

Beoordelen met het oog op de toekomst

□ Dominique Sluijsmans

Studenten ervaren beoordelen vooral als een sluitstuk van het onderwijs. Vanuit de gedachte 'de voldoende is binnen, dus het leren is klaar' zal beoordelen slechts een korte termijn effect op leren hebben en is er nauwelijks sprake van beklijving van kennis en kunde. Deze summatieve functie van beoordelen is niet meer toereikend gezien de hedendaagse ontwikkelingen zoals een leven lang leren. In deze bijdrage worden tien kenmerken van beoordelen benoemd die gericht zijn op duurzaam leren.

Tien kenmerken van duurzaam beoordelen

1. De unieke student staat centraal
2. Het accent ligt op leren voor de toekomst
3. Betekenisvolle beoordelingstaken
4. Zelfsturing wordt gestimuleerd
5. De lerende voelt zich eigenaar van de beoordeling
6. De mate van duurzaamheid van kennis wordt erkend
7. Beoordelen is start van een nieuwe leercyclus
8. Heldere criteria en standaarden
9. De resultaten van beoordelingen zijn informatierijk
10. In de beoordeling wordt informeel en non-formeel leren erkend

Inleiding

Tot op de dag van vandaag heeft beoordelen in de ogen van studenten een afrekenende, selecterende of certificerende rol. Deze summatieve functie is gezien de complexiteit en heterogeniteit van onderwijsdoelen en arbeidscontexten niet meer toereikend om op een valide en betrouwbare wijze te bepalen hoe en wat studenten geleerd hebben en wat het effect van het geleerde is op de lange termijn. Het onderwijs zal zich meer moeten richten op een holistisch ontwerp van onderwijs waarin instructie, leren en beoordelen een coherent geheel vormen (Sluijsmans, 2008). Daarbij wordt het belangrijker de lerende te voorzien van rijke informatie die verder leren voedt, de zogenaamde formatieve functie. Naast de summatieve en formatieve functie kan een derde functie van beoordelen worden benoemd die gericht is op de individuele ontwikkeling van de student met het oog op zijn of haar toekomst. Dit zogenaamde duurzaam beoordelen hangt zeer nauw samen met het leerproces en geeft

de student een hoge mate van verantwoordelijkheid. Wat precies met duurzaam beoordelen wordt bedoeld, wordt met de volgende tien kenmerken verduidelijkt.

Kenmerken van duurzaam beoordelen

Kenmerk 1. De unieke student staat centraal

Huidige wijzen van beoordelen sluiten nog weinig aan bij de ontwikkeling van de individuele student. Zo worden prestaties van studenten in het 'normgerichte beoordelingsmodel' afgezet tegen de gemiddelde prestatie van de groep. Zit je bijvoorbeeld als laag presterende student in een hoog presterende groep, dan is de kans op een voldoende klein. In het 'criteriumgerichte beoordelingsmodel' worden prestaties niet zozeer vergeleken met de normprestatie van de groep, maar worden studenten beoordeeld op grond van een aantal beoordelingscriteria en standaarden. Resultaat van criteriumgerichte beoordelingen is dat sommige studenten



Duurzaam beoordelen geeft de student een hoge mate van verantwoordelijkheid

de vastgestelde standaarden op de criteria bereiken en andere studenten niet of met moeite.

In zowel het normgerichte als criteriumgerichte beoordelingsmodel schuilt het gevaar van het bekende toetsgedrag ('wash-back effect'; Alderson & Wall, 1993). De boeken en het portfolio verdwijnen in de kast als de voldoende binnen is. De individuele student met zijn unieke profiel verdwijnt volledig in een groepsgemiddelde of wordt slechts summatief afgevinkt op een lijst van beoordelingscriteria die niet op het niveau van de individuele student zijn afgestemd. Bij duurzaam beoordelen is de beoordeling afgestemd op de 'zone van naaste ontwikkeling' van de unieke student, om verder leren mogelijk te maken. Dit betekent dat studenten niet meer op hetzelfde moment met dezelfde beoordelingscriteria en op hetzelfde niveau aan de slag gaan.

Een beoordelingsmodel dat hierbij past is het ipsatief beoordelingsmodel ('ipse' is Latijn voor 'zelf'), waarbij beoordelingen van studenten, die zijn aangepast aan het niveau van de student, worden vergeleken in de tijd. Met andere woorden, de prestatie van de student wordt steeds vergeleken met zijn of haar eigen prestaties. Er is sprake van duurzaam beoordelen omdat criteria en standaarden flexibel kunnen worden aangepast en de beoordelingen zich niet meer beperken tot een momentopname maar continu plaatsvinden. Uiteraard blijven de criteria en standaarden essentieel om de kwaliteit te kunnen borgen. Wanneer de afstand tussen de ontwikkeling van de student te groot wordt ten opzichte van de minimale standaarden, kan tijdig worden gezien of bijvoorbeeld moet worden gekozen voor een andere standaard c.q. andere opleiding.

Kenmerk 2. Het accent ligt op leren voor de toekomst

In het huidige onderwijs heeft het beoordelen vooral betrekking op uitspraken over hoe de studenten in het verleden hebben gepresteerd en presteren op het moment van toetsafname.

Een tweede kenmerk van duurzaam beoordelen is dat de beoordeling informatie verschaft over de groei die is doorgemaakt ten aanzien van het vorige toetsmoment en tot welk nieuw leerdoel deze leidt. Dit betekent dat beoordelen niet langer het sluitstuk van het onderwijs is, maar steeds fungeert als een startpunt van een nieuwe leerfase.

Kenmerk 3. Betekenisvolle beoordelingstaken

In het huidige onderwijs bestaat een gebrekkige aansluiting tussen leren en beoordelen die onder andere zichtbaar is in de ontworpen verhouding tussen 'onderwijstijd' en 'beoordelingstijd'. Voorbeelden zijn de geclusterde proefwerkweken, (her)tentamens en de halfjaarlijkse portfoliobeoordelingen. Zolang onderwijstijd en beoordelingstijd zwak zijn geïntegreerd, zullen studenten zich vooral inspannen in de periodes vlak voor de beoordeling met het bekende uitstelgedrag als gevolg. Juist voor complexere kennisgebieden en vaardigheden is het cruciaal de relevante beoordelingscriteria en standaarden optimaal te integreren.

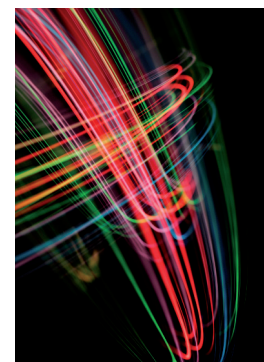
Een derde kenmerk van duurzaam beoordelen zijn betekenisvolle beoordelingstaken. In deze beoordelingstaken - die kunnen worden geschaard onder de noemer 'onderwijs' - worden leren en beoordelen geïntegreerd, omdat ze studenten aanzetten tot het werken aan gedragsdoelen die gericht zijn op het beklijven van kennis, vaardigheden en houdingen. Beoordelingstaken kunnen de student in vele vormen worden aangeboden, zoals opdrachten, situaties, problemen of projecten. Beoordeling- en onderwijstijd zijn door de integratie volledig in balans.

Kenmerk 4. Zelfsturing wordt gestimuleerd

Zelfsturing wordt als een essentiële vaardigheid beschouwd om leren voor de lange termijn mogelijk te maken. Zelfsturing veronderstelt dat studenten kunnen plannen en zichzelf (en anderen) beoordelen. Er is echter nog maar weinig empirisch bewijs dat studenten in staat zijn hun leervraag goed te definiëren en zichzelf te beoordelen. Zelfsturing is moeilijk. Onderzoek wijst uit dat studenten hun kennen en kunnen overschatten door een gebrek aan kennis over de beoordelingscriteria (Krueger & Dunning, 1999).

Het (leren) vormgeven aan het eigen leren is een vierde kenmerk van duurzaam beoordelen. Voor studenten betekent dit dat zij niet langer meer achteroverleunend het onderwijs consumeren, maar dat zij zich veel meer opstellen als 'prosumers', proactieve gebruikers die mede vormgeven aan het beoordelen door op eigen initiatief feedback te zoeken (Martens, 2007). Om zelfsturing te bevorderen zal in het ontwerp van beoordelingstaken ook het leren plannen (selecteren van doelen en taken) en het leren (zelf)beoordelen van belang worden. De ondersteuning die de student wordt geboden beperkt

Huidige vormen van beoordelen hebben dikwijls een korte termijn effect op leren



zich niet tot de beoordelingstaak, maar strekt zich uit tot leren plannen en beoordelen.

Kenmerk 5. De lerende voelt zich eigenaar van de beoordeling

Huidige beoordelingsmethoden zijn overwegend prescriptief van aard. Omdat de opleiding en de docent bepalen wat, hoe en wanneer wordt beoordeeld, hebben zij nog steeds de volledige controle over het onderwijs-leerproces van de student. Omdat de student nauwelijks betrokken wordt bij de voorbereiding en uitvoering van de beoordeling, zal het effect van de beoordeling op lange termijn leren minimaal zijn. De laatste jaren zijn veel initiatieven genomen om studenten meer te betrekken bij beoordelen, zoals zelf-, peer- en portfoliobeoordelingen. Maar ook hier gaat het vooral om vormen van beoordelen die de student worden opgelegd. Zo wordt per docent vastgelegd wanneer de reflectierapporten of het portfolio moeten worden geschreven, in hoeveel woorden en volgens welke structuur. Het invoeren van zelfbeoordeling op deze manier demotiveert studenten eerder en leidt tot 'reflectie-moeheid'. Reflecteren wordt een routine en studenten schrijven op wat de docent graag in het reflectieverslag terug wil zien. Het geven van controle aan studenten over de wijze waarop ze worden beoordeeld, wanneer ze worden beoordeeld en door wie ze worden beoordeeld, kan leiden tot een actievere betrokkenheid van studenten.

Het creëren van eigenaarschap is een vijfde kenmerk van duurzaam beoordelen. Dit kan vervolgens tot een hogere motivatie en tot meer leren leiden dan wanneer het beoordelen extern wordt aangestuurd. Dit perspectief sluit goed aan bij motivatietheorieën, zoals de 'zelfdeterminatie-theorie' (Ryan & Deci, 2000), waarin competentie, autonomie en sociale verbondenheid de basisbehoeften zijn. Studenten kunnen bijvoorbeeld zelf kiezen op welke criteria en standaarden ze worden beoordeeld.

Kenmerk 6. De mate van duurzaamheid van kennis wordt erkend

Gezien de opkomende kenniseconomie is het van belang dat studenten leren nieuwe kennis te construeren en deze kennis adequaat toe te passen in situaties. Recente ontwikkelingen zoals het vaststellen van diverse canons en de wens vanuit het ministerie voor een krachtige kennisbasis voor diverse beroepende tonen aan dat kennis cruciaal is.

Een zesde kenmerk van duurzaam beoordelen is dat de duurzaamheid van kennis wordt erkend. Hiermee wordt bedoeld dat de waarde van kennis voor de lange termijn wordt vastgesteld.

Kenmerk 7. Beoordelen is start van een nieuwe leercyclus

Doordat veel beoordelingen worden afgesloten met een voldoende of een cijfer, wordt een soort 'eindigheid' in leren gesuggereerd. In het normgerichte beoordelingsmodel is er bijvoorbeeld voor diegenen die het beste zijn in hun groep geen reden om verder te leren als de norm is behaald. Ook in het criteriumgerichte beoordelingsmodel is er geen reden om verder te leren als de criteria en standaarden zijn behaald. In het ipsatief beoordelingsmodel zoals beschreven bij het eerste kenmerk, zal de lerende niet alleen informatie zoeken en krijgen over zijn prestaties van nu, maar zal vooral worden geïnvesteerd in het bieden van handvatten voor de volgende stap in het leerproces.

Een zevende kenmerk van duurzaam beoordelen is dat beoordelen niet het sluitstuk van leren is, maar vooral de functie heeft een nieuwe leercyclus op te starten.

Kenmerk 8. Heldere criteria en standaarden

Helder geformuleerde beoordelingscriteria en standaarden vormen nog te weinig de basis in het onderwijs. Ze worden meestal per vak door de individuele docent bepaald, vaak op een moment dat het onderwijs al 'achter de rug is' en er nog een beoordeling moet plaatsvinden. De beoordelingscriteria en standaarden zitten daarbij vooral in het hoofd van de docent ('tacit knowledge'; Sadler, 1989), waardoor de student niet kan begrijpen waaraan moet worden voldaan en zich dus ook niet adequaat kan voorbereiden. Uit onderzoek blijkt dat studenten nogal eens verrast worden door de inhoud van een beoordeling, omdat de verwachtingen van de student over de beoordelingscriteria niet overeenkomen met de verwachtingen van de docent (Broekkamp & van Hout-Wolters, 2007). Naast heldere beoordelingscriteria zijn ook heldere standaarden ('niveau') nodig. Wat betreft de standaarden is er niet altijd helderheid, blijkt uit reacties van studenten als 'De toets was veel te moeilijk!'. Helderheid over beoordelingscriteria en standaarden is het achtste kenmerk van duurzaam beoordelen. Hiervoor is het van belang zogenaamde gedragsdoelen te definiëren. Gedragsdoelen worden afgeleid van de gewenste



Duurzaam beoordelen past bij een leven lang leren

vaardigheden in een beroep of bepaald kennisdomein. Goed gedefinieerde gedragsdoelen beschrijven de condities waaronder kennis en kunde moeten worden gedemonstreerd, en de beoordelingscriteria en -standaarden. De condities waaronder de student een taak uitvoert bepalen in veel gevallen de standaard van een criterium. Vervolgens is een groot aantal beoordelingen met duidelijke beoordelingscriteria en -standaarden als basis essentieel. Alleen dan kunnen resultaten op deze taken met elkaar worden vergeleken en is het ontwerp van een ipsatief beoordelingsmodel (zie eerste kenmerk) mogelijk.

Kenmerk 9. De resultaten van beoordelingen zijn informatierijk

Over de feilbaarheid van de menselijke beoordelaar is al veel gepubliceerd (zie bijvoorbeeld De Groot, 1966). Vele studies hebben aangetoond dat docenten met vergelijkbare deskundigheden dezelfde taak verschillend beoordelen. Tot op heden is het gebrek aan overeenstemming tussen beoordelaars een heikel punt. In het bijzonder als het gaat om certificerende beslissingen wordt de mens als beoordelaar in twijfel getrokken. Iedere docent hanteert zo zijn eigen 'beoordelingsregels' en '-strategieën', maar expliciteert deze helaas niet of nauwelijks. In beoordelingen wordt de aanvankelijk rijke informatie van beoordelaars vaak teruggefilterd tot een saai, simpel, nietszeggend cijfer, de student nauwelijks motiverend tot verder leren (Schuwirth, 2007). Ook wordt het cijfer op een moment verstrekt dat de student de beoordelingstaak nog maar nauwelijks voor de geest weet te halen. Niet alleen de beoordelingscriteria waren bij het verstrekken van de opdracht onvoldoende specifiek gecommuniceerd, ook is lang niet altijd sprake van specifieke en constructieve feedback. Hoewel onderzoek al meerdere malen heeft aangetoond dat cijfers de meest zwakke vorm van feedback zijn, zien we deze nog veel in het onderwijs. Omdat het becijferen zich vooral richt op het vaststellen van een prestatie op een bepaald tijdstip, zijn beoordelingen per definitie summatief. Beoordelingen die nu vaak als formatief worden bestempeld krijgen die benaming vooral als het cijfer 'niet meetelt'. Zolang echter informatie ontbreekt over de betekenis van de beoordeling en het toekomstig leertraject dat de student zal doorlopen is er geen sprake van duurzaam beoordelen.

Een negende kenmerk van duurzaam beoordelen is de informatierijkheid van beoordelingen. Docenten expliciteren hun beoordelingsaanpak naar studenten, zodat zij beter begrijpen waarop en hoe zij worden beoordeeld. De communicatie in duurzaam beoordelen betekent dat niet alleen het resultaat van een beoordeling wordt gecommuniceerd (voldoende, cijfer, geslaagd, goed gedaan), maar dat deze communicatie de student handvaten biedt voor verdere ontwikkeling. Het benadrukken van de individuele vooruitgang bevordert niet alleen het geloof in eigen kunnen, maar reduceert ook de neiging zichzelf continu te vergelijken met medestudenten.

Kenmerk 10. In de beoordeling wordt informeel en non-formeel leren erkend

Het laatste maar zeker niet minst belangrijke kenmerk van duurzaam beoordelen is dat in de beoordeling ook de informele en non-formele leeromgeving van de student wordt erkend. Was het tot nu toe zo dat alleen formeel leren werd beoordeeld, het is relevant te kijken naar andere leerervaringen van studenten die kunnen bijdragen aan een goede beoordeling van kennis en kunde (Joosten-ten Brinke, Sluijsmans, Brand-Gruwel & Jochems, 2008). Ook kan de wijze waarop studenten leren relevant zijn om te bepalen waarom zij op een bepaalde manier presteren. Tot nu toe wordt informeel en non-formeel leren erkend bij procedures voor het erkennen van eerder of elders verworven competenties (EVC). Gezien het belang van continuïteit in beoordelen (zie het derde kenmerk), is het betrekken van leer- en leefwereld van de student doorlopend van belang.

What's new?

Een terechte vraag is wat duurzaam beoordelen toevoegt aan bestaande concepten als formatief of ontwikkelingsgericht beoordelen. Onderzoek naar de kracht van formatieve beoordelingen (Black & William, 1998) heeft al bewezen dat deze wijze van beoordelen het leren bevordert. Al in 2002 heeft de Assessment Reform Group (ARG) principes voor het veranderen van de beoordelingspraktijk in de klas geformuleerd. Hoewel deze principes overeenkomsten vertonen met de kenmerken zoals uiteengezet in deze bijdrage (bijvoorbeeld het belang van heldere criteria en oog voor motivatie), schieten zij nog tekort als het gaat om het ontwerp van beoordelingstaken, de effecten van beoordelen op de

Bij duurzaam beoordelen wordt de waarde van kennis voor de lange termijn vastgesteld



lange termijn, het belang van non- en informeel leren, de informatierijkheid van beoordelingen, de communicatie over beoordelen en de continuïteit van beoordelen. Wanneer onderwijsgeven en de huidige beoordelingsmethoden onder de loep nemen blijken deze vaak nog onvoldoende duurzaam ontworpen: beoordelingen zijn informatiearm door het cijfersysteem, de inhoudsvaliditeit is laag, lerenden worden nauwelijks betrokken bij de beoordeling en de leereffecten van beoordelen worden te weinig onderkend en onderzocht. De ontwikkeling van een duurzame beoordelingsvisie die past bij huidige en toekomstige ontwikkelingen, zal voor alle niveaus en in alle sectoren van onderwijs cruciaal worden, maar staat op dit moment nog in de kinderschoenen.

Afsluitend

Samenvattend kunnen drie pijlers worden benoemd waarop duurzaam beoordelen rust:

- (1) een goed ontwerp van onderwijs waarbij leren en beoordelen samensmelten,
- (2) een heldere communicatie met studenten en collega's over het waarom, wat en hoe van beoordelen en
- (3) het stimuleren van een zelfsturende houding.

Uit de tien beschreven kenmerken is te herleiden dat duurzaam beoordelen van studenten en docenten een blik over de grenzen van het diploma heen vraagt. Het leren houdt immers niet op na een opleiding. Studenten zullen moeten leren hoe ze blijven leren, simpelweg omdat de kenniseconomie dat van ze vraagt. De praktijk van alledag maakt het zeer lastig om op individueel niveau studenten informatie te verschaffen over hun kennen en kunnen, op een manier die ook nog eens past bij hun leervraag en persoonlijkheid. Omdat deze flexibilisering ook vanuit logistiek perspectief veel van het onderwijs zal vragen, zal beoordelen bij vraaggestuurd leren moeilijk te realiseren zijn op korte termijn. Dit maakt het lastig te voorspellen wanneer duurzaam beoordelen vruchten zal afwerpen. Voor nu is een eerste waardevolle stap om onderwijsgeven naar hun visie op beoordelen te vragen: Waarom wil je beoordelen en op welke manier zou je dat graag willen vormgeven? Door deze vraag centraal te stellen in het debat over beoordelen komt vaak (het gebrek aan) de aansluiting tussen onderwijs en beoordelen naar de oppervlakte. Het 'waarom' van beoordelen is echter voorwaardelijk voor het 'wat' en 'hoe' om tot structureel andere beelden over beoordelen in

het onderwijs te komen. Bewustwording van de huidige beoordelingspraktijk is de eerste essentiële stap naar structurele verandering.

Literatuur

- Alderson, J.C. & Wall, D. (1993). Does washback exist? *Applied Linguistics*, 14(2), 115-129.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7-74.
- Broekkamp, H. & van Hout-Wolters, B.H.A.M. (2007). Students' adaptation of study strategies when preparing for classroom tests. *Educational Psychology Review*, 19, 401-428.
- De Groot, A. (1966). Vijven en zessen: *Cijfers en beslissingen*. Groningen: Wolters-Noordhoff.
- Joosten-ten Brinke, D., Sluijsmans, D.M.A., Brand-Gruwel, S. & Jochems, W. (2008). The quality of procedures to assess and credit prior learning: implications for design. *Educational Research Review*, 1(3), 51-65.
- Krueger, J. & Dunning, D. (1999). Unskilled, and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 82, 180-188.
- Martens, R.L. (2007). *Positive learning met multimedia. Onderzoeken, toepassen en generaliseren*. Inaugurele rede. Heerlen/Hilversum: Open Universiteit Nederland/Teleac/NOT.
- Ryan, R.M. & Deci, E.L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, 55, 68-78.
- Sadler, R.D. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18, 119-144.
- Schuwirth, L.W.T. (2007). *Metten de maat genomen*. Inaugurele rede. Maastricht: Universiteit Maastricht.
- Sluijsmans, D.M.A. (2008). *Betrokken bij beoordelen*. Lectorale rede. Nijmegen: Hogeschool Arnhem en Nijmegen.
- The Assessment Reform Group (2002). *Assessment for Learning: 10 principles. Research-based principles to guide classroom practice*. Gevonden op <http://www.qcda.gov.uk/4335.aspx>.

□ Mw. dr. D.M.A. Sluijsmans is als lector 'Duurzaam beoordelen in vraaggestuurd leren' verbonden aan de Faculteit Educatie van de Hogeschool Arnhem Nijmegen. E-mail: dominique.sluijsmans@han.nl.

Gesignaleerd

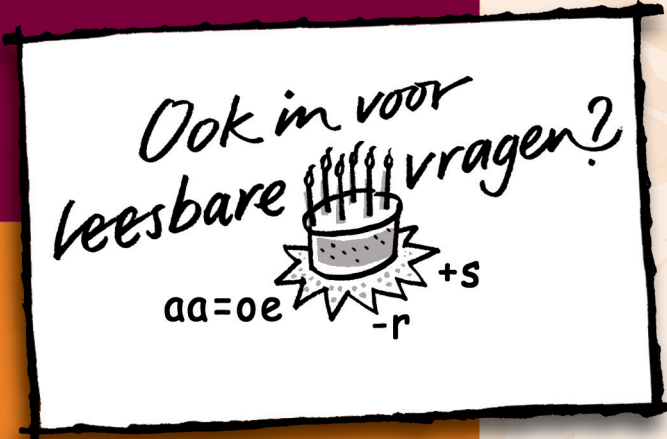
Positie examencommissies hoger onderwijs wordt versterkt

Persbericht INSPECTIE van het ONDERWIJS – 26 november 2009

De positie van examencommissies in het hoger onderwijs wordt op korte termijn versterkt. Dit kwam naar voren bij de conferentie 'Examencommissies en examenkwaliteit' van de Inspectie van het Onderwijs op 25 november 2009. Nieuwe wetgeving beoogt grotere deskundigheid en onafhankelijkheid van de commissies.

Aanleiding voor de conferentie was onder andere het rapport 'Boekhouder of wakend oog' van de onderwijsinspectie. Hieruit bleek dat een groot deel van de commissies onvoldoende is toegerust voor een gezaghebbende positie.

Inspecteur-generaal van het Onderwijs Annette Roeters noemt examencommissies het scharnier tussen opleiding en maatschappij. Volgens de inspectie moet de betrouwbaarheid van het getuigschrift buiten kijf staan. Nu nog laat het interne toezicht op toetsing en examinering te vaak te wensen over.



Ook in voor
leesbare vragen?

aa=oe -r +s



TEELN
Kennis MANAGEMENT

WWW.TEELN.NL

Ook de EVC procedures van



zijn beoordeeld door
de Stichting Examenkamer

