

# De validiteit van het beoordelen van docentcognities en docentgedrag in docentportfoliobeoordelingen<sup>1</sup>

M. van der Schaaf, K. Stokking en N. Verloop

## Samenvatting

In dit onderzoek staat de relatie tussen docentcognities en docentgedrag in docentportfoliobeoordelingen centraal. In het onderzoek stelden 18 ervaren docenten portfolio's samen. De portfolio's hadden betrekking op cognities en gedrag van docenten bij het instrueren, begeleiden en beoordelen van onderzoeksvaardigheden van leerlingen in de gammavakken. We analyseerden op kwalitatieve wijze de in de portfolio's opgenomen cognities en docentgedragingen. Tevens beoordeelden zes getrainde beoordelaars paarsgewijs de 18 portfolio's op acht eerder vastgestelde beoordelingscriteria. Daarnaast werd elke docent door de eigen leerlingen ( $n = 317$ ) beoordeeld in een leerlingvragenlijst over docentgedrag. De verkregen data (van de onderzoekers, de leerlingen en de beoordelaars) vormden de input voor lineaire meerniveau-analyse van de samenhang tussen docentcognities en gedrag. Daaruit bleek dat de beoordelingen door de leerlingen van het gedrag van hun docenten (over hoe zij onderzoeksoopdrachten instrueren en begeleiden) significant konden worden voorspeld vanuit de beoordelingen door de externe beoordelaars. Verder bleken docenten die volgens de beoordelaars een hoge score behaalden op het beoordelingscriterium 'vooraf nadenken over instructie en begeleiding' ook door hun leerlingen positiever te zijn beoordeeld op hun gedrag dan de docenten die op dit criterium volgens de beoordelaars een lagere score behaalden. We bediscussiëren de implicaties van de resultaten en geven suggesties voor verder onderzoek.

decennia verschoven van een benadering gebaseerd op behavioristische uitgangspunten, voornamelijk gericht op het beoordelen van docentgedragingen, naar een meer cognitief psychologische benadering waarin gedrag wordt gezien als geïntegreerd met docentcognities en de context van het onderwijzen (Clark & Peterson, 1986; Reynolds, 1992; Shulman, 1986; Verloop, 2003). Thans worden docentcompetenties veelal opgevat als een integratie van cognities, vaardigheden, houdingen en persoonskenmerken die nodig zijn voor het uitvoeren van professionele taken in een beroepsmatige context (Eraut, 1994; Gonczy, 1994). Wij sluiten bij deze omschrijving aan. Fundamenteel in deze omschrijving is de aanname dat docentgedrag onlosmakelijk is verbonden met docentcognities over leren en onderwijzen in een bepaalde context. Dit leidt ertoe dat ook docentcognities, naast docentgedrag, tegenwoordig regelmatig worden betrokken bij docentbeoordelingen (Beijaard & Verloop, 1996). Docentcognities kunnen worden beschreven vanuit de inhoud van de cognities (bijvoorbeeld met een focus op de vakinhoudelijke of de pedagogische kennis van docenten), het type cognities (bijvoorbeeld opvattingen of feitenkennis) en de aard van de cognities (bijvoorbeeld meer of minder bewust, persoonlijk en contextgebonden). Hoewel we een relatie veronderstellen tussen docentcognities en docentgedrag is het echter onduidelijk onder welke omstandigheden deze zich voordoet (Mathijsen, 2006).

Voor het in kaart brengen van docentcognities en gedrag (cognities als onderdeel van het competentiebeprijp) is een brede gegevensbasis nodig. Daartoe wordt steeds meer gewerkt met portfolio's waarin docenten zowel hun gedragingen als hun cognities (bijvoorbeeld cognities over onderwijzen) expliciteren in een geselecteerde neerslag van hun dagelijks werk in een bepaalde context. Portfoliobeoordeling is een middel om na te gaan in welke mate docenten voldoen aan ge-

222

PEDAGOGISCHE  
STUDIËN

2008 (85) 222-239

## 1 Inleiding

Tegenwoordig neemt de beoordeling van docentcompetenties in de beroepsgroep van docenten een prominente plaats in. Het perspectief op docentbeoordelingen is afgelopen

wenste beoordelingscriteria (summatief) en om feedback te kunnen geven met het oog op hun verdere professionele ontwikkeling (formatief). Wij richten ons in dit onderzoek op beide functies.

Veelal wordt aangenomen dat portfolio's, afhankelijk van hun inhoud en vorm, een authentieke weerspiegeling geven van docentcompetenties in een werkcontext. Als gevolg daarvan worden portfolio's vaak als valide beoordelingsinstrumenten beschouwd. Een fundamentele voorwaarde voor een valide beoordeling is dat de scores die in een beoordeling aan een portfolio worden toegekend daadwerkelijk zijn gebaseerd op de competenties van de betreffende docent. Deze voorwaarde is van wezenlijk belang, omdat het de grondslag vormt waarop de beoordeling en de daaruit voortvloeiende feedback en consequenties zijn gebaseerd. De gegevens in een portfolio zijn echter vaak uniek, situationeel, persoonlijk en beschrijvend van aard en vergen veel interpretatie voordat ze kunnen worden beoordeeld. Een consequentie daarvan is dat veel gebruikte methoden om de validiteit van beoordelingen aan te tonen, bijvoorbeeld gebaseerd op het gebruik van paralleltoetsen, minder geschikt zijn, aangezien dergelijke toetsen niet voorhanden zijn en het persoonlijke en contextgebonden karakter van een portfolio zulke paralleltoetsen ook minder zinvol maken.

Om toch een indicatie te krijgen van de validiteit van portfoliobeoordelingen hantieren we in dit onderzoek daarom een andere aanpak. We vergelijken de inhoud van de verschillende docentportfolio's (inclusief zelfbeschrijvingen en videoregistraties van lessen), zoals geanalyseerd door twee onderzoekers, met de beoordelingen van leerlingen en beoordelaars. Daarbij maken we een onderscheid tussen gedrag (zoals bijvoorbeeld blijkt uit videoregistraties) en cognities zoals door docenten weergegeven in hun portfolio's, en de beoordelingen van dat gedrag en die cognities door de beoordelaars.

De onderzoeksvragen zijn:

- 1) Hoe kunnen in portfolio's weergegeven docentcognities worden beoordeeld?
- 2) Wat is de relatie tussen cognities zoals weergegeven door docenten in hun portfolio en gedrag zoals beoordeeld door leerlingen en beoordelaars?

Het onderzoek is onderdeel van een meer omvattend onderzoek naar het beoordelen van docentcompetenties van ervaren docenten bij het instrueren, begeleiden en beoordelen van onderzoeksvaardigheden van leerlingen in de bovenbouw havo-vwo in de gammavakken. Onderzoeksvaardigheden maken deel uit van de examens in de Tweede Fase havo-vwo. Het ontwikkelen van onderzoeksvaardigheden bij leerlingen sluit aan bij nieuwe opvattingen over leren, waarin het zelf actief construeren van kennis wordt benadrukt. Dit vraagt van docenten een rolverandering waarin naast kennisoverdracht ook de begeleiding van meer zelfstandig lerende leerlingen en meer authentieke vormen van beoordelen belangrijk zijn. Dit doet een beroep op deels nieuwe cognities en gedragingen van docenten. Hoe deze cognities en gedragingen kunnen worden beoordeeld in een portfolio is nog niet onderzocht. Inzicht hierin is nodig om meer begrip te krijgen van portfoliobeoordeling en hoe de kwaliteit van deze vorm van beoordelen kan worden verbeterd.

## 2 Validiteit van het beoordelen van portfolio's

Portfoliobeoordeling moet uiteraard aan bepaalde kwaliteitseisen voldoen. Het is echter een cruciale en actuele kwestie welke die eisen zouden moeten zijn. Met name bestaat discussie over de vraag in hoeverre bij portfoliobeoordeling traditionele kwaliteitscriteria van betrouwbaarheid en validiteit kunnen worden losgelaten en minder hoge eisen zouden moeten gelden. Naar onze mening moet ook portfoliobeoordeling voldoen aan kwaliteitseisen met betrekking tot betrouwbaarheid en validiteit. Betrouwbaarheid verwijst naar de stabiliteit over de tijd (bijvoorbeeld bij herhaalde metingen) en de consistentie tussen beoordelingen (bijvoorbeeld tussen afzonderlijke beoordelingen van twee of meer beoordelaars). Validiteit verwijst naar de vraag of de beoordeling de beoogde constructen, in dit geval docentcompetenties, meet en niet iets anders.

Deels gebaseerd op Messick (1989, zie ook Stokking, Van der Schaaf, Jaspers, &

Erkens, 2004), vatten we validiteit op als een begrip dat een aantal facetten dekt:

- 3) Construct. Het te beoordelen construct (domein, taak) moet zo goed mogelijk worden gedekt. Dit houdt in dat een beoordeling de relevante docentcompetenties zou moeten dekken.
- 4) Proces. De beoordeling moet zijn gebaseerd op een model van (de uitvoering van) de betreffende taak, respectievelijk de ontwikkeling van de relevante kennis en vaardigheden. De beoordeling van competenties zou moeten aansluiten bij de uitvoering van de beoogde onderwijstaken.
- 5) Scoring. Het scoringsmodel en het scoringsvoorschrift moeten de structuur van het construct, domein of taak weerspiegelen. Wanneer de onderwijstaken worden opgevat als bestaande uit afzonderlijke fasen (bijvoorbeeld onderwijzen in termen van pre-actieve, interactieve en post-actieve fasen), zou dit moeten worden weerspiegeld in de scoringsinstructies.
- 6) Specificiteit. De scores moeten alleen de beoogde competenties representeren en niet ook iets anders meten. Dit betekent dat beoordelingen enkel voldoende onderscheid moeten maken tussen meer of minder competente docenten en niet ook op irrelevant punten een onderscheid moeten maken (bijvoorbeeld sekse of etniciteit).
- 7) Convergentie. Beoordelingen van hetzelfde construct, domein of taak met verschillende methoden moeten zoveel mogelijk corresponderen.
- 8) Generaliseerbaarheid. De reikwijdte van de resultaten (de geldigheid voor een bepaalde range van taken en condities) moet kunnen worden aangegeven.

Al met al is beoordelen van docentcompetenties een complex proces. Hierin staat constructvalidering (facet a) centraal, omdat deze de conceptuele afbakening en definitie betreft van de te beoordelen constructen (bijvoorbeeld docentcompetenties) en het empirische bewijs moet leveren dat deze beoordelingen adequaat zijn. Docentcompetenties zijn contextgebonden en persoonlijk en daarom is het extra lastig om een indruk te krijgen van de validiteit van docentportfolio-

beoordelingen. In dit onderzoek hanteren we daarom een specifieke aanpak om indicaties voor deze validiteit te krijgen, namelijk door de inhoud van docentportfolio's kwalitatief te analyseren en de uitkomsten daarvan te vergelijken met beoordelingen van leerlingen en beoordelaars. We spitsen dit toe op de cognities van docenten (als onderdeel van hun competenties) en de relatie tussen cognities en gedrag.

### 3 Het beoordelen van docentcognities

#### 3.1 Docentcognities

Docentcognities beïnvloeden wat docenten in bepaalde omstandigheden waarnemen en wat ze, in deze omstandigheden, als mogelijk gedrag beschouwen en de kennis die ze in dergelijke omstandigheden aanwenden (Clark & Peterson, 1986). Het wordt algemeen verondersteld dat docentcognities, afhankelijk van de situatie en de manier waarop deze wordt geïnterpreteerd, verschillen in specificiteit en sterkte. Er wordt aangenomen dat cognities worden geactiveerd in clusters en dat alleen de meest prominente cognities binnen een bepaalde, specifieke situatie het uiteindelijke gedrag van de docent beïnvloeden (Aguirre & Speer, 2000; Ajzen, 2002; Schoenfeld, 1998). Dat impliceert dat niet alle docentcognities een rol spelen bij het uiteindelijke docentgedrag.

Ook kunnen cognities docentgedrag oproepen op zowel onbewuste als bewuste wijze. Bij een onbewuste wijze worden cognities routinematig geactiveerd. De clusters van cognities die bepaald docentgedrag oproepen zijn hierbij sterk context- en persoonsgebonden en bestaan voornamelijk uit *tacit* kennis die niet eenvoudig expliciet kan worden gemaakt (Eraut, 1994). Op een bewuste manier worden cognities met inspanning geconstrueerd in een bepaalde context en ze worden verondersteld aan te zetten tot doelgericht (intentioneel) gedrag.

In aansluiting hierop vatten we doelformuleringen door de docent (bijvoorbeeld leerdoelen) op als manifestaties van hun cognities omtrent de effecten die ze op langere of kortere termijn beogen te bereiken (Aguirre

& Speer, 2000). Op langere termijn gaat het om langetermijndoelen die vooraf aan het onderwijzen worden geformuleerd (in de pre-actieve fase van onderwijzen, Clark & Peterson, 1986). Ze betreffen bijvoorbeeld doelstellingen die docenten met een vooraf geformuleerde onderzoeksopdracht bij de leerlingen nastreven, en zijn daarom in de context van ons onderzoek relevant. Ook kan het gaan om intenties van docenten die doorgaans op korte termijn een rol spelen, tijdens het onderwijzen (in de interactieve fase van onderwijzen) en bij bepaalde (contextgebonden) gebeurtenissen in de klas (Saxe, 1991). In dit onderzoek beschouwen we de eerste (langetermijndoelen) als doelen van docenten en de tweede (kortetermijnintenties) als intenties ten aanzien van een bepaald handelen.

### **3.2 Geconstrueerde docentcognities**

Cognities zijn deels impliciet, en impliciete cognities kunnen alleen worden bestudeerd als ze worden geëxpliciteerd. We gaan ervan uit dat docenten in hun portfolio's hun cognities weergeven op basis van een selectie, interpretatie en (re-)constructie van hun eigen cognities.

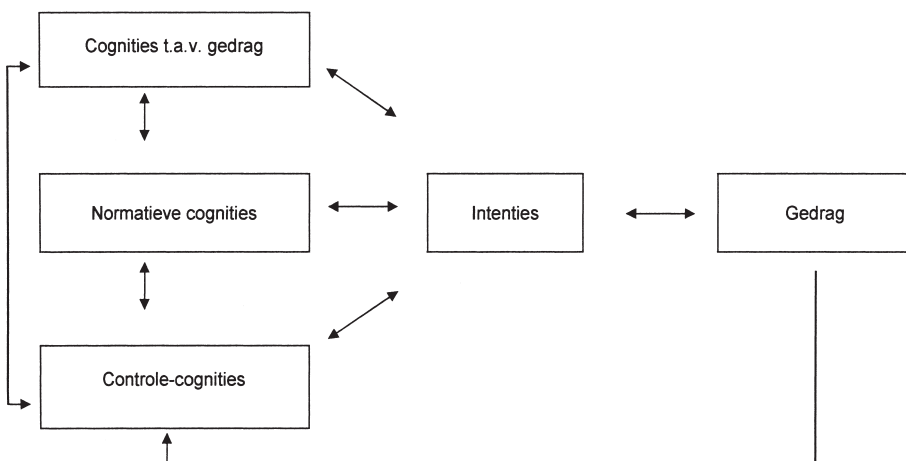
In het onderzoek interpreteren we de cognities voorafgaande aan het onderwijzen (doelen) en tijdens het onderwijzen (uitmondend in intenties tot bepaald handelen) van docenten in hun werkcontext, zoals zij die hebben geëxpliciteerd in hun portfolio's (bijvoorbeeld in reflectieverslagen). We doen dit aan de hand van de Theory of Planned Behavior (TPB, Ajzen, 1985; Ajzen & Fishbein, 2005). De TPB is reeds eerder waardevol gebleken in onderzoek naar de relatie tussen geëxpliciteerde cognities en gedrag in de (sociale) wetenschappen en bij onderwijsveranderingsprocessen (vgl. Crawley, 1990; Haney, Czerniak, & Lumpe, 1996) en is ook relevant voor de context en het doel van ons onderzoek. De TPB is gebaseerd op drie clusters van cognities die variantie veroorzaken in de bewust nagestreefde doelen en intenties van mensen ten aanzien van bepaald gedrag (Fishbein, Triandis, Kanfer, Becker, Middletstadt, & Echler, 2001) (zie Figuur 1). Binnen de TPB wordt gesproken van opvattingen in plaats van cognities. In dit artikel prefereren we echter de term cognities. Deze clusters

zijn: cognities over de verwachte gevolgen van een bepaald gedrag (cognities ten aanzien van bepaald gedrag); cognities over de verwachtingen van significante anderen, bijvoorbeeld collega's, ten aanzien van een bepaald gedrag (normatieve cognities); en cognities over mogelijk aanwezige factoren die bepaald gedrag kunnen faciliteren of belemmeren (controlecognities). Tezamen leiden cognities ten aanzien van bepaald gedrag, normatieve cognities en controlecognities tot nagestreefde doelen en intenties om bepaald gedrag uit te voeren. Zo zal bijvoorbeeld bij docenten die denken over de competenties en support te beschikken om onderzoeksvaardigheden bij leerlingen te ontwikkelen, de waargenomen controle groter zijn, evenals de intentie om het gedrag uit te voeren. Docenten die verwachten niet over dergelijke competenties en support te beschikken, zullen vermoedelijk zwakkere controlecognities hebben en als een gevolg daarvan zwakkere doelen en intenties om onderzoeksvaardigheden bij leerlingen te ontwikkelen.

### **3.3 Beoordeelde docentcognities**

Uit onderzoek van Van der Schaaf (2005), mede gebaseerd op beoordelingsmodellen in de sociaal cognitieve psychologie, blijkt dat beoordelaars gebruik maken van schemata om de cognities die docenten in hun portfolio weergeven te beoordelen, te voorspellen en te begrijpen (Carlston, 1992, 1994; Jones & Davis, 1965). Deze schemata zijn te vergelijken met persoonlijke constructen (Kelly, 1955), waarmee tijdens het beoordelen de gepercipieerde docentcompetenties worden gefilterd door de persoonlijke ervaringen en cognities van de beoordelaar. Daarom maken we een onderscheid tussen de door de docent weergegeven cognities en de door de beoordelaar beoordeelde cognities. Uit onderzoek blijkt dat de mate waarin deze persoonlijke ervaringen en cognities van beoordelaars in de beoordeling een rol spelen kan worden verminderd door beoordelaarstrainingen (Lievens, 2001; Woehr & Huffcutt, 1994).

Wanneer kan worden aangetoond dat beoordelaars zich bij hun beoordeling daadwerkelijk baseren op de door de docenten weergegeven cognities is dat een indicatie voor de validiteit van de beoordeling, in het bijzonder



Figuur 1. Relaties tussen docentcognities, intenties en gedrag (gebaseerd op Ajzen & Fishbein, 2005).

ten aanzien van de facetten construct, proces en scoring, zoals geformuleerd in paragraaf 2. Overigens zouden beoordelaars ook omwille van de aanvaardbaarheid van de beoordeling voor de docent duidelijk moeten maken dat hun oordeel aansluit bij de feitelijke inhoud van het portfolio. Dit is ook voor het kunnen geven van voorbeelden in de feedback aan docenten cruciaal.

#### 4 Beoordelingscriteria en portfolio-inhoud

De beoordelingscriteria en de portfolio-inhoud zijn in ons onderzoek bouwstenen bij het verkrijgen van valide portfolio-beoordelingen. Beide, criteria en inhoud, zijn ontwikkeld en empirisch onderzocht in eerdere onderzoeken (Van der Schaaf, 2005). In deze paragraaf volgt daarvan een samenvatting.

##### 4.1 Beoordelingscriteria

In de eindtermen voor het vwo worden onderzoeksvaardigheden in de Tweede Fase behandeld in termen van een serie van activiteiten, die als volgt kan worden beschreven: 1. onderzoeksonderwerp formuleren, passend bij het betreffende vakgebied; 2. onderzoeksvragen formuleren; 3. onderzoeksplan ontwikkelen; 4. data verzamelen en selecteren; 5 data analyseren; 6. resultaten formuleren; 7. conclusies trekken; 8. onderzoek

evalueren; 9. onderzoek rapporteren en presenteren (vergelijk Stokking & Van der Schaaf, 2000).

Bij het instrueren en begeleiden van onderzoeksvaardigheden zullen docenten keuzes moeten maken ten aanzien van de mate waarin zij de leeromgeving van leerlingen willen structureren en sturen. Sterke sturing door de docent, waarbij leerlingen amper of geen ruimte hebben om eigen keuzes te maken, ligt daarbij niet voor de hand (Roth & Roychoudhury, 1993). Dat geldt evenmin voor een zeer open leeromgeving. Leerlingen lijken het meest gebaat bij gedeelde sturing (niet te gestructureerd en niet te open), omdat beide extremen kunnen leiden tot oppervlakkig leren (Kanselaar, Galen, Beemer, Erkens, & Gravemeijer, 1999).

Bij het helpen ontwikkelen van onderzoeksvaardigheden van leerlingen vervullen docenten professionele taken. Deze taken zijn in te delen naar pre-actieve taken (taken voorafgaande aan het onderwijzen, bijvoorbeeld het stellen van doelen), interactieve taken (taken gedurende het onderwijzen, bijvoorbeeld instrueren en coachen), en post-actieve taken (taken na het onderwijzen, bijvoorbeeld beoordelen en reflecteren op eigen gedrag, Reynolds, 1992; Van der Schaaf, 2005).

In een eerder onderzoek bestudeerden we de taken die docenten zouden moeten uitvoeren om onderzoeksvaardigheden bij leerlingen te helpen ontwikkelen en de bijbe-

horende docentcompetenties, benodigd om deze taken uit te voeren. We ontwikkelden beoordelingscriteria, samengesteld uit indicatoren, die beschrijven waaraan docenten bij het ontwikkelen van onderzoeksvaardigheden bij leerlingen in de Tweede Fase zouden moeten voldoen. Met behulp van een literatuurstudie en een Delphistudie met 21 experts (Van der Schaaf, 2005) zijn acht beoordelingscriteria ontwikkeld:

#### A. Pre-actieve fase

- 1) Hanteren van lange termijn doelen voor het aanleren van onderzoeksvaardigheden door leerlingen (DOEL);
- 2) Selecteren of maken van een authentieke onderzoekopdracht gericht op vakinhoudelijke en algemene vaardigheden, die aansluit bij de voorkennis van de leerlingen en hen voldoende keuzemogelijkheden biedt (OPDR);
- 3) Voorbereiden en organiseren van het werken aan de opdracht door de leerlingen (ORGA);

#### B. Interactieve fase

- 4) Vooraf nadenken over instructie en begeleiding die aansluit bij de voorkennis van de leerlingen en passen bij het leerdoel, het onderzoeksonderwerp en de beoordeling (DENK);
- 5) Gebruiken van instructie en begeleiding die zelfstandig onderzoek bevordert (INSTRU);
- 6) Creëren van een veilige en stimulerende leeromgeving (KLIM);

#### C. Postactieve fase

- 7) Vaststellen van beoordelingsdoelen, gebruiken van heldere beoordelingscriteria en het daaraan verbinden van de juiste consequenties (BEO), en
- 8) Reflecteren op het onderwijsprogramma en op het eigen handelen (REFL).

## 4.2 Portfolio-inhoud

De bedoeling van het portfolio was dat docenten elk in een paar maanden tijd materiaal verzamelden over hoe ze onderzoeksvaardigheden van leerlingen instrueren, begeleiden en beoordelen. De docenten werden vooraf geïnstrueerd over de inhoud van het samen te stellen portfolio. De portfolio's bestonden uit minimaal zeven elementen (zie Figuur 2), die samen een aantal typen bewijs bevatten: pro-

ducties, documenten die speciaal zijn ontwikkeld voor het portfolio; een reproductie, een registratie van het dagelijks werk van een docent, gereproduceerd voor het portfolio; en verschillende artefacten, reguliere producten van het dagelijks werk van de docent. De inhoud van het portfolio was gericht op onderdelen van de TPB. Het portfolio bevatte informatie over: cognities ten aanzien van bepaald gedrag van docenten bij het onderwijzen van onderzoeksvaardigheden, bijvoorbeeld gerelateerd aan bepaalde doelen; normatieve cognities, bijvoorbeeld gericht op samenwerking met collega's; controlecognities, bijvoorbeeld de inschatting van beschikbare faciliteiten om leerlingen onderzoek te laten doen, en de inschatting van knelpunten en problemen die leerlingen bij het werken aan een onderzoekopdracht kunnen ervaren; gedrag, bijvoorbeeld werkvormen om leerlingen te instrueren en te begeleiden bij het ontwikkelen van onderzoeksvaardigheden. Daarnaast werden contextgegevens verzameld, zoals het aantal jaren onderwijservaring, het vak, de klas, en de gevolgde scholing in het onderwijzen van onderzoeksvaardigheden.

Het portfolio bestond uit de volgende elementen:

- 1) een zelfbeschrijving van de ervaring van de docent en zijn of haar visie op het ontwikkelen van onderzoeksvaardigheden;
- 2) een serie onderzoekopdrachten die de docent de leerlingen in opeenvolgende leerjaren in de Tweede Fase geeft;
- 3) een concrete onderzoekopdracht, inclusief de leerdoelen van de opdracht en de motieven van de docent voor de inhoud en vorm van de opdracht;
- 4) de resultaten van twee interviews over de praktijkkennis van de docent en zijn of haar bedoelingen met het geven van instructie en coaching bij onderzoeksvaardigheden van leerlingen;
- 5) twee video-opnames van lessen waarin de docent leerlingen instrueert en coacht bij het doen van onderzoek;
- 6) beoordelingen van het werk van leerlingen, en
- 7) reflecties op eigen zwakke en sterke kanten, en op hoe het onderwijs kan worden verbeterd.

Fase en beoordelingscriterium	Element	Type bewijs	Inhoud
<i>Pre-actief</i>			
1) DOEL	1. Zelf beschrijving	productie	Achtergrond van de docent, context, cognities t.a.v. gedrag
2) OPDR	2. Serie opdrachten	artefact en productie	Docentgedrag, normatieve cognities, controle-cognities
3) ORGA	3. Onderzoeksopdracht met onderbouwing	artefact en productie	Docentgedrag, intenties, controle-cognities
<i>Interactief</i>			
4) DENK	4a. Interview voor de les	productie	Normatieve cognities, controle-cognities
5) INSTRU	5. Video-opnames	reproductie	Docentgedrag
6) KLIM	4b. Interview na de les	productie	Controle-cognities
<i>Post-actief</i>			
7) BEO	6. Beoordeling	artefact	Docentgedrag
8) REFLECT	7. Reflecties	productie	Normatieve cognities, controle-cognities

Figuur 2. Beoordelingscriteria, elementen, typen bewijs en inhoud van docentportfolio's.

In het onderzoek was het behulpzaam te werken met een matrix waarin de criteria werden afgezet tegen de portfolio-elementen. De meeste criteria worden beoordeeld op basis van meerdere portfolio-elementen. In Figuur 2 wordt dit zichtbaar gemaakt.

## 5 Methode

### 5.1 Deelnemers

#### *Selectie van docenten*

Uit een aselechte steekproef van 115 scholen voor voortgezet onderwijs in Nederland benaderden we de vaksectieleders van de vakken aardrijkskunde, economie en geschiedenis met informatie over het onderzoek en een verzoek om deelname. Vanwege het veeleisende karakter van het onderzoek in termen van vereiste motivatie, tijd en ervaring in het onderwijzen van onderzoeksvaardigheden is noodzakelijkerwijs gewerkt met een kleine gemotiveerde groep met kennis van zaken. Eenentwintig docenten van 21 scholen waren bereid tot en geschikt voor deelname.

Drie docenten namen deel aan een voorstudie, en 18 docenten namen deel aan de hoofdstudie waarover we in dit artikel rapporteren (3 aardrijkskundedocenten, 6 economiedocenten, en 9 geschiedenisdocenten). De gemiddelde leeftijd van de docenten was 44 jaar en ze hadden gemiddeld 15 jaar onderwijservaring.

#### *Selectie van leerlingen*

Elke deelnemende docent koos zelf een klas uit waarin de leerlingen vragenlijsten invulden over de begeleiding door de docent bij een onderzoeksopdracht. In totaal waren 317 leerlingen bij het onderzoek betrokken (gemiddeld 18,6 leerling per docent). Het aantal leerlingen varieerde per docent. Deze variatie was afhankelijk van het aantal leerlingen dat het vak van de betreffende docent had gekozen als eindexamenvak.

#### *Selectie van beoordelaars*

We selecteerden beoordelaars met een onderwijsachtergrond, omdat deze vaak minder moeite hebben met het beoordelen van cognities en gedrag van docenten dan beoordelaars zonder onderwijsachtergrond (Pula & Huot, 1993). Verder kozen we voor externe beoordelaars die de te beoordelen docenten niet kenden, om zo bevoordeling te voorkomen. De volgende beoordelaars namen deel aan het onderzoek: twee ervaren aardrijkskundeleraars, tevens nascholers; een ervaren aardrijkskundedocent, tevens docent geschiedenis; twee ervaren geschiedenisdocenten; een ervaren docent in de gammavakken, tevens lerarenopleider economie; en een schoolleider, tevens ervaren docent. Geen van de beoordelaars had eerder docentportfolio's beoordeeld. Alle beoordelaars namen eerder deel aan onderzoeken waarin de beoordelingscriteria, standaarden en procedures voor het beoordelen van docentportfolio's in



dit onderzoek zijn ontwikkeld (Van der Schaaf, 2005). De beoordelaars konden zich, desgevraagd, goed vinden in de criteria, standaarden en beoordelingsprocedure.

## 5.2 Beoordelaarstraining

De beoordelaars werden zorgvuldig getraind in het beoordelen van docentportfolio's. Ten eerste bestudeerden ze een handleiding met een overzicht van de doelen, planning en procedures van het onderzoek en een volledige omschrijving van de beoordelingscriteria, de bijbehorende ankerpunten, en het portfolio-materiaal. Ten tweede bestudeerden ze individueel een voorbeeldportfolio. Ten derde namen ze daarna deel aan een trainingssessie (vier uur plenair) waarin ze werden geïnstrueerd in de te volgen beoordelingsprocedure en de bijbehorende beoordelingsformulieren (zie Bijlage 2 voor een voorbeeld van een beoordelingsformulier betreffende een criterium). De beoordelingsformulieren bestonden uit: een omschrijving van de criteria met bijbehorende scoring, de portfolio-onderdelen waar de beoordelaars zich per beoordelingscriterium op zouden moeten baseren, een indicatie voor de weging van de betreffende portfolio-onderdelen per criterium en instructies voor de uitvoering van de beoordeling. Elke beoordelaar oefende individueel met het uitvoeren van beoordelingen. Ten vierde beoordeelden de beoordelaars na de training individueel drie portfolio's. Dit waren de portfolio's van de drie docenten die deelnamen aan deze pilot. Deze portfolio's waren willekeurig geselecteerd per vak (aardrijkskunde, economie, geschiedenis). Deze beoordelingsronde fungeerde als voorstudie voor de beoordelingsprocedure in het onderzoek en was mede bedoeld om de beoordelaars feedback te kunnen geven op hun beoordelingen. Ten vijfde kregen de beoordelaars feedback op hun scoring en ontvingen ze suggesties voor verbetering van hun beoordelingen. De suggesties impliceerden het accurate gebruik van beoordelingscriteria en de beoordelingsprocedure en aandachtspunten met betrekking tot de interpretatie. We hebben alle beoordelaars gevraagd naar in hoeverre ze de beoordelingsprocedure bruikbaar vonden. Dat was in sterke mate het geval.

## 5.3 Dataverzameling

*Geconstrueerde docentcognities en -gedrag*  
Omdat de 18 docenten in ons onderzoek geen ervaring hadden met het samenstellen van een portfolio structureerden we de gewenste inhoud van het portfolio, terwijl de docenten vrij waren in de vorm van het portfolio materiaal (bijvoorbeeld zelf geschreven of in een gesprek verwoord en daarna uitgetypt) en ook vrij in het toevoegen van materiaal. De docenten werden geïnstrueerd over de doelen van het onderzoek, de beoordelingscriteria en de beoordelingsprocedure. De docenten kregen naar behoefte ondersteuning in het samenstellen van hun portfolio, bijvoorbeeld bij het maken van video-opnames en bij het uittypen van gesprekken. Elk van de 18 portfolio's werd in een paar maanden tijd samengesteld. Alle docenten gaven aan dat hun portfolio hun cognities en gedrag representeerde, dus we veronderstellen dat de portfolio's redelijk authentiek zijn. Na de beoordeling ontvingen de docenten geschreven feedback van de beoordelaars op hun sterke en minder sterke kanten in relatie tot de beoordelingscriteria. Om de relevantie van de feedback te evalueren vulden de docenten vervolgens een vragenlijst in met een aantal vragen met antwoordmogelijkheden op een 5-puntsschaal (van 1 = *sterk mee oneens* tot 5 = *sterk mee eens*).

### *Beoordeelde docentcognities en -gedrag*

In aansluiting op de beoordelingscriteria ontwikkelden we voor *beoordelingen door leerlingen* een vragenlijst voor leerlingen met 15 uitspraken over het gedrag van hun docent tijdens de gegeven instructie en begeleiding bij een onderzoeksopdracht, waarmee zij hun positieve of minder positieve beoordeling van dit gedrag tot uitdrukking konden brengen (zie Bijlage 1). De leerlingen beantwoordden de vragen op een 4-puntsschaal (van 1 = *geheel niet overeenkomstig mijn leeraar* tot 4 = *sterk overeenkomstig mijn leeraar*). We probeerden de vragenlijst met succes uit in een voorstudie met 45 leerlingen van 3 docenten (Cronbach's  $\alpha = 0,84$ ).

Wat betreft de *beoordelingen door beoordelaars* werden de 18 docentportfolio's elk beoordeeld door twee beoordelaars. Dit



deden ze onafhankelijk van elkaar. De samenstelling van de beoordelaarsparen gebeurde op basis van vakexpertise. De meeste beoordelaars hadden expertise in meer vakken, en de samenstelling van de beoordelaarsparen kon daardoor mede worden afgestemd op de beschikbare tijd bij het binnenkomen van nieuwe portfolio's. De 18 portfolio's werden in negen maanden tijd beoordeeld in volgorde van beschikbaar komen. De beoordelaars gebruikten bij het beoordelen van de portfolio's beoordelingsformulieren, waarin ze elk beoordelingscriterium illustreerden met verwijzingen naar bepaalde elementen van het portfolio. Daarna beschreven ze hun interpretaties van deze elementen. Vervolgens gaven ze een score op elk van de criteria, op een 3-puntsschaal met ankerpunten (analytische beoordeling). Ten slotte gaven ze per portfolio een 'overall' holistische beoordeling, ook weer op een 3-puntsschaal. De beoordeling resulteerde dus in twee typen scores: een score gebaseerd op een analytische beoordeling per beoordelingscriterium en een score gebaseerd op een holistische beoordeling over de acht beoordelingscriteria samen. De beoordelaars zijn gevraagd in hoeverre ze de beoordelingsprocedure zoals getraind opvolgden. Dat was bij alle beoordelaars in sterke mate het geval.

#### 5.4 Data analyse

##### *Geconstrueerde docentcognities en -gedrag*

Het interactieve gedrag van docenten werd gecodeerd op basis van video-opnames van twee lessen per docent. De docenten namen twee instructie- of begeleidingslessen op van maximaal 30 minuten elk. De video-opnames werden vervolgens volledig uitgeschreven. We deelden elke opname op in segmenten per onderzoeksdeelvaardigheid of onderzoeksactiviteit zoals beschreven in paragraaf 4.1. De onderzoeker (eerste auteur) en een onderzoeksassistent analyseerden de video-opnames op inhoud (de negen onderzoeksdeelvaardigheden), de mate van sturing in de leeromgeving (docentgestuurd, gedeelde sturing, leerlinggestuurd), de context (bijvoorbeeld klas of mediatheek), en het leerjaar van de leerlingen. Per segment per onderzoeksdeelvaardigheid telden we het aantal malen

dat aan de betreffende onderzoeksdeelvaardigheid aandacht werd besteed en we berekenden ook de frequenties van de vormen van interactie tussen docenten en leerlingen (docentgestuurd, gedeelde sturing, leerlingsturing).

Wat betreft de beoordeling van de mate van sturing in de leeromgeving richten we ons hier alleen op de sturing door middel van het docentgedrag en niet op de onderzoeksopdracht of andere middelen die de mate van sturing in de leeromgeving mede kunnen bepalen. De beoordelingen werden door de onderzoekers uitgevoerd op een schaal van 1 (*docentgestuurd*) tot 3 (*leerlinggestuurd*) en waren gebaseerd op een gedetailleerde beschrijving van het docentgedrag gerelateerd aan de mate van sturing in een leeromgeving (Van der Schaaf, 2000; Vermunt & Verloop, 1999). Hoge scores werden gegeven wanneer docenten leerlingen mogelijkheden boden om zelfstandig te werken (bijvoorbeeld hun keuzemogelijkheden aanbieden en hen niet onnodig begeleiden). Lage scores werden gegeven wanneer de docent de onderzoeksactiviteiten van de leerlingen sterk controleerde of voordeed.

Gegevens met betrekking tot contextfactoren, cognities over docentgedrag bij het onderwijzen van onderzoeksvaardigheden, normatieve cognities, en controlecognities, werden kwalitatief geanalyseerd en van de resulterende categorieën werden frequenties en gemiddelden berekend.

##### *Beoordeelde docentcognities en -gedrag*

Wat betreft de *beoordelingen door leerlingen* analyseerden we de schaalbaarheid van de vragen in de vragenlijst voor de leerlingen door het berekenen van Cronbach's  $\alpha$ . We berekenden frequenties en analyseerden de verschillen tussen de docenten met enkelvoudige variantie-analyse op de via de vragenlijst door de leerlingen gegeven beoordelingen van het gedrag van hun docenten.

Voor de *beoordelingen door beoordelaars* gebruikten we het percentage volledige overeenstemming in de holistische beoordelingen (over alle beoordelingscriteria samen) en de Cronbach's  $\alpha$  (jury  $\alpha$ ) over de analytische beoordelingen (op de afzonderlijke beoordelingscriteria) per portfolio binnen de beoor-

delaarsparen om de interbeoordelaarsovereenstemming respectievelijk de betrouwbaarheid te berekenen.

#### *De relatie tussen docentcognities en gedrag*

Voor het analyseren van de relatie tussen cognities en gedrag van de docenten zoals volgens de onderzoekers in de portfolio's weergegeven en zoals beoordeeld door de externe beoordelaars en de leerlingen, gebruikten we meerniveau-analyse, aangezien sprake was van geneste data (leerlingen binnen docenten). We gebruikten de volgende sets predictoren: 1) de docentcognities in termen van doelen om onderzoeksvaardigheden bij leerlingen te ontwikkelen; 2) de onderzoeksdeelvaardigheden waar de docent in zijn of haar gedrag aandacht aan besteedt en de mate waarin de docent sturend en structurend is bij het instrueren en begeleiden van de onderzoeksactiviteiten van leerlingen (op een 3-puntsschaal); 3) de beoordelingen door de leerlingen van het gedrag van hun docent (zoals gegeven in de vragenlijst voor leerlingen op een 4-puntsschaal). De criteriumvariabelen waren de gemiddelde analytische scores van de beoordelaars op een 3-puntsschaal.

Om de structuur van de gegeven beoordelingen nader te onderzoeken, zijn we nagegaan of docenten konden worden gegroepeerd op basis van de aan hen door de beoordelaars toegekende holistische beoordelingen. Daartoe voerden we een exploratieve Q-principale componentenanalyse (Q-PCA) uit op de beoordelingen van de portfolio's waarbij de interbeoordelaarsbetrouwbaarheid voldoende was gebleken. Q-PCA is een variant van factoranalyse waarbij de datamatrix wordt gespiegeld en de respondenten worden behandeld als variabelen die laden op componenten.

Vervolgens gingen we na of de resulterende groepen docenten ook op de afzonderlijke beoordelingscriteria verschilden. Daartoe berekenden we per docent per criterium het gemiddelde van de beoordelingen van beide beoordelaars en toetsten we of de gemiddelden van de gevonden groepen per criterium significant van elkaar verschilden. Aangezien de Q-PCA tot een indeling in twee groepen leidde, gebruikten we de t-test

voor onafhankelijke groepen. Eenzelfde analyse voerden we ook uit op de beoordelingen van de leerlingen van het gedrag van hun docent.

De antwoorden van de docenten op de vragenlijst over de relevantie van de door de beoordelaars gegeven feedback zijn beschrijvend geanalyseerd.

## 6 Resultaten

### 6.1 Geconstrueerde docentcognities en -gedrag

#### *Docentcognities*

Uit de analyses van de portfolio's door de onderzoekers blijkt dat de docenten over het algemeen de volgende doelen nastreven: leerlingen deelvaardigheden van onderzoek aanleren ( $n = 16$ ) en leerlingen laten oefenen met zelfstandig leren ( $n = 13$ ). Eenderde van de docenten beoogde leerlingen kennis bij te brengen over het vakgebied ( $n = 5$ ).

Met betrekking tot de normatieve cognities van de docenten, zegt slechts de helft duidelijke afspraken met collega's te maken over onderzoeksopdrachten. De afspraken die gemaakt worden, hebben voornamelijk betrekking op de inhoud en tijdsplanning (roostering) van het onderzoek.

Met betrekking tot controlecognities, biedt hun school volgens alle docenten voldoende faciliteiten om onderzoeksvaardigheden bij leerlingen te helpen ontwikkelen. Volgens 16 van de 18 docenten ervaren leerlingen enkele knelpunten bij het werken aan onderzoeksopdrachten. De belangrijkste hebben betrekking op de haalbaarheid van de opdrachten in de tijd ( $n = 9$ ) en de dataverzameling ( $n = 6$ ). Negen docenten geven suggesties om deze problemen op te lossen. Daarbij werd het geven van meer coaching en formatieve feedback het meest genoemd ( $n = 5$ ). Twee docenten verklaarden dat ze geen interventie door de docent prefererden omdat: "Het doen van onderzoek het eigen leerproces van de leerling is". De helft van de docent had een een- of tweedaagse training gevolgd in het helpen ontwikkelen van onderzoeksvaardigheden bij leerlingen.

### *Docentgedrag*

De twee videoregistraties per docent ( $n = 36$ ) hadden betrekking op (plenaire) klassikale begeleiding van leerlingen ( $n = 11$ ), begeleiding van groepjes leerlingen ( $n = 11$ ), of individuele begeleiding ( $n = 3$ ). Elf opnames bevatten instructie-activiteiten, voornamelijk het uitleggen van de onderzoeksopdracht. Vijf van de opnames vonden plaats in de mediatheek. De andere opnames vonden plaats in klaslokalen.

Het instrueren en begeleiden door docenten had voornamelijk betrekking op de volgende onderzoeksdeelvaardigheden: formuleren van een probleem, formuleren van een onderzoeksvraag; dataverzameling (in tweederde van de video-opnames) en rapporteren en presenteren van onderzoek (in de helft van de opnames). In eenderde van de opnames werd aandacht besteed aan de data-analyse en in slechts twee opnames besteedden docenten (ook) aandacht aan het trekken van conclusies. In de video-opnames werd geen aandacht besteed aan de onderzoeksdeelvaardigheid evalueren van onderzoek. Dit resultaat is mogelijk veroorzaakt door de geselecteerde lessen; de meeste docenten kozen voor de opname van lessen en begeleidingssessies in de eerste helft van de uitvoering van de onderzoeksopdracht die centraal stond. De leeromgevingen die de docenten creëerden, verschilden significant in hun mate van sturing (range 1,30 – 2,93,  $SD = 0,70$ ) ( $F = 3,11$ ;  $df = 35$ ;  $p = 0,019$ ). Gemiddeld genomen waren de leeromgevingen sterk docentgestuurd. In de meeste opnames stond de docent centraal en in totaal waren ze 76% van de geregistreerde tijd aan het woord. De verschillen tussen de docenten op dit punt waren echter groot (range 43%-95%;  $SD = 0,13$ ) ( $F = 7,79$ ;  $df = 35$ ;  $p \leq 0,001$ ).

In de interviews na de video-opnames bevestigden 17 van de 18 docenten dat ze hun voorgenomen doelen met de les of begeleidingssessie hadden bereikt. Ze waren over het algemeen tevreden met de opgenomen lessen en begeleiding. Tien docenten gaven aan deze constatering te baseren op de waargenomen houding, opmerkingen en vragen van de leerlingen. Geen enkele docent gaf aan expliciet te controleren of de leerlingen de les of begeleiding hadden begrepen.

### **6.2 Beoordeelde docentcognities en -gedrag**

#### *Leerlingbeoordelingen*

De items in de leerlingvragenlijst vormden een betrouwbare schaal (Cronbach's  $\alpha = 0,81$ ;  $M = 3,13$ ;  $SD = 0,89$ ; zie Bijlage 1). Gemiddeld scoorden de docenten het hoogste op de items "Mijn leraar is enthousiast" ( $M = 3,43$ ;  $SD = 0,77$ ) en "Mijn leraar is bereid iets voor een tweede maal uit te leggen" ( $M = 3,38$ ;  $SD = 0,79$ ). De docenten kregen over het algemeen de laagste scores op de items "Mijn leraar gaat na of we de opdracht begrijpen" ( $M = 2,74$ ;  $SD = 0,96$ ) en "Mijn leraar gaat na of we wel aan de onderzoeksopdracht werken" ( $M = 2,88$ ;  $SD = 0,95$ ). De resultaten van de vragenlijst bevestigden de resultaten van de interviews na de video-opname. Daaruit bleek dat docenten nauwelijks de onderzoeksactiviteiten van de leerlingen op expliciete wijze monitoren. Een enkelvoudige variantie-analyse toont op dit punt wel significante verschillen tussen docenten ( $F = 10,19$ ;  $df = 17$ ;  $p \leq 0,001$ ).

#### *Beoordelingen door beoordelaars*

De jury  $\alpha$ 's voor de afzonderlijke gemiddelde analytische scores waren redelijk voor 12 beoordelaarsparen (range van 0,39 tot 0,76). De jury  $\alpha$ 's waren laag of zelfs negatief voor zes paren (range van -0,80 (in één geval) tot 0,22). De beoordelaarsparen gaven in 35% van de beoordelingen exact dezelfde holistische scores op de gebruikte 5-puntsschaal. Een verschil van een half punt trad op in 12% van de beoordelingen, een verschil van één punt in 47% van de beoordelingen, en een verschil van 1,5 punt in 6% van de beoordelingen (één beoordelaarspaar). De beoordelingscriteria vormden samen een betrouwbare schaal (Cronbach's  $\alpha = 0,76$ ).

Alle docenten in de voorstudie ( $n = 3$ ) en in de hoofdstudie ( $n = 18$ ) retournerden de vragenlijst over de relevantie van de feedback van de beoordelaars. De items in de vragenlijst vormden drie betrouwbare schalen. De eerste schaal bevat vragen over de relevantie van de geschreven feedback (4 items,  $n = 19$ ), bijvoorbeeld "Ik herken mijn handelen in de gegeven feedback" (Cronbach's  $\alpha = 0,84$ ;  $M = 3,68$ ;  $SD = 0,19$ ). De tweede schaal betreft

de transparantie van de beoordeling (3 items,  $n = 20$ ), bijvoorbeeld “Het is helder op welke beoordelingscriteria de beoordeling van mijn portfolio is gebaseerd” (Cronbach’s  $\alpha = 0,84$ ;  $M = 4,00$ ;  $SD = 0,78$ ). De derde schaal betreft de mate waarin de docent door de feedback werd gemotiveerd tot verbetering van het eigen handelen (3 items,  $n = 21$ ), bijvoorbeeld “De gegeven feedback geeft informatie om mijn handelen te verbeteren” (Cronbach’s  $\alpha = 0,81$ ;  $M = 3,80$ ;  $SD = 0,98$ ). De resultaten laten zien dat docenten tamelijk tevreden zijn over de gegeven feedback en de beoordelingen door de beoordelaars tamelijk relevant vinden.

### 6.3 De relatie tussen docentcognities en -gedrag

Analyse door de onderzoekers van de inhoud van de 18 docentportfolio’s toont, noch per portfolio, noch op geaggregeerd niveau, duidelijke relaties tussen de cognities van de docenten en hun gedrag zoals weergegeven in de portfolio’s. Dit betekent dat de beoordelingen door de externe beoordelaars niet significant kunnen worden voorspeld vanuit de cognities en het gedrag van de docenten zoals in de analyse door de onderzoekers aangetroffen in de portfolio’s. Wel konden de beoordelingen door de leerlingen van het gedrag van hun docent significant worden voorspeld vanuit de gemiddelde analytische beoordelingen door de externe beoordelaars (de regressiecoëfficiënt van de beoordeling door de leerlingen van het gedrag van hun docent op de beoordeling door de externe beoordelaars in de meerniveau-analyse was 2,52 bij een standaardfout van 0,32).

Om na te gaan of de docenten zijn te groeperen op basis van de aan hen toegekende beoordelingen werd een Q-principale componentenanalyse uitgevoerd met betrekking tot de 12 portfolio’s die voldoende betrouwbaar zijn beoordeeld. We gebruikten 0,60 als een minimum voor een significante lading. Dit resulteerde in twee groepen van docenten (zie Tabel 1). De componenten verklaarden 34% (component 1) en 22% (component 2) van de variantie. Met enige voorzichtigheid kunnen we dus twee groepen onderscheiden.

Om na te gaan op welke afzonderlijke beoordelingscriteria de twee groepen van elkaar

Tabel 1

Resultaten van een exploratieve Q-principale componentenanalyse ( $n = 12$ )

Docent	Component 1	Component 2
1	-0,02	-0,23
2	0,01	<b>0,62</b>
3	-0,45	0,28
4	<b>0,61</b>	0,51
5	<b>0,93</b>	0,04
6	<b>0,75</b>	-0,02
7	<b>0,81</b>	-0,13
8	-0,07	<b>0,64</b>
9	-0,16	<b>0,92</b>
10	0,11	<b>0,65</b>
11	<b>0,93</b>	0,04
12	<b>0,71</b>	0,46

Noot. Ladingen hoger dan 0,60 zijn vet gedrukt.

verschillen zijn t-toetsen voor onafhankelijke steekproeven uitgevoerd. Daaruit bleek dat de twee groepen verschillen in de beoordelingen van de externe beoordelaars op het criterium Vooraf nadenken over instructie en begeleiding die aansluit bij de voorkennis van de leerlingen en passen bij het leerdoel, het onderzoeksonderwerp en de beoordeling (DENK). De docenten in de eerste groep (docent 4, 5, 6, 7, 11, 12) krijgen een hogere score op dit criterium ( $M = 0,13$  op een 3-puntsschaal;  $SD = 0,22$ ) dan de docenten in de tweede groep (docent 2, 8, 9, 10) ( $M = 1,55$ ;  $SD = 0,14$ ;  $t = 2,78$ ;  $df = 9$ ;  $p = 0,02$ ). De twee groepen verschillen niet significant in de overall-beoordelingen door de beoordelaars zoals die tot uitdrukking komen in de gemiddelde analytische scores en de holistische scores.

Eenzelfde t-toets is uitgevoerd op de beoordeling door de leerlingen in de door hen ingevulde vragenlijst. Ook ten aanzien van deze beoordeling krijgen de docenten in de eerste groep gemiddeld een significant hogere score ( $n = 102$ ;  $M = 3,21$ ;  $SD = 0,49$ ) dan de docenten in de tweede groep ( $n = 109$ ;  $M = 2,93$ ;  $SD = 0,53$ ;  $t = 4,05$ ;  $df = 208$ ;  $p \leq 0,001$ ).

## 7 Conclusie en discussie

Algemeen wordt aangenomen dat docentportfolio’s vanwege hun authenticiteit valide beoordelingsinstrumenten zijn. Een voorwaarde voor een valide beoordeling is dat de

beoordelingen die aan een portfolio worden toegekend daadwerkelijk zijn gebaseerd op de gegevens zoals door de docenten in hun portfolio zijn weergegeven. Verschillende kenmerken van de data in een portfolio (uniek, situationeel, persoonlijk, beschrijvend) maken echter dat het beoordelen van portfolio's veel interpretatie vergt en dat validering van portfolio-beoordeling moeilijk via traditionele statistische verificatiemethoden kan verlopen. Verder bestaat ondanks de inmiddels onderkende relatie tussen docent-cognities en -gedrag veel onduidelijkheid over de aanwezigheid van deze relatie bij docentportfolio-beoordelingen. Inzicht in de vraag of en, zo ja, hoe deze relatie kan worden aangetoond is van belang om meer grip te krijgen op validering van docentportfolio-beoordelingen.

In dit artikel stonden de volgende twee vragen centraal:

- 1) Hoe kunnen in portfolio's weergegeven docentcognities worden beoordeeld?
- 2) Wat is de relatie tussen cognities zoals weergegeven door docenten in hun portfolio en gedrag zoals beoordeeld door leerlingen en beoordelaars?

In het onderzoek stelden 18 ervaren docenten (aardrijkskunde, economie en geschiedenis in de Tweede Fase) portfolio's samen. De portfolio's hadden betrekking op een relatief nieuw onderdeel van de Tweede Fase, namelijk het instrueren, begeleiden en beoordelen van onderzoeksopdrachten van leerlingen door docenten. Dit vergt van docenten nieuwe cognities over leren, gericht op het zelf actief construeren van kennis, en het verlangt van docenten dat ze leerlingen instrueren, begeleiden en beoordelen bij opdrachten met een open karakter waaraan ze zelfstandig en in samenwerking werken.

De portfolio's werden door zes beoordelaars beoordeeld. De beoordelaars hadden eerder deelgenomen aan onderzoeken waarin de criteria, standaarden en de beoordelingsprocedure waren ontwikkeld en bleken zich daarin, desgevraagd, dan ook goed te kunnen vinden. De beoordelaars werden uitgebreid getraind in hun beoordelingstaak. Daarbij bestudeerden ze onder meer een beoordelaars-handleiding en beoordelingsformulieren, oefenden ze met het uitvoeren van beoorde-

lingen en kregen ze feedback op hun beoordelingen. Vervolgens beoordeelden zes beoordelaars paarsgewijs de 18 portfolio's op acht eerder vastgestelde beoordelingscriteria. De beoordelaars gaven desgevraagd aan dat ze de beoordelingsprocedure zoals geoefend in de training goed werkbaar vonden en stevast toepasten bij hun beoordelingen. De beoordelingsprocedure met bijbehorende training waren dus bruikbaar.

Bij 12 van de 18 portfolio's was de interbeoordelaarsbetrouwbaarheid redelijk en variantie-analyse liet amper een beoordelaars-effect zien. De docentportfolio's in het onderzoek zijn dus redelijk betrouwbaar beoordeeld. Lineaire meerniveau-analyse wees uit dat de beoordelingen door de leerlingen van het gedrag van hun docenten significant konden worden voorspeld vanuit de beoordelingen door de externe beoordelaars. Hiermee hebben we een indicatie verkregen voor de betrouwbaarheid en de validiteit van de portfolio-beoordelingen in ons onderzoek.

Daarnaast zochten we naar verschillen in cognities en gedrag tussen docenten. Daartoe voerden we Q-principale componenten-analyse uit op de gemiddelde beoordelingen van de twee beoordelaars per criterium per portfolio (voor de 12 voldoende betrouwbaar beoordeelde portfolio's). Dit resulteerde in twee groepen docenten. T-toetsen wezen uit dat zes docenten in de eerste groep gemiddeld significant hogere score kregen op het beoordelingscriterium Vooraf nadenken over instructie en begeleiden, dan docenten in de tweede groep. Dat betekent dat volgens de beoordelaars de docenten in de eerste groep gemiddeld beter zijn in het expliciteren van de rationale achter hun handelen dan de tweede groep. De eerste groep docenten werd ook gemiddeld significant positiever beoordeeld door de eigen leerlingen op hun gedrag dan de tweede groep. Dit resultaat duidt erop dat er sprake is van een relatie tussen docent-cognities en docentgedrag. Bij de interpretatie van deze bevinding moet rekening worden gehouden dat het hier gaat om de relatie tussen cognities en gedrag bij een vrij recente onderwijsvernieuwing. De twee groepen docenten verschilden overigens niet significant op de overall-beoordelingen van de portfolio's door de beoordelaars.

Hoewel we in ons onderzoek indicaties hebben verkregen van de betrouwbaarheid en de validiteit van het beoordelen van docent-cognities en docentgedrag op basis van door docenten samengestelde portfolio's is meer bewijs nodig om hierover daadwerkelijk uitspraken te kunnen doen. In ons onderzoek hebben we ons bijvoorbeeld toegespitst op docentcognities zoals expliciet weergegeven in de portfolio's. Het gebruikte model van de Theory of Planned Behavior (Ajzen, 1985; 2002; Ajzen & Fishbein, 2005) en de methode van dataverzameling (een gedocumenteerde neerslag van cognities in portfolio's) richten zich alleen op bewuste cognities en kunnen geen impliciete *tacit* cognitieve processen in kaart brengen en verklaren. Verder waren de deelnemende docenten ervaren en namen ze vrijwillig deel aan het onderzoek. Mede hierdoor kunnen de resultaten niet worden beschouwd als representatief voor de gehele populatie van docenten in de gamavakken of andere vakgebieden.

Voor het ontwikkelen van valide portfolio-beoordelingen is verder onderzoek nodig. Ten eerste kan een valide docentportfolio-beoordeling worden ondersteund door nadere theorievorming. Dergelijke theorievorming zou om te beginnen betrekking moeten hebben op de te meten constructen, dat wil zeggen de conceptualisering van docentcompetenties (inclusief cognities en gedrag) en de ontwikkeling van docentcompetenties in de werkcontext. Onderzoek naar de aard en ontwikkeling van docentcompetenties is dan ook essentieel (vgl. Bakkenes, Vermunt, & Wubbels, 2004; Kwakman, 1999). Ten tweede is nadere theorievorming nodig op het vlak van de cognities van beoordelaars tijdens het beoordelen van portfolio's. Een belangrijk aandachtspunt bij docentportfolio-beoordeling is dat een dergelijke beoordeling gepaard gaat met complexe interacties tussen docentcompetenties, het portfolio, de gebruikte beoordelingscriteria, kenmerken van beoordelaars en interpretaties van de beoordelaars. Als gevolg daarvan hebben portfolio-beoordelingen doorgaans een beperkte generaliseerbaarheid. Voor het gebruik van docentportfolio-beoordelingen zijn generaliseerbaarheidsstudies dan ook essentieel (Straetmans & Sanders, 2001). Deze zouden ten minste

moeten zijn gericht op de facetten beoordelaars en onderwijsituaties. Te denken valt aan onderzoek naar de cognities van beoordelaars tijdens beoordelingsprocessen (Van der Schaaf, Stokking, & Verloop, 2005) en de mate waarin beoordelaars zich bij het beoordelen daadwerkelijk op de beoordelingscriteria baseren (Van der Schaaf & Stokking, 2008).

Ten derde is om portfolio-beoordelingen te verbeteren meer inzicht nodig in de manier waarop en de mate waarin verschillende kenmerken van portfolio-beoordelingen (bijvoorbeeld het werken met een meer of minder gestructureerd portfolio, of met een meer of minder analytisch of holistisch scoringmodel) van invloed zijn op de kwaliteit van zulke beoordelingen. Quasi-experimentele designs waarin wordt gevarieerd op verschillende kenmerken zijn voor dit doel zinvol.

## Noot

- 1 Het onderzoek waarover hier wordt gerapporteerd, is gefinancierd door de Nederlandse Organisatie voor Wetenschappelijk Onderzoek (projectnummer NWO-PROO 411-21-204).

## Literatuur

- Aguirre, J., & Speer, N. M. (2000). Examining the relationship between beliefs and goals in teacher practice. *Journal of Mathematical Behavior*, 18, 327 - 356.
- Ajzen, I. (1985). From intentions to actions: A Theory of Planned Behavior. In J. Kuhl & J. Beckman (Eds.), *Action control: from cognition to behavior* (pp. 11 - 39). Heidelberg, Germany: Springer.
- Ajzen, I. (2002). Perceived behavioral control, self-efficacy, locus of control, and the Theory of Planned Behavior. *Journal of Applied Social Psychology*, 32, 665 - 683.
- Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. Albarracín, B.T. Johnson, & M.P. Zanna (Eds.), *The handbook of attitudes* (pp. 173 - 221). Mahwah, NJ: Erlbaum.
- Bakkenes, I., Vermunt, J., & Wubbels, T. (2004,



- juni). *Leren van docenten in de beroepspraktijk vanuit een theoretisch perspectief*. Paper gepresenteerd op de Onderwijs Research Dagen, Utrecht, Nederland.
- Beijaard, D., & Verloop, N. (1996). Assessing teachers' practical knowledge. *Studies in Educational Evaluation*, 22, 275 - 286.
- Carlston, D. (1992). Impression formation and the modular mind: the associated systems theory. In L.L. Martin & A. Tesser (Eds). *The construction of social judgments* (pp. 301 - 341). Hillsdale, NJ: Erlbaum.
- Carlston, D. (1994). Associated systems theory: A systematic approach to cognitive representations of persons. *Advances in Social Cognition*, 7, 1 - 78.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In M.C. Wittrock (Ed.), *Handbook of research on teaching* (3<sup>rd</sup> ed.) (pp. 255 - 296). New York: Macmillan.
- Crawley, F. E. (1990). Intentions of Science teachers to use investigative teaching methods: A test of the Theory of Planned Behavior. *Journal of Research in Science Teaching*, 27, 685 - 697.
- Eraut, M. E. (1994). *Developing professional knowledge and competence*. London: Falmer Press.
- Fishbein, M., Triandis, H. D., Kanfer, F. H., Becker, M., Middlestadt, S. E., & Eichler, A. (2001). Factors influencing behavior and behavior change. In A. Baum, T.A. Revenson & J.E. Singer (Eds.), *Handbook of health psychology* (pp. 3 - 17). Mahwah, NJ: Lawrence Erlbaum Associates.
- Gonczi, A. (1994). Competency based assessment in the professions in Australia. *Assessment in Education* 1(1), 27-45.
- Haney, J. J., Czerniak, C. M., & Lumpe, A. T. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. *Journal of Research in Science Teaching*, 33, 971 - 993.
- Jones, E. E., & Davis, K. E. (1965). From acts to dispositions: the attribution process in person perception. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, 2 (pp. 219 - 266). New York: Academic Press.
- Kanselaar, G., Galen, F., van, Beemer, H., Erkens, G., & Gravemeijer, K. (1999). *Grafieken leren met de computer*. Utrecht, Nederland: ICO-ISOR, Universiteit Utrecht.
- Kelly, G. A. (1955). *The psychology of personal constructs*. New York: Norton.
- Kwakman, K. (1999). *Leren van docenten tijdens de beroepsloopbaan*. Dissertatie. Katholieke Universiteit van Nijmegen, Nijmegen, Nederland.
- Lievens, F. (2001). Assessor training strategies and their effects on accuracy, interrater reliability, and discriminant validity. *Journal of Applied Psychology*, 86, 255 - 264.
- Mathijssen, I. C. H. (2006). *Denken en handelen van docenten*. Dissertatie. Universiteit Utrecht, Utrecht, Nederland.
- Messick, S. (1989). Validity. In R.L. Linn (Ed.), *Educational measurement* (3<sup>rd</sup> ed.) (pp. 13 - 103). New York: Macmillan.
- Pula, J. J., & Huot, B. A. (1993). A model of background influences on holistic raters. In M.M. Williamson & B.A. Huot (Eds.), *Validating holistic scoring for writing assessment. Theoretical and empirical foundations* (pp. 237 - 265). Cresshill, NJ: Hampton Press.
- Reynolds, A. (1992). Getting to the core of the apple: a theoretical view of the knowledge base of teaching. *Journal of Personnel Evaluation in Education*, 6, 41 - 55.
- Roth, W. M., & Roychoudhury, A. (1993). The development of science process skills in authentic contexts. *Journal of Research in Science Teaching*, 30, 127 - 152.
- Saxe, G. B. (1991). *Culture and cognitive development: Studies in mathematical understandings*. Hillsdale, NJ: Lawrence Erlbaum.
- Schaaf, M. F. van der. (2000). *Vaardig in het begeleiden van zelfstandig leren? Monitorings-instrument*. Utrecht, Nederland: Afdeling Onderwijskunde, Universiteit Utrecht.
- Schaaf, M. F. van der. (2005). *Construct validation of teacher portfolio assessment. Procedures for improving teacher competence assessment illustrated by teaching students research skills*. Dissertatie. Universiteit Utrecht, Utrecht, Nederland.
- Schaaf, M. F., van der, & Stokking, K. M. (2008). Developing and validating a design for teacher portfolio assessment. *Assessment & Evaluation in Higher Education*, 33, 245-260.
- Schaaf, M. F., van der, Stokking, K. M., & Verloop, N. (2005). Cognitive representations in raters' assessment of teacher portfolio's. *Studies in Educational Evaluation*, 31, 27 - 55.

- Schoenfeld, A. H. (1998). Toward a theory of teaching-in-context. *Issues in Education*, 4(1), 1 - 94.
- Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: a contemporary perspective. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3<sup>rd</sup> ed.) (pp. 3 - 36). New York: Macmillan.
- Stokking, K. M., & Schaaf, M. F. van der. (2000). *Ontwikkeling en beoordeling van onderzoeksvaardigheden*. Utrecht, Nederland: Afdeling Onderwijskunde, Universiteit Utrecht.
- Stokking, K. M., Schaaf, M. F., van der, Jaspers, J., & Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal*, 30, 93-116.
- Straetmans, J. J. M., & Sanders, P. F. (2001). *Beoordelen van competenties van docenten*. Utrecht, Nederland: Programmamanagement Educatief Partnerschap.
- Verloop, N. (2003). De leraar. In N. Verloop & J. Lowyck (Eds.), *Onderwijskunde: een kennisbasis voor professionals* (pp. 194 - 248). Groningen, Nederland: Wolters-Noordhoff.
- Vermunt, J. D., & Verloop, N. (1999). Congruence and friction between learning and teaching. *Learning and Instruction*, 9, 257 - 280.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189 - 205.

Manuscript aanvaard: 28 maart 2008

## Auteurs

**Marieke van der Schaaf** is universitair docent bij het departement Pedagogische en Onderwijskundige Wetenschappen van de Universiteit Utrecht.

**Karel Stokking** is hoogleraar onderwijskunde bij het departement Pedagogische en Onderwijskundige Wetenschappen van de Universiteit Utrecht.

**Nico Verloop** is hoogleraar onderwijskunde en hoogleraar-directeur van het Interfacultair Centrum voor Lerarenopleiding, Onderwijsontwikkeling en Nascholing (ICLON), Universiteit Leiden.

*Correspondentieadres:* Marieke van der Schaaf, departement Pedagogische en Onderwijskundige Wetenschappen, Universiteit Utrecht, Postbus 80140, 3508 TC Utrecht, e-mail: *m.f.vanderschaaf@uu.nl*

## Abstract

### **The validity of the assessment of teacher beliefs and teacher behaviour in portfolio assessment**

This study focuses on the relation between teacher beliefs and teacher behaviour in teacher portfolio assessment. Eighteen experienced teachers developed portfolios about their cognitions and behaviour in instructing, coaching and assessing students' research skills in Social Sciences. We qualitatively analysed the beliefs and behaviour of 18 teachers as described in their portfolios. In addition, each portfolio was independently assessed by two trained raters on eight content standards (assessment criteria) and the teachers' classroom behaviour was assessed by their own students in a questionnaire ( $n = 317$ ). Linear multilevel analysis showed that the students' assessments of their teachers' behaviour could be significantly predicted by the raters' assessments of the teachers' beliefs and behaviour as described in their portfolios. Teachers with high raters' assessments on the content standard THINK (the deliberate choice of teaching strategies that meet students' abilities) had significantly higher student assessments than teachers who were judged low on this content standard. Implications of the results are discussed and suggestions for further research are given.

## Bijlage 1

### Psychometrisch rapport voor de schaal in de leerlingvragenlijst

De leerlingen gaven antwoord op een 4-puntsschaal:

- 1 = *geheel niet overeenkomstig mijn leraar*;  
 2 = *komt een beetje overeen met mijn leraar*;  
 3 = *komt redelijk overeen met mijn leraar*;  
 4 = *sterk overeenkomstig mijn leraar*.

Items	<i>M</i>	<i>SD</i>	<i>Rir</i>
1. Het doel van de onderzoeksopdracht is helder	3,32	0,77	0,46
2. We hebben voldoende faciliteiten om de opdracht uit te voeren	3,27	0,76	0,38
3. De beoordelingscriteria bij de opdracht zijn helder	3,26	0,85	0,32
4. De opdracht is moeilijk (gehercodeerd)	2,88	0,90	0,25
5. Mijn leraar geeft praktische aanwijzingen	3,11	0,78	0,52
6. Mijn leraar gaat na of we de onderzoeksopdracht begrijpen	2,74	0,96	0,53
7. Mijn leraar geeft heldere aanwijzingen	3,24	0,78	0,55
8. Mijn leraar is enthousiast	3,43	0,77	0,52
9. Mijn leraar is bereid iets voor een tweede maal uit te leggen	3,38	0,79	0,50
10. Mijn leraar heeft vertrouwen in ons	3,17	0,85	0,56
11. Mijn leraar heeft hoge verwachtingen van ons	2,93	0,79	0,28
12. Mijn leraar is geïnteresseerd in ons	3,20	0,78	0,64
13. Mijn leraar toont belangstelling voor onze ideeën	3,16	0,83	0,64
14. Mijn leraar let op samenwerking tussen leerlingen	2,91	1,91	0,27
15. Mijn leraar gaat na of we wel aan de onderzoeksopdracht werken	2,88	0,95	0,42

Statistieken voor de schaal		
<i>N</i>	=	317
<i>n</i> items	=	15
<i>M</i>	=	3,13
<i>SD</i>	=	0,89
Scheefheid	=	-0,26
Kurtosis	=	0,97
Cronbach's $\alpha$	=	0,81

## Bijlage 2

### Voorbeeld van een beoordelingsformulier

#### criterium e)

#### Gebruiken van instructie en begeleiding die zelfstandig onderzoek bevorderen (INSTRU)

A – De docent gebruikt instructie- en begeleidingsvormen die zelfstandig onderzoek in de gammavakken bevorderen op juiste wijze. Hij besteedt in zijn instructie en begeleiding aandacht aan leeractiviteiten die bij zelfstandig onderzoek aan de orde zijn en sluit aan bij de fase van onderzoek waarin de leerlingen zich bevinden. Hij houdt in de gaten of leerlingen de instructie hebben begrepen en niet teveel sturing krijgen.

*De docent komt volledig of in sterke mate tegemoet aan deze omschrijving.*

B – De docent gebruikt op redelijke of op enige wijze instructie- en begeleidingsvormen die zelfstandig onderzoek in de gammavakken bevorderen; hij besteedt in zijn instructie en begeleiding in redelijke of enige mate aandacht aan leeractiviteiten die bij zelfstandig onderzoek aan de orde zijn en sluit redelijk of enigszins aan bij de fase van onderzoek waarin de leerlingen zich bevinden. Hij houdt redelijk of enigszins in de gaten of leerlingen de instructie hebben begrepen en niet teveel sturing krijgen. *Een andere mogelijkheid is dat de docent bij het gebruik van instructie- en begeleidingsvormen*

wel veel aandacht besteedt aan de leeractiviteiten en fasen van onderzoek (zie A), maar minder goed in de gaten houdt of leerlingen de instructie hebben begrepen en niet teveel sturing krijgen (of andersom).

C – De docent gebruikt nauwelijks of geen instructie- en begeleidingsvormen die zelfstandig onderzoek in de gammavakken bevorderen; hij besteedt in zijn instructie en begeleiding nauwelijks of geen aandacht aan leeractiviteiten die bij zelfstandig onderzoek aan de orde zijn en sluit niet goed aan bij de fase van onderzoek waarin de leerlingen zich bevinden. Hij houdt nauwelijks of niet in de gaten of leerlingen de instructie hebben begrepen en niet teveel sturing krijgen.

*Een andere mogelijkheid is dat de docent bij het gebruik van instructie- en begeleidingsvormen wel enige aandacht besteedt aan de leeractiviteiten en fasen van onderzoek (zie B), maar geheel niet in de gaten houdt of leerlingen de instructie hebben begrepen en niet teveel sturing krijgen (of andersom).*

Scoring:

3 = A. Docent komt volledig of in sterke mate tegemoet aan het criterium; (zeer) goed, sterk punt;

2 = B. Docent komt in redelijke of in enige mate tegemoet aan het criterium, middelmatig; vergt verdere ontwikkeling;

1 = C. Docent komt in geringe mate of niet tegemoet aan het criterium; onvoldoende, te weinig.

Portfolio-element	Omschrijving
Interview	In het interview voor de les introduceert de docent de les die op video wordt opgenomen. Hij of zij geeft aan wat leerlingen gaan doen, wat hij of zij gaat doen, welke faciliteiten worden gebruikt en hoe de les aansluit bij andere lessen. In het interview na de les becommentarieert de docent de les die is opgenomen. Hij of zij blikt terug en beredeneert wat goed ging en waarom hij of zij heeft gehandeld zoals hij of zij heeft gedaan.
Video	Tweemaal wordt een door de docent gekozen les opgenomen waarbij hij of zij instructie en begeleiding geeft aan leerlingen met betrekking tot onderzoekswerk. De docent kiest zelf de lessen uit en geeft aan welke aaneengesloten 20" van de betreffende lessen zullen worden opgenomen in het portfolio.
Opdracht	De docent wordt gevraagd de aan de leerlingen verstrekte opdracht, inclusief de eventuele schriftelijke instructie in het portfolio te voegen en de opdracht toe te lichten.

### Instructies aan beoordelaars bij het beoordelen

1. Markeer in de portfolio-onderdelen de passages die iets zeggen over zelfstandig onderzoek instrueren en begeleiden. Nummer de passages als e.1, e.2, e.3, enz.
2. Hoe komt de docent in de gemarkeerde passages tegemoet aan het criterium? m.a.w. hoe interpreteer je de passages?
3. Vat je interpretaties samen in een eindscore (1-3). Daarbij weegt de video-opname iets zwaarder mee (zie onderstaande figuur). Motiveer waarom deze eindscore volgens jou het meest passend is.

#### Indicatie voor weging bij criterium e)

#### Gebruiken van instructie en begeleiding die zelfstandig onderzoek bevorderen (INSTRU)



■ interview

■ video

□ opdracht