.

Brouwers, T., Halfwerk, E., Spek, W., Visser, T., Legierse, A., & Oosterbaan, M. (2003). *Onderzoeken als basis voor beroep*. Enschede: SLO.

Brown, C.R., Moore, J.L., Silkstone, B.E., & Botton, C. (1996). The construct validity and context dependency of teacher assessment of practical skills in some pre-university level science examinations. *Assessment in Education* 3 (3), 377-391.

Brown, J.S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher* 18 (1), 32-42.

Buffler, A., Allie, S., Lubben, F., & Campbell, B. (2001). The development of first year physics students' ideas about measurement in terms of point and set paradigms. *International Journal of Science Education* 23 (11), 1137-1156.

Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1989). 'An experiment is when you try it and see if it works': a study of grade 7 students' understanding of the construction of scientific knowledge. *International Journal of Science Education* 11 (special issue), 514-529.

Chalmers, A. (1999). *Wat heet wetenschap?* Amsterdam: Boom, 10e druk. Vertaling van A. Chalmers (Ed.), *What is this thing called science?* Queensland: University of Queensland Press, 3rd ed.

Chin, C. (2004). Students' questions: fostering a culture of inquisitiveness in science classrooms. *School Science Review*, 86 (314), 107-112.

Chinn, C.A., & Malothra, B.A. (2002). Epistemological authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education* 86, 175-218.

Cito (2006). *TiaPlus users manual*. Arnhem: Cito.

Claessens, S.J.A., & Griethuysen, G.A. van (2001). *Veldwerk in de tweede fase*. Utrecht / Zutphen: Thieme Meulenhoff.

Collins, A., Brown, J.S., & Newman, S.E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In L.B. Resnick (Ed.), *Knowing, learning and instruction, essays in honor of Robert Glaser* (453-494). Hillsdale: Lawrence Erlbaum, Publishers.

Commissie Vernieuwing Biologie Onderwijs (2006a) *Concept examenprogramma Biologie HAVO*. CVBO.

Commissie Vernieuwing Biologie Onderwijs (2006b). *Concept Examenprogramma Biologie VWO*. CVBO.

Crawford, B.A. (2000). Embracing the essence of inquiry: New roles for science teachers. *Journal of Research in Science Teaching*, 37 (9), 916-937.

Crawford, B.A., Krajcik, J.S., & Marx, R.W. (1999). Elements of a community of learners in a middle school science classroom. *Science Education* 83 (6), 701-723.

Crombie, A.C. (1994). *Styles of scientific thinking in the European tradition: the history of argument and explanation especially in the mathematical and biomedical sciences and arts*. London: Duckworth.

Dam, G. ten, & Volman, M. (2002). Het sociale karakter van kritisch denken: didactische richtlijnen. *Pedagogische Studiën* 79, 167-183.

Dewey, J. (1933). The process and product of reflective activity: Phychological process and logical forms. In J. Boydston (Ed.), *The later works of John Dewey* (17 -186). Carbondale: Southern Illinois University Press.

Dewey, J. (1938). The school and society. In G.E. Allen, & J.J.W. Baker (Eds.), *Biology. Scientific process and social issues*. Bethesda, Md., [etc.]: Fitzgerald Science Press.

Dirks, H. (1988). Handleiding bij het praktisch schoolonderzoek biologie. In H. Griethuysen-Elfers, J.M. Kapteijn, C.M. Prop-van den Berg, J.J. Westra, & P.J.A. van Wijk (Eds.), *Practicum biologie 2* (229-247). Groningen: Wolters-Noordhoff.

Duggan, S., & Gott, R. (1996). Scientific evidence: the new emphasis in the practical science curriculum in England and Wales. *The Curriculum Journal* 7 (1), 17-33.

Edutekst (1999/2000). *Algemene vaardigheden exact, deel 4 havo.vwo* en *deel 5 havo.vwo en 6 vwo*. Den Bosch: Malmberg.

Eijck, M. van (2005). Er is geen methode! Leve de colaproef! *Niche* 36 (2), 5-9.

Eijck, M. van (2006). *Teaching quantitative concepts with ICT in pre-university biology education. The case of datalogging the heart.* Proefschrift Universiteit van Amsterdam.

Ergazaki, M., & Zogza, V. (2005). From a causal question to stating and testing hypotheses: exploring the discursive activity of biology students: exploring the discursive activity of biology students. In K. Boersma, M. Goedhart, O. de Jong, & H. Eijkelhof (Eds.), Research and the Quality of Science Education (407-418). Dordrecht: Springer.

Falk, D.F., Keuchenius, F.D., & Saaltink, H.J. (1974). *De bioloog als leraar*. Groningen: Wolters-Noordhoff.

Ferket, H.G. (2003). *Reflectie op kwaliteit van leerlingenonderzoek*. Stageverslag. Amsterdam: IDO/VU.

Feyerabend, P.K. (1975). *Against method: outline of an anarchistic theory of knowledge*. London: NLB.

Geene, W. (2004). *Goed gezien. Onderzoeksvaardigheden in het onderwijs*. Den Haag: Uitgeverij Pak.

Giddings, G.J., Hofstein, A., & Lunetta, V. (1991). Assessment and evaluation in the science laboratory. In B.E. Woolnough (Ed.), *Practical Science* (167 – 177). Buckingham UK: Open University Press.

Giere, R.N., Sloep, P.B., Reddingius, J., & Voorzanger, B. (1990). *Vat op wetenschap*. Groningen / Heerlen: Wolters-Noordhoff / Open Universiteit.

Goldsworthy, A., Watson, R., & Wood Robinson, V. (2000). *Investigations: Developing understanding in scientific enquiry.* Hatfield / Herts: ASE.

Goodlad, J.I., Klein, J.F., & Tye, K.A. (1979). The domains of curriculum and their study. In J.I. Goodlad (Ed.), *Curriculum inquiry* (43-76). New York: McGraw-Hill Book Company.

Goor, A. van, & Hertog, J. den (2001). Redden we het practicum? *Niche* 31 (3), 15-20.

Goor, A. van, & Hertog, J. den (2002). Veldwerk Meyendel. In I. Mommers, (Ed.), *Verslag van de 16e conferentie voor het biologieonderwijs 'Biologie uitbuiten'.* Utrecht: NIBI.

Gott, R., & Duggan, S. (1995a). *Investigative Work in the Science Curriculum.* Buckingham / Philadelphia: Open University Press.

Gott, R., & Duggan, S. (1995b). The place of investigations in practical work in the UK National Curriculum for Science. *International Journal of Science Education* 17 (2), 137-147.

Gott, R., & Duggan, S. (1996). Practical work: its role in the understanding of evidence in science. *International Journal of Science Education* 18 (7), 791-806.

Gott, R., Duggan, S., & Johnson, P. (1999). What do practising applied scientists do and what are the implications for science education? *Research in Science & Technology Education* 17 (1), 97-107.

Gott, R., Duggan, S., & Roberts, R. (2006). Research into understanding scientific evidence (1-14). Bron: www.dur.ac.uk/richard.gott/Evidence/cofev.htm. Durham: School of Education.

Gott, R., & Duggan, S. (2003). *Understanding and Using Scientific Evidence*. Sage Publications Ltd.

Gott R., Foulds K., & Johnson P. (1997). *Science Investigations Book 1*. Collins Educational.

Gott R., Foulds K., & Jones M. (1998). *Science Investigations Book 2.* Collins Educational.

Gott R., Foulds K., & Roberts R. (1999). *Science Investigations Book 3*. Collins Educational.

Gott, R., & Johnson, P. (1999). Science in schools: time to pause for a thought? *School Science Review* 81 (295), 21-28.

Gott, R., & Roberts, R. (2003*). A written test for procedural understanding: a case study into pupils' performance.* Paper gepresenteerd tijdens de vierde European Science Education Research Association conferentie, Noordwijkerhout: The Netherlands.

Griethuysen-Elfers, H., Kapteijn, J.M., Prop-van den Berg, C.M., Westra, J.J., & Wijk, P.J.A. van (1991). *Practicum biologie 2*. Groningen: Wolters-Noordhoff.

Hannay, M. (red.) (1988). *Van Dale handwoordenboek Engels-Nederlands*. Utrecht: Van Dale Lexicografie.

Hennessy, S. (1993). Situated cognition and cognitive apprenticeship: implications for classroom learning. *Studies in Science Education* 22 (1), 1-41.

Heuvelmans, A.P.J.M., & Sanders, P.F. (1993). Beoordelaarovereenstemming. In T.J.H.M. Eggen, & P.F. Sanders (Eds.), *Psychometrie in de praktijk* (443-470). Arnhem: Cito Instituut voor Toetsontwikkeling.

Hodson, D. (1992). In search of a meaningful relationship: an exploration of some issues relating to integration in science and science education. *International Journal of Science Education* 14 (5), 541-562.

Hodson, D. (1993). Re-thinking old ways: towards a more critical approach to practical work in school science. *Studies in Science Education* 22, 85-142.

Hodson, D. (1996). Laboratory work as scientific method: three decades of confusion and distortion. *Journal of Curriculum Studies*, 115-135.

Hodson, D., & Hodson, J. (1998a). From constructivism to social constructivism: a Vygotskian perspective on teaching and learning science. *School Science Review* 79 (289), 33-41.

Hodson, D., & Hodson, J. (1998b). Science education as enculturation: some implications for practice. *School Science Review* 80 (290), 17-24.

Hofman, H. (2006). *Occurrence and quality of a critical discourse in a student inquiry project.* Stageverslag. Amsterdam: Onderwijscentrum VU.

Hogan, K. (1999). Thinking aloud together: a test of an intervention to foster students' collaborative scientific reasoning. *Journal of Research in Science Teaching* 36, 1085-1109.

Hubers, S.T.T.H. (2003). *Individuele leertheorieën en het leren onderzoeken in de Tweede Fase.* Proefschrift Technische Universiteit Eindhoven.

Hubers, S.T.T.H. (2004). *Onderzoeken en ontwerpen van PO naar HO in een doorlopende leerlijn*. Enschede: SLO.

Hullu, E. de, Wulp, D. van der, & Sloep, P. (1999). Hypothesen in de klas en op het examen. *NVOX* 24 (4), 188-193.

Inspectie van het Onderwijs (2003). Rapport regulier schooltoezicht 03.6458. Bron: www.owinsp.nl.

Inspectie van het Onderwijs (2004). Rapport periodiek kwaliteitsonderzoek 44481. Bron: www.owinsp.nl.

Jones, M., & Gott, R. (1998). Cognitive acceleration through science education: alternative perspectives. *International Journal of Science Education* 20 (7), 755-768.

Keys, C.W. (1998). A study of grade six students generating questions and plans for open-ended investigations. *Research in Science Education* 28, 301-316.

Knecht-van Eekelen, A. de, & Pohlmann-Nepveu, J. (1994). Vernieuwing van het biologie-examen vwo. *NVOX* 19 (9), 424-425.

Knippels, M.C.P.J. (2002). *Coping with the abstract and complex nature of genetics in biology education. The yo-yo learning and teaching strategy*. Utrecht: CDβ -Press.

Koningsveld, H. (2006). *Het verschijnsel wetenschap*. Amsterdam: Boom.

Koninklijke Nederlandse Akademie van Wetenschappen (1997). *Biologie: het leven centraal. Eindrapport van de Verkenningscommissie Biologie.* Amsterdam: KNAW.

Koninklijke Nederlandse Akademie van Wetenschappen (2001). *Biologie: een vitaal belang. Strategische visie op de universitaire biologie in Nederland.* Amsterdam, KNAW.

Koninklijke Nederlandse Akademie van Wetenschappen (2003). *Biologieonderwijs: een vitaal belang. Advies van de Biologische Raad.* Amsterdam: KNAW.

Korthagen, F.A.J. (2001). A reflection on reflection. In F.A.J. Korthagen, J. Kessels, B. Koster, B. Lagerwerf, & Wubbels, T. (Eds.), *Linking practice and theory: the pedagogy of realistic teacher education* (51-68). Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.

Korthagen, F.A.J., Kessels, J., Koster, B., Lagerwerf, B., & Wubbels, T. (2001). *Linking practice and theory: the pedagogy of realistic teacher education.* Mahwah: Lawrence Erlbaum Associates, Inc., Publishers.

Kortland, J. (2001). A *problem posing approach to teaching decision making about the waste issue*. Utrecht: CDβ Press,

Kress, G., Jewitt, C., Ogborn, J., & Tsatsarelis, C. (2001). *Multimodal teaching and learning: the rhetoric's of the science classroom*. London: Continuum.

Kuiper, W.A.J.M. (1993). *Curriculumvernieuwing en lespraktijk*. Proefschrift Universiteit Twente.

Kwa, C. (2005). *De ontdekking van het weten. Een andere geschiedenis van de wetenschap.* Amsterdam: Boom.

Lagerwerf, B., & Korthagen, F. (2006). *Een leraar van klasse: een goede docent worden en blijven.* Soest: Uitgeverij Nelissen.

Lammertsma, D.R., S. Broekhuizen, & Muskens, G.J.D.M (1994). Verminderde voortplanting bij steenmarters *Martes fiona* in Nijmegen. *Lutra* 37, 29-45.

Landis, J.R., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics* 33, 159-174.

Lave, J. (1997). The Culture of Acquisition and the Practice of Understanding. In D. Kirshner, & J. Whitson, J. (Eds.), *Situated Cognition: Social, Semiotic, and Psychological Perspectives* (17-35). Mahwah NJ: Lawrence Erlbaum Associates.

Lawson, A.E. (1995). *Science teaching and the development of thinking.* Belmont, California: Wadsworth Publishing Company.

Lawson, A.E. (2001). Using the learning cycle to teach biology concepts and reasoning patterns. *Journal of Biological Education* 35 (4), 165-169.

Lawson, A.E. (2002). Sound and faulty arguments generated by preservice biology teachers when testing hypotheses involving unobservable entities. *Journal of Research in Science Teaching* 39 (3), 237-252.

Lawson, A.E. (2003). The nature and development of hypothetical-predictive argumentation with implications for science teaching. *International Journal of Science Education* 25 (11), 1387-1408.

Legierse, A. (2006). *Handreiking schoolexamen biologie havo/vwo*. Enschede: SLO.

Lemke, J.L. (1998). Analysing verbal data: principles, methods and problems. In B.J. Fraser, & K.G. Tobin (Eds.), *International Handbook of Science Education* (1175-1189). Dordrecht: Kluwer Academic Publishers.

Leo, E.L., & Galloway, D. (1996). Conceptual links between cognitive acceleration through science education and motivational style: a critique of Adey and Shayer. *International Journal of Science Education* 18 (1), 35-49.

Lijnse, P.L. (1995). "Developmental research" as a way to an empirically based "didactical structure" of science. *Science Education* 79 (2), 189-199.

Lijnse, P., & Klaassen, K. (2004). Didactical structures as an outcome of research on teaching-learning sequences? *International Journal of Science Education* 26 (5), 537-554.

Lock, R. (1990). Open-ended, problem-solving investigations, what do we mean and how can we use them? *School Science Review* 71 (256), 63-72.

Lubben, F., & Millar, R. (1996). Children's ideas about the reliability of experimental data. *International Journal of Science Education* 18 (8), 955-968.

Lui, J., Pysarchik, D.T., & Taylor, W.W. (2002). Peer review in the classroom. *Bioscience* 52 (9), 824-829.

Lunsford, E., & Melear, C. T. (2004). Using scoring rubrics to evaluate inquiry. *Journal of College Science Teaching* 34 (1), 34-39.

Maier, E., & Wijk, P. van (Eds.) (1999/2000/2001). *Nectar vwo bovenbouw biologie 1* en *2 (deel 1* en *2), uitwerkingen* en *docentenhandleiding.* Groningen: Wolters-Noordhoff.

Mallory, B.L., & New, R.S. (1994). Social constructivist theory and principles of inclusion: challenges for early childhood special education. *The Journal of Special Education* 28 (3), 322-337.

Marbach-Ad, G., & Claassen, L.A. (2001). Improving students' questions in inquiry labs. *The American Biology Teacher* 63 (6), 410-419.

Marijnissen, J. (Ed.) (2005). *Bruggen tussen VO en universiteiten*. Nijmegen: Radboud Universtieit / Stichting Weten / VSNU.

Mayo, D. (1996). *Error and the growth of experimental knowledge*. Chicago: Chicago University Press.

Mayr, E. (1997). *This is biology, the science of the living world*. Cambridge (Mass.) / London: Belknap Press of Harvard University Press.

McPherson, G.R. (2001). Teaching & learning the scientific method. *The American Biology Teacher* 63 (4), 242-245.

Millar, R., & Driver, R. (1987). Beyond processes. *Studies in Science Education* 14, 33-62.

Millar, R., Lubben, F., Gott, R., & Duggan, S. (1994). Investigating in the school science laboratory: conceptual and procedural knowledge and their influence. *Research Papers in Education* 9 (2), 207-249.

Millenaar, H., Mooldijk, A., & Boer, G. den (1997). Het HVC. In W. Bustraan, (Ed.), *Kamer in het studiehuis. Onderzoekend leren bij de natuurwetenschappen*. Utrecht: APS.

Molenaar, P., & Schermer, A. (Eds.) (1999). *ViaDELTA Biologie: Werkboek 1 vwo B1*. Leiden: Spruyt, Van Mantgem & De Does bv.

Mols, B. (2002). Hersenen zijn aaibaar geworden. *Natuur & Techniek* 70 (12), 18-21.

Morélis, H. (2001). *Verslag enquête natuurwetenschappelijke vakken 2001 (biologie).* Enschede: SLO.

Mortimer, E.F., & Scott, Ph.H. (2000). Analysing discourse in the science classroom. In R.J. Millar Leach, & J. Osborne (Eds.), *Improving Science Education: The Contribution of Research* (126-142). Buckingham: Open University Press.

Mortimer, E.F., & Scott, Ph.H. (2003). *Meaning Making in Secondary Science Classrooms*. Maidenhead: Open University Press.

Oers, B. van (1998a). From context to contextualizing. *Learning and Instruction* 8 (6), 473-488.

Oers, B. van (1998b). The fallacy of decontextualisation. *Mind, Culture, and Activity* 5 (2), 135-142.

Perkins.D.N., & Salomon, G. (1989). Are cognitive skills context-bound? *Educational Researcher* 18 (1), 16-25.

Piccinini, C., & Martins, I. (2003). *Multimodal communication in the science classroom: an exercise of analysis.* Paper gepresenteerd tijdens de vierde European Science Education Research Association conferentie, Noordwijkerhout, Nederland.

Piers, L. (2000). Broedvogels in De Tip/Kruishammen, *Niche* 31 (5), 5-9.

Pieters, J.M., & Verschaffel, L. (2003). *Beïnvloeden van leerprocessen*. Groningen / Houten: Wolters-Noordhoff.

Pohlmann, J., Dietvoorst, P., Knecht, A. de, & Lieverse, M. (1999). Hypothesen in centraal examen en schoolexamen: goed of fout? *NVOX* 24 (6), 306-311.

Pohlmann-Nepveu, J. (1996). *Examens tussen basisvorming en vernieuwde tweede fase HAVO en VWO. Analyse van de huidige examens en mogelijkheden voor aanpassing*. Arnhem: Cito.

Rens, E.M.M. van (2005). *Effectief scheikundeonderwijs voor 'leren onderzoeken' in de tweede fase van het vwo. Een chemie van willen, weten en kunnen*. Proefschrift Vrije Universiteit Amsterdam.

Rens, E.M.M. van, & Dekkers, P.J.J.M. (2000). Leren onderzoeken – de rol van de docent. *Tijdschrift voor Didactiek der ß-wetenschappen* 17 (1), 76-94.

Rens, E.M.M. van, & Dekkers, P.J.J.M. (2001). Learning about investigations - the teacher's role. In H. Behrendt, H. Dahncke, R. Duit, W. Gräber, M.Komorek, A. Kross, & P. Reiska (Eds.), *Research in Science Education – Past, Present, and Future* (325-330). Dordrecht / Boston / London: Kluwer Academic Publishers.

Resnick, L.B. (1987). Learning in school and out. *Educational Researcher* 16, 13-20.

Roberts, R. (2001). Procedural understanding in biology: the 'thinking behind the doing'. *Journal of Biological Education* 35 (3), 113-117.

Roberts, R., & Gott, R. (1999). Procedural understanding: its place in the biology curriculum. *School Science Review* 81 (294), 19-25.

Roberts, R., & Gott, R. (2003a). Written tests for procedural understanding in science: why? And would they work? *Education in Science*, February 2003, 16-18.

Roberts, R., & Gott, R. (2003b). Assessment of biology investigations. *Journal of Biological Education* 37 (3), 114-121.

Roberts, R., & Gott, R. (2004a). A written test for procedural understanding: a way forward for asessment in the UK science curriculum? *Research in Science & Technology Education* 22 (1), 5-21.

Roberts, R., & Gott, R. (2004b). Assessment of Science 1: alternatives to coursework. *School Science Review* 85 (313), 103-108.

Rose, S. (1998). *Lifelines: biology, freedom, determinism*. Penguin.

Roth, W.M. (1995). *Authentic School Science*. Dordrecht / Boston / London: Kluwer Academic Publishers.

Roth, W.M., & Lee, S. (2004). Science education as/for participation in the community. *Science Education* 88 (2), 263-291.

Roth, W.M., & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching* (30) 2, 127-152.

Sang, D. & Wood-Robinson, V. (Eds.) (2002). *Teaching secondary scientific enquiry*. London: John Murray.

Schalk, H. (2001). Begrip van bewijs, de essentie van leren onderzoeken? *Niche* 32 (1), 28-32.

Schalk, H. (2002a). Vragen en verklaren, begrip van bewijs bij docenten. *Niche* 33 (1), 13-17.

Schalk, H. (2002b). Veeleisende variabelen, begrip van bewijs bij docenten (2). *Niche* 33 (5), 32-36.

Schalk, H. (2002c). Hoezo hypothese? Over het begrijpen van biologie als wetenschap. *Niche* 33 (6), 21-24.

Schalk, H. (2005a). Bewaken van de kwaliteit van leerling-onderzoek. In: I. Mommers (Ed.), *Verslag van de 19e conferentie voor het biologieonderwijs 'Biologie extreem, op zoek naar de grenzen'*. Utrecht: NIBI.

Schalk, H. (2005b). Wetenschappelijke methode? Zeker weten! Een reactie op de colaproef. *Niche* 36 (4), 12-15.

Schalk, H., Kapteijn, M., Gelderen, C. van, Neerincx, T., & Wagenaar, F. (2002). Compewattes? Een zoektocht naar competenties. *Niche* 33 (3), 4.

Schalk, H., Schermer, A., & Bax, W. (2004). De delen van een onderzoek. Stap voor stap van hypothese naar conclusie? *Niche* 35 (5), 12-17.

Schalk, H., & Yuksel, A. (2004). Argumenteren over onderzoek. Begrip van bewijs in argumentaties van leerlingen. *Niche* 35 (3), 11-15.

Schee, J.A. van der, Rens, L. van, & Beishuizen, J.J. (2006). *Onderzoekend leren in een community of learners van leerlingen, docenten en onderzoekers.* Paper divisie Curriculum Onderwijs Research Dagen, Amsterdam, Nederland.

Schee, J.A van der., & Rijborz, J.D. (2003). Coaching students in research skills: a difficult task for teachers. *European Journal of Teacher Education* 26 (2), 229-237.

Schön, D.A. (1983). *The reflective practitioner*. Aldershot, England: Arena, Ashgate Publishing.

Scott, Ph. (1998). Teacher talk and meaning making in science classrooms: a Vygotskian analysis and review. *Studies in Science Education* 32, 45-80.

Simons, P.R.J. (1990). *Transfervermogen*. Nijmegen: Katholieke Universiteit.

Sinkeldam, R. (1998). *Handleiding profielwerkstuk*. Arnhem: Cito.

Sliepen, I. (2002). *Het scoren van begrip van bewijs bij leerlingen*. Stageverslag. Amsterdam: IDO/VU.

Smits, G., & Waas, B. (1999/2000/2001). *Biologie voor jou vwo B1 en B2 (deel 1, 2* en *3), leerlingenboeken, antwoorden, examenvragen* en *docentenhandleiding*. Den Bosch: Malmberg.

Smits, Th. (1997). Kwaliteit van open onderzoek: criteria voor begeleiding en beoordeling. *NVOX* 22 (7), 334-337.

Smits, Th., Lijnse, P.L., & Bergen, Th. (2000). Leerlingonderzoek met kwaliteit. *Tijdschrift voor Didactiek der β-wetenschappen* 17 (1), 14-30.

Smits, Th.J.M. (2003). *Werken aan kwaliteitsverbetering van onderwijs*. Utrecht: CDβ Press.

Steen, W.J. van der (1982). *Algemene methodologie voor biologen*. Utrecht /Antwerpen: Bohn, Scheltema & Holkema.

Steen, W.J. van der (1993). *A Practical Philosophy for the Life Sciences*. Albany: State University of New York Press.

Steen, W.J. van der, & Ho, V.K.Y. (2001). *Methods and morals of the life sciences. a guide for analysing and writing texts.* Westport, Conn., [etc.]: Praeger

Stichting Leerplan Ontwikkeling (1996). Examenprogramma's havo/vwo. Enschede: SLO.

Stokking, K.M., & Schaaf, M.F. van der (2000). *Ontwikkeling en beoordeling van onderzoeksvaardigheden*. ISOR Rapport 00.13. Utrecht: Universiteit Utrecht (Onderwijskunde/ISOR).

Strauch, B. (2003). *The Primal Teen*. New York: Double Day, Random house, Inc.

Sutton, C. (1996). The scientific model as a form of speech. In G. Welford, J. Osborne, & Ph. Scott (Eds.), *Research in Science Education in Europe, current issues and themes* (143-152). London: Falmer Press.

Swaak, J. (1998). *What-if; Discovery Simulations and the assessment of intuitive knowledge*. Proefschrift Universiteit Twente.

Takao, A.Y., & Kelly, G.J. (2003). Assessment of evidence in university students' scientific writing. *Science & Education* 12, 341-363.

Tiberghien, A., Veillard, L., Le Marechal, J.-F., Buty, C., & Millar, R. (2001). An analysis of labwork tasks used in science teaching at upper secondary school and university levels in several European countries. *Science Education* 85 (9), 483-508.

Tinbergen, J.M., & Verhulst, S. (2000). Aanpassing kent zijn grenzen. In J.M. Tinbergen, J.P. Bakker, T. Piersma, & J. van den Broek (Eds.), *De onvrije natuur, verkenningen van natuurlijke grenzen* (107-113). Utrecht: KNNV.

Toulmin, S. (1958). *The uses of argument*. Cambridge: University Press.

Valk, T. van der, Broekman, H., Frederik, I., Abels, M., & Jambroes, A. (2005). Een discourse community van docenten rond onderzoekende houding in de bètavakken. *Tijdschrift voor didactiek der β-wetenschappen* 22 (1&2), 22-50.

Valk, A.E. van der, & Soest, M.F.N. van (2004). *Onderzoek leren doen in de bètavakken. Elementen van een leerlijn in de onderbouw van twee scholen.* Utrecht: Universiteit Utrecht (CDβ/ Onderwijskunde/ICO-ISOR).

Veenhoven, J. (2004). *Begeleiden en beoordelen van leerlingonderzoek*. Proefschrift Universiteit Utrecht.

Veldhuijzen, N.H., Godebeld, P., & Sanders, P.F. (1993). Klassieke testtheorie en generaliseerbaarheidstheorie. In T.J.H.M. Eggen, & P.F. Sanders (Eds.), *Psychometrie in de praktijk* (33-82). Arnhem: Cito.

Verhelst, N.D. (2000). *Estimating the reliability of a test from a single test administration*. Measurement and Research Department Report 98-1. Arnhem: Cito.

Verhoeff, R.P. (2003). *Toward systems thinking in cell biology education*. Utrecht: CDβ-Press.

Verloop, N., & Lowyck, J. (2003). *Onderwijskunde*. Groningen: Wolters-Noordhoff.

Vermunt, J.D.H.M. (1992). *Leerstijlen en sturen van leerprocessen in het hoger onderwijs naar procesgerichte instructie in zelfstandig denken.* Lisse: Zwets & Zeitlinger.

Verschaffel, L. (1995). Beïnvloeden van leerprocessen. In J. Lowyck, & N. Verloop (Eds.), *Onderwijskunde. Een kennisbasis voor professionals* (152-187). Groningen: Wolters-Noordhoff.

Vervaecke, H. (2002). De *bonobo's, schalkse apen met menselijke trekjes*. Leuven: Davidsfonds.

Volman, M.L.L. (2006). *Jongleren tussen traditie en toekomst: de rol van docenten in leergemeenschappen*. Amsterdam: Vrije Universiteit.

Vos de, W., & Genseberger, R. (2000). 'Onderzoek doen' in de natuurwetenschappelijke vakken. *Tijdschrift voor Didactiek der ß-wetenschappen* 17 (1), 4-13.

Vries, B. de (2004). *Opportunities for reflection: E-mail and the web in the primary classroom*. Proefschrift Universiteit Twente.

Vygotsky, L.S. (1978). *Mind in society: The development of higher psychological processes.* Cambridge, MA: Harvard University Press.

VZZ (2006). Hoe schrijf je een artikel voor Lutra? Bron: www.vzz.nl/lutra/lutra-auteur-nl.pdf.

Walsarie Wolff-Cox, N.M., Dertien, B.K., & Krüger, J.H.J. (2000/2001/2002). *Biologie Actief vwo B2 profielvak (deel A en B), informatieboeken, activiteitenboeken* en *antwoordenboeken*. Baarn: Nijgh Versluys.

Walsarie Wolff-Cox, N.M., & Krüger, J.H.J. (1999). *Biologie Actief vwo B1 deelvak, informatieboek, activiteitenboek, antwoordenboek* en *docentenboek.* Baarn: Nijgh Versluys.

Weezenbeek, R. (1999). *Zo doen we dat, leerlingenversie* en *docentenversie*. Baarn: Nijgh Versluys.

Wijgerinck, M. (1998). *Vaardigheden compact Tweede Fase*. Loenen a/d Vecht: Edumedia.

Windschitl, M. (2002). Inquiry Projects in Science Teacher Education: What Can Investigative Experiences Reveal About Teacher Thinking and Eventual Classroom Practice? *Science Teacher Education* 87 (1), 112-143.

Windschitl, M. (2003). Folk Theories of "Inquiry": How Preservice Teachers Reproduce the Discourse and Practices of an Atheoretical Scientific Method. *Journal of Research in Science Education* 41 (5), 481-512.

Windschitl, M., & Buttemer, H. (2000). What should the inquiry experience be for the learner? *The American Biology Teacher* 62 (5), 346-350.

Wood, D., Bruner, J.S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry* 17, 89-100.

Wouters, A.G. (1999). *Explanation without a cause*. Utrecht: Zeno.

Wubbels, Th., Brekelmans, M., Brok, P. den, & Tartwijk, J. van (2006). An interpersonal perspective on Classroom Management in Secondary Classrooms in the Netherlands. In M. Carolyn Evertson, S. Carol, C.S. Weinstein (Eds.), *Handbook of Classroom Management: Research, Practice, and Contemporary Issues* (1161-1192). Mahwah, NJ: Lawrence Erlbaum Associates.

Yip, D.Y. (1999). Assessing the concept of controlled experiments in science teachers. *Journal of Biological Education* 33 (4), 204-208.

Yuksel, A. (2004). *Begrip van bewijs in leerlingargumentaties.* Stageverslag. Amsterdam: Onderwijscentrum VU.

Zeiss, F. (2000). Het belang van functionele verklaringen. *Niche* 31 (1), 20-22.

Zohar, A. (1996). Transfer and retention of reasoning strategies taught in biological contexts. *Research in Science & Technological Education* 14 (2), 205-219.

Zohar, A.(2004). *Higher order thinking in science classrooms: students' learning and teachers' professional development.* Dordrecht: Kluwer Academic Publishers.

Lijst van bijlagen

Lijst van bijlagen

1	Vragenlijst voor docenten	(par. 3.2)
2	Momenten van Expliciete Reflectie: opdrachten en scenario's	(par. 5.2)
3	Interview-opdrachten eerste cyclus	(par. 5.3)
4	Toets, antwoordblad en antwoordmodel eerste cyclus	(par. 5.3)
5	Scenario's impliciete en expliciete strategie tweede cyclus	(par. 7.1)
6	Scenario startles	(par. 7.1)
7	Boekjes impliciete en expliciete strategie (inclusief reflectieopdrachten)	(par. 7.1)
8	Scenario's nabespreking reflectieopdrachten	(par. 7.2)
9	Vragenlijst onderzoekservaringen	(par. 7.3)
10	Toets, antwoordblad en antwoordmodel tweede cyclus	(par. 7.3)
11	Evaluatievragenlijst	(par. 7.3)
12	Herkomst van de citaten	(par. 5.3, 7.4 en 7.5)

Samenvatting

Samenvatting

Dit proefschrift gaat over het leren bewaken van de kwaliteit van onderzoek dat leerlingen doen bij het vak biologie in het vwo. Leerlingen doen vaak met enthousiasme een eigen onderzoek, maar dat enthousiasme is geen garantie voor kwaliteit. Leerlingen beschouwen onderzoek doen vooral als 'iets te weten komen', 'een vermoeden controleren' of 'het vinden van een antwoord op een vraag'. De kwaliteit van het antwoord houdt hen minder bezig. Die valt dan ook vaak tegen, zoals het voorbeeld in de **proloog** laat zien.

Waarom 'zeker weten?', het doen van onderzoek door leerlingen is immers al enkele decennia gebruik op ten minste een deel van de scholen voor vwo in Nederland. In het voorbereidend wetenschappelijk onderwijs worden leerlingen niet opgeleid tot onderzoekers, maar maken ze wel kennis met het doen van onderzoek. Dat gebeurt met het oog op de voorbereiding op een wetenschappelijke opleiding, maar ook omdat de westerse samenleving doorspekt is met resultaten van onderzoek. Om die beide redenen vinden wij het kunnen beoordelen van de kwaliteit van onderzoek belangrijk. De eindtermen in het examenprogramma biologie beschrijven een aantal deelvaardigheden c.q. stappen van een onderzoek en vermelden ook dat gegevens, resultaat en conclusie geëvalueerd moeten kunnen worden. Toch wordt dat laatste aspect vaak onderbelicht in de praktijk van het onderwijs.

Om die reden is in het kader van het onderzoeksprogramma 'Kennisontwikkeling tussen theorie en praktijk' van het Onderwijscentrum VU dit promotieonderzoek gestart met als hoofdvraag:

Hoe kunnen leerlingen in het biologieonderwijs in de tweede fase VO leren de kwaliteit van het door hen uitgevoerde eigen onderzoek beter te bewaken?

Voor het bewaken van de kwaliteit van natuurwetenschappelijk onderzoek is begrip nodig van de empirische bewijsvoering die in alle fasen van een onderzoek een rol speelt. Het PACKS-model van Millar e.a. (1994) laat dat goed zien. Empirische bewijsvoering kan uitgewerkt worden in een aantal *concepts of evidence*, een geheel van procedurele kennis over onderzoek doen.

Het onderzoek in dit proefschrift is opgezet als een ontwikkelingsonderzoek, dat wil zeggen dat het gericht is op het ontwerpen en uitproberen van onderwijs dat *goed genoeg* is om bovengenoemd doel te bereiken. Tegelijkertijd wordt daarmee echter ook een visie ontwikkeld op de optimale didactiek voor het leren bewaken van de kwaliteit van onderzoek. In de exploratieve fase van het onderzoek zijn verkenningen uitgevoerd naar het begrip 'kwaliteit van onderzoek', naar de plaats van onderzoek doen in het biologieonderwijs en naar de manier waarop kwaliteitsbewaking geleerd zou kunnen worden. In twee ontwikkelcycli is vervolgens een onderwijsleerstrategie ontworpen en uitgeprobeerd. De bevindingen in die cycli hebben niet alleen geleid tot bijstelling van het ontwerp (na de eerste cyclus) en aanwijzingen voor verdere bijstelling (na de tweede cyclus), maar ook tot reflectie op de uitgangspunten voor een visie op het leren bewaken van de kwaliteit van onderzoek. Het mondt uit in een terugblik en een vooruitblik: wat lijkt mogelijk en wat is er nodig om dat te bereiken?

Om meer grip te krijgen op het begrip 'kwaliteit van onderzoek', is het zinvol om te kijken wat **kwaliteit van onderzoek** inhoudt, in het bijzonder in de biologie, en hoe je die kunt beoordelen en bewaken.

Wetenschappelijk onderzoek is zeer divers, zelfs als je het beperkt tot natuurwetenschappelijk onderzoek. In de loop der tijd hebben zich verschillende, elkaar niet uitsluitende wetenschappelijke stijlen ontwikkeld, elk met hun, deels eigen, geaccepteerde methoden. Dat

betekent dat er niet één universele methode bestaat, maar het betekent niet dat 'alles kan', zoals ook wel eens beweerd wordt. Maar wat is dan het verschil tussen wetenschap en niet-wetenschap? Ons inziens ligt dat verschil vooral in de nadruk die in de natuurwetenschap gelegd wordt op zowel empirie en logica als op samenhangende verklaringen voor natuurlijke verschijnselen; wetenschap vraagt 'empirical evidence'. Verder is kenmerkend dat wetenschap probeert vooringenomenheid (bias) te identificeren en te voorkomen. Bovendien is wetenschappelijk onderzoek een onderneming van een gemeenschap van onderzoekers. Dat betekent dat wetenschappelijk onderzoek voortbouwt op eerder onderzoek en dat de resultaten ook in die gemeenschap gewogen worden.

Biologisch onderzoek houdt zich bezig met de levende natuur, zoals bouw en functie van organismen, hun ontstaanswijze en hun relaties met hun levende en dode omgeving. Biologisch onderzoek kan op vele manieren ingedeeld worden. Wij kiezen voor een indeling aan de hand van de vraagstelling, omdat we onderzoek primair zien als het zoeken van een antwoord op een vraag. We onderscheiden drie typen van biologisch onderzoek: beschrijvend, hypothesetoetsend en ontwerpend onderzoek. Beschrijvend onderzoek richt zich op het in kaart brengen van de biologische werkelijkheid, zoals de anatomie van een vleermuis, de verschillende soorten vleermuizen die voorkomen in een bos of de volgorde van de basen in hun DNA. Hypothesetoetsend onderzoek onderwerpt een mogelijk verband tussen verschijnselen of een mogelijke verklaring van een verschijnsel aan een kritische test. Er worden (deductief) logische consequenties uit de hypothese afgeleid: als de hypothese waar is en ik dát doe (en aan een aantal randvoorwaarden is voldaan), dan zal ik die en die resultaten uit een experiment moeten krijgen of die en die waarnemingen moeten kunnen doen; dit wordt de hypothetisch-deductieve aanpak genoemd. Ontwerpend onderzoek betreft het ontwerpen van materialen, instrumenten en procedures die de (biologische) wetenschap vooruit helpen, zoals de Skinnerbox of de polymerase kettingreactie. De drie typen biologisch onderzoek hebben niet elk een eigen of vaste methode. Toch valt er wel degelijk iets te zeggen over wanneer een onderzoek gedegen is uitgevoerd en wanneer niet, wanneer je meer vertrouwen mag hebben in de resultaten en wanneer minder. Anders gezegd: er zijn wel degelijk criteria te formuleren om te beoordelen wanneer een onderzoek valide is en de resultaten betrouwbaar zijn. Dergelijke criteria, *concepts of evidence*, zijn door Gott e.a. (2006) uitgebreid omschreven, maar voor gebruik in het Nederlandse biologieonderwijs was er een vertaalslag nodig. De deelvraag W*at is de biologische invulling van 'begrip van bewijs'?* is beantwoord met een lijst van 23 elementen van begrip van bewijs (zie bladzijde 250). Deze elementen zijn niet allemaal op elk onderzoek van toepassing, maar vormen ons inziens samen wel een voldoende compleet gereedschap om de kwaliteit van elk (biologisch) onderzoek mee te beoordelen.

De vraag naar de plaats van **'onderzoek doen' in het biologieonderwijs** wordt aan de orde gesteld om te beoordelen of leerlingen in het vwo eigenlijk wel moeten leren de validiteit en de betrouwbaarheid van onderzoek te bewaken en als dat zo is, of ze daar eigenlijk al niet genoeg gelegenheid voor krijgen.

'Onderzoek doen' bevindt zich in een spanningsveld met drie hoeken: 'biologie leren', kennis opdoen van biologische concepten en theorieën, 'over biologie leren', kennis opdoen van het proces van kennisverwerving en 'biologie doen', het opdoen van ervaring en het leren met je handen. Onderzoekspractica hebben in de praktijk verschillende doelen en kunnen daarom eigenlijk op elke plek tussen deze drie hoeken geplaatst worden, afhankelijk van hun doel en de invulling daarvan. Het examenprogramma omschrijft onderzoek doen als een reeks deelvaardigheden en het beheersen daarvan wordt dus voldoende geacht. Als onderzoek doen echter wordt opgevat als een competentie, een samenhangend geheel van kennis, vaardigheden en houdingen – zoals verwoord in een concept van een nieuw examenprogramma – dan valt daar ook het bewaken van de kwaliteit onder. Docenten die ernaar gevraagd zijn, waren van mening dat de meeste van de 23 elementen van begrip van bewijs in het kader van het

biologieonderwijs in de bovenbouw van het vwo belangrijk zijn, geen enkele noemden ze ónbelangrijk. Het meeste belang kenden ze toe aan een duidelijke onderzoeksvraag (element 1), een daarop aansluitende (3) en toetsbare (4) hypothese, helderheid over variabelen (8, 9) en blanco (10), een conclusie die de onderzoeksvraag beantwoordt (18) en de notie van validiteit en betrouwbaarheid van het onderzoek (21).

De docenten is ook gevraagd hoeveel aandacht ze de elementen geven in hun onderwijs en de beoordeling van de leerlingen. Dat levert in grote lijnen hetzelfde beeld op als de vraag naar het belang, al scoorde de gegeven aandacht systematisch lager dan het toegekende belang. Docenten doen dus wat ze belangrijk vinden of vinden belangrijk wat ze doen. Het positieve beeld dat deze inventarisatie oplevert moet echter van kanttekeningen voorzien worden: de steekproef is getrokken uit leden van een docentenvereniging en de respons was laag. Dus komen de antwoorden waarschijnlijk van docenten die meer dan gemiddeld betrokken zijn op en geïnteresseerd zijn in het onderwerp 'onderzoek doen'.

Om meer zicht te krijgen of leerlingen al niet leren de kwaliteit van onderzoek te bewaken zijn biologiemethoden en leerlingverslagen geanalyseerd. In de drie geanalyseerde biologiemethoden komen de elementen van begrip van bewijs weinig ter sprake en voor het grootste deel impliciet. Relatief veel aandacht wordt besteed aan het opzetten van een onderzoek: de onderzoeksvraag en de hypothese, slechts zelden worden validiteit en betrouwbaarheid aan de orde gesteld. In de geanalyseerde verslagen van leerlingonderzoek werd gemiddeld in ruim 60 procent van de gevallen dat een element van begrip van bewijs van toepassing was, eraan voldaan of beargumenteerd waarom er niet aan was voldaan. Ook hier moeten echter kanttekeningen gemaakt worden. De verslagen zijn eindproducten waar docenten ook hun invloed op hebben gehad en de scholen weerspiegelen niet de situatie van het gehele biologieonderwijs in Nederland. De conclusie luidt dat op scholen waar serieus gewerkt wordt aan onderzoek doen door leerlingen, een redelijke mate van begrip van bewijs gerealiseerd wordt. De vraag blijft echter of dat voldoende is om te kunnen stellen dat leerlingen de kwaliteit van hun onderzoek weten te bewaken.

Een onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek zou ons inziens eigenlijk alle elementen van begrip van bewijs moeten omvatten. Vanwege praktische randvoorwaarden moest echter een keus gemaakt worden. Daarbij werd gekozen voor de elementen waar docenten het meeste belang aan hechten. De selectie omvat ook de kern van de hypothetisch-deductieve aanpak:

- *onderzoek geeft een duidelijk antwoord op een heldere vraag* (element 1 en 18);
- *hypothesetoetsend onderzoek onderwerpt een mogelijke verklaring, een mogelijk antwoord aan een kritische toets* (element 3, 4, 5 en 6);
- *in een onderzoek is duidelijk wat onderzocht wordt en wat niet* (element 8, 9, 10 en 14);
- *onderzoek is valide en betrouwbaar* (element 21).

Leren – het toevoegen van nieuwe begrippen aan je mentale structuur – vindt plaats in een combinatie van denken, doen en reflecteren in interactie met anderen. Voor het **leren bewaken van de kwaliteit van onderzoek** betekent dat zowel het nadenken over eisen aan goed onderzoek, het toepassen daarvan in eigen onderzoek als het reflecteren op de kwaliteit ervan: is aan de eisen voldaan? Volgens de 'spiral formation of mental actions' van Galperin (Arievitch & Haenen, 2005) worden begrippen opgenomen in de conceptuele structuur (het mentale niveau) door het handelen met objecten (het materiële niveau), en het onder woorden brengen van dat handelen (het verbale niveau). Het opbouwen van kennis is echter geen individuele zaak, maar een interpersoonlijke, sociale gebeurtenis. Betekenisverlening vindt plaats in interactie en wordt daarna geïnternaliseerd.

Kritisch denken behoeft criteria en houdt ook in dat de denker herkent wanneer hij die criteria moet toepassen. Dat betekent kennis nemen van de criteria voor goed onderzoek en

inzicht in het gebruik ervan. Die kennis en inzicht worden het best verworven door er actief mee aan de slag te gaan, dus door zelf onderzoek te doen en dan te reflecteren de kwaliteit. Dilemma in de gedidactiseerde praktijk van onderzoek doen in het onderwijs is echter of die reflectie het best kan plaatsvinden in en naar aanleiding van (problemen in) het eigen onderzoek, of als een stap opzij door een ander probleem op te werpen en dat toe te passen op het eigen onderzoek. Beide benaderingen zijn in dit onderzoek uitgeprobeerd. In de *expliciete onderwijsleerstrategie* wordt de reflectie gestimuleerd door leerlingen opdrachten te laten maken, waarvoor ze hun eigen onderzoek onderbreken, en ze met de criteria die uit die opdrachten komen te laten kijken naar hun eigen onderzoek: *reflection-on-action* (Schön, 1983). In de *impliciete onderwijsleerstrategie* worden alleen criteria aan de orde gesteld – hetzij door de docent, hetzij door de leerlingen – als de kwaliteit van het leerlingonderzoek daarom vraagt: *reflection-in-action*.

De eerste cyclus van ontwerpen en uitproberen heeft plaatsgevonden op twee scholen, de *Amsterdamse School* en de *School in Amstelveen* in de 5-vwoklassen (samen 50 leerlingen) van twee redelijk ervaren docenten. In wisselwerking tussen docenten, begeleiders en onderzoeker is een eerste ontwerp van een praktische opdracht tot stand gekomen met een omvang van negen lessen, waarin leerlingen in groepjes van twee of drie aan een eigen onderzoek binnen het vak biologie werken. De docenten hebben daarbij een ondersteunende rol met betrekking tot de uitvoering én een initiërende rol waar het het reflecteren op de kwaliteit betreft. In de expliciete variant van deze praktische opdracht zijn vier Momenten van Expliciete Reflectie (MERs) ingebouwd, waarin door middel van een opdracht de beoogde elementen van begrip van bewijs aan de orde kwamen. In de impliciete variant zijn deze MERs er niet, daar hadden de leerlingen meer tijd om te werken aan het eigen onderzoek.

De uitvoerbaarheid en effectiviteit van de onderwijsleerstrategieën zijn in kaart gebracht aan de hand van observaties door de onderzoeker, opnames van de gesprekken van de docenten met de leerlingen, evaluaties met de docenten, werkbladen bij de MERs, een voor- en natoets, de verslagen van de leerlingen en interviews met enkele van hen.

Een eigen onderzoek doen in een periode van tien weken terwijl ook ander schoolwerk moest gebeuren, bleek een moeilijke taak voor de leerlingen. De meeste van hen lukte het niet om de gegeven planning te halen, hetgeen tot tijdsdruk leidde aan het eind van de beschikbare periode. In de begeleiding in zowel de expliciete als de impliciete onderwijsleerstrategie werd veel tijd besteed aan het bespreken van de uitvoerbaarheid van de onderzoeken van de leerlingen, bijvoorbeeld hoe je aan voldoende proefdieren of proefpersonen kwam. Elementen van begrip van bewijs kwamen daarbij soms ter sprake, maar er werden ook kansen gemist. Uit observatie bleek dat meer elementen aan de orde gesteld hadden kunnen worden of explicieter gemaakt hadden kunnen worden. De momenten van expliciete reflectie verliepen niet geheel zoals gepland en de uitvoerbaarheid ervan viel tegen. Hoewel het wel lukte om de opdrachten en discussies uit te voeren, namen ze als geheel te veel tijd, werd het belang ervan onvoldoende duidelijk en pasten de leerlingen de betrokken elementen van begrip van bewijs slechts mondjesmaat toe op hun eigen onderzoek.

De toets kon helaas niet betrouwbaar genoeg leereffecten vaststellen, maar liet wel zien dat leerlingen in hun antwoorden slechts in beperkte mate elementen van begrip van bewijs gebruikten. Uit de leerlingverslagen kwam een minder positief beeld naar voren dan in de voorstudie. De verslagen waren vaak erg informeel van toon en bewaken van de kwaliteit werd door de leerlingen niet als hoofdzaak gezien. De interviews zijn geanalyseerd met behulp van de argumentatietheorie van Toulmin (1958). Dat leverde een beeld op van het gebruik van elementen van begrip van bewijs in verschillende delen van de argumentaties van de leerlingen, maar het bleek weinig inzicht te verschaffen in het leerproces, de ontwikkeling van begrip van bewijs.

Dat er weinig leereffect waargenomen kon worden zegt zowel iets over de praktische opdracht zelf als over de manier van analyseren. Voor het tweede ontwerp zijn daarom een aantal verbeterpunten geformuleerd: meer aandacht voor het motief om de kwaliteit van onderzoek te bewaken en een betere oriëntatie van de leerlingen daarop, meer structuur en overzicht voor de leerlingen en ze een artikel laten schrijven in plaats van een verslag, herzien en decentraliseren van de MERs, beter omschrijven van het beoogde docentgedrag en het herzien van de meetinstrumenten. Met name moest de ontwikkeling van begrip van bewijs bij de leerlingen beter zichtbaar worden.

Bij de **aanscherping** van de uitgangspunten is de probleemstellende benadering (Lijnse & Klaassen, 2004) gebruikt om de plaats en rol van het motief in het ontwerp te verduidelijken. Het motief wordt daarin geacht de motor te zijn die het leerproces op gang houdt: de centrale vraag 'zeker weten?' leidt tot de behoefte een zo goed mogelijk eigen onderzoek te doen en tot de bereidheid op de kwaliteit ervan te reflecteren; dat leidt weer tot de behoefte meer te weten over criteria en die toe te passen op het eigen onderzoek en resulteert in de behoefte om in meer algemene zin iets te weten over kwaliteit van onderzoek. Om leerlingen het motief te laten ontwikkelen is besloten aan het begin van de praktische opdracht een 'startles' te zetten om de aanwezige kennis over onderzoek doen bij de leerlingen te mobiliseren en de centrale vraag 'zeker weten?' stevig neer te zetten.

Om meer zicht te krijgen op de ontwikkeling van begrip van bewijs bij de leerlingen is een analyse-instrument ontwikkeld, mede op basis van het analytisch raamwerk van Mortimer en Scott (2003). Uitspraken van docent en leerlingen met betrekking tot de elementen van begrip van bewijs worden ondergebracht in de categorieën *problematiseren, beschrijven, expliciteren, generaliseren* en *toepassen*. Beschrijven (op het concrete niveau van het eigen onderzoek), expliciteren (het onder woorden brengen van de achterliggende gedachte) en generaliseren (het algemene criterium verwoorden) komen in zekere mate overeen met de het materiële, het verbale en het mentale niveau uit het model van de spiraalsgewijze vorming van mentale acties van Galperin. Als er in een gesprek zowel op het beschrijvende, het expliciterende als het generaliserende niveau over een element van begrip van bewijs gesproken wordt, wordt dat een *volledig patroon* genoemd. Komen slechts twee van die niveaus voor, dan is er sprake van een *onvolledig patroon*, terwijl bij slechts één van de drie over een *rudimentair patroon* wordt gesproken.

In **de tweede cyclus** is het ontwerp van de praktische opdracht aangepast naar aanleiding van de bevindingen van de eerste cyclus. De transcripten van de startlessen – waar de vraag hoe je goed zou kunnen onderzoeken of blonde meisjes dommer zijn dan andere meisjes centraal stond – laten zien dat de beoogde criteria voor goed onderzoek in ruime mate aan de orde zijn geweest, zowel op het beschrijvende en expliciterende als op het generaliserende niveau. Ook droegen zowel docenten als leerlingen daar aan bij. De startlessen hebben daarmee een brede, gezamenlijke basis opgeleverd in de zin dat leerlingen zich actief met de vraag 'wat is goed onderzoek?' hebben beziggehouden. De vraag 'zeker weten?' hebben de leerlingen wel meegenomen als motief om het eigen onderzoek zo goed mogelijk (= in overeenstemming met de criteria) te gaan uitvoeren, maar minder om zoveel mogelijk te leren over wat goed onderzoek is (de criteria zelf). De vraag 'hoe weet ik of ik het zeker weet?' hield de leerlingen niet bezig. De leerlingen zelf gaven na afloop van de praktische opdracht echter wel aan dat ze vooral hadden geleerd 'wat goed onderzoek doen is', meer dan over het onderwerp van hun onderzoek. In dit opzicht is de praktische opdracht dus zeker effectief, het is overgekomen dat het erom ging om te leren over het doen van goed onderzoek.

Het is de leerlingen allemaal gelukt om een eigen onderzoek op te zetten, al hadden ze daar moeite mee. Ondanks het lange nadenken over een onderwerp werd de grote vrijheid om te kiezen gewaardeerd. Deze vrijheid van de leerlingen leidde wel tot een grote druk op de

docenten om veel uiteenlopende onderzoeken te begeleiden, die ze niet alle qua inhoud of technieken even goed beheersten. Bijna alle groepjes leerlingen hebben uiteindelijk een artikel ingeleverd, dat echter niet altijd aan de gestelde vormeisen voldeed. De docenten hebben daar ook niet veel nadruk op gelegd en spraken zelf ook regelmatig van 'verslag' in plaats van 'artikel'. De leerlingen ontleenden vooral in het begin veel steun aan het boekje dat de opdracht structureerde; toen ze verder waren met hun onderzoek gebruikten ze het bijna niet meer.

De docenten waren zich beter bewust van hun rol als initiator en stimulator van het denken van de leerlingen dan in de eerste ronde en ook waren ze alerter op het aan de orde stellen van de criteria voor goed onderzoek. Directe feedback, zowel mondeling als schriftelijk (in de impliciete strategie) werd door de leerlingen zeer gewaardeerd, ook al was die vaak in vragende vorm. Leerlingen vonden wel dat ze vaak lang moesten wachten tot ze de docent konden spreken. De reflectieopdrachten fungeerden zoals beoogd, in de zin dat ze aanleiding waren om over de criteria te praten. Ze bleken redelijk tot goed uitvoerbaar voor de leerlingen, maar kostten toch nog veel tijd, vooral door het nabespreken. Ondanks de gedecentraliseerde opzet, werden ze toch vaak als lastige onderbreking van het eigen onderzoek ervaren, meer dan als steun erbij.

Bij de ontwikkeling van begrip van bewijs gedurende de praktische opdracht bleek de rol van de docent cruciaal voor het expliciteren en generaliseren van de criteria voor goed onderzoek. Dat expliciteren en generaliseren vond namelijk hoofdzakelijk in docent-leerlinggesprekken plaats. In hun onderlinge gesprekken bleven de leerlingen vrijwel geheel op het concrete niveau van beschrijven en problematiseren. Dientengevolge waren er ook maar zelden volledige patronen en af en toe onvolledige patronen waar te nemen. De gesprekken over de reflectieopdrachten bleken zich beter te lenen voor het expliciteren en generaliseren van criteria – en daarmee voor meer volledige patronen – dan de gesprekken over het eigen onderzoek. Daarbij ontwikkelde begrip van bewijs zich vaker langs de inductieve weg dan langs de deductieve, maar ook vaak via een heen-en-weer weg tussen concreet en algemeen.

Gezien de zware wissel die op de docenten werd getrokken, is het niet verwonderlijk dat ze niet alle kansen benutten voor het expliciteren en generaliseren van de criteria. Desondanks kwamen alle beoogde elementen van begrip van bewijs wel aan de orde, zij het niet in dezelfde mate. Veel nadruk viel op de startfase van het onderzoek: onderzoeksvraag, variabelen en hypothese. Ook was er aandacht voor eerlijk tellen en meten. Het overkoepelende element 21 (over validiteit en betrouwbaarheid) kreeg slechts weinig aandacht. Helaas bleek ook dat de docenten niet alle criteria even goed beheersten. Teveel is ervan uitgegaan dat een eerstegraads docent – die dus een wetenschappelijke opleiding heeft gehad – de criteria wel beheerst of snel oppikt. Dit betekent dat in elke volgende versie van de praktische opdracht de inhoudelijke voorbereiding van de docenten op de elementen van begrip van bewijs beter moet zijn. Het verdient dan ook serieuze overweging om nascholing voor docenten op te zetten waarin zowel de criteria voor goed onderzoek aan de orde komen als de manier om ze te benoemen, te expliciteren en te generaliseren.

De definitieve artikelen bevatten al een meer dan redelijke hoeveelheid begrip van bewijs. In ongeveer tweederde van de voorkomende gevallen werd een criterium juist gehanteerd. De impliciete onderwijsleerstrategie deed het hierin iets beter. Er was voor de groep leerlingen uit 5 vwo als geheel een klein, maar positief verschil waar te nemen ten opzichte van in de voorstudie beoordeelde leerlingverslagen, grotendeels uit 6 vwo, waarbij het grootste verschil betrekking had op de evaluatie van het eigen onderzoek.

De leerlingen hebben een kleine, maar significante stap vooruit gedaan in het beoordelen van ander onderzoek, zoals in toetsvragen aan hen voorgelegd. Ze hebben dus zeker iets geleerd ten aanzien van het beoordelen van onderzoek. De criteria met betrekking tot het opzetten van een onderzoek (onderzoeksvraag, hypothese en variabelen) beheersten ze beter dan criteria met betrekking tot latere fasen. De leerwinst van de leerlingen uit de expliciete on-

derwijsleerstrategie was iets groter dan die van de leerlingen uit de impliciete groep, maar de leerwinst in Amstelveen was ook groter dan in Amsterdam. Een al te stellige uitspraak dat de expliciete groep het beter deed is daardoor niet verantwoord.

De uitvoering van de praktische opdracht is in beide strategieën in grote lijnen gegaan zoals gepland, maar vrijwel van meet af aan moest er onder tijdsdruk gewerkt worden, zowel door de leerlingen als door de docent. De uitvoering vraagt veel van de docenten en de leerlingen, maar ons inziens niet té veel. De conclusie over de effectiviteit van de praktische opdracht is genuanceerd. De leerlingen hebben geleerd over goed onderzoek doen en hebben dat ook als zodanig ervaren. De opbrengst ten opzichte van de verslagen uit de voorstudie en ten opzichte van de voortoets is positief en dat terwijl de voorstudie grotendeels is gebaseerd op producten uit 6 vwo. De praktische opdracht lijkt daarmee *goed genoeg* om de leerlingen te laten leren over onderzoek doen met kwaliteit. Maar er is ook twijfel, omdat de uitvoering tegen grenzen aan leek te lopen, zowel qua tijdsinvestering als qua eisen aan de docent. De elementen van begrip van bewijs zijn aan de orde geweest, maar niet alle even goed en vaak niet in gegeneraliseerde vorm. De geschetste cyclus van *uitvoeren – reflecteren – ontwikkelen van criteria* blijkt weinig doorlopen te worden. Zowel het motief van de leerlingen als de focus van de gesprekken over de kwaliteit van onderzoek bleven op het concrete niveau van het eigen onderzoek. De vraag of de opdracht dan *goed genoeg* is om de leerlingen te leren de kwaliteit van onderzoek te bewaken kan helaas niet onvoorwaardelijk bevestigend beantwoord worden.

De uitvoerbaarheid en de effectiviteit zouden ons inziens echter wel verbeterd kunnen worden door een aantal aanpassingen, zoals de planning (nog geconcentreerder), de onderwerpskeus (meer inperken), de nadruk op het vereiste eindproduct, maar dan met meer focus op begrip van bewijs. Het zou zowel de docenten als de leerlingen nog duidelijker voor ogen moeten staan dat het primair gaat om het leren bewaken van de kwaliteit van onderzoek. Ook de functie van het boekje zou beter uit de verf kunnen komen, bijvoorbeeld door bij de verschillende fasen nog meer prikkelende vragen over de kwaliteit op te nemen.

De expliciete en de impliciete onderwijsleerstrategie hebben elk hun sterke kant en een derde versie van de praktische opdracht zou daarom moeten zoeken naar een manier om beide te verenigen. In onze ogen schiet de impliciete strategie tekort als de criteria in de gesprekken over het eigen onderzoek alleen maar op het concrete niveau van het eigen onderzoek aan de orde komen. De aanleiding die het eigen onderzoek geeft om over begrip van bewijs te praten is wel voldoende – in die zin zijn de reflectieopdrachten dus niet nodig – maar de diepgang c.q. de mate van explicitering en veralgemenisering niet. In een volgende versie van de praktische opdracht zou er naar gestreefd moeten worden de reflectie naar een hoger plan te tillen. De docent zou de opdrachten als facultatief in te zetten reflectie achter de hand kunnen houden om die zonodig door leerlingen te laten maken die moeite hebben met bepaalde elementen van begrip van bewijs. Het peer review uit de laatste reflectieopdracht zou behouden moeten worden, en terugkijkend op volledig afgesloten onderzoeken zouden in een 'slotles' de criteria nogmaals expliciet aan de orde kunnen komen, eventueel aan de hand van een nieuw voorbeeldonderzoek.

Een **terugblik en vooruitblik** vraagt in eerste instantie om een duidelijk antwoord op de onderzoeksvraag. De kenmerken van een uitvoerbare en effectieve onderwijsleerstrategie voor het leren bewaken van de kwaliteit van onderzoek in het biologieonderwijs in de tweede fase VO zijn volgens dit onderzoek de volgende:

- duidelijkheid over wat bewaken van kwaliteit van onderzoek inhoudt, zowel voor docent als voor leerlingen;
- stevige inbedding van het doen van onderzoek in het wetenschappelijke verhaal over onderzoek doen, niet als slagroom op de taart, maar als gist in het beslag;

- gestalte geven aan de wisselwerking tussen onderzoek doen en leren over onderzoek doen via *reflection-in-action* en *reflection-on-action*;
- genoeg 'eigenheid' voor de leerlingen in het doen van onderzoek, er moet een belang zijn om een zo zeker mogelijk antwoord te krijgen;
- voldoende expertise van de docent in het doen en begeleiden van onderzoek én in het bewaken van de kwaliteit ervan.

De aard van het onderwerp van dit onderzoek – de kwaliteit van onderzoek – maakt de vraag 'hoe zeker is dat antwoord eigenlijk?' natuurlijk onontkoombaar. Twijfel zou kunnen bestaan over het gebruikte begrippenkader, de onderzoekspopulatie, het ontwikkelde materiaal en de onderzoeksinstrumenten. We denken met de 23 elementen van begrip van bewijs 'kwaliteit van onderzoek' – met daarin de kern van de in de biologie zo belangrijke hypothetisch-deductieve werkwijze – goed geoperationaliseerd te hebben. Met de begrippen *problematiseren, beschrijven, expliciteren, generaliseren* en *toepassen* hebben we het proces van ontwikkeling van begrip van bewijs in beeld proberen te brengen. De vraag kan gesteld worden of het daarmee gelukt is. Bevordert hardop uitgesproken explicitering en generalisatie het opnemen van kwaliteitscriteria in je mentale structuur? Op die vraag kunnen we helaas geen robuust antwoord geven. Verder onderzoek daarnaar lijkt zeker nuttig. Een andere onderzoekspopulatie of een andere opdracht zouden de resultaten van het onderzoek ongetwijfeld hebben beïnvloed, maar naar onze mening niet wezenlijk. De ontwikkelde onderzoeksinstrumenten leveren geen van alle spijkerharde gegevens op, maar gezamenlijk geven ze ons inziens wel een betrouwbaar genoeg beeld van de uitvoerbaarheid en effectiviteit van de ontwikkelde onderwijsleerstrategieën. De onzekerheden betekenen dat we de vraag 'zeker weten?' niet ronduit bevestigend kunnen beantwoorden. Toch zijn we van mening dat er voldoende grond is om de conclusies te trekken die we trekken, zij het dat we alert moeten blijven op signalen die in een andere richting wijzen.

Maar is met het beschrijven van de kenmerken van een uitvoerbare en effectieve onderwijsleerstrategie een afdoend antwoord gegeven op de onderzoeksvraag? Naar onze mening niet helemaal. Want het in dit proefschrift beschreven onderzoek laat zien dat het niet eenvoudig is om in het voortgezet onderwijs aan deze voorwaarden te voldoen. Misschien moet de conclusie wel luiden dát er in de huidige tweede fase niet in voldoende mate aan is te voldoen. Het leren bewaken van de kwaliteit van onderzoek lijkt in ieder geval meer nodig te hebben dan één eigen onderzoek en meer dan de schoolpraktijk als context. Wat nodig lijkt, is het vinden van meer motief en meer gelegenheden om de spiraal van beschrijven, expliciteren en generaliseren te laten draaien. We zien twee mogelijkheden om dat te stimuleren. Ten eerste een doorgaande lijn, waarbij spiraalsgewijs op concreet, verbaal en mentaal niveau begrip van kwaliteit van onderzoek opgebouwd kan worden. 'Volgende keer beter' wordt dan een drijvende kracht. Ten tweede door het directe belang, de directe betekenis van het onderzoek te vergroten, bijvoorbeeld door een grotere authenticiteit van de (grotere) onderzoekscontext of door beoordeling door anderen. 'Zeker weten?' krijgt dan extra belang en betekenis. Deze 'lijnen naar betekenisvol leren' zouden door verder onderzoek ondersteund moeten worden, waarbij het onderhavige onderzoek een bouwsteen kan zijn.

Geven de bescheiden resultaten van dit onderzoek eigenlijk niet aan dat het bewaken van de kwaliteit van biologie-onderzoek een te hoog doel is voor vwo-leerlingen? Ons inziens niet. Er zal wel samenhang en samenwerking gecreëerd moeten worden tussen vakken, tussen jaren en ook tussen mensen. Leerlingen laten leren om kwaliteit van onderzoek te bewaken vereist serieuze, langdurige inspanning. Dus als een docent, een sectie, een school, een schoolbestuur, een examenprogrammacommissie of een minister serieus werk willen maken van het leren bewaken van de kwaliteit van onderzoek in de hierboven beschreven zin, zal de eerste vraag aan hen moeten zijn: 'Zeker weten?'

Summary

Summary

This dissertation is about ensuring the quality of research[35] that pre-university students carry out in the subject of biology. Students often carry out their own research projects with enthusiasm, but that enthusiasm does not guarantee quality. Students think of their research merely as 'finding something out', 'verifying an assumption' or 'answering a question'. The quality of the answer is of less concern to them. It is often disappointing, as the example in the **prologue** shows.

Why 'being certain?', after all, in the past few decades, it has been common for students to carry out research in at least some of the schools throughout the Netherlands. Although they do become familiar with doing research, pre-university students are not trained to *be* researchers. This is in part due to preparation for further scientific education, but also because in contemporary Western society results of scientific research are dominating. Therefore, ensuring the quality and control of research is important. The examination program for biology requires a number of sub-skills or steps within the research process. Students should be able to evaluate data, results, and conclusions; however, often too little emphasis is put on the last aspect in this educational practice.

As part of the research program, 'Knowledge development between theory and practice', a study has been conducted at the Centre for Educational Training, Assessment and Research of the Vrije Universiteit Amsterdam. Its main research question is:

How can students in upper secondary biology education learn to better ensure the quality of their own research?

To ensure the quality of scientific research one needs to understand the empirical argumentation which plays a role in every phase of research, as shown by the PACKS model (Millar *et al.*, 1994). Empirical argumentation can be described by *concepts of evidence,* more specifically, the comprehensive procedural knowledge about conducting research.

The study is set up as developmental research, which means that its aim is to design and implement a method of education that is *good enough* to reach the goal previously mentioned. Throughout this process, an idea about the best way of teaching and learning to ensure the quality of research is developed. The investigatory phase of this study has inquired into the concept of 'quality of research', dealing with research projects within biology education and ways to learn to ensure quality. In two research cycles a teaching and learning strategy has been designed and implemented. The results of the first research cycle led to the revision of the designs of the teaching and learning strategies. The second research cycle led to clues for further changes. In turn, this led the researchers to reflect and reconsider their original objectives and vision with respect to learning to ensure the quality of research. Finally, looking back and looking ahead, we consider the possibilities and what is needed to achieve these goals.

To clarify the concept of 'quality of research', especially in the field of biology, it is necessary to consider what **quality of research** means, and how one can judge and ensure it.

Scientific research is very diverse, even within the natural sciences. Throughout the course of time, several overlapping scientific styles have developed, each with its – partially –

[35] In biology and science education, research conducted by students is known by various names, such as investigations, enquiry, and research projects. They do not mean exactly the same, but partly, the variety is due to the variety of educational contexts. We chose for 'research' as an overarching term.

own, criteria and acknowledged methods. This means that no universal method exists, but does not mean that 'everything goes', which has sometimes been claimed. What, then, is the difference between science and non-science? The difference is where science places its primary emphasis. Science emphasizes empirical data and logic as well as coherent explanations of natural phenomena. Furthermore, it is typical that science attempts to identify and prevent bias, and demands empirical evidence. Moreover, science is a social enterprise, undertaken by a community of researchers. That means that science builds upon earlier research and that the results are weighed in that community.

Biological research focuses on living nature, such as the form and function of organisms, their origin, and their relations within their living and non-living environment. Biological research can be categorized in many ways. We choose a categorization based upon the type of question asked, because in our view, research is primarily the search for an answer to a question. We distinguish three types of biological research: description, hypothesis testing and designing. Descriptive research is focused on mapping the biological world, for example, exploring the anatomy of a bat, the different species of bats that live in the woods, or the sequence of nucleotide bases present in their DNA. Hypothesis testing research puts relations between phenomena or possible explanations of phenomena to the test. Logical consequences are (deductively) inferred from the hypothesis. For example, if the hypothesis which I have projected is true and I do thát (and a number of premises are met), then I should obtain those results from an experiment or should be able to make those observations. This is called the hypothetico-deductive approach. Design research deals with the design of materials, instruments and procedures that drive (biological) science forward, like the Skinner box or the polymerase chain reaction. The three types of biological research do not have their own exclusive or fixed method. However, there is certainly something to say about whether research is carried out soundly or not and when to trust the results.

In other words: certainly, criteria can be formulated and accordingly, one can judge whether research is valid and the results are reliable. Such criteria, *concepts of evidence*, have been formulated extensively by Gott *et al.* (2006), but for use in the Dutch biology education system, a ´translation´ was needed. The following sub-question; *What is the biological content of ´concepts of evidence´?* has been answered with a list of 23 concepts (see page 245). These concepts do not all apply to each and every investigation, but in our view, when observed together, they serve as a sufficiently complete set of tools in order to evaluate the quality of every (biological) investigation.

Questions pertaining to the place of **'doing research' in biological education** are being raised to judge whether students in pre-university education should really learn to ensure the validity and reliability of research. If they should, questions remain pertaining to whether or not they do already have enough opportunity to learn it.

'Doing research' can be placed in a triangle with the cornerstones being 'learning biology', (gaining knowledge about biological concepts and theories), 'learning *about* biology', (gaining knowledge about the process of knowledge acquisition), and 'doing biology', (gaining experience and hands-on learning). In everyday practice investigative practicals can have different goals and therefore can actually be placed within any of these cornerstones, depending on the goal and the content. The examination program describes conducting research as a series of skills and mastering these is expected to be satisfactory. However, if doing research is considered a competency issue, one which includes a comprehensive knowledge, in addition to skills and attitudes – as described in a draft of a new examination program – then it includes ensuring the quality of the conducted research. When asked, teachers consider most of the 23 concepts of evidence as being important in pre-university biology education. Most important was a clear research question (concept nr. 1), a related (3) and testable (4) hypothe-

sis, clarity about variables (8, 9) and control (10), a conclusion that fits the research question (18) and the notion of validity and reliability (21).

Also, teachers were asked about how much attention they devote to the concepts within their instruction and assessment of the students. The answers to this question illustrate a trend roughly similar to the answers concerning the question of importance, although systematically the scores for attention devoted to the concepts within instruction are lower than those for attributed importance. Therefore, teachers do what they consider important or they consider what they do to be important. However, these positive results must not be simply taken at face value, because the sample of teachers consisted of members of an association of teachers and the amount of responses was low, which suggests that the answers probably came from teachers who were, on the average, more actively engaged and interested in 'doing research'.

To find out whether students are already learning to ensure the quality of research, biology textbooks and students' reports have been analyzed. The concepts of evidence are not mentioned often in the three different textbooks (including assignments and teacher guides) that have been analyzed, and when mentioned, they are merely implicitly referenced. Relatively much attention is given to the design of the research: the research question and the hypothesis; validity and reliability were addressed only rarely. In the student reports which were analyzed, about 60 percent of the cases in which a concept of evidence was applicable, it was either met or students argued why this was not the case. It is of importance to note that the reports are end-products which teachers have influenced and the schools do not reflect the situation of all of biology education in the Netherlands. It is concluded that in schools where students work seriously on doing research, 'understanding of evidence' is realized , at least to a reasonable extent. Whether that is enough to claim that students can ensure the quality of their research remains an open question.

In our opinion, a teaching and learning strategy to learn to ensure the quality of research should include all concepts of evidence. However, due to practicality, a choice had to be made. We, therefore, chose the concepts which teachers attributed the most importance to. The selection includes the heart of the hypothetico-deductive approach:

- *research provides an clear answer to a clear question* (concepts 1 and 18);
- *hypothesis testing research puts a possible explanation, a possible answer to a question to the test* (concepts 3, 4, 5 and 6);
- *during research, it is clear what is investigated and what isn't* (concepts 8, 9, 10 and 14);
- *research is valid and reliable* (concept 21)

Learning – adding new concepts to your mental structure – is a combination of thinking, doing and reflecting while interacting with others. Moreover, **learning to ensure the quality of research,** means thinking about requirements for good investigatory research, and applying those requirements to one's own research as well as reflecting on its quality; more specifically, have the requirements been met? According to Galperin's 'spiral formation of mental actions' (Arievitch & Haenen, 2005), concepts are integrated into one's conceptual structure (the mental level) by acting with objects (the material level), and putting them into words (the verbal level). The construction of knowledge, however, does not take place at the level of an individual, but it is an interpersonal, social event. Construction of meaning, therefore, takes place during interaction, after which it becomes internalized.

Criteria are needed for critical thinking and critical thinking implies that the individual should recognize when to apply those criteria. Critical thinking involves taking note of the criteria of good research and insight of their use. This knowledge and understanding are best acquired by actively acting upon it, which is done by actually doing research and then reflecting upon the quality. There is, however, a dilemma in the didactical practice of doing research

in school. The dilemma is whether reflection should take place in and as a consequence of (problems within) one's own research, or by taking a step aside and instead posing an other problem and applying it to one's own research. Both approaches have been used in this research. In the *explicit teaching and learning strategy,* reflection is stimulated by giving assignments to students. During these assignments, students are temporarily diverted from their own research, and evaluate their own research with the criteria that result from the assignment, *reflection-on-action* (Schön, 1983). In the *implicit teaching and learning strategy,* criteria are only raised – either by the teacher or by the students – when it is demanded by the quality of the research: *reflection-in-action.*

The first research cycle of designing and implementing has taken place in two schools, the *Amsterdam School* and the *School in Amstelveen*, in two 11th grade classes (combined consisting of 50 students) of two reasonably experienced teachers. In interaction between teachers, the researcher, the promotores and a panel of teachers, a preliminary design of a practical assignment consisting of nine lessons was made, in which students in small groups of two or three conducted research, according to their own interests within the subject of biology. The teachers' role was to support students when carrying out research and initiating reflection on the quality of the research. In the explicit version of the practical assignment, four Moments of Explicit reflection (MERs) have been built in, in which the intended concepts of evidence surfaced by means of an assignment. In the implicit version there were no such MERs, but instead students had more time to carry out their own research.

The feasibility and effectiveness of the teaching and learning strategies have been evaluated through observations conducted by the researcher, recordings of conversations between teachers and students, teacher evaluations, notes from MERs, pre and post-tests, students' reports and interviews.

Conducting research within a period of ten weeks while simultaneously doing schoolwork, appeared to be a difficult task for the students. Most of them didn't succeed in reaching their initial proposed time frame, which led to a time crunch at the end of the period. When coaching the explicit as well as the implicit teaching and learning strategy, a lot of time was spent on talking about the feasibility of students' research projects. For example, issues pertaining to how to obtain enough laboratory animals or subjects were mentioned. Within these sessions, concepts of evidence came up, yet sometimes opportunities for discussing these concepts were missed. Observations indicated that more concepts could have been discussed or could have been more explicitly raised. The course of the MERs did not fully go as planned and the feasibility was disappointing. Although students and teachers succeeded in carrying out the assignments and discussions, overall, the MERs took too much time, its importance was not clear enough, and students did not apply the concepts in question within their own research enough.

Unfortunately, the test was not reliable enough to measure the learning effects, but it did, however, illustrate that students used concepts of evidence in their answers to a very limited extent. The students' reports showed a less positive image as in the explorative research. The reports were often very informal and ensuring quality was not considered to be a central issue by the students. The interviews have been analysed according to the argumentation theory of Toulmin (1958) which showed students' use of concepts of evidence in different parts of the argumentations, but it turned out that it showed little insight into the learning process, i.e. the development of understanding of evidence.

The fact that the observed learning effect was small says something about the practical assignment itself as well as the way it is analyzed. Therefore, for a number of issues improvement has been formulated: more attention for the motive to ensure the quality of investigations and a better orientation of the students towards it, more structure and overview for the students, letting them write an article in stead of a report, revision and decentralisation of the

MERs, a better description of the intended teacher behaviour and revision of the research instruments. Especially, students' development of the understanding of evidence had to be made more visible.

To **accentuate** the points of departure within the assignment, the problem posing approach (Lijnse & Klaassen, 2004) has been used to clarify the place and role of the motive in the design of the strategies. The motive is supposed to be the driving force that keeps the learning process going: the central question 'are you certain?' facilitates the need for research to be as good as possible in addition to the willingness to reflect on its quality. This in turn, requires further knowledge about the criteria and how to apply them to the students' research.

Finally, it also results in students needing to know more about the quality of research in general. In order to let the students develop a motive, we decided to place an 'introductory lesson' at the beginning of the practical assignment to mobilise students' existing knowledge about conducting research as well as emphasize the central question 'are you certain?'

To gain insight into the development of students' understanding of evidence, an instrument for analysis has been developed, based in part, on an analytical framework of Mortimer and Scott (2003). Utterances of teacher and students with respect to the concepts of evidence are categorised as *problematizing, describing, explaining, generalizing* and *applying.* Describing (at the concrete level of one's own research), explaining (the verbalising of the thoughts behind) and generalizing (verbalising the general criterion) correspond to a certain degree with the material, verbal and mental level from Galperin's model of the spiral formation of mental actions. In a conversation, speaking about a concept of evidence on the descriptive, on the explanatory as well as on the general level, it is called a *complete pattern*. If only two of those three levels occur, then it is called an *incomplete pattern*, whereas a *rudimentary pattern* includes only one of those three.

In **the second research cycle,** the design of the practical assignment has been adapted according to the results from the first cycle. The transcripts of the introductory lessons – in which the central question was, how one could properly research whether blonde girls are dumber than other girls – show that the intended criteria for good research came up amply, on the descriptive, the explanatory and the general level. Teachers as well as students contributed to it. Therefore, the introductory lessons have yielded a broad, common basis in the sense that students were actively engaged with the question 'what is good research?'. In addition, the question 'are you certain?' has been taken up by the students as a motive to carry out their research projects as best (in accordance with the criteria) as possible but to a lesser extent in order to learn as much as possible about what good research is (about the criteria itself). Originally, the students were not engaged in the question 'how do I know if I am certain?' However, afterwards, the students reported that they had learned about 'what good research is', more than just about the subject of their research. So surely, in this respect, the practical assignment is effective, because students understood that the aim was to learn about the particulars concerning doing good research.

The students all succeeded in designing their own research projects, although it was difficult for them. Despite the length of contemplation about a subject, the flexibility and freedom to choose was appreciated. This freedom placed a high amount of pressure on the teachers to coach diverse types of research, because not all of the teachers master to the same level the content or techniques. Nearly all groups of students eventually handed in an article, although it did not always meet the requirements. In part, this was because teachers did not emphasize it much and often spoke about 'reports' instead of 'articles'. In the beginning, students mainly derived support from the booklets which structured the assignment; but when they were further along in their research they hardly used it anymore.

The teachers were better aware of their role as initiators and motivators of the students' ways of thinking than in the first cycle. The teachers were also more alert to bringing up the criteria for good research during the second cycle. Direct feedback, both verbally and written (in the implicit strategy) were much appreciated, even though it was often in an interrogative fashion. Students had the opinion, however, that they had to wait long before they could receive feedback from the teacher. The reflective assignments functioned as intended because they functioned as a vehicle for talking about the criteria.

Their feasibility for the students was reasonably good, but it still took too much time, especially because of the evaluations afterwards. Despite of the decentralized design, students often found the assignments as an annoying interruption of their own research rather than something which aided their research.

In the development of the understanding of evidence, the role of the teacher is crucial for making the criteria for good research both general and explicit. The explaining and generalizing particularly took place in conversations between teacher and students. In their mutual conversations the students mainly rested on the concrete level of describing and problematizing their research. Consequently, only seldom complete and incomplete patterns were recorded. The conversations about the reflective assignments appeared to be better occasions for making the criteria both explicit and general – and consequently for more complete patterns – than the conversations about their own research. Also, understanding of evidence developed more often by way of an inductive route than a deductive one, just as often as going to and fro between the concrete and the general level.

Because the teachers were heavily drawn upon, it is not surprising they did not seize every opportunity to explain and generalise the criteria. Even though, all intended concepts of evidence were brought up, although not to the same extent. Much emphasis was laid upon the starting phase of research: research question, variables and hypothesis. There was also an emphasis on fair testing and measuring. The overarching concept number 21 (about validity and reliability) received little attention. Unfortunately it also appeared that the teachers did not master all criteria equally well. It had been too easily assumed that a – scientifically educated – teacher in upper secondary education knows the criteria or acquires them quickly. That means that in every subsequent version of the practical assignment preparation of the teachers with respect to the content should be better. Serious consideration should be given to set up teacher training in which criteria for good research are addressed as well as ways to describe, explain and generalize them.

The students' final articles contained more than a reasonable amount of understanding of evidence. In about two thirds of the cases a criterion was met correctly. The implicit teaching and learning strategy functioned a bit better in this respect. For the group of the total 11th-graders, a small, but positive difference could be observed compared to the reports analyzed in the explorative study, mainly based on 12th-graders. The greatest difference was related to the evaluation of their own research.

The students made a small, but significant step forward in evaluating other research, as presented to them in test items. They showed more mastery of the criteria with respect to the research design (research question, hypothesis and variables) than to criteria with respect to later phases. The progress of the students in the explicit teaching and learning strategy was greater than that of the students in the implicit group. Yet, the progress in Amstelveen was also greater than in Amsterdam. Therefore, declaring that the explicit group performed better is not completely justified.

In broad outlines, the practical assignment was carried out as planned in both strategies, but had to be done under time pressure by both teachers and students from the start of the assignment clear until the end. The assignment draws heavily upon teachers and students, but in our opinion not too heavily. The conclusion with respect to the effectiveness is nuanced. The students learned about doing good research and experienced it as such. The results

compared to the reports in the explorative study and pre-test are positive, notwithstanding the fact that the explorative study is mainly based upon the work of 12th-graders. Therefore, the practical assignment seems *good enough* to let students learn about doing good research. But there is also doubt, because the implementation came with particular obstacles with regard to time investment as well as demands on the teacher. The concepts of evidence were brought up, but not all of the same quality and often in a non-generalized way. The drafted cycle of *acting – reflecting – developing criteria* did not appear often. The motive of the students as well as the focus of the conversations remained on the concrete level of their own research. Unfortunately, the question whether the assignment is *good enough* to teach the students to ensure the quality of research cannot be affirmed unconditionally.

In our opinion, the feasibility and effectiveness could be improved by means of particular adaptations, including more focused planning, narrowing the choice of subject, and placing emphasis on the required product, yet with a better focus on the understanding of evidence. Teachers as well as students should even be more aware that the focus is primarily about learning to ensure the quality of research. The function of the booklet could be more clear and articulate. For instance, it could include more stimulating questions about the quality of the research throughout the different phases.

The explicit and the implicit teaching and learning strategies each have their strengths and a third version of the practical assignment should try to find a way to combine both. In our view, the implicit strategy falls short if the criteria in the conversations are only raised at the concrete level of the students' research. The extent to which their own research gives rise to discussions about the criteria is sufficient – the reflective assignments aren't needed for that – but the depth i.e. the extent to which they are made explicit and generalized, is not. A next version of the practical assignment should aim at elevating reflection to a higher level.

The teacher could keep the assignments as an optional reflection for students that have difficulty with certain concepts of evidence. The peer review from the last reflective assignment should be retained. Also in a 'concluding lesson' looking back on fully finished research, the criteria could once again be brought up explicitly, possibly by means of a new exemplary investigation.

Looking back and looking ahead in the first place demands an answer to the research question. According to the research in this thesis, the characteristics of a feasible and effective strategy for teaching and learning to ensure the quality of research in pre-university biology education are:

- clarity about what it means to ensure the quality of research, for teachers as well as for students;
- a sound imbedding of conducting research in the scientific story about doing research, not merely as icing on the cake, but as yeast in the dough;
- forming to the interaction between doing research and learning about doing research by way of *reflection-in-action* and *reflection-on-action*;
- there should be an interest to get an answer as certain as possible in order for students to feel a sense of ownership to the research they are engaged in.
- sufficient expertise on the part of the teacher in doing and coaching research as well as ensuring the research quality.

Because of the nature of the subject of this research – the quality of research – the question 'how certain is the above answer really?' is inevitable. There could be doubt with respect to the concepts used, the research population, the developed materials and the research instruments. We believe our 23 concepts of evidence cover the 'quality of research' in the right way, including the core of the hypothetico-deductive approach, which is very important in biology studies. By means of the concepts *problematizing, describing, explaining, general-*

izing and *applying,* we have tried to map the process of developing an understanding of evidence. One could ask whether we have succeeded or not. Does explicating and generalizing aloud really enhance the integration of criteria for quality in the mental structure of students? Unfortunately, to that question we cannot give a robust answer. Further research into that question seems useful. A different research population or a different assignment would have affected the outcomes of the research, but in our opinion, not in a fundamental way. None of the developed research instruments yields decisive results by its own, but together they sketch the feasibility and effectiveness of the teaching and learning strategies in a way that we feel is reliable enough. These uncertainties suggest that we cannot fully affirm the question 'are you certain?' Still, in our opinion, there are enough grounds to draw the conclusions that we do. However, we should be alert and receptive to signals which point in a different direction.

But has the research question been fully and validly answered by describing the characteristics of a feasible and effective teaching and learning strategy? In our opinion it has not yet been fully answered because the research in this thesis shows that it is not simple to meet these characteristics in the present secondary education in the Netherlands. Perhaps the conclusion should be that in the present upper secondary education, the requirements can not be met in a sufficient way. At least, learning to ensure the quality of research seems to require more than one student's research project and more than school as a context. It seems necessary to find more motives and more opportunities to put the spiral of describing, explaining and generalising into motion. We see two possibilities to encourage that. First, a continuing track on which understanding of evidence can be built – in a spiral way – through the concrete, verbal and mental levels. 'Next time better', then becomes a driving force. Second, by augmenting the direct interest and the direct meaning of the investigation, (e.g. by greater authenticity of the context of research or by the reviewing by others). The question, 'Are you certain?' then gains more interest and meaning. These 'lines to meaningful learning' should be supported by further research, to which the present research could then contribute to.

Do the modest results of this research imply that ensuring the quality of biology research is too high a goal for students in pre-university education? In our opinion they don't. Coherence and cooperation will have to be created between school subjects, grade level and people. Letting students learn to ensure the quality of research demands a serious, lengthy effort. So, if a teacher, a department, a school, school governors, an examination board or education ministry seriously want to enhance the learning and teaching with respect to ensuring the quality of research in the sense described above, the first question directed towards them should be: 'Are you certain?'

23 concepts of evidence

Related to the research question

1. A research question:
   - consists only of unambiguous terms and formulations;
   - is sufficiently specific and confined.
2. It is important to distinguish whether the research question demands a description or the testing of a hypothesis.

Related to the hypothesis

3. A hypothesis should be formulated that fits the research question: it is an expected result or a possible explanation.
4. The hypothesis should be testable.

5. It is (nearly) always possible to postulate more than one hypothesis.
6. Based on the hypothesis (and a number of assumptions), a prediction is formulated about which observations or measurements can be expected (if the hypothesis [and the assumptions] is true, then …).

Related to the design of the enquiry

7. In descriptive studies, the features to be observed should be stated explicitly.
8. In hypothesis testing research, the dependent and the independent variables should be identified.
9. In hypothesis testing research, all other variables that may have an influence should be identified and kept constant if possible.
10. In hypothesis testing research, a 'control experiment' should be included.

Related to observations and measurements

11. Observations and measurements should not influence the outcome of those observations or measurements themselves, and any unavoidable influence must be made explicit.
12. The range and the intervals of the chosen values of the independent variable should match the expected variation of the dependent variable.
13. The sensitivity of the instrument should be appropriate to the measurements to be taken.
14. The sample should be representative of the population
    - large enough
    - random or stratified, and the method of sampling should have nothing to do with the subject of research.
    The number of measurements or observations should be large enough.
15. Aberrant or anomalous results should be examined to decide whether they should be used further (can they be qualified as outliers?).
16. If measurements are averaged, it should be in accordance with the content.
17. One should only speak of differences between measurements if the difference is statistically significant.

Related to conclusions and explanation

18. The conclusions should match the research question and (in the case of hypothesis testing) the hypothesis.
19. A correlation is not the same as a causal relationship.
20. It should be stated whether the explanation is causal (related to the causes) or functional (related to the consequences).

Related to evaluation

21. If an enquiry is not carried out validly and reliably (in other words: if the preceding conditions are not met), the results are open to dispute.
22. If possible, the results of the enquiry should be confirmed by data from other research.
23. The conclusion should be compared with accepted ideas (theories), common sense and experience.

Curriculum vitae

Curriculum vitae

Herman Schalk werd geboren op 1 januari 1956 te Dordrecht. In 1978 behaalde hij het kandidaatsexamen biologie (*cum laude*) en in 1984 het doctoraalexamen aan de Vrije Universiteit, met als hoofdvak Celbiologie van Planten en Micro-organismen en als bijvakken Didactiek van de Biologie en Milieuplanning (TU Delft). Tijdens zijn studie behaalde hij de eerstegraads onderwijsbevoegdheid en was hij onder andere actief in de onderzoekscommissie van de Faculteit Biologie en de Initiatiefgroep Wetenschapswinkel van de Vrije Universiteit.

In het schooljaar 1984/1985 was hij als docent biologie verbonden aan de Chr. Avondscholengemeenschap Rotterdam e.o. en de CSG Comenius in Capelle a/d IJssel; in het schooljaar 1985/1986 aan de OSG Het Baken te Almere.

Van 1986 tot 2000 was hij werkzaam bij het Cito Instituut voor Toetsontwikkeling te Arnhem, op de afdelingen 'beginfase voortgezet onderwijs' (later 'basisvorming') en 'centrale examens havo/vwo'. Hij was vanuit het Cito lid van de commissies kerndoelen voor de basisvorming voor de vakken biologie en natuur- en scheikunde.

In de periode 1998 tot 2003 was hij hoofdredacteur van *Niche*, een tijdschrift voor biologiedocenten in het voortgezet onderwijs, uitgegeven door het Nederlands Instituut voor Biologie. Sinds 1998 is hij bij dat instituut ook lid van de onderwijscommissie, vanaf 2004 als voorzitter. Tevens is hij sinds 2000 secretaris van de Vereniging voor Didactiek van de Biologie.

In juni 2000 is hij – in deeltijd – bij het Onderwijscentrum VU (voorheen Instituut voor Didactiek en Onderwijspraktijk) begonnen aan het in dit proefschrift beschreven onderzoek. Daarnaast begeleidt hij bij dat instituut bètadocenten in opleiding bij hun praktijkonderzoek.

Vanaf september 2006 is hij betrokken bij de ontwikkeling van lesmateriaal in het kader van de concept-examenprogramma's biologie voor havo en vwo van de Commissie Vernieuwing Biologieonderwijs. Hij werkt samen met docenten aan het opzetten, ontwikkelen en uitproberen van onderwijs volgens de concept-contextbenadering.

23 elementen van begrip van bewijs

- *Elementen van begrip van bewijs m.b.t. onderzoeksvraag*

1. Een onderzoeksvraag:
   - bevat alleen ondubbelzinnige termen en formuleringen;
   - is voldoende specifiek en afgeperkt.
2. De beantwoording van de onderzoeksvraag vereist een beschrijving en/of de toetsing van een hypothese.

- *Elementen van begrip van bewijs m.b.t. hypothese*

3. Bij hypothesetoetsend onderzoek wordt er een hypothese (een mogelijke verklaring of voorlopig antwoord) opgesteld die bij de onderzoeksvraag aansluit.
4. De hypothese moet toetsbaar zijn.
5. (Bijna) altijd is er meer dan één hypothese mogelijk.
6. Op basis van de hypothese (en een aantal aannames) wordt een voorspelling geformuleerd over welke observatie(s) of meting(en) verwacht kunnen worden (als de hypothese waar is [en …], dan …); bij beschrijvend onderzoek kan een verwachting van de resultaten geformuleerd worden.

- *Elementen van begrip van bewijs m.b.t. onderzoeksopzet en variabelen*

7. Voor beschrijvend onderzoek is het belangrijk de te beschrijven kenmerken (van organismen, ecosystemen e.d.) te benoemen.
8. Het is noodzakelijk de afhankelijke en de onafhankelijke variabele(n) te onderscheiden.
9. Het is noodzakelijk alle andere mogelijk beïnvloedende variabelen te expliciteren en zo mogelijk constant te houden.
10. Om de invloed van een variabele te kunnen beoordelen is een controle of een ‘blanco’ noodzakelijk.

- *Elementen van begrip van bewijs m.b.t. metingen en observaties*

11. De observatie(s) of meting(en) zelf mogen geen invloed hebben op de uitkomst van die observatie(s)/meting(en) of die invloed moet expliciet gemaakt worden.
12. Het bereik en de intervallen van de gekozen waarden van de onafhankelijke variabele moeten passen bij de verwachte variatie van de afhankelijke variabele.
13. De gevoeligheid van het instrument moet passen bij de te verrichten meting(en).
14. De steekproef moet representatief zijn voor de populatie:
    - groot genoeg
    - aselect of gestratificeerd gekozen waarbij de manier van kiezen niets met het object te maken mag hebben.

    Het aantal metingen of observaties moet groot genoeg zijn.
15. Van uit de toon vallende resultaten (‘uitbijters’) moet beoordeeld worden of ze in de verwerking meegenomen moeten worden (zijn het artefacten?).
16. Wanneer uitkomsten van metingen gemiddeld worden, moet dit inhoudelijk verantwoord zijn.
17. Er mag alleen van een verschil tussen metingen gesproken worden als dat verschil statistisch significant is.

- *Elementen van begrip van bewijs m.b.t. conclusie(s) en verklaring*

18. De conclusie(s) moet(en) aansluiten bij de onderzoeksvraag en de hypothese.
19. Een correlatie is nog geen oorzakelijk verband en mag dus niet zonder meer als zodanig geïnterpreteerd worden.

20. Er wordt aangegeven of een verklaring een oorzakelijke verklaring (verklaring vanuit de oorzaken) of een functionele verklaring (verklaring vanuit de gevolgen) is.

- *Elementen van begrip van bewijs m.b.t. evaluatie*

21. Als een onderzoek niet valide en betrouwbaar is uitgevoerd (met andere woorden: als niet aan de voorgaande criteria is voldaan), hebben de resultaten weinig waarde.
22. Indien mogelijk worden de resultaten van een onderzoek vergeleken met gegevens uit ander onderzoek.
23. De conclusie wordt vergeleken met geaccepteerde ideeën (theorieën), gezond verstand en ervaring.