

De effecten van de evaluatie-ervaring op de percepties van studenten over assessment

K. Struyven, F. Dochy, S. Janssens, W. Schelfhout en S. Gielen

Samenvatting

Deze studie onderzoekt de effecten van de concrete ervaring met een ongekende evaluatievorm op de percepties van studenten ten aanzien van deze evaluatievorm en 'assessment'. Daartoe werd een cursus Ontwikkelingspsychologie ingericht waarbij vier evaluatievormen – namelijk een meerkeuzetoets, een casusexamen, een 'peer assessment' en een portfolio – werden toegekend aan vijf onderzoeksgroepen ($N = 612$). Data werden verzameld via een 'pretest-posttest design' met vragenlijsten. De resultaten tonen aan dat onbekende evaluatievormen aanvankelijk onbemind zijn. Onzekerheid over de procedure en de invloed op de prestaties leidt tot een kritische, weerhoudende attitude. Niettegenstaande ondervindt elke evaluatievorm na het examen een toename van positieve percepties, evenals verwante evaluatievormen. De proportionele toename verschilt echter van methode tot methode. Hoe positiever de ervaringen met de evaluatievorm, des te hoger is de toename van positieve percepties. Het nastreven van positieve ervaringen, zelfs met ongekende evaluatievormen, lijkt bijgevolg een aangewezen innovatiestrategie.

1 Inleiding

Dat het moeilijk is om veranderingen in het onderwijs te realiseren, is reeds bekend. Van den Berg en Vandenberghe (1999) stellen dat de percepties of betekenissen die betrokken personen aan innovaties geven van groot belang zijn voor het slagen van innovatieprocessen. Nieuwe situaties stoten op weerstand. Nochtans, naargelang succeservaringen zich opstapelen en percepties worden bijgesteld, worden subjectieve gevoelens van onzekerheid en passiviteit gereduceerd en heeft de innovatie kans op slagen. Deze principes van initiële weerstand en geleidelijke adaptatie staan centraal in dit onderzoek, met name de

effecten van de innovatie-ervaring (d.i. een ongekende evaluatievorm) op de percepties van studenten (d.i. de appreciatie van studenten ten aanzien van die evaluatievorm en gerelateerde evaluatiemethodes). Omdat instructie en evaluatie gerelateerd zijn, houdt de studie bovendien rekening met de onderwijssetting waarin de evaluatievorm wordt gebruikt en vergelijkt zij studentactiverende instructie met leerkrachtsturing.

2 Theoretisch kader

Studentactiverende didactische werkvormen worden sedert de doorbraak van constructivistische theorieën in onderzoek en praktijk (Kinnucan-Welsch & Jenlink, 1998; Oxford, 1997; Terwel, 1999; Vermunt, 1998; Von Glasersfeld, 1988) aangevoerd als verantwoorde methoden ter vervanging van leerkrachtgestuurde werkvormen, zoals doceren en onderwijsleergesprek. Deze laatste slagen er volgens deze theorieën betreffende leren zelden of niet in om studenten te brengen tot "actief", diepgaand, transitief en constructief leren (Bonwell & Sutherland, 1996; De Corte, 2000; Hatch & Farris, 1989; Holt-Reynolds, 2000; Kroll & Laboskey, 1996; Tynjälä, 1997). Studentactiverende werkvormen daarentegen, stellen deze doelstellingen centraal. Volgens de gehanteerde definitie, bestaat de activering van de student erin dat hij/zij via authentieke taken en praktijkrelevante opdrachten uitgedaagd wordt om beschikbare informatie actief te verwerken, dat wil zeggen selecteren, interpreteren en toe te passen op levensechte casussen en om complexe praktijkproblemen op te lossen (Jacobson & Mark, 1995; Meyers & Jones, 1993; Silberman, 1996; Tenenbaum, Naidu, Jegede, & Austin, 2001; White, 1996).

Deze veranderingen in didactische werkvormen brachten ook evoluties en innovaties in evaluatievormen teweeg. Gekende evaluatiepraktijken met korte antwoordvragen die

veelal peilen naar kennis en/of inzicht sluiten zelden aan bij de leerprocessen en vaardigheden die van studenten vereist worden binnen de voornoemde studentactiverende werkvormen (Birenbaum, 1996). Evaluatie binnen deze visie is een proces van kennisconstructie, eerder dan kennisreproductie (Sambell, McDowell, & Brown, 1997; Tynjälä, 1997). Evaluatievormen die deze doelstellingen nastreven, worden samengebracht onder de noemer *assessment*, als tegenhanger van *testen* of *toetsen*. Assessments omvatten activerende opdrachten, complexe cases of levens-echte probleemtaken die zowel individueel als collaboratief kunnen worden beantwoord (Davies & LeMahieu, 2003; Segers, 2003). Voorbeelden zijn: portfolio, peer assessment, casusgebaseerde evaluatie, simulatieoefeningen en assessmentcenters. Studenten kunnen daarbij vaak expliciet beroep doen op bronnen, zoals handboeken, tijdschriften, internet, etc. (Segers, Dochy, & Cascallar, 2003). Bovendien omsluiten deze evaluaties veelal zowel een formatieve functie (*assessment for learning*) als een summatieve functie (*assessment of learning*) (MacLellan, 2001). Het werken van studenten aan activerende taken kan bijgevolg worden beschouwd als onderdeel van het leerproces, en worden opgenomen in het evaluatieproces (Birenbaum, 2003; Black & William, 1998). Meer dan traditionele evaluatievormen, zoals schriftelijke of mondelinge geslotenboek-examens of meerkeuzetoetsen bevorderen en verdiepen deze assessments het leren van studenten (Birenbaum, 2003; Sambell, McDowell, & Brown, 1997).

Er is echter vaak een discrepantie tussen het onderwijs als bedoeld en het onderwijs zoals gepland door de leraar, of als beleefd door de betrokken studenten (Van den Berg & Vandenbergh, 1999). Niet zozeer de onderwijsrealiteit op zich, maar de realiteit zoals deze ervaren wordt door de student sorteert effecten op het leren van studenten (Entwistle, 1991). Niet zozeer de kenmerken van de leeromgeving op zich, zoals didactische werkvormen en verwachte evaluatiemethoden, zijn bepalend voor het studeergedrag, maar wel de manier waarop studenten deze werkvormen en evaluatievormen “ervaren” (bijvoorbeeld wat vinden ze hiervan? (Niet)

gepast? (Niet) leerrijk? etc.) (Birenbaum & Feldman, 1998; Elen & Lowyck, 2000). Dit betekent dat verschillende didactische ervaringen een differentiërend effect hebben op de percepties van studenten en uiteindelijk op het leren van die studenten. In dit onderzoek wordt nagegaan hoe de groeiende ervaring van studenten met een ongekende evaluatievorm een differentiërend effect heeft op de appreciaties van studenten ten aanzien van deze evaluatievorm en gerelateerde evaluatiemethodes. Vier evaluatievormen staan daarbij centraal: een meerkeuzetoets, een casusexamen, een peer assessment-procedure en een portfolio. Er wordt bovendien rekening gehouden met het verschil in instructievorm, namelijk leerkrachtsturing via hoorcolleges en studentactiverende werkvormen via authentieke casussen en probleemtaken. Specifiek werd de literatuur met betrekking tot evaluatievoorkeuren of ‘assessment preferences’ onder de loep genomen.

Wanneer studenten gevraagd wordt om twee evaluatiemethodes met elkaar te vergelijken, concludeert Zeidner (1987) dat studenten bij een examen de voorkeur geven aan meerkeuzevragen, waarbij het juiste antwoord moet worden aangestipt, boven essayvragen, waarbij het juiste antwoord door de student wordt uitgeschreven. Ook Traub en MacRury's (1990) constateren dat studenten de voorkeur geven aan de meerkeuzetoets als evaluatievorm, omdat deze volgens hen makkelijker is voor te bereiden, eenvoudiger is af te leggen, en studenten verwachten hoger te zullen scoren. Deze voorkeuren zijn volgens het onderzoek van Birenbaum en Feldman (1998) niet van toepassing voor de gehele groep van studenten. Deze onderzoekers ontdekten dat studenten met goede leerstrategieën, die bovendien een hoog vertrouwen stellen in hun academische kwaliteiten, essayvragen en -evaluatie verkiezen boven meerkeuzetoetsen. Omgekeerd prefereren studenten met zwakke leerstrategieën en die weinig vertrouwen hebben in het eigen kunnen, de meerkeuzetoets boven essayexamens. De resultaten tonen eveneens aan dat lage scores voor faalangst een significante samenhang vertonen met een positieve houding ten aanzien van essays, terwijl hoge faalangst-scores zijn gerelateerd aan eerder negatieve

percepties over open vragen en een voorkeur uiten voor meerkeuze- en dus gesloten vragen. In tegenstelling tot Zeidner vinden Birenbaum en Feldman wel een effect van geslacht, met mannelijke studenten die een meer uitgesproken voorkeur vertonen ten aanzien van meerkeuzetoetsen dan hun vrouwelijke collega's (Struyven, Dochy, & Janssens, 2005). Bij andere koppels van vergelijking vinden Sander, Stevenson, King en Coates (2000) dat een permanente evaluatie gebaseerd op taken ('coursework assessment') een beperkte voorkeur krijgt van studenten in vergelijking met examens aan het einde van de cursus ('end-of-course examinations'). Macdonald (2002) demonstreert gelijkaardige resultaten bij studenten in het hoger onderwijs, wanneer een gesloten-boekexamen vergeleken werd met een project dat summatief werd geëvalueerd. In tegenstelling tot hun collega's in het secundair onderwijs waren deze studenten in het algemeen tevreden over het projectwerk en de evaluatieve waarde ervan voor hun leerprestaties. Wanneer studenten de keuze kregen tussen feedback geschreven door de leerkracht of feedback verkregen via corrigerende software op de computer, na het schrijven van een paper en nadat ze beide soorten van feedback eerder kregen, vonden Dwyer en Sullivan (1993) dat 87% van de studenten de feedback van de leerkracht verkoos boven de feedback van de computersoftware, omdat deze persoonlijker was, de feedback makkelijker te begrijpen is, en de leerkracht als ultieme evaluator wordt gezien die de uiteindelijke leerprestaties bepaalt.

Studenten wordt echter niet alleen gevraagd om twee evaluatiemethodes met elkaar te vergelijken. Vaak wordt studenten een lijst van methodes voorgelegd waarbij ze via Likert-schalen hun appreciatie voor iedere evaluatievorm moeten aangeven. Op die manier vond Birenbaum (1997) dat de hoogste appreciatie werd geuit voor een "leerkracht-gestuurde voorbereiding van de examens" ('teacher guided test preparation'). Vervolgens kregen "alternatieve assessment en inspraak/participatie van studenten" hoge appreciaties, gevolgd door "permanente evaluatie en taken die hogere-ordedenken vereisen". De categorie die de laagste appreciatiescores kreeg, was "mondelinge exa-

mens", waarbij de gemiddelde score significant lager was dan de scores die de voorgaande categorieën kregen. Ook Zoller en Ben-Chaim (1997) vonden dat mondelinge examens de laagste voorkeurscores kregen van studenten, zowel in de Verenigde Staten als in Israël. Het hoge stressniveau en vermeende lage(re) resultaten die geassocieerd worden met deze evaluatievorm zijn volgens studenten de voornaamste redenen om dit evaluatietype weinig of niet te appreciëren. In tegenstelling met de resultaten van Birenbaum blijken de alternatieve, schriftelijke examenvormen, voor dewelke geen tijds-limieten zijn en alle bronnen mogen gebruikt worden, bovenaan de voorkeurslijst van de studenten te staan. Er zijn echter ook culturele verschillen: met name Amerikaanse studenten geven een significante, hogere appreciatie aan het traditionele schriftelijke examen in vergelijking met hun Israëlische collega's. Deze verschillen worden verklaard door de perceptie die studenten hebben over leren. Leren is volgens Israëlische studenten een middel om diepgaand en breed begrip te verkrijgen en levenssechte problemen op te lossen, terwijl Amerikaanse studenten leren beschouwen als een doel op zich. Deze verschillen in visie op leren worden eveneens weerspiegeld in de volgorde van evaluatievoorkeuren voor de overige evaluatiemethodes in de lijst.

Een belangrijk verschil tussen bovenvermeld onderzoek en de huidige studie is, dat de laatste de voorkeuren die studenten hebben ten aanzien van evaluatiemethodes niet beschouwt als stabiele, onveranderbare studentkenmerken. Integendeel, concrete ervaringen met ongekende onderwijspraktijken en evaluatiemethodes in het bijzonder, kunnen deze voorkeuren aanzienlijk veranderen. De onderwijsliteratuur levert bewijsmateriaal voor deze dynamiek. Bijvoorbeeld, Feigenbaum en Friend (1992) vinden dat de ervaring van studenten met grote klassen positief samenhangt met de wenselijkheid en waarschijnlijkheid van het volgen van lessen in een grote groep. Deze samenhang geldt niet voor het volgen van lessen in kleine groepen. Studenten die lessen volgen in grote groepen geraken eraan gewend (in vergelijking met secundair onderwijs) en kiezen ze uiteindelijk boven kleine lesgroepen. De onder-

zoeksresultaten van Aday en Campbell (1995) tonen deze dynamiek in percepties aan voor de werkplekvoorkeuren van studenten in een sociale opleiding naar aanleiding van het volgen van een cursus over gerontologie gericht op het veranderen van vooroordelen en vooronderstellingen over ouderen. Studenten aan het einde van de cursus hielden er minder negatieve stereotype opvattingen over ouderen op na in vergelijking met de start van de cursus. In relatie tot de percepties van studenten over assessment argumenteert Maclellan (2001) dat de visies die studenten hebben over evaluatie deprimerend zijn. Studenten gebruiken evaluatie zelden of niet als middel tot leren en houden er bovendien onderontwikkelde concepties over assessment op na. Het principe van initiële weerstand die evolueert naar positieve appreciatie, adaptatie en implementatie door positieve ervaring biedt interessante perspectieven in dit verband.

Het doel van dit onderzoek is aantonen dat de concrete ervaring van studenten met een ongekende evaluatiemethode een positief effect heeft op de percepties van studenten ten aanzien van die evaluatievorm en gerelateerde evaluatiemethodes, uiteraard op voorwaarde dat de ervaringen succeservaringen zijn. Bovendien wordt het differentiërend effect van vier evaluatiemethodes beschouwd, namelijk een meerkeuzetoets, een casusgebaseerd examen, een peer assessment-procedure en een portfolio, rekening houdend met de instructievorm. Twee concrete hypothesen worden daarbij getoetst:

- Ongekende evaluatiemethodes zijn onbeminde evaluatiemethodes.
- Naarmate de ervaring met de ongekende evaluatievorm toeneemt, veranderen de percepties van studenten positief. Dit wil zeggen, de voorkeuren (uitgedrukt in appreciaties) voor deze evaluatievorm vertonen een stijgende tendens. Uiteraard dienen ervaringen positief te zijn.

3 Onderzoeksdesign en -methode

3.1 Onderzoeksgroep

Voor het onderzoek werd een lessenpakket Ontwikkelingspsychologie uitgewerkt, be-

staande uit een infomap (Struyven, Sierens, Dochy, & Janssens, 2003), een werkmap (Gielen & Struyven, 2004a/b; Struyven & Gielen, 2004a/b) en een evaluatievorm. De cursus Ontwikkelingspsychologie werd gegeven aan eerstejaarsstudenten in de opleiding tot leraar Lager Onderwijs binnen acht Vlaamse lerarenopleidingen ($N = 608$).

3.2 Onderzoekscondities en leermaterialen

Elke hogeschool verbond zich tot een van vijf onderzoeksgroepen (zie Tabel 1). De eerste onderzoeksgroep kreeg les via hoorcolleges betreffende de infomap en werd geëvalueerd aan de hand van een meerkeuzetoets ($N = 114$). De werkmap werd in deze onderzoeksgroep niet gebruikt. De studentactiverende leeromgeving werd op eenzelfde manier gerealiseerd binnen de vier resterende onderzoeksgroepen ($N = 494$). Via authentieke, complexe probleemtaken en casusopdrachten in de werkmap werden studenten aangewezen op de infomap om de leerstof samen met hun collega-studenten te doorworstelen. Deze vier onderzoeksgroepen verschilden van elkaar door de evaluatievorm die het lessenpakket vergezelde. De vier evaluatievormen zijn: (1) de meerkeuzetoets ($N = 109$), (2) het casusgebaseerde examen ($N = 107$), (3) peer/coöperatieve assessment ($N = 172$) en (4) portfolio assessment ($N = 106$).

De cursus Ontwikkelingspsychologie bestond uit 10 lessen van ongeveer $1\frac{1}{2}$ uur, die volledig werden voorgestructureerd in de leermaterialen (d.i. infomap, werkmap en evaluatievormen). De *infomap* zorgde voor de inhoudelijke overeenstemming en werd gebruikt in de vijf onderzoeksgroepen. De informatie doorliep de ontwikkeling van het kind vanaf de conceptie tot en met de adolescentie, met de klemtoon op het lagere-schoolkind. In de leerkrachtgestuurde leeromgeving werden deze inhouden voorgesteld op transparante wijze die de lectoren leidden door de hoorcolleges Ontwikkelingspsychologie.

De *werkmap* diende om dezelfde studentactiverende werkvormen te realiseren in de vier activerende onderzoekscondities (A-Mk, A-Ca, A-Pe en A-Po). Week per week werden de lessen opgebouwd aan de hand van uitsluitend studentactiverende werkvormen.

Tabel 1

Onderzoeksgroepen, bijbehorende instructie- en evaluatievorm, en aantal studenten (N)

Onderzoeksgroep	Instructie (Lessen)	Evaluatie (Examen (+lessen))	N
H-Mk:	Leerkrachtgestuurde leeromgeving	meerkeuzetoets	114
A-Mk:	Student-activerende instructie	meerkeuzetoets	109
A-Ca:	Student-activerende instructie	casusexamen	107
A-Pe:	Student-activerende instructie	peer/coöperatieve assessment	172
A-Po:	Student-activerende instructie	portfolio assessment	106

Een brede waaier van methodes werd daarbij gebruikt, bijvoorbeeld probleemtaken, educatief spel, casussen, rollenspel en reflectietaken. De meeste taken moesten in groepen van zes tot acht studenten worden uitgevoerd. Studenten waren aangewezen op de inhouden in de infomap om de activerende taken tot een goed einde te brengen. Waar nodig, werden correctiesleutels voorzien, opdat studenten zelfstandig hun werk konden verbeteren. De rol van de leerkracht concentreerde zich vooral op het begeleiden van de groepswerken, bijvoorbeeld door vragen te beantwoorden, gesprekken te volgen en/of taken te verbeteren. Ter informatie vindt u in Tabel 2 een overzicht van de inhouden en geassocieerde activerende werkvormen voor de 10 lessen Ontwikkelingspsychologie.

Er werden vier *evaluatievormen* bij dit onderzoek gebruikt. Enerzijds was er de eerder traditionele evaluatievorm, namelijk de meerkeuzetoets, en anderzijds werden drie assessments geïmplementeerd, met name het casusexamen, peer en portfolio assessment.

Omdat de leerkrachtgestuurde leeromgeving geen gebruikmaakte van studentactiverende taken, werd uitsluitend de meerkeuzetoets aan deze instructievorm gekoppeld. Studenten hadden bij het begin van de studie geen of weinig concrete ervaring met de evaluatievormen in het onderzoek. Omdat de resultaten van dit onderzoek niet losstaan van de kwaliteit van de evaluatie, wordt iedere evaluatievorm uitgebreid beschreven.

De meerkeuzetoets

De meerkeuzetoets bestond uiteraard uit meerkeuzevragen. De deelnemende leerkrachten kregen een lijst met 60 mogelijke meerkeuzevragen en een verdeelsleutel. Op basis daarvan kon een examen met 20 vragen worden opgesteld. De verdeelsleutel zorgde voor een evenredige verdeling van vragen zowel naar inhoud (gespreid over de verschillende hoofdstukken) als naar soort van vragen (kennis-, inzichts-, toepassings- en probleemoplossingsvragen). Elke vraag bevatte vier keuzantwoorden waarvan slechts

Tabel 2

Inhouden en bijhorende didactische werkvormen voor de lessen Ontwikkelingspsychologie in de student-activerende condities

Les	Inhoud	Didactische werkvorm
Les 1	Inleiding bij lesaanpak/ evaluatie Inleiding in Ontwikkelingspsychologie	Klassikale instructie Inleidende casus Levenslijnopdracht Stellingspel
Les 2	Prenatale periode Pasgeborene en baby	Contractwerk Opdracht bij video
Les 3	Pasgeborene en baby Peuter en kleuter	Rollenspel Probleemtaak 1 (start)
Les 4	Peuter en kleuter	Probleemtaak 1 (afronding) Probleemtaak 1' (uitbreiding)
Les 5	Peuter en kleuter Lagerschoolkind	Probleemtaak 1' (afronding) Probleemtaak 2 (start)
Les 6	Lagerschoolkind	Probleemtaak 2 (afronding) Taak bij spelontwikkeling
Les 7	Lagerschoolkind	Hoekenwerk
Les 8	Lagerschoolkind Adolescentie	Reflectietaak Casus (start)
Les 9	Adolescentie Synthese/ Samenvatting	Casus (afronding) Educatief spel (ganzenbord)
Les 10	Informatie bij examen Gelegenheid tot vragen stellen	Klassikale instructie Klassikale instructie

één antwoord de gepaste oplossing representeerde. Een correctie voor raden werd ingebouwd: juiste antwoorden scoorden 1 punt, niet ingevulde antwoorden 0 punten en foutieve antwoorden kregen een score $-1/3^{\text{de}}$ punt.

In contrast met de meerkeuzetoets waren de drie assessments open-boekexamens waarbij studenten gebruik mochten maken van de infomap en andere bronnen die ze waardevol achtten.

Het casusexamen

Het casusexamen omvatte een set van fictieve documenten uit “jouw klas”, d.i. een fictieve klas uit het vierde leerjaar, zoals bijvoorbeeld: het puntenoverzicht van je leerlingen, een weekplanning, de plattegrond van het klaslokaal, brieven van ouders, een onderschepte liefdesbrief, een medisch verslag van de schoolarts, leerlingenobservaties en klasscreenings uit het leerlingvolgsysteem. Studenten kregen dit bundeltje met praktijkmaterialen mee aan het einde van de laatste les als studiemateriaal voor het open-boek-casusexamen samen met enkele voorbeeldvragen. De examenvragen bleven echter geheim tot het examen, maar waren allen gerelateerd aan de informatie in de casus, namelijk “jouw klas”.

De peer (coöperatieve) assessment

De peer assessment of coöperatieve assessment bestond uit twee delen, namelijk de taken en opdrachten tijdens de cursus en een mondeling groepsexamen betreffende het ingediende werk. Enerzijds werkten studenten tijdens het jaar aan de activerende opdrachten in groepen van zes tot acht studenten. Naar aanleiding van drie taken (probleemtaken 1, 2 en de casus in Tabel 2) werden er drie peer assessment-momenten voorzien binnen de opbouw van de cursus. Studenten werkten de opdracht in groepen af en dienden het groepsproduct in bij de leerkracht, die vervolgens een groepscore toekende. De peerscores maakten de groepscore tot een individuele score. Studenten moesten op een scoreformulier aangeven voor ieder groepslid en voor zichzelf of de persoon voor een bepaald criterium (bijvoorbeeld inbreng in de discussie, voorbereiding van gestelde leer-

doelen) een beter dan gemiddelde bijdrage leverde (score 3), een gemiddelde bijdrage (score 2), minder dan gemiddelde bijdrage (score 1), of geen bijdrage (score 0). Deze scores werden verrekend in de peer-factor, een persoonlijke multiplicator. Bovendien hield de procedure in dat voor elke persoon zowel de hoogste als de laagste peer-scores niet werden meegenomen in de berekening van de peer-factor, om eventuele vriendschap of vijandigheden te omzeilen. Groepsleden met een peer-factor lager, gelijk of hoger dan 1 zullen respectievelijk lager, gelijk en hoger scoren dan de door de leerkracht toegekende groepscore. Na elke opdracht en berekening van de peer-factor was een formatieve terugkoppeling van de factorscore voorzien, zodat studenten hun gedrag eventueel konden bijstellen bij een volgende groepsopdracht. Deze feedback werd door tijdsdruk van de leerkrachten echter niet gerealiseerd binnen de onderzoeksgroep. Anderzijds werd aan het einde van de cursus een mondeling openboek-groepsexamen gepland dat de groepsleden verder zou ondervragen betreffende het ingediende werk. De definitieve score voor de cursus Ontwikkelingspsychologie bracht de drie groepscores, gekoppelde peer-factoren en dit assessment-gesprek in rekening.

Het portfolio

De portfolio assessment bestond eveneens uit twee delen, namelijk enerzijds de constructie van het portfolio en het proces dat daarbij wordt doorlopen en anderzijds het individuele mondelinge assessment-gesprek bij het definitieve portfolio. Het portfolio kan het best voorgesteld worden als een soort verzamelmapp waarin studenten doorheen het verloop van de cursus hun taken presenteren, hierop reflecteren en deze koppelen aan eigen ervaringen. Aan het begin van de cursus werden er evaluatiecriteria afgesproken in samenspraak met de studenten. Deze vormden de basis voor de “tussentijdse feedback”-sessie, waarbij het (tijdelijke) portfolio op voorhand werd ingediend, besproken en eventueel bijgestuurd. Tijdens de examenperiode, ten slotte, onderging iedere student een individueel assessment-gesprek met de leerkracht over het definitieve portfolio. Pas

na dit gesprek werd er een definitieve score van het vak Ontwikkelingspsychologie toegekend die rekening hield met het doorlopen portfolio-proces tijdens de lessen.

Aan iedere assessment ging uiteraard een voorbereidende fase vooraf. Tijdens de eerste les van de cursus Ontwikkelingspsychologie werd de evaluatieprocedure klassikaal en gedetailleerd toegelicht. In de aanloop naar het examen werd deze informatie herhaald en werden bovendien voorbeeldvragen gedemonstreerd. Voorafgaand aan de peer assessment-procedure werd een 'try out'-versie gebruikt, die studenten konden invullen en waarbij de berekening van de multiplicator (en leerresultaat) onmiddellijk gebeurde. Op die manier wilden we studenten informeren over de evaluatie, zodat een gerichte voorbereiding kon plaatsvinden.

De leerresultaten voor de cursus Ontwikkelingspsychologie werden dus uitsluitend bepaald door het examen in de meerkeuzetoetscondities (H-Mk en A-Mk) en in de casusexamengroep (A-Ca), terwijl zowel bij de peer assessment (A-Pe) als bij het portfolio (A-Po) het werk van studenten aan de activerende taken in de leerprestaties werden verrekend. Bij deze laatste groepen (A-Pe en

A-Po) diende de assessment zowel een formatieve als een summatieve functie. Met uitzondering van de meerkeuzetoets waren alle examens of assessment-gesprekken "open boek". Studenten mochten zowel de informatie in de infomap als andere interessant geachte bronnen consulteren. De meerkeuzetoets en het casusexamen werden schriftelijk afgelegd, in tegenstelling tot de peer assessment en het portfolio waarbij de assessment-gesprekken mondeling gebeurden, respectievelijk in een groep en individueel. Deze gesprekken gingen voornamelijk over het ingediende werk, namelijk de probleemtaken en het definitieve portfolio. Tabel 3 geeft een overzicht van de belangrijkste kenmerken per evaluatiemethode. Het bovenste gedeelte van de tabel beschrijft de specifieke kenmerken van het examen of het assessment-gesprek, het onderste gedeelte daarentegen, bespreekt de kenmerken behorende bij de evaluatieprocedure. De procedure is uitgebreider dan het eindexamen en integreert veelal instructie en evaluatie.

Niettegenstaande de evaluatiemethoden gebaseerd waren op dezelfde inhoud en/of gerelateerde taken, zijn de doelstellingen en vaardigheden die worden vereist van studen-

Tabel 3

Overzicht van de kenmerken van de vier evaluatievormen

Kenmerk evaluatie	Meerkeuzetoets	Casusexamen	Peer assessment	Portfolio
Open/ gesloten boek examen	Gesloten	Open	Open	Open
Individueel/ groepsexamen	Individueel	Individueel	Groep	Individueel
Schriftelijk/ mondeling examen	Schriftelijk	Schriftelijk	Mondeling	Mondeling
Vragen van het examen handelen over?	Infomap	Casus (in relatie tot infomap)	Ingediende groepstaken (in relatie tot infomap)	Ingediende portfolio (in relatie tot infomap)
Vorbereiding op evaluatie tijdens lessen?				
Les 1	Toelichting	Toelichting	Toelichting + Try out	Toelichting
Tussentijds	/	Voorbeeld Casus Adolescentie	(Tussentijdse feedback) ^a	Tussentijdse feedback
Les 10	Voorbeeldvragen	Casus naar huis + voorbeeldvragen	Voorbeeldvragen	Voorbeeldvragen
Formatief/ summatief gebruik van evaluatie	Summatief	Summatief	Formatief + Summatief	Formatief + Summatief
Tussentijdse feedback	Neen	Neen	(Ja, 3x, na elke peer assessment) ^a	Ja, 1x, over het tussentijdse ingediende portfolio
Taken tellen mee?	Neen	Neen	Ja	Ja
Leerresultaten	Examen	Examen	3 Groepstaken + Peer assessment + Examen	3 Groepstaken + Individuele reflecties + Examen

^aDoor tijdgebrek, werd deze formatieve feedback niet gegeven aan studenten.

ten niet gelijklopend. Bij het zoeken naar de effecten van de evaluatie-ervaring op de voorkeuren van studenten over assessment wordt hier echter abstractie van gemaakt. Bij de bespreking van de commentaren van studenten (zie verder) worden deze verschillen expliciet opgenomen, omwille van de differentiërende rol en verklarende functies die ze dienen.

3.3 Meetmomenten en instrumenten

De dataverzameling over de percepties van studenten betreffende evaluatie werd gerealiseerd via een pretest-posttest design. Drie vragenlijsten werden gebruikt op drie meetmomenten (zie Tabel 4): (1) bij de start van de cursus Ontwikkelingspsychologie, onmiddellijk na Les 1; (2) aan het einde van de laatste les (Les 10) en (3) onmiddellijk na het examen of assessment-gesprek voor de cursus Ontwikkelingspsychologie (Examen). De vragenlijsten peilden algemeen naar de percepties van studenten en naar de assessment-voorkeuren en evaluatie van de verwachte/ervaren methode in het bijzonder. Er werden voornamelijk gesloten, maar ook open vragen gesteld.

De evaluatievoorkeuren van studenten voor verschillende evaluatiemethodes werden gemeten in de eerste (pretest 1) en de derde (posttest) vragenlijst. Op beide momenten kregen studenten een lijst met 13 identieke evaluatievormen voorgelegd, waarbij voor iedere methode de persoonlijke appreciatie werd aangegeven op een vijfpunts Likert-schaal. De antwoorden varieerden van *deze evaluatievorm kan ik volledig appreciëren*, naar *deze evaluatievorm kan ik in de meeste gevallen wel appreciëren; ik weet het niet; ik ben (nog) onbeslist; deze evaluatievorm kan ik in de meeste gevallen niet appreciëren; en deze evaluatievorm kan ik niet ap-*

preciëren. De 13 evaluatiemethodes die in de lijst werden opgenomen, waren: (1) meerkeuzetoets, (2) individueel schriftelijk (gesloten-boek)examen, (3) individueel mondeling (gesloten-boek)examen, (4) openboekexamen, (5) ‘take home’-examen, (6) praktisch examen (bijvoorbeeld stage, simulatie), (7) casusexamen, (8) paper/essay, (9) portfolio, (10) mondeling groepsexamen, (11) peer assessment, (12) zelfevaluatie en (13) coöperatieve assessment. Bij elke evaluatievorm werd een korte beschrijving gegeven, opdat studenten de termen op eenzelfde manier interpreteren.

De tweede vragenlijst (pretest 2) peilde eveneens naar de percepties van studenten over de ervaren cursus, meer in het bijzonder naar hun percepties betreffende de evaluatiemethode bij het examen Ontwikkelingspsychologie. Studenten werd gevraagd op een schaal weer te geven “wat ze vonden van de evaluatiemethode bij het vak Ontwikkelingspsychologie”. Eén van de volgende antwoorden werd aangeduid: *zeer goed, goed, matig, zwak en zeer zwak*. Er werd open ruimte voorzien om de gemaakte keuze eventueel schriftelijk toe te lichten.

De derde vragenlijst (posttest) vroeg studenten niet enkel naar hun appreciatie van de 13 evaluatiemethodes in de bovenvermelde lijst, hun werd eveneens gevraagd aan te geven of ze de evaluatievorm die ze ervoeren bij het examen Ontwikkelingspsychologie een “gepaste evaluatievorm vonden voor de cursus”. Een *ja-* of *neen-* antwoord moest worden aangestipt. Ook hier werd schrijfruimte voorzien voor een eventuele verantwoording bij de keuze. Meer zelfs, indien studenten de ervaren evaluatiemethode ongeschikt vonden, werd gevraagd om andere, meer gepaste evaluatiemethodes te suggereren.

Tabel 4

Meetmomenten en instrumenten voor de meting van de voorkeuren en appreciaties van studenten ten aanzien van evaluatie, en de ervaren evaluatievorm bij de cursus in het bijzonder

	Meetmoment		
	Les 1 (Pretest 1)	Les 10 (Pretest 2)	Examen (Posttest)
Instrument	Vragenlijst 1	Vragenlijst 2	Vragenlijst 3
Inhoud	Lijst met 13 evaluatievormen	Vraag naar <i>beoordeling</i> van evaluatievorm bij cursus	Lijst met 13 evaluatievormen Vraag over <i>geschiktheid</i> van ervaren evaluatievorm

Tabel 5

Beschrijvende statistieken van de uitgedrukte appreciatie van studenten voor de evaluatiemethodes uit de lijst bij de start van de cursus Ontwikkelingspsychologie (pretest 1)

Evaluatiemethode	N	M	SD	Volgorde
Individueel schriftelijk examen	668	4.24	.73	1
Praktisch examen (bv. stages)	667	4.10	.84	2
Individueel mondeling examen	665	3.78	1.12	3
Portfolio	662	3.66	1.14	4
Take home-examen	660	3.38	1.20	5
Openboekexamen	661	3.29	1.20	6
Paper/ essay	663	3.29	1.19	6
Zelfevaluatie	668	3.27	1.17	8
Coöperatieve assessment	665	3.22	1.25	9
Meerkeuzetoets	669	3.11	1.27	10
Casusexamen	661	3.03	1.06	11
Mondeling groepsexamen	667	2.70	1.23	12
Peer assessment	666	2.67	1.24	13

4 Resultaten

De resultaten worden beschreven in twee delen, naargelang de hypothese die ze testen. Vooraf wordt de algemene volgorde ('ranking') van de lijst met evaluatiemethodes getoond in Tabel 5, met als doel een algemeen beeld te schetsen betreffende de appreciaties van studenten voor de assessments op het ogenblik van de start van het project, namelijk Les 1.

De meest gekende en gebruikte evaluatiemethode, namelijk het individueel schriftelijk gesloten-boekexamen krijgt de hoogste appreciatiescores van studenten, gevolgd door het praktische examen en het individuele, mondelinge gesloten-boekexamen. De appreciaties werden gemeten op een vijfpuntsschaal, wat concreet betekent dat de evaluatiemethodes die lager scoren dan 3, eerder niet worden geapprecieerd door studenten, zoals het mondelinge groepsexamen en de peer assessment. Hoewel de meeste

assessments die het gebruik van bronnen toelaten (open-boek), thuis of tijdens het examen worden beschouwd als stress- en angst-reducerend (Zoller & Ben-Chaim, 1988), scoren ze slechts middelmatig in de voorkeurenlijst van studenten.

4.1 Hypothese 1: Ongekende assessments zijn onbeminde assessments

Zijn ongekende evaluatievormen initieel onbeminde evaluatievormen, zoals innovatietheorieën (Van den Berg & Vandenbergh, 1999) ons doen vermoeden? Wanneer studenten op het moment van de laatste les gevraagd wordt om de verwachte evaluatievorm voor de cursus Ontwikkelingspsychologie te evalueren, worden de resultaten in Tabel 6 gevonden.

Tabel 6 toont aan dat de leerkrachtgestuurde studenten (H-Mk) en hun collega's in de actieve portfolio groep (A-Po) het meest tevreden zijn met hun verwachte evaluatiemethode, met ongeveer de helft van

Tabel 6

De appreciatie van de "verwachte" evaluatiemethode op het moment van de laatste les in de vijf onderzoekscondities, weergegeven in percentages (N= 612)

Conditie	N	++	+	+ / -	-	--	Non-respons	Totaal
Leerkrachtsturing + Meerkeuzetoets (H-Mk)	114	6.14	46.49	35.96	8.77	1.75	.88	100
Activering + Meerkeuzetoets (A-Mk)	109	6.42	22.94	30.28	22.94	10.09	7.34	100
Activering + Casusexamen (A-Ca)	111	1.80	29.73	22.52	24.32	6.31	15.32	100
Activering + Peer assessment (A-Pe)	172	4.65	20.93	25.58	25.58	17.44	5.81	100
Activering + Portfolio (A-Po)	106	9.43	38.68	20.75	19.81	8.49	2.83	100

de studenten die de evaluatiemethode beschouwt als “goed” en “zeer goed”. Studenten in de actieve meerkeuzegroep (A-Mk), casusexamenconditie (A-Ca) en de peer assessment-groep (A-Pe) daarentegen, denken eerder in negatieve termen over de evaluatiemethode in het verschiep. De Cochran-Mantel-Haensel-statistieken wijzen op een statistisch significante algemene associatie tussen de variabelen *evaluatie van de verwachte methode* en *onderzoeksgroep* ($df = 20$, $GA = 83.4274$, $p < .0001$).

De verantwoording die studenten geven bij deze vraag, onthullen zowel argumenten in positieve als negatieve richting. Omdat de open vraag naar verantwoording vrijblijvend was, werd ervoor gekozen de antwoorden niet te kwantificeren. Niettegenstaande, geven deze antwoorden zicht op de karakteristieken (bijvoorbeeld eigenschappen, doelstellingen) die samengaan met de evaluatiemethode en de appreciatie ervan bepalen. Vandaar worden de meest voorkomende argumenten kort opgesomd. De afkortingen tussen haakjes geven de onderzoeksgroep(en) weer die dit argument aanhaalden.

Argumenten in de positieve richting zijn de volgende: “niet vanbuiten blokken” (A-Po), “het blijft niet bij theorie alleen” / “toepassing nodig” (A-Po, A-Ca), “leerrijk” (A-Po, A-Pe, A-Ca), “sluit aan bij mijn interesse en eigen ervaringen” (A-Po), “geen écht examen” (A-Pe, A-Po), “interessante manier van evalueren” (A-Pe), “leren omgaan met beoordelingen van anderen” (A-Pe), “openboek” (A-Ca), “opzoekwerk boeit” (A-Ca), “teveel info om te blokken” (A-Ca), “niet veel schrijven” (H-Mk, A-Mk), “denkvragen” (H-Mk), “vragen omvragen hele cursus” (H-Mk), “je weet of je alles begrepen hebt” (H-Mk, A-Mk) en “antwoorden worden aangeboden” / “juiste antwoord staat er al” (H-Mk, A-Mk). Argumenten in de negatieve richting daarentegen zijn eerder van deze aard: “veel werk” / “tijdrovend” (A-Po), “onvoldoende” / “wat ken ik van de inhoud?” (A-Po), “vriendjespolitiek” / “eerlijkheid?” (A-Pe), “degelijke evaluatie?” (A-Pe), “moeilijk” (A-Pe), “niet iedereen is altijd aanwezig” (A-Pe), “nooit inhouden echt geblokt” (A-Ca), “makkelijk” (A-Ca), “werkjes zouden moeten meetellen” (A-Ca, A-Mk),

“antwoorden liggen dicht bij elkaar” (H-Mk), “moeilijk” (H-Mk, A-Mk), “geen extra uitleg mogelijk” (H-Mk, A-Mk) en “teveel info om in te studeren” (H-Mk, A-Mk). Bij deze resultaten dient opgemerkt te worden dat de peer assessment-groep (A-Pe) en de portfolio-groep (A-Po), beiden reeds vertrouwd zijn met het formatieve deel van de evaluatie. Ook de leerkrachtgestuurde groep (H-Mk) legde reeds een meerkeuzetoets (partieel examen) af op het ogenblik dat deze vraag werd beantwoord.

4.2 Hypothese 2: Als de ervaring met de ongekende evaluatiemethode toeneemt, zullen ook de percepties van studenten positief veranderen

Voor de bewijsvoering van deze hypothese wordt het pretest-posttest design van de studie gebruikt, waarbij de lijst met evaluatievormen en geassocieerde appreciaties worden vergeleken. Gepaarde *t*-testen tussen Vragenlijst 1 (Les 1) en Vragenlijst 3 (Examen) voor alle evaluatievormen geven de evoluties in appreciatie weer, zoals gedemonstreerd in Tabel 7. Voor elke onderzoeksgroep werd het getal behorende bij de ervaren evaluatievorm en de gerelateerde methodes vet gedrukt. Bijvoorbeeld: de casusgroep (A-Ca) maakte uiteraard gebruik van het casusexamen, dat eveneens een openboekexamen was.

De resultaten zijn opvallend. Negatieve *t*-waarden wijzen op een toename aan appreciatie, terwijl de positieve *t*-waarden een afname van appreciatie aanduiden. Het blijkt dat iedere evaluatievorm in de onderzoeksofzet significant “profiteert” van de ervaring van studenten met die evaluatievorm, evenals de gerelateerde evaluatievormen, met uitzondering van het portfolio en het geassocieerde openboekexamen in de actieve portfolio-groep (A-Po).

De meerkeuzetoets wordt na het examen meer geprefereerd door de studenten in vergelijking met de start van de lessen in beide condities, H-Mk en A-Mk. Meer zelfs, de meerkeuzetoets werd de meest populaire evaluatievorm in de activerende groep (A-Mk) en neemt de tweede positie in in de Top 3 bij de hoorcollegestudenten (H-Mk). Dezelfde evolutie wordt vastgesteld bij de studenten in

Tabel 7

De voorkeuren van studenten voor elke evaluatievorm in de lijst voor de vijf onderzoekscondities, weergegeven in de gepaarde t-test-resultaten (Les 1 min Examen) en bijhorende significanties

	Gepaarde t-testwaarden (Les 1- Examen) en significanties				
	H-Mk	A-Mk	A-Ca	A-Pe	A-Po
Meerkeuzetoets	-2.44*	-7.98***	-.07	-3.02**	-.83
Individueel schriftelijk examen	2.49*	5.89***	2.69**	3.11**	5.42***
Individueel mondeling examen	3.30***	7.55***	3.50***	.25	-.31
Open-boekexamen	.37	-3.62***	-5.43***	-7.27***	.21
Take home-examen	1.53	-1.62	.08	-3.70***	-.09
Praktisch examen	7.66***	6.64***	4.98***	2.06*	2.26*
Casusexamen	3.52***	2.11*	-7.37***	1.18	-1.71
Paper/essay	4.46***	3.24**	3.57***	1.70	1.33
Portfolio	3.72***	3.12**	6.42***	2.88	.50
Mondeling groepsexamen	1.24	1.58	.07	-5.95***	.10
Peer assessment	1.35	3.99***	1.12	-3.26**	1.93
Zelfevaluatie	2.75**	2.07*	2.16*	-3.77***	1.88
Coöperatieve assessment	1.43	2.82**	4.20***	-4.17***	1.99

Noot. De vetgedrukte waarden duiden de betrokken en gerelateerde evaluatievormen aan per onderzoeksgroep, * $p < .05$; ** $p < .01$; *** $p < .001$.

de casusexamenconditie (A-Ca), waarbij zowel het casusexamen zelf als het geassocieerde open-boekexamen behoorlijk aan appreciatie winnen. De toename leidt ertoe dat deze twee evaluatievormen eveneens Top 3, zelfs Top 2 evaluatievormen zijn geworden. Hoewel de Top 3 niet bereikt werd, winnen de coöperatieve assessment en gerelateerde vormen, namelijk zelfevaluatie, peer assessment en mondeling groepsexamen, significant aan appreciatie na de concrete ervaring. Ook het open-boekformaat en het take home-examen winnen aan populariteit binnen deze conditie (A-Pe). Uitsluitend de portfolioconditie (A-Po) noteerde geen statistisch significante evoluties in appreciatie voor het portfolio, noch voor het open-boekexamen. Niettegenstaande, stijgen ook hier de gemiddelde appreciatiescores. Het portfolio bereikt zelfs de tweede plaats in de Top 3 van evaluatievormen binnen deze groep, na het individuele, mondelinge examen dat gelijkennissen vertoont met het individueel assessmentsgesprek betreffende het definitieve portfolio.

Opmerkelijk is dat de appreciatie-'levels' van de gekende evaluatievormen, zoals nummers 2 en 3 in de voorkeurslijst (Tabel 4), zijn gedaald. Meer bepaald, het individuele, schriftelijke (gesloten-boek)examen onderzocht een significante daling in appreciatie door de studenten, in alle onderzoeksgroe-

pen. Ook het voorkeursniveau van het individuele, mondelinge (gesloten-boek)examen is teruggevallen in de meerkeuzetoetsgroepen (H-Mk en A-Mk) en de casusexamengroep (A-Ca). Tot slot kende ook het praktische examen een opmerkelijke daling in appreciatie bij de studenten.

Het examen Ontwikkelingspsychologie was uiteraard niet het enige examen dat de studenten in de onderzoeksgroepen aflegden in de examenperiode. Verschillende resultaten in Tabel 7 lijken dan ook veroorzaakt door ervaringen van studenten die zich buiten het bestudeerde vak bevinden. Bijvoorbeeld: de leerkrachtgestuurde meerkeuzegroep (H-Mk) apprecieert het casusexamen, de paper of essay als evaluatievorm en het portfolio heel wat minder dan bijvoorbeeld hun collega's in de actieve meerkeuzeconditie (A-Mk). Deze laatste zijn dan weer meer het open-boekexamen gaan prefereren en geven minder de voorkeur aan de peer assessment na het examen. Ook in de andere groepen zijn er tal van voorbeelden te vinden waarvan de resultaten niet consistent zijn met de hypothesen in dit onderzoek. Bovendien zijn de evoluties verschillend voor de diverse onderzoeksgroepen. Niet onlogisch, want deze evoluties zijn veelal schoolafhankelijk. De verklarende factoren bevinden zich buiten de onderzoeksofzet en daarom worden deze resultaten niet verder geëxpliciteerd.

Tabel 8

ANOVA-analyses voor de vier evaluatiemethodes in het onderzoek, vanuit de lijst met evaluatievormen, voor de onderzoekscondities op beide meetmomenten (MM: L1 en Ex)

Evaluatiemethodes (condities)	MM ^a	df _{model}	df _{error}	F	p	F ²	Sign. verschillen ^b
Meerkeuzetoets (H-Mk and A-Mk)	L1 ^c	4	664	6.01	<.0001	.0349	A-Mk < H-Mk, A-Ca, A-Pe, A-Po
	Ex ^d	4	588	9.04	<.0001	.0579	A-Mk > A-Pe, A-Po, A-Ca H-Mk > A-Po, A-Ca
Casusexamen (A-Ca)	L1	4	656	.56	.6907	.0034	
	Ex	4	637	31.54	<.0001	.1653	A-Ca > A-Po, A-Pe, A-Mk, H-Mk
Portfolio (A-Po)	L1	4	657	2.29	.0588	.0137	
	Ex	4	636	12.50	<.0001	.0729	A-Po, A-Pe > A-Mk, H-Mk, A-Ca
Coöperatieve assessment (A-Pe)	L1	4	660	.31	.8729	.0019	
	Ex	4	631	13.19	<.0001	.0772	A-Pe > A-Po, H-Mk, A-Mk, A-Ca

^aMM = Meetmomenten. ^bBonferroni (Dunn) t-tests voor elk koppel van condities ($p < .05$). ^cL1 = Einde van Les 1. ^dEx = Onmiddellijk na het examen.

Tabel 7 bevestigt ook de eerste hypothese, namelijk dat ongekende evaluatievormen initieel ook onbeminde evaluatievormen zijn. Bijkomend bewijs wordt geleverd door de ANOVA's van de vier evaluatiemethodes in dit onderzoek op beide meetmomenten naar onderzoeksgroep, zoals weergegeven in Tabel 8.

Deze analyses demonstreren duidelijk de stijging in appreciatie van de evaluatievorm die door de betreffende studenten werd ervaren. Terwijl studenten bij de start van de cursus (L1) in de verschillende onderzoeksgroepen gelijkaardig denken over de evaluatievormen in termen van appreciatie, onderscheiden de onderzoeksgroepen zich van elkaar na het examen (Ex). Eén uitzondering is de actieve groep met een meerkeuzetoets (A-Mk) die op voorhand de meerkeuzetoets opmerkelijk minder apprecieert dan hun collega's in de vier andere onderzoeksgroepen. Het omgekeerde fenomeen wordt vastgesteld na het examen, waarbij beide meerkeuzetoetsgroepen (A-Mk en H-Mk) de meerkeuzetoets meer appreciëren dan de groepen met een andere assessment (resp. A-Pe, A-Po,

A-Ca en A-Po, A-Ca). Het casusexamen, de peer assessment en het portfolio onderscheiden op eenzelfde, significante wijze de studenten die de evaluatievorm ervoeren van de andere onderzoeksgroepen. Bij het portfolio echter, tonen de portfolio-studenten en de peer assessment-groep gelijkaardige percepties van appreciatie. De evoluties in voorkeuren bepaald door de ervaring met één evaluatievorm worden eveneens kracht bijgezet door de gestegen R^2 -waarden, die een maat zijn voor de verklaarde variantie en dus het effect van de onafhankelijke variabele. Tussen Les 1 en het Examen is werkelijk "iets" gebeurd.

Verondersteld werd dat evoluties in appreciatie of evaluatievoorkeur positief zouden zijn, indien de ervaring die opgedaan wordt met de ongekende evaluatievorm eveneens positief zijn. Om die reden werd studenten na het examen weerom gevraagd naar een evaluatie van de ervaren evaluatiemethode. Deze keer werd hun beoordeling betreffende de geschiktheid van de evaluatiemethode nagegaan. Er werd congruentie verwacht tussen enerzijds de evolutie in appreciatie en ander-

Tabel 9

De inschatting van studenten betreffende de geschiktheid van de ervaren evaluatievorm voor de doeleinden van de cursus, na het examen; cijfers zijn percentages (N = 625)

Conditie	Ja	Ja/Neen	Neen	Non-respons	N
Leerkrachtsturing + Meerkeuzetoets (H-Mk)	64.03	7.91	25.90	2.16	139
Activering + Meerkeuzetoets (A-Mk)	78.76	2.65	16.81	1.77	113
Activering + Casusexamen (A-Ca)	79.09	1.82	10.00	9.09	110
Activering + Peer (A-Pe)	52.55	6.12	39.80	1.53	196
Activering + Portfolio (A-Po)	80.60	8.96	10.45	0.00	67

Noot. De Cochran- Mantel- Haensel-statistieken tonen een significante associatie ($df = 16$, $GA = 85$. 8302 $p < .0001$).

zijds de percepties van studenten betreffende de geschiktheid van de ervaren evaluatievorm. Tabel 9 geeft de antwoorden van de studenten weer.

Na het examen is de meerderheid van de studenten overtuigd van de geschiktheid van de ervaren evaluatievorm voor de cursus Ontwikkelingspsychologie. De actieve meerkeuzetoetsstudenten (A-Mk), de casusexamenstudenten (A-Ca) en de studenten die werkten met het portfolio (A-Po) zijn zeer positief over respectievelijk de meerkeuzetoets, het casusexamen en het portfolio. Deze evaluatiemethoden stootten ook door tot de Top 3 in de appreciatielijst. De leerkrachtgestuurde studenten lijken iets minder enthousiast over hun meerkeuzetoets in vergelijking met hun geactiveerde collega's. Niettegenstaande worden de opmerkelijkste resultaten genoteerd bij de peer assessment-groep (A-Pe), waar slechts de helft van de studenten tevreden lijkt met deze methode voor de cursus Ontwikkelingspsychologie.

Wanneer studenten gevraagd werd bij ontevredenheid om alternatieve evaluatiemethodes voor te stellen, werden vooral de gebruikelijke methodes van het individueel schriftelijk of mondeling examen voorgesteld. Menigmaal werden ook suggesties voor verandering voorgesteld, zoals schrijfruimte bij de meerkeuzevragen om bijkomende verantwoording te geven, of een combinatie van meerkeuzevragen en open vragen in één examen.

5 Discussie en verklaringen

Wanneer de voorkeurenlijst van evaluatiemethodes in zijn algemeenheid wordt aanschouwd, dan prefereren studenten niet één methode, terwijl ze de andere eerder afwijzen. Integendeel, studenten lijken een mix van evaluatiemethodes te appreciëren. De mix die door de studenten in het onderzoek werd verkozen, toont gelijkenissen, maar eveneens duidelijke verschillen met de bevindingen in de onderwijsliteratuur. Bijvoorbeeld: wanneer studenten hun voorkeur moeten aangeven voor een meerkeuzetoets of een examen met open/essay-vragen, dan wordt de voorkeur gegeven aan de meerkeuzetoets (Traub

& MacRury, 1990; Zeidner, 1987). Het omgekeerde werd vastgesteld in de resultaten: zowel het individueel schriftelijk en mondeling examen als de paper of essay worden meer gewaardeerd dan de meerkeuzetoets. De verklaring ligt ten gronde aan de eerste hypothese, namelijk ongekende evaluatiemethodes zijn initieel onbeminde evaluatiemethodes. De meerkeuzetoets kent in Vlaanderen geen gangbare praktijk in het basis- en secundair onderwijs, gewoonlijk de studies die voorafgaan aan het eerste jaar van de lerarenopleiding. Bijgevolg hebben studenten hooguit vage voorstellingen van deze methode, zijn er allerminst vertrouwd mee, en appreciëren deze methode minder. Tegengesteld aan de bevindingen van Sander, Stevenson, King en Coates (2000) en Macdonald (2002) prefereren deze studenten methodes gebaseerd op het werken met taken zoals het portfolio of de paper/essay, niet boven de traditionele methodes, zoals het individuele schriftelijke of mondeling gesloten-boekexamen, tenminste niet aanvankelijk. Er is echter één uitzondering, het praktische examen (bijvoorbeeld stages, simulaties) dat om onbekende redenen erg hoog scoort in termen van appreciatie. Is het de beroepsgerichte oriëntatie van studenten in de lerarenopleiding die deze keuze beïnvloedt, of positieve ervaringen in de vooropleiding? Hoewel het praktische examen bij de start van de cursus een topper is in de voorkeurenlijst, heeft deze evaluatiemethode aan charmes verloren na het examen. Hebben studenten misschien ondervonden dat andere evaluatiemethodes ook praktische vaardigheden kunnen beoordelen of hadden studenten negatieve ervaringen met praktische evaluaties bijvoorbeeld op de stage? Ook de bevinding van Birenbaum (1997), Zoller en Ben-Chaim (1997) dat het mondeling examen het minst wordt geprefereerd door studenten, is niet van toepassing op de resultaten in dit onderzoek. Integendeel, het individueel mondeling examen staat zelfs in de Top 3 van de lijst. Misschien ligt het in de aard van het leerkrachtenpubliek? Geven toekomstige leraren liever of makkelijker een mondelinge toelichting bij een vraag dan dat ze een schriftelijk betoog neerpennen? Het mondeling groepsexamen spreekt studenten weinig

aan, sterker nog, studenten appreciëren deze evaluatievorm zelden. Ook de peer assessment-methode wordt erg laag ingeschat. Vermoedelijk speelt de afhankelijkheid van de leerprestaties van andere personen buiten de leerkracht – de ultieme evaluator volgens de studenten in de studie van Dwyer en Sullivan (1993) – daarbij een bepalende rol. Bovendien bevestigen de resultaten niet de constatering van Birenbaum dat studenten alternatieve vormen van assessment prefereren boven eerder traditionele evaluatiemethodes. Integendeel, de resultaten in de lijst tonen eerder overeenkomsten met de Amerikaanse studenten in de studie van Zoller en Ben-Chaim, die vooral geïnteresseerd lijken in traditionele evaluatievormen als methodes om hoge punten te scoren, gegeven de vermeende lagere inspanningen die deze evaluatiemethodes vereisen in vergelijking met alternatieve assessments. Of is het de vertrouwde die studenten doorheen hun vooropleiding met deze evaluatiemethodes hebben opgebouwd die leidt tot de hogere appreciatiescores? Deze laatste verklaring zou alleszins overeenkomstig de hypothesen in deze studie zijn.

Onbekende evaluaties zijn onbemind. Onzekerheid en het “niet weten wat te verwachten” verklaart vermoedelijk de lage inschattingen van de verwachte evaluatiemethode bij de actieve meerkeuzetoets (A-Mk) en bij het casusexamen (A-Ca) tijdens de laatste les. Vooral de scores bij de studenten met een meerkeuzetoets (A-Mk) zijn opmerkelijk. Waarschijnlijk schrok de boodschap dat de evaluatie zou gebeuren met een meerkeuzetoets deze studenten erg af. Niettegenstaande, verdwenen deze zorgen als sneeuw voor de zon, na de eigenlijke ervaring van de meerkeuzetoets op het moment van het examen. Niet alleen stijgt de appreciatie van studenten voor deze evaluatievorm in beide onderzoeksgroepen (H-Mk en A-Mk), de meerkeuzetoets stoot door tot de eerste en tweede positie in de Top 3 van methodes bij deze studenten.

Daarnaast stond de dynamiek van de percepties van studenten ten aanzien van evaluatie en naar aanleiding van een concrete evaluatie-ervaring met een ongekende evaluatievorm centraal in de onderzoeksopzet.

De resultaten tonen duidelijk deze dynamiek aan die in het verlengde ligt van de resultaten van Feigenbaum en Friend (1992), Aday en Campbell (1995). De evaluatie van studenten betreffende de verwachte evaluatievorm laat deze evoluties al vermoeden. Zowel de leerkrachtgestuurde meerkeuzegroep (H-Mk) als de portfolio-groep (A-Po) zijn beide enthousiast over de evaluatiemethode. De verklaring ligt bij de argumentatie die gestalte geeft aan de tweede hypothese, namelijk dat naarmate de ervaring groeit, ook de percepties van studenten positief evolueren. De studenten in de leerkrachtgestuurde leeromgeving hadden reeds een meerkeuzetoets afgelegd tijdens de partiële examens, waardoor hun initiële appreciatie reeds hoger lag in vergelijking met hun geactiveerde collega's (A-Mk). Ook het formatieve luik dat gekoppeld is aan het portfolio (A-Po) maakte dat studenten reeds uitgebreid kennismaakten met het werken aan een portfolio tijdens de cursus Ontwikkelingspsychologie en waren waarschijnlijk hard bezig aan de afwerking van hun definitieve portfolio op het ogenblik van de bevraging. Studenten hadden tussentijds ook een assessment-gesprek over hun portfolio-ontwikkeling en waren daardoor reeds vertrouwd met het individuele assessment-gesprek met de lector over het definitieve portfolio. Opvallend is echter dat dezelfde redenering ook van toepassing moet zijn op de peer assessment-onderzoeksgroep (A-Pe), wat echter niet blijkt uit de resultaten.

Ook de beschrijvende statistieken en gepaarde *t*-testen voor de pretest-posttest lijst van evaluatiemethodes toont deze dynamiek en de differentiërende invloed van de verschillende evaluatievormen. Want hoewel elke evaluatiemethode “profiteert” van de evaluatie-ervaring tijdens het examen en dus wint aan appreciatie, zijn er verschillen in de mate waarin de methodes winnen aan populariteit. Zo is de peer assessment-procedure de enige evaluatievorm in het onderzoek die na de ervaring niet doorstoot tot de Top 3 van assessments. De commentaren van studenten suggereren dat problemen zoals subjectiviteit, vriendjespolitiek en afwezigheden deze percepties in de hand werkten. Ook wanneer studenten wordt gevraagd om de geschiktheid van het examen in te schatten, is de

groep van peer assessment-studenten de laagst scorende. Een plausibele verklaring is het door tijdgebrek ontbreken van formatieve feedback naar studenten tijdens de lessen. Een logisch gevolg is dat studenten niet weten hoe deze evaluatiemethode de leerprestaties gaat beïnvloeden; ze weten niet hoe (en hoe eerlijk) hun medestudenten de geleverde bijdrages waardeerden, noch hoe de groepswerken evolueerden, noch hoe de peer-factoren de eigen leerprestaties gaan beïnvloeden. Zelfs op het ogenblik van het examen zijn studenten onwetend hierover, met de gemiddeld lagere appreciaties voor de ervaren methode in vergelijking met de andere groepen tot gevolg. Een adequate(re) toepassing van het formatieve luik van de peer assessment had betere appreciaties kunnen uitlokken. Hoewel beide meerkeuzegroepen (H-Mk en A-Mk) niet significant verschillen van elkaar in appreciatiescores na het examen, wordt vastgesteld dat de geactiveerde studenten (A-Mk) de geschiktheid van de meerkeuzetoets hoger inschatten dan hun leerkrachtgestuurde collega's (H-Mk). Vermoedelijk heeft de autonomie die lectoren kregen om de meerkeuzevragen te selecteren, ondanks het respecteren van de verdeelsleutel, geleid tot een verschillende invulling van het examen met een andere beleving door studenten als gevolg. Vooral de praktische invulling van de vragen werd door de geactiveerde studenten erg gewaardeerd, terwijl leerkrachtgestuurde studenten eerder commentaren uitten zoals "moeilijke, verwarrende vragen", "antwoorden gelijken op elkaar" of "soms weet ge het antwoord op een deelvraag, maar niet op twee deelvragen en dus op de vraag" (bij dubbele vragen).

Hoewel de gegevens in deze studie de hypothesen bevestigen, blijkt eveneens dat niet iedere student in de groepen er dezelfde percepties op nahoudt. Hoewel tendensen worden onderkend, tonen de resultaten ook antwoorden in tegengestelde richting, bijvoorbeeld erg lage appreciaties. Een gevarieerde mix van verschillende evaluatievormen lijkt dan ook aangewezen, niet alleen om tegemoet te komen aan verschillende doelstellingen, maar ook aan de verscheidenheid van studenten. Variatie is mogelijk binnen een vak, maar is zeker wenselijk binnen een

vakdomein of binnen een bepaalde opleiding. Eerder dan ons af te vragen "what is preferred by most?", moet ook de vraag "who prefers what = which personal characteristics affect students' assessment preferences and how?" worden gesteld (Birenbaum, 1997). Een combinatie van enerzijds deze studentkenmerken die recht doet aan de heterogeniteit van studenten en anderzijds de doelstellingen die nagestreefd worden met bepaalde evaluatiemethodes, maken best het didactische assessment-beleid van een leerkracht, leerkrachtengroep of school uit. Vooral de samenwerking met vak- en schoolcollega's moet de praktische haalbaarheid van deze "mix" garanderen, zowel voor leerkrachten als voor studenten.

Bovendien suggereren de resultaten van dit onderzoek dat, gegeven de ontdekte dynamiek van percepties van studenten, zeker wanneer het gaat om ongekende assessments, het weinig zinvol is om evaluatievoorkeuren te bestuderen als stabiele studentkenmerken (geslacht, nationaliteit, leerstijlen) zoals gebeurde bij Zoller en Ben-Chaim (1997), Birenbaum (1997) en Birenbaum en Feldman (1998). Deze voorkeuren hangen, zoals aangetoond, onlosmakelijk samen met de ervaringen van studenten met de evaluatievorm of het ontbreken van deze praktische expertise. Op termijn (lees: na veel ervaringen) kan stabiliteit wel worden verwacht, zo blijkt uit de continuerende populariteit van gekende evaluatiemethodes zoals het individueel schriftelijk of mondeling examens. Deze klassiekers werden als eventuele alternatieve, meer gepaste methode het meest vermeld door studenten.

Een veronderstelling die in dit onderzoek wordt gemaakt, is dat ervaringen leiden tot veranderende percepties, waarbij positieve ervaringen positievere percepties veroorzaken, en negatieve ervaringen zullen leiden tot een voortzetting (of verslechtering) van de initieel kritische percepties bij ongekende, en dus onbeminde, evaluatievormen. Maar wat maakt een concrete ervaring tot een positieve ervaring? De commentaren van studenten geven suggesties in de gewenste richting. Bovendien worden verschillende evaluatiemethodes geapprecieerd om uiteenlopende redenen (bijvoorbeeld antwoorden niet ver-

sus wel uitschrijven, praktische toepassingen versus inhouden “vanbuiten blokken”, objectiviteit versus leren van elkaar, etc.). Andere plausibele verklaringen, die gestalte geven aan een positieve ervaring, waaraan wordt gedacht zijn bijvoorbeeld de grote hoeveelheden aan aandacht en tijd die wordt besteed aan de evaluatiemethode tijdens de lessen of de hoge inschatting van de resultaten door studenten na de evaluatie-ervaring. Aangezien instructie en evaluatie op elkaar afgestemd waren in de studie, geeft ook het instructieproces en leerproces van de studenten gestalte aan de appreciaties van studenten. Ongetwijfeld zal de mening van de leerkracht of de aandacht en tijd die hij/zij aan de evaluatie geeft, beïnvloedend werken. Hoe lineair deze relatie echter is, is bevestigbaar. Niet alleen kozen de leerkrachten in de studie voor een bepaalde evaluatievorm (i.e. gelijkwaardige motivatie), bovendien zouden de appreciaties van de evaluatievormen die formatieve en summatieve evaluatie (A-Pe en A-Po) combineren hoger moeten zijn dan de appreciaties van de summatieve eindexamens (A-Ca, H-Mk en A-Mk) en dat wordt niet door de resultaten bevestigd. Ook de resultaten van studenten voor de ervaren evaluatievorm kunnen bepalend zijn. De studie beschikt echter niet over de “effectieve” of “gepercipieerde” resultaten van studenten voor de cursus Ontwikkelingspsychologie, dus daarover kunnen geen uitspraken worden gedaan. Er zijn echter wel de resultaten op onverwachte gestandaardiseerde testen voor deze cursus (Struyven, Dochy, Janssens, Schelfhout, & Gielen, in press). Indien de percepties overeenkomstig zijn met de eindresultaten op deze test, dan zouden de meerkeuzetoetsgroepen (H-Mk en A-Mk) hogere appreciaties vertonen ten opzichte van de casusstudenten (A-Ca), peer (A-Pe)- en portfolio (A-Po)-groepen. Deze volgorde wordt niet bevestigd door deze resultaten. De testen waren echter “gesloten boek”, wat mogelijk de meerkeuzestudenten heeft bevoordeeld. Nochtans, indien de resultaten inderdaad bepalend zijn, dan zouden ook grote verschillen tussen de andere groepen bijvoorbeeld peer assessment (A-Pe)- en casus/portfolio (A-Ca en A-Po)-groepen verwacht worden; verschillen die in de testen niet significant zijn.

Tot slot, kunnen er ook andere verklaringen worden geopperd die onafhankelijk zijn van de positieve waarde van de evaluatie-ervaring, zoals bijvoorbeeld “is het plausibel dat het enkel de naambekendheid is, die studenten een evaluatiemethode meer doet appreciëren en niet zozeer de ervaring op zich?”. Een uitdagende vraag, die niet eenvoudig kan worden weerlegd. Een tegenargument wordt gevonden bij de meerkeuzetoets. Ook al hebben studenten weinig ervaring met de meerkeuzetoets als evaluatievorm in het basis- en secundair onderwijs, studenten zijn niet onbekend met meerkeuzevragen, bijvoorbeeld in psychometrische testen of vragenlijsten allerlei. Bovendien roept de naam *meerkeuzetoets* op zich al een beeld op. Toch wordt ook hier de dynamiek vastgesteld. Meer zelfs, de geactiveerde studenten die geen ervaring hadden, apprecieerden de meerkeuzetoets gevoelig minder dan de andere groepen die eveneens geen (A-Ca, A-Pe, A-Po) of beperkte ervaring (H-Mk) hadden, alsof de aankondiging van de meerkeuzetoets deze studenten extra onzekerheid en angst inboezemde. Ook het differentiatievermogen van de verschillende evaluatiemethoden pleit tegen de hypothese dat het dynamische effect enkel door “naambekendheid” wordt veroorzaakt.

6 Conclusies

Innovaties zijn vaak angstaanjagend wanneer ze op ons afkomen. Nochtans, wanneer de ervaringen met vernieuwingen positief zijn, veranderen onze percepties in dezelfde positieve richting, zoals werd aangetoond voor de voorkeuren in assessment naar aanleiding van de concrete ervaring met één van vier ongekende evaluatiemethodes. Het omgekeerde, namelijk de persistentie (of groei) van kritische gevoelens en negatieve percepties kan eveneens worden verwacht indien de ervaringen van studenten negatief zijn, zoals gedeeltelijk wordt geobserveerd bij de peer assessment-studenten (A-Pe), die zich beduidend minder positief uitspreken over de ervaren evaluatiemethode. De hypothese dat naarmate de evaluatie-ervaring groeit ook de percepties van studenten positief evolueren,

wordt dus (grotendeels) bevestigd door de resultaten.

Opmerkelijk is dat deze conclusie van toepassing is op zowel de eerder klassieke vormen van evaluatie, namelijk de meerkeuzetoets in dit onderzoek, als de nieuwe vormen van assessment, zoals het casusexamen, de peer assessment en het portfolio. Gegeven de onpopulariteit en de kritieken die de meerkeuzetoets ondergaat (Zeidner, 1987), zijn deze resultaten verrassend. De verklaring wordt gevonden in de kwaliteit van het instrument. De commentaren van studenten suggereren dat de oriëntatie van de meerkeuzetoets op kennisconstructie – en niet (enkel) op kennisreproductie zoals vaak wordt aangenomen – wordt gewaardeerd door studenten. De vragen in de categorieën *toepassing* en *probleemoplossing* belichamen deze oriëntatie en maakten “dat het niet enkel om theorie ging” en “veel toepassingen = goed”. Misschien komt klassieke toetsing toch beter tegemoet aan de vereisten van nieuwe vormen van instructie dan voorondersteld door Birenbaum (1997), of heeft het tenminste dat potentieel. Omgekeerd kan een lage kwaliteit van het evaluatie-instrument, of in ons geval een slechte uitvoering van de assessment-procedure, leiden tot eerder matige effecten, zoals het ontbreken van de voorziene formatieve feedback bij de peer assessment. Ook Clifford (1999) vindt deze ambigue percepties van studenten naar aanleiding van inadequate gebruik van het instrument. Bijvoorbeeld: studenten uitten frustraties over de tijd die werd geïnvesteerd in het bepalen van de evaluatiecriteria en sommige studenten vonden dat leerkrachten hun verantwoordelijkheden ontlieden. Studenten voelden zich eveneens onzeker over hun eigen en andermans capaciteiten om elkaars werk te beoordelen.

Deze bevindingen over het gebruik en de kwaliteit van de evaluatiemethodes belichten een eerste belangrijke beperking van de onderzoeksopzet in deze paper, namelijk: de conclusies van deze studie zijn (enkel) van toepassing voor de methodes zoals ze werden geoperationaliseerd. De kwaliteit evenals een (in)adequaat gebruik van de instrumenten bepalen de resultaten van deze studie. Bijvoorbeeld: dit onderzoek maakte gebruik van een

meerkeuzetoets die een correctie voor raden omvatte, ook het portfolio en de peer assessment zijn onlosmakelijk verbonden met de activerende opdrachten die werden uitgevoerd, en het casusexamen heeft betrekking op één casus en diens kenmerken. Bovendien werden de evaluatievormen verbonden aan een instructiemethode, die eveneens deel uitmaakt van de operationalisatie van de “evaluatiemethode” en dus bijgevolg ook haar invloed heeft. Conclusies betreffende een bepaalde evaluatiemethode moeten voorzichtig worden aangewend. Tot slot zijn evaluatievoorkeuren slechts één manier om percepties van studenten te bestuderen. Voorzichtigheid bij de generalisatie van deze resultaten lijkt bijgevolg aangewezen. Replicaties en verificatie van de resultaten, ook in andere onderwijssettings, zijn nodig.

Een tweede limitatie die met de voorgaande samenhangt, is dat de huidige resultaten besluiteloos zijn betreffende de processen, de doelstellingen en de karakteristieken die iedere evaluatiemethode onderbouwen. Verschillende vragen kunnen worden gesteld: welke kwaliteiten en kenmerken van de instrumenten hebben de positieve of negatieve ervaringen van studenten uitgelokt? Of welke leerprocessen hebben zich afgespeeld tijdens het voorbereiden van en/of het werken aan een bepaalde evaluatiemethode? En welke leerresultaten komen daaruit voort? De commentaren van studenten geven hooguit enkele richtingen en suggesties voor toekomstig onderzoek, maar toekomstig onderzoek dat een antwoord zoekt op bovenstaande vragen blijft noodzakelijk indien verdergaande conclusies ten aanzien van deze methodes en hun instrumenten het doel is. Dit aankomend onderzoek vormt een zinvolle aanvulling op de huidige constatering.

Tot slot werden de studenten binnen een onderzoeksgroep als groep onderzocht, en werden studentkenmerken, leerkrachtkenmerken, schoolkenmerken of eventuele andere onafhankelijke variabelen buiten beschouwing gelaten. Vragen zoals: zijn er gender-effecten? of is er een relatie tussen de leerprestaties van studenten en hun assessment-voorkeuren? of in welke mate heeft de leerkrachtstijl, de klassfeer of de collegialiteit tussen leerkrachten de voorkeuren van

studenten beïnvloed? blijven onbeantwoord. Ook de invloed van de instructievorm is niet eenduidig te bepalen in deze onderzoeksopzet, aangezien er maar één vergelijkingsgroep is. Toekomstig onderzoek dat deze vragen adequaat kan beantwoorden, wordt warm aanbevolen. Tot die tijd is aangetoond dat de concrete evaluatie-ervaring een (voornamelijk) gunstige invloed heeft op de percepties van studenten ten aanzien van deze evaluatievorm. Hoewel het zoeken naar de concrete factoren die gestalte hebben gegeven aan de “succeservaring” blijft voortduren, zijn de resultaten van dit onderzoek bemoedigend voor actoren die het onderwijspraktijkveld willen verrijken met innovatieve projecten.

Literatuur

- Aday, R. H., & Campbell, M. J. (1995). Changes in nursing students' attitudes and work preferences after a gerontology curriculum. *Educational Gerontology, 21* (3), 247-260.
- Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievements, learning processes and prior knowledge. Evaluation in education and human services* (pp. 3-29). Boston, MA: Kluwer Academic Publishers.
- Birenbaum, M. (1997). Assessment preferences and their relationship to learning strategies and orientations. *Higher Education, 33*(1), 71-84.
- Birenbaum, M. (2003). New insights into learning and teaching and their implications for assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 13-36). Dordrecht: Kluwer Academic Publishers.
- Birenbaum, M., & Feldman, R. A. (1998). Relationships between learning patterns and attitudes towards two assessment formats. *Educational Research, 40*(1), 90-97.
- Black, P., & William, D. (1998). Inside the Black Box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*(2), 139-148.
- Bonwell, C. C., & Sutherland, T. E. (1996). The active learning continuum: Choosing activities to engage students in the classroom. *New Directions for Teaching and Learning, 67*, 3-16.
- Clifford, V. A. (1999). The development of autonomous learners in a university setting. *Higher Education Research and Development, 18*(1), 115-128.
- Davies, A., & LeMahieu, P. (2003). Reconsidering portfolios and research evidence. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: in search of qualities and standards* (pp. 141-170). Dordrecht: Kluwer Academic Publishers.
- De Corte, E. (2000). Marrying theory building and the improvement of school practice: A permanent challenge for instructional psychology. *Learning and Instruction, 10*(3), 249-266.
- Dwyer, H. J., & Sullivan, H. J. (1993). Student preferences for teacher and computer composition marking. *Journal of Educational Research, 86*(3), 137-141.
- Elen, J., & Lowyck, J. (2000). Instructional meta-cognitive knowledge: A qualitative study on conceptions of freshman about instruction. *Journal of Curriculum Studies, 32*(3), 421-444.
- Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment. Introduction to the special issue. *Higher Education, 22*, 201-204.
- Feigenbaum, E., & Friend, R. (1992). A comparison of freshmen and upper division students' preferences for small and large psychology classes. *Teaching of Psychology, 19* (1), 12-15.
- Gielen, S. & Struyven, K. (2004a). *Groot worden. De ontwikkeling van baby tot adolescent (Werkboek voor toekomstige leerkrachten Basisonderwijs, Basisdeel)*. Leuven: LannooCampus.
- Gielen, S. & Struyven, K. (2004b). *Groot worden. De ontwikkeling van baby tot adolescent (Werkboek voor toekomstige leerkrachten Secundair Onderwijs, Basisdeel)*. Leuven: LannooCampus.
- Hatch, D. H., & Farris, C. R. (1989). Helping TAs use active learning strategies. *New Directions for Teaching and Learning, 39*, 89-97.
- Holt-Reynolds, D. (2000). What does the teacher do? Constructivist pedagogies and prospective teachers' beliefs about the role of a teacher. *Teaching and Teacher Education, 16*(1), 21-32.

- Jacobson, T. E., & Mark, B. L. (1995). Teaching in the information age: Active learning techniques to empower students. *Reference Librarian, 51-52*, 105-120.
- Kinnucan-Welsch, K., & Jenlink, P. M. (1998). Challenging assumptions about teaching and learning: Three case studies in constructivist pedagogy. *Teaching and Teacher Education, 14(4)*, 413-427.
- Kroll, L. R., & Laboskey, V. K. (1996). Practicing what we preach: Constructivism in a teacher education program. *Action in Teacher Education, 18(2)*, 63-72.
- Macdonald, J. (2002). Getting it together and being put on the spot: synopsis, motivation and examination. *Studies in Higher Education, 27(3)*, 329-338.
- Maclellan, E. (2001). Assessment for learning: The differing perceptions of tutors and students. *Assessment and Evaluation in Higher Education, 26(4)*, 307-318.
- Meyers, C., & Jones, T. B. (1993). *Promoting active learning. Strategies for the college classroom*. San Francisco: Jossey-Bass.
- Oxford, R. L. (1997). Constructivism: Shaping, shifting, substance and teacher education applications. *Peabody Journal of Education, 72(1)*, 35-66.
- Sambell, K., McDowell, L., & Brown, S. (1997). 'But is it fair?': An exploratory study of student perceptions of the consequential validity of assessment. *Studies in Educational Evaluation, 23(4)*, 349-371.
- Sander, P., Stevenson, K., King, M., & Coates, D. (2000). University students' expectations of teaching. *Studies in Higher Education, 25(3)*, 309-323.
- Segers, M. (2003). Evaluating the OverAll Test: Looking for multiple validity measures. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 119-140). Dordrecht: Kluwer Academic Publishers.
- Segers, M., Dochy, F., & Cascallar, E. (2003). The era of assessment engineering: Changing perspectives on teaching and learning and the role of new modes of assessment. In M. Segers, F. Dochy, & E. Cascallar (Eds.), *Optimising new modes of assessment: In search of qualities and standards* (pp. 1-12). Dordrecht: Kluwer Academic Publishers.
- Silberman, M. (1996). *Active learning: 101 strategies to teach any subject*. Needham Heights, MA: Allyn & Bacon
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment and Evaluation in Higher Education, 30(4)*, 331-347.
- Struyven, K., Dochy, F., Janssens, S., Schelfhout, W., & Gielen, S. (in press). The effects of the method of assessment on student performance: A comparison between multiple choice testing, peer assessment, case based assessment and portfolio assessment. *Studies in Educational Evaluation*.
- Struyven, K. & Gielen, S. (2004a). *Groot worden. De ontwikkeling van baby tot adolescent (Werkboek voor toekomstige leerkrachten Basisonderwijs, Uitbreidingsdeel)*. Leuven: LannooCampus.
- Struyven, K. & Gielen, S. (2004b). *Groot worden. De ontwikkeling van baby tot adolescent (Werkboek voor toekomstige leerkrachten Secundair Onderwijs, Uitbreidingsdeel)*. Leuven: LannooCampus.
- Struyven, K., Sierens, E., Dochy, F., & Janssens, S. (2003). *Groot worden. De ontwikkeling van baby tot adolescent (Handboek voor toekomstige leerkrachten)*. Leuven: LannooCampus.
- Tenenbaum, G., Naidu, S., Jegede, O., & Austin, J. (2001). Constructivist pedagogy in conventional on-campus and distance learning practice: An exploratory investigation. *Learning and Instruction, 11(2)*, 87-111.
- Terwel, J. (1999). Constructivism and its implications for curriculum theory and practice. *Journal of Curriculum Studies, 31(2)*, 195-199.
- Traub, R. E., & MacRury, K. (1990). Multiple choice vs. free response in the testing of scholastic achievement. In K. Ingenkamp & R. S. Jager (Eds.), *Test und Trends 8: Jahrbuch der pädagogischen Diagnostik* (pp. 128-159). Weinheim und Base: Beltz Verlag.
- Tynjälä, P. (1997). Developing education students' conceptions of the learning process in different learning environments. *Learning and Instruction, 7(3)*, 277-292.
- Van den Berg, D., & Vandenberghe, R. (1999). *Succesvol leiding geven aan onderwijsinnovaties. Investeren in mensen*. Alphen aan den Rijn: Samsom.
- Vermunt, J. D. (1998). The regulation of constructive learning processes. *British Journal of*

Educational Psychology, 68(2), 149-171.

- Von Glasersfeld, E. (1988). Constructivism as a scientific method. *Scientific Reasoning Research Institute Newsletter*, 3(2), 8-9.
- White, C. (1996). Merging technology and constructivism in teacher education. *Teacher Education and Practice*, 12(1), 62-70.
- Zeidner, M. (1987). Essay versus multiple-choice type classroom exams: The student's perspective. *Journal of Educational Research*, 80(6), 352-358.
- Zoller, U., & Ben-Chaim, D. (1988). Interaction between examination-type anxiety state and academic achievement in college science: An action-oriented research. *Journal of Research in Science Teaching*, 26(1), 65-77.
- Zoller, U., & Ben-Chaim, D. (1997). Examination type preferences of college science students and their faculty in Israel and USA: A comparative study. *School Science and Mathematics*, 97(1), 3-12.

Manuscript aanvaard: 9 november 2005

Auteurs

Katrien Struyven, Filip Dochy, Steven Janssens, Wouter Schelfhout en Sarah Gielen zijn werkzaam bij het Centrum voor Opleidingsdidactiek van de katholieke Universiteit Leuven.

Correspondentieadres: K. Struyven, KU Leuven, Centrum voor Opleidingsdidactiek, Dekenstraat 2, 3000 Leuven, België, katrien.struyven@ped.kuleuven.be

Abstract

Effects of hands-on experience with new modes of assessment on students' assessment preferences

This study investigates the effects of hands-on experience on student teachers' preferences of assessment methods. A course on Child Development within the first-year of the elementary teacher education program provides the quasi-experimental learning/teaching setting. Five research conditions ($N = 612$) are linked to four assessment modes, namely: portfolio, case-based, peer assessment, and multiple choice evaluation. Data collection was obtained by a pretest-posttest design with the help of three questionnaires. Results show negative responses to unknown assessment methods. Feelings of uncertainty make students feel uncomfortable with and critical towards the expected assessment method. Though the extent differs, every assessment method significantly benefited from the hands-on experience with the format. As familiarity with the assessment tool grew, students' preferences changed positively. Moreover, students' perceptions of the appropriateness of the assessment method are congruent with students' preferences. Consequently, to change students' preferences for unknown assessments, hands-on experiences need to be positive.