De professionele ontwikkeling van universitaire docenten

Een evaluatie van de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent', aangeboden door het CNA

1ste beoordelaar/begeleider: Drs. Yvette Sol
2de beoordelaar: Dr. Frans Prins

Opdrachtgever: Centrum Nascholing Amsterdam

Groep 24
Henriëtte Braakmann 3304353
Bregje Jennen 3369056
Susanne Matena 0011975

Bachelorthesis Onderwijskunde
Universiteit Utrecht, juni 2010

Samenvatting

Sinds 2008 worden er bekwaamheidseisen aan universitaire docenten gesteld, middels de certificering Basiskwalificatie Onderwijs. Een onderdeel voor het behalen van deze certificering is vaak het volgen van één of meerdere cursussen. Om een indruk te krijgen wat de opbrengsten van dergelijke cursussen zijn, is gekozen om de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent', aangeboden door het Centrum Nascholing Amsterdam, te evalueren. In dit evaluatieonderzoek is, aan de hand van een self-assessment, in kaart gebracht of en in welke mate de deelnemers hun eigen didactische kennis en vaardigheden anders percipiëren na afloop van de cursus, ten opzichte van de door henzelf waargenomen beginsituatie. Daarnaast wordt bepaald of de vooraf gestelde persoonlijke leerdoelen volgens de deelnemers zijn behaald. Een pretest-posttest design en een retrospectief design zijn naast elkaar gezet om een response-shift bias te ondervangen. Resultaten laten zien dat er op de cursus als geheel aantoonbaar sprake is van een toename van didactische kennis en vaardigheden. Tevens kan gesteld worden dat de cursus meer dan voldoende heeft bijgedragen aan het behalen van de persoonlijke leerdoelen.

Kernwoorden: *professionele ontwikkeling, nascholingscursus, didactiek, self-assessment, response-shift bias, retrospectieve voormeting.*

Inleiding en probleemstelling

Onder het credo 'Een leven lang leren' neemt de aandacht voor professionele ontwikkeling van docenten heden ten dage een belangrijke plaats in binnen het onderwijsbeleid. De snel veranderende kenniseconomie vraagt om docenten die goed zijn opgeleid en hun kennis en vaardigheden voortdurend ontwikkelen (Onderwijsraad, n.d.). In dit kader heeft de overheid een aantal maatregelen genomen om de professionele ontwikkeling van docenten te stimuleren. Sinds 2008 worden er, middels een overeenkomst tussen verschillende universiteiten, bekwaamheidseisen aan universitaire docenten gesteld, uitmondend in een certificering Basiskwalificatie Onderwijs (hierna: BKO) (VSNU, n.d.).

Een onderdeel voor het behalen van deze certificering is vaak het volgen van één of meerdere cursussen. Om een indruk te krijgen wat de opbrengsten van dergelijke cursussen zijn, is gekozen om de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' te evalueren. Deze cursus wordt aangeboden door het Centrum Nascholing Amsterdam (hierna: CNA) en is specifiek gericht op docenten van de Universiteit van Amsterdam (hierna: UvA). De cursus resulteert in een deelcertificaat BKO en is bedoeld als een eerste stevige handreiking ten aanzien van didactiek voor beginnende docenten.

In het licht van de professionele ontwikkeling van universitaire docenten wordt in dit evaluatieonderzoek nagegaan wat de opbrengsten zijn van de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' voor de deelnemers ten aanzien van de toename van hun didactische kennis en vaardigheden. De methode die daartoe gehanteerd wordt is een self-assessment, waarbij de zelfperceptie ten aanzien van de eigen didactische kennis en vaardigheden als uitgangspunt

genomen wordt. Om een zo zuiver mogelijke meting te waarborgen zijn een pretest-posttest design en een retrospectief design naast elkaar gezet om een response-shift bias te ondervangen. Middels de antwoorden op de gestelde onderzoeksvragen kan het CNA inzicht verkrijgen in de door de deelnemers zelf waargenomen opbrengst van de cursus ten aanzien van didactische kennis en vaardigheden. De toegepaste aanpak bij de evaluatie van deze cursus en de hierbij gebruikte instrumenten kunnen bijdragen aan het verder vormgeven van een meer systematische evaluatie van soortgelijke cursussen binnen het CNA.

Vanuit wetenschappelijk oogpunt kan dit onderzoek bijdragen aan het verkrijgen van meer inzicht in het al dan niet optreden van een response-shift bias bij het gebruik van een self-assessment. Aan de hand van de uitkomsten van dit onderzoek kunnen aanbevelingen worden geformuleerd voor soortgelijk onderzoek bij andere cursussen.

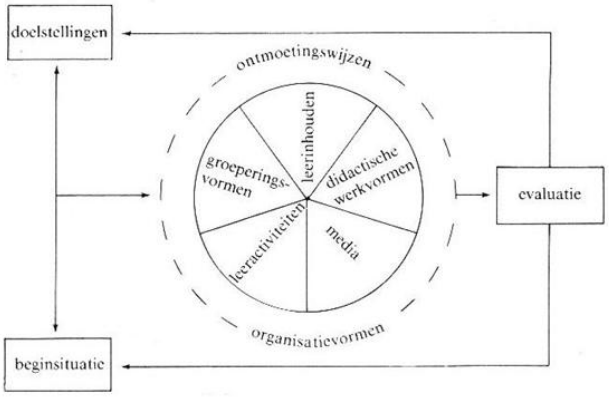
Theoretische achtergrond

Didactiek

De driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' richt zich op didactiek. De cursusdoelen zijn 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'. Het oefenen van vaardigheden in de praktijk staat centraal. De cursusopzet is dat er zoveel mogelijk wordt aangesloten bij de bestaande onderwijservaringen van de deelnemers.

Didactiek, in ons taalgebied vaak opgevat als 'een geheel van opvattingen over en vuistregels voor het praktisch handelen in de schoolsituatie' kent geen eenduidige definitie (De Corte, 1981 in: Lagerweij & Lagerweij-Voogt, 2004). Er bestaat geen algemeen aanvaarde theorie van het onderwijzen en leren**.** Volgens Lagerweij en Lagerweij-Voogt (2004) wordt daarom veelal gebruik gemaakt van didactische modellen. Deze modellen kennen een brede invulling, waarbij de methodiek meestal centraal staat. Een methodiek omvat doorgaans de volgende kenmerken: doelstellingen, beginsituatie, leerstof, didactische werkvormen, onderwijsleermiddelen, leeractiviteiten en bepaling van resultaten.

Een bekend model dat tot op heden nog gebruikt wordt en ook in de cursus naar voren komt is het didactische analysemodel van Van Gelder uit 1973 (figuur 1). Dit model is een beschrijving van het onderwijsleerproces. Het bevat een aantal didactische sleutelvragen die verwijzen naar relevante aspecten voor lesgeven. Sleutelvragen zijn onder meer: 'Wat moeten de leerlingen bereiken?', 'Hoe kan ik aansluiten bij de leerlingen?', 'Hoe kan ik het onderwijs geven?' en 'Met welk resultaat heb ik mijn onderwijs gegeven?' (Van Gelder, Oudkerk Pool, Peters & Sixma, 1979). Op een aantal aspecten van het model is op cursusdag 1 nader ingegaan, waaronder het kiezen en uitvoeren van didactische werkvormen.



Figuur 1. Het didactisch analysemodel van Van Gelder

Didactische werkvormen kiezen en uitvoeren

De term didactische werkvorm kent een veelheid aan definities. Veelal ligt de nadruk binnen de definitie op activiteiten van de docent, terwijl ook de activiteiten van studenten centraal kunnen staan (Hoogeveen & Winkels, 2009). In dit onderzoek is gekozen voor de definitie van Milius, Oost en Holleman (2001) waarbij de nadruk ligt op academische vorming en actief leren, te weten: "de aanpak die een docent kiest om tijdens contacturen studenten te ondersteunen in hun studie".

Werkvormen kunnen op verschillende manieren ingedeeld worden, bijvoorbeeld naar bekwaamheden, naar studietaken, naar vormen van academisch denken of naar didactische principes. Gezien de inhoud van de cursus is gekozen voor de indeling naar didactische principes. Daarbij worden de volgende aspecten onderscheiden: 'samenwerken of individueel werken', 'docentsturing of studentsturing', 'schoolse context of realistische context' en 'complete taken of deeltaken' (Milius, Oost & Holleman, 2001).

Het kiezen van werkvormen wordt bepaald door de leerdoelen die worden nagestreefd (cognitief, affectief, regulatief) en/of relevante situatiekenmerken. Relevante situatiekenmerken zijn kenmerken die te maken hebben met de student (zoals kennis, leerstijl of motivatie), de docent (zoals didactische vaardigheden, onderwijsstijl of beschikbare tijd) en randvoorwaarden (zoals inrichting en organisatie van de school, onderwijsleermiddelen en groepsgrootte) (Hoogeveen & Winkels, 2009).

Universiteiten bieden in toenemende mate studieprogramma's aan waarin werkvormen worden gebruikt waarbij de studenten een meer actieve rol vervullen. Binnen veel opleidingen wordt activerend onderwijs verzorgd onder de noemer van werkcolleges, in samenhang met hoorcolleges (Kallenberg, van der Grijspaarde & ter Braak, 2009). Dit past binnen het hedendaags sociaal-constructivistische onderwijsconcept. Hierbij wordt leren steeds minder opgevat als een passief, kennisabsorberend en extern gestuurd proces, maar gezien als een actief, constructief en zelf gestuurd proces (Lagerweij & Lagerweij-Voogt, 2004).

Het kiezen en uitvoeren van didactische werkvormen komt zowel theoretisch als praktisch aan bod binnen de cursus. Het model van Van Gelder is ook uitgangspunt voor zelf te ontwerpen microlessen op de tweede cursusdag. Het cursusdoel 'helder uitleggen en goed presenteren' staat dan centraal.

Helder het onderwijs uitleggen en goed presenteren

Een activerend college kan gedefinieerd worden als een college waarin een docent een stimulerende leeromgeving creëert en aanzet tot actief leergedrag van studenten (Van Dijk, 2000). Binnen het sociaal-constructivistische onderwijsconcept wordt hiermee gedoeld op het mentaal actief zijn van studenten. Het is niet alleen de taak van een docent om studenten aan het werk, maar ook aan het denken te zetten (Stokking, 2003). Een goed lesplan ontwerpen, variëteit in instructie geven, het efficiënt indelen van beschikbare tijd, het betrekken van de studenten bij het leerproces en de successcores behaald door de hiervoor genoemde vaardigheden, kunnen bijdragen aan effectief onderwijzen en de transfer van kennis (Mosca, Paul & Skiba, 2006).

Een drijfveer voor het gedrag van een student, dat vertoond wordt als respons op interne dan wel externe stimuli, is motivatie (Skinner, 2000 in: Van Dijk, 2000). Om studenten te motiveren is het van belang de aandacht te krijgen en vast te houden, daarbij rekening houdend met een spanningsboog van 15 tot 20 minuten (Driscoll, 1994; McKeachie & Svinicky, 2006). In de basis is aandacht gerelateerd aan de mate waarin iemand zich openstelt voor nieuwe kennis en gemotiveerd is om hernieuwde kennis tot zich te nemen. Hieraan ten grondslag ligt het vermogen te focussen op de complexiteit van de omgeving en de complexiteit van de taak (Driscoll, 2005 in: Woolfolk, Hughes & Walkup, 2008).

De houding en het enthousiasme van een docent die de te verwerken stof aandraagt, heeft een belangrijk effect op leren en motiveren van studenten. Het enthousiasme en de expressiviteit, evenals het talent om als van nature lesstof over te dragen, zal verschillen van docent tot docent. Het potentiële talent om te presenteren is aan te leren door techniek, oefening en het uitvoeren in de praktijk (McKeachie & Svinicky, 2006). De vaardigheid van een docent in het afwisselen van verbale en non-verbale communicatie, het afwisselen van onderwijsleermiddelen en afwisseling in didactische werkvormen, zal bijdragen aan het trekken van aandacht en het verhogen van belangstelling om actief leergedrag te bewerkstelligen (Van Dijk, 2000).

De tweede bijeenkomst van de cursus didactiek, aangeboden door het CNA, nodigde de cursisten uit een kleine presentatie te geven over een onderwerp voortkomend uit de eigen specialistische achtergrond. Het gebruik van stem, spreektempo, lichaamshouding en het maken van oogcontact werden als deelaspect van een observatieschema, ten behoeve van didactische en presentatievaardigheden, kritisch beschouwd door de groep cursisten en de trainer.

Een groep studenten begeleiden

Op cursusdag drie stond het begeleiden en beoordelen van studenten centraal. Binnen de huidige ontwikkelingen in het onderwijs is er veel aandacht voor het stimuleren van het zelfstandig leren van studenten. Een nieuwe uitdaging voor docenten is het goed begeleiden hiervan. Centraal bij het zelfstandig leren staat het actief betrokken zijn van de student bij het sturen van het eigen leerproces. De mate van zelfstandigheid van de student is een belangrijk criterium voor de mate van begeleiding die een docent kan bieden. Een meer sturende rol van de docent is gewenst wanneer de student nog onbewust onbekwaam is ten aanzien van zijn aanpak van de taak en manier van leren. De mate van sturing kan tijdens het leerproces worden aangepast, afhankelijk van het bereikte niveau van de student (Sol & Stokking, 2008). Volgens van der Schaaf en Stokking (1999) verloopt de ontwikkeling van het zelfstandig leren van aanvankelijk veel docentsturing, via gedeelde sturing naar meer leerlingsturing.

Afhankelijk van de complexiteit van de uit te voeren taken en de daarbij benodigde begeleiding, kan een docent specifieke begeleidingsstrategieën inzetten. Belangrijke begeleidings-strategieën zijn 'scaffolding', 'modeling' en 'coaching' (Sol & Stokking, 2008). Bij scaffolding geeft de docent ondersteuning daar waar nodig, waarbij deze ondersteuning geleidelijk afneemt. Bij modeling laat de docent zien hoe hij of zij een complexe taak aanpakt als voorbeeld voor de student. In het geval van coaching geeft de docent interactieve en gesitueerde feedback en suggesties (Collins, Brown en Newman, 1989 in: Sol & Stokking, 2008). Het doel van scaffolding, modeling en coaching is het ontwikkelen van cognitieve en metacognitieve vaardigheden bij studenten met behulp van observatie en begeleid oefenen. Om deze (meta)cognitieve vaardigheden te stimuleren, kunnen het geven van feedback en het stimuleren tot het zelf monitoren en evalueren van het eigen leerproces elementen zijn die docenten in de begeleiding bewust kunnen aanbieden en doseren (Sol & Stokking, 2008).

De microlessen op dag 2 boden een authentieke context waarbinnen het oefenen van feedback geven als onderdeel van begeleiding werd gepraktiseerd.

Studenten beoordelen

In de literatuur maakt men ten aanzien van beoordelen een onderscheid tussen 'summatief' en 'formatief' beoordelen (Gipps, 1994). Binnen het huidige onderwijs vindt formatief beoordelen, in de vorm van het geven van feedback, het tussentijds beoordelen en evalueren van het bereikte niveau van de student, steeds meer opgang (Sol & Stokking, 2008). Deze vorm van beoordeling wordt tijdens een periode of cursus gebruikt, hoofdzakelijk om terug te koppelen naar het onderwijsleerproces (Gipps, 1994). Summatieve beoordeling vindt plaats aan het einde van een periode of cursus en wordt gebruikt om te toetsen of studenten aan een vooraf gestelde norm voldoen (Gipps, 1994). De beoordeling resulteert vaak in een cijfer.

Binnen de cursus valt het cursusdoel 'studenten beoordelen' onder het summatief beoordelen. Het formatief beoordelen wordt gezien als een aspect van het cursusdoel 'studenten begeleiden'. Belangrijke vaardigheden die een docent moet kunnen toepassen bij summatief beoordelen zijn

‘selecteren en beslissingen nemen’, ‘opstellen van criteria’ en ‘hanteren van criteria’. Binnen de cursus is aandacht besteed aan de twee laatstgenoemde vaardigheden.

Voor het bepalen van de mate waarin de cursusdoelen op basis van zelfperceptie door de cursisten worden beheerst en toegepast, is voor dit onderzoek gebruik gemaakt van een self-assessment.

Self-assessment

Binnen evaluatieonderzoek wordt een onderscheid gemaakt tussen ‘cursusuitkomsten’ en ‘cursuseffectiviteit’. Hierbij verwijst ‘cursusuitkomsten’ naar de vraag wat de leerwinst is voor de deelnemers. ‘Cursuseffectiviteit’ behelst de vraag of de beoogde cursusdoelen behaald zijn (Lam, 2009). Aangezien het hier een driedaagse cursus betreft, is het niet mogelijk om feitelijke cursuseffectiviteit vast te stellen. De opgedane kennis en vaardigheden zullen op langere termijn moeten beklijven en in praktijk gebracht worden willen de cursusdoelen zoals omschreven getoetst kunnen worden (Lam, 2009). Met het oog op de uitvoerbaarheid van het onderzoek is gekozen om de cursusuitkomsten te evalueren.

Om cursusuitkomsten te meten kan self-assessment een goede evaluatiemethode zijn (D’eon, Sadownik, Harrison & Nation, 2008). Een self-assessment is relatief eenvoudig uit te voeren en vereist minimale middelen in vergelijking met andere evaluatiemethoden als het doen van observaties, het maken van tests of het voeren van gesprekken (D’eon et al., 2008; Stufflebeam & Wingate, 2005). Voor het begrip self-assessment worden in de literatuur verschillende definities gehanteerd. In dit onderzoek wordt onder self-assessment verstaan “ het betrekken van deelnemers bij het nemen van beslissingen over het eigen leerproces” (Boud & Falchikov, 1989).

Een kritische noot ten aanzien van self-assessments is dat resultaten niet generaliseerbaar zijn. Self-assessments vormen daarom geen goede basis om harde besluiten op te nemen ten aanzien van eventuele verander- of verbetertrajecten (Lam, 2009). Een belangrijke factor die hieraan ten grondslag ligt, is dat er gebruik gemaakt wordt van de zelfperceptie van deelnemers. Hieronder wordt verstaan “het inschatten van eigen kennis en vaardigheden vanuit een intrinsieke beleving, gebaseerd op eigen ervaringen en concepties” (Van Dijk, 2000). Veel deelnemers kunnen zichzelf niet accuraat beoordelen, veelal veroorzaakt door het hebben van onvoldoende kennis met betrekking tot het te meten construct van de cursus. Dit kan ertoe leiden dat deelnemers zichzelf vaak over- of onderschatten. Van een overschatting van eigen kennis en vaardigheden is sprake als de deelnemer vooraf denkt over meer kennis en vaardigheden te beschikken aangaande het construct van de cursus dan daadwerkelijk het geval is (D’eon et al., 2008). Afhankelijk van het gekozen design kan een over- of onderschatting leiden tot systematische meetfouten en vormt daarmee een bedreiging voor de validiteit van de meting (Lam, 2009).

Traditioneel gezien wordt er bij het evalueren van een cursus door middel van self-assessment gebruik gemaakt van een voormeting aan het begin van een cursus (pretest) en een nameting aan het einde van een cursus (posttest) (Drennan & Hyde, 2008). Dit zogenoemde pretest-posttest design

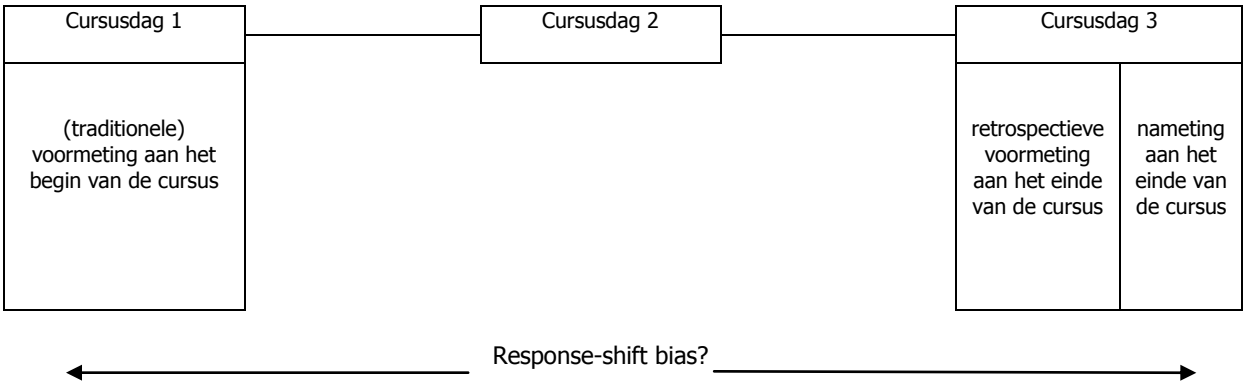
kent evenwel beperkingen. Zo neemt het afnemen van een pretest en posttest vragenlijst vaak veel tijd in beslag (Marshak, de Silva & Silberstein, 1998), kunnen deelnemers bewust een lagere score bij de voormeting aangeven om de te meten toename in kennis of vaardigheid te vergroten (Lam & Bengo, 2003) en moeten deelnemers aan het begin en het einde van de cursus aanwezig zijn om een zinvolle vergelijking te kunnen maken (Pratt, McGuigan & Katzev, 2000). De belangrijkste beperking echter is dat er sprake kan zijn van een response-shift bias. Dit kan worden gedefinieerd als een "programmageproduceerde verandering in het deelnemersbegrip van het construct dat wordt gemeten" (Howard & Daily, 1979). Het volgen van de cursus heeft geleid tot meer inzicht in het construct dat gemeten wordt. Daarnaast heeft een bijstelling plaatsgevonden van de eigen perceptie van de deelnemers ten aanzien van het oorspronkelijke niveau van functioneren. Als gevolg hiervan zijn de scores op de voor- en nameting minder valide (Hoogstraten, 1982). Om een dergelijke response-shift bias te voorkomen kan beter gebruik worden gemaakt van een 'retrospectieve voormeting' aan het einde van de cursus (Howard & Daily, 1979).

Retrospectieve voormeting

Het evalueren van een cursus aan de hand van de retrospectieve voormeting is een meer legitieme beoordeling van de resultaten dan met behulp van het pretest-posttest design (Pratt et al., 2000). Bij de retrospectieve voormeting worden de voor- en nameting op hetzelfde moment afgenomen, aan het einde van de cursus. De response-shift bias neemt af doordat vanuit een zelfde intern referentiekader de self-assessment wordt uitgevoerd (Howard, 1980).

Onderzoeksvragen

In figuur 2 is te zien dat er sprake is van drie meetmomenten. Onderzocht zal worden of er een verschil is tussen de traditionele voormeting en de retrospectieve voormeting, tussen de traditionele voormeting en de nameting en tussen de retrospectieve voormeting en de nameting van de cursisten ten aanzien van hun didactische kennis en vaardigheden.



Figuur 2. Meetmomenten voor de cursus 'Didactiek Basiskwalificatie UvA-docent'.

De volgende vragen staan centraal binnen dit onderzoek:

1. Is er een verschil tussen de voormeting en de retrospectieve voormeting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden en in het bijzonder ten aanzien van de afzonderlijke cursusdoelen 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'?
2. Is er een verschil tussen de voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden en in het bijzonder ten aanzien van de afzonderlijke cursusdoelen 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'?
3. Is er een verschil tussen de retrospectieve voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden en in het bijzonder ten aanzien van de afzonderlijke cursusdoelen 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'?
4. In welke mate heeft deze cursus vanuit de perceptie van de deelnemers bijgedragen aan het behalen van de persoonlijke leerdoelen 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'?

Verwachtingen

Op basis van bovenstaande onderzoeksvragen zijn de volgende drie verwachtingen geformuleerd:

Verwachting ten aanzien van de voormeting en de retrospectieve voormeting

Er is sprake van een verschil tussen de voormeting en de retrospectieve voormeting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Dit verschil is te verklaren vanuit een over- of onderschatting van de deelnemers aangaande de eigen kennis en vaardigheden rondom het construct didactiek. Voorafgaande aan de cursus is weinig informatie verstrekt aan de deelnemers omtrent de inhoud van de cursus en over wat er verstaan wordt onder de gestelde cursusdoelen. Op basis hiervan verwachten wij een response-shift bias, ofwel een veranderd referentiekader ten opzichte van het construct didactiek en een bijstelling van de eigen perceptie van de deelnemers ten aanzien van het oorspronkelijke niveau van functioneren. Indien er sprake is van een response-shift bias, zal gebruik worden gemaakt van de gegevens uit de retrospectieve voormeting om het door de deelnemers waargenomen verschil in mogelijke uitbreiding van didactische kennis en vaardigheden binnen dit onderzoek in kaart te kunnen brengen. Indien het verschil tussen de voormeting en de retrospectieve voormeting niet significant is, zullen de uitkomsten van de retrospectieve voormeting toch meegenomen worden, aangezien uit de theorie blijkt dat deze meting meer valide is.

Verwachting ten aanzien van de voormeting en de nameting

Er is sprake van een verschil tussen de voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Een toename van didactische kennis en vaardigheden heeft plaatsgevonden door het volgen van de cursus. De verwachting is dat de meeste deelnemers over weinig onderwijservaring beschikken waardoor een toename een logische uitkomst zou zijn.

Verwachting ten aanzien van de retrospectieve voormeting en de nameting

Er is sprake van een verschil tussen de retrospectieve voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Een toename van didactische kennis en vaardigheden heeft plaatsgevonden door het volgen van de cursus, welke bij het gebruik van een retrospectieve voormeting op theoretische gronden ook meer valide is. Deelnemers met weinig onderwijservaring kunnen door meer kennis over het construct didactiek beter bepalen welke didactische kennis en vaardigheden zij onvoldoende beheersten bij aanvang van de cursus.

Methode

Deelnemers

Deelnemers zijn 14 cursisten (n=14) aan de driedaagse CNA nascholingscursus 'Didactiek Basiskwalificatie UvA-docent'. Kenmerkend voor deze heterogene groep, met verschil in onderwijservaring, werkzaamheden en motivatie, is dat het vaak (jonge) beginnende docenten, dan wel assistenten in opleiding (aio's) met een kleine onderwijsaanstelling aan de Universiteit van Amsterdam zijn. Deelnemers zijn aangemeld door hun werkgever of hebben zelf een verzoek ingediend deel te mogen nemen.

Instrumenten

Er is gebruik gemaakt van schriftelijke vragenlijsten. De keuze voor schriftelijke vragenlijsten komt voort uit het feit dat binnen een korte tijd gevraagd kan worden naar verschillende gegevens. De vragen hebben zowel een open kwalitatief als gesloten kwantitatief karakter en zijn, afhankelijk van de complexiteit van de kenmerken, (in)direct bevraagd. Het afnemen van de vragenlijsten geschiedde bij aanvang van de cursus op dag 1 en aan het einde van de cursus op dag 3, waarbij in totaal sprake was van drie meetmomenten: de voormeting, de retrospectieve voormeting en de nameting. Deze zullen achtereenvolgens toegelicht worden.

Meetmoment 1

De eerste meting betrof de voormeting aan het begin van cursusdag 1, welke twee delen omvatte. Het eerste deel bestond uit open en gesloten vragen betreffende de achtergrond van de cursist op het gebied van geslacht, leeftijd, onderwijservaring, onderwijstaken en motivatie ten aanzien van het volgen van de cursus. De antwoorden op deze vragen zijn gebruikt voor de interpretatie van de onderzoeksgegevens achteraf. Daarnaast bestond dit deel uit open vragen ten aanzien van de cursusdoelen zoals gesteld door het CNA, te weten 'helder het onderwijs uitleggen en goed presenteren', 'didactische werkvormen kiezen en uitvoeren', 'een groep studenten begeleiden' en 'studenten beoordelen'. Een voorbeeldvraag is: 'Wat verstaat u onder: "U kunt helder het onderwijs uitleggen en goed presenteren"?'. Met deze vragen wilden we inzicht krijgen in de interpretatie van de cursisten ten aanzien van deze cursusdoelen. Het eerste deel eindigt met open vragen aangaande de persoonlijke leerdoelen. Doel was het achterhalen van de persoonlijke leerdoelen van de cursist. Een voorbeeld is: 'Waar loopt u in de praktijk tegenaan waarover u tijdens deze cursus zou willen leren?'.

Het tweede deel van de voormeting bestond uit gesloten vragen betreffende de vier cursusdoelen. De vragen zijn gericht op het handelen van docenten, waarbij een tweetal vijfpuntsschalen is gehanteerd, te weten 'ik beheers dit' 'onvoldoende', 'matig', 'voldoende', 'redelijk', 'goed' of 'ik doe dit' 'nooit', 'bijna niet', 'soms', 'regelmatig', 'meestal'. De vragen zijn gebaseerd op het instrument van Sol en Stokking (2008), welke is ontwikkeld om (nieuwe) vaardigheden van docenten in kaart te brengen. De items die zijn gebruikt bij de samenstelling van de vragenlijst in dit onderzoek, zijn geselecteerd op basis van de operationalisering van de vier cursusdoelen 'uitleggen en presenteren', 'didactische werkvormen kiezen en uitvoeren', 'begeleiden' en 'beoordelen', zoals in het theoretisch kader gedefinieerd. Doel was het vaststellen van de beginsituatie van de deelnemers ten aanzien van de cursusdoelen.

Om het cursusdoel 'uitleggen en presenteren' te meten zijn 22 items gebruikt. Deze items zijn geformuleerd rond presentatievaardigheden, gericht op gedragingen van de docent bij het geven van hoor- en werkcolleges om aandacht en belangstelling van studenten vast te houden. Een voorbeeld is: 'Ik ben in staat om een gepast spreektempo te hanteren'. Om het cursusdoel 'didactische werkvormen kiezen en uitvoeren' te meten zijn 11 items gebruikt, waarbij de aspecten 'samenwerken of individueel werken', 'docentsturing of studentsturing', 'schoolse context of realistische context' en 'complete taken of deeltaken' als uitgangspunt zijn genomen. Een voorbeeld is: 'Ik ben in staat om studenten te stimuleren in verschillende groepjes samen te werken'. Om het kenmerk 'begeleiden' te meten zijn 43 items gebruikt waarbij de aspecten 'scaffolding', 'modeling' en 'coachen' centraal stonden. Een voorbeeld van een dergelijk item is: 'Ik ben in staat om, bij het geven van feedback, voorbeelden te noemen over hoe iets beter kan'. Om het kenmerk 'beoordelen' te meten zijn 15 items gebruikt op basis van de aspecten, 'opstellen van criteria', 'hanteren van criteria', en 'selecteren en beslissingen nemen'. Een voorbeeld van een dergelijk item is: 'Ik ben in staat om individuele prestaties te beoordelen aan de hand van vastgestelde criteria'.

Meetmoment 2

De tweede meting betrof de retrospectieve voormeting aan het einde van cursusdag 3. De deelnemers kregen de gesloten vragenlijst betreffende de cursusdoelen uit de voormeting voorgelegd. Vraag was om deze vragenlijst opnieuw in te vullen aan het einde van de cursus, terugkijkend naar de eigen beginsituatie bij aanvang van de cursus. Doel was het vaststellen van een response-shift bias.

Meetmoment 3

Meetmoment drie betrof de nameting aan het einde van cursusdag 3. Deze bestond uit twee delen. Het eerste deel bestond uit vragen betreffende de persoonlijke leerdoelen. Het behalen van de persoonlijke leerdoelen is gesloten bevraagd op een schaal van 1 'onvoldoende' tot 5 'goed'. De cursusinhoud en de aanpak van de trainer zijn gesloten bevraagd ten behoeve van de interpretatie van de onderzoeksgegevens. Het tweede deel van de vragenlijst omvatte wederom de gesloten vragen aangaande de cursusdoelen uit de voormeting. De vraag was om te beoordelen wat de kennis en vaardigheden zijn aan het einde van de cursus. Doel was om aan de hand van de gedane metingen een verschil te berekenen om daarmee de door de cursisten zelf waargenomen opbrengst te kunnen bepalen.

Design en procedure

Design

Het betreft hier een (summatief) evaluatieonderzoek, waarbij zowel gebruik gemaakt is van een kwantitatieve als een kwalitatieve benadering. De kwalitatieve gegevens zijn complementair gebruikt voor de interpretatie van de kwantitatieve resultaten. De methode die gehanteerd is is een self-assessment, waarbij de zelfperceptie ten aanzien van de eigen didactische kennis en vaardigheden als uitgangspunt genomen is. Een pretest-posttest design en een retrospectief design zijn naast elkaar uitgevoerd om een response-shift bias te ondervangen.

Procedure

Na het benaderen van het CNA voor het uitvoeren van een onderzoek omtrent de professionele ontwikkeling van docenten werd ons de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' toegewezen, aangeboden op 12, 19 en 26 april 2010. Volledige medewerking van de trainer en deelnemers werd toegezegd. Op cursusdag 1 heeft een korte uitleg over het doel van dit onderzoek plaatsgevonden en is de voormeting afgenomen. Het beantwoorden van de vragen nam ongeveer 20 minuten in beslag. De nameting op de laatste cursusdag nam ongeveer 30 minuten in beslag. Het beantwoorden van de vragen geschiedde niet anoniem om voor- en nameting te kunnen vergelijken. De verwerking is wel anoniem uitgevoerd.

Een tweetal zieken, waarvan één op dag 1 en één op dag 3, heeft tot non-response geleid op de voormeting en op de retrospectieve voormeting en nameting. Hoewel deelnemers aan begin en einde van de cursus aanwezig moeten zijn voor een zinvolle vergelijking, is gekozen deelnemers niet uit te sluiten. Daarnaast zijn een aantal vragenlijsten niet volledig ingevuld. Dit is over het algemeen

te wijten aan het feit dat niet elke deelnemer (praktijk)ervaring bleek te hebben met de vier cursusdoelen. Een aantal cursisten heeft bijvoorbeeld nog nooit beoordeeld. Dit verklaart veelal de non-respons op dit cursusdoel. Ook de hoeveelheid vragen leidde - naar eigen zeggen van de deelnemers - tot het onvolledig of niet serieus invullen van de vragen. Gezien de relatief kleine n van dit onderzoek, welke van invloed kan zijn op de uitkomsten, is gekozen om alle vragenlijsten mee te nemen bij verdere verwerking.

Analyse

Om de onderzoeksvragen te beantwoorden, is gebruik gemaakt van verschillende soorten analyses. De resultaten van de vragenlijsten met betrekking tot de gesloten vragen zijn ingevoerd in het kwantitatieve analyseprogramma SPSS. De betrouwbaarheid van de vier afzonderlijke schalen 'uitleggen en presenteren', 'didactische werkvormen kiezen en uitvoeren', 'begeleiden' en 'beoordelen' is uitgerekend aan de hand van het coëfficiënt Cronbach's alpha. Hierbij is een gewenste minimale alpha van .80 bij minder complexe begrippen gehanteerd. Aan de hand van een schaalanalyse is voor elk item nagegaan in hoeverre de antwoorden daarop samenhangen met de antwoorden op alle andere items van de schaal. Een item is uit de schaal verwijderd als de item-restcorrelatie lager was dan .35. Om de gemiddelden op de beginsituatie omtrent de achtergrond van de deelnemende cursisten, de afzonderlijke vier cursusdoelen en de cursus als geheel, de persoonlijke leerdoelen, de cursusinhoud en de rol van de trainer te kunnen aflezen, is een frequentieanalyse uitgevoerd. Om de verschillen tussen de voormeting, de retrospectieve voormeting en de nameting te beoordelen is tot slot een gekoppelde t-toets gehanteerd.

Resultaten

Uitkomsten betrouwbaarheidsanalyse

Voor de betrouwbaarheid van de vier schalen is een betrouwbaarheidsanalyse uitgevoerd, welke terug te vinden is in tabel 1.

Tabel 1. *Cronbach's alpha en item-restcorrelaties bij de vier cursusdoelen*

Cursusdoel	N	*k*	α	r_{ir}
Uitleggen en presenteren	11	14	.92	.39
Didactische werkvormen kiezen en uitvoeren	11	5	.82	.50
Begeleiden	10	24	.93	.37
Beoordelen	9	15	.89	.38

Een eerste analyse op het cursusdoel 'uitleggen en presenteren' (n=11) leverde een Cronbach's alpha op van .89. Op basis van de statistische analyses en de inhoudelijke verantwoording is besloten 14 items te behouden, resulterend in een Cronbach's alpha van .92. De statistische samenhang tussen de verschillende items loopt uiteen van .39 tot .86. Verwijderde items verwezen meer naar het aspect 'instructie' en minder naar het aspect 'uitleggen en presenteren'. Een voorbeeld van een item dat verwijderd is, is: 'Ik ben in staat om bij het geven van opdrachten duidelijk te maken wanneer iets af moet zijn'.

Een eerste analyse op het cursusdoel didactische werkvormen kiezen en uitvoeren (n=11) leverde een Cronbach's alpha op van .70. Op basis van de statistische analyses en de inhoudelijke verantwoording is besloten 5 items te behouden, resulterend in een Cronbach's alpha van .82. De statistische samenhang tussen de verschillende items loopt uiteen van .50 tot .72. Items die verwijderd zijn pasten minder in de indeling naar didactische principes en meer bij andere indelingen als bijvoorbeeld studietaken. Een voorbeeld is: 'Ik bied de lesstof vooral schriftelijk aan'.

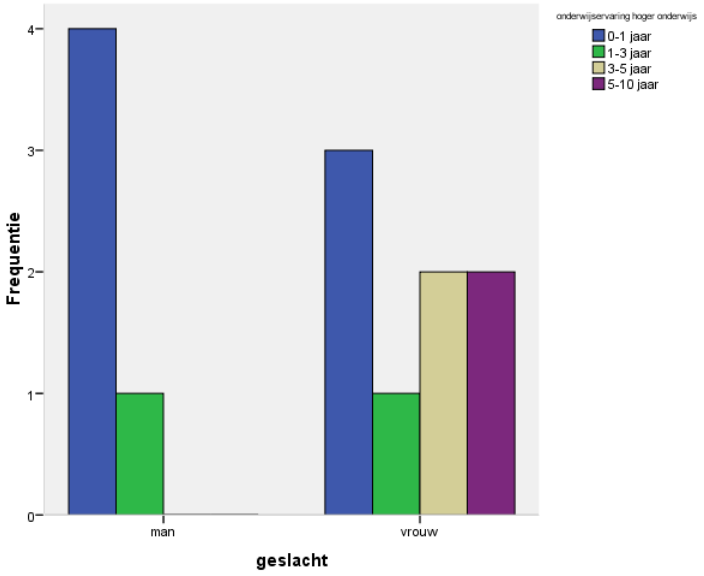
Een eerste analyse op het cursusdoel begeleiden (n=10) leverde een Cronbach's alpha op van .91. Op basis van de statistische analyses en de inhoudelijke verantwoording is besloten 24 items te behouden, resulterend in een Cronbach's alpha van .93. De statistische samenhang tussen de verschillende items loopt uiteen van .37 tot .89. Items die zijn verwijderd konden niet worden ingedeeld onder de aspecten 'scaffolding', 'modeling' of 'coachen' Een voorbeeld van een dergelijk item is: 'Ik ben in staat om studenten uit te laten wisselen welke keuzes ze hebben gemaakt bij het maken van opdrachten en waarom'.

Een eerste analyse op het cursusdoel beoordelen (n=9) leverde een Cronbach's alpha op van .84. Op basis van de statistische analyses en de inhoudelijke verantwoording is besloten 15 items te behouden, resulterend in een Cronbach's alpha van .89. De statistische samenhang tussen de verschillende items loopt uiteen van .38 tot .95. De verwijderde items waren niet onder te verdelen onder de aspecten 'selecteren en beslissingen nemen', 'opstellen van criteria' en 'hanteren van criteria'. Een voorbeeld van een dergelijk item is: 'Ik ben in staat om studenten te begeleiden bij het beoordelen van elkaars werk als dat aan de orde is'.

Uitkomsten frequentieanalyse

Om de beginsituatie omtrent de karakteristieken van de deelnemende cursisten (geslacht, leeftijd, motivatie, onderwijservaring hoger onderwijs en onderwijstaken) vast te kunnen stellen en de uitkomsten van de vragen met betrekking tot de vier cursusdoelen en de cursus als geheel, de persoonlijke leerdoelen, de cursusinhoud en de rol van de trainer te kunnen aflezen, is een frequentieanalyse op groepsniveau uitgevoerd.

Deelnemers zijn 8 vrouwen en 6 mannen met een gemiddelde leeftijd tussen de 25 en 35 jaar. Twee derde van de cursisten volgde deze cursus op eigen initiatief al dan niet aangegeven bij de werkgever. De overige deelnemers zijn door de werkgever verzocht de cursus te volgen. Zoals te zien is in figuur 3 hebben de meeste deelnemers (7) minder dan een jaar ervaring in het hoger onderwijs. Slechts een tweetal cursisten zijn meer dan 5 jaar als docent aan het hoger onderwijs verbonden.



Figuur 3. Frequentieverdeling onderwijservaring hoger onderwijs op basis van geslacht.

Uit tabel 2 blijkt dat voor het merendeel van de cursisten de onderwijstaak bestaat uit het begeleiden van groepen studenten (10), gevolgd door het begeleiden van individuele studenten (8), het geven van colleges (5) en het klassikaal begeleiden (5).

Tabel 2. *Frequentieverdeling onderwijstaken*

Onderwijstaak	Ja	Nee	Geen antwoord	Totaal
Geven van colleges	5	8	1	14
Begeleiden individuele student	8	5	1	14
Begeleiden groepen studenten	10	3	1	14
Klassikaal begeleiden	5	8	1	14

Tabel 3 laat zien dat de gemiddelde groepsscores zowel variëren op de vier cursusdoelen en de cursus als geheel, als op de drie metingen. Opvallend ten aanzien van de voormeting en de retrospectieve voormeting is dat er op de cursusdoelen 'uitleggen en presenteren', 'begeleiden' en 'beoordelen' sprake lijkt te zijn van een onderschatting van de deelnemers ten aanzien van kennis en vaardigheden, aangezien op de retrospectieve voormeting (m=3.80; m=3.63; m=3.72) hoger wordt gescoord dan op de voormeting (m=3.79; m= 3.60; m= 3.53). Dit verschil geldt in het bijzonder het cursusdoel 'beoordelen'. Ten aanzien van het cursusdoel 'het kiezen en uitvoeren van didactische werkvormen' en de cursus als geheel lijkt er sprake te zijn van een overschatting van kennis en vaardigheden. Op de retrospectieve voormeting (m=3.40; m=3.58) wordt namelijk lager gescoord dan op de voormeting (m=3.53; m=3.61). Met betrekking tot de nameting wordt er op alle cursusdoelen (m=4.10; m=3.70; m=3.94; m=4.12) en op de cursus als geheel (m=3.91) altijd hoger

gescoord dan op de beide voormetingen, wat een toename in kennis en vaardigheden lijkt te veronderstellen.

Tabel 3. *Groepsgemiddelden cursusdoelen en cursus als geheel op de drie metingen*

	Voormeting			Retrospectieve voormeting			Nameting		
	gem	sd	N	gem	sd	N	gem	sd	N
Uitleggen en presenteren	3.79	.74	13	3.80	.62	13	4.10	.57	13
Didactische werkvorm kiezen en uitvoeren	3.53	.77	13	3.40	.64	13	3.70	.61	13
Begeleiden	3.60	.54	13	3.63	.70	12	3.94	.49	12
Beoordelen	3.53	1.05	12	3.72	1.04	10	4.12	.43	9
Cursus als geheel	3.61	.64	13	3.58	.59	13	3.91	.47	13

Zoals uit tabel 4 naar voren komt kan, met 3.55 als laagste gemiddelde waarde (voldoende tot redelijk) en 4.20 (redelijk tot goed) als hoogste gemiddelde waarde, gesteld worden dat de cursus meer dan voldoende heeft bijgedragen aan het behalen van de persoonlijke leerdoelen met betrekking tot de vier cursusdoelen en het eigen praktijkprobleem.

Tabel 4. *Groepsgemiddelden en standaarddeviatie persoonlijke doelen nameting*

	gem	sd	N
Persoonlijk doel uitleggen en presenteren	4.20	1.03	10
Persoonlijk doel didactische werkvorm kiezen en uitvoeren	3.55	1.29	11
Persoonlijk doel begeleiden	3.91	.70	11
Persoonlijk doel beoordelen	3.70	.95	10
Praktijkprobleem	3.78	1.30	9

De cursusinhoud krijgt van ruim de helft van de cursisten het cijfer 8 (60%). Lager dan een 6 wordt niet gegeven. Tweederde van de cursisten (67%) geeft aan dat de trainer heeft bijgedragen aan het behalen van de persoonlijke doelen en een ruime meerderheid (60%) beloont dat met een cijfer 9.

Uitkomsten gekoppelde t-toets

Om de verschillen tussen gemiddelden van de voormeting, de retrospectieve voormeting en de nameting te beoordelen is een gekoppelde t-toets gehanteerd. Er is sprake van een verschil tussen de voormeting en de retrospectieve voormeting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden . Op de voormeting (m=3.92; m=3.64; m=3.70; m=3.66; m=3.69) wordt gemiddeld genomen altijd hoger gescoord dan op de retrospectieve voormeting

(m=3.76; m=3.45; m=3.55; m=3.60; m=3.54). Dit verschil is echter niet significant. Er is niet aantoonbaar sprake van een response-shift bias. De uitkomsten van de voormeting versus de retrospectieve voormeting zijn achtereenvolgens per cursusdoel en op de cursus als geheel beschreven in tabel 5.

Tabel 5. *Toetsing van verschillen tussen groepsgemiddelden voormeting en retrospectieve voormeting*

	Voormeting		Retrospectieve voormeting		Toetsing van verschillen tussen gemiddelden voormeting en retrospectieve voormeting			
	gem	sd	gem	sd	N	df	t	p
Uitleggen en presenteren	3.92	.60	3,76	.62	12	11	1.43	.180
Didactische werkvorm kiezen en uitvoeren	3.64	.68	3.45	.64	12	11	1.11	.293
Begeleiden	3.70	.52	3.55	.67	11	10	.86	.412
Beoordelen	3.66	1.16	3.60	1.03	9	8	.36	.730
Cursus als geheel	3.69	.60	3.54	.60	12	11	1.42	.184

Noot. $p < .05$.

Er is eveneens sprake van een verschil tussen de voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Op de nameting (m=4.05; m=3.73; m=3.88; m=4.04; m=3.87) wordt gemiddeld genomen altijd hoger gescoord dan op de voormeting (m=3.93; m=3.64; m=3.70; m=4.00; m=3.69). Dit verschil is echter niet significant. Er is niet aantoonbaar sprake van een toename van didactische kennis en vaardigheden van de deelnemers. De uitkomsten van de voormeting versus de nameting zijn achtereenvolgens per cursusdoel en op de cursus als geheel beschreven in tabel 6.

Tabel 6. *Toetsing van verschillen tussen groepsgemiddelden voormeting en nameting*

	Voormeting		Nameting		Toetsing van verschillen tussen gemiddelden voormeting en nameting			
	gem	sd	gem	sd	N	df	t	p
Uitleggen en presenteren	3.93	.60	4.05	.55	12	11	-1.31	.217
Didactische werkvorm kiezen en uitvoeren	3.64	.68	3.73	.63	12	11	- 0.54	.601
Begeleiden	3.70	.52	3.88	.46	11	10	- 1.56	.149
Beoordelen	4.00	.64	4.04	.37	8	4	- 0.27	.793
Cursus als geheel	3.69	.60	3.87	.47	12	11	- 1.59	.141

Noot. $p < .05$.

Op de nameting (m=4.10; m=3.70; m=3.94; m=4.12; m=3.91) wordt gemiddeld genomen altijd hoger gescoord dan op de retrospectieve voormeting (m=3.00; m=3.40; m=3.63; m=4.02; m=3.58). Dit verschil is significant voor het cursusdoel 'uitleggen en presenteren', 'didactische werkvormen kiezen en uitvoeren' en het cursusdoel 'begeleiden'. Er is geen significant verschil ten aanzien van het cursusdoel 'beoordelen'. Ten aanzien van de cursus als geheel is eveneens sprake van een significant verschil. Er is aantoonbaar sprake van een toename van didactische kennis en vaardigheden van de deelnemers. De uitkomsten van de retrospectieve voormeting versus de nameting zijn achtereenvolgens per cursusdoel en op de cursus als geheel beschreven in tabel 7.

Tabel 7. *Toetsing van verschillen tussen groepsgemiddelden retrospectieve voormeting en nameting*

	Retrospectieve voormeting		Nameting		Toetsing van verschillen tussen gemiddelden retrospectieve voormeting en nameting			
	gem	sd	gem	sd	N	df	t	p
Uitleggen en presenteren	3.00	.62	4.10	.57	13	12	- 4.86	.000
Didactische werkvorm kiezen en uitvoeren	3.40	.64	3.70	.61	13	12	- 3.54	.004
Begeleiden	3.63	.70	3.94	.49	12	11	- 3.04	.011
Beoordelen	4.02	.44	4.12	.43	9	8	- 2.00	.081
Cursus als geheel	3.58	.59	3.91	.47	13	12	- 5.64	.000

Noot. $p < .05$.

Uitkomsten open vragen vier cursusdoelen

Door middel van open vragen is de interpretatie van de cursusdoelen door de deelnemers omtrent didactische kennis en vaardigheden in kaart gebracht. De antwoorden zijn gerubriceerd en dienen ter interpretatie van de kwantitatieve resultaten. Opgemerkt dient te worden dat van de 13 cursisten een aantal deelnemers meerdere antwoorden per vraag hebben gegeven.

De kwalitatieve gegevens betreffende de interpretatie van de cursisten ten aanzien van het cursusdoel 'helder het onderwijs uitleggen en presenteren' zijn in drie categorieën antwoorden te rubriceren, te weten 'activerend handelen' (11 keer genoemd), 'presentatievaardigheden' (7 keer genoemd) en 'kennisoverdracht' (10 keer genoemd). De kwalitatieve gegevens omtrent de interpretatie van de cursisten ten aanzien van het cursusdoel 'didactische werkvormen kiezen en uitvoeren' zijn in drie categorieën te rubriceren, te weten 'kennis van werkvormen' (9 keer genoemd), 'toepassen van werkvormen' (5 keer genoemd) en 'relevante situatiekenmerken' (3 keer genoemd). De kwalitatieve gegevens omtrent de interpretatie van de cursisten ten aanzien van het cursusdoel 'een groep studenten begeleiden' zijn in twee categorieën te rubriceren, te weten 'activerende rol docent' (6 keer genoemd) en 'sturende rol docent' (7 keer genoemd). De kwalitatieve gegevens

betreffende de interpretatie van de cursisten ten aanzien van het cursusdoel 'studenten beoordelen' zijn in drie categorieën te rubriceren, namelijk 'opstellen van criteria' (5 keer genoemd), 'hanteren van criteria' (8 keer genoemd) en 'formatief beoordelen' (3 keer genoemd).

Opvallend ten aanzien van de antwoorden van de cursisten is dat deze vaak onder te verdelen zijn in rubrieken welke overeenkomen met de aspecten die in de gesloten vragenlijst zijn bevraagd. Cursisten lijken in hun antwoorden blijk te geven van de theorie betreffende de vier cursusdoelen. Deze kennis is echter veelal onvolledig en gefragmenteerd.

Conclusie

Voormeting versus retrospectieve voormeting

Op basis van de uitkomsten van de voormeting versus de retrospectieve voormeting wordt de eerste verwachting verworpen. De eerste onderzoeksvraag leverde een verschil op tussen de voormeting en de retrospectieve voormeting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Dit verschil is echter niet significant. Er is niet aantoonbaar sprake van een response-shift bias op de vier cursusdoelen en de cursus als geheel.

Deze uitkomst is verrassend omdat, zoals uit de uitkomsten van de toetsing van verschillen tussen groepsgemiddelden op de voormeting en de retrospectieve voormeting blijkt (tabel 5), er wel sprake is van een zekere overschatting van de deelnemers. Voorafgaande aan de cursus is weinig informatie verstrekt aan de deelnemers betreffende de inhoud van de cursus en over wat er verstaan wordt onder de gestelde cursusdoelen. Volgens Hoogstraten (1982) zou het volgen van de cursus moeten leiden tot meer inzicht in het construct dat gemeten wordt en het bijstellen van de eigen perceptie van de deelnemers ten aanzien van het oorspronkelijke niveau van functioneren. De lagere gemiddelden bij de retrospectieve voormeting lijken hier dienovereenkomstig op te wijzen.

Voormeting versus nameting

Op basis van de uitkomsten van de voormeting versus de nameting wordt de tweede verwachting eveneens verworpen. De tweede onderzoeksvraag leverde een verschil op tussen de voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van hun didactische kennis en vaardigheden. Dit verschil is echter niet significant. Er is niet aantoonbaar sprake van een toename van didactische kennis en vaardigheden van de deelnemers op de vier cursusdoelen en de cursus als geheel.

Deze uitkomst is verrassend. Zowel uit de uitkomsten van de frequentieanalyse op de beginsituatie omtrent de karakteristieken van de cursisten (figuur 3), als uit de uitkomsten op de open vragen ten aanzien van de vier cursusdoelen blijkt dat de meeste deelnemers over weinig onderwijservaring beschikken en een onvolledig en gefragmenteerd beeld hebben ten aanzien van de vier cursusdoelen. Een toename in didactische kennis en vaardigheden door het volgen van de cursus zou een logische uitkomst zijn.

Retrospectieve voormeting versus nameting

Op basis van de uitkomsten van de retrospectieve voormeting versus de nameting wordt de derde verwachting aangenomen. Onderzoeksvraag drie laat een significant verschil zien tussen de retrospectieve voormeting en nameting voor het cursusdoel 'helder het onderwijs uitleggen en presenteren', het cursusdoel 'didactische werkvormen kiezen en uitvoeren' en het cursusdoel 'een groep studenten begeleiden'. Er is geen significant verschil tussen de retrospectieve voormeting en de nameting van de zelfperceptie van de cursisten ten aanzien van het cursusdoel 'studenten beoordelen'. Ten aanzien van de cursus als geheel is wel sprake van een significant verschil. Er is aantoonbaar sprake van een toename van didactische kennis en vaardigheden van de deelnemers.

Deze uitkomst ligt in de lijn der verwachting. Een toename van didactische kennis en vaardigheden heeft plaatsgevonden door het volgen van de cursus, welke bij het gebruik van een retrospectieve voormeting op theoretische gronden ook meer valide is. Deelnemers met weinig onderwijservaring en een onvolledig en gefragmenteerd beeld ten aanzien van de vier cursusdoelen kunnen door meer kennis over het construct didactiek beter bepalen welke didactische kennis en vaardigheden zij onvoldoende beheersten bij aanvang van de cursus.

Persoonlijke leerdoelen

Het behalen van de persoonlijke leerdoelen, onderzoeksvraag vier, is gesloten bevraagd op een schaal van 1 'onvoldoende' tot 5 'goed'. Met 3.55 als laagste gemiddelde waarde (voldoende tot redelijk) en 4.20 (redelijk tot goed) als hoogste gemiddelde waarde, kan gesteld worden dat de cursus meer dan voldoende heeft bijgedragen aan het behalen van de persoonlijke leerdoelen met betrekking tot didactische kennis en vaardigheden.

Samenvattend kan gesteld worden dat de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' – zoals waargenomen door de deelnemers - zowel geleid heeft tot een toename van didactische kennis en vaardigheden ten aanzien van de cursusdoelen 'helder het onderwijs uitleggen en presenteren', het cursusdoel 'didactische werkvormen kiezen en uitvoeren' het cursusdoel 'een groep studenten begeleiden' en de cursus als geheel, als tot het behalen van de persoonlijke leerdoelen. De hoge beoordelingen ten aanzien van de cursusinhoud en de aanpak van de trainer lijken deze uitkomsten te ondersteunen. Een response-shift bias is tegen verwachting in niet vastgesteld. In de discussie zal worden stilgestaan bij de evaluatie van deze cursus. Hiertoe zullen aanbevelingen gedaan worden welke gebruikt kunnen worden voor het verder vormgeven van een meer systematische evaluatie van soortgelijke cursussen.

Discussie

Evaluatie en aanbevelingen

In dit evaluatieonderzoek naar de opbrengsten van de driedaagse cursus 'Didactiek Basiskwalificatie UvA-docent' ten aanzien van de toename van didactische kennis en vaardigheden van de deelnemers, zijn een aantal factoren die de uitkomsten en conclusies kunnen hebben beïnvloed. Deze factoren zijn te herleiden tot de methode en de cursus zelf. In het licht van eventueel vervolgonderzoek is het van belang deze factoren in acht te nemen, welke hieronder toegelicht zullen worden.

Methode

De kleine omvang van de onderzoeksgroep (n=14) bleek een belangrijke factor bij de bepaling van de betrouwbaarheid en de validering van de vragenlijsten en de uitkomsten. Dit heeft er toe geleid dat de resultaten niet generaliseerbaar zijn, wat afbreuk doet aan het wetenschappelijk belang. Een poweranalyse voorafgaand aan het onderzoek, welke helpt om een minimumomvang voor een steekproef te bepalen, zou dit resultaat kunnen ondervangen. Wanneer de vragenlijsten uitgezet waren binnen de door een poweranalyse aangegeven minimumomvang van eenheden en getoetst op validiteit en betrouwbaarheid, zouden de resultaten van de metingen binnen een kleine groep cursisten tot meer significante interpretaties kunnen leiden.

Een andere factor om rekening mee te houden, is de dubbelrol van de cursist ten aanzien van het 'deelnemer' zijn en het in zekere zin afgedwongen 'onderzoeksobject' zijn. Dit kan leiden tot irritatie en verwarring, omdat de cursist vanuit een andere intentie de cursus volgt. De kans bestaat hierdoor dat deelnemers de vragenlijsten sociaalwenselijk, niet serieus of onvolledig invullen, wat niet ten goede komt aan de betrouwbaarheid. Om een dergelijke dubbelrol te ondervangen is naast een waarborging van de anonimiteit van de deelnemers, tevens een goede toelichting op het doel van het onderzoek nodig.

Ten aanzien van het instrument leidde de hoeveelheid vragen in de gesloten vragenlijst na een intens beleefde cursusdag tot vermoeide en emotionele reacties. Dit leidde - naar eigen zeggen van de deelnemers - veelal tot het onvolledig invullen en het geven van willekeurige en zogenoemde 'fake' antwoorden, welke tot een ondermijning van de betrouwbaarheid kunnen leiden. De open vragen vergden evenzoveel concentratie, zowel inhoudelijk als taaltechnisch als qua stijl. Om een cognitieve overload en het geven van onvolledige, willekeurige 'fake' antwoorden te vermijden, is een kortere vragenlijst aan te bevelen.

Een belangrijke factor die speelde ten aanzien van het instrument was de inhoud en formulering van de vragen. Inhoudelijk waren de vragen binnen de vragenlijst gericht op vermeende kennis en vaardigheden, geformuleerd als 'Ik ben in staat om' en 'Ik kan'. Het korte tijdsbestek van de cursus bleek voor veel deelnemers onvoldoende om het geleerde in de eigen beroepspraktijk tot uitvoer te brengen. Dit heeft mogelijk geleid tot vertroebelde antwoorden en resultaten. Een herformulering van de vragen is aan te bevelen.

Een laatste opmerking ten aanzien van het instrument betreft de antwoordschalen, welke mogelijkerwijs uitgebreid kunnen worden met de antwoordmogelijkheid 'niet van toepassing'. Uit de uitkomsten bleek namelijk dat niet elke deelnemer (praktijk)ervaring heeft met de vier cursusdoelen. Een aantal cursisten heeft bijvoorbeeld nog nooit beoordeeld, wat veelal de non-respons op dit cursusdoel verklaart.

Ten aanzien van dit evaluatieonderzoek is gekozen voor een kwantitatief-kwalitatatieve benadering waarbij de kwalitatieve gegevens complementair zijn gebruikt. Bij een gelijkblijvende doelgroep zou ook gekozen kunnen worden voor een volledig kwalitatieve opzet. Het onderzoek zou bij meerdere cursussen afgenomen kunnen worden, zodat er een grotere verscheidenheid ontstaat in uitkomsten. Het blijkt lastig uitspraken te doen over de leeropbrengsten van één cursus wanneer de resultaten niet vergeleken kunnen worden met die van andere cursussen. Een andere mogelijkheid is om dit onderzoek kwalitatief uit te breiden om zo meer verdieping van de resultaten te verkrijgen. Door één of meerdere diepte interviews af te nemen bij de cursisten, kan meer informatie verkregen worden over de perceptie van de cursisten. Deze informatie is te gebruiken voor het 'duiden' van de gevonden resultaten.

Zoals uit het theoretisch kader blijkt, kan het gebruik van een self-assessment een goede evaluatiemethode zijn om (korte) cursusuitkomsten te evalueren. Een kritische noot ten aanzien van self-assessments is echter dat resultaten niet generaliseerbaar zijn, gezien de zelfperceptie van deelnemers. Een self-assessment an sich levert beperkte informatie ten aanzien van de leerwinst van deelnemers en is daarmee onvoldoende geschikt om cursusuitkomsten en cursuseffectiviteit te bepalen. Een self-assessment komt het meest tot zijn recht wanneer deze uitgevoerd wordt in combinatie met andere evaluatiemethoden (Stufflebeam & Wingate, 2005).

Het gevoelsmatig niet kunnen voldoen aan beoogde verwachtingen zou een valkuil kunnen zijn ten aanzien van overschatting binnen self-assessment. Niet aanwezige kennis en vaardigheden rondom het construct didactiek zouden een schaamtegevoel kunnen oproepen, welke een te hoge inschatting weergeeft van eigen kennis en vaardigheden. Alhoewel dit schaamtegevoel niet volledig ondervangen kan worden, zou de waarborging van anonimiteit dit schaamtegevoel kunnen verminderen.

In dit onderzoek zijn een pretest-posttest design en retrospectief design naast elkaar uitgevoerd. Zoals uit het theoretisch kader naar voren komt zijn belangrijke beperkingen ten aanzien van een pretest-posttest design; het tijdsaspect, de aanwezigheid van deelnemers en de aannemelijkheid van een response shift bias, binnen deze specifieke cursus ervaren. Het bewust lager scoren bij de voormeting om de te meten toename in kennis of vaardigheid te vergroten is niet waargenomen.

Binnen de cursus stond het oefenen van vaardigheden in de praktijk centraal. Een kort tijdsbestek van drie dagen geeft onvoldoende mogelijkheden om het geleerde in de eigen beroepspraktijk te brengen en te oefenen. Voor een transfer van kennis en vaardigheden is dit een noodzakelijke voorwaarde (Van Merrienboer & Kirschner, 2007). Om meer betrouwbare uitspraken te

kunnen doen over de opbrengst van de cursus – ook in relatie tot de cursusinhoud en de trainer – kunnen een folluw-up 'meeting' en 'meting' uitkomst bieden.

Een exploratieve factoranalyse zou uitgevoerd kunnen worden om vast te stellen of op basis van de verzamelde data dezelfde set van items betekenisvolle schalen blijken te vormen. Gezien het kleine aantal deelnemers aan de cursus en de grote hoeveelheid items, welke mogelijk kunnen leiden tot een onherkenbaar aantal factoren, is gekozen om deze analyse niet uit te voeren. Een confirmerende factoranalyse na de betrouwbaarheidsanalyse is om zojuist genoemde redenen ook niet uitgevoerd.

Aandachtspunt ten aanzien van de uitgevoerde betrouwbaarheidsanalyse is de mogelijke aanwezigheid van subschalen door het niet juist hebben kunnen uitvoeren van een factoranalyse. Dit kan leiden tot multicollineariteit, items binnen de schaal hebben een te hoge consistentie, waardoor de betrouwbaarheid van de analyse in het geding kan komen.

Cursus

De cursus en daarmee de cursusinhoud, weerspiegelt in mindere mate de gestelde doelstellingen. Binnen deze specifieke cursus stond het uitwisselen van kennis en vaardigheden rond didactiek centraal. De wijze van formuleren van de cursusdoelen suggereert een te bereiken bekwaamheid in uitleggen en presenteren, het kiezen van didactische werkvormen, het begeleiden en beoordelen. Gezien de duur van de cursus is er slechts een tipje van de sluier rond een didactisch model opgelicht, het trainen en oefenen van vaardigheden is enigszins onderbelicht gebleven.

Referentielijst

Boud, D. & Falchikov, N. (1989). Quantitative studies of student self-assessment in higher education: a critical analysis of findings. *Higher Education, 18,* 529-549.

D'eon, M., Sadownik, L., Harrison, A. & Nation, J. (2008). Using self-assessments to detect workshop success*. American Journal of Evaluation 29 (1)*, 92-98.

Dijk, L. A. van (2000). *Activeren in colleges, mogelijkheden en effecten onderzocht.* Delft: Technische Universiteit Delft.

Drennan, J. & Hyde, A. (2008). Controlling response shift bias: the use of the retrospective pre-test design in the evaluation of a master's programme. *Assessment & Evaluation in Higher Education, 33 (6)*, 699-709.

Driscoll, M.P. (1994). *Psychology of learning for instruction.* Boston: Allyn and Bacon.

Gelder, L. van, Oudkerk Pool, T., Peters, J. & Sixma, J. (1979). *Didactische analyse. Studieboek.* Groningen: Wolters-Noordhoff.

Gipps, C.V. (1994). *Beyond testing. Towards a theory of educational assessment.* Londen: Falmer Press.

Hoogeveen, P. & Winkels, J. (2009). *Het didactische werkvormenboek. Variatie en differentiatie in de praktijk.* Assen: Van Gorcum.

Hoogstraten, J. (1982). The retrospective pre-test in an educational training context. *Journal of Experimental Education, 50,* 200-204.

Howard, G. (1980). Response shift bias: A problem in evaluating interventions with pre/post self-reports. *Evaluation Review, 4*, 93-106.

Howard, G. & Dailey, P. (1979). Response shift bias: A source of contamination of self-report measures. *Journal of Applied Psychology, 64*, 144-150.

Kallenberg, T., Grijspaarde, L. van & Braak, A. ter (2009*). Leren (en) doceren in het hoger onderwijs.* Den Haag: Lemma.

Lagerweij, N. & Lagerweij-Voogt, J. (2004). *Anders kijken. De dynamiek van een eeuw onderwijsverandering.* Apeldoorn: Garant.

Lam, T.C.M. (2009). Do self-assessments work to detect workshop success? An analysis of argument and recommendation by D'eon et al. *American Journal of Evaluation, 30 (1)*, 93-105.

Lam, T.C.M. & Bengo, P. (2003). A comparison of three retrospective self-reporting methods of measuring change in instructional practice. *American Journal of Evaluation, 24 (1),* 65–80.

Marshak, H.H., Silva, P. de & Silberstein, J. (1998). Evaluation of a peer-thaught nutrition education program for low-income parents. *Journal of Nutrition Education, 30 (5),* 314-322.

Mckeachie, W.J. & Svinicki, M. (2006). *Teaching Tips. Strategies, Research, and Theory for College and University Teachers.* Boston New York: Houghton Mifflin Company.

Merrienboer, J.J.G. van & Kirschner, P.A. (2007). *Ten steps to complex learning. A systematic approach to four-component instructional design.* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Milius, J., Oost, H. & Holleman, W. (2001). *Werken aan academische vorming. Ideeën voor actief leren in de onderwijspraktijk.* Utrecht: Zuidam & Uithof.

Mosca, J., Paul, D. & Skiba, M. (2006). Technology Increases Students' Attention**.** *Journal of College Teaching & Learning, 3 (8),* 7-14.

Onderwijsraad. (n.d.) Opgeroepen op 22 04, 2010 van http://www.onderwijsraad.nl/dossiers/een-leven-lang-leren.

Pratt, C.C., McGuigan, W.M., & Katzev, A.R. (2000). Measuring program outcomes: Using retrospective pretest methodology. *American Journal of Evaluation, 21 (3)*, 341-349.

Schaaf, M.F. van der, & Stokking, K.M. (1999). *Vaardig in het begeleiden van zelfstandig leren? Constructie van een monitoringsintrument voor het begeleiden van zelfstandig werken en leren van leerlingen door docenten.* Utrecht: ISOR.

Sol, Y. & Stokking, K. (2008). *Het handelen van docenten in scholen met een vernieuwend Onderwijsconcept. Ontwikkeling, gebruik en opbrengsten van een instrument.* Utrecht: Universiteit Utrecht, IVLOS en Onderwijskunde.

Stokking, K. (2003). *Organiseren van ontwikkeling tussen vraag en aanbod. Notities bij de Bachelor-Master ambities van de Universiteit Utrecht.* Utrecht: Universiteit Utrecht.

Stufflebeam, D.L. & Wingate, L.A. (2005). A Self-Assessment Procedure for Use in Evaluation Training. *American Journal of Evaluation, 26,* 544-561.

VSNU. (n.d). Opgeroepen op 22 03, 2010, van http://www.vsnu.nl/web/show

Woolfolk, A., Hughes, M. & Walkup, V. (2008). *Psychology in Education.* Harlow Essex England: Pierson Education Limited.