

University of Groningen

## Teacher evaluation through observation

van der Lans, Rikkert

**IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.**

*Document Version*

Publisher's PDF, also known as Version of record

*Publication date:*

2017

[Link to publication in University of Groningen/UMCG research database](#)

*Citation for published version (APA):*

van der Lans, R. (2017). *Teacher evaluation through observation: Application of classroom observation and student ratings to improve teaching effectiveness in classrooms*. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

### Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

### Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

# Samenvatting



## Algemene introductie

De leraar speelt een grote rol in het leren van leerlingen. Mede om deze reden leggen politici en beleidsmakers steeds meer nadruk op het evalueren van leraren als middel om het onderwijs te verbeteren. Evaluatie dient daarin twee doelen. Ten eerste zou evaluatie moeten leiden tot feedback waarmee leraren hun lesgeven kunnen verbeteren. Ten tweede zou evaluatie moeten bijdragen aan besluiten over leraren met betrekking tot bijvoorbeeld salarisschalen, contractverlening of, in extreme gevallen, ontslag. Het geven van feedback en het nemen van zulke evaluatieve besluiten is niet nieuw. Alle scholen moeten immers personeelsbeleid voeren. Nieuw in het beleid rondom evaluatie van leraren is de relatief grote nadruk op hun bekwaamheid in lesgeven en ook nieuw is de hogere frequentie waarin dit zou moeten worden geëvalueerd (het streven is dat iedere docent jaarlijks een functioneringsgesprek heeft<sup>3</sup>). Met deze nieuwe focus is het belangrijk geworden om de bekwaamheid in lesgeven valide en betrouwbaar te evalueren. Wanneer evaluaties onbetrouwbaar zijn dan is er een onacceptabel grote kans dat leraren verkeerde feedback ontvangen of dat er over hen verkeerde evaluatieve besluiten worden genomen. Verkeerde feedback zal bovendien niet leiden tot een verbetering in lesgeven.

In dit proefschrift wordt de validiteit en betrouwbaarheid onderzocht van de feedback aan leraren en de evaluatieve besluiten over leraren op basis van lesobservaties en leerlingenvragenlijsten. In lijn met de duale doelstelling van de evaluatie van leraren worden er twee verschillende vragen van validiteit behandeld. Om informatie uit lesobservaties en leerlingenvragenlijsten te kunnen gebruiken om leraren feedback te geven is het belangrijk om wetenschappelijk bewijs te vinden hoe de vaardigheid in lesgeven zich ontwikkelt. Alleen wanneer evaluatie van leraren gekoppeld is aan een algemene theorie die voorspelt hoe de vaardigheid van lesgeven van een leraar zich zal ontwikkelen kan de informatie over het lesgeven van een leraar worden gerelateerd aan een concreet advies voor verdere ontwikkeling. Daarom is in 3 verschillende steekproeven onderzocht hoe vaardigheid van lesgeven zich in het algemeen ontwikkelt bij docenten. Dit onderzoek naar de algemene ontwikkeling wordt beschreven in Hoofdstuk 2, 3, en 4. Om de informatie uit lesobservaties en leerlingenvragenlijsten te gebruiken in functioneringsgesprekken is het belangrijk wetenschappelijk aan te tonen of en onder welke voorwaarden deze informatie hiervoor betrouwbaar genoeg is. Goede besluitvorming vereist dat er voldoende informatie

---

<sup>3</sup> Zie bijvoorbeeld: <https://www.delerarenagenda.nl/de-lerarenagenda/scholen-als-lerende-organisaties>

is verzameld vanuit verschillende gezichtspunten (zodat falsificatie mogelijk is) en dat inaccurate informatie wordt opgemerkt en niet wordt meegewogen. Om deze reden worden in Hoofdstuk 5 en 6 vraagstukken rondom de betrouwbaarheid beschreven.

Vanuit bovenstaande is de volgende hoofdvraag geformuleerd die vervolgens is opgedeeld in 5 deelvragen.

*Hoe kunnen lesobservaties en leerlingenvragenlijsten op zo'n manier worden ingezet in scholen dat ze leraren op een valide en betrouwbare manier van feedback kunnen voorzien en dat scholen op een valide en betrouwbare manier evaluatieve besluiten over leraren kunnen nemen?*

Deelvragen:

1. Kunnen observaties van het pedagogisch en didactisch handelen van docenten tijdens de les stapsgewijs geordend worden, zodat minder complexe handelingen eerst getoond moeten worden alvorens meer complexe handelingen kunnen worden getoond? En: Hoe ziet deze stapsgewijze ordening eruit? **(Hoofdstuk 2)**
2. Kunnen leerlingenantwoorden op items die refereren naar het pedagogisch en didactisch handelen tijdens de les stapsgewijs geordend worden? En: Hoe ziet deze stapsgewijze ordening eruit? **(Hoofdstuk 3)**
3. In hoeverre is er overeenstemming tussen lesobservatoren en leerlingen in hoe zij de stapsgewijze ontwikkeling indelen? En: vinden lesobservatoren en leerlingen gelijkwaardig gedrag ongeveer even complex? **(Hoofdstuk 4)**
4. Hoeveel lesobservaties door lesobservatoren zijn noodzakelijk voor een bescheiden betrouwbaarheid, voldoende om te gebruiken voor feedback aan leraren? En hoeveel lesobservaties door lesobservatoren zijn noodzakelijk voor een hoge betrouwbaarheid, voldoende om te gebruiken om de informatie te gebruiken in functioneringsgesprekken? **(Hoofdstuk 5)**
5. Zijn er leraren die zich anders ontwikkelen dan de stapsgewijze volgorde beschreven in hoofdstuk 2, 3 en 4? En: Hoe kunnen we deze unieke gevallen herkennen? **(Hoofdstuk 6)**

## Hoofdstuk 2 en 3

De hoofdstukken 2 en 3 richten zich op hoe de vaardigheid in lesgeven zich ontwikkelt. De hoofdstukken zijn gebaseerd op dezelfde theorie en in deze hoofdstukken worden ongeveer dezelfde methoden gebruikt. Daarom worden ze hier gezamenlijk samengevat.

### Introductie

Onderwijsbeleid richt zich steeds meer op lerarenevaluatie als een middel om leraren feedback te geven over hun functioneren. Daarom is een aantal instrumenten ontwikkeld waarmee scholen de vaardigheid van lesgeven kunnen evalueren. Er is echter kritiek op deze instrumenten omdat alleen het terugkoppelen van de huidige vaardigheid in lesgeven leraren weinig informatie biedt over hoe ze zich kunnen verbeteren en doorontwikkelen. Net zoals leerlingen in stappen de lesstof leren lijkt het aannemelijk dat leraren leren lesgeven in stappen. Feedback zou daarom beter gericht kunnen worden op waar in de ontwikkeling de leraar zich nu bevindt en in welke pedagogische en didactische handelingen de leraar zich zou moeten bekwaamen om een stap verder te komen.

Om meer inzicht in de stappen te verkrijgen wordt in de hoofdstukken 2 en 3 bestudeerd of observaties van het pedagogisch en didactisch handelen van de leraar tijdens de les stapsgewijs (cumulatief) kunnen worden geordend en of deze ordening overeenstemt met de eerdere bevindingen beschreven in theorie over de ontwikkeling van leraren. Onder observaties worden zowel lesobservaties door collega's of inspecteurs verstaan (Hoofdstuk 2) als observaties door leerlingen (Hoofdstuk 3). Op basis van eerder onderzoek wordt de hypothese geformuleerd dat de pedagogische en didactische vaardigheid in lesgeven zich ontwikkelt in grofweg de volgende zes stappen: (1) veilig en stimulerend leerklimaat, (2) efficiënte lesorganisatie en (3) duidelijke en gestructureerde instructie, (4) intensieve en activerende les, (5) leerstrategieën aanleren en (6) afstemmen van instructie. Hierbij is de hypothese ook dat een zeker mate van bekwaamheid in de handelingen beschreven in een lagere stap voorwaardelijk is om bekwaam te worden in de handelingen beschreven in een hogere stap (bijvoorbeeld kundigheid in stap 1 is voorwaardelijk voor kundigheid in stap 2). Als deze hypothese klopt dan zouden observaties van lesgeven grofweg het patroon moeten volgen, zoals weergegeven in Figuur S1. In deze Figuur is een vinkje genoteerd in het geval dat een docent bekwaam is in de handelingen passend bij deze stap. Het patroon geeft weer dat sommige docenten zich al in meer stappen bekwaam tonen dan collega-

docenten. Het is ook duidelijk dat geen van de docenten zich bekwaam toont in een hogere stap zonder dat ze zich ook bekwaam tonen in een lagere stap.

**Figuur S1**

De veronderstelde stapsgewijze ontwikkeling van leraren. Een vinkje betekent dat een leraar de (meeste) handelingen behorend bij deze stap voldoende beheerst.

	veilig en stimulerend klimaat	efficiënt les- organisatie	duidelijke en gestructureerde instructie	intensieve en activerende les	leerstrategieën aanleren	afstemmen van instructie
Docent A	✓					
Docent B	✓	✓				
Docent C	✓	✓	✓			
Docent D	✓	✓	✓	✓		
Docent E	✓	✓	✓	✓	✓	
Docent F	✓	✓	✓	✓	✓	✓

*Method*

Voor het onderzoek in Hoofdstuk 2 zijn 878 lesobservaties uitgevoerd op 119 scholen verspreid over Nederland. Van alle observaties is 46.6% uitgevoerd door inspecteurs van de onderwijsinspectie, de overige 53.4% is uitgevoerd door collega-leraren. Alle observaties zijn uitgevoerd met hetzelfde instrument: de ICALT. Het ICALT-lesobservatie formulier bestaat uit 32 items (zie appendix D). De lesobservaties zijn uitgevoerd in het schooljaar 2010-2011. In de analyse wordt gebruik gemaakt van het Rasch-model dat specifiek ontwikkeld is om cumulatieve item volgorde te toetsen.

Voor het onderzoek in Hoofdstuk 3 zijn 1590 leerlingenvragenlijsten geanalyseerd die zijn ingevuld voor 68 leraren werkzaam op één school. Deze vragenlijsten zijn afgenomen in het schooljaar 2011-2012. De zogenoemde “Mijn Leraar” vragenlijst, die is gebruikt voor dit onderzoek bestond in een eerste versie uit 59 items. Items beschrijven een specifieke handeling van de leraar en de observator of leerling wordt gevraagd of de handeling in voldoende mate wordt uitgevoerd of niet (voor de specifieke instrumenten zie

appendix D en E). Om te onderzoeken of we de pedagogische en didactische handelingen kunnen ordenen volgens een stapsgewijs patroon is gebruik gemaakt van het Rasch-model.

### **Resultaten**

Uit de resultaten komt naar voren dat 31 van de 32 items van de ICALT passen binnen het stapsgewijs patroon en dat 32 van de 59 vragen op de “Mijn Leraar” vragenlijst passen binnen een stapsgewijs patroon. De volgorde van de handelingen komt grotendeels overeen met de hypothese van de zes stappen in ontwikkeling: (1) *veilig en stimulerend leerklimaat*, (2) *efficiënte lesorganisatie* (3) *duidelijke en gestructureerde instructie*, (4) *intensieve en activerende les*, (5) *leerstrategieën aanleren* en (6) *afstemmen van instructie*. Toch blijkt wel dat de stappen niet helemaal los van elkaar kunnen worden gezien. Het lijkt bijvoorbeeld niet zo te zijn dat een leraar zich eerst compleet moet bekwamen in alle handelingen behorend bij stap 2 voordat de leraar zich kan bekwamen in de pedagogische en didactische handelingen uit stap 3.

### **Conclusie & Implicaties**

Zowel de ICALT-lesobservaties als de “Mijn Leraar” leerlingenvragenlijst bevestigen de hypothese. Daarmee bieden deze studies ondersteunend bewijs voor de stelling dat de ontwikkeling in pedagogische en didactische vaardigheid van leraren zich cumulatief ontwikkelt.

Met dit onderzoek is meer inzicht verkregen in de ontwikkeling in pedagogische en didactische vaardigheid van leraren. Eerdere inzichten waren meestal beperkt tot minder ervaren leraren (<3 jaar), terwijl dit onderzoek leraren van alle ervaringsjaren meegenomen heeft. Met de gevonden stapsgewijze ontwikkeling in pedagogisch-didactische handelingen is het bovendien mogelijk om preciezer feedback te geven aan leraren, zodat duidelijk kan worden aangegeven in welke pedagogische en didactische handelingen de leraar al vaardig is, welke handelingen de leraar nu als eerste zou moeten gaan leren en welke handelingen vooralsnog te moeilijk lijken voor de leraar om te leren.

## **Hoofdstuk 4**

### **Introductie**

Evaluatie in scholen kent een variëteit aan methoden. Aan de ene kant is deze variëteit wenselijk omdat het vanuit verschillende invalshoeken informatie kan bieden over het



lesgeven van een specifieke leraar. Aan de andere kant blijken de methoden vaak heel verschillende informatie te bieden waarvan onzeker is of deze met elkaar vergeleken of gecombineerd mag worden.

De evaluatiesystematiek die is uitgevoerd in het kader van dit proefschrift maakt gebruik van twee evaluatiemethodes: Lesobservaties en leerlingenvragenlijsten. Sommige wetenschappers hebben gesteld dat het vergelijken van lesobservaties met leerlingenvragenlijsten wat weg heeft van het vergelijken van appels met peren. Ze stellen dat leerlingenvragenlijsten en lesobservaties tot compleet verschillende inzichten over lesgeven kunnen leiden. Anderen stellen het tegenovergestelde, namelijk dat deze twee methodes vergelijkbare informatie bieden en juist gecombineerd moeten worden tot één evaluatie. Zij bieden bewijs voor deze stelling door te laten zien dat evaluaties gebaseerd op een combinatie van lesobservaties en leerlingenvragenlijsten meer betrouwbaar zijn dan evaluaties op basis van enkel lesobservaties.

Uit het onderzoek in Hoofdstuk 2 en 3 is al gebleken dat zowel lesobservatoren als leerlingen een soortgelijke stapsgewijze volgorde aan pedagogische en didactische handelingen toekennen. Dit betekent dat zowel leerlingen als lesobservatoren ‘veilig en stimulerend leerklimaat’ het minst complex vinden (zodat dit het gemakkelijkste als eerste geleerd kan worden), terwijl ze beiden differentiatie en lesgeven in leerstrategieën het meest complex vinden. In dit onderzoek wordt de vergelijkbaarheid van de leerlingenvragenlijst en het lesobservatieformulier verder onderzocht. Om dit na te gaan, worden de items van de vragenlijst en van de lesobservaties samen geanalyseerd alsof ze van hetzelfde instrument komen. Het is dan mogelijk om te onderzoeken of leerlingen en lesobservatoren alle stappen ongeveer even complex vinden, of dat lesobservatoren alle stappen minder/meer complex vinden dan leerlingen. Bijvoorbeeld, zou het kunnen dat lesobservatoren en leerlingen dezelfde stapsgewijze volgorde observeren, maar dat lesobservatoren de stap (2) *efficiënte lesorganisatie* net zo complex vinden als de leerlingen de stap (4) *intensieve en activerende les*? In dit voorbeeld zouden leerlingen hun leraren altijd twee stappen hoger evalueren dan dat de lesobservatoren doen. Dit zou betekenen dat de lesobservatoren en leerlingen geen overeenkomstige interpretatie geven aan de afzonderlijke handelingen en dus dat de resultaten uit de leerlingenvragenlijst niet zomaar kunnen worden opgeteld bij de resultaten van de lesobservaties.

## Methode

Voor dit onderzoek is een steekproef gebruikt die bestond uit 269 lesobservaties en 2.876 leerlingenvragenlijsten. De lesobservaties vonden plaats in de klas die ook de leerlingenvragenlijst invulde zodat de observatoren dezelfde lessen observeerden als de leerlingen. Deze data zijn verkregen in de schooljaren 2013-2014 en 2014-2015. Het ICALT-lesobservatieformulier kent 32 items, waarvan 31 items werden meegenomen in de analyse (zie Hoofdstuk 2). De gebruikte versie van de “Mijn Leraar” vragenlijst kent 40 items waarvan – op basis van de resultaten in Hoofdstuk 3 – 28 items zijn meegenomen in de analyse (zie Appendix E). Om te toetsen of de handelingen wanneer gevraagd aan leerlingen en aan lesobservatoren een gelijke stapsgewijze ontwikkeling laten zien is een multiniveau Rasch analyse uitgevoerd.

## Resultaten

De resultaten geven aan dat de 31-items van de ICALT en de 28-items van de leerlingenvragenlijst samengenomen kunnen worden tot een eendimensionale stapsgewijze volgorde. Wel blijkt dat een aantal items op de “Mijn Leraar” vragenlijst gerelateerd aan leerstrategieën de ordening wat verstoren. Ook blijkt dat de handelingen gerelateerd aan de stap (of het domein) “differentiatie” wanneer geobserveerd met de ICALT te vaak dezelfde score krijgt. Dus alhoewel de items in de stap differentiatie verschillende handelingen beschrijven worden deze verschillen te weinig opgemerkt door de observatoren. De leerlingen merken meer verschillen op in de handelingen binnen de stap differentiatie.

Een tweede belangrijk resultaat is dat ondanks dat de observatoren en leerlingen grotendeels overeenstemmen in hun interpretatie van de items, ze wel heel verschillend kunnen denken over of de leraar ook bekwaam is in een pedagogisch of didactische handeling beschreven door dat item. Wanneer een evaluatieprocedure wordt gebruikt waarin de observator slechts één les bezoekt en de uitkomsten van deze ene lesobservatie worden vergeleken met de observaties van leerlingen die alle lessen hebben gezien dan is de correlatie laag ( $r = .26$ ).

## Conclusie & Implicaties

De studie in Hoofdstuk 4 biedt aanvullend empirisch bewijs dat lesobservatoren en leerlingen wel degelijk vergelijkbare informatie *kunnen* verschaffen, waaruit geconcludeerd kan worden dat lesobservaties en leerlingenvragenlijsten geen ‘appels en peren’ zijn. Toch

biedt het ook aanvullend bewijs dat niet verwacht kan worden dat één lesobservatie een voldoende representatief beeld geeft van de bekwaamheid van de leraar in lesgeven (zie ook Hoofdstuk 5), waardoor leerlingen – die veel meer lessen hebben gezien – de leraar behoorlijk anders kunnen evalueren.

Met dit onderzoek is een verdere stap gezet in hoe scholen gegevens verkregen met verschillende evaluatiemethoden zouden kunnen combineren. Het blijkt mogelijk om de gegevens verkregen met verschillende methoden op eenzelfde meetschaal te plaatsen. Sommige evaluatie methoden kunnen gemakkelijker en goedkoper op grotere schaal worden ingezet dan anderen. Een mogelijke implicatie van dit onderzoek is dat zulke meer efficiënte methoden zouden kunnen worden gebruikt om een globaal overzicht te krijgen in welke lessen en bij welke leraren de inzet van duurdere (maar ook voor de betreffende docent meer informatieve) vormen van evaluatie waardevol lijkt.

### **Hoofdstuk 5**

#### **Introductie**

Wanneer scholen docenten feedback willen geven ofwel hun functioneren willen beoordelen dan is het belangrijk om na te gaan of de evaluaties, waarop deze feedback en besluiten worden gebaseerd, betrouwbaar zijn. Wanneer evaluaties betrouwbaar zijn, dan betekent dit dat de uitkomsten gerepliceerd kunnen worden. Dus wanneer een andere observator een andere les zou hebben geobserveerd, zou de uitkomst voor de betreffende leraar niet anders uitvallen. Wanneer een evaluatie te weinig betrouwbaar is, dan neemt een school een onacceptabel risico dat het een leraar verkeerde feedback teruggeeft of dat zij op oneigenlijke gronden evaluatieve besluiten neemt over de leraar. Vanuit het voorgaande volgt dat er een norm gesteld dient te worden voor wat ‘te weinig’ betrouwbaar is. Het is ook duidelijk dat deze norm anders ligt voor feedback – wat doorgaans geen directe persoonlijke consequenties heeft – dan voor evaluatieve besluiten – wat doorgaans wel directe persoonlijke consequenties hebben. Op basis van deze redentatie zijn er twee criteria gesteld: een betrouwbaarheid van .70 voldoet voor feedback, een betrouwbaarheid van .90 voldoet voor evaluatieve besluiten.

#### **Methode**

De studie is gebaseerd op 198 lesobservaties uitgevoerd door 62 lesobservatoren bij 69 leraren werkzaam op 8 verschillende scholen. De lesobservaties zijn gedaan met het

ICALT-observatie instrument dat 32 items telt. Op basis van de resultaten in Hoofdstuk 2 is besloten om 31 items in de analyse mee te nemen.

De analyse van betrouwbaarheid combineert de principes van generaliseerbaarheidstheorie (G-theorie) met de principes van de item response theorie (IRT), tezamen wordt verwezen naar deze analysetechniek als GIRT.

### **Resultaten**

De resultaten tonen aan dat één lesobservatie onvoldoende betrouwbaar is om leraren van feedback te voorzien en zeker onvoldoende betrouwbaar is om te gebruiken voor de onderbouwing van besluiten over leraren in functioneringsgesprekken. Wanneer vier lessen worden bezocht door vier verschillende observatoren is hun gemiddelde uitkomst voldoende betrouwbaar om leraren feedback te geven (hoger dan .70). Na meer dan 10 verschillende lesbezoeken begint de stijging in betrouwbaarheid nihil te worden. De extra lesbezoeken brengen dus weinig nieuwe informatie in. Ook bij 10 lesbezoeken is het gemiddelde niet betrouwbaar genoeg om te gebruiken als input op een functioneringsgesprek (d.w.z. de betrouwbaarheid blijft beneden de .90).

### **Conclusies en Implicaties**

Scholen worden geadviseerd minimaal vier lesbezoeken door vier verschillende observatoren te verzamelen alvorens de leraar feedback te geven. Wanneer scholen dit zonder gebruik te maken van eventueel aanvullende informatie niet doen, nemen ze een onacceptabel groot risico dat ze leraren verkeerde feedback geven op basis waarvan de leraar zich niet kan verbeteren. Als gevolg zal de leraar misschien deelnemen aan trainingen of meedoen in intervisiesessies met verminderde kans op resultaat. Scholen worden verder geadviseerd niet meer dan 10 lesbezoeken per jaar te doen. Het effect van nog meer lesbezoeken is verwaarloosbaar. Het is beter om in plaats van nog meer lesbezoeken andere methodes voor evaluatie in te zetten. Een combinatie van lesobservatie en andere evaluatiemethodes is noodzakelijk om te zorgen dat informatie betrouwbaar genoeg is voor besluiten over het functioneren van de leraar. Wederom geldt dat de school een onacceptabel risico loopt om op verkeerde grond evaluatieve besluiten te nemen over leraren wanneer ze deze baseren op bijvoorbeeld alleen lesbezoeken, ongeacht hoeveel.

### **Hoofdstuk 6**

#### **Introductie**

Sommige van de theorieën over de ontwikkeling van leraren berust op de aanname dat er (relevante) individuele verschillen zijn in de ontwikkeling. Hierbij is soms gesuggereerd dat er ‘exceptionele’ leraren zijn die een talent hebben voor het vak waardoor ze zich het vak op een andere manier eigen maken dan hun minder getalenteerde collega’s. Gezien de implicaties die evaluaties kunnen hebben voor individuele leraren is het een morele verplichting na te gaan of er sprake is van individuele verschillen in de ontwikkeling van lesgeven. Als er reden is om aan te nemen dat sommige leraren zich op een exceptionele manier ontwikkelen dan zal de feedback op basis van de gewone ontwikkeling – zoals gespecificeerd in Hoofdstuk 2 en 3 – voor deze leraren niet bijdragen aan hun ontwikkeling. Een leraar die zich exceptioneel ontwikkelt wordt in deze studie gedefinieerd als een leraar die eerst de meer complexere vaardigheden ontwikkelt (bijv. aanleren van leerstrategieën of het differentiëren) voordat de basale vaardigheden zijn ontwikkeld (bijv. een efficiënt klassenmanagement).

#### **Methode**

De steekproef bestaat uit dezelfde 198 lesobservaties die ook in Hoofdstuk 5 zijn onderzocht. Eerst werd van alle lessen vastgesteld of er in deze les relatief complexe pedagogische en didactische handelingen werden verricht terwijl in diezelfde les de minder complexe handelingen niet werden geobserveerd. Zulke lessen werden gediagnosticeerd als ‘afwijkend’. Vervolgens is nagegaan of lessen waarin sprake was van afwijkend beeld, telkens werden gedoceerd door dezelfde leraren. De analyse – uitgevoerd met de  $G_{NORMED}$  person fit statistiek – kan niet worden uitgevoerd voor lesobservaties waarin één item een missende waarde had. Hierdoor konden er van de 198 lesobservaties 141 worden meegenomen. Er is wel nagegaan of verwacht mag worden dat de resultaten wijzigen wanneer de lesobservaties met een missende waarde wel zouden worden meegenomen. Dit lijkt niet het geval.

#### **Resultaten**

Van alle 141 lessen werd 15% gediagnosticeerd als afwijkend, maar er kon geen overtuigend bewijs worden gevonden dat afwijkende lessen vaker gedoceerd werden door dezelfde groep ‘exceptionele’ leraren. De 21 (15%) afwijkende lessen werden gedoceerd

door 18 verschillende leraren. Het lijkt er daarom op dat iedere leraar incidenteel een afwijkende les kan hebben.

### **Conclusies & Implicaties**

De resultaten bevestigen eerdere resultaten dat sommige lessen afwijken of ‘exceptioneel’ zijn. Wat er gebeurt in die lessen stemt niet overeen met wat op basis van theorie over de ontwikkeling van leraren wordt voorspeld. Het is op basis van deze steekproef mogelijk dit resultaat verder te analyseren. Hieruit blijkt dat er weinig bewijs is dat zulke afwijkende lessen vaker voorkomen bij een groep ‘exceptionele’ leraren die zich dan anders zouden ontwikkelen dan de reguliere leraar. Daarmee bieden de resultaten weinig bewijs voor de stelling dat er relevante individuele verschillen zijn in de ontwikkeling van leraren waarmee rekening gehouden zou moeten worden in de feedback.

Omdat het erop lijkt dat afwijkende lessen bij iedere leraar kunnen voorkomen is het belangrijk dat scholen leren zulke lessen adequaat te herkennen en verwijderen om zodoende te voorkomen dat deze incidentele afwijkingen de validiteit van de feedback en evaluatieve besluiten verminderen. Deze conclusie bevestigt wat we in hoofdstuk 5 al concludeerde, namelijk dat het nodig is om meerdere lesobservaties bij iedere leraar uit te voeren.

### **Algemene conclusie proefschrift**

De centrale onderzoeksvraag van dit proefschrift is:

*Hoe kunnen lesobservaties en leerlingenvragenlijsten op zo'n manier worden ingezet in scholen dat ze leraren op een valide en betrouwbare manier van feedback kunnen voorzien en dat scholen op een valide en betrouwbare manier evaluatieve besluiten over leraren kunnen nemen?*

Op basis van de deelstudies kan worden geconcludeerd dat zowel lesobservaties als leerlingenvragenlijsten valide en betrouwbare feedback kunnen geven over het lesgeven van leraren. Zowel de “Mijn Leraar” vragenlijst als de ICALT-lesobservaties tonen eenzelfde stapsgewijze volgorde van het ontwikkelen van pedagogisch en didactische handelingen van leraren. Dit maakt het mogelijk om leraren feedback te geven over in welke vaardigheden ze al bekwaam zijn, welke vaardigheden ze zich nu eerst moeten

proberen te bekwalen, en welke vaardigheden vooralsnog te complex zijn om zich in te bekwalen. Op deze manier kan meer specifieke feedback gegeven worden.

Op basis van de deelstudies kan ook worden geconcludeerd dat één enkele lesobservatie onvoldoende is om leraren van feedback te voorzien. Wanneer feedback wordt gegeven op basis van één lesobservatie dan is de kans onacceptabel groot dat een leraar verkeerde feedback ontvangt. Dit kan ertoe leiden dat een leraar bijvoorbeeld wordt verteld bekwaam te zijn in de eerste 2 stappen, zodat zijn leren zich zou moeten richten op stap 3, terwijl de leraar eigenlijk bekwaam is in de eerste 4 stappen. Zulke feedback kan leiden tot het verkeerd besteden van professionaliseringsgelden en ook tot demotivatie bij de betreffende leraar. Betrouwbare feedback vereist minimaal 4 lesobservaties door 4 verschillende observatoren.

Het lijkt ook mogelijk om lesobservaties in te zetten voor functioneringsgesprekken. Echter, dit moet wel op een verantwoorde, dus betrouwbare, wijze gebeuren. Wanneer een schoolbestuur op een betrouwbare manier besluiten wil nemen over haar personeel, dan wordt zij geadviseerd om meer dan 10 lesobservaties uitgevoerd door verschillende observatoren te verzamelen in combinatie met meerdere leerlingenvragenlijsten. Dit lijkt niet mogelijk binnen één schooljaar. We adviseren daarom om deze hoeveelheid te spreiden over enkele jaren. In de tussentijd kunnen de jaarlijks verzamelde lesobservaties wel worden ingezet voor het geven van feedback tijdens besprekingen over de ontwikkeling van een leraar, bijvoorbeeld in de vorm van een voortgangsgesprek.