

Onderwijsevaluaties: fit for use, fit for purpose?

door Ouke J. Pijl

Het lijkt erop dat er meer wordt geëvalueerd dan ooit. En vooropgesteld: dat is een goede zaak! In het afgelopen half jaar ontving ik alleen al via de e-mail zo'n 25 verzoeken om mee te doen met een digitale enquête. Mijn mening werd gevraagd over kleine zaken als de online bestelde oordopjes voor mijn telefoon, maar ook over grotere zaken als een vakantie, een verzekering of zelfs een uitvaart van een dierbare van jaren geleden. De enquêtes worden vaak automatisch verstuurd en bevatten soms flink lange vragenlijsten.

Er lijkt daarbij soms een disbalans tussen de moeite die het de verzender kost om de vragenlijst (automatisch) te versturen en de tijd die het de ontvanger kost om het allemaal in te vullen. Soms wordt zelfs de enquête nog geëvalueerd... Maar er is nog iets dat opvalt: aanzienlijk veel vragen gaan over 'hoe kunnen wij onze organisatie verbeteren' en niet over 'wat heeft u er als klant aan gehad'. In dit artikel uit de onderwijspraktijk zullen we aan de hand van evaluaties in het hoger onderwijs zien dat het veel kan opleveren onderscheid te maken tussen evalueren van het middel (de onderwijsleeromgeving) en van het resultaat (gediplomeerde studenten die als professional het verschil uitmaken in de beroepspraktijk). En te zoeken naar een verband tussen middel en doel...

PDCA in het onderwijs

Ook in het (hoger) onderwijs wordt met regelmaat geëvalueerd. Soms lijken we bijna vergeten te zijn dat in het onderwijsproces al een keurige plan-do-check-act cyclus is geïntegreerd, zie figuur 1 bovenaan. Elk onderwijsproces start met het *plan*: het vaststellen van de leerdoelen, de beoogde kwalificaties. Op basis daarvan wordt een onderwijsprogramma opgesteld dat studenten doorlopen, de *do*. Na afloop van elke onderwijseenheid krijgt iedere student een toets, de *check*. De resultaten van de toets worden vervolgens gebruikt om vast te stellen of de student voldoende heeft geleerd, de *act*. Bij een onvoldoende resultaat gaat de student terug in het onderwijsproces om opnieuw te leren en een nieuwe toets af te leggen. En wanneer het helemaal niet gaat, wordt

het onderwijsproces voor die student afgebroken. In de oneerbiedige termen van een productieproces wordt in het onderwijs dus elk product – elke student – afzonderlijk intensief gekeurd: tussentijds en met een allesbeslissende eindcontrole. Compleet met certificaat.

Ging het goed? Heeft het geholpen?

– In een enquête naar aanleiding van een fysiotherapeutische behandeling werden 35 vragen gesteld. Daarvan gingen 11 vragen over persoonlijke achtergronden en 24 vragen over het behandelingstraject. Vragen over de behandeling zelf voerden daarbij de boventoon: over telefonische bereikbaarheid van de therapeut, privacy in de behandelruimte, of de uitleg duidelijk was, of er goed werd geluisterd, of er belangstelling was voor het al dan niet uitvoeren van de thuisoefeningen, enz. Slechts enkele vragen gingen over het *resultaat* van de behandeling, al was het wat indirect: of er was toegewerkt naar de behandel doelstelling, of de resultaten van de behandeling waren besproken, en of de resultaten overeenkwamen met de verwachtingen. De *ultimate question* 'Vindt u dat de behandeling heeft geholpen?' ontbrak.

Metten is weten?

In het hoger onderwijs vindt bovenop de zo mooi geïntegreerde PDCA-cyclus op procesniveau een groot aantal evaluaties plaats, zie figuur 1 onderaan. Alle 430 hogere opleidingen (wo en hbo)¹ worden voor de vereiste *NVAO-accreditatie* eens in de zes jaar beoordeeld op de 'standaarden' voor beoogde eindkwalificaties, de onderwijsleeromgeving, het toetsstelsel en de gerealiseerde eindkwalificaties. Veel instellingen voeren – bijvoorbeeld via *interne audits* – ook tussentijds een dergelijke beoordeling uit. Jaarlijks worden honderdduizenden studenten in het wo en hbo benaderd met de *Nationale Studenten Enquête* die bestaat uit een vragenlijst van rond de 100 vragen (response ongeveer 40%)². In het hbo worden bovendien jaarlijks een kleine 50.000 alumni via de *HBO Monitor* bevraagd over hun kijk op de opleiding vanuit het perspectief van



Ouke J. Pijl

hun beroepsloopbaan (response ongeveer 30%)³. Naast deze landelijke evaluaties hebben alle opleidingen nauwe banden met (internationale) *branches* en *kennisinstellingen* om de eindkwalificaties en de inhoud van hun curricula te toetsen aan nieuwe ontwikkelingen. Binnen de instellingen voeren *examencommissies* hun wettelijke taak uit de kwaliteit van de examinering en het eindniveau van studenten zeker te stellen, onder meer door de kwaliteit van toetsen te beoordelen⁴. Veel opleidingen laten studenten na afloop van een cursus eigen *evaluatie* invullen waarbij ze ook hun mening kunnen geven over docenten. Ook het studiesucces (*rendement*) in het onderwijs wordt nauwgezet gemeten en in het hbo gepubliceerd op de website van de Vereniging Hogescholen⁵. In het kader van de Prestatieafspraken is ook de onderwijsintensiteit (*contacturen*) in het eerste jaar een belangrijk gegeven. En – 'last but not least' – verschijnen jaarlijks de 'rankings' van de *Keuzegids* (hbo, universiteiten en masters) en *Elsevier Beste studies* (hbo en wo). Al met al een hele uitdaging voor opleidingsmanagers en docententeams om zich tot al deze cijfers te verhouden.

Fit for use, fit for purpose?

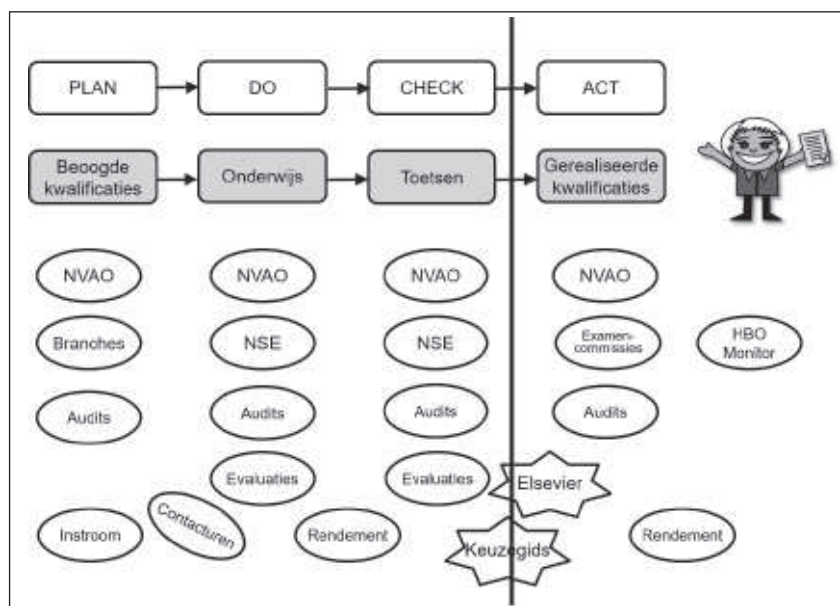
Bij Hogeschool Utrecht is onlangs een eerste stap gezet om orde te scheppen in dit woud aan evaluatiegegevens. Dat werd gedaan in drie stappen:

1. Aan de hand van het onderwijsmodel van Scheerens (2010) is onderscheid gemaakt tussen de kenmerken van de

onderwijsleeromgeving en de kwaliteit van de afgestudeerden. In figuur 1 het deel links, respectievelijk rechts van de verticale lijn. Vanuit het perspectief van de student is het onderscheid: 'Hoe gaat het op school?' versus 'Wat kan ik met mijn diploma in de beroepspraktijk?'. Oftewel 'fit for use' versus 'fit for purpose'.

- Alle opleidingen zijn geplot in een grafiek met een verticale as 'kwaliteit' en een horizontale as 'rendement'. Bij 'kwaliteit' ging het daarbij uitsluitend om de kwaliteit van afgestudeerden, geoperationaliseerd in een construct waarbij de beoordeling (voldoende, goed, excellent) op het NVAO-criterium 'gerealiseerd niveau' (standaard 4) werd samengenomen met alumnitevredenheid (HBO Monitor). Voor 'rendement' is het afstudeerrendement gehanteerd. Beide assen snijden elkaar rond het gemiddelde. Het resultaat is te zien in figuur 2. We zien hier vier kwadranten waarbij een positie in het kwadrant 'KR' rechtsboven (hogere kwaliteit en hoger rendement) het meest wenselijk is. Een positie in het kwadrant 'kr' linksonder heeft de minste voorkeur.
- Vervolgens is gekeken of er verschillen waren te zien tussen de kenmerken van de onderwijsleeromgeving in de verschillende kwadranten. En die waren er. In het gunstigere kwadrant KR bleek de onderwijsintensiteit (contacturen) iets lager dan in de andere kwadranten. De studenttevredenheid (Nationale Studenten Enquête) bleek in elk kwadrant ongeveer gelijk. Maar de specifieke tevredenheid van studenten over de docenten bleek in het KR kwadrant aanmerkelijk hoger.

De opbrengst van deze analyse maakt allereerst duidelijk wat de positie van opleidingen is op *resultaat*, de kenmerken die er uiteindelijk echt toe doen: de kwaliteit van afgestudeerden en het rendement. Dat zijn de twee kenmerken die in de (landelijke) discussies over onderwijskwaliteit ook steeds naar voren komen. Bovendien lijkt deze analyse aan te geven dat algemene studenttevredenheid over de onderwijs-



Figuur 1: PDCA in het onderwijs: van Beoogde eindkwalificaties (plan) via Onderwijs (do) en Toetsen (check) naar Gerealiseerde eindkwalificaties (act). Daaronder in de ovalen een overzicht van gehanteerde evaluaties en kengetallen in het hbo. Links van de verticale lijn gaat het voornamelijk om de evaluatie van de onderwijsleeromgeving (wat op school gebeurt, 'fit for use') en rechts van de verticale lijn om de uitkomsten daarvan (kwaliteit afgestudeerde studenten en hun aantal, 'fit for purpose').

leeromgeving blijkbaar niet van invloed is op het eindresultaat maar dat docentkwaliteit (zoals beoordeeld door studenten) er blijkbaar wel toe doet.

In de komende periode zal een tweede analyse worden uitgevoerd met nieuwe data om na te gaan of de conclusies grotendeels gelijk blijven. Er zijn bovendien nog meer analyses nodig om het gebruikte construct voor 'kwaliteit' te toetsen, na te gaan of er geen andere tussenliggende variabelen zijn die de uitkomsten kunnen verklaren (bijvoorbeeld het aantal studenten per opleiding), wat het effect is van het verleggen van het nulpunt in het assenkruis, enzovoort. Ook daarom zal de analyse vooral worden gebruikt om meer betekenis te geven aan de dialoog over onderwijskwaliteit in plaats van een nieuwe 'ranking' te introduceren. De cijfers geven geen verklaring; in een dialoog kan er betekenis aan worden gegeven.

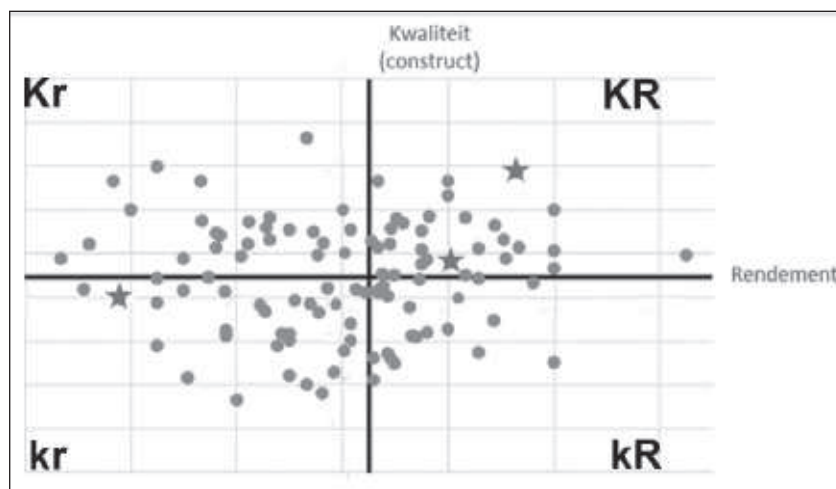
Werken aan onderwijskwaliteit
Ongetwijfeld wordt in het onderwijs vrijwel dagelijks de ambitie geuit: wij willen de

kwaliteit van het onderwijs verbeteren. Met de beschreven analyse in het achterhoofd roept zo'n uitspraak meteen vragen op als: wat wil je dan precies verbeteren en waarom? De dialoog over onderwijskwaliteit krijgt meer diepgang door een verbinding te maken met concrete resultaten en verbeteringen te plaatsen in het perspectief van een kwaliteitscultuur. Bijvoorbeeld aan de hand van de zeven lessen over de kwaliteitscultuur van de Onderwijsraad (2015):
1) Een duidelijke, gedeelde en doorleefde visie op goed onderwijs.
2) Verbeteringsgerichtheid vanuit collectief en individueel lerend vermogen.
3) Leiderschap op opleidingsniveau.
4) Een ondersteunende organisatiecultuur die aanzet tot samenwerken.
5) Hrm-beleid dat kwaliteitsbeleid ondersteunt.
6) Grote studentenbetrokkenheid.
7) Een externe oriëntatie.

Gedeelde visie en eigenaarschap
Werken aan onderwijskwaliteit vraagt allereerst om een gedeelde visie op wat je met elkaar wilt bereiken (beoogd resultaat, 'outcome') en hoe je dat wilt doen in de (onderwijsleeromgeving)⁶. Dat voorkomt

Visie op resultaat – Een gedeelde visie op beoogd resultaat is niet altijd vanzelfsprekend. Dat zien we bijvoorbeeld bij de Nederlandse Spoorwegen (NS) en de manier waarop het begrip vertraging wordt gehanteerd. Blijkbaar ziet men vertraging als een belangrijke ‘dissatisfier’, want het is dan mogelijk geld terug te vragen. Gelukkig rijden de meeste treinen op tijd. Maar natuurlijk gaat het wel eens mis. De ultieme vertraging is wanneer er helemaal geen treinen rijden. In dat geval kan er echter geen geld worden teruggevraagd omdat je eenvoudigweg geen reis hebt gemaakt. Voor reizigers die werken in de trein is het hebben van een zitplaats vaak veel bepalender dan de vertraging van de trein. Zitten in een vertraagde trein betekent langer kunnen werken zodat er in feite geen sprake hoeft te zijn van tijdsverlies – maar er kan wél een vergoeding worden gevraagd. Omgekeerd kan het moeten staan in een niet vertraagde trein wel degelijk tijdsverlies betekenen – maar daarvoor kan weer geen vergoeding worden gevraagd. Het beoogde resultaat van ‘treinen op tijd’ is dus lang niet altijd in lijn met ‘voorkomen tijdsverlies reiziger’.

ook dat bijvoorbeeld elke lage score op een vraag uit de Nationale Studenten Enquête meteen tot een (losse) verbeteractie leidt. Verbeteren is allereerst analyseren en prioriteren: “Wat zijn onze doelstellingen? Wat willen we bereiken? Wat leren we van evaluaties? En waar raken die aan onze doelstellingen? En waar juist niet?” Daarbij is ook een veilige omgeving waarin openlijk kan worden gesproken over goede en tegenvallende resultaten een randvoorwaarde. Bij een kwaliteitscultuur hoort ook eigenaarschap. Instellingsbrede gestandaardiseerde cursus- en docentevaluaties kunnen dat eigenaarschap bemoeilijken. Wanneer docententeams hun ‘eigen’ evaluaties hanteren en aan de hand daarvan met elkaar in gesprek gaan, kan dat heel veel opleveren (zie ook Vaal, K. de (2013)). En daarvoor zijn lang niet altijd ingewikkelde vragenlijsten met rapportages in grafieken en tabellen nodig, getuige een gesprek dat ik ooit meemaakte toen ik gastdocent was bij een collegereeks waarbij de docenten



Figuur 2: Plot van opleidingen op de dimensies Kwaliteit (verticaal, construct van gerealiseerd niveau NVAO standaard 4 en alumnevredeheid HBO Monitor) en Rendement (horizontaal, afstudeerrendement). Het snijpunt van de assen is het gemiddelde. In het kwadrant rechtsboven (KR) is sprake van zowel een gemiddeld hogere kwaliteit als een gemiddeld hoger rendement. De sterren zijn opleidingen met een toppositie in landelijke ‘rankings’.

elkaars colleges bijwoonden. Voorafgaand aan een nieuw college stonden ze bij elkaar, kopje koffie in de hand: “Ging goed hé, vorige keer?” “Ja, prima. Mooie presentatie. Alleen op het einde was het wat te theoretisch. Toen raakte je de studenten kwijt.” “Klopt, ik ga er wat sheets uithalen en wat vragen stellen aan de groep.”

drs. ing. Ouke J. Pijl

Teamleider kwaliteitszorg onderwijs en onderzoek bij Hogeschool Utrecht, bestuurslid Nederlands Netwerk voor Kwaliteitsmanagement (NNK), zelfstandig adviseur bij 2Reflect. Reacties en opmerkingen welkom: ouke.pijl@hu.nl. Met dank aan Marianne Polderdijk en Fiona Grooters.

Referenties

Onderwijsraad (2015). *Kwaliteit in het hoger onderwijs - Evenwicht in ruimte, regels en rekenschap*.

Scheerens, J., H. Luyten en J. van Ravens (2010). *Visies op onderwijskwaliteit*, Universiteit Twente.

Vaal, K. de, O. Pijl en B. van Schijndel (2013). *Kwaliteitsmanagement in de praktijk*, de hoofdstukken 10 ‘Meten is weten’ en 13 ‘Wat vindt uw stakeholder?’. Hilversum, Concept uitgangspunt.

¹ Bron: <http://www.kiesjestudie.nl/>.

² Bron: Studiekeuze 123, <http://www.studiekeuzeinformatie.nl/nse>.

³ Bron: Researchcentrum voor Onderwijs en Arbeidsmarkt (ROA), <http://roa.sbe.maastrichtuniversity.nl/?portfolio=school-leaver-surveys>.

⁴ Wet op het hoger onderwijs en wetenschappelijk onderzoek (WHW), artikel 7.12.

⁵ <http://cijfers.vereniginghogescholen.nl/index.htm>.

⁶ Zie bijvoorbeeld: Dolfing, R., e.a. (2012). ‘Domain-Specific Expertise of Chemistry Teachers on Context-Based Education About Macro–Micro Thinking in Structure–Property Relations’, *Research in Science Education*, June 2012, Volume 42, Issue 3, pp 567-588.