

# **Effecten van selectie ten behoeve van de lerarenopleidingen**

**Een literatuurreview in opdracht van NRO**

**Roeland van der Rijst, Dineke Tigelaar, Jan van Driel,  
ICLON, Universiteit Leiden**

**Marco Snoek, Marloes van Verseveld  
Kenniscentrum Onderwijs en Opvoeding Hogeschool van Amsterdam**



Deze verkenning is in opdracht van NRO uitgevoerd door:

Roeland van der Rijst, Dineke Tigelaar en Jan van Driel (Universiteit Leiden, ICLON)  
Marco Snoek en Marloes van Verseveld (Hogeschool van Amsterdam, Kenniscentrum Onderwijs en Opvoeding)

December 2014

© Universiteit Leiden / Hogeschool van Amsterdam



Universiteit Leiden  
ICLON  
Postbus 905  
2300 AX Leiden  
[Driel@iclon.leidenuniv.nl](mailto:Driel@iclon.leidenuniv.nl)  
Tel. 071-5274027



Hogeschool van Amsterdam  
Kenniscentrum Onderwijs en Opvoeding  
Postbus 2009  
1000CA Amsterdam  
[M.Snoek@hva.nl](mailto:M.Snoek@hva.nl)  
Tel. 06-53649824

## Inhoudsopgave

Inleiding .....	4
Methode.....	7
Resultaten van de literatuurreview.....	16
Overzicht van criteria die onderzocht zijn bij selectie voor de lerarenopleidingen.....	16
Overzicht van studies naar de effecten van integrale selectieprocedures bij lerarenopleidingen...	20
Discussie en conclusie .....	23
Implicaties voor de Nederlandse praktijk .....	28
Referenties .....	30
Bijlage 1: Uitgebreide samenvatting van studies naar integrale selectieprocedures .....	32

# Effecten van selectie ten behoeve van de lerarenopleidingen

## Een literatuur review

### Inleiding

De toegang tot het Nederlandse hoger onderwijs wordt in beginsel bepaald door de vooropleiding van aankomende studenten. Een VWO- of HAVO-diploma met een bepaald profiel volstaat in de regel om toegelaten te worden tot een universitaire respectievelijk HBO-opleiding. Uitzonderingen hierop worden gevormd door opleidingen met een beperkt aantal plaatsen (numerus fixus), of opleidingen die specifieke kwaliteiten vergen die bij de examinering in het voortgezet onderwijs niet of nauwelijks worden meegenomen. Opleidingen met een numerus fixus, zoals geneeskunde, hanteerden jarenlang een systeem van gewogen loting, waarbij kandidaten met een hoger gemiddelde examencijfer in het VWO een grotere kans hadden om ingeloot te worden. Dit systeem wordt binnenkort vervangen door een systeem waarbij de opleidingen zelf mogen bepalen op welke gronden zij studenten selecteren. Met deze decentrale selectie bestaat al op enkele plaatsen ervaring. Opleidingen die speciale kwaliteiten vergen komen vooral voor in de domeinen kunst en sport, maar ook bij exclusieve programma's zoals de University colleges. Bij zulke opleidingen worden aankomende studenten beoordeeld op talenten en kwaliteiten die nodig zijn om de betreffende opleiding met succes te kunnen volgen. In al deze gevallen staat de vraag centraal hoe studenten het beste geselecteerd kunnen worden op basis van hun geschiktheid voor de betreffende opleiding, en de hiermee verbonden beroepen.

Wat betreft de lerarenopleidingen voor basis- en voortgezet onderwijs, was de toelating tot voor kort een kwestie van de juiste vooropleiding. De laatste jaren is hierin verandering gekomen, vooral vanwege klachten over het gebrekkige kennisniveau van sommige instromende studenten. Zo zijn er diverse maatregelen genomen om het startniveau van studenten die aan de lerarenopleiding voor het basisonderwijs (Pabo) beginnen, te verhogen. Reken- en taaltoetsen zijn ingevoerd als bindend onderdeel van de propedeuse en recent zijn eisen gesteld ten aanzien van het kennisniveau met betrekking tot aardrijkskunde, geschiedenis en natuur & techniek. In de discussies over de kwaliteit van de leraar en de lerarenopleiding klinkt echter regelmatig een pleidooi om verdere stappen te zetten ten aanzien van selectie voor de poort van de lerarenopleidingen: niet alleen bij de pabo, maar ook bij de tweede- en eerstegraads lerarenopleidingen, en niet alleen ten aanzien van vakinhoudelijke eisen, maar ook ten aanzien van geschiktheid voor het beroep.

In beschouwingen rond selectie en instroomeisen voor de lerarenopleidingen lopen vaak drie perspectieven door elkaar heen. In het eerste perspectief staat de match tussen de student en het toekomstige beroep centraal. Centrale vraag binnen dit perspectief is welke kwaliteiten het beroep vraagt, welke daarvan tijdens de opleiding niet of nauwelijks te ontwikkelen zijn en of het mogelijk is om op voorhand studenten die die bepaalde kwaliteiten missen uit te sluiten van de opleiding. In het tweede perspectief staat verbetering van het toekomstig beroep centraal. Achterliggende gedachte is dat door de kwaliteit van de instroom te verhogen het eindniveau van de opleiding ook verhoogd wordt wat een kwaliteitsimpuls zou betekenen voor het onderwijs. Een selectievere toelating van alleen de beste studenten tot de lerarenopleiding zou ook bijdragen aan een hogere status, beter imago en aantrekkelijker beroep. In het derde perspectief staat het rendement van de opleiding centraal. Centrale vraag is dan welke studenten efficiënt en zonder vertraging de opleiding kunnen doorlopen en afronden. Deze drie perspectieven zijn niet zonder meer van elkaar te scheiden. Als studenten bijvoorbeeld bepaalde kwaliteiten missen, maar die met inspanning tijdens de opleiding

wel kunnen verwerven, vraagt dat tijd en aandacht tijdens de opleiding, zowel van student als opleider, wat ten koste kan gaan van de studieduur en/of van het uitstroomniveau van de student. In andere landen is selectie voor toelating tot de lerarenopleiding staande praktijk. In dergelijke landen (zoals Finland, de USA, Taiwan en Ierland) worden aan de hand van bepaalde criteria aanstaande leraren met behulp van specifieke procedures en instrumenten geselecteerd voor de lerarenopleidingen.

Het invoeren van selectie voor de poort is een ingrijpende maatregel, zowel vanuit het perspectief van de individuele student - die wel of niet wordt toegelaten -, als vanuit het perspectief van de opleiding - vanwege de mogelijke impact op de instroom voor de opleidingen en vanwege de tijd en het geld dat geïnvesteerd moet worden in de ontwikkeling en uitvoering van selectieprocedures. Zorgvuldigheid is daarom van belang. Kernvraag daarbij is in welke mate een selectieprocedure (qua criteria en instrumenten) betrouwbaar en valide is: Levert een selectieprocedure eenduidige uitspraken over toelaatbaarheid op en hebben die uitspraken ook een voorspellende waarde ten aanzien van het doorlopen van de opleiding en ten aanzien van het toekomstige functioneren in het beroep? Om zicht te krijgen op het antwoord op deze kernvraag is het belangrijk om na te gaan in hoeverre er al eerder onderzoek gedaan is naar deze vraag, bijvoorbeeld in (internationale) contexten waar reeds selectie-instrumenten voor toelating tot de (leraren)opleidingen worden toegepast.

Dit rapport betreft een internationale literatuurreview naar studies over voorspellende effecten van selectie voorafgaand aan de lerarenopleiding. Er is literatuur verzameld waarin aan de hand van empirisch bewijsmateriaal wordt besproken in hoeverre bepaalde selectie-instrumenten en – procedures voorspellende waarde hebben ten aanzien van uitkomstmaten die van belang zijn voor het functioneren als leraar. Teneinde de problematiek van de lerarenopleiding in een breder kader te kunnen bespreken, is bovendien via aanvullende literatuurstudie nagegaan wat de voorspellende effecten zijn van selectieprocedures die worden gehanteerd bij enkele andere opleidingen die relevant zijn voor lerarenopleidingen. Hierbij gaat het concreet om medische opleidingen en de politieacademie omdat deze opleiden tot beroepen die, net als dat van leraar, worden gekenmerkt door (mondelinge) interactie met groepen en/of individuen, soms in stressvolle situaties. Selectieve opleidingen op het gebied van kunst of sport zijn buiten beschouwing gelaten.

De in dit rapport beschrijven literatuurstudie bouwt voort op een eerder verrichte studie naar de wenselijkheid van de invoering van selectie op basis van geschiktheid voor het beroep van leraar (Snoek et al., 2014). Deze studie werd uitgevoerd in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen en was voornamelijk gebaseerd op gesprekken met stakeholders uit het domein van de lerarenopleidingen (Pabo, tweedegraads lerarenopleiding en ULO). Uit deze studie kwam naar voren dat de opvattingen ten aanzien van de (on)wenselijkheid om te selecteren op basis van geschiktheid voor het beroep in belangrijke mate worden gekleurd door de opvatting of men hoger onderwijs als een recht of als een voorrecht ziet. Voorstanders van selectie gaven vooral aan dat selectie voor de poort van de lerarenopleidingen kwaliteits-verhogend zou zijn omdat zwakkere studenten uitgeselecteerd worden. Tegenstanders vonden het vooral onwenselijk dat het risico bestaat dat studenten uitgeselecteerd zouden kunnen worden die mogelijk wel goede leraren kunnen worden. Respondenten gaven ook aan dat het inzetten van selectie-instrumenten aan de poort van de lerarenopleiding effecten zou kunnen hebben op de kwaliteit van de uitstroom (hogere instroomeisen kunnen leiden tot hoger uitstroomniveau en kunnen daarmee een kwaliteitsimpuls geven in het onderwijs), het imago en daarmee de status en aantrekkelijkheid van het beroep en de lerarenopleiding, en het opleidingsrendement (doordat er betere studenten aangetrokken worden en in de opleiding minder tijd nodig is voor remediëring en sneller op een hoger niveau gewerkt kan worden). Deze opvattingen sluit aan bij de constatering van Barber en Mourshed (2007) dat de best

presterende onderwijsstelsels in de wereld gekenmerkt worden door selectieve toelating tot de lerarenopleiding. Hoewel er geen onderwijswetenschappelijke evidentie is dat deze effecten daadwerkelijk in Nederland op zullen treden, kunnen deze verwachte effecten wel een reden zijn om over te gaan tot selectie aan de poort. De literatuurreview in dit rapport is dan ook bedoeld om een goed beeld te krijgen van wat er uit internationaal onderzoek bekend is over de voorspellende effecten van bepaalde selectiecriteria, -instrumenten en –procedures die worden toegepast bij selectie tot toegang van lerarenopleidingen.

Het rapport is als volgt opgebouwd. In het volgende hoofdstuk wordt de gevolgde methode beschreven, waarin de systematiek die gevolgd is bij de literatuurreview centraal staat. In hoofdstuk 3 worden de uitkomsten van de analyses van de gevonden literatuur besproken. Dit hoofdstuk wordt aangevuld met een appendix waarin de meest relevante studies uitvoerig worden besproken. Dit betreft 6 empirische studies naar de effecten van selectieprocedures bij lerarenopleidingen. Hoofdstuk 4 bevat een discussie van de bevindingen op basis waarvan enkele conclusies worden getrokken, waarna het rapport afsluit met een kort hoofdstuk over implicaties voor de Nederlandse lerarenopleidingen.

## Methode

Om inzicht te krijgen in wat er uit internationaal onderzoek bekend is over de voorspellende effecten van bepaalde selectiecriteria, -instrumenten en –procedures die worden toegepast bij selectie tot toegang van lerarenopleidingen, is een systematische verkenning van literatuur uitgevoerd.

### *Bepalen van de zoektermen*

In een pilotverkenning van de literatuur, gericht op het definiëren van zoektermen en inclusiecriteria, werd allereerst nagegaan in hoeverre er studie is gedaan naar voorspellende effecten van selectie in de lerarenopleidingen. De zoektocht in literatuur bestond uit een aantal stappen: eerst zijn relevante databases (te weten: ERIC, PsychINFO, PiCarta, Web of Science, Academic search premier, Google scholar) doorzocht met een aantal zoektermen. Vervolgens werden de gevonden artikelen uit de pilotverkenning doorgenomen om mogelijke andere relevante zoektermen op het spoor te komen. Hierna zijn via een sneeuwbalmethode de referenties nagekomen die voorkwamen in de artikelen die tot dusver waren gevonden, en zijn publicaties gezocht waarin deze artikelen worden geciteerd.

Op basis van de bevindingen van deze pilotverkenning werden de uiteindelijke zoektermen gedefinieerd. In Tabel 1 zijn deze zoektermen en criteria voor uitsluiting opgenomen.

**Tabel 1**

*Overzicht van de centrale zoektermen medische (basis- en vervolg)opleidingen en lerarenopleidingen en inclusiecriteria*

<i>Inhoudelijke zoekterm</i>	<i>Lerarenopleidingen</i>	<i>Medische opleidingen</i>	<i>Inclusiecriteria</i>
Select	Teacher Education	Medical education	Engelstalig
Predict	Teacher training	Residency Education	Tijdschrift artikelen
Student performance*			Peer-reviewed
Resident performance*			Over voorspellende waarde van selectiecriteria, -instrumenten, -procedures voorafgaand aan de opleiding

\*Student resp. resident performance zijn alleen meegenomen als zoekterm in het medische domein

#### *Zoekopdracht in het domein van lerarenopleidingen*

De zoekopdrachten betroffen alle jaren. Met deze zoektermen werden 198 manuscripten geïdentificeerd binnen het gebied van lerarenopleidingen. De literatuurverkenning werd ten tweede ingeperkt door enkel die manuscripten te selecteren die gebaseerd waren op onderzoeksresultaten en gepubliceerd zijn in Engelstalige peer-reviewed tijdschriften. Deze inperking is gemaakt om de kwaliteit van de bestudeerde onderzoeksresultaten te waarborgen. Na doornemen van titel en keywords op de selectiecriteria: teacher education, teacher characteristics, instructors, en educational setting werden 61 artikelen over selectie bij lerarenopleidingen geïdentificeerd.

Na bestudering van de samenvattingen van deze 61 artikelen, bleek dat het merendeel inhoudelijk ging over beoordeling en assessment *tijdens* de lerarenopleidingen. Deze publicaties zijn niet meegenomen in deze literatuurverkenning omdat we ons hebben gefocused op voorspellende waarde van selectiecriteria *voorafgaand* aan de lerarenopleidingen.

Uiteindelijk werden 19 publicaties geïdentificeerd die een of meerdere onderwerpen behandelen met betrekking tot de voorspellende effecten van selectie van studenten voorafgaand aan de lerarenopleidingen. Deze verzameling omvatte vier literatuurreviews, negen studies naar de voorspellende effecten van een enkele variabele en zes studies naar de effecten van selectieprocedures. Een inhoudelijk overzicht van de bestudeerde publicaties, verdeeld over deze drie groepen, is opgenomen in de tabellen 2 en 3 en in bijlage 1 (voor wat betreft de zes studies naar selectieprocedures). Het volgende hoofdstuk bevat een inhoudelijke interpretatie van de resultaten uit deze studies.



**Tabel 2***Overzicht van relevante reviewstudies naar individuele predictoren van studentprestaties in lerarenopleidingen*

<i>Studie (naam eerste auteur)</i>	<i>Context</i>	<i>Onderzoeksvraag</i>	<i>Methode</i>	<i>Analyse</i>	<i>Verklaarde variantie</i>
Archer (1946) en Archer (1949); <i>Review of Educational Research</i>	Teacher education in colleges and universities in US	Recruitment, Institutional Selection, and Guidance of Teachers	Description of literature about all three topics; in this summary we focus on the predictive value of selection criteria.	Qualitative description of results from previous studies	<ul style="list-style-type: none"> <li>- A combination of variables is superior to a single one in prediction of academic success in teacher education.</li> <li>- Achievement test and test of English have sufficient predictive value; adding intelligence tests raised the predictive value.</li> <li>- Personality traits and neurotic tendency did not correlate with GPA.</li> </ul>
Barr (1949); <i>Review of Educational Research</i>	Teacher education in colleges and universities in US	Prediction of teacher efficiency and teacher success	Description of literature about measurement techniques from a psychological and methodological perspective	Qualitative description of results from previous studies	<ul style="list-style-type: none"> <li>- increase in systematic studies into teacher success.</li> <li>- yet, no reports of the differential predictions of teaching efficiency have appeared in the literature.</li> <li>- No relations between intelligence, academic achievement in previous studies or high school grades and teaching success (university examiner estimate of teaching ability, professional tests).</li> </ul>
Black (1958); <i>Review of Educational Research</i>	Teacher education in colleges and universities in US	Recruitment and Selection procedures at primary and secondary TE	Description of literature about recruitment of teacher candidates and selection.	Qualitative description of results from previous studies	<ul style="list-style-type: none"> <li>-in general, the studies do not indicate the existence of individual testing instruments which provide a valid basis for selection.</li> <li>- Only 1 study was found about evaluation of selection procedure; This one year study concluded that (a) an organized recruitment program is more effective than none, (b) a personal program is more effective than an impersonal program, (c) a combination of personal and impersonal techniques is more effective than either technique used alone, and (d) commonly employed recruitment techniques have no differential effect.</li> </ul>

**Tabel 3**

*Overzicht van relevante studies naar individuele predictoren van studentprestaties in lerarenopleidingen*

<i>Studie (naam eerste auteur)</i>	<i>Context</i>	<i>Onderzoeksvraag</i>	<i>Methode</i>	<i>Analyse</i>	<i>Afhankelijke variabele</i>	<i>Onafhankelijke variabele (predictor)</i>	<i>Verklaarde variantie</i>
Duckworth et al (2009); <i>The Journal of Positive Psychology</i>	Effectivity of starting teachers (TFA)	Do the positive traits of grit, life satisfaction, and optimistic explanatory style independently predict performance among TFA teachers?	Study on TFA teachers (n =390)	Regression and correlation	Student learning gains	Grit, life satisfaction, and optimistic explanatory style	All three positive traits individually predicted teacher performance. These findings suggest that positive traits should be considered in the selection and training of teachers.
Durflinger (1943); <i>The Journal of Experimental Education</i>	Predictive value of several achievement test in Nebraska	Which entrance tests predict grades at the end of first semester in a teacher education program?	Survey of pre-service teachers (n=321)	Regression analysis	First semester grades	6 entrance test (college aptitude, elementary achievement, English, personal data, personality inventory, music)	Only elementary achievement and English, have enough factors in common with first semester grades
Gorow (2005); <i>Education</i>	Undergraduate pre-service education methods class	This study was designed to examine the relationship between pre-service teachers' perceived body sizes and confidence in their ability to teach.	102 prospective teacher candidates both major in primary education and physical education	Correlational analyses; independent t-test; two way MANOVA	Self-perceived body weight; discrepancy between perceived weight and preferred weight; gender; subject major	Anxiety about teaching; self-confidence in ability to teach	- the results indicate that body size influences pre-service teachers' confidence in their ability to teach.  - Analyses of data by academic major revealed no significant differences between these two groups.  - Anxiety about teaching does not relate to gender, major, perceived body weight and discrepancy score.
Heinz (2013); <i>Educational Assessment, Evaluation &amp; Accreditation</i>	Irish Professional Diploma in Education programs	How far and in what way are demo-graphic variables and selection criteria (primary degree result and prior teaching experience) associated with students' results in teacher education (academic and teaching practice grades)?	Cohort study of students for teacher education in Ireland (2000-2007; n=1,023)	Correlational analysis	GPA; grades for teaching practicum; grades content courses;	Primary degree result and prior teaching experience	- This study showed a lack of evidence as regards the predictive value of prior academic achievement for students' performance in both the academic and practical components of the PDE programme.
Holmgren (1984); <i>Education</i>	US university preservice in professional areas for public education	The purpose of this investigation was to examine the critical thinking ability and interpersonal values of three groups of pre-service educational workers.	60 pre-service students with major in elementary education, special education and speech correction	Correlational analyses; One-way analyses of variance; multiple stepwise regression	GPA; English proficiency (EP); student background variables (age)	critical thinking ability and interpersonal values	- Only on the Benevolence subscale (doing things for others in an altruistic manner) of Interpersonal values special educ majors scored significantly higher than elementary education majors.  - Critical thinking ability and age were positively related predictors for GPA and EP.
Shechtman (1992); <i>Journal of Personnel Evaluation in Education</i>	Three contexts: 1. Programs of School Counselling and Special Education. 2. Teacher college. 3. Teaching positions in the army.	What is the interrater reliability of group assessment procedure for evaluating teacher-education candidates .	1. 109 teachers SE-candidates / 13 pairs of professors. 2. 94 teacher-students / eleven pairs of teachers. 3. 69 candidates teacher positions / nine pair of recruiters.	Correlational analysis; Bivariate correlation analysis (Fisher's Z transformation)	Group assessment procedure; Interrater agreement of GAP using assessors with different backgrounds.	Overall rating score during group assessment (based on verbal ability; interpersonal skills; leadership qualities)	Overall rating score reaches the highest agreement of interrater agreement (.62<R <sup>2</sup> <.66) (p<.0.1). Fisher Z Transformation found no significant differences between Overall rating scores (only for Leadership and Human Relations)

Shechtman (1998); <i>Journal of Personnel Evaluation in Education</i>	a. Special education and counseling program at University. b. Teacher training college	1. Whether self-evaluations significantly differ from both peer and professional evaluations. 2. Whether peer evaluations will be similar to those of professional evaluations	a. 98 female first-year students b. 61 female prior admission to teacher's college.	1. T-tets 2. Correlation analysis	Group assessment procedure; Self-evaluation scores, peer evaluation scores, professional evaluation scores.	Rating scores of professionals on dimensions of GAP: Oral communication; Human interaction; Leadership; Overall rating.	1. GAP (all 4 dimensions) reaches high agreement between professional evaluations and peer evaluations. ( $30 < R^2 < .49$ ) ( $p < .05$ )  2. Self-evaluations are significantly higher than peer evaluations which are higher than professional evaluation.
Turner (2000); <i>Journal of Personnel Evaluation in Education</i>	Postgraduate Certificate in Education (PGCE) courses	Whether the selection process is reliable (judged by the number of candidates who complete the course successfully)	145 respondents from a single cohort of science students from application (respondents from interview) until completion. Monitoring withdrawals and performance on the written and practical teaching elements of the course.	Analysis of IHE records; questionnaire	Evaluation of student-teacher performance	1. Class of degree; 2. GCE A-levels. 3. Practical teaching component of the course. 4. Written component of the course.	1. Qualification with a class 2.2 degree (GPA) or above could be a predictor for successful student-teachers on the program (88% of successful students)  2. GCE A-levels could be a predictor for successful students (83% One grade A-C and 45% Three grade A-C). With these qualifications, sign (?) more % white respondents were offered a place (84%).  3. Poor teaching skills was a key factor in the withdrawal of ethnic minority students
Viskic-Stalec (1998); <i>Review of Psychology</i>	Faculty of Physical Education UniZagreb '93/'94	1. Whether the selection process as measured by GPA and a battery of tests predicts later measures of academic success (who have managed to enrol in the 5 <sup>th</sup> semester)	182 undergraduate students who passed the entrance examination for admission.  No fee students (n=103) fee students (n=79); successful (n=88) unsuccessful (n=94) .	Discriminant analysis; MANOVA	Evaluation of academic success in the physical education teacher program (whether students enroll in the 5 <sup>th</sup> semester or not)	GPA; assessment of cognitive abilities, conative characteristics (personality), motor knowledge (sport specific skills) and motor abilities.	The only predictor for academic success in the physical education teacher program is GPA (Structure coefficient = .55)

### *Zoekopdracht in andere domeinen: Medisch onderwijs en politieacademie*

In de verkenning van de literatuur is ook nagegaan in hoeverre er onderzoek is gedaan naar voorspellende effecten van selectie in bepaalde specifieke opleidingen die relevant zijn voor lerarenopleidingen. De volgende opleidingen zijn hierbij betrokken: de medische basis en – vervolgoopleidingen en de politieacademie. Specifiek is nagegaan in hoeverre er onderzoek is gedaan naar voorspellende effecten van selectie in deze opleidingen. Aangezien in het medische domein erg veel studies op dit gebied voorhanden zijn (bij een aanvankelijke search met dezoektermen uit Tabel 1 werden 112 artikelen geïdentificeerd), is besloten om uitsluitend reviewstudies en meta-analyses in het onderzoek te betrekken die samenvatten wat uit internationaal onderzoek bekend is over de voorspellende effecten van selectieprocedures in de medische opleidingen. Uiteindelijk werden vijf van dergelijke publicaties geïdentificeerd. Tabel 4 geeft een inhoudelijk overzicht van de bestudeerde publicaties. In deze rapportage geven we een overzicht en een interpretatie van de resultaten uit deze studies. Bij de verkenning van de literatuur over de opleiding tot politieagent werd één rapport gevonden waarin op een systematische manier predictoren voor prestatie in de politieacademie werden bestudeerd, plus twee relevante wetenschappelijke publicaties over predictoren van politiewerk en politieopleiding (zie Tabel 5).

**Tabel 4**

*Overzicht van relevante overzichtsstudies studies naar predictoren tijdens selectie voor medische basis- en vervolgopleidingen*

<i>Studie (naam eerste auteur)</i>	<i>Context</i>	<i>Onderzoeksvraag</i>	<i>Methode</i>	<i>Analyse</i>	<i>Dependent measures</i>	<i>Independent measures (predictor)</i>	<i>Verklaarde variantie</i>
Salvatory (2001); <i>Advances in Health Sciences Education</i>	Health care education	To review the health professions literature on the reliability and validity of both cognitive and non-cognitive measures used to select students to health profession education programs, with particular attention to their inter-rater reliability as selection tools and their predictive validity in terms of academic and clinical performance.	A total of 83 articles were reviewed for this paper (from the fields of medical education, nursing, physiotherapy, occupational therapy, respiratory care, medical imaging, and midwifery)	Review	Academic and clinical performance	Pre-admission academic grades, aptitude tests, interviews, written submissions, and letters of reference	Pre-admission overall grade point average (GPA) is the best predictor of academic performance in all of the health professions; however, the relationship between pre-admission GPA and clinical performance is less clear. The Medical College Admission Test is a good predictor of performance of medical students in terms of in-course grades and licencing examination scores but a similar test does not exist in the other health professions. Controversy remains as to the value of personal interviews and written submissions as selection tools, although it is clear that training of assessors and explicit rating guidelines enhance their reliability and validity.
Hamdy (2006). <i>Medical Teacher</i>	Medical education and medical practice	To assess the value of measurements obtained in medical schools in predicting future performance in medical practice.	Of 569 studies retrieved with our search strategy, 175 full text studies were reviewed. A total of 38 studies met our inclusion criteria and 19 had sufficient data to be included in a meta-analysis of correlation coefficients.	Quantitative meta-analysis and qualitative approaches were used for data analysis and synthesis including the methodological quality of the studies included.	Residency supervisor ratings, NBME III, residency in-training examinations, American Specialty Board examination scores, and on-the-job practice performance.	National Board Medical Examinations (NBME) I and II, preclinical and clerkship grade-point average, Observed Standardized Clinical Examination scores and Undergraduate Dean's rankings and honors society.	<p>The highest correlation between predictor and outcome was NBME Part II and NBME Part III, <math>r=0.72</math>, 95% CI 0.30–0.49 and the lowest between NBME I and supervisor rating during residency, <math>r=0.22</math>, 95% CI 0.13–0.30.</p> <p>The approach to studying the predictive value of assessment tools varied widely between studies and no consistent approach could be identified.</p> <p>Overall, undergraduate grades and rankings were moderately correlated with internship and residency performance. Performance on similar instruments was more closely correlated. Studies assessing practice performance beyond postgraduate training programs were few.</p>
Henry (2009). <i>Rural and Remote Health</i>	Medical education in colleges and universities in Australia	To assess the effectiveness of selection practices to predict 1) successful graduation and 2) the impact of rural pipeline components on eventual rural practice.	<p>1. review of the literature on the power of selection practices to predict successful graduation;</p> <p>2. review of the literature on the impact of the components of the rural pipeline, incl. student selection, on the choice of rural practice as a career.</p> <p>(number of studies: 48)</p>		<p>1. Medical course academic performance.</p> <p>2. The choice of a rural career.</p>	<p>1. Undergraduate academic performance; interviews other measures of non-cognitive performance or the content of the undergraduate degree.</p> <p>2. Prior rural residence; extended rural exposure during medical training; professional support at national, state and local levels; career pathway opportunities</p>	<p>1. Undergraduate academic performance is the strongest predictor of medical course academic performance. The predictive power of interviews is modest. There are limited data on the predictive power of other measures of non-cognitive performance or the content of the undergraduate degree.</p> <p>2. Prior rural residence is the strongest predictor of choice of a rural career but extended rural exposure during medical training also has a significant impact. The most significant influencing factors are: professional support at national, state and local levels; career pathway opportunities; contentedness of the practitioner's spouse in rural communities; preparedness to adopt a rural lifestyle;</p>

Kreiter (2014). <i>Teaching and Learning in Medicine</i>	Medical Education in US	To discuss factors impeding the incorporation of research findings on how best to select applicants to study medicine, and to offer a perspective on 25 years of research evidence that we believe has important implications for structuring a better selection process.	Number of studies: 61	Five areas of inquiry are discussed: (1) the interview and related techniques, (2) admission tests, (3) other measures of personal competencies, (4) the decision process, and (5) defining and measuring the criterion.	Academic and Clinical Performance	(1) the interview and related techniques, (2) admission tests, (3) other measures of personal competencies	<p>educational opportunities for children; and proximity to extended family and social circle.</p> <ul style="list-style-type: none"> <li>- The traditional interview has been shown to lack both reliability and validity. Alternatives have been developed that display promising measurement characteristics.</li> <li>- Admission test scores have been shown to predict academic and clinical performance and are generally the most useful measures obtained about an applicant.</li> <li>- Due to the high-stakes nature of the admission decision, it is difficult to support a logical validity argument for the use of personality tests. Although standardized letters of recommendation appear to offer some promise, more research is needed.</li> <li>- The methods used to make the selection decision should be responsive to validity research on how best to utilize applicant information.</li> <li>- Few resources have been invested in obtaining valid criterion measures. Future research might profitably focus on composite score as a method for generating a measure of a physician's career success.</li> </ul>
Kreiter (2007). <i>Teaching and Learning in Medicine</i>	Medical education and medical practice	To summarize the validity evidence for using Medical College Admissions Test (MCAT) scores and undergraduate grade point average (GPA) to select applicants to study medicine and become physicians. To characterize uGPA and MCAT's relationship across medical training outcomes and to interpret the observed variance in uGPA and MCAT validity across levels of educational and professional attainment.	29 studies that provided evidence related to the validity generalization research question of using uGPA and MCAT for selection.	Meta-analysis: validity generalization (VG) methodology to summarize and clarify previous research investigating the validity of using uGPA and MCAT scores to predict performance outcomes in medical school and professional practice.	Knowledge and clinical reasoning; clinical performance in medical school and medical practice.	Medical College Admissions Test (MCAT) and undergraduate grade point average (GPA).	<p>The validity of uGPA and MCAT for predicting written tests measuring knowledge and clinical reasoning appears to diminish only slightly over the 7 years for which there was adequate evidence. For nonwritten, performance-based clinical measures, there was evidence that MCAT and uGPA are predictive of clinical skills. However, evidence is insufficient regarding the predictive trend across the attainment levels.</p> <p>Only one study examined the long-term relationship (professional practice) between MCAT and uGPA and clinical performance. The two studies that reported the reliabilities of the clinical performance measures both suggest that MCAT and uGPA have a positive and statistically significant predictive relationship with clinical skills displayed in medical school and residency.</p>

**Tabel 5**

*Overzicht van relevante studies naar predictoren in het domein van politie en politieacademie*

<i>Studie (naam eerste auteur)</i>	<i>Context</i>	<i>Onderzoeksvraag</i>	<i>Methode</i>	<i>Analyse</i>	<i>Afhankelijke variabele</i>	<i>Onafhankelijke variabele (predictor)</i>	<i>Verklaarde variantie</i>
Van der Linden (2013); <i>Rapport EUR</i>	Evaluatie van selectieprocedure bij Politieacademie	Relevante onderzoeksvraag: Wat is de voorspellende waarde van de meetinstrumenten in de selectieprocedure?	Selectieprocedure bestond uit: 1) cognitieve test, 2) persoonlijkheidsvragenlijst, 3) selectiegesprek, en 4) praktijkopdracht.	Correlatie analyse en r-waarde (maat voor verklaarde variantie)	Behalen of falen bij 'proeve van bekwaamheid' tijdens de opleiding (90 % van studenten slaagt in een keer!); uitval	Cognitie; persoonlijkheid (big five); competenties (bepaald in selectiegesprek en praktijkopdracht)	- Geen van de selectieonderdelen blijken een sterke voorspellende waarde te hebben.  - Persoonlijkheidstest geeft in vele domeinen een lage voorspellende waarde voor functioneren in het beroep; mogelijk werkt een domeinspecifieke persoonlijkheidsmaat wel.
Girodo (1997); <i>Journal of Social Behavior and Personality</i>	Selectie van undercover agenten in US	Hoe effectief is de procedure om kandidaten te selecteren die integer zijn en ook goed kunnen misleiden/liegen?	Selectieprocedure bestaande uit 1) Persoonlijkheidstest; 2) interview; 3) rollenspel	Kwalitatieve analyse; uitval; job performance verslagen	Job performance	Persoonlijkheid; motivatie;	- Meten van de vaardigheden (skills) in plaats van de persoonlijkheid (trait) lijkt een betere inschatting te geven van de mogelijkheden van de kandidaten om te misleiden/ liegen.  - De persoonlijkheidsmaat moet specifieker toegespitst worden op de sociale context waarvoor gemeten wordt.
King (1959); <i>Educational and Psychological Measurement</i>	Politie-academie US (in 1959!)	Deze studie beoogde om een multiple regressie vergelijking op te stellen die gebruikt kan worden om het GPA van studenten van de politieacademie te voorspellen	Test batterij bestond uit: lees test, Engelse taal test, reken test, persoonlijkheidstest; 114 deelnemers	multiple regressie analyse	GPA gedurende het eerste semester	Lezen, Engelse taal en rekenvaardigheden; persoonlijkheid	De enige twee variabelen met voorspellende waarde waren taal en lezen.

Het totale aantal studies dat is geanalyseerd ten behoeve van deze review betreft daarmee 27.

*Systematische analyse van geselecteerde artikelen*

De analyse van de gevonden literatuur vond plaats volgens een aantal systematische stappen. Eerst is een Excel bestand gemaakt waarin voor iedere studie werd samengevat: onderzoeksvraag, doelgroep, context, criteria, instrumenten en procedures die gebruikt zijn in het kader van de selectie met het oog op het (toekomstige) beroep, uitkomstmaten en bevindingen.

De analyse van de geselecteerde literatuur heeft geresulteerd in een overzicht van selectiecriteria, -instrumenten en -procedures en hun effecten in de betreffende context, in elk geval wat betreft voorspellende waarde t.a.v. succes in de opleiding, en soms ook t.a.v. de beroepsuitoefening. Hierbij is gespecificeerd wat uit onderzoek wel en niet bekend is, en in welke context het onderzoek heeft plaatsgevonden.

Om deze uitkomst te toetsen op relevantie voor de Nederlandse situatie is dit overzicht voorgelegd aan een panel van ervaringsdeskundigen op het gebied van selectie in de Nederlandse lerarenopleidingen. Dit panel bestond uit drie vertegenwoordigers<sup>1</sup> van verschillende typen lerarenopleidingen (Pabo, tweedegraads en universitaire lerarenopleidingen) die actief betrokken

<sup>1</sup> Robert Viëtor, opleidingsmanager Pabo Hogeschool Leiden; Alessandra Corda, opleidingsmanager cluster Talen, Hogeschool van Amsterdam; Annemarie Thomassen, hoofd afdeling VO, ICLON, Universiteit Leiden.

zijn bij het vraagstuk van selectie. Twee van hen maken deel uit de zgn. Voorlopersgroep over selectie in de lerarenopleidingen die door het Ministerie van OCW is ingesteld in het kader van haar Lerarenagenda 2013-2017. In een gesprek van ca. 1,5 uur hebben de leden van het panel gereageerd op de uitkomsten van de literatuurreview in het licht van hun eigen ervaringen met selectie bij lerarenopleidingen. Zij hebben aangegeven wat de betekenis en bruikbaarheid van de literatuurstudie in hun ogen is, en waar zij beperkingen ervaren. In het verlengde hiervan hebben zij mogelijkheden en wensen voor toekomstig onderzoek in de Nederlandse context benoemd.

## Resultaten van de literatuurreview

De systematische analyse van de internationale literatuur over de voorspellende waarde van indicatoren voor prestaties van studenten in de lerarenopleidingen leverde 19 publicaties op. Deze zijn inhoudelijk samengevat in de tabellen 2 en 3 en bijlage 1. Als aanvulling hierop zijn vijf reviewstudies uit het medische domein (Tabel 4) en drie op het gebied van politie-opleidingen (Tabel 5) geïdentificeerd. In de volgende paragraaf worden de afzonderlijke criteria en variabelen besproken waarvan de voorspellende waarde is onderzocht in de gevonden studies. In 3.2 worden zes studies besproken naar selectieprocedures waarin meerdere variabelen waren gecombineerd.

### Overzicht van criteria die onderzocht zijn bij selectie voor de lerarenopleidingen

In de bestudeerde internationale studies naar selectie voor lerarenopleidingen worden verschillende criteria beschreven. Deze selectiecriteria kunnen gecategoriseerd worden in vier clusters van criteria waarop studenten geselecteerd worden voor lerarenopleidingen, namelijk 1) kennis- en denkniveau, 2) motivatie, 3) professionele houding en persoonlijke eigenschappen, en 4) beroepsgerelateerde vaardigheden.

*Kennis- en denkniveau* omvat vakinhoudelijke kennis, maar ook taalvaardigheid, abstractievermogen, hogere-orde denken, cultureel kapitaal en oordeelsvermogen.

*Motivatie* heeft niet alleen betrekking op motivatie voor het volgen van een lerarenopleiding, maar vooral op motivatie voor het beroep van leraar en het bijdragen aan het leren en de ontwikkeling van kinderen en jongeren. Bij selectie van aankomende studenten gaat het niet alleen om de mate van motivatie, maar ook in hoeverre hun motivatie gebaseerd is op een realistische verwachting ten aanzien van het beroep en de eigen rol als leraar.

*Professionele houding en persoonlijke eigenschappen* omvatten waarden en attituden die bepalend zijn voor een professionele houding die gekenmerkt wordt door reflectie, een kritische en onderzoekende houding, nieuwsgierigheid en leergierigheid, dienstbaarheid, ondernemendheid en omgaan met complexiteit, en persoonskenmerken zoals empathie, openheid, flexibiliteit en creativiteit.

*Beroepsgerelateerde vaardigheden* omvat interpersoonlijke, pedagogische en didactische vaardigheden die kenmerkend zijn voor het beroep van leraar, zoals communicatieve vaardigheden, sociale competenties, samenwerken, structureren, organiseren, en orde houden.

Hieronder bespreken we de resultaten uit de bestudeerde studies aan de hand van deze clusters van selectiecriteria.

#### *Kennis- en denkniveau*

##### Kennisniveau

Enkele van de bestudeerde studies in deze literatuurreview nemen het gemiddelde cijfer (Grade point average, GPA) in opleidingen voorafgaand aan de lerarenopleiding mee als onafhankelijke variabele, ofwel potentiële voorspeller van de prestaties van studenten in de lerarenopleiding (Byrnes et al., 2003; Durflinger, 1943; Viskic-Stalec, Sadura & Horga, 1998; Turner & Turner, 2000; Heinz, 2013 ). Het merendeel van deze studies vindt inderdaad een sterk positieve correlatie tussen het gemiddelde cijfer dat studenten in eerdere studies behaalden en het succes in de



lerarenopleiding. Omdat de prestaties in de lerarenopleiding (afhankelijke variabele) in deze studies vaak gedefinieerd is als het gemiddelde cijfer in de opleiding, en niet altijd duidelijk uit wat voor soort vakken het gemiddelde cijfer is opgebouwd (academische vakken, stage en/of praktijk), is het niet altijd inzichtelijk wat de positieve correlatie inhoudt. In een review uit 1943(!) laat Durflinger bijvoorbeeld zien dat in een studie waarin een batterij aan vragenlijsten (waaronder GPA, college aptitude, personal data, English language proficiency en personality) werd uitgezet onder 321 pre-service leraren-in-opleiding (US) alleen het gemiddelde cijfer in eerdere opleiding en de Engelse taalvaardigheid gerelateerd konden worden aan de cijfers in het eerste semester van de lerarenopleiding. Deze gelijksoortige uitkomst werd gereproduceerd door Viskic-Stalec, Sadura en Horga (1998) voor leraren lichamelijke opvoeding in opleiding (n=182) en door Turner en Turner (2000) voor bèta-afgestudeerde leraren-in-opleiding (n=145). Slechts één studie vond geen bewijs voor een positieve relatie tussen prestaties in eerdere opleidingen en de prestaties van studenten in de lerarenopleiding (Heinz, 2013). In deze studie werden de prestaties van de studenten in het praktijk gedeelte van de lerarenopleiding meegenomen als variabele onafhankelijk van het succes in de academische componenten van de opleiding. Deze uitkomst ondersteunt het vermoeden dat de prestaties van studenten in eerdere studies een voorspellende waarde hebben voor de academische component van lerarenopleidingen, maar niet voor het praktijk gedeelte, ofwel stage, behorende bij de opleiding. Afgeleid hiervan kunnen we vermoeden dat de prestaties in eerdere studies mogelijk ook een minder sterke voorspeller is voor succes in het beroep. Evenzogoed laten de bestudeerde literatuurreviews in het medische domein zien dat GPA in eerder opleidingen, zoals de medische bacheloropleiding, een bepaalde voorspellende waarde hebben voor de prestaties van studenten in de academische onderdelen van de medische vervolgoopleidingen (Henry et al., 2009; Kreiter & Kreiter, 2007). Maar gelijk aan de studies in het domein van de lerarenopleidingen, zijn de resultaten over de relatie tussen het gemiddelde cijfer in eerdere opleidingen en prestatie in het klinische gedeelte van de opleidingen tegenstrijdig (Salvatory, 2001; Hamdy et al., 2006).

#### Algemene denkvaardigheden

In enkele studies worden ook algemene denkvaardigheden bestudeerd als mogelijke predictieve variabele voor succes in de lerarenopleidingen (Holmgren & Covin, 1984; Andrew et al., 2005). Het vermoeden is dat algemene denkvaardigheden, zoals kritische en analytische denkvaardigheden, noodzakelijke voorwaarden zijn voor leraren-in-opleiding om hun stage, en later hun beroep, effectief te kunnen uitvoeren. Leraren moeten in staat zijn om voor, tijdens en na het lesgeven kritisch en analytisch te denken, keuzes te maken tijdens het lesgeven en te reflecteren op hun eigen handelen. Holmgren en Covin (1984) lieten in een studie onder 60 studenten in een lerarenopleiding in de VS (met majors in speciaal onderwijs, primair onderwijs en logopedie) zien dat kritische denkvaardigheden (gemeten met de Wason Glaser Thinking Appraisal) en leeftijd positief gerelateerd waren aan gemiddelde cijfer (GPA tijdens de lerarenopleiding) en Engelse taalvaardigheid (Holmgren & Covin, 1984). Alhoewel we kunnen vermoeden dat het soort van denkvaardigheden die leraren-in-opleiding gebruiken varieert tussen de academische en de praktische onderdelen van de opleiding, werd in deze studie geen onderscheid gemaakt tussen de verschillende gedeelte van de lerarenopleiding. The Graduate Record Examinations (GRE) is een gestandaardiseerde test welke bij meerdere opleidingen in de VS gebruikt wordt als toelatingsexamen. Dit toelatingsexamen meet verbale, rekenkundige, analytische, kritische schrijf- en denkvaardigheden die zich over een langere termijn hebben ontwikkeld en niet gerelateerd zijn aan specifieke opleidingen. In een studie van Andrew et al. (2005) bleek opvallend genoeg dat verbale vaardigheden niet voorspellend waren voor de beoordelingen van de studenten tijdens het praktijkonderdeel van de lerarenopleiding. De analytische schrijfvaardigheden, daarentegen, van aanstaande studenten voor lerarenopleidingen waren wel voorspellend voor het succes tijdens de opleiding. Dit onderdeel van de GRE meet de analytische en logische denkvaardigheden (zoals scheiden van hoofd- en bijzaken en het maken van

weloverwogen keuzes op basis van beperkte informatie). Dat deze vaardigheden het succes van docenten voorspellen wordt door de auteurs toegeschreven aan het gegeven dat deze vaardigheden gerelateerd zijn aan het maken van keuzes tijdens het lesgeven (Andrew et al., 2005). In geen enkele van de gevonden studies wordt de relatie van predictoren met succes in de beroepsuitoefening aangetoond. Veel auteurs hebben het hier wel over, maar er is geen evidentie (empirisch of theoretisch) dat de indicatoren voor prestatie in de opleidingen gerelateerd zijn aan prestaties in het latere beroep.

### *Motivatie*

Geen van de bestudeerde studies neemt studentmotivatie mee als unieke voorspellende factor voor prestaties in de lerarenopleidingen. Dit is opvallend omdat in talrijke studies naar de relatie tussen motivatie en studieprestaties is gebleken dat motivatie voor de opleiding en motivatie voor het latere beroep van invloed zijn op de uitval en het succes in de opleiding (zie bijvoorbeeld Brophy, 2004; Pintrich, 2003). Daarnaast kan een verkeerde afstemming tussen de verwachtingen van studenten en het beleefde opleidingsprogramma aanleiding zijn voor studenten om voortijdig te stoppen met hun studie. In enkele bestudeerde selectie procedures werden motivatie en verwachtingen van studenten wel meegenomen in de vorm van motivatiebrieven en interviewgesprekken met studenten (Caskey, Peterson, & Temple, 2001; Bieri & Schuler, 2011; Valli & Johnson, 2007; Shechtman & Sansbury, 1989; zie verder paragraaf 3.2). Soms maakt motivatie min of meer verdekt uit van een selectieprocedure, zonder dat dit apart gemeten is. Dit is bijvoorbeeld het geval bij procedures waarbij studenten een extra inspanning moeten leveren om aan de eisen voor deelname aan de procedure te voldoen (bijv. een portfolio samenstellen, een demonstratieles voorbereiden). De bestudeerde studies, evenwel, geven geen uitsluitsel over de voorspellende waarde van de motivatie en verwachtingen van studenten voorafgaand aan de opleiding als zodanig voor succes in de opleiding, stage of in het beroep.

Ook de literatuurreviews in het medische domein versterken het vermoeden dat toelatingsinterviews weinig tot geen voorspellende waarde hebben voor de studentprestaties in de opleiding. Salvatory (2001) bestudeerde 83 artikelen over de betrouwbaarheid en validiteit van selectie-instrumenten die worden gebruikt in het medische onderwijs en komt tot de conclusie dat de waarde van individuele interviews en geschreven motivatiebrieven niet eenduidig vast ligt. Henry et al. (2009) beschrijven na bestudering van de medisch onderwijskundige literatuur over selectiepraktijken dat de voorspellende waarde van interview gematigd ('modest') is. Kreiter en Axelson (2013) bestudeerden 25 jaar aan onderzoeksgegevens over selectiepraktijken in het medische domein en concludeerde dat het traditionele toelatingsinterview een gebrek vertoont zowel aan betrouwbaarheid als aan validiteit.

### *Professionele houding en persoonlijke eigenschappen*

In de jaren 40 van de vorige eeuw zijn diverse studies uitgevoerd naar de voorspellende waarde van persoonlijkheidseigenschappen voor studiesucces in lerarenopleidingen. Toentertijd was de aanname dat 'goede' docenten een specifieke persoonlijkheidsstructuur hadden en dat het daarom mogelijk was om aanstaande docenten te selecteren op basis van specifieke persoonlijkheidseigenschappen. Eind jaren 40 liet Archer (1946, 1949) in een kwalitatieve literatuurstudie zien dat studies geen eenduidige resultaten lieten zien over de voorspellende waarde van persoonlijkheidseigenschappen, zoals neuroticisme, introversie-extraversie, dominantie-ondergeschiktheid, voor de prestaties als docent of docent-in-opleiding. Alhoewel in een recente studie Duckworth et al. (2009) correlaties vaststelden tussen enkele persoonlijkheidseigenschappen van beginnende docenten (te weten doorzettingsvermogen ('grit'), tevredenheid en optimisme) en succes als docent, zijn er geen studies gevonden die gericht waren op de voorspellende waarde van persoonlijkheidseigenschappen voor studiesucces van aanstaande studenten voor de lerarenopleiding.

De resultaten uit de studies in het domein van de politieacademie zetten vraagtekens bij de voorspellende waarde van algemeen gemeten persoonlijkheidskenmerken van kandidaat-agenten. In de evaluatie van de selectieprocedure bij de politieacademie in Nederland blijkt persoonlijkheid, gemeten met een algemene persoonlijkheidsmaatgeen voorspellende waarde te hebben (alle andere selectiematen trouwens ook niet). De auteurs geven aan dat een domeinspecifieke persoonlijkheidsmaat mogelijk wel voorspellende waarde zou kunnen hebben (Van der Linden et al., 2013). Ook een oudere studie in het domein van de politieacademie in de VS (vergelijkbaar met HBO) vond geen voorspellende waarde van persoonlijkheidsmaten (trouwens wel van taalvaardigheid) (King et al., 1959).

### *Beroepsgerelateerde vaardigheden*

Tijdens de lerarenopleidingen wordt gewerkt aan diverse beroepsgerelateerde bekwaamheden, zoals inter-persoonlijke bekwaamheid, pedagogische bekwaamheid en vakdidactische bekwaamheid (zie SBL bekwaamheidseisen in Wet BIO). We zouden kunnen vermoeden dat deze bekwaamheden reeds voor de opleiding in een bepaalde mate aanwezig moet zijn om tijdens de beperkte tijd van de opleidingen deze bekwaamheid op afstudeerniveau te kunnen ontwikkelen. In de bestudeerde literatuur wordt slechts door één studie aandacht gegeven aan de voorspellende waarde van de verbale vaardigheden van aanstaande studenten bij lerarenopleidingen voor het succesvol doorlopen van de lerarenopleiding (Andrew et al., 2005). Deze auteurs laten in deze studie bij leraren-in-opleiding voor primair en secundair onderwijs in de VS (n=116) zien dat er *geen* eenduidig positief verband is tussen verbale vaardigheden, gemeten met de GRE-V test<sup>2</sup> en de doceervaardigheden die de leraren-in-opleiding laten zien in hun stage tijdens de opleiding. Dit resultaat wijkt af van een oude studie die wel voorspellende waarde van taalvaardigheidsaspecten voor het succes in de opleiding wist te identificeren (Durflinger, 1943; deze studie werd meegenomen in de literatuurstudie van Archer, 1946). Zij liet in de VS context zien dat Engelse taalvaardigheid (gemeten voorafgaand aan de opleiding) gecorreleerd is aan de cijfers (GPA) die studenten in het eerste semester van de lerarenopleiding haalden (incl. inhoudelijke vakken en stages). Uiteraard is het de vraag wat dit resultaat heden ten dage betekent. Holmgren en Covin (1984) namen inter-persoonlijke vaardigheden als onafhankelijke variabele mee in hun studie en toonden aan dat deze, in tegenstelling tot kritische denkvaardigheden en leeftijd, geen voorspellende waarde hebben voor studiesucces in de lerarenopleiding. In de bestudeerde literatuur werden geen andere studies gevonden die aandacht schonken aan de voorspellende waarde van beroepsgerelateerde bekwaamheden die reeds (in een bepaalde mate) aanwezig zijn bij aanstaande studenten van de lerarenopleidingen.

### *Tot slot*

Samenvattend laten deze studies zien dat individuele predictoren slechts beperkte voorspellende waarde voor prestaties van studenten in de lerarenopleiding laten zien. Verder steunen deze studies het vermoeden dat predictoren voor prestatie in de academische componenten van de lerarenopleiding anders zijn (of een andere sterkte hebben) dan de predictoren voor het praktijk gedeelte van de opleiding. Opvallend is dat in deze studies niet de motivatie van aanstaande studenten meegenomen is als potentiële predictor voor prestaties in de opleidingsonderdelen. Verschillende studies benadrukken dat een mix van criteria de meest voorspellende waarde heeft. In de volgende paragraaf worden deze studies besproken.

---

<sup>2</sup> Volgens de makers van de GRE-testen, *The Educational Testing Service* (ETS), test de GRE-V de vaardigheid om geschreven teksten te analyseren en te evalueren en relevante informatie samen te voegen; om relaties tussen componenten van zinnen te analyseren; en om relaties tussen woorden en concepten te herkennen (ETS, 2003, p.4, als geciteerd in Andrew et al. (2005).

## Overzicht van studies naar de effecten van integrale selectieprocedures bij lerarenopleidingen

Er zijn zes artikelen gevonden waarin onderzoek gedaan is naar selectieprocedures bij lerarenopleidingen, waarin een mix van criteria is geïntegreerd. Bij de bespreking van deze studies is gebruik gemaakt van dezelfde vier clusters van criteria als in par. 3.1: 1) kennis- en denkniveau, 2) motivatie, 3) professionele houding en persoonlijke eigenschappen en 4) beroepsgerelateerde vaardigheden. Daarnaast zal waar mogelijk worden toegelicht welke vormen van studieprestaties door deze criteria worden voorspeld. In bijlage 1 wordt ieder van deze zes studies uitgebreid samengevat.

Bieri en Schuler (2011), Byrnes et al. (2003), Caskey et al. (2001), Shechtman & Sansbury (1989) en Valli & Johnson (2007) onderzochten allemaal of een selectieprocedure met een mix van criteria een voorspellende waarde kan geven voor het studiesucces aan het einde van de lerarenopleiding. Alle studies richtten zich hierbij op de praktische vaardigheden die studenten aan het eind van de opleiding moeten bezitten. Selectieprocedures die inzetten op het nabootsen van een les (Valli & Johnson, 2007), of op vaardigheden waarbij het probleemoplossend vermogen en andere beroepsgerelateerde vaardigheden worden aangesproken, blijken een voorspellende waarde te hebben m.b.t. de geschiktheid van studenten om voor de klas te staan (*teaching skills*). Uit de studie van Caskey et al. (2001) blijkt dat een simulatie-activiteit waarin kandidaten de opdracht krijgen om in kleine groepjes een fictief probleem gezamenlijk op te lossen, op zichzelf een zwakke voorspeller is, maar dat deze samen met allerlei andere selectiecriteria (waaronder een motivatiebrief, referenties en gestandaardiseerde testen) een grotere voorspellende waarde heeft (gezamenlijke verklaarde variantie van 38%). Het studiesucces werd beoordeeld aan de hand van een beoordeling op de criteria 'kwaliteit van prestaties in de opleiding' en 'kwaliteit als docent'. Wellicht dat de academische criteria eerder de kwaliteit van opleidingsprestaties voorspellen en dat de simulatieactiviteit de meer praktijkgerichte kwaliteit als docent voorspelt, maar dit wordt in deze studie niet duidelijk. Hier zou verder onderzoek op gedaan moeten worden.

Valli en Johnson (2007) onderzochten of een demonstratieles, een groepstaak en een interview een goede voorspellende waarde hebben voor de eindprestaties van studenten. Tijdens de demonstratieles werden kandidaten beoordeeld op de mate waarin zij het leren van leerlingen begeleidden en op de effectiviteit van hun interacties met de leerlingen. Tijdens de groepstaak moesten kandidaten deelnemen aan een groepsdiscussie, een dialoog en een analyse. De groepstaak komt in grote lijnen overeen met de groepsprocedure van Shechtman en Sansbury (1989). In het interview werden kandidaten beoordeeld op hun algemene geschiktheid en moesten zij voor aanvang hiervan een motivatiebrief sturen ('ik als docent'). Vervolgens werd gekeken of deze scores een voorspellende waarde hadden met betrekking tot de academische en praktische vaardigheden van studenten aan het einde van hun opleiding. Voor de academische vaardigheden werden studenten beoordeeld op hun masterthesis waarin het wetenschappelijk redeneren en analyseren werd geëvalueerd en voor de praktische vaardigheden werden studenten beoordeeld op hun onderwijsvaardigheden tijdens een onderwijspracticum. De resultaten van deze studie laten zien dat een demonstratieles het beste voorspelt hoe succesvol studenten aan het einde van hun opleiding hun onderwijspracticum doen, maar dat die demonstratieles niet de prestaties op de masterthesis voorspelt. Daarnaast blijkt dat hoge beoordelingen op het interview en de groepstaak nog beter voorspellen of studenten goed of zwak zullen scoren op hun praktische vaardigheden aan het eind van de opleiding. Valli en Johnson (2007) concluderen dan ook dat het gebruik van een combinatie van selectiemethoden belangrijk is en bevelen een selectieprocedure in fasen aan waarbij de demonstratieles in de eerste fase wordt gehouden en het interview en de groepstaak in de tweede

fase. Een kanttekening hierbij is dat de groepstaak zwakke punten kent. De groepstaak komt in grote lijnen overeen met de groepsprocedure van Shechtman en Sansbury (1989) die wél valide en betrouwbaar is gebleken. Om die reden wordt de groepsprocedure hieronder nauwkeuriger besproken.

### *Groepsprocedure*

Shechtman en Sansbury (1989) zijn grondlegger van een groepsprocedure waarin meer wordt gemeten dan enkel cognitieve vaardigheden. Een dergelijke procedure bleek succesvol te worden gebruikt in managementberoepen, en de verwachting was dat de gemeten 'persoonlijke componenten' ook belangrijk zijn voor succesvolle docenten. In het onderzoek van Shechtman en Sansbury (1989) en later in een vergelijkende studie van Byrnes en collega's (2003) is onderzocht of de groepsprocedure een voorspellende waarde heeft voor de praktijkvaardigheden van studentleraren.

In een aantal studies heeft Shechtman met collega's de groepsprocedure onderzocht op lerarenopleidingen in Israël en heeft daarbij zowel de voorspellende waarde (1989) als de interbeoordelaarsbetrouwbaarheid (1992, 1998) ervan gemeten, beiden met een positief resultaat. In 2003 heeft Shechtman met collega's Byrnes en Kiger de groepsprocedure ook in de VS getoetst. Ook hier bleek deze een voorspellende waarde te hebben.

De groepsprocedure is oorspronkelijk gebaseerd op 9 dimensies, namelijk: verbale expressie, denkvaardigheden, motivatie, zelfvertrouwen, intermenselijke vaardigheden, leiderschapskwaliteiten, flexibiliteit, creativiteit en een algemeen beeld van geschiktheid voor het leraarschap. In de latere studie van Byrnes en collega's (2003) zijn de dimensies echter teruggebracht naar drie dimensies, waarin de 9 dimensies verwerkt zijn, namelijk: 1) verbale vaardigheden (duidelijkheid en organisatie van gedachten, focus op essentiële zaken, logische presentatie en opeenvolgende ontwikkeling van gedachten, verbale expressie, vloeiend spreken, precisie, en een uitgebreide vocabulaire); 2) interpersoonlijke vaardigheden (warme expressie, vriendelijkheid, ondersteuning, respectvolle houding, sensitiviteit en goede verstandhouding met de klas); en 3) leiderschapskwaliteiten (dynamisch, alertheid, initiatief nemen, enthousiasme, verantwoordelijkheid, zelfverzekerdheid en zelfsturendheid). Naast deze dimensies werd er ook een beoordeling gegeven op basis van een algemene indruk van geschiktheid voor het leraarschap. Op deze manier dekt deze beoordeling voornamelijk de clusters van criteria *professionele houding en persoonkenmerken* en *beroepsgerelateerde vaardigheden* (zie 3.1). *Kennis- en denkniveau* worden gemeten aan de hand van GPA. *Motivatie* wordt mogelijk indirect bepaald via de houding van kandidaten. Zo gezien dekt de groepsprocedure dus de meeste in 3.1 besproken criteria.

De groepsprocedure bestaat in de studie van Byrnes et al. (2003) uit een 1,5 uur durende sessie waarin vijf groepsactiviteiten werden gedaan: 1) een niet-gestuurde groepsintroduktie waarbij kandidaten zichzelf op hun eigen (vrije) manier introduceren; 2) een gestructureerd groepsinterview gericht op houding en waarden van de kandidaten waarbij de groep twee controversiële onderwijskundige en pedagogische onderwerpen krijgt om over te discussiëren (bijv: 'Welke rol moeten openbare scholen spelen met betrekking tot seksuele voorlichting?' en 'Moeten ouders afgerekend worden op crimineel gedrag van hun kinderen?'); 3) een ongeleide groepsdiscussie waarbij de groep een beslissing moet nemen over een belangrijk en lastig onderwerp (bijv. bepalen wie van de vele patiënten een niertransplantatie zou moeten krijgen of welke leerkracht van een school ontslagen zou moeten worden bij bezuinigingen); 4) het geven van mondelinge feedback aan elkaar; 5) een discussie over de groepservaring waarbij kandidaten reflecteren over hun eigen prestaties tijdens de procedure.

De kandidaten werden beoordeeld door twee assessoren (i.c. docenten van de universiteit) die een training van twee halve dagen hebben gekregen gericht op beoordeling van de groepsprocedure. Zij

beoordeelden de kandidaten onafhankelijk en bespraken hun oordelen daarna om tot consensus te komen over één gezamenlijke score per dimensie.

Aan het eind van de opleiding werden studenten beoordeeld op hun praktijkvaardigheden door hun praktijkdocent en een supervisor van de universiteit. De beoordeling die zij hiervoor kregen werd gebaseerd op hun prestaties op de volgende acht criteria: a) effectief lesgeven; b) klassenmanagement; c) assessment van leren/benadering van leren; d) persoonlijke en professionele ontwikkeling; e) ouder- en communitybetrokkenheid; f) aandacht voor diversiteit; g) aandacht voor de leerling; h) begrip van het curriculum. De eindbeoordeling kon variëren van een 0 tot 4, waarbij 0 betekent dat de student geen enkele competenties heeft voor het leraarschap en 4 betekent dat een student net zo goed presteert als een ervaren docent.

De algemene indruk-score van assessoren bij de groepsprocedure blijkt een (zwakke) voorspellende waarde te hebben voor de praktijkvaardigheden van studenten aan het eind van hun opleiding na een intensieve stagperiode. Deze score was bovendien het meest betrouwbaar (hoge interrater reliability). De andere drie dimensies (verbale vaardigheden, interpersoonlijke vaardigheden en leiderschapskwaliteiten) bleken in de studie van Byrnes et al. (2003) wel een positief verband te hebben met de studieprestaties aan het eind van de opleiding, maar dit verband was niet significant. Dit was wel het geval in de studie van Shechtman en Sansbury uit 1989, maar ook daar bleek de algemene-indruk-score de best voorspellende variabele.

#### *Tot slot*

Uit de analyse van selectieprocedures als selectiemethoden voor lerarenopleidingen blijkt dat een mix van criteria nodig is voor het voorspellen van succesvolle studenten aan het einde van hun opleiding. De praktische vaardigheden lijken het dichtst bij de beroepspraktijk te liggen en deze worden niet zozeer voorspeld door academische criteria (zoals GPA of een schrijftest), maar eerder door het gebruik van groepsprocedures, interviews en demonstratielessen. Groepsprocedures zijn op basis van onderzoek in verschillende landen het meest betrouwbaar gebleken. Studies waarin geïntegreerde selectieprocedures worden bestudeerd, kunnen dus van bijzonder belang zijn voor ons begrip van de voorspellende waarde van clusters van criteria voor prestaties van studenten in de verschillende opleidingsonderdelen van de lerarenopleidingen.

## Discussie en conclusie

In het voorgaande hoofdstuk is antwoord gegeven op de vraag in hoeverre internationaal wetenschappelijk onderzoek zicht geeft op betrouwbare en valide methoden voor selectie aan de poort van de lerarenopleiding. Daartoe zijn drie stappen gezet: in eerste instantie is aan de hand van gerichte zoektermen gezocht naar relevante onderzoekspublicaties. Aan de hand van de abstracts en nadere bestudering van de artikelen is de set van gevonden publicaties teruggebracht naar 19 publicaties die daadwerkelijk ingingen op (aspecten van) de onderzoeksvraag. In de tweede stap zijn deze 19 publicaties geanalyseerd. Deze analyse is aangevuld met uitkomsten van 8 (review)studies rond selectie bij opleidingen die een vergelijkbaar beroep doen op een mix van cognitieve, affectieve en interpersoonlijke kwaliteiten als de lerarenopleiding: medische opleidingen en de politie-opleidingen.

De gevonden artikelen kennen een redelijke spreiding over landen en over contexten (drie/vierjarige undergraduate trajecten vs. eenjarige post-graduate trajecten).

### *Beperkingen van de aanpak*

De keuze om binnen dit onderzoek wat betreft lerarenopleidingen alleen te kijken naar wetenschappelijke peer-reviewed en Engelstalige publicaties, leidt tot enige beperkingen. De inperking tot Engelstalige publicaties kan tot een culturele bias leiden. Onze aanname is echter dat relevant wetenschappelijk onderzoek uiteindelijk zijn weg weet te vinden in de Engelstalige tijdschriften, zeker voor wat betreft het meer recente onderzoek.

De keuze voor wetenschappelijke artikelen leidt ook tot een beperking. Deze artikelen kennen over het algemeen een vast format met een beperkte lengte. Daarbij is voor de onderzoeksvraagstelling en de methode van onderzoek meer aandacht dan voor de precieze beschrijving van de context. Gevolg is dat de artikelen weinig specifieke informatie geven over de context en de kenmerken van de lerarenopleidingen waar de selectieinstrumenten zijn toegepast, over de precieze selectieprocedures en -instrumenten die gebruikt zijn en over de beoordelingscriteria die gehanteerd zijn. Daarmee geeft deze literatuurreview wel inzicht in de algemene stand van zaken ten aanzien van wetenschappelijke kennis over de voorspellende waarde van selectie voor de lerarenopleidingen, maar zij geeft slechts beperkt concrete handvatten voor de praktijk.

De inperking van de zoektermen tot de voorspellende waarde van selectie binnen de specifieke context van de lerarenopleidingen (aangevuld met medische opleidingen en politie-opleidingen) heeft ook consequenties voor de uitkomsten. Algemene wetenschappelijke kennis ten aanzien van effectieve selectie is niet meegenomen in de review. Ook is niet gekeken naar neven-effecten van selectie. Zo is de indruk uit de huidige ervaringen met selectie in de context van de Nederlandse lerarenopleidingen dat het hanteren (of het invoeren) van selectie op zich een selecterend effect heeft: de wetenschap dat er een selectieprocedure is, schrikt al een bepaalde studentengroep af. De gevonden publicaties gaan niet in op dergelijke neven-effecten.

### *Weinig onderzoek*

Het geringe aantal publicaties rond selectie voor de lerarenopleidingen is opvallend. Meerdere landen kennen al geruime tijd selectie bij de toegang tot lerarenopleidingen, maar er is verrassend weinig onderzoek gedaan naar de betrouwbaarheid en validiteit van de gebruikte selectieprocedures (criteria en instrumenten). Dit lijkt te maken te hebben met het feit dat de meeste landen waar selectie voor de poort wordt toegepast, voornamelijk gebruik maken van reeds beschikbare informatie ten aanzien van cognitieve prestaties (met name eindexamencijfers van het voortgezet onderwijs) zonder dat dit ter discussie staat. Selectie voor de lerarenopleidingen heeft in de afgelopen decennia bovendien slechts beperkt beleidsmatige en onderzoeksmatige aandacht gehad.

Gevolg is dat er in de periode van 1945 tot 2014 weinig voortgang is in wetenschappelijke inzichten ten aanzien van effectieve selectie voor de lerarenopleidingen.

De gevonden studies hebben een beperkte scope. De uitkomsten van selectieprocedures vergeleken met prestaties tijdens de opleiding of net na de opleiding. Studies die de voorspellende waarde van selectie op de langere termijn (gedurende de beroepsloopbaan) in kaart brengen, zijn er niet gevonden.

Tenslotte merken we op dat een belangrijk deel van de bevindingen (met name de gevonden criteria) niet uniek of specifiek van betekenis zijn voor selectie bij lerarenopleidingen, maar ook kunnen gelden voor selectie in het hoger onderwijs in meer generieke zin.

### *Selectie voor de opleiding of voor het beroep*

In de bestudeerde publicaties zijn twee van de drie perspectieven op selectie die in de inleiding genoemd zijn (de match met het toekomstige beroep, verbetering van het onderwijs in het algemeen, verhoging van studiesucces) terug te vinden. Veel studies bepalen de voorspellende waarde van selectie door de uitkomsten van selectieprocedures te vergelijken met studieresultaten en studiesucces tijdens de opleiding (Andrew et al., 2005; Archer, 1946, 1949; Bieri & Schuler, 2011; Caskey et al., 2001; Durflinger, 1943; Holmgren & Covin, 1984; Viskic-Stalec et al., 1998). De voorspellende waarde van selectie zal dan afhangen van de mate waarin de selectiecriteria matchen met de kwaliteiten die centraal staan in de opleiding. Het ligt dan in de lijn der verwachting dat selectie aan de hand van kennis- en denkniveau een grote voorspellende waarde heeft voor de academische component van de opleiding. Wanneer kennis- en denkniveau ook centraal staan in de totale opleiding, impliceert dit een voorspellende waarde voor de studie als geheel. Doordat de bestudeerde artikelen weinig inzicht geven in de curricula van de lerarenopleidingen waar de selectie plaats vond, kunnen onderzoeksuitkomsten niet zondermeer naar de huidige Nederlandse situatie vertaald worden. Dit beperkt met name de betekenis van de oudere publicaties (voor 1980) waarin de voorspellende waarde van selectie voor studieresultaten in de opleiding onderzocht is. De curricula van de lerarenopleidingen zijn in de afgelopen decennia immers steeds praktijkgerichter geworden (Darling-Hammond & Lieberman, 2012). Succesvol afronden van de opleiding heeft nu dan ook een andere betekenis dan twintig of dertig jaar geleden.

In een ander deel van de studies is de voorspellende waarde van selectieprocedures bepaald in relatie tot het functioneren in concrete beroepssituaties (Andrew et al., 2005; Byrnes et al., 2003; Caskey et al., 2001; Shechtman & Sansbury, 1989; Turner & Turner, 2000; Valli & Johnson, 2007). Het ging hier met name om gesimuleerde beroepssituaties tijdens de opleiding (stages). Daarmee beogen deze selectieprocedures een voorspellende waarde te hebben ten aanzien van de geschiktheid voor het beroep. Hier kunnen andere selectiecriteria (bijvoorbeeld criteria gekoppeld aan beroepsspecifieke vaardigheden) een sleutelrol spelen dan bij een focus op (academische) studieresultaten tijdens de opleiding.

### *Validiteit van selectieprocedures*

#### Sleutelindicatoren

Bij de bestudering van de geselecteerde artikelen hebben we een onderscheid gemaakt tussen vier clusters van criteria:

- Kennis- en denkniveau: het merendeel van de studies waarin onderzoek gedaan is naar de voorspellende waarde van het kennis- en denkniveau van studenten bij aanvang van de lerarenopleidingen vindt een sterke positieve correlatie tussen het gemiddelde cijfer dat studenten bij vooropleidingen behaalden en het succes in de lerarenopleiding. Daarbij gaat het met name om de academische component van de opleiding.
- Motivatie: Motivatie is een impliciete factor in die studies waar uitgebreide selectieprocedures worden gehanteerd. Studenten die zich voorafgaand aan een opleiding



onderwerpen aan een uitgebreide selectieprocedure (met motivatiebrieven, interviews, aanvullende toelatingstoetsen of groepsopdrachten) laten daarmee al een grote mate van motivatie zien voor de opleiding en het beroep. De bestudeerde publicaties geven echter weinig uitsluitsel over motivatie als expliciete voorspellende factor voor succes in de studie of het beroep en daarmee voor het belang van motivatie als criterium voor selectie. De weinige studies die dit aspect als factor hebben meegenomen, tonen geen voorspellende waarde aan (Shechtman & Sansbury, 1989; Caskey et al., 2001).

- Professionele houding en persoonlijke eigenschappen: Ook ten aanzien van persoonlijkheidseigenschappen zijn er feitelijk geen studies die de voorspellende waarde hiervan als zodanig hebben meegenomen of aangetoond. Dit weerspiegelt de nadruk die in de afgelopen jaren in discussies over en onderzoek naar lerarenkwaliteit gelegd is op concreet leraarsgedrag in plaats van op onderliggende persoonskenmerken. Duckworth et al. (2009) vonden wel een positief verband tussen bepaalde persoonlijkheidskenmerken en succes als leraar, maar hun studie betrof beginnende leraren uit het programma Teach for America. Hun resultaten zijn zeker niet zonder meer toepasbaar op reguliere aankomende leraren-in-opleiding.
- Beroepsgerelateerde vaardigheden: Van de bestudeerde artikelen gaan alleen Archer (1946, 1949) en Andrew et al. (2005) expliciet in op beroepsspecifieke vaardigheden en dan met name op communicatieve vaardigheden. Deze onderzoeken suggereren dat communicatieve vaardigheden wel voorspellend zijn ten aanzien van het studiesucces (Andrew et al., 2005), maar niet t.a.v. het functioneren in het beroep (Archer, 1946, 1949).

Ondanks de geringe hoeveelheid wetenschappelijk onderzoek, kan er uit de bestudeerde literatuur één hoofdconclusie getrokken worden: Er is niet een enkel criterium dat een beslissende voorspellende waarde heeft ten aanzien van studiesucces of kwaliteit als (toekomstig) leraar. Een mix van criteria lijkt de meest valide manier van selecteren op te leveren (Archer, 1946; Bieri & Schuler, 2011; Byrnes et al., 2003; Caskey et al., 2001; Shechtman & Sansbury, 1989; Valli & Johnson, 2007). Die mix kan dan bestaan uit variabelen waarvan de predictieve waarde in bepaalde studies is aangetoond, met name:

- Cognitieve capaciteiten (bijv. aan de hand van eindexamencijfers of toelatingstoetsen) (Caskey et al., 2001; Durflinger, 1943; Turner & Turner, 2000, Valli & Johnson, 2007; Viskic-Stalec et al., 1998)
- Analytische capaciteiten/intelligentie (Andrew et al., 2005, Archer, 1946/1949; Caskey et al., 2001; Shechtman & Sansbury, 1989)
- Taalvaardigheid (Archer, 1946, 1949; Caskey et al., 2001; Durflinger, 1943)
- Verbale capaciteiten (Andrew et al., 2005; Byrnes et al., 2003; Caskey et al., 2001; Shechtman & Sansbury, 1989)
- Interpersoonlijke vaardigheden (Bieri & Schuler, 2011; Byrnes et al., 2003; Shechtman & Sansbury, 1989)
- Leiderschapskwaliteiten (Bieri & Schuler, 2011; Byrnes et al., 2003; Shechtman & Sansbury, 1989)
- Creativiteit en flexibiliteit (Shechtman & Sansbury, 1989)
- Motivatie (Shechtman & Sansbury, 1989)

Overigens zijn deze variabelen niet uniek voor lerarenopleidingen maar kunnen ook relevant zijn bij de selectie voor andere opleidingen in het hoger onderwijs.

### Instrumenten

De bestudeerde artikelen geven weinig informatie over de concrete instrumenten die gebruikt zijn. Naast cognitieve testen of reeds beschikbare cognitieve scores (cijfers uit vooropleidingen), worden met name individuele of groepsinterviews of groepssimulaties genoemd (Bieri & Schuler, 2011;

Byrnes et al., 2003; Caskey et al., 2001; Shechtman & Sansbury, 1989; Valli & Johnson, 2007). De groepsprocedure die ontwikkeld is door Shechtman en collega's is tamelijk gedetailleerd beschreven in Byrnes et al. (2003; zie par. 3.2). Daar waar interviews werden toegepast, geven de bestudeerde artikelen weinig zicht op de kenmerken waar deze interviews op focusten. Slechts in één publicatie wordt gerapporteerd over onderzoek naar een selectieprocedure waarin het verzorgen van een demonstratieles was inbegrepen (Valli & Johnson, 2007).

#### *Betrouwbaarheid van selectieprocedures*

Het hanteren van (groeps)interviews en demonstratielessen vraagt een hoge kwaliteit van beoordelaars en in veel gevallen ook het inzetten van meerdere assessoren. Van belang is het vermogen van assessoren om het opleidingspotentieel van studenten betrouwbaar te kunnen bepalen. Slechts twee studies gaan expliciet in op de betrouwbaarheid van de selectieprocedure. Shechtman en Sansbury (1992) onderzochten de interbeoordelaarsbetrouwbaarheid (in welke mate komen assessoren tot vergelijkbare oordelen) van de groepsassessment procedures die zij eerder ontwikkeld hadden (Shechtman & Sansbury, 1989). Deze interbeoordelaarsbetrouwbaarheid is een belangrijk element van de kwaliteit van selectieprocedures en is voorwaarde voor bredere acceptatie van die procedures.

Daarnaast is voor dezelfde beoordelingsprocedure onderzocht in welke mate verschillende typen actoren tot vergelijkbare uitkomsten komen: de kandidaat zelf, medekandidaten (peers) en assessoren (Shechtman & Sansbury, 1998). Ook deze vorm van alignment is een belangrijke factor ten aanzien van de acceptatie van selectieprocedures en -uitkomsten.

#### *Aandachtspunten*

In verschillende artikelen wordt nog de aandacht gevestigd op enkele specifieke aspecten van selectieprocedures.

#### Culturele bias

Turner en Turner (2000) constateerden dat de voorspellende waarde van de selectieprocedure die zij onderzochten niet voor alle culturele groepen even groot is. Dat vestigt de aandacht op mogelijke culturele bias van selectieprocedures. Dat kan met name betrekking hebben op taalvaardigheid, verbale capaciteiten, interpersoonlijke vaardigheden of leiderschapskwaliteiten.

Het risico van culturele bias is ook benadrukt in een Europese werkconferentie rond selectie voor de lerarenopleidingen (Thematic Working Group Professional Development of Teachers, 2012). Die bias kan zowel ontstaan bij de feitelijke selectie, als bij de vooraanmelding in de vorm van zelfselectie doordat bepaalde groepen zich bij voorbaat terugtrekken als zij weten welke aspecten in de selectie aan de orde komen. Als culturele diversiteit van lerarenteams in scholen een belangrijk streven is in het onderwijs, dan moet scherp gekeken worden naar een mogelijke culturele bias van de selectieprocedure en naar het effect op de samenstelling van de groep die toegelaten wordt voor de lerarenopleidingen. Ditzelfde geldt ook voor een mogelijke gender bias van selectieprocedures.

#### Variatie tussen programma's

Programma's van lerarenopleidingen kunnen onderling sterk van elkaar verschillen. Dat kan te maken hebben met de aard van het programma (undergraduate of postgraduate opleidingen hebben bijvoorbeeld te maken met instroom die sterk kan verschillen qua leeftijd en opleidingsniveau), maar ook met het type leraarschap waar voor opgeleid wordt. Van een leraar in de onderbouw van het basisonderwijs worden andere accenten in kwaliteiten verwacht dan van een leraar in de bovenbouw van het vwo of in het speciaal onderwijs (Holmgren & Covin, 1984). Dat kan dus tot verschillende accenten leiden in selectieprocedures en criteria voor verschillende typen opleidingen. Met andere woorden: de bevindingen met een bepaald type selectieprocedure zijn geldig voor het type

lerarenopleiding waarvoor die procedure is onderzocht, en kunnen niet worden gegeneraliseerd naar andere lerarenopleidingen.

*Betekenis voor de Nederlandse context*

Om na te gaan in welke mate de uitkomsten van de uitgevoerde literatuurreview betekenis heeft voor de Nederlandse context zijn de uitkomsten in concept versie voorgelegd aan een panel van drie ervaringsdeskundigen (zie hoofdstuk 2). Volgens dit panel sluit de uitkomst van de literatuurreview dat geïntegreerde selectieprocedures waarin een mix van selectiecriteria gehanteerd worden de grootste voorspellende waarde hebben, goed aan bij de huidige discussies over selectie in de context van lerarenopleidingen. Binnen veel pabo's wordt de selectie aan de hand van kennis ten aanzien van aardrijkskunde, geschiedenis, natuur en techniek en ten aanzien van taal en rekenen (binnen de propedeuse) als te beperkt en te weinig voorspellend ervaren voor het effectief handelen als leraar. Verschillende lerarenopleidingen experimenteren dan ook met bredere selectie-instrumenten. Een selectieprocedure met daarin een opdracht rond academische houding en vaardigheden (op basis van een leesopdracht, gevolgd door een college vakdidactiek en een toets), een kennistoets en een korte videopresentatie wordt komend jaar bij alle tweedegraads lerarenopleidingen Engels ingevoerd. Ook selectieprocedures zoals gehanteerd bij *Eerst de Klas* sluiten aan bij de bevindingen van de review. Daarmee ervaren de panelleden de uitkomsten als een bevestiging voor de ingeslagen weg. Tegelijk geven zij aan behoefte te hebben aan valide instrumenten die gebruikt kunnen worden om op een goede wijze persoonlijke vaardigheden en beroepsvaardigheden te kunnen beoordelen ten behoeve van selectie. Op dit moment domineert de kenniscomponent nog sterk in de gebruikte selectieprocedures. Tenslotte constateren zij dat er behoefte is aan meer onderbouwde kennis over de effecten van selectie voor de lerarenopleidingen in de Nederlandse context.

## Implicaties voor de Nederlandse praktijk

Uit de beperktheid van deze literatuurreview wordt duidelijk dat er weinig wetenschappelijke handvatten zijn voor het ontwerpen van goede selectieprocedures die een voorspellende waarde hebben ten aanzien van studiesucces in de lerarenopleidingen en ten aanzien van de kwaliteit van het handelen in de beroepspraktijk. Dat is op zich geen reden om niet over te gaan tot invoering van strengere selectie voor de lerarenopleidingen, maar het is wel aanleiding om dit met de nodige voorzichtigheid te doen. Dit sluit aan bij de terughoudendheid die bepleit is in ons eerdere rapport (Snoek, van der Rijst & van Driel, 2014).

De conclusie die uit de literatuurreview te trekken is - dat een selectieprocedure waarin aandacht wordt besteed aan een mix van variabelen de meest valide en betrouwbare vorm van selectie oplevert - sluit goed aan bij de experimenten die momenteel in verschillende opleidingen plaats vinden. De beperkte ervaringen met dergelijke selectieprocedures is in onze ogen aanleiding om lokale experimenten met elkaar te verbinden om zo kennis, instrumenten en ervaringen te bundelen en gezamenlijk kennis op te bouwen.

De nadruk die in de onderzochte studies gelegd wordt op een rijke vorm van selectie waarbij via een mix van instrumenten verschillende criteria en aspecten aan bod komen, leidt tot een beeld waarin assessoren een sleutelrol spelen in (groeps)interviews en (les)simulaties. Dat stelt hoge eisen aan assessoren (Bieri & Schuler, 2011; Shechtman, 1992). Selectie voor de poort vraagt daarmee om een investering in de kwaliteit van assessoren binnen hogescholen en universiteiten. Dat is niet alleen van belang bij de eventuele invoering van systematische selectie voor de poort, maar ook voor de huidige niet-selectieve intakeprocedures die universiteiten en hogescholen nu reeds hanteren en die leiden tot een advies aan aankomende studenten ten aanzien van hun geschiktheid voor de studie en het beroep. Deze aanbeveling geldt niet alleen voor de lerarenopleidingen; de kwaliteit van assessoren is ook bij selectieprocedures voor andere opleidingen van groot belang. Het werk dat reeds binnen de HBO-lerarenopleidingen gedaan is ten aanzien van assessorenkwaliteit en assessorentrainingen kan hier een belangrijke bijdrage aan leveren (Oudkerk Pool, 2013).

Het geringe aantal studies rond betrouwbare en valide selectieprocedures is aanleiding om te pleiten voor nader onderzoek, zoals dat bijvoorbeeld al gedaan wordt rondom de decentrale selectieprocedures die bij universitaire medische opleidingen zijn (en worden) ingevoerd. Daarbij hoeft niet gewacht te worden op de formele invoering van selectieprocedures bij lerarenopleidingen. Wij zien verschillende mogelijkheden:

- Sinds dit jaar worden binnen hogescholen en universiteiten 'studiekeuzechecks' (SKC) uitgevoerd, waarin studenten voorafgaand aan de opleiding een intakeprocedure doorlopen die leidt tot een advies ten aanzien van de geschiktheid voor de betreffende opleiding en daarbij passende beroepen. Deze studiekeuzecheck is te beschouwen als een selectieprocedure, echter zonder harde uitspraak.
- De tweedegraads lerarenopleidingen Engels gaan landelijk een systeem van decentrale selectie invoeren aan de hand van een vergelijkbaar assessment-instrument.
- De universitaire lerarenopleiding van de Universiteit Leiden is van plan om de selectieprocedure die bij Eerst de Klas gebruikt wordt, als basis voor een pilot te gebruiken voor de toelating tot de lerarenopleidingen geschiedenis.

Het zou in onze ogen uitermate waardevol zijn om rond deze ontwikkelingen en pilots systematisch en longitudinaal data te verzamelen rond de volgende vragen:

- Welke criteria hanteren opleidingen tijdens de selectie/intakeprocedure?
- Welke instrumenten hanteren opleidingen tijdens de selectie/intakeprocedure?
- Wat is de relatie tussen de uitkomsten van de selectie/intake en het studiesucces tijdens de opleiding?
- Wat is de relatie tussen de uitkomsten van de selectie/intake en het handelen in de beroepspraktijk (tijdens stages of na afloop van de opleiding)?

Vergelijkbaar onderzoek is uit te voeren rond de traditionele selectieinstrumenten bij de pabo om goed zicht te krijgen op de impact en voorspellende waarde van die instrumenten.

Om de voorspellende waarde goed te kunnen bepalen, is het noodzakelijk om longitudinaal te meten. Dat vraagt het verzamelen en vastleggen van data voor een langere periode. In veel gevallen zijn die data wel aanwezig, en nemen opleidingen ook initiatieven om cohortonderzoek te doen, maar die data zijn veelal alleen toegankelijk binnen hogescholen/universiteiten en de huidige onderzoeksinitiatieven beperken zich nog tot de eigen opleiding of eigen instelling. Het is van belang om deze data systematisch te ontsluiten, zodat ze als instellingsoverstijgende onderzoeksbron gebruikt kunnen worden. NRO zal een rol moeten spelen om dergelijk onderzoek te initiëren en te faciliteren.

## Referentias

- Andrew, M. D., Cobb, C. D., & Giampietro, P. J. (2005). Verbal Ability and Teacher Effectiveness. *Journal Of Teacher Education*, 56(4), 343-354.
- Archer, C. P. (1946). Recruitment, Institutional Selection, and Guidance of Teachers. *Review of Educational Research*, 16, 209- . DOI: 10.3102/00346543016003209
- Archer, C.P. (1949). Recruitment, Institutional Selection, and Guidance of Teachers. *Review of Educational Research*, 19, 191- . DOI: 10.3102/00346543019003191
- Barber, M., & Mourshed, M. (2007). *How the world's best-performing schools come out on top*. London: McKinsey.
- Barr, A.S. (1949). Measurement and Prediction of Teaching Success. *Review of Educational Research*, 19, 185- . DOI: 10.3102/00346543019003185
- Bieri, C., & Schuler, P. (2011). Cross-curricular competencies of student teachers: A selection model based on assessment centre admission tests and study success after the first year of teacher training. *Assessment & Evaluation In Higher Education*, 36(4), 399-415. doi:10.1080/02602938.2011.557146
- Black, W.J. & Page, M.J. (1958). Recruitment and Selection of Elementary-and Secondary-School Teachers. *Review of Educational Research*, 28, 198- . DOI: 10.3102/00346543028003198
- Brophy, J. (2004). *Motivating students to learn* (Second edition). Mahwah: Laurence Erlbaum Associates.
- Byrnes, D. A., Kiger, G., & Shechtman, Z. (2003). Evaluating the use of group interviews to select students into teacher-education programs. *Journal of Teacher Education*, 54(2), 163-172.
- Caskey, M. M., Peterson, K. D., & Temple, J. B. (2001). Complex admission selection procedures for a graduate preservice teacher education program. *Teacher Education Quarterly*, 28(4), 7-21.
- Darling-Hammond, L., & Lieberman, A. (Eds.), *Teacher Education around the world: Changing policies and practices*. New York: Routledge.
- Duckworth, A.L., Quinn, P.D. & Seligman, M.E.P. (2009) Positive predictors of teacher effectiveness, *The Journal of Positive Psychology: Dedicated to furthering research and promoting good practice*, 4 (6), 540-547, DOI: 10.1080/17439760903157232
- Durflinger, G.W. (1943). Scholastic Prediction in a Teachers College. *The Journal of Experimental Education*, 11(4), 257- .
- Girodo, M (1997). Undercover agent assessment centers: Crafting vice and virtue for impostors. *Journal of Social Behavior and Personality*, 12(5), 237-260.
- Gorow, T. R., Muller, S. M., & Schneider, S. R. (2005). The Relationship between Perceived Body Size and Confidence in Ability to Teach among Preservice Teachers. *Education*, 126(2), 364-373.
- Hamdy, H., Prasad, K., Anderson, M. B., Scherpbier, A., Williams, R., Zwierstra, R., & Cuddihy, H. (2006). BEME systematic review: Predictive values of measurements obtained in medical schools and future performance in medical practice. *Medical Teacher*, 28(2), 103-116.
- Heinz, M. (2013). Tomorrow's teachers – selecting the best: An exploration of the quality rational behind academic and experiential selection criteria for initial teacher education programmes. *Educational Assessment, Evaluation & Accountability*, 25, 93-114.
- Henry, J. A., Edwards, B. J., & Crotty, B. (2009). Why do medical graduates choose rural careers. *Rural Remote Health*, 9(1), 1083.
- Holmgren, B. R., & Covin, T. M. (1984). Selective Characteristics of Preservice Professionals. *Education*, 104(3), 321-28.
- King, P., Norrell, G., & Erlandson, F. L. (1959). The prediction of academic success in a police administration curriculum. *Educational and Psychological Measurement*, 19, 649 – 651.

- Kreiter, C.D. & Kreiter, Y. (2007) A Validity Generalization Perspective on the Ability of Undergraduate GPA and the Medical College Admission Test to Predict Important Outcomes, *Teaching and Learning in Medicine: An International Journal*, 19(2), 95-100, DOI: [10.1080/10401330701332094](https://doi.org/10.1080/10401330701332094)
- Kreiter, C.D. & Axelson, R.D. (2013). A Perspective on Medical School Admission Research and Practice Over the Last 25 Years. *Teaching and Learning in Medicine: An International Journal*, 25, 50-56, DOI:[10.1080/10401334.2013.842910](https://doi.org/10.1080/10401334.2013.842910)
- Oudkerk Pool, I. (2013). *Expertiseontwikkeling en Professionalisering van de Assessor. Ontwikkeling van een rubriek voor assessorenkwaliteit Criteriumgericht Beoordelen*. Amsterdam: Hogeschool van Amsterdam. (<http://www.hva.nl/kc-onderwijs-opvoeding/publicaties/content/publicaties-algemeen/expertiseontwikkeling-en-professionalisering-van-de-assessor.html>)
- Pintrich, R.P. (2003). A motivational science perspective on the role of student motivation in learning and teaching contexts. *Journal of Educational Psychology*, 95, 667-686.
- Salvatory, P. (2001). Reliability and validity of admissions tools used to select students for the health professions. *Advances in Health Sciences Education*, 6(2), 159-175.
- Shechtman, Z. (1992). Interrater reliability of a single group assessment procedure administered in several educational settings. *Journal Of Personnel Evaluation In Education*, 6(1), 31-39. doi:10.1007/BF00126918
- Shechtman, Z. (1998). Agreement between lay participants and professional assessors: Support of a group assessment procedure for selection purposes. *Journal Of Personnel Evaluation In Education*, 12(1), 5-17. doi:10.1023/A:1007932515469
- Shechtman, Z., & Sansbury, D. (1989). Validation of a group assessment procedure for the selection of teacher-education candidates. *Educational And Psychological Measurement*, 49(3), 653-661. doi:10.1177/001316448904900318
- Snoek, M., van der Rijst, R.M., & van Driel, J.H. (2014). *Verkenning naar de wenselijkheid en mogelijkheid van selectie op geschiktheid voor het beroep van leraar en van verzwaring van vakinhoudelijke eisen voor de tweedegraads lerarenopleidingen*. Leiden/Amsterdam: Universiteit Leiden/ Hogeschool van Amsterdam. [http://www.delerarenagenda.nl/wp-content/uploads/2014/10/Eindrapportage\\_Geschiktheid\\_Leraar\\_versie\\_1\\_1.pdf](http://www.delerarenagenda.nl/wp-content/uploads/2014/10/Eindrapportage_Geschiktheid_Leraar_versie_1_1.pdf)
- Thematic Working Group Professional Development of Teachers (2012). *Policies to select the best for teaching. Report of a Peer Learning Activity in The Hague, The Netherlands, 2 - 6 December 2012*. Brussels: European Commission.
- Turner, S., & Turner, T. (2000). Who Succeeds? A Case Study of Science Graduates Following Postgraduate Certificate in Education (Secondary) Courses. *European Journal of Teacher Education*, 23(2), 159-174.
- Valli, R., & Johnson, P. (2007). Entrance examinations as gatekeepers. *Scandinavian Journal Of Educational Research*, 51(5), 493-510. doi:10.1080/00313830701576631
- Van der Linden, D., Born, M., Phielix, L., & Touw, L. (2013). *Het selecteren van aspirant-agenten: Evaluatie van de selectieprocedure voor agenten op niveau 2 tot en met 4*. Rotterdam: Erasmus University.
- Viskic-Stalec, N., Sadura, T. & Horga, S. (1998). Predictive value of the entrance examination for the academic efficiency of students at the faculty of physical education in Zagreb. *Review of Psychology*, 5(1-2), 13-21.

## Bijlage 1: Uitgebreide samenvatting van studies naar integrale selectieprocedures

(n=6; chronologische volgorde)

### Shechtman, Z. & Sansbury, D. (1989). Validatie van een groepsbeoordelingprocedure voor de selectie van studenten voor een docentenopleiding.

<i>Studie</i>	In het onderzoek van Shechtman en Sansbury (1989) werd de voorspellende waarde van een groepsprocedure getest op het studiesucces in een docentenopleiding.
<i>Context</i>	Aanleiding van dit onderzoek waren de aanwijzingen in eerdere literatuur dat enkel het meten van cognitieve vaardigheden niet toereikend zou zijn voor het selecteren van succesvolle toekomstige docenten. De groepsprocedure bleek effectief te worden gebruikt voor de selectie van managementberoepen, waarvan de gemeten 'persoonlijke componenten' ook nodig zijn voor succesvolle docenten.
<i>Onderzoeksvraag</i>	Shechtman en Sansbury (1989) onderzochten in hun onderzoek of de scores van de groepsprocedure het studiesucces tijdens de docentenopleiding kon voorspellen. Verder onderzochten zij of de gevonden correlaties sterker waren dan de correlaties tussen de traditionele selectiemethode en het studiesucces van studenten. Ten slotte hebben de onderzoekers onderzocht of de globale score van de groepsprocedure extreme scores in studiesucces kon voorspellen.
<i>Methode</i>	Het onderzoek van Shechtman en Sansbury (1989) vond plaats op een grote docentenopleiding in Israël. De steekproef bestond uit 97 aanmelders die door middel van de groepsprocedure in 1979 of 1980 werden toegelaten tot de opleiding en die drie jaar later ook weer afstudeerde. Om deel te mogen nemen aan de groepsprocedure moesten zij eerst een IQ-test succesvol doorlopen.
<i>Analyse</i>	Om de voorspellende waarde van de groepsprocedure te testen werd er eerst een correlatieanalyse gedaan tussen de negen dimensies van de groepsprocedure en de vijf vaardigheden voor studiesucces. Ook werden de correlaties berekend tussen de drie traditionele selectiemethoden en de vijf vaardigheden voor studiesucces. Vervolgens werd een multi-pele regressieanalyse gedaan voor de voorspellende waarde van de groepsprocedure en traditionele 16PF schaal op de vijf variabelen voor studiesucces.
<i>Interventie</i>	De groepsprocedure was een 1 tot 1,5 uur durende sessie waarin vijf groepsactiviteiten werden gedaan: 1) een vrije groepsintroductie, 2) een gestructureerd groepsinterview gericht op houding en waarden, 3) een groepsdiscussie zonder groepsleider leidend naar een besluit, 4) het mondeling feedback geven onder participanten, 5) een discussie over de groepservaring door de aangemelde studenten. Elk groepje bestond uit acht deelnemers en zij werden beoordeeld door twee getrainde faculteitsleiders op een zes-puntenschaal.
<i>Afhankelijke en onafhankelijke variabelen</i>	De <b>groepsprocedure</b> is gebaseerd op 9 variabelen die worden gemeten, namelijk: globaal, verbale expressie, denkvaardigheden, motivatie, zelfvertrouwen, intermenselijke vaardigheden, leiderschap, flexibiliteit en creativiteit. Ter vergelijking van de groepsprocedure werden ook de scores van de studenten op traditionele selectiemethoden geanalyseerd in het onderzoek: De <b>Bagrut-a score</b> die bestond uit een combinatie van de GPA van de middelbare school en de scores van een aantal uitgebreide toetsen van het einde van de middelbare school. Een <b>IQ-score</b> die afkomstig was van een nationale IQ-toets van de onderwijsraad. Deze toets is verplicht voor alle kandidaten die zich aanmelden voor een docentenopleiding in Israël. De <b>16-PF scores</b> , afkomstig van acht schalen van de 16-PF-vragenlijst, namelijk intelligentie, innerlijke spanning, zelfovereenstemming, situationele gespannenheid, mensgerichtheid, assertiviteit en leiderschap, super-ego eisen en mate van lef/brutaliteit. Het <b>studiesucces</b> werd gemeten aan de hand van vijf variabelen: onderwijspraktijk evaluatie, motivatie voor het docentschap, intermenselijke vaardigheden,



	leiderschapsvaardigheden, en de GPA. De evaluatie van de onderwijspraktijk was de belangrijkste variabele in het meten van studiesucces, omdat deze score is gebaseerd op intensieve praktijkervaring in de klas en dit is geëvalueerd door experts op het gebied van docentenevaluatie.
<i>Bevindingen</i>	Uit de correlatieanalyse bleek dat een groot aantal van de 9 dimensies van de groepsprocedure significant correleerden met variabelen van het gemeten studiesucces (onderwijspraktijkevaluatie, GPA, leiderschap en intermenselijke relaties). Opvallend was dat geen van de dimensies van de groepsprocedure correleerde met motivatie van studenten. De rest van de correlaties waren echter zo sterk, dat de voorspellende waarde van de groepsprocedure in dit onderzoek sterk werd ondersteund. Dit was niet het geval voor traditionele selectiemethoden. Deze hadden slechts op enkele punten een significante correlatie met enkele dimensies van het studiesucces. Op basis van dit onderzoek is de groepsprocedure dus een betere methode dan het gebruik van traditionele selectiemethoden zoals de Bagrut-a score, een IQ-test en een 16-PF-schaal. Verder bleek dat de globale score uit de groepsprocedure bruikbaar was om de meest veelbelovende potentiële docenten te identificeren.
<i>Opmerkingen</i>	Het verschil tussen de 16PF vragenlijst en de groepsprocedure is interessant, omdat ze af en toe ongeveer dezelfde schalen meten (mensgerichtheid en leiderschap) en hierop de vragenlijst WEL significant correleert met het studiesucces → zijn dus belangrijke dimensies om te meten.

**Caskey, M., Peterson, K. & Temple, J. (2001). Complexe selectieprocedures voor een universitaire basisonderwijsopleiding.**

<i>Studie</i>	Het doel van het onderzoek van Caskey, Peterson en Temple (2001) was om een complexe toelatingsprocedure te testen voor een master docentenopleiding. Hierbij hebben zij gekeken naar de betrouwbaarheid en van de verschillende onderdelen van de procedure en de mate waarin deze de toelating beïnvloeden en het studiesucces tijdens de opleiding voorspellen.
<i>Context</i>	Caskey et al. (2001) waren voor hun onderzoek geïnteresseerd in de vraag hoe men het beste nieuwe studenten kan werven op basis van hun persoonlijke achtergrond en studieprestaties. Een interview is een populaire methode om studenten te beoordelen voor toelating, maar uit de door Caskey et al. (2001) beoordeelde literatuur blijkt het gebruik van enkel deze methode niet toereikend. Het onderzoek van Caskey et al. (2001) richtte zich daarom op het onderzoeken van een breed scala aan toelatinginstrumenten.
<i>Onderzoeksvraag</i>	Een van de onderzoeksvragen die Caskey et al. (2001) onderzochten en die interessant is voor deze review, was de mate waarin belangrijke selectievariabelen studiesucces kunnen voorspellen.
<i>Methode</i>	Het onderzoek van Caskey et al. werd uitgevoerd op een master docentenopleiding in Portland in de Verenigde Staten. Er namen 141 aanmelders deel aan het onderzoek en zij hadden allemaal een bachelortitel in een bepaald vak. Uiteindelijk werden er 82 kandidaten toegelaten tot de opleiding en zij zullen worden meegenomen in het onderzoek.
<i>Analyse</i>	Voor het onderzoeken van de voorspellende waarde van belangrijke selectievariabelen op het studiesucces van studenten, is een multiële regressieanalyse uitgevoerd.
<i>Interventie</i>	De interventie kan omschreven worden als het totaalpakket van een complexe selectieprocedure: Tijdens de simulatie-activiteit kregen de kandidaten de opdracht om in kleine groepjes van 4-6 studenten een fictief probleem op te lossen. Elk groepje kreeg 20 minuten de tijd om te overleggen en om een gezamenlijk antwoord te geven op het probleem (bijv. een lessenplan). Daarnaast moesten de kandidaten een persoonlijke motivatie aanleveren, waarin hun doelen, interesses, motivatie en achtergrond met betrekking tot het docentschap waren beschreven. Verder waren de kandidaten verplicht om drie referenties aan te leveren over de geschiktheid van de kandidaat voor het docentschap en de bekwaamheid van de kandidaat voor de masteropleiding.

	<p>Om te worden toegelaten moesten de kandidaten ook een nationale en gestandaardiseerde vragenlijst invullen die gericht was op de basisonderwijsvaardigheden voor basisonderwijskandidaten en de vakinhoudelijke kennis voor eerste- en tweedegraads kandidaten.</p>
<i>Afhankelijke &amp; onafhankelijke variabelen</i>	<p>De onafhankelijke variabelen in het onderzoek van Caskey et al. (2001) bestonden uit de zes selectievariabelen die een correlatie hadden met het besluit om studenten toe te laten tot de opleiding:</p> <p><b>Persoonlijke motivatie (statement):</b> twee beoordelaars gaven onafhankelijk scores voor de kwaliteit van de motivaties op een zevenpuntschaal. Ze keken hierbij naar de inhoud en naar de schrijfvaardigheid. De som van de twee onafhankelijke werd gebruikt als beoordelingsscore in het onderzoek.</p> <p><b>Referenties:</b> twee beoordelaars gaven onafhankelijk scores voor de kwaliteit van de referenties op een zevenpuntschaal. Ze keken hierbij met name naar referenties van oud-docenten die de student heeft zien lesgeven, maar ook naar referenties die werkervaring op het terrein van lesgeven beschreven. De som van de twee onafhankelijke werd gebruikt als beoordelingsscore in het onderzoek.</p> <p><b>Simulatie activiteit:</b> de beoordelaars waren allen werkzaam aan de faculteit of in het onderwijs en onderzoek. Zij observeerden de studenten en gaven hen scores op groeps- en procesvaardigheden, waaronder communicatie en idee-ontwikkeling. Zij werden beoordeeld op een vijfpuntschaal waarbij de beoordelaars in consensus kwamen over de beoordeling.</p> <p><b>Gestandaardiseerde testen:</b> de gemeten basisonderwijsvaardigheden waren 'lezen', 'schrijven' en rekenen' en de gemeten vakinhoudelijke kennis was gericht op 'algemene kennis' voor eerstegraads kandidaten en op 'specifieke vakkennis' voor tweedegraads kandidaten.</p> <p><b>Totale aanmeldingsbestand:</b> twee beoordelaars gaven onafhankelijk scores voor de globale kwaliteit van de aanmeldingsbestanden op een zevenpuntschaal. Hierbij werd gelet op de GPA, referenties, cursussen, test scores op de gestandaardiseerde testen, persoonlijke motivatie, ervaringen en unieke informatie die studenten aanleverden. De som van de twee onafhankelijke scores werd gebruikt als beoordelingsscore in het onderzoek. De afhankelijke variabele was de beoordeling die studenten kregen aan het eind van de opleiding van hun faculteit leiders.</p> <p>De afhankelijke variabele in het onderzoek van Caskey et al. (2001) was het studiesucces van studenten aan het eind van de eenjarige opleiding:</p> <p><b>Studieprestaties:</b> Aan het eind van de eenjarige opleiding werden de studenten beoordeeld op het criterium 'kwaliteit van prestaties in de opleiding' door twee faculteit leiders. Dit hield zowel de kwaliteit van deelname en prestaties in, als de kwaliteit als docent. De beoordelingsschaal was een vijfpuntschaal. De faculteit leiders kenden de studenten goed en waren ook betrokken bij het selectieproces.</p>
<i>Bevindingen</i>	<p>Caskey et al. (2001) vonden dat de voorspellende waarde van de selectievariabelen positief maar matig waren en dat deze met verdere aanpassingen veelbelovend was. De zes getoetste selectievariabelen verklaarden samen 38% van de variantie in de eindbeoordelingen. De schrijftoets uit de gestandaardiseerde test en de referenties wogen hierbij het zwaarst (<math>R^2=.304</math>, <math>R^2=.402</math>, <math>p?</math>). De simulatietoets is licht vergelijkbaar met de eerder besproken groepsbeoordelingsprocedure, maar bleek in dit onderzoek op zichzelf een zwakke voorspellende waarde te hebben voor een succesvolle eindprestatie (<math>R^2=.196</math>, <math>p?</math>).</p>
<i>Opmerkingen</i>	<p>Opvallend in dit onderzoek is dat de simulatietoets een vergelijkbare verklaarde variantie heeft als in het onderzoek van Byrnes et al. (2003). Toch is een schrijftest en het gebruik van referenties is hier een sterkere voorspeller. Samen zijn ze het sterkst.</p>

**Byrnes, D., Kiger, G. & Shechtman, Z. (2003). Evaluatie van het gebruik van groepsinterviews voor het selecteren van studenten voor de docentenopleidingen.**

<i>Studie</i>	Byrnes, Kiger en Shechtman (2003) onderzochten of een groepsbeoordeling aan het begin van de opleiding kan voorspellen hoe studenten in de opleiding gaan presteren en of een groepsbeoordeling een beter instrument is om de studenten te selecteren voor de opleiding dan de nu gebruikte academische criteria.
<i>Context</i>	Vaak worden studenten die zich aanmelden voor een docentenopleiding geselecteerd op basis van academische vaardigheden, zoals GPA en GRE-scores. Uit eerdere onderzoeken vonden Byrnes et al. (2003) dat verbale, interpersoonlijke en leiderschapsvaardigheden van belang zijn voor effectief docentschap. Het onderzoek van Shechtman en Sansbury (1989), hierboven besproken, in Israël liet eerder al zien dat een groepsprocedure die deze vaardigheden meet, een potentieel goed instrument is om studenten mee te selecteren. Byrnes et al. (2003) gebruikten dit instrument in hun studie om de validiteit en betrouwbaarheid ook in Utah (USA) te meten.
<i>Onderzoeksvraag</i>	Byrnes et al. (2003) onderzochten of een groepsbeoordeling aan het begin van de opleiding kan voorspellen hoe studenten in de opleiding gaan presteren en of een groepsbeoordeling een beter instrument is om studenten te selecteren voor de opleiding dan, nu vaak gebruikte, academische criteria.
<i>Methode</i>	Het onderzoek werd uitgevoerd op de docentenopleiding voor het basisonderwijs op de universiteit van Utah. Er namen 68 studenten deel aan de groepsbeoordeling, waarbij studenten in groepjes van acht werden beoordeeld door twee getrainde beoordelaars van de universiteit. In totaal namen elf verschillende docenten van de universiteit namen deel als beoordelaar. Daarnaast werden de GPA-scores van studenten verzameld en deden zij een ACT-toets om toegelaten te worden tot de opleiding. Nadat de studenten werden toegelaten volgden zij gemiddeld de tweejarige opleiding. Studenten werden tijdens of na de opleiding beoordeeld door een directe docent en een supervisor op hun studieprestaties.
<i>Analyse</i>	Byrnes et al. (2003) hebben een regressieanalyse uitgevoerd om te onderzoeken of een groepsbeoordeling een voorspellende waarde heeft met betrekking tot studieprestaties en om te onderzoeken of een groepsbeoordeling een beter instrument is om als toelatingscriteria te gebruiken dan de academische criteria. Daarnaast hebben de onderzoekers door middel van een correlatieanalyse onderzocht of de scores die verschillende beoordelaars geven met elkaar samenhangen en op die manier onderzocht of de groepsbeoordeling een betrouwbaar instrument is.
<i>Interventie</i>	De beoordelingsprocedure bestond uit een 90-minuten durende sessie waarin drie dimensies van docentgedrag werden geëvalueerd, namelijk: verbale, interpersoonlijke, en leiderschapsvaardigheden. Ook werd de gehele prestatie (overall rating) gemeten.
<i>Afhankelijke &amp; onafhankelijke variabelen</i>	De <b>prestaties van studenten</b> (afhankelijke variabele) werden beoordeeld door hun directe docent en door de supervisor van de universiteit door middel van evaluaties. De scores die hieruit voort kwamen zijn gebaseerd op de volgende criteria: effectief docentschap; klassenmanagement; kenniscontrole; persoonlijke en professionele ontwikkeling; ouder- en gemeenschapsbetrokkenheid; diversiteitsbewust; leerlinggericht; begrip van het curriculum. De scores konden variëren van 0 tot 4, waarbij score 4 gelijk staat aan de prestaties van een succesvolle en ervaren docent en 0 betekent dat de student onvoldoende scoort op alle competentiecriteria die hierboven genoemd zijn. De <b>academische criteria</b> (onafhankelijke variabele) werden gemeten aan de hand van de Grade Point Average-score en de American College Test-score. De GPA-scores hebben studenten behaald op een aantal verplichte algemene onderwijs cursussen voordat zij zich aanmeldden voor de docentenopleiding. De GPA werd gemeten op een vierpuntschaal. De ACT-test bestaat uit de onderdelen sociale wetenschappen, wiskunde en natuurwetenschappen en werd gebruikt als toelatingstest. De <b>groepsbeoordelingscriteria</b> waren gebaseerd op vier dimensies: verbale vaardigheden, interpersoonlijke vaardigheden, leiderschapskwaliteiten en een algemene score. Deze dimensies werden gemeten op een zes-puntschaal. Beoordelaars gaven onafhankelijk een score per dimensie, waarvan zij later één score maakten.

<i>Bevindingen</i>	<p>In het eerder besproken onderzoek van Shechtman en Sansbury (1989) bleek dat de groepsprocedure een goede voorspeller is voor succesvolle eindevaluaties van studenten in Israël. Uit het onderzoek van Byrnes et al. (2003) bleek dit ook het geval voor de deelnemende studenten in de Verenigde Staten. De algemene score bleek succesvolle beoordelingen van studenten het beste te voorspellen (12%-22% verklaarde variantie R<sup>2</sup>). De andere onderdelen van de groepsprocedure (verbale, interpersoonlijke en leiderschapsvaardigheden) hadden ook positieve associaties met succesvolle prestaties van studenten, maar deze waren niet significant.</p> <p>Wat opvallend was in het onderzoek van Byrnes et al. (2003), is dat niet alleen de academische criteria geen voorspellende waarde hadden op succesvolle beoordelingen van studenten, maar dat de ACT-toets zelfs negatief geassocieerd was met prestatiesucces van studenten. Hoe slechter studenten scoorden op deze toets, hoe aannemelijker het was dat deze studenten een hogere beoordeling kregen op de studieprestaties.</p>
<i>Opmerkingen</i>	De verklaarde variantie van de groepsprocedure is zwak, dus geen sterke voorspeller. Toch wordt in het onderzoek de groepsprocedure wel als sterk bevonden.

**Andrew, M. D., Cobb, C. D., & Giampietro, P. J. (2005). Verbale vaardigheden en effectief docentschap**

<i>Studie</i>	Andrew, Cobb en Giampietro (2005) onderzochten in hun studie de relatie tussen verbale vaardigheden en effectief leraarschap. Dit deden zij naar aanleiding van het debat tussen enerzijds aanhangers van de traditionele onderwijsprogramma's, die sterk geloven in de pedagogische training van leerkrachten, en anderzijds aanhangers van de school die gelooft dat eigenschappen, zoals verbale vaardigheden en inhoudelijke vak kennis, belangrijker zijn dan pedagogische training.
<i>Context</i>	<p>Uit een analyse van eerdere onderzoeken naar de relatie tussen verbale vaardigheden en effectief leraarschap concluderen Andrew et al. (2005) dat er een positieve relatie bestaat, maar dat deze met veel voorzichtigheid geïnterpreteerd moet worden. Zo zijn er veel kanttekeningen te plaatsen bij de testen die gebruikt zijn om de verbale vaardigheden vast te stellen in de onderzoeken en hebben de geanalyseerde onderzoeken veel methodologische beperkingen. Verbale vaardigheden werden in dit onderzoek gedefinieerd als de vaardigheid om ideeën om te zetten in woorden, zowel mondeling als schriftelijk. Daarnaast verwijst deze term naar de vaardigheid om woorden op een coherente manier te organiseren.</p> <p>Het onderzoek van Andrew et al. (2005) vond plaats in de Verenigde Staten op de universiteit van New Hampshire bij een master-lerarenopleiding voor basis- en secundair onderwijs (master).</p>
<i>Onderzoeksvraag</i>	Wat is de voorspellende waarde van verbale vaardigheden op effectief leraarschap van stagiaires tijdens hun stage als leerkracht?
<i>Methode</i>	De steekproef die Andrew et al. (2005) gebruikten voor hun onderzoek bestond uit 116 stagiaires die zijn toegelaten tot de lerarenopleiding aan de universiteit van New Hampshire. De masteropleiding was gericht op zowel reguliere bachelorstudenten die een vijfjarig programma volgen, als op post-bachelorstudenten die een verkorte opleiding volgen. Om de opleiding te behalen moesten zij vijf cursussen van het kernlesprogramma (vakspecifieke major) afronden en een stage van een jaar afronden. Deze kandidaten waren geselecteerd voor de opleiding op basis van hun scores op de GRE-testen (V: verbaal, Q: kwantitatief en A: analytisch) en hun <i>Bachelor-Grade Point Average</i> (GPA). 40% van de stagiaires beoogden een diploma voor het primair onderwijs en 60% van de participanten streefden een diploma voor tweedegraads leraarschap na. De stagiaires varieerden in specialisaties, waaronder wiskunde, Engels en geschiedenis en waren op het moment van het onderzoek in hun laatste fase van het masterprogramma, namelijk het afronden van een eenjarige stage met begeleiding van een ervaren docent en een supervisor vanuit de universiteit.
<i>Analyse</i>	Andrew et al. (2005) hebben een correlatieanalyse gedaan naar de relatie tussen de verschillende GRE-testen (V, Q en A) en de docentprestaties van stagiaires om te kijken of

	deze twee variabelen met elkaar samenhangen. Vervolgens is onderzocht in welke mate de GRE-V-test een voorspellende waarde heeft op de hier onderzochte prestaties van stagiaires.
<i>Interventie</i>	N.v.t.
<i>Afhankelijke variabelen &amp; onafhankelijke variabelen</i>	<p>De prestaties op effectief leraarschap van stagiaires is de variabele (afhankelijke variabele) waarvan Andrew et al. (2005) onderzoeken of deze afhankelijk is van de verbale vaardigheden van stagiaires (onafhankelijke variabele).</p> <p><b>Verbale vaardigheden</b> werden gemeten aan de hand van de GRE-V-test. Deze test is specifiek gericht op het meten van verbale vaardigheden van bachelor-afgestudeerden en wordt gebruikt als voorspeller van academisch succes in masteropleidingen. Volgens de makers van de GRE-testen, <i>The Educational Testing Service</i> (ETS), “test de GRE-V de vaardigheid om geschreven teksten te analyseren en te evalueren en relevante informatie samen te voegen; om relaties tussen componenten van zinnen te analyseren; en om relaties tussen woorden en concepten te herkennen” (ETS, 2003, p.4, als geciteerd in Andrew et al. (2005).</p> <p>Naast de GRE-V-test werden ook de GRE-Q (Kwantitatief) en de GRE-A (Analytisch) testen meegenomen in het onderzoek.</p> <p><b>Effectief leraarschap</b> werd gemeten aan de hand van de beoordelingscriteria die supervisors gebruiken om de prestatiedoelen en verwachte uitkomsten van hun stagiaires te beoordelen. Deze criteria zijn gebaseerd op zeven hoofdoelen van de opleiding (o.a.: ‘Onze afgestudeerden zullen deskundig zijn in het vak dat zij onderwijzen en in de manier waarop zij de vakkennis onderwijzen aan leerlingen’). Deze doelen zijn weer verder geoperationaliseerd in subdoelen.</p> <p>De supervisors werd aan het eind van het jaar gevraagd de stagiaires te beoordelen op een schaal van <i>acceptabel, goed, heel goed, of buitengewoon goed</i>. Andrew et al. (2005) beargumenteerden dat de professionele beoordeling door supervisors een van de betrouwbare en valide methodes is om effectief leraarschap te meten.</p>
<i>Bevindingen (verklaarde variantie)</i>	<p>Uit dit onderzoek blijkt geen overtuigend bewijs voor een relatie tussen verbale vaardigheden en effectief docentschap. Uit de correlatieanalyse bleek dat de GRE-V, de test om de verbale vaardigheden te meten, een zwakke en non-significante correlatie heeft met de prestaties van stagiaires. Het is dus onduidelijk of er daadwerkelijk een zwakke samenhang bestaat of dat deze samenhang op toeval berust. De GRE-A-test bleek wel een significante correlatie te hebben met de prestaties van de stagiaires. Deze test meet de analytische- en logische denkvaardigheden, oftewel de vaardigheden om relevante informatie van irrelevante informatie te scheiden en om geschikte keuzes te maken.</p> <p>Wanneer de groepen worden onderverdeeld in <i>acceptabel, goed, heel goed, of buitengewoon goed</i>, blijkt dat stagiaires die hierop hoger scoorden, ook substantieel hoger scoorden op de GRE-V-test. Deze verschillen zijn echter niet significant gebleken. Dit komt door de grote variatie van scores op verbale vaardigheden. Zo is er in elke beoordeelde groep sprake van lage scores.</p> <p>Andrew et al. (2005) concludeerden hieruit dat verbale vaardigheden wel degelijk belangrijk zijn om goed les te kunnen geven, maar dat verbale vaardigheden alleen/op zichzelf niet toereikend genoeg zijn.</p>
<i>Opmerkingen</i>	<p>De GRE-V test lijkt eerder schriftelijke taalvaardigheid te testen dan verbale vaardigheden, maar ook schriftelijke taalvaardigheid valt in dit onderzoek onder verbale vaardigheden. De validiteit van de GRE-V kan laag zijn wat zou verklaren waarom de scores hierop in het onderzoek niet significant zijn. Zo is er een grote spreiding in scores hierop voor elk type student.</p>

**Valli, R. & Johnson, P. (2007). Selectieprocedures als poortwachters.**

<i>Studie</i>	Valli en Johnson (2007) hebben onderzoek gedaan naar de waarde van een demonstratieles als selectiemethode voor een docentenopleiding in Finland.
<i>Context</i>	In Finland is er vanwege grote populariteit van docentenopleidingen weinig plaats voor alle kandidaten die zich aanmelden. Er bestaan hierdoor veel verschillende selectiemethoden in Finland. Het onderzoek van Valli en Johnson (2007) vindt plaats op het Chyndenius Instituut, gericht op onderwijs voor volwassenen. De eerste fase van de selectieprocedure werd gedaan aan de hand van aanmeldingsformulieren. Kandidaten voor de opleiding moeten een achtergrond in een universitaire studie hebben en minimaal vier maanden werkervaring. In de tweede fase van de selectieprocedure moesten alle kandidaten, ook degenen die niet aan bovenstaande criteria voldeden, een demonstratieles, een gesprek en een groepstaak doen om toelating te krijgen tot de opleiding. De opleiding duurde circa twee jaar.
<i>Onderzoeksvraag</i>	Valli en Johnson (2007) onderzochten in welke mate de selectieprocedure op het Chyndenius Instituut het studiesucces van studenten aan het einde van de opleiding voorspelt.
<i>Methode</i>	Het onderzoek van Valli en Johnson (2007) was gericht op 78 geselecteerde studenten (58 vrouwen en 20 mannen) die tussen 1997 en 2004 zijn toegelaten tot het Chyndenius Instituut in Kokkola in Finland.
<i>Analyse</i>	Valli en Johnson (2007) hebben een correlatieanalyse uitgevoerd.
<i>Interventie</i>	Tijdens de <b>demonstratieles</b> mogen de kandidaten kiezen uit een van drie curriculumonderwerpen voor hun les. Vervolgens kregen zij 45 minuten de tijd om de les voor te bereiden. De demonstratieles werd gegeven aan 10 tot 12 kinderen van 9 tot 11 jaar en mocht 10-15 minuten duren. De kandidaten werden geobserveerd door getrainde beoordelaars die hen beoordelen op de mate waarin zij het leren van leerlingen begeleiden en de effectiviteit van de interacties met de leerlingen op vijfpuntenschaal. Het <b>interview</b> werd gegeven door de meest ervaren beoordelaars. Zij hadden jarenlange ervaring met het doen van interviews. De interviews werden afgenomen door twee beoordelaars met als doel om een compleet beeld te krijgen van een persoon en om een voorstelling te krijgen van de kandidaat als docent. Voor aanvang moesten kandidaten een schrijfpodracht aanleveren met als onderwerp <i>Ik als docent</i> . De <b>groepstaak</b> bestond uit het onderzoeken van een vraag over onderwijs in het algemeen, een groepsdiscussie, een dialoog en een analyse. In groepjes van vier kandidaten kregen zij hiervoor een half uur de tijd. De onderzoekers geven in hun onderzoek duidelijk aan waar de zwakke punten van elk onderdeel van de interventie zitten. Dit probeerden zij te ondervangen door ervaren of getrainde beoordelaars te gebruiken
<i>Afhankelijke en onafhankelijke variabelen</i>	Het <b>studiesucces</b> werd gemeten aan de hand van twee afhankelijke variabelen: het onderwijspracticum en een masterthesis. Tijdens het onderwijspracticum werd gekeken naar de onderwijsvaardigheden in de klas. Bij het evalueren van de masterthesis werd gekeken naar de mate van wetenschappelijk redeneren en wetenschappelijk analyseren. De onafhankelijke variabelen is de <b>selectieprocedure</b> , bestaande uit de bovengenoemde demonstratieles, het interview en de groepstaak.
<i>Bevindingen</i>	De demonstratieles bleek het beste instrument om geschiktheid voor het docentschap te voorspellen. Studenten die hoog scoorden op de demonstratieles, scoorden ook hoog op onderwijsvaardigheden tijdens het onderwijspracticum aan het einde van de opleiding. Er werd geen verband gevonden tussen de demonstratieles en scores op de evaluatie van de masterthesis. Verder bleek dat de studenten die de hoogste totaalscores hadden in de totale selectieprocedure, ook de beste onderwijsvaardigheden hadden. Het omgekeerde gold voor de zwakste studenten. De verschillen tussen deze groepen bleek significant. Er bleek geen correlatie tussen de demonstratieles en de evaluatiescores van de masterthesissen. Zwakke studenten in de gehele selectieprocedure behaalden de laagste evaluatiescores op de masterthesissen, maar dit was niet significant. Valli en Johnson (2007) bevelen op basis van hun onderzoek aan dat de demonstratieles het beste als eerste fase kan worden gebruikt in de selectieprocedure. Dit houdt in dat de

	kandidaten die hier positief voor beoordeeld worden mogen doorgaan naar fase twee van de selectieprocedure, namelijk het interview en de groepsprocedure. Deze selectiecriteria zijn op zichzelf niet toereikend genoeg bevonden en daarom bevelen de onderzoekers het gebruik van een combinatie van selectiemethoden aan.
<i>Opmerkingen</i>	Er werden veel kanttekeningen genoemd voor de groepstaak in het onderzoek van Valli en Johnson (2007). Ook uit dit onderzoek blijkt dat het gebruik van verschillende selectiemethoden belangrijk is en dat voor het meten van interactievaardigheden een actieve methode belangrijk is. Het is in dit artikel echter niet helemaal duidelijk welke maten zijn bepaald en hoeveel de verklaarde variantie precies is.

**Bieri, C. & Schuler, P. (2011). Cross-curriculaire competenties van kandidaten voor docentenopleidingen: een selectiemodel dat gebaseerd is op toelatingstesten van beoordelingscentra en het studiesucces na het eerste jaar op de docentenopleiding.**

<i>Studie</i>	Bieri en Schuler (2011) onderzochten in hun studie of het succesvol doorlopen van een beoordelingsprocedure van toekomstige studenten voor de docentenopleiding, betekent dat zij ook succesvol zijn in hun studieloopbaan.
<i>Context</i>	Studenten zonder formele kwalificaties dienen een beoordelingsprocedure te doorlopen (gebaseerd op een competentiemodel bij een beoordelingscentrum) om toegang te verkrijgen tot de opleiding. Het doel van deze procedure is om inzicht te krijgen in de sterke en zwakke punten in werk-of studie-gerelateerde competenties van deze potentiële docent-studenten. Deze beoordelingsprocedure is gebaseerd op een competentiemodel dat ontwikkeld is binnen de docentenopleiding van Zürich. Het hier besproken onderzoek van Bieri en Schuler (2011) vond plaats in de Pädagogische Hochschule Zürich.
<i>Onderzoeksvraag</i>	De onderzoeksvragen van Bieri en Schuler (2011) luiden: 1) Tot welke hoogte hebben de variabelen opleidingsachtergrond en prestatie op de beoordelingsprocedure een voorspellende waarde voor het al dan niet slagen voor de examens aan het eind van het eerste studiejaar? en 2) 'Werden de studenten die hoog scoorden voor de beoordelingsprocedure ook als geschiktere docenten beschouwd door hun mentoren dan de studenten die een lagere prestatie lieten zien voor de beoordelingsprocedure na hun eerste studiejaar?'
<i>Methode</i>	De steekproef die Bieri en Schuler (2011) gebruiken voor hun onderzoek bestond uit 266 docent-studenten (197 vrouwen, 69 mannen) die de beoordelingsprocedure in de periode 2005-2007 succesvol hadden doorlopen.
<i>Analyse</i>	Om de invloed van de beoordelingsprocedure op de mate van succes van studenten na het eerste studiejaar te meten, hebben Bieri en Schuler (2011) een logistische regressie analyse uitgevoerd.
<i>Interventie</i>	De toekomstige studenten deden de opdrachten van de beoordelingsprocedure in groepjes van zes personen en werden geobserveerd door drie beoordelaars. De beoordelingsprocedure bestond uit vier onderdelen: 1) een groepsdiscussie zonder leider, waarvoor studenten individueel een plan voor een schoolproject moesten maken en vervolgens tot een overeenkomst met elkaar moesten komen over hun beslissingen. 2) een probleemoplossend-scenario waarin een constructie moest worden gemaakt met het hele groepje. 3) het houden van een korte presentatie over een complex probleem en deze informatie vervolgens evalueren voor het creëren van een lessenserie. 4) een gestructureerd interview met situationele en biografische vragen over motivatie voor het lesgeven en het leren.
<i>Afhankelijke variabelen &amp; onafhankelijke variabelen</i>	Bieri en Schuler (2011) onderzochten voor de eerste onderzoeksvraag of de beoordelingsprocedure een invloed hadden op het al dan niet slagen voor de toetsen na het eerste studiejaar. Voor hun tweede onderzoeksvraag onderzochten de onderzoekers of de beoordelingsprocedure een voorspellende waarde heeft voor de mate waarin de mentoren hun studenten als geschikte docenten beschouwden. De <b>docentcompetenties</b> werden gemeten tijdens de beoordelingsprocedure door

	<p>beoordelaars en vlak na en na het eerste studiejaar aan de hand van zelfrapportage-vragenlijsten. De volgende vijf competenties werden bepaald:</p> <p><i>Communicatie</i>: de mate waarin een individu geschreven of gesproken informatie duidelijk kan overbrengen.</p> <p><i>Coöperatie in termen van 'bewustzijn van anderen'</i>: de mate waarin de acties van een individu afgestemd zijn op de behoeften van anderen en de mate waarin een individu bewust is van de relevantie en de impact van zijn of haar handelen op de bredere context.</p> <p><i>Assertiviteit, waaronder het overtuigen van anderen</i>: de mate waarin een individu een ander overtuigt om iets te doen of een ander overtuigt om een bepaalde zienswijze over te nemen om op die manier bepaalde doelen te behalen. Ook wordt de mate gemeten waarin een individu zich laat leiden door zijn of haar eigen overtuigingen en zich niet makkelijk laat beïnvloeden.</p> <p><i>Motivatie</i>: de mate waarin een individu acties kan aanmoedigen over een langere periode om een bepaald doel te bereiken.</p> <p><i>Feitenonderzoek</i>: interpretatie van informatie en wordt gezien als onderdeel van kennisverwerving en omvat de volgende onderdelen: selectie, vergelijking, evaluatie, combineren en uitwisseling van informatie. Ook wordt rekening gehouden met de inclusie van voorkennis en probleemoplossend vermogen.</p> <p>Tijdens de beoordelingsprocedure geven drie beoordelaars individueel een score op een vierpuntschaap (zwak (1) – sterk (4)) op 6-10 items per gemeten competentie. Na de procedure voegden de beoordelaars hun scores na onderling overleg samen tot één beoordeling per competentie. De beoordelaars hadden voor de beoordelingsprocedure een gedrags-gedreven training van vijf uur gevolgd met video-simulatie oefeningen. De training was gericht op competentiegerichte observaties, classificeren, beoordelen en rapporteren. Het overgrote deel van de beoordelaars hadden een psychologische, pedagogische of onderwijskundige, of wetenschappelijke achtergrond.</p> <p>De zelfrapportage-vragenlijsten bestonden uit dezelfde items als die de beoordelaars van het beoordelingscentrum gebruikten voor het vaststellen van hun score en hadden een vierpuntschaal (helemaal mee oneens (1) – helemaal mee eens (4)). De mentoren van de studenten werd gevraagd dezelfde vragenlijst af te nemen.</p> <p>De <b>studieresultaten</b> aan het einde van het eerste studiejaar werden gemeten aan de hand van het behalen of niet behalen van een toets waarin de kennis van wiskunde, taal, onderwijs en psychologie werd gemeten.</p> <p>De <b>evaluatie van mentoren op studieprestaties</b> was gebaseerd op de visie van mentoren op de mate van beroepsgeschiktheid van hun studenten. Zij werden gevraagd een score te geven op de vraag: 'in welke mate ziet u de student als een geschikte docent?'. Daarnaast moesten zij een rapport schrijven over de ontwikkeling van de competenties van de studenten tijdens onderwijsactiviteiten gedurende de opleiding. Zo hebben studenten gedurende het studiejaar een dag per week les gegeven in een zogenaamde 'coöperatieve school'. Daarnaast kregen zij gedurende twee à drie weken een praktijkgerichte training in doceren en een speciaal-onderwijsstraining samen met hun mentoren. De mentoren waren de supervisors van de studenten tijdens hun opleiding en waren eerste- of tweedegraads opgeleid met een academische achtergrond in sociale wetenschappen, geesteswetenschappen of natuurkunde.</p>
<p><i>Bevindingen (verklaarde variantie)</i></p>	<p>Uit het onderzoek van Bieri en Schuler (2011) blijkt dat het selectiemodel van competenties, gebaseerd op principes van beoordelingscentra, een betrouwbaar instrument is om succesvolle studenten op de docentenopleiding te kunnen voorspellen. De resultaten uit de logistische regressie analyse laten zien dat de gemiddelde scores op de beoordelingsprocedure en opleidingsachtergrond een significante, maar matige invloed hebben op de prestaties van de studenten tijdens de examens na het eerste studiejaar (27% verklaarde variantie (R<sup>2</sup>). Daarnaast laat een vergelijking tussen studenten die hoog en laag hebben gescoord tijdens de beoordelingsprocedure zien dat studenten met lagere scores significant meer moeite hadden om te slagen voor de examens na het eerste studiejaar dan studenten met een hoge score. Dit verschil was significant (<math>X^2=15.15</math>,</p>



	<p>p&lt;.001). Ten slotte lieten scores van de mentoren zien dat studenten die hoger scoorden tijdens de beoordelingsprocedure voortdurend succesvol waren tijdens het eerste jaar. Deze studenten werden door hun mentoren significant vaker beoordeeld als geschikte docenten met hoge competenties, met name op het gebied van communicatie en assertiviteit na het eerste studiejaar, dan studenten die laag scoorden tijdens de beoordelingsprocedure.</p> <p>Bieri en Schuler (2011) concludeerden in hun artikel dat het meten van docentcompetenties in een beoordelingsprocedure een goede manier is om de toekomstige competenties van beginnende studenten te meten.</p>
<i>Opmerkingen</i>	<p>Ook in dit onderzoek blijkt een beoordelingsprocedure een matige maar valide methode om docentcompetenties te beoordelen bij kandidaten voor een docentenopleiding. De competenties die hiervoor gemeten zijn verschilden van de competenties die gemeten werden in de onderzoeken van Byrnes et al. en Shechtman en Sansbury, maar hebben ongeveer dezelfde voorspellende waarde. De manier die Bieri en Schuler gebruikten om studiesucces te meten verschilt ook van andere onderzoeken, waarbij deze betrouwbaar lijkt omdat studiesucces met twee maten gemeten wordt (een toets en een persoonlijke evaluatie).</p>