

# ACHT JAAR ONDERWIJSEVALUATIE: IS ER IETS BEREIKT?

J.A. Smal  
Psycholoog

Correspondentieadres:  
Afdeling Onderzoek en  
Ontwikkeling van het Medisch  
Onderwijs  
Faculteit Geneeskunde  
Bijlhouwerstraat 6  
3511 ZE Utrecht

**V**ijf criteria waaraan een onderwijsevaluatiesysteem getoetst kan worden passeren de revue. Waar kom ik op uit? Een criterium als 'werd het onderwijs beter?' is niet bruikbaar, omdat een causaal verband tussen evaluatie en kwaliteitsverbetering onwaarschijnlijk is. Een beter criterium is de vraag of de opdrachtgever de evaluatiegegevens heeft gebruikt.

## HET EVALUATIESYSTEEM IN UTRECHT

Bij het invoeren van de tweefasenstructuur in 1982 besloot de Onderwijscommissie van de Faculteit Geneeskunde in Utrecht om al het onderwijs voortaan systematisch te evalueren. Hiertoe wordt viermaal per jaar een enquête gehouden onder een steekproef van 40 studenten per studiejaar. De enquête geeft jaarlijks informatie over ongeveer 100 verschillende cursussen, over de colleges van meer dan 200 docenten, 75 practica, 36 tentamens, 6 junior co-assistentenschappen en 10 senior co-assistentenschappen in 19 ziekenhuizen. Gezien dit grote aantal moet de enquête beperkt blijven tot slechts enkele vragen per onderdeel. Om toch geen belangrijke gegevens te missen, wordt de student uitgenodigd om zoveel mogelijk commentaar in eigen woorden te geven. Deze opmerkingen worden ongecensureerd aan de betrokken docenten doorgespeeld. Op grond van het totaal van de verkregen gegevens adviseert een Werkgroep Evaluatie aan de Onderwijscommissie welke maatregelen wenselijk zijn voor de verbetering van het onderwijs.

## EVALUATIE VAN EEN EVALUATIESYSTEEM

Na acht jaar kan de vraag gesteld worden wat deze evaluatie heeft opgeleverd. Het ligt voor de hand om het nut van een evaluatiesysteem af te meten aan de verbeteringen die in het onderwijs zijn opgetreden. Het bezwaar van dit criterium is dat een verbetering of verslechtering van het onderwijs nooit direct aan een evaluatiesysteem kan worden toe-

geschreven. Veel factoren beïnvloeden de kwaliteit van het onderwijs.

Welke andere criteria komen dan in aanmerking om een evaluatiesysteem te toetsen? De vraag naar 'evaluatie van evaluatie' is een recursieve vraag. Recursiviteit is te illustreren aan het pakje met Droste cacao. Op het pakje staat een verpleegster met een pakje cacao, waarop weer een verpleegster met een pakje cacao, en zo ad infinitum. Computerprogrammeurs zijn bekend met het verschijnsel dat bij recursiviteit dezelfde regels herhaald kunnen worden op lagere en hogere niveaus. Dit biedt een welkom houvast bij evaluatie van evaluatie. De criteria gehanteerd bij de evaluatie van het onderwijs kan men opnieuw gebruiken voor de evaluatie van het evaluatiesysteem. Daarom nu eerst iets over criteria voor onderwijsevaluatie.

## CRITERIA VOOR DE EVALUATIE VAN HET ONDERWIJS

In 1975 formuleerde Smith een handzame lijst met criteria voor de evaluatie van onderwijs.<sup>1</sup> Toegepast op het medisch onderwijs ziet deze lijst er als volgt uit.

1. Welk onderwijs wordt er volgens het rooster (aan)geboden?
2. Wordt er aan een aantal minimale voorwaarden van kwaliteit voldaan: is de docent verstaanbaar, zijn de teksten verkrijgbaar, enzovoorts?
3. Door hoeveel studenten wordt het onderwijs daadwerkelijk gevolgd: hoe is het collegebezoek, hoeveel mensen verschijnen op de practica?
4. Hoe zijn de tentamenresultaten en de doorstroomcijfers?
5. Is de gezondheidszorg verbeterd door het gegeven onderwijs?

Gewoonlijk richt de onderwijsevaluatie zich op het tweede criterium: voldoet het onderwijs aan (minimale) kwaliteitseisen? Op dit gebied kan de docent aangesproken worden. De aspecten die als criterium drie, vier en vijf genoemd worden, vallen vaak buiten de invloedssfeer van de onderwijsgeveenden. Zo staan het collegebezoek en de cijfers die de student behaalt vaak los van de kwaliteit van het gegeven onderwijs.

#### **CRITERIA VOOR DE EVALUATIE VAN EEN EVALUATIESYSTEEM**

Bij het beoordelen van een evaluatiesysteem is het moeilijk een causaal verband aan te tonen tussen het evaluatiesysteem en de effecten ervan. Verbetering van het onderwijs behoeft helemaal niet het gevolg te zijn van een evaluatie van het onderwijs. De vraag naar effecten op korte en lange termijn is niet te beantwoorden zonder een speciaal hierop gerichte onderzoekszet, een soort onderwijskundige 'trial'. Dit maakt de laatste twee hierboven genoemde criteria, vertaald naar evaluatie van de evaluatie, moeilijk toepasbaar. De eerste drie criteria daarentegen, aanbod, kwaliteit en consumptie van de evaluatie, bieden een goed houvast.

##### *Criterium 1.*

Aanbod: wordt er geëvalueerd?

Deze vraag is eenvoudig te beantwoorden. Viermaal per jaar wordt er een enquête verstuurd, de resultaten worden onder alle betrokkenen verspreid en de Onderwijscommissie ontvangt een reeks adviezen van de Werkgroep Evaluatie.

##### *Criterium 2.*

Kwaliteit: beantwoordt de evaluatie aan de minimale eisen die aan een evaluatie gesteld mogen worden?

De minimale eisen die aan de kwaliteit van een evaluatie gesteld moeten worden zijn betrouwbaarheid en validiteit. Zonder een gericht onderzoek is men voornamelijk

aangewezen op 'critical incidents', zoals opvallende veranderingen in het onderwijs, die zich weerspiegelen in de enquête-uitslagen, en op min of meer toevallig parallelonderzoek om een indruk te krijgen van de betrouwbaarheid en validiteit van de enquête-uitslagen. De volgende vier verschijnselen lijken in ieder geval nauw samen te hangen met betrouwbaarheid en validiteit: (a) de samenstelling van de non-response-groep, (b) de consistentie van antwoorden, en het optreden van (c) 'spookstudenten en (d) spookdocenten'.

a) Gegevens over de non-response-groep. Zeer zelden worden alle enquêtes ingevuld en geretourneerd. Gewoonlijk antwoordt 60% tot 90% van de studenten. Het gevaar is niet denkbeeldig dat dit een 'selecte' groep is, bijvoorbeeld de groep ijverige studenten. De enquête-uitslagen zouden dan een vertekend beeld geven van het onderwijs. In 1988 kon de uitslag van de evaluatie vergeleken worden met de uitslagen van een andere enquête, waaraan wel alle studenten hadden deelgenomen. In beide enquêtes werd gevraagd naar het collegebezoek bij de vakken Histologie, Methodologie en Oogheelkunde, gegeven in respectievelijk het 1e, 2e en 4e studiejaar. Er bleek geen significant verschil te zijn tussen het opgegeven collegebezoek bij de evaluatie-enquête en bij de andere enquête.

b) Een tweede indicatie voor de betrouwbaarheid en de validiteit van de enquête-uitslagen is de consistentie waarmee bepaalde antwoorden terugkeren. Heel opvallend is de consistentie in het aantal opgegeven uren voorbereidingstijd bij de verschillende tentamens. Hoewel er van jaar tot jaar kleine schommelingen optreden, vallen de verschillen tussen de jaren in het niet bij de verschillen tussen de tentamens. In tabel 1 is te zien hoe voor een vak als Farmacologie gewoonlijk bijna viermaal zo lang wordt gestudeerd als voor Histologie. Omgekeerd zijn veranderingen in het onderwijspakket terug te vinden in vermeerdering of vermindering van de voorbereidingstijd voor tentamens. In de tabel is bijvoorbeeld te zien hoe de voorbereidings-

Tabel 1.  
Gemiddelde voor-  
bereidingstijd tentamens  
(in uren)

		1986	1987	1988	1989
Histologie	(prop)	47	46	53	41
Farmacologie	(doc2)	155	150	159	168
Dermatologie	(doc3)	41	52	33	32
Inw. Geneesk.	(doc3)	104	118	183	96
Neurologie	(doc3)	57	76	78	94

tijd Inwendige Geneeskunde in 1988 veel hoger was dan daarvoor, en het jaar daarna scherp was gedaald. Veranderingen in de omvang van het tentamen zijn hier een aannemelijke verklaring.

c) Spookstudenten. Het verschijnsel 'spookstudent' treedt op als een student NIET heeft deelgenomen aan colleges, maar toch een oordeel geeft over de kwaliteit van de presentatie. Vooral bij enquêtes per post is dit gevaar niet denkbeeldig. Uit alles blijkt echter dat de enquêtes zeer consciëntieus worden ingevuld. Het is in acht jaar slechts enkele keren voorgekomen dat een student invulde dat hij geen colleges had gelopen, maar toch een oordeel gaf over de presentatie van de colleges. Men onthoudt zich nagenoeg altijd

van een mening als men zelf niet naar een college geweest is. Wel komt het voor dat iemand schrijft: "Ik ben zelf niet geweest, maar heb horen vertellen dat het goede colleges waren".

d) Spookdocenten. Van 'spookdocenten' is sprake wanneer studenten een oordeel geven over docenten die in werkelijkheid nooit college hebben gegeven. Dit komt gewoonlijk een- of tweemaal per jaar voor. De fout is meestal te verklaren doordat de naam van de docent niet voldoende bekend was. Eenmaal echter werd een docente door verschillende studenten geprezen, met vermelding van haar naam, terwijl zij naar eigen zeggen nooit college had gegeven .... Dit raadsel is nooit opgelost.

Deze vier analyses ondersteunen de betrouwbaarheid en de validiteit van de enquêtegegevens; het evaluatiesysteem lijkt derhalve te voldoen aan deze minimale kwaliteitseisen.

### Criterium 3.

Consumptie: worden de gegevens uit de evaluatie gebruikt?

Het antwoord hierop is te vinden in de archieven van de Onderwijscommissie. In de periode 1985-1989 werden door de Werkgroep Evaluatie 76 adviezen geformuleerd. Deze adviezen zijn globaal in twee groepen in te delen: (a) verzoeken om aandacht aan een probleem te schenken, (b) suggesties tot verbeteringen of concrete adviezen om actie te ondernemen.

Een voorbeeld hiervan is het advies om alle docenten richtlijnen toe te sturen over het gebruik van overheadsheets. Een ander advies was het toesturen van de studiegids aan alle docenten in plaats van alleen aan de vakgroepen.

In tabel 2 is een overzicht gemaakt van het lot dat deze adviezen beschoren was. In 11 gevallen is in het archief niet terug te vinden of de Onderwijscommissie iets met het advies heeft gedaan ("niets"). In de overige gevallen, 86%, is er zeker aandacht aan besteed; in veel situaties is er contact gezocht met docenten of vakgroepen en is inderdaad actie ondernomen. Hieruit mag worden geconcludeerd dat de evaluatie voldoet aan het derde criterium, namelijk dat de resultaten van de evaluatie gebruikt zijn.

Tabel 2.  
Adviezen van de  
Werkgroep Evaluatie en  
maatregelen van de  
Onderwijscommissie in  
de periode 1985 - 1989

Inhoud advies werkgroep	aandacht schenken	advies tot gerichte actie	totaal
Genomen maatregel door de Onderwijs- commissie:			
niets	2	9	11
besproken	6	5	11
doorverwezen naar andere werkgroep	4	10	14
brief aan docent	2	12	14
contact met docent	7	12	19
invitatie docent voor gesprek met Onderwijs- commissie	1	1	2
de Onderwijscommissie neemt zelf initiatief	0	5	5
<b>Totaal</b>	<b>22</b>	<b>54</b>	<b>76</b>

#### *De overige criteria*

Zoals boven reeds vermeld, is de interpretatie van de effecten op het onderwijs moeilijk. Zaken die gemakkelijk te meten en te veranderen zijn, zoals duidelijkheid in de aankondiging van het programma, lijken in de loop der jaren beter te worden. Andere aspecten tonen geen duidelijke verbetering. Toe- en afname van het collegebezoek lijken meer samen te hangen met studieduur-beperking of tentamenroosters dan met de kwaliteit van de colleges. Voor de beoordeling van de presentatie van de colleges zijn geen absolute normen te geven. Zolang er geen oorzakelijk verband aangetoond kan worden, is het riskant om verbeteringen in het onderwijs toe te schrijven aan het evaluatiesysteem. Effecten op korte en lange termijn zijn dan niet geschikt als evaluatie-criterium.

mate van onverschilligheid: slechts weinigen zullen zich verdiepen in de kwaliteit van alle onderwijsonderdelen. Dat ligt anders bij de opmerkingen die direct betrekking hebben op het eigen onderwijs. Docenten lezen deze zeer kritisch en zoeken meestal naar wegen om gesignaleerde problemen op te lossen. Soms echter spreken de evaluatiegegevens elkaar tegen: de ene helft van de studenten wil het anders dan de andere helft. Dit werkt wel eens ontmoedigend. De studentenvertegenwoordigers in de werkgroepen en facultaire commissies benadrukken telkens weer dat de evaluatie moet doorgaan, want de evaluatie-uitkomsten vormen een legitieme reden om knelpunten aan de orde te stellen.

#### **CONCLUSIE EN SLOTBESCHOUWING**

De waarde van een evaluatiesysteem kan niet zonder meer afgeleid worden uit de verbeteringen die in het onderwijs zijn opgetreden. Het optreden of achterwege blijven van onderwijsverbeteringen is afhankelijk van veel andere factoren, en hoogstens indirect van de evaluatie zelf.

De vraag of de evaluatiegegevens niet in de la verdwijnen, maar daadwerkelijk worden gebruikt voor het beleid van de Onderwijscommissie, is wel een geschikt criterium voor de beoordeling van een evaluatiesysteem. Een analyse van de notulen heeft aangetoond dat de Onderwijscommissie in 86% van de gevallen de adviezen van de Werkgroep Evaluatie heeft gevolgd. In dit opzicht heeft het evaluatiesysteem aan zijn doel beantwoord.

Een enkele opmerking nog over de reacties die een evaluatieproject als hier beschreven oproept in de faculteit. Op den duur is ieder aan de evaluatie gewend. Het is normaal dat het onderwijs beoordeeld wordt. Hier staat tegenover dat de herhaling leidt tot een zekere

#### **LITERATUUR**

1. Smith MC. Evaluation and evaluative research in pharmacy education. Am J Pharm Educ 1975; 39: 428-35.