

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 bewijsstuk, *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 beaanbare 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, onderzoeksoptzet & 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 voorgesteld 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 biologietoelichtingsonderwijs, al is het taalgebruik 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)?'

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

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 wintervlinder rups 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

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) mogelijkwijs in elkaar zit of werkt. In een onderbouwing van die hypothese kan aangegeven worden waardoor die hypothese aannemelijk is.⁶

Vaak zijn verschillende hypothesen mogelijk die elkaar al dan niet uitsluiten. Bijvoorbeeld de hypothesen 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 hypothesen 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 stereomicroscop 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.