Stagebeoordelingen

Hoe kunnen deze beter verlopen?

Intern rapport

2018

Auteurs:

- Duriez, Bart (projectleider)
- Coppens, Serge
- Heynderickx, Thomas
- Verheyen, Helena

Karel de Grote Hogeschool

INHOUD

1 Samenvatting 1
2 Inleiding 4
2.1 Het belang van stages 4
2.2 Het belang van stagebeoordelingen 4
2.2.1 Courante problemen met stagebeoordelingen 4
2.2.2 Richtlijnen voor betere stagebeoordelingen 5
2.2.2.1 Betrouwbaarheid en validiteit 5
2.2.2.2 Omvattendheid en flexibiliteit 7
2.2.2.3 Haalbaarheid en tijdigheid 8
2.2.2.4 Relevantie en verantwoordbaarheid 8
2.2.2.5 Samenvattend 8
2.2.3 Naar kwaliteitscontrole 9
3 Voorstudie 1 13
3.1 Methode 13
3.1.1 Procedure 13
3.1.2 Steekproef 14
3.1.3 Effectmetingen 15
3.1.3.1 Kwaliteitscriteria 15
3.2 Resultaten 16
3.2.1 Kwaliteitscriteria 16
3.3 Conclusie 17
4 Voorstudie 2 21
4.1 Methode 21
4.1.1 Procedure 21
4.1.2 Steekproef 21
4.1.3 Effectmetingen 22
4.1.3.1 Kwaliteitscriteria 22
4.2 Resultaten 23
4.2.1 Kwaliteitscriteria 23
4.3 Conclusie 23
5 Studie 1 27
5.1 Methode 27
5.1.1 Procedure 27

5.1.2 Steekproef 30
5.1.3 Effectmetingen 32
5.1.3.1 Kwaliteitscriteria voor de specifieke interventie 32
5.1.3.2 Kwaliteitscriteria voor de algemene beoordeling 33
5.1.3.3 Uitdieping van het Leereffectcriterium 33
5.1.3.4 Uitdieping van het Selectiecriterium 35
5.2 Verwachtingen 35
5.3 Resultaten 36
5.3.1 Kwaliteitscriteria voor de specifieke interventie 36
5.3.2 Kwaliteitscriteria voor de algemene beoordeling 36
5.3.3 Uitdieping van het Leereffectcriterium 37
5.3.4 Uitdieping van het Selectiecriterium 38
5.4 Conclusie 38
6 Studie 2 54
6.1 Methode 54
6.1.1 Procedure 54
6.1.2 Steekproef 57
6.1.3 Effectmetingen 58
6.1.3.1 Kwaliteitscriteria voor de specifieke interventie 58
6.1.3.2 Kwaliteitscriteria voor de algemene beoordeling 59
6.1.3.3 Uitdieping van het Leereffectcriterium 60
6.1.3.4 Uitdieping van het Selectiecriterium 61
6.2 Verwachtingen 61
6.3 Resultaten 62
6.3.1 Kwaliteitscriteria voor de specifieke interventie 62
6.3.2 Kwaliteitscriteria voor de algemene beoordeling 62
6.3.3 Uitdieping van het Leereffectcriterium 63
6.3.4 Uitdieping van het Selectiecriterium 63
6.4 Conclusie 64
7 Studie 3 79
7.1 Methode 80
7.1.1 Procedure 80
7.1.2 Steekproef 82
7.1.3 Effectmetingen 82

7.1.3.1 Kwaliteitscriteria voor het beoordelingsformulier 82
7.1.3.2 Kwaliteitscriteria voor de leergroepen 83
7.1.3.3 Uitdieping van het Leereffectcriterium 84
7.1.3.4 Uitdieping van het Selectiecriterium 86
7.2 Verwachtingen 87
7.3 Resultaten 87
7.3.1 Kwaliteitscriteria voor het beoordelingsformulier 87
7.3.2 Kwaliteitscriteria voor de leergroepen 88
7.3.3 Uitdieping van het Leereffectcriterium 88
7.3.4 Uitdieping van het Selectiecriterium 89
7.4 Conclusie 89
8 Discussie 104
8.1 Samenvatting van de resultaten 105
8.1.1 De perceptie van de interventies 105
8.1.2 Het globaal effect van de interventies 106
8.1.3 Het leereffect van de interventies 107
8.1.4 Het selectie-effect van de interventies 108
8.2 Bespreking van de resultaten 109
8.3 Aanbevelingen 110
8.4 Beperkingen 111
9 Referenties 113
10 Appendices 117
10.1 Kwaliteitscriteria zoals gebruikt in Voorstudie 1 117
10.2 Kwaliteitscriteria zoals gebruikt in Voorstudie 2 120
10.3 Studie 1: Kwaliteitscriteria voor de specifieke interventie 123
10.4 Kwaliteitscriteria voor de algemene beoordeling 125
10.5 Studie 2: Kwaliteitscriteria voor de specifieke interventie 127
10.6 Studie 3 (Studenten): Kwaliteitscriteria voor de specifieke interventies 129
10.7 Studie 3 (Begeleiders): Kwaliteitscriteria voor de specifieke interventies 133
10.8 De competentiegroeimeter voor stage voor docenten 137
10.9 De competentiegroeimeter voor stage voor studenten 138

DANKWOORD

Als teamcoördinator wens ik in het kader van dit onderzoek alvast een heel aantal mensen van harte te bedanken. Vooreerst de leden van mijn team (Serge Coppens, Thomas Heynderickx en Helena Verheyen) voor de geleverde inspanningen doorheen de volledige duur van het project, om mee te denken over zowel inhoudelijke als praktische zaken, en voor de rust en relativeringsvermogen tijdens stressvolle momenten. Verder wens ik Geert Speltincx te bedanken voor zijn inhoudelijke en logistieke ondersteuning en om altijd paraat te staan als sparringpartner en doorhakker van knopen. Ook de andere leden van de stuurgroep, Ann Cleerbout, Jill Schoonvliet, Sophie Toelen, en Tine van Daal, waren heel belangrijk voor het goede verloop van ons project. Verder gaat mijn dank ook uit naar de mensen van het D-PAC team van Universiteit Antwerpen, en naar de mensen binnen de opleidingen PBKO & PBLO (en in het bijzonder Peggy Roumans), Verpleegkunde, en Accountancy en Fiscaliteit die ervoor hebben gezorgd dat we ons project tot een goed einde konden brengen. Bedankt voor jullie hulp en steun!

1 SAMENVATTING

Dit onderzoek richt zich op de verbetering van stagebeoordelingen. Vanuit de literatuur kan men mogelijke interventies indelen in drie categorieën: (1) gebruik meer beoordelingsmethoden, (2) gebruik meer ratings en observaties, en (3) organiseer trainingen voor docenten en mentoren. Tegenwoordig wordt er in de literatuur veel nadruk gelegd op het procesmatige karakter van stages. De meeste interventies willen daarom niet enkel een effect hebben op de kwaliteit van de beoordeling maar ook op de kwaliteit van het hele stageproces. Voor het onderzoek werden vier opleidingen geselecteerd (de Professionele Bachelors Kleuter- en Lager Onderwijs, Verpleegkunde, en Accountancy & Fiscaliteit). Voor elk van deze opleidingen werd de stagebeoordeling van het laatste jaar van de opleiding doorgelicht op zoek naar verbeteropties, telkens in samenspraak met de opleiding in kwestie. Voor elke opleiding werd vervolgens een onderzoeksdesign uitgewerkt voor het academiejaar 2017-2018. Een belangrijke aanbeveling uit de literatuur werd daarbij voor ogen te houden: een goede stagebeoordeling dient niet enkel een summatief maar ook een formatief aspect te hebben. Om de interventies te evalueren, kozen we voor het Competentie Assessment Wiel van Baartman. Hierin worden 15 kwaliteitscriteria naar voor geschoven. De vragenlijst die we opstelden om deze criteria te meten werd een eerste maal getest in 2016-2017 tijdens een onderzoek van UAntwerpen in het Flextraject van de Professionele Bachelors Kleuteronderwijs. In Voorstudie 2 werd onze vragenlijst op punt gesteld tijdens de stage van de tweedejaars Verpleegkunde.

In Studie 1 en 2, werd het aantal assessmentsituaties en beoordelaars opgedreven in een experimenteel design. In Studie 1 (in de Professionele Bachelors Kleuter- en Lager Onderwijs) werd voor de helft van de studenten meer begeleiders ingeschakeld die ook vaker op lesbezoek gingen. In Studie 2 (in Verpleegkunde) werd voor de helft van de studenten de medestudenten ingeschakeld. De effecten werden gemeten aan de hand van bovenvermelde kwaliteitscriteria. Die werden zowel gemeten op het niveau van specifieke interventie (in de experimentele groep) als op het niveau van de gehele beoordeling (in de controlegroep en de experimentele groep). Tot slot werden de controlegroep en de experimentele groep ook vergeleken qua Leereffect (aan de hand van competentiegroei) en Selectie (de mate waarin sterke van zwakke studenten kunnen worden onderscheiden). Voor Studie 1 leerde de bevraging in de experimentele groep ons dat de interventie goed onthaald werd. Wanneer er gepolst werd naar de bijdrage van de interventie aan de gehele beoordeling, waren zowel begeleiders als studenten het er gemiddeld genomen over eens dat de interventie een positieve bijdrage leverde aan de stagebeoordeling. Desalniettemin kwamen er uit de groepsvergelijkingen weinig effecten van de interventie naar voor: de experimentele groep verschilde nagenoeg niet van de controlegroep in termen van de scores op de kwaliteitscriteria. Ook voor Leereffect en Selectie vonden we geen effecten. Los van het feit dat er in deze studie niet expliciet werd ingezet op (andere) competenties, bleek dat in de controleconditie alle competenties gemiddeld genomen een zeer hoge groeiscore vertoonden. De verbetermogelijkheden op dit vlak waren dan ook beperkt.

Ook voor Studie 2 leerde de bevraging in de experimentele groep ons dat de interventie goed onthaald werd. Wanneer er gepolst werd naar de bijdrage van de interventie aan de gehele beoordeling, zagen we dat begeleiders het er gemiddeld genomen voor bijna alle criteria over eens waren dat de interventie een positieve bijdrage leverde aan de stagebeoordeling. Voor studenten was dit niet zo. Studenten vonden wel dat de interventie het zicht van begeleiders

op hun cognitieve vaardigheden groter maakte, waardoor ze de beoordeling ook eerlijker en meer acceptabel vonden, maar voor de rest stonden ze eerder neutraal tegenover de interventie. Het waren dus vooral de begeleiders die hier positief tegenover stonden. Uit de groepsvergelijking kwamen opnieuw weinig effecten van de interventie naar voor. De experimentele groep scoorde niet hoger dan de controlegroep op de kwaliteitscriteria. Eerder integendeel: waar er verschillen waren, deed de controlegroep het beter. De studenten in de experimentele conditie waren minder positief over de interventie. Studenten lijken dus geen voorstanders van peer assessment. Mogelijks was het minder positieve oordeel van studenten in de experimentele conditie een gevolg van het feit dat het peer assessment niet meetelde, waardoor studenten het misschien veel gedoe voor niets vonden. Voor Leereffect vonden we geen effecten. Los van het feit dat er in deze studie niet expliciet werd ingezet op (andere) competenties, bleek dat in de controleconditie alle competenties gemiddeld genomen een zeer hoge groeiscore vertoonden. De verbetermogelijkheden op dit vlak waren dus ook hier beperkt. Ook voor Selectie vonden we geen effecten. Zelfs al hadden de scores op het peer assessment meegeteld, dan had dit niets wezenlijks veranderd. Het peer assessment bracht niets in rekening wat niet sowieso al in rekening werd gebracht bij het bepalen van de eindscore. Dit kan een gevolg zijn van de specifieke situatie. Het peer assessment werd hier ingevoerd in het kader van de leerwerkplek. In dit kader is het functioneren van (de studenten in) het team essentieel. De interactie tussen studenten en hun oordeel over elkaar werd daarom ook nog langs andere wegen bevraagd.

Wanneer de resultaten van Studie 1 en Studie 2 naast elkaar gelegd worden, zou geconcludeerd kunnen worden dat beoordelaars toevoegen niet altijd unaniem positief onthaald wordt. Hoe dit onthaald wordt, hangt af van wie als beoordelaars worden ingezet en van hoe die worden ingeschakeld. In Studie 1 werden extra begeleiders ingeschakeld die ook vaker op stagebezoek gingen, meer feedback gaven, en onderling overlegden. Ook al had deze interventie geen direct aantoonbare impact op de beoordeling van de stagebeoordeling (in de zin dat de experimentele groep niet positiever stond ten aanzien van de stagebeoordeling dan de controlegroep), en ook al bracht ze geen extra competentiegroei teweeg, toch werd de interventie zowel door begeleiders als door studenten positief onthaald. In Studie 2 daarentegen werden medestudenten ingeschakeld als extra beoordelaars. Deze interventie werd minder positief onthaald. Begeleiders zagen hier wel een meerwaarde in, maar onder studenten was dit minder het geval. Een voor de hand liggende verklaring is dat studenten onvoldoende vertrouwen hebben in het oordeel van hun medestudenten. Dit is echter geen reden om peer assessment te mijden. Veel hangt er immers van af hoe je dit vormgeeft. In de literatuur worden een aantal factoren geïdentificeerd die mee bepalen hoe tevreden studenten zullen zijn met peer assessment. Zo speelt niet enkel het type peer assessment een rol, maar ook (de kwaliteit van) de training van studenten die hieraan voorafgaat, want die is belangrijk om ervoor te zorgen dat de feedback kwaliteitsvol is en niet te hard beïnvloed wordt door bijvoorbeeld bestaande vriendschapsbanden.

In Studie 3 (in Accountancy & Fiscaliteit) gooiden we het over een andere boeg. Via leergroepen werkten studenten aan soft skills. Een nieuw beoordelingsformulier gaf aan deze soft skills een expliciete plaats in het eindoordeel. Omwille van het geringe studentenaantal was hier geen controlegroep en werden de kwaliteitscriteria enkel gemeten op niveau van de specifieke interventies. Ook hier werd ingezoomd op Leereffect en Selectie. Voor beide werd een vergelijking gemaakt met vorig academiejaar. Beide interventies werden over het algemeen positief ont-

haald en de betrokkenen zagen over het algemeen een meerwaarde in zowel het nieuwe beoordelingsformulier als de leergroepen. Voor Selectie vonden we geen effecten. Er is dus geen aanwijzing voor het feit dat de interventies een beter onderscheid zouden helpen maken tussen sterke en zwakke studenten. Voor Leereffect, zagen we dan weer duidelijk de meerwaarde van de interventies. In 2017-2018 verwachtten docenten voor zo goed als alle competenties meer groei dan in 2016-2017. Studenten rapporteerden in 2017-2018 dan weer voor alle competenties significant meer groei dan in 2016-2017. Het zijn de leergroepen die deze meerwaarde creëren. Hieruit kan geconcludeerd worden dat de leergroepen een succes waren: de leergroepen zorgden ervoor dat docenten meer overtuigd waren van de door de stage te verwachten leereffecten en studenten ervaarden ook daadwerkelijk meer competentiegroei.

Wil men de stagebeoordeling verbeteren, dan doet men er, zoals ook uit de literatuur naar voor komt, goed aan om de stagebeoordeling niet los te zien van de stage als leerproces. Studie 1 en 2 suggereren dat, wanneer je enkel inzet op de beoordeling zelf – zelfs al heb je hierbij oog voor de formatieve aspecten van de beoordelingswijze – je in het beste geval wel een breder draagvlak creëert voor de beoordeling, maar dat je op die manier het leerproces van studenten geen extra boost geeft. Het lijkt dus niet voldoende om enkel in te spelen op de in de literatuur aangereikte verbeteringsmogelijkheden van de beoordeling zelf. Het lijkt, zoals eveneens in de literatuur gesteld wordt, ook nodig om de stage zelf grondig onder de loep te nemen om dit te bewerkstelligen. Op welke competenties werkt de stage in? Zijn dit wel alle competenties die je in studenten wil zien ontwikkelen? Als dat niet het geval is, dan is het nodig om de stage te hervormen zodat ook die competenties expliciet aan bod komen. Niet enkel in de stage maar ook in de uiteindelijke beoordeling. Dit was het opzet van Studie 3. In Studie 3 werden onderontwikkelde competenties geïdentificeerd en werd nagedacht hoe die versterkt konden worden, en dit zowel vanuit de instelling (= door de invoering van leergroepen) als vanop de stageplek zelf (= door mentoren te wijzen op het belang van goede feedback). Deze competenties werden in overeenstemming met het principe van constructive alignment ook expliciet betrokken in het eindoordeel (= het nieuwe beoordelingsformulier). Zo winnen ze ook in het hoofd van studenten aan belang. Een dergelijke brede interventie, die niet alleen inspeelt op elk van de in de literatuur aangereikte verbeteringsmogelijkheden, maar ook op de pijnpunten van de stage zelf, lijkt meest effectief om te bekomen wat men beoogt: het opleiden van studenten tot competente professionals die in staat zijn levenslang zelfstandig te leren.

Voor opleidingen die zich afvragen of hun stage(beoordeling) kan of dient te worden verbeterd, lijkt het een goed idee om bij alle betrokkenen (i.e., studenten, docenten en mentoren) een voorafname te doen van zowel de kwaliteitscriteria als de competenties waar een stage op inzet. Op basis daarvan kan men dan zien of er pijnpunten of lacunes zitten in de stage(beoordeling). Zijn er kwaliteitscriteria die zwak scoren? Zijn er competenties die onvoldoende aan bod komen? Eens dergelijke lacunes geïdentificeerd zijn, kan men nadenken over remediërende interventies. De beschikbare literatuur biedt handvaten en inspiratie ter verbetering van de stage(beoordeling). Uiteraard is het niet per definitie zo dat er problematische aspecten zijn aan een stage(beoordeling). Een voorafname kan ook uitwijzen dat men "goed bezig" is. Iets wat zich ook toont in de huidige studies. Zo zagen we in Studie 1 en Studie 2 dat alle competenties ook reeds in de controleconditie erg hoge groeiscores vertoonden.

2 INLEIDING

In verschillende opleidingen leeft de vraag of de stagebeoordelingen niet nog beter kunnen. In 2016-2017 startte KdG een onderzoek naar mogelijke pistes ter verbetering van de stagebeoordelingen. Hoe zorg je dat de stagebegeleider op de werkplek (= de mentor) werkt vanuit de opleidingsdoelen? Hoe bewaak je als opleiding de beoordeling van stagiairs? Hoe leg je de lat juist (niet te laag en niet te hoog)? Hoe zorg je dat je de student en niet (het begeleidende werk van) de mentor beoordeelt? Allemaal vragen die in dit project aan bod komen. Om ze te beantwoorden, werden een aantal opleidingen geselecteerd. Voor elk van deze opleidingen werd een onderwijskundige interventie uitgewerkt. Vooraleer we overgaan tot een beschrijving van het onderzoek en de resultaten ervan, gaan we in dit hoofdstuk op zoek naar verschillende manieren waarop stagebeoordelingen kunnen worden verbeterd, op de uiteindelijke selectie van interventies, en op de onderzoeksmethode die we zullen hanteren om het effect van deze interventies te testen.

2.1 Het belang van stages

Stages bieden studenten de kans om de kloof te overbruggen tussen de verwachtingen in de klas en de realiteit van een werkomgeving (Verney et al., 2009). Stages zorgen ervoor dat studenten de ervaring opdoen die ze later nodig hebben wanneer ze een job beginnen zoeken. Daarom is het ook belangrijk dat het werk dat studenten op hun stage voorgeschoteld krijgen relevant is en afgestemd op hun niveau en het door het werkveld verwachtte niveau (Pijl-Zieber, 2014). Ook voor werkgevers zijn stages voordelig. Ze voorzien bedrijven of instellingen van goedkope werkkrachten en vormen een interessant rekruteringskanaal.

2.2 Het belang van stagebeoordelingen

Het belang van een goede stagebeoordeling is tweeërlei (Turnbull et al., 2000). Voor de stagiair biedt die, zeker in combinatie met goede tussentijdse feedback, de mogelijkheid om diens vaardigheden en kennis te verbeteren. Voor de samenleving biedt die de garantie dat er enkel bekwame werkkrachten op de arbeidsmarkt komen. Voor de school biedt de stage ook een test van het curriculum. Doen studenten het goed aan het einde van de rit?

2.2.1 Courante problemen met stagebeoordelingen

Eind vorige eeuw lijsten Turnbull et al. (1998) een aantal courante problemen met het beoordelen van stages op. Een eerste probleem is dat er te weinig competenties beoordeeld worden. Vaak wordt bijvoorbeeld geen rekening gehouden met competenties die wel relevant zijn voor een bepaald werk maar niet eigen zijn aan dat werk. Denk aan meer algemene vaardigheden zoals sociale vaardigheden, communicatievaardigheden, en reflectievaardigheden. Een tweede probleem is dat er vaak te weinig observaties gedaan worden en dat er te weinig relevant gedrag gerapporteerd wordt. Zo krijgen leerlingen in een lerarenopleiding bijvoorbeeld typisch maar een beperkt aantal lesbezoeken. Tijdgebrek is hier de belangrijkste oorzaak van (Hunt, 1992). Een derde probleem is dat stagebeoordelingen vaak weinig betrouwbaar zijn, in de zin dat beoordelingen aanzienlijk kunnen fluctueren tussen beoordelaars en binnen beoordelaars (= het probleem van, respectievelijk, de inter- en intra-beoordelaarsbetrouwbaarheid; zie ook Atkinson et al, 1987; Downing, 2004).

De onderliggende reden voor betrouwbaarheids- en validiteitsproblemen ligt voor de hand: beoordelaars worden zelden voorbereid op hun dubbele rol, enerzijds als mentor, anderzijds als beoordelaar. De persoon die een punt moet toekennen, moet immers vaak iemand beoordelen waar hij ondertussen een persoonlijke band mee heeft opgebouwd. Dit maakt het moeilijk om 'streng' te zijn. Hierdoor liggen de punten voor stages in vele gevallen systematisch erg hoog, is er weinig variatie in de toegekende punten, en wordt het moeilijk om sterke van minder sterke studenten te onderscheiden. Vertekende oordelen kunnen verschillende oorzaken hebben. Wanneer de punten beïnvloed worden door de meest recente ervaringen of door de gemoedstoestand van de beoordelaar, spreekt men van een "recency effect". Wanneer het zo is dat studenten die als "goede" studenten gekend zijn op basis van voorgaande prestaties systematisch hogere punten krijgen, spreekt men van een "halo effect" of "carryover effect". Wanneer niet de hele schaal gebruikt wordt, spreekt men van "distribution errors". Dit is vaak het geval wanneer beoordelaars niet streng genoeg zijn, wat een vaak voorkomend probleem is (zie ook Carraccio et al., 2002; Govaerts et al., 2007).

2.2.2 Richtlijnen voor betere stagebeoordelingen

Om bovenvermelde problemen op te vangen, zijn er in de vakliteratuur tal van aanbevelingen te vinden waaraan stagebeoordelingen best voldoen of waar ze op zijn minst tegen dienen te worden afgetoetst. Deze aanbevelingen kunnen in feite onderverdeeld worden in vier categorieën (gebaseerd op Turnbull et al., 1998; 2000): betrouwbaarheid en validiteit, omvattendheid en flexibiliteit, haalbaarheid en tijdigheid, en relevantie en verantwoordbaarheid. We bespreken ze hieronder elk van deze vier categorieën.

2.2.2.1 Betrouwbaarheid en validiteit

Een beoordeling wordt klassiek als betrouwbaar beschouwd als het zo is dat, wanneer dezelfde beoordelaar de student meermaals zou beoordelen, de uitkomst telkens redelijk vergelijkbaar zou zijn (= intra-beoordelaarsbetrouwbaarheid) en/of wanneer eenzelfde student door meerdere personen beoordeeld zou worden, de student in kwestie telkens nagenoeg dezelfde beoordeling zou krijgen (= inter-beoordelaarsbetrouwbaarheid). Validiteit slaat dan weer op de mate waarin het resulterend oordeel daadwerkelijk verwijst naar wat het beweert te verwijzen (i.e., het functioneren tijdens de stage).

In het onderzoek naar betrouwbaarheid en validiteit werd druk gezocht naar het ideale meetinstrument om iemands functioneren tijdens een stage te beoordelen. De nadruk lag daarbij op de psychometrische eigenschappen van de gebruikte meetinstrumenten. De aanbevelingen die men in dit kader in de literatuur vindt, komen grosso modo hierop neer: kijk naar waarneembaar gedrag en niet naar (veronderstelde) onderliggende competenties, rapporteer betekenisvolle verschillen (zowel tussen als binnen individuen) en rapporteer deze op gestandaardiseerde wijze (Anema & McCoy, 2010; Turnbull et al., 1998; 2000). Doorgaans zocht men zijn heil in checklists en rating scales, al dan niet met ruimte voor het nodige extra commentaar (zie o.a., Gray, 1996; Turnbull et al., 2000; Young et al., 2011; Prescott-Clements et al., 2008; Tigelaar & van der Vleuten, 2014). De uitdaging bestond erin om te komen tot een duidelijke definitie van de te meten competenties en een vertaling daarvan naar bijhorend gedrag en gepaste antwoordcategorieën met voldoende variatie (Atkinson & Murray, 1987; Verney et al., 2009).

Vervolgens kwam het erop aan beoordelaars te trainen in het gebruik van deze antwoordcategorieën (Atkinson & Murray, 1987; Gray, 1996). Een evident gevaar van deze manier van werken, bestaat erin dat competenties op dusdanige manier opgebroken worden in gedragingen, dat dit uitmondt in een eindeloos lange lijst te beoordelen gedragingen die verre van handig en overzichtelijk zijn (Frank et al., 2010).

Gaandeweg zijn diverse auteurs meer en meer vraagtekens gaan plaatsen bij deze in hun ogen overdreven bekommernis om de psychometrische eigenschappen van de meetinstrumenten (e.g., Govaerts et al., 2007; Govaerts & van der Vleuten, 2013, Tigelaar & van der Vleuten, 2014; van der Vleuten & Schuwirth, 2005; van Lohuizen et al., 2010). De zoektocht naar het ideale meetinstrument vertrekt namelijk van de assumptie dat het functioneren van een stagiair stabiel is, dat er een "ware" score bestaat die dit functioneren kan weerspiegelen, en dat deze "ware" score bekomen kan worden wanneer men de meetfouten tot nul kan herleiden. Volgens deze auteurs klopt deze assumptie echter niet. Volgens hen is het beoordelen van een stagiair uitermate complex en daardoor ook inherent subjectief. Verschillende oordelen wijzen volgens hen niet per se op meetfouten. Verschillende uitkomsten en oordelen kunnen immers minstens drie andere oorzaken hebben: (1) een specifieke methode kan de aandacht op bepaalde zaken vestigen terwijl een andere methode eerder andere dingen belicht (= elke methode zal bepaalde specifieke inzichten verschaffen, eigen aan die methode), (2) verschillende beoordelaars kunnen tot een ander oordeel komen omdat ze vanuit hun eigen achtergrond en ervaring nu eenmaal op andere dingen letten, en (3) oordelen over een stagiair die op een ander moment gemaakt worden kunnen verschillen omdat de stagiair in kwestie gewoon niet op elk moment even goed presteert (diens prestatie kan niet alleen van dag tot dag fluctueren maar ook mede afhankelijk zijn van de kansen en beperkingen eigen aan de context waarin de stagiair opereert). Verschillende oordelen kunnen dus net heel betekenisvol zijn. Het komt er volgens deze auteurs dan ook niet op aan te zoeken naar de ideale methode om het functioneren van een student op een stageplek te meten. Het komt er eerder op aan verschillende methodes te combineren en zoveel mogelijk observaties bij het beoordelingsproces te betrekken, liefst verzameld door meerdere beoordelaars. Ze doelen daarmee niet alleen meer beoordelaars maar ook meer verschillende soorten beoordelaars (i.e., mensen met verschillende rollen). McNamara (2013) raadt in dit kader aan te werken met een mix van informatie uit verscheiden bronnen, met name informatie bekomen via minstens 1 docent of begeleider binnen de onderwijsinstelling, minstens 1 mentor of begeleider uit het werkveld, en de student zelf. Meerdere beoordelaars dienen daarbij niet in eerste instantie om de "objectiviteit" van een oordeel te verhogen maar wel om tot een "rijker" oordeel te komen. Een "rijker" oordeel is in hun ogen meer betrouwbaar (niet in strikt psychometrische zin maar in de zin dat het meer vertrouwen wekt). Bovendien heeft een "rijker" oordeel ook een grotere validiteit. Omdat het niet eenvoudig is om dat "rijker" oordeel te vertalen naar een eindquotering, raden deze auteurs het gebruik van een expertpanel aan die al deze de informatie tracht samen te brengen, in woorden, eerder dan in cijfers.

Samengevat kan gesteld worden dat de hedendaagse visie meer dynamisch is. Deze visie spitst zich namelijk niet zozeer toe op het oordeel zelf, maar op het stage*proces*. Daarbij is het niet enkel zo dat de beoordeling een weerspiegeling moet bieden van het proces, maar ook dat ze moet bijdragen aan dat proces: idealiter is het een instrument dat het leren bevordert (Dochy et al. 2007; Higgs, 2014; Hodges, Eames, & Coll, 2014; McNamara, 2013). Tussentijdse feedback is in deze visie essentieel (Hattie & Timperly, 2007). Niet enkel omdat het de motor is die

ervoor zorgt dat studenten de kans krijgen om hun competenties te ontwikkelen (zie 2.2.2.2), maar ook omdat die voorbereidt op het eindoordeel (zie 2.2.2.4). Meerdere auteurs pleiten er vanuit die optiek voor om trainingen voor beoordelaars en mentoren te organiseren (Govaerts et al., 2007; 2013; McNamara, 2013). Daarbij dient men niet enkel te focussen op het leren gebruiken van meetinstrumenten, maar ook op het begeleidingsproces en het geven van constructieve feedback. Als men bovendien wil bekomen dat iedereen die betrokken is bij het beoordelingsproces de stagiairs op vergelijkbare wijze beoordeelt, dan is het volgens hen tevens belangrijk dat alle beoordelaars en begeleiders dezelfde verwachtingen hebben van stagiairs, dezelfde of gelijkwaardige kansen bieden, en dezelfde standaarden en criteria hanteren. Staan niet in alle neuzen in dezelfde richting, dan dreigt de beoordeling natte-vinger-werk te worden of een weerspiegeling van de capaciteiten van de mentor eerder dan van de student. Een training voor beoordelaars en begeleiders gaat dus best ook in op verschillen in voor het beoordelings- en begeleidingsproces relevante bedenkingen, op de impliciete theorieën en verwachtingen van beoordelaars en begeleiders, en op omgevingsfactoren die de motivatie van beoordelaars en begeleiders beïnvloeden.

2.2.2.2 Omvattendheid en flexibiliteit

Van oudsher vindt men in de literatuur rond stagebeoordelingen de bekommernis om ervoor te zorgen dat alle relevante competenties aan bod komen in de beoordeling (Turnbull et al., 1998; 2000). Gaandeweg is daarbij ook meer aandacht gevraagd voor "algemene" competenties zoals sociale vaardigheden, communicatievaardigheden en reflectievaardigheden (cfr. Anema & McCoy, 2010; Biggs, 2003; Epstein & Hundert, 2002; Govaerts et al., 2007; Norcini et al., 2011; van der Vleuten & Schuwirth, 2005). Hier hoort ook de competentie om zelfstandig nieuwe kennis te verwerven bij (Fastré et al., 2012; Frank et al., 2010; Epstein & Hundert, 2002; Govaerts & van der Vleuten, 2013; Higgs, 2014). Hierbij dient men er uiteraard rekening mee te houden dat niet elke stagecontext dezelfde is. Elke stage is specifiek en niet alle competenties zullen in elke context even relevant zijn. Met een dergelijke specificiteit moet rekening gehouden worden in de beoordelingen (zie ook Carraccio et al., 2002).

Naarmate de aandacht in de literatuur verschoof van het louter psychometrische naar het procesmatige karakter van stages, werd de roep ook luider om ervoor te zorgen dat dit procesmatige karakter zich ook zou weerspiegelen in de beoordelingen. Van daaruit wordt er ook op gehamerd dat het noodzakelijk is om te specifiëren welke competenties je wil zien ontwikkelen tijdens een bepaalde stage (McNamara, 2013). Een explicitering hiervan zou ook de motivatie van studenten ten goede komen (Turnbull et al., 1998, Verney et al., 2009), zeker als die explicitering gebeurt in samenspraak met de student en de mentor (Zegwaard, Coll & Hodges, 2003). Vervolgens dient men ervoor te zorgen dat dergelijke evoluties meetbaar zijn (van Lohuizen et al., 2010). Recent is er dan ook meer aandacht voor methoden die oog hebben voor dergelijke evoluties (voor een voorbeeld, zie Prescott-Clements et al., 2008). En ook in dit kader benadrukken verschillende auteurs het belang van tussentijdse feedback. Feedback is immers de motor bij uitstek die ervoor zorgt dat studenten de kans krijgen om hun competenties te ontwikkelen (Aronson, 2011; Dannefer, 2013; Embo et al., 2015; Govaerts et al., 2007; Norcini et al., 2011; Norcini & Burch, 2007; Tigelaar & van der Vleuten, 2014; van der Vleuten et al., 2010; van Lohuizen et al., 2010). Feedback kan studenten ook helpen een onderscheid te maken tussen het werk op zich en wat ze daaruit geacht worden te leren (Walsh 2007).

Het ultieme doel is ervoor te zorgen dat studenten uitgroeien tot professionals die in staat zijn tot levenslang zelfstandig leren (Boud & Falchikov, 2006; Fastré et al., 2013). Daartoe moeten studenten zowel hun eigen performantie en hun eigen leren beoordelen als leerdoelen leren formuleren. Opdat ze deze vaardigheden zouden ontwikkelen, is het noodzakelijk dat de beoordeling afgestemd is op wat studenten zouden moeten leren (= constructive alignment; Biggs, 1996, 2007). Wat ze zouden moeten leren moet ook opgesteld zijn met het oog op wat er in de toekomst van hen verwacht zal worden (zie ook Boud & Falchikov, 2006; Kicken et al., 2009). Bovendien moeten studenten de beoordelingscriteria begrijpen en op zichzelf leren toepassen zodat ze hun eigen leerproces in handen leren nemen en sturen (iets waartoe feedback heel belangrijk is; zie ook Hattie & Timperly, 2007).

2.2.2.3 Haalbaarheid en tijdigheid

De literatuur benadrukt dat het belangrijk is dat er voldoende tijd voorzien wordt voor de beoordelaars en dat ervoor gezorgd wordt dat wat gemeten moet worden, kan worden gemeten zonder grote bijkomende inspanningen (Hunt, 1992; Turnbull et al., 1998; 2000). Hierbij wordt ook benadrukt dat het belangrijk is ervoor te zorgen dat rapporteringen in de tijd best zo dicht mogelijk aansluiten bij het te observeren gedrag opdat er zo weinig mogelijk vertekening in de rapportering zou sluipen (Govaerts et al., 2007, Sanches & DeLa Torre, 1996; Turnbull et al., 1998; 2000). Idealiter gebeuren observaties dan ook onmiddellijk (= ter plekke) en wordt er een systeem bedacht waarbij deze snel en gemakkelijk geregistreerd kunnen worden. Dit biedt dan ook de mogelijkheid om meer frequent observaties te rapporteren. Dit is belangrijk omdat het functioneren van een student nu eenmaal geen stabiel gegeven is maar van dag tot dag of van taak tot taak kan fluctueren (zie hoger). Wil men een juist beeld van het functioneren van een stagiair krijgen, dan dient men die fluctuaties in rekening te brengen.

2.2.2.4 Relevantie en verantwoordbaarheid

Tot slot wordt er in de literatuur benadrukt dat de evaluatie voor alle betrokken partijen relevant moet zijn (Turnbull et al., 1998; 2000). Zowel positieve als negatieve beoordelingen moeten daarom consequenties hebben, ook al is het hard is om studenten tegen te houden (Short, 1993). Een hele reeks auteurs benadrukt in dit kader het belang van tussentijdse feedback. Tussentijdse feedback geeft studenten een duidelijke richting van de pijnpunten waar ze aan moeten werken, en kan er tegelijkertijd voor zorgen dat een eventuele negatieve eindevaluatie niet als een verrassing komt, wat (juridische) problemen kan besparen (Govaerts et al., 2007; Hunt, 1992; Short 1993; Tigelaar & van der Vleuten, 2014; van der Vleuten et al., 2010; van der Vleuten & Schuwirth, 2005). Omwille van die consequenties moeten evaluaties ook verdedigbaar zijn. Vanuit die optiek pleiten verschillende auteurs voor een duidelijk narratieve component en een shift van cijfers naar woorden (zie o.a. Govaerts et al., 2007; Govaerts & van der Vleuten, 2013; van der Vleuten & Schuwirth, 2005).

2.2.2.5 Samenvattend

Mogelijke interventies ter verbetering van stagebeoordelingen kan men indelen in drie categorieën. Omwille van de nadruk die tegenwoordig gelegd wordt op het procesmatige karakter van stages, beogen deze interventies niet enkel een effect te hebben op de kwaliteit van de beoordeling maar ook op de kwaliteit van het hele stageproces. We sommen hieronder de drie categorieën op waar dergelijke interventies in onderverdeeld kunnen worden.

1. Gebruik meer beoordelingsmethoden
   - Checklists en rating scales, al dan niet met ruimte voor commentaar.
   - Observaties (Govaerts & van der Vleuten, 2013), videofragmenten (van der Vleuten & Schuwirth, 2005) of high fidelity tasks zoals simulaties waarin de nodige skills kunnen worden getoond (McNamara, 2013; Tigelaar & van der Vleuten, 2014).
   - Kwalitatieve informatie (van der Vleuten & Schuwirth, 2005) zoals een log- of dagboek, een reflectieverslag, een eindwerk of een (e)portfolio (Anema & McCoy, 2010; Comfort & Ferns, 2014; Kicken et al., 2009; Tigelaar & van der Vleuten, 2014). Kwalitatieve informatie kan ook bekomen worden uit gesprekken of individuele presentaties (McNamara, 2013), of via werkvormen zoals supervisie of intervisie.
2. Gebruik meer ratings en observaties
   - Laat meer mensen beoordelen
     - Meer docenten (= stagebegeleiders binnen de onderwijsinstelling)
     - Meer mentoren (= stagebegeleiders op de werkplek)
     - Collega's en leidinggevenden op de werkplek
     - Derden (klanten, patiënten, …)
     - Medestudenten (= peer assessment)
     - De student zelf (= self-assessment)
   - Laat beoordelaars meer ratings doen
   - Gebruik verschillende types beoordelaars
3. Organiseer trainingen voor docenten en mentoren
   - Leer hen de beoordelingsmethoden te hanteren.
   - Probeer de manier waarop ze feedback geven te verbeteren.
   - Stem de criteria en verwachtingen van de actoren op elkaar af.

Bij het uitwerken van interventies ter verbetering van stagebeoordelingen is het belangrijk om de andere aanbevelingen die werden opgesomd niet uit het oog te verliezen.

- Verbeteringen moeten niet enkel gericht zijn op het meten van maar ook op het ondersteunen van het ontwikkelingsproces van studenten. Het einddoel dat men voor ogen moet houden is dat de student moet uitgroeien tot een professional die in staat in levenslang zelfstandig te leren. Tussentijdse feedback is hiervoor onmisbaar. Om dit ontwikkelingsproces adequaat te kunnen ondersteunen dient men rekening te houden met de specificiteit van een stage (= omvattendheid en flexibiliteit).
- Tussentijdse feedback dient gebaseerd te zijn op relevante observaties. Hiertoe is een systeem nodig dat toelaat dergelijke observaties snel en efficiënt te rapporteren opdat begeleiders niet overbelast zouden worden (= haalbaarheid en tijdigheid).
- Het eindoordeel moet relevant zijn en consequenties hebben. Een goede verantwoording van dit oordeel is dan ook heel belangrijk. Ook hiervoor is tussentijdse feedback essentieel. Tussentijdse feedback biedt immers niet alleen groeikansen, het bereidt de student ook voor op het eindoordeel (= relevantie en verantwoordbaarheid).

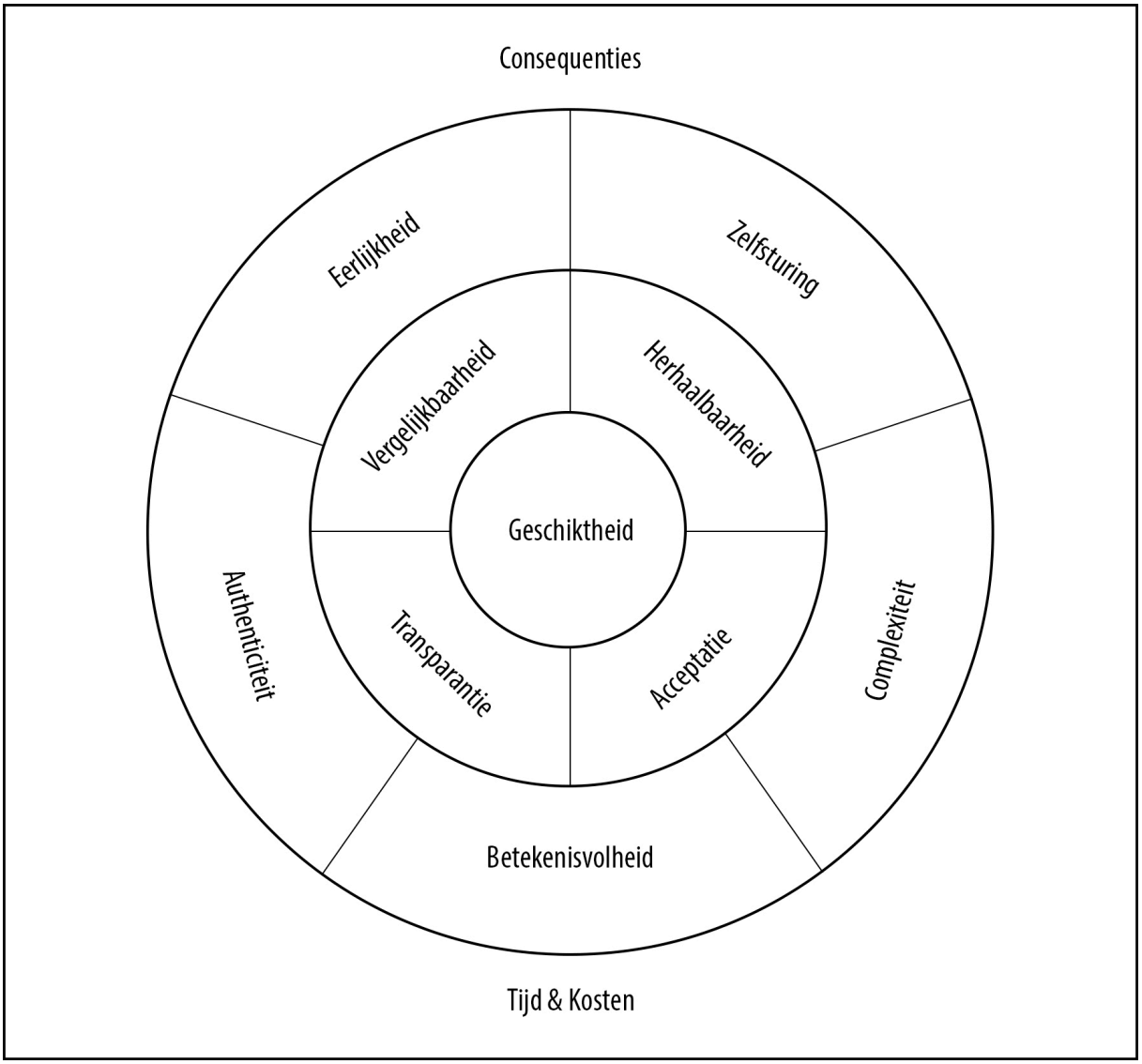
2.2.3 Naar kwaliteitscontrole

Om de kwaliteit van een beoordeling te meten, ging men vroeger vanuit een klassieke, psychometrische benadering op zoek naar het ideale meetinstrument om de "ware" score te bepalen

die het functioneren van een student eenduidig weerspiegelt. Vanuit een procesmatige benadering wijzen verschillende oordelen echter niet per se op meetfouten en doet een exclusieve focus op het meetinstrument zelf afbreuk aan de complexiteit van het beoordelingsproces dat altijd en onherroepelijk tot op zekere hoogte een subjectief gegeven is. Wanneer men er echter vanuit een procesmatige benadering voor kiest om betrouwbaarheid in de klassieke, psychometrische zin van het woord los te laten als kwaliteitscriterium voor een beoordeling, dan stelt er zich een probleem. Wat zijn dan wel nog mogelijke kwaliteitscriteria?

In hun "Competentie Assessment Wiel" schuiven Baartman et al. (2006, 2007) twaalf kwaliteitscriteria naar voor. Deze worden in lagen gerepresenteerd omwille van het feit dat ze aan elkaar gerelateerd zijn (zie Figuur 2.1 en Tabel 2.1). De binnenste lagen worden gezien als voorwaarden voor de buitenste lagen. In dit model komen ook alle eerder opgesomde aanbevelingen terug. Elk van de twaalf criteria kan een indicator zijn voor de effectiviteit van een interventie ter verbetering van een stagebeoordeling, al zullen niet alle criteria altijd even relevant zijn. Baartman et al. (2006, 2007) benadrukken immers dat een beoordelingsprogramma meestal uit verschillende beoordelingsmethoden bestaat en dat het niet zo is dat elke beoordelingsmethode aan al deze criteria moet voldoen. Het beoordelingsprogramma als geheel zou echter wel aan alle criteria moeten voldoen. Voor de beoordeling van een specifieke interventie zullen sommige criteria dus meer relevant zijn dan andere.

Figuur 2.1: De kwaliteitscriteria uit het Competentie Assessment Wiel



Tabel 2.1: Toelichting bij de kwaliteitscriteria uit het Competentie Assessment Wiel

Criterium	Toelichting
01. Geschiktheid	De beoordeling is gebaseerd op de competenties die studenten dienen te ontwikkelen.
02. Vergelijkbaarheid	De opdrachten voor studenten, de beoordelingsprocedure en de beoordelingscriteria zijn vergelijkbaar voor alle studenten.
03. Herhaalbaarheid	De beslissingen die worden genomen over studenten zijn gebaseerd op meerdere assessmentsituaties en meerdere beoordelaars.
04. Transparantie	Betrokkenen (= studenten, docenten en mentoren) weten wat en hoe wordt beoordeeld, wat de criteria en procedures zijn en hoe de beoordeling meetelt voor de diplomering.
05. Acceptatie	De beoordelingscriteria zijn voor alle betrokkenen aanvaardbaar.
06. Authenticiteit	De beoordelingsvormen stemmen overeen met de toekomstige beroepssituatie (bijv. wat betreft fysieke en sociale omstandigheden).
07. Eerlijkheid	Alle studenten krijgen een eerlijke kans hun competenties te demonstreren (bijv. door verschillende opdrachten en bekwame beoordelaars te gebruiken) en ervaren de beoordeling ook als eerlijk.
08. Complexiteit	De beoordeling laat toe de denkprocessen te beoordelen die een student doorloopt als die aan een opdracht werkt. Hierdoor wordt zichtbaar of een student de nodige cognitieve vaardigheden heeft.
09. Betekenisvolheid	De beoordeling levert waardevolle feedback op voor studenten en is naast een beoordelingsmoment ook een leermoment.
10. Zelfsturing	Door de beoordeling leren studenten hoe ze zelfstandig hun leerproces kunnen sturen. Ze leren zichzelf en elkaar beoordelen, ze leren reflecteren en ze leren hun eigen leerdoelen formuleren.
11. Tijd & Kosten	De beoordeling is haalbaarheid voor alle betrokkenen.
12. Consequenties	De beoordeling beïnvloedt het leren van studenten en de begeleiding die wordt gegeven door docenten en mentoren. Zoals verderop zal blijken, zijn er aan aantal mogelijke consequenties die in rekening gebracht kunnen worden: (1) Selectie (de mate waarin goede van slechte studenten kunnen worden onderscheiden), Leereffect (het leereffect dat een beoordeling genereert), (3) Onderwijseffect (het effect dat de beoordeling heeft op de wijze waarop de opleiding vorm wordt gegeven), en (4) Motivatie (het effect dat een beoordeling heeft op de motivatie van studenten en begeleiders).

3 VOORSTUDIE 1

Aan het einde van het academiejaar 2016-2017 werd door het D-PAC-team van de Universiteit Antwerpen een onderzoek gevoerd in het Flex-traject van de opleiding Professionele Bachelor Lager Onderwijs (PBLO). De bedoeling was om een alternatieve beoordelingswijze (i.e., D-PAC; zie o.a., Lesterhuis et al., 2017) uit te proberen in een nieuwe context, met name die van de stage. Aangezien de doelstelling van dit onderzoek aansloot bij de doelstelling van ons project, besloten we dit onderzoek in ons project in te schakelen. We waren hierbij geïnteresseerd in of deze alternatieve beoordelingswijze als een "verbetering" kon worden aanzien ten aanzien van de gangbare manier van werken. Meer specifiek wilden we te weten komen hoe deze werkwijze onthaald zou worden door de betrokken actoren (i.e., studenten, docenten en mentoren) in termen van de kwaliteitscriteria uit het model van Baartman. Voor dat doel construeerden we een vragenlijst die deze kwaliteitscriteria probeert te meten. Naast een evaluatie van D-PAC beoogde deze studie ook een eerste test te zijn van deze vragenlijst.

Op welke wijze sluit deze interventie aan bij het theoretisch kader? In eerste instantie wordt hier geëxperimenteerd met de beoordelingsmethode (zie "Gebruik meer beoordelingsmethoden"). Meer bepaald wordt D-PAC als beoordelingsmethode geïntroduceerd. Daarnaast zet deze interventie ook in op ratings en observaties (zie "Gebruik meer ratings en observaties"). Specifiek probeert men meer mensen te betrekken in het beoordelingsproces en gebruik te maken van verschillende types beoordelaars (i.e., niet alleen docenten en mentoren maar ook de studenten zelf). Verder zet D-PAC ook in op de wijze waarop mensen de beoordelingsmethode hanteren en de criteria die ze daarbij gebruiken (zie "Organiseer trainingen voor docenten en mentoren"), zij het op alternatieve wijze. In plaats van de criteria strikt te definiëren, houdt men deze bewust vaag zodat iedere beoordelaar ze kan invullen zoals ie zelf verkiest. Een deel van de interventie bestaat er immers in zicht te krijgen op de criteria die mensen gebruiken bij het beoordelen door afwijkende antwoordpatronen te identificeren en onder de loep te nemen. Dit is een eerste stap richting het bespreekbaar maken van dergelijke verschillen.

3.1 Methode

3.1.1 Procedure

In een voorbereidende fase leverden de studenten uit het Flex-traject elk vier videofragmenten van zichzelf voor de klas aan met de bedoeling hun startbekwaamheid als leerkracht aan te tonen. In een latere fase kregen deze studenten, evenals een aantal docenten (= stagebegeleiders binnen KdG) en mentoren (= stagebegeleiders op de werkplek), paarsgewijs een aantal videofragmenten aangeboden (Lesterhuis et al., 2017). Binnen elk paar werd gevraagd het filmpje te kiezen dat de meest startbekwame leerkracht toont. De onderliggende assumptie is dat het voor mensen gemakkelijker is om dergelijke vergelijkingen te maken dan om één enkele prestatie te beoordelen (Thurstone, 1927) omdat een oordeel nu eenmaal altijd resulteert uit een vergelijkingsproces (Laming, 2003). In totaal moest elke beoordelaar 10 dergelijke vergelijkingen maken. Het aantal keer dat een videofragment als winnaar uit een dergelijke vergelijking komt, verschaft informatie over de rangorde van dat videofragment. Na toepassing van het Bradley-Terry-Luce model (BTL; Bradley & Terry, 1952; Luce, 1959) resulteert deze procedure in een rangorde van studenten van minst naar meest startbekwaam (Lesterhuis et al., 2017). Indien gewenst kan de opleiding in deze rangorde een cesuur aanbrengen (= aangeven

wat het punt is waarboven studenten wel en waaronder studenten als niet geslaagd dienen te worden beschouwd) en er een puntenverdeling aan toekennen. Verder laat het BTL-model ook toe te berekenen hoe afwijkend beoordelaars zich gedragen ten aanzien van elkaar en hoe groot de consensus is over de verschillende videofragmenten (Lesterhuis et al., 2017).

Na de interventie kregen de deelnemers via e-mail een elektronische vragenlijst toegestuurd waarin naar de perceptie van de beoordelingsprocedure gepeild werd aan de hand van de kwaliteitscriteria uit het model van Baartman (zie Tabel 2.1). Na Periode 4 werden van de deelnemende studenten ook nog een aantal relevante achtergrondkenmerken uit Bamaflex gehaald, evenals de punten die deze studenten voor hun stage verkregen. Dit laatste liet ons toe na te gaan in welke mate de beoordeling van de kwaliteitscriteria gekleurd werd door het eindcijfer. Het is immers niet onlogisch dat studenten die slecht scoren via een bepaalde beoordelingswijze negatiever zullen staan ten aanzien van die beoordelingswijze.

3.1.2 Steekproef

Van de 18 studenten die waren ingeschreven in het Flex-traject deden er 12 (= 66%) mee aan de interventie. Verder namen 13 begeleiders deel (6 docenten en 7 mentoren). In totaal namen dus 25 mensen deel aan de interventie. Alle docenten en het overgrote deel van de mentoren (71%) waren vrouwen. De studenten bestonden voor 100% uit vrouwen. Hun gemiddelde leeftijd was 22.62 jaar (SD = 5.18). Qua vooropleiding komen 9 studenten uit het ASO (75%) en 3 uit het TSO (25%). Op basis van deze gegevens werd een variabele "educatie" geconstrueerd waarop studenten een 1 kregen als ze uit het BSO kwamen, een 2 als ze uit het TSO of het KSO kwamen of als hier geen duidelijke gegevens over te vinden waren, en een 3 als ze uit het ASO kwamen. De restcategorie (= 2) vormt de baseline waartegen effecten van een lage (= BSO) en een hoge educatie (= ASO) worden afgezet (zie Duriez et al., 2016). Alle studenten hadden de Belgische nationaliteit, zijn in België geboren, en hebben een grootmoeder langs moederszijde die in België geboren is. Op basis van deze gegevens werd een variabele "etniciteit" geconstrueerd waarop studenten een 0 kregen als ze de Belgische nationaliteit hebben, in België geboren zijn en een grootmoeder van Belgische nationaliteit hebben en een 1 als aan één van deze voorwaarden niet is voldaan (Duquet, Gloriaux, Laurijssen, & Van Dorsselaer, 2006).

Om te weten of het hier om een aselecte steekproef van studenten ging, onderzochten we via Oneway Anova's of de groep die deelnam aan de interventie vergelijkbaar was in termen van achtergrondkenmerken met de groep die niet deelnam (N = 6). De groep die aan de interventie deelnam heeft een hogere educatie dan de groep die niet deelnam (Mean = 2.75 en 2.17; $F(1, 17) = 7.06$, $p < .05$). Er waren geen significante verschillen in termen van geslacht (Mean = 0.00 en 0.17; $F(1, 17) = 2.13$, ns), leeftijd (Mean = 26.41 en 27.50; $F(1, 17) = 0.21$, ns) en etniciteit (Mean = 0.00 en 0.17; $F(1, 17) = 2.13$, ns). Hieruit dient te worden geconcludeerd dat de studenten die aan de interventie deelnamen niet helemaal representatief zijn voor de totale populatie: ze hebben gemiddeld genomen een hogere educatie.

De studenten die deelnamen aan de interventie haalden een gemiddelde score van 10.83 (SD = 3.88) op hun stage. Slechts 1 student slaagde niet voor de stage (die student haalde een 0). Door de kleine steekproefgrootte verschilden de studenten die deelnamen aan de interventie niet significant van de studenten die niet deelnamen in termen van stageresultaat (Mean = 10.83 en 6.50; $F(1, 17) = 3.46$, ns), iets wat te wijten is aan. De deelnemende studenten

waren dus representatief voor de totale studentenpopulatie. Via regressieanalyse onderzochten we of het stageresultaat kon worden voorspeld op basis van de achtergrondkenmerken. Hoewel de achtergrondkenmerken 30% van de variantie in stageresultaat konden verklaren, was dit, gegeven de kleine steekproef, niet significant ($R^2 = .30$; $F(4, 13) = 2.78$, ns).

3.1.3 Effectmetingen

3.1.3.1 Kwaliteitscriteria

Na de interventie kregen de studenten, docenten en mentoren (N = 25) een zelf-geconstrueerde vragenlijst die de perceptie van de interventie poogde te meten aan de hand van de kwaliteitscriteria uit het model van Baartman (zie Tabel 2.1). In later werk van Baartman kreeg Criterium #12 (= consequenties) meer aandacht (Baartman, Kloppenburg & Prins, 2014). Dit werd uitgesplitst in drie deelcriteria: (1) Selectie, of de mate waarin de beoordeling geschikt is om sterke van zwakke studenten te onderscheiden, (2) Leereffect, of het effect dat de beoordeling heeft op de ontwikkeling van studenten, en (3) Onderwijseffect, of het effect dat de beoordeling heeft op de wijze waarop de opleiding vorm wordt gegeven. Op basis van motivatietheorieën, zoals de Zelf-Determinatie Theorie (Deci & Ryan, 2000), besloten we hier nog een vierde deel aan toe te voegen dat aansluit bij het Leereffectcriterium: Motivatie, of de mate waarin de beoordeling de motivatie van studenten beïnvloedt. Het is immers plausibel dat een beoordeling een effect heeft op de interesse van studenten in het vakgebied en dus ook op hun motivatie. In totaal beoogt de gebruikte vragenlijst dus 15 criteria te meten: (1) Geschiktheid, (2) Vergelijkbaarheid, (3) Herhaalbaarheid, (4) Transparantie, (5) Acceptatie, (6) Authenticiteit, (7) Eerlijkheid, (8) Complexiteit, (9) Betekenisvolheid, (10) Zelfsturing, (11) Tijd & Kosten, (12) Selectie, (13) Leereffect, (14) Motivatie, en (15) Onderwijseffect. Voor elk criterium werden 5 vragen opgesteld die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 1. Negatief geformuleerde items zijn aangeduid met "(-)". De scoring van deze items werd uiteraard eerst geïnverteerd alvorens onze analyses aan te vatten.

Als indicator van de betrouwbaarheid van een schaal (= diens interne consistentie), wordt doorgaans de Cronbach alfa van die schaal berekend (Cronbach, 1951). De Cronbach alfas voor de 15 schalen in deze vragenlijst waren .89 voor Geschiktheid, .65 voor Vergelijkbaarheid, .52 voor Herhaalbaarheid, .74 voor Transparantie, .86 voor Acceptatie, .93 voor Authenticiteit, .92 voor Eerlijkheid, .68 voor Complexiteit, .96 voor Betekenisvolheid, .89 voor Zelfsturing, .86 voor Tijd & Kosten, .87 voor Selectie, .89 voor Leereffect, .95 voor Motivatie, en .69 voor Onderwijseffect. Voor Vergelijkbaarheid, Herhaalbaarheid, Complexiteit en Onderwijseffect haalde de Cronbach alfa niet het gangbaar vereiste minimum van .70. Voor Vergelijkbaarheid steeg de interne consistentie evenwel tot .80 wanneer item 09 en 10 werden weggelaten. Voor Herhaalbaarheid steeg de interne consistentie tot .82 wanneer item 11 en 12 werden weggelaten. Voor Complexiteit steeg de interne consistentie tot .73 wanneer item 36 en 37 werden weggelaten. Voor Onderwijseffect steeg de interne consistentie tot .73 wanneer item 64 werd weggelaten. Deze items werden dan ook weggelaten in de berekening van deze schalen.

De correlaties tussen deze schalen zijn te vinden in Tabel 3.1. Uit deze tabel blijkt dat er sterke correlaties zijn tussen heel wat schalen. Zo is er een uitermate sterke correlatie tussen zelfsturing en leereffect. Niet verwonderlijk, want inhoudelijk liggen deze twee schalen heel erg dicht

bij elkaar liggen. Verder zijn er onder andere ook sterke correlaties tussen bijvoorbeeld Geschiktheid, Acceptatie, Authenticiteit, Eerlijkheid, Selectie en Tijd en kosten. Wie een beoordeling geschikt vindt, lijkt die ook acceptabel, authentiek, eerlijk en geschikt voor selectie te vinden en de zinvolheid van de vereiste tijd en kosten in te zien.

Gezien de substantiële correlaties tussen heel wat schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. De scree plot van een exploratorische factoranalyse wees op een oplossing met drie factoren. Deze factoren vielen echter moeilijk te interpreteren. Hetzelfde kon gezegd worden van een oplossing met twee en vier factoren. Wat (uiteraard) wel te interpreteren viel, was een oplossing met één factor. Een dergelijke oplossing kon 44% van de totale variantie verklaren. De enige schalen met een factorlading onder de gebruikelijke drempel van .40 waren Transparantie en Herhaalbaarheid. Alle andere schalen hadden een factorlading > .40 op deze factor. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze algemene score had een Cronbach alfa van .90.

Binnen de groep studenten gingen we via regressieanalyse na of deze algemene score alsook de score op de 15 onderliggende schalen kon worden voorspeld op basis van stageresultaat en de achtergrondkenmerken leeftijd en educatie (geslacht en etniciteit werden niet in de analyses meegenomen omdat de studenten die aan de interventie deelnamen niet onderling verschilden in termen van geslacht en etniciteit). De resultaten van deze analyses zijn te vinden in Tabel 3.2. Noch de algemene score noch de 15 onderliggende schalen konden significant worden voorspeld door de achtergrondkenmerken en het stageresultaat. Uiteraard is dit voor een groot stuk te wijten aan de wel erg kleine steekproefgrootte.

3.2 Resultaten

3.2.1 Kwaliteitscriteria

De gemiddelden en standaarddeviaties voor de algemene score en de subschalen zijn te vinden in Tabel 3.3. Algemeen genomen werd deze interventie vrij neutraal beoordeeld: T-tests wezen uit dat het gemiddelde op de algemene score voor alle deelnemers niet significant verschilde van 0 (= het neutraal middelpunt van de schaal). De interventie werd gemiddeld genomen wel als Authentiek en geschikt voor Zelfsturing ervaren (op deze criteria wijkt het gemiddelde significant af van 0 in positieve richting), maar ook als niet Eerlijk, niet Herhaalbaar, weinig Transparant en niet het potentieel hebbend om een Onderwijseffect te genereren (op deze criteria wijkt het gemiddelde significant af van 0 in negatieve richting). Oneway Anova's legden echter belangrijke verschillen bloot tussen studenten en begeleiders. Hieruit bleek dat het enkel de begeleiders waren die positieve punten aan deze beoordelingswijze zagen. Het waren de begeleiders die de interventie als Authentiek en geschikt voor Zelfsturing zagen, niet de studenten. Daarenboven zagen zij de interventie ook als Betekenisvol. Minpunten voor begeleiders waren Transparantie en Herhaalbaarheid. Studenten zagen niet echt positieve punten aan deze methode maar wel een aantal minpunten. Zij zagen minpunten op vlak van Herhaalbaarheid en Complexiteit, maar vooral op vlak van Eerlijkheid. Studenten zagen de interventie dus negatiever dan begeleiders, al was het verschil tussen beiden enkel significant voor Eerlijkheid.

3.3 Conclusie

Over het algemeen werd deze interventie door begeleiders eerder neutraal en door studenten eerder negatief beoordeeld. Begeleiders zagen de interventie als Authentiek, Betekenisvol en geschikt voor Zelfsturing. Studenten traden dit echter niet bij. De begeleiders zagen ook minpunten aan deze methode. Ze zagen problemen inzake Herhaalbaarheid en Transparantie. Studenten zagen eveneens problemen wat Herhaalbaarheid betreft, maar ook op vlak van Complexiteit en Eerlijkheid. Zowel studenten als begeleiders vonden dus dat deze methode te weinig assessmentsituaties bevatte en/of te weinig beoordelaars gebruikte. Begeleiders maakten zich bovendien zorgen om de transparantie, terwijl studenten vooral bezorgd waren dat ze op deze manier geen eerlijke kans kregen om hun competenties te tonen. Zij vonden deze methode ook niet geschikt om hun cognitieve vaardigheden mee te beoordelen. De bedenkingen die begeleiders en studenten bij deze methode hadden, kunnen uiteraard ook deels het gevolg zijn van de manier waarop deze methode in deze studie werd aangewend. Mogelijks zouden de resultaten anders geweest zijn mocht er meer tijd genomen zijn om de deelnemers vertrouwd te maken met (de onderliggende filosofie van) deze manier van beoordelen.

Naast een evaluatie van D-PAC beoogde deze studie ook een eerste test te zijn voor de vragenlijst die de kwaliteitscriteria uit het model van Baartman beoogt te meten. De interne consistentie van de schalen was over het algemeen goed. Voor vier schalen (i.e., Vergelijkbaarheid, Herhaalbaarheid, Complexiteit en Onderwijseffect) werd de gebruikelijke norm echter slechts gehaald na verwijdering van een aantal items. Wat vergelijkbaarheid betreft, dienden die vragen verwijderd te worden die verwijzen naar de vergelijkbaarheid tussen beoordelaars. Deze bleken dus iets anders te meten dan de vragen die polsten naar de vergelijkbaarheid van de taak op zich (i.e., de opdrachten en de criteria). Wat herhaalbaarheid betreft, dienden die vragen te worden verwijderd die iets objectief meten: of er al dan niet meerdere beoordelaars betrokken waren. De andere vragen verwezen naar de objectiviteit van de oordelen op zich. Voor Complexiteit en Onderwijseffect was het niet duidelijk waarom bepaalde items dienden te worden verwijderd. Dit is op zich niet zo vreemd, want ook met de inclusie van deze items was de interne consistentie nog steeds behoorlijk. Verder viel het op dat er sterk positieve correlaties waren tussen de meeste schalen. Een factoranalyse wees dan ook uit dat een oplossing met één factor een acceptabele oplossing was. Alle schalen, met uitzondering van Transparantie en Herhaalbaarheid, hadden een acceptabele lading op deze factor. Deze factor kan dus gebruikt worden als indicatie van de algemene perceptie van een stagebeoordeling, waarna er meer in detail naar de verschillende subschalen kan worden gekeken. Omdat het hier om een wel erg kleine steekproef ging, hebben we echter weinig zekerheid over de nauwkeurigheid van de geschatte interne consistenties, de kwaliteit van de items, en de berekende factoroplossing. Daarom kozen we ervoor een tweede voorstudie te doen. Die heeft drie doelen: (1) het testen van een aantal vervangitems, (2) een nieuwe evaluatie van de interne consistenties en factorstructuur, en (3) komen tot een onderbouwde reductie van het aantal items.

Tabel 3.1: Correlaties tussen de kwaliteitscriteria

Criteria	01	02	03	04	05	06	07	08	09	10	11	12	13	14
01. Geschikt														
02. Vergelijk	.33													
03. Herhaalbaar	.35	-.06												
04. Transparant	.34	.44 [1]	.48 [1]											
05. Acceptatie	.78 [2]	.31	.51 [2]	.47 [1]										
06. Authentiek	.71 [2]	.46 [1]	.21	.39	.77 [2]									
07. Eerlijk	.78 [2]	.34	.33	.25	.71 [2]	.70 [2]								
08. Complex	.39	.20	.07	.29	.20	.45 [1]	.43 [1]							
09. Betekenis	.34	.12	.11	.26	.48 [1]	.42 [1]	.37	.27						
10. Zelfsturing	.13	-.08	-.33	-.01	.31	.23	.18	.15	.63 [2]					
11. Tijd/Kosten	.68 [2]	.40 [1]	.22	.22	.71 [2]	.63 [2]	.70 [2]	.30	.33	.31				
12. Selectie	.75 [2]	.45 [1]	-.02	.16	.66 [2]	.73 [2]	.59 [2]	.34	.33	.26	.60 [2]			
13. Leereffect	.29	.14	-.25	-.05	.33	.36	.09	.25	.65 [2]	.81 [2]	.45 [1]	.45 [1]		
14. Motivatie	.51 [2]	.15	-.30	.15	.39	.45 [1]	.45 [1]	.17	.37	.61 [2]	.56 [2]	.66 [2]	.66 [2]	
15. Onderwijs	.42 [1]	.55 [2]	-.15	.05	.36	.48 [1]	.43 [1]	.20	.36	.27	.45 [1]	.61 [2]	.62 [2]	.56 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 3.2: Regressieanalyses op de kwaliteitscriteria

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Resultaat
00. Algemeen	.01	-	.14	.05	-	.17
01. Geschikt	.14	-	-.40	.14	-	-.53
02. Vergelijk	.38	-	.34	-.61	-	.26
03. Herhaalbaar	.04	-	.23	.11	-	.26
04. Transparant	.23	-	.15	-.44	-	.28
05. Acceptatie	.11	-	-.09	.33	-	-.07
06. Authentiek	.09	-	.35	.13	-	.14
07. Eerlijk	.24	-	.64	.22	-	.53
08. Complex	.37	-	.85	-.13	-	.90
09. Betekenis	.31	-	.07	-.45	-	.35
10. Zelfsturing	.11	-	-.41	.06	-	-.12
11. Tijd/Kosten	.08	-	-.15	.26	-	-.19
12. Selectie	.28	-	-.34	.50	-	-.16
13. Leereffect	.13	-	-.21	.26	-	.10
14. Motivatie	.21	-	-.15	.30	-	-.43
15. Onderwijs	.03	-	.23	.06	-	.10

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het stageresultaat. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 3.3: Gemiddelden per groep op de kwaliteitscriteria

Criteria	Iedereen	Studenten	Begeleiders	ΔBS
00. Algemeen	-0.14 (0.94) [0]	-0.41 (0.64) [1]	0.11 (1.13) [0]	0
01. Geschikt	-0.21 (1.54) [0]	-0.57 (1.11) [0]	0.12 (1.82) [0]	0
02. Vergelijk	-0.21 (1.60) [0]	0.03 (1.35) [0]	-0.44 (1.82) [0]	0
03. Herhaalbaar	-1.17 (1.16) [2]	-1.19 (1.32) [1]	-1.15 (1.04) [2]	0
04. Transparant	-0.62 (1.28) [1]	-0.25 (1.44) [0]	-0.97 (0.92) [2]	0
05. Acceptatie	0.06 (1.34) [0]	-0.15 (1.21) [0]	0.25 (1.48) [0]	0
06. Authentiek	0.72 (1.65) [1]	0.22 (1.54) [0]	1.18 (1.66) [1]	0
07. Eerlijk	-1.38 (1.54) [2]	-2.10 (0.99) [2]	-0.72 (1.69) [0]	1
08. Complex	-0.36 (1.35) [0]	-0.80 (0.96) [1]	0.15 (1.52) [0]	0
09. Betekenis	0.70 (1.84) [0]	0.13 (1.80) [0]	1.22 (1.78) [1]	0
10. Zelfsturing	0.82 (1.33) [2]	0.63 (1.17) [0]	1.00 (1.49) [1]	0
11. Tijd/Kosten	-0.23 (1.46) [0]	-0.70 (1.18) [0]	0.20 (1.60) [0]	0
12. Selectie	0.25 (1.43) [0]	-0.33 (1.25) [0]	0.51 (1.59) [0]	0
13. Leereffect	0.30 (1.36) [0]	0.03 (0.99) [0]	0.55 (1.64) [0]	0
14. Motivatie	-0.24 (1.81) [0]	-0.73 (1.51) [0]	0.22 (2.00) [0]	0
15. Onderwijs	-0.55 (1.22) [1]	-0.67 (1.28) [0]	-0.44 (1.21) [0]	0

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Iedereen" is met een subscript aangegeven of het gemiddelde voor alle deelnemers verschilt van 0. Ook in de kolommen "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

4 VOORSTUDIE 2

Voor onze uiteindelijk studies is het van belang dat onze metingen van een goede kwaliteit zijn. Als de metingen beter zijn, moeten we minder vragen stellen om te meten wat we willen meten, waardoor de deelnemers aan onze studies ook minder vragen moeten beantwoorden. Voorstudie 1 liet een eerste test toe van de vragenlijst die we ontwikkelden om de kwaliteitscriteria uit het model van Baartman te meten. Helaas was de steekproef dermate klein dat die niet geschikt was om op basis hiervan definitieve keuzes betreffende deze vragenlijst te maken. We besloten dan ook een tweede voorstudie te doen. Deze heeft drie doelen: (1) het testen van een aantal vervangitems (i.e., voor Vergelijkbaarheid en Herhaalbaarheid), (2) een herevaluatie van de interne consistenties en factorstructuur, en (3) komen tot een onderbouwde reductie van het aantal items met het oog op onze eigenlijke studies. Wat dat laatste betreft, stelden we ons tot doel te komen tot een vragenlijst waarin elk criterium gemeten wordt aan de hand van 3 (in plaats van 5) items. Uit praktische overwegingen gingen we op zoek naar een opleiding waarvan de studenten een volledige stage liepen in het najaar van 2017. Dit was immers een vereiste om deze herevaluatie op tijd rond te krijgen voor onze eigenlijke studies. Zo kwamen we terecht bij het tweede jaar van de opleiding Verpleegkunde. Om praktische redenen beperkten we de bevraging tot de studenten. Die konden we namelijk collectief bereiken.

4.1 Methode

4.1.1 Procedure

Na de stageperiode kregen de studenten van het tweede jaar Verpleegkunde tijdens een verplicht lesmoment een papieren vragenlijst aangeboden die bestond uit twee onderdelen.

1. Eerst werden een aantal persoonsgegevens bevraagd (i.e., naam, voornaam en e-mailadres) alsook een aantal achtergrondkenmerken (i.e., leeftijd, geslacht, vooropleiding, en etniciteit). Vervolgens werd ook gevraagd wat volgens hen het punt is dat ze verdienen op hun stage (= een self-assessment).
2. Daarna werd gepolst naar de perceptie van de algemene beoordelingsprocedure van de stage aan de hand van de kwaliteitscriteria uit het model van Baartman.

4.1.2 Steekproef

In totaal vulden 101 tweedejaarsstudenten Verpleegkunde onze vragenlijst in. De studenten bestonden voor 83% uit vrouwen. Hun gemiddelde leeftijd was 20.87 jaar (SD = 4.13). Qua vooropleiding komen 8 studenten uit het BSO (7.9%), 52 uit het TSO (51.5%), 1 uit het KSO (1%) en 37 uit het ASO (36.6%). Over 3 (3%) was niets geweten. Op basis van deze gegevens werd een variabele "educatie" geconstrueerd waarop studenten een 1 kregen als ze uit het BSO kwamen, een 2 als ze uit het TSO of het KSO kwamen of als hier geen duidelijke gegevens over te vinden waren, en een 3 als ze uit het ASO kwamen (zie Voorstudie 1). De restcategorie (= 2) vormt de baseline waartegen effecten van een lage (= BSO) en een hoge educatie (= ASO) worden afgezet. Van alle studenten hadden er 98 de Belgische nationaliteit (97%), waren er 91 in België geboren (90,1%), en hadden er 81 een grootmoeder langs moederszijde die in België geboren is (80,2%). Op basis van deze gegevens werd een variabele "etniciteit" geconstrueerd waarop studenten een 0 kregen als ze de Belgische nationaliteit hebben, in België geboren zijn

en een grootmoeder van Belgische nationaliteit hebben en een 1 als aan één van deze voorwaarden niet is voldaan (zie Voorstudie 1). Gemiddeld genomen gaven de studenten zichzelf een score van 14.36 op 20 (SD = 1.62) op hun stage. Geen enkele student gaf zichzelf een score kleiner dan 10. Via regressieanalyse onderzochten we of deze score kon worden voorspeld op basis van de achtergrondkenmerken. Dit was niet het geval. De achtergrondkenmerken verklaarden slechts 2% van de variantie ($R^2 = .02$; $F(4, 96) = 0.59$, ns).

4.1.3 Effectmetingen

4.1.3.1 Kwaliteitscriteria

De studenten kregen een herwerkte versie van de vragenlijst uit Voorstudie 1. Wijzigingen in vergelijking met Voorstudie 1 waren de volgende: (1) we vervingen twee items voor Vergelijkbaarheid door twee nieuwe items, (2) we vervingen twee items voor Herhaalbaarheid door twee nieuwe items, (3) we lieten het criterium Zelfstudie vallen omdat dit inhoudelijk sterk overlapt met het criterium Leereffect (wat ook bleek uit de hoge correlatie tussen beide criteria in Voorstudie 1), en (4) we lieten het criterium Onderwijseffect achterwege. De reden voor deze laatste beslissing is dat studenten niet bezig zijn met het mogelijke onderwijseffect van een beoordeling en dat zij hoogstwaarschijnlijk ook niet in staat zijn om deze vragen op de bedoelde wijze te beantwoorden. In totaal beoogde deze herwerkte vragenlijst zodoende 13 criteria te meten: (1) Geschiktheid, (2) Vergelijkbaarheid, (3) Herhaalbaarheid, (4) Transparantie, (5) Acceptatie, (6) Authenticiteit, (7) Eerlijkheid, (8) Complexiteit, (9) Betekenisvolheid, (10) Tijd & Kosten, (11) Selectie, (12) Leereffect, en (13) Motivatie. Elk criterium werd gemeten aan de hand van 5 vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 2. Negatief geformuleerde items zijn aangeduid met "(-)". De scoring van deze items werd uiteraard eerst geïnverteerd alvorens onze analyses aan te vatten.

Als indicator van de betrouwbaarheid van een schaal (= diens interne consistentie), wordt doorgaans de Cronbach alfa van die schaal berekend. Het gangbaar vereiste minimum van om van een intern consistente schaal te kunnen spreken is .70. De Cronbach alfas van de originele schalen zijn te vinden in Tabel 4.1. Alle schalen haalden het gangbaar vereiste minimum. Om deze schalen in te korten, volgden we de volgende procedure. In Stap 1 werd per schaal gekeken welk item we konden laten vallen om tot een 4-item-schaal met een zo groot mogelijke interne consistentie te komen. In Stap 2 werd deze procedure herhaald op de 4-item-schaal om tot een 3-item-schaal met een zo groot mogelijke interne consistentie te komen. Van elke schaal kon een 3-item-versie gemaakt worden met een goede Cronbach alfa. De Cronbach alfas van de verkorte schalen en de correlaties tussen deze schalen zijn te vinden in Tabel 4.1. De items die op die manier wegvielen zijn in Appendix 2 doorstreept.

Gezien de substantiële correlaties tussen heel wat schalen besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. De scree plot van een exploratorische factoranalyse wees op een oplossing met twee factoren. Deze factoren vielen moeilijk te interpreteren en waren hoog gecorreleerd. Hetzelfde kon gezegd worden van een oplossing met drie factoren. Wat (uiteraard) wel te interpreteren viel, was een oplossing met één factor. Deze kon 44% van de totale variantie verklaren. De enige schalen met een factorlading < .40 waren Vergelijkbaarheid en Herhaalbaarheid. Alle

andere schalen hadden een factorlading > .40 op deze factor. Omdat zo'n factor handig is om een overzicht te krijgen van de algemene perceptie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze score had een Cronbach alfa van .88.

Binnen de groep studenten gingen we via regressieanalyse na of de algemene score alsook de score op de 13 onderliggende schalen kon worden voorspeld op basis van het self-assessment en de achtergrondkenmerken (i.e., leeftijd, geslacht, educatie en etniciteit). De resultaten zijn te vinden in Tabel 4.2. Noch de algemene score noch de 14 onderliggende schalen konden significant worden voorspeld door de achtergrondkenmerken en het self-assessment.

4.2 Resultaten

4.2.1 Kwaliteitscriteria

De gemiddelden en standaarddeviaties voor de algemene score en de subschalen zijn te vinden in Tabel 4.3. Algemeen genomen werd de stagebeoordeling eerder positief beoordeeld: T-tests wezen uit dat het algemeen gemiddelde voor alle deelnemers significant hoger was dan 0 (= het neutraal middelpunt van de schaal). Dit was het geval voor alle subschalen behalve voor Herhaalbaarheid. Voor Herhaalbaarheid lag het gemiddelde significant lager dan 0.

4.3 Conclusie

Alle originele schalen vertoonden een goede interne consistentie. Van elke schaal kon ook een 3-item-versie gemaakt worden met een goede interne consistentie. Verder viel het op dat er sterk positieve correlaties waren tussen de meeste schalen. Een factoranalyse wees dan ook uit dat een oplossing met één factor een acceptabele oplossing was. Alle schalen, met uitzondering van Vergelijkbaarheid en Herhaalbaarheid, hadden een acceptabele lading op deze factor. Deze factor kon dus gebruikt worden als indicatie van de algemene perceptie van een stagebeoordeling, waarna er meer in detail naar de verschillende subschalen kon worden gekeken.

Zijdelings liet deze voorstudie ook toe te concluderen dat de stagebeoordeling in het tweede jaar Verpleegkunde door de studenten over het algemeen vrij positief beoordeeld werd. Voor op één na alle criteria week de score significant af van 0 in positieve richting. Enkel op vlak van Herhaalbaar gaven de studenten aan ontevreden te zijn. De studenten waren dus van mening dat de beslissingen die over hen worden genomen op onvoldoende assessmentsituaties en beoordelaars gebaseerd waren.

Tabel 4.1: Cronbach alfas van en correlaties tussen de kwaliteitscriteria

Criteria	Alfa1	Alfa2	01	02	03	04	05	06	07	08	09	10	11	12
01. Geschikt	.80	.79												
02. Vergelijk	.78	.79	.20 [1]											
03. Herhaalbaar	.79	.86	.11	.37 [2]										
04. Transparant	.82	.82	.25 [1]	.10	.26 [2]									
05. Acceptatie	.82	.84	.39 [2]	.25 [1]	.33 [2]	.47 [2]								
06. Authentiek	.81	.78	.41 [2]	.17	.14	.43 [2]	.49 [2]							
07. Eerlijk	.87	.86	.25 [1]	.38	.39 [2]	.36 [2]	.66 [2]	.37 [2]						
08. Complex	.81	.82	.27 [2]	.08	.29 [2]	.32 [2]	.45 [2]	.50 [2]	.53 [2]					
09. Betekenis	.92	.91	.20 [1]	.03	-.04	.39 [2]	.37 [2]	.47 [2]	.34 [2]	.33 [2]				
10. Tijd/Kosten	.79	.89	.33 [2]	.19	.19	.24 [1]	.54 [2]	.53 [2]	.50 [2]	.60 [2]	.61 [2]			
11. Selectie	.87	.83	.19	.22 [1]	.29 [2]	.24 [1]	.49 [2]	.44 [2]	.59 [2]	.46 [2]	.40 [2]	.57 [2]		
12. Leereffect	.88	.86	.34 [2]	.08	.12	.42 [2]	.52 [2]	.54 [2]	.44 [2]	.51 [2]	.71 [2]	.58 [2]	.53 [2]	
13. Motivatie	.93	.91	.31 [2]	.09	.08	.34 [2]	.46 [2]	.48 [2]	.33 [2]	.54 [2]	.47 [2]	.66 [2]	.44 [2]	.66 [2]

Noot: Alfa1 = Cronbach alfa van de originele schalen; Alfa2 = Cronbach alfa van de verkorte schalen.

[1] $p < .05$; [2] $p < .01$.

Tabel 4.2: Regressieanalyses op de kwaliteitscriteria

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.01	-.00	.04	-.01	.09	.05
01. Geschikt	.07	-.12	.11	-.19	-.06	.06
02. Vergelijk	.02	-.08	-.04	.00	.11	.03
03. Herhaalbaar	.07	-.06	-.02	-.02	.06	.25 [1]
04. Transparant	.01	.04	.05	.03	-.03	-.08
05. Acceptatie	.02	.08	.01	-.05	.07	.01
06. Authentiek	.01	-.09	.05	-.03	-.01	-.01
07. Eerlijk	.00	.01	-.01	-.02	-.04	-.05
08. Complex	.05	.00	.10	.10	.19	-.08
09. Betekenis	.05	.13	.01	.03	.07	-.14
10. Tijd/Kosten	.04	-.03	.10	-.08	.12	.01
11. Selectie	.03	-.05	.02	-.02	.05	.16
12. Leereffect	.03	.09	.11	.12	.07	.03
13. Motivatie	.05	.03	-.08	.02	.13	.16

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 4.3: Gemiddelden per groep op de kwaliteitscriteria

Criteria	Studenten
00. Algemeen	0.68 (0.70) [2]
01. Geschikt	1.09 (0.93) [2]
02. Vergelijk	0.34 (1.30) [2]
03. Herhaalbaar	-0.43 (1.25) [2]
04. Transparant	0.59 (1.16) [2]
05. Acceptatie	1.05 (0.99) [2]
06. Authentiek	1.12 (0.82) [2]
07. Eerlijk	0.40 (1.10) [2]
08. Complex	0.49 (1.08) [2]
09. Betekenis	1.38 (1.15) [2]
10. Tijd/Kosten	0.35 (1.11) [2]
11. Selectie	0.57 (1.05) [2]
12. Leereffect	1.20 (1.09) [2]
13. Motivatie	0.66 (1.20) [2]

Noot: Standaarddeviaties zijn weergegeven tussen de haakjes. De subscripts in de kolom "Studenten" geven aan of het algemeen gemiddelde verschilt van 0.

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

5 STUDIE 1

Voor academiejaar 2017-2018 werd voor het laatste jaar van de Professionele Bachelors Kleuter en Lager Onderwijs (PBKO & PBLO) gekeken op welke manier de stagebeoordeling verbeterd kon worden. Hierbij werd enerzijds vertrokken van ons theoretisch kader en anderzijds vanuit de noden van de opleidingen zelf. In samenspraak met de opleidingen werd een interventie uitgewerkt en een experimenteel onderzoeksdesign dat toeliet de effecten van die interventie na te gaan. Het was de bedoeling om het effect na te gaan van het toevoegen van extra beoordelaars en beoordelingen aan de stagebeoordeling. Specifiek wilden we te weten komen hoe het toevoegen van extra beoordelaars en beoordelingen onthaald zou worden door studenten en begeleiders en wat daar de eventuele leereffecten voor studenten van zouden zijn. Dat deze interventie aan een reëel probleem tegemoetkomt, werd duidelijk in Voorstudie 2. Daar gaven studenten Verpleegkunde aan dat zij vinden dat de beslissingen die over hen worden genomen op onvoldoende assessmentsituaties en beoordelaars gebaseerd waren. Het is uiteraard niet omdat dit problematisch bevonden werd in het tweede jaar Verpleegkunde dat dit per definitie ook problematisch is binnen het derde jaar PBKO & PBLO, maar het onderstreept wel het feit dat onze interventie aansluit bij een reëel probleem binnen (sommige) opleidingen.

Op welke wijze sluit deze interventie aan bij het theoretisch kader? Deze interventie zet, net als D-PAC (zie Voorstudie 1), in op "Gebruik meer ratings en observaties". Specifiek werd hier geprobeerd (1) meer mensen te betrekken in het beoordelingsproces, (2) beoordelaars meer observaties te laten doen (= meer lesbezoeken), en (3) verschillende types beoordelaars te gebruiken. Qua onderliggende filosofie staat deze interventie echter vrij haaks op de manier waarop D-PAC gebruikt werd in Voorstudie 1. Waar D-PAC, zoals dit in Voorstudie 1 gebruikt werd, de criteria die de beoordelaar dient te hanteren bewust in het midden hield, is het hier net de bedoeling deze te expliciteren. Het is namelijk de bedoeling beoordelaars een zo rijk mogelijk oordeel te laten vormen en hen rond de tafel te zetten opdat de rijkdom van de verschillende oordelen optimaal tot zijn recht zou komen. Zijdelings zet deze interventie ook in op "Gebruik meer beoordelingsmethoden" en "Organiseer trainingen voor docenten en mentoren". Het is immers de bedoeling om additionele bronnen van kwalitatieve informatie aan te boren en om beoordelaars te ondersteunen in de voor hen nieuwe aspecten van deze taak, zoals actief deelnemen aan een groepsgesprek waarin de student beoordeeld wordt.

Omwille van de nadruk die tegenwoordig in de literatuur gelegd wordt op het procesmatige karakter van stages, werd bewaakt dat de interventie niet enkel een effect zou hebben op de kwaliteit van de beoordeling maar ook op de kwaliteit van het stageproces. Er werd met andere woorden voor gezorgd dat de interventie niet enkel zou inspelen op de stagebeoordeling zelf (= het summatief aspect) maar ook op het formatieve, procesmatige aspect. De extra beoordelaars werden daartoe niet alleen als beoordelaars maar ook als begeleiders ingezet. Zij kregen namelijk de instructie om hun bevindingen telkens ook in feedback om te zetten.

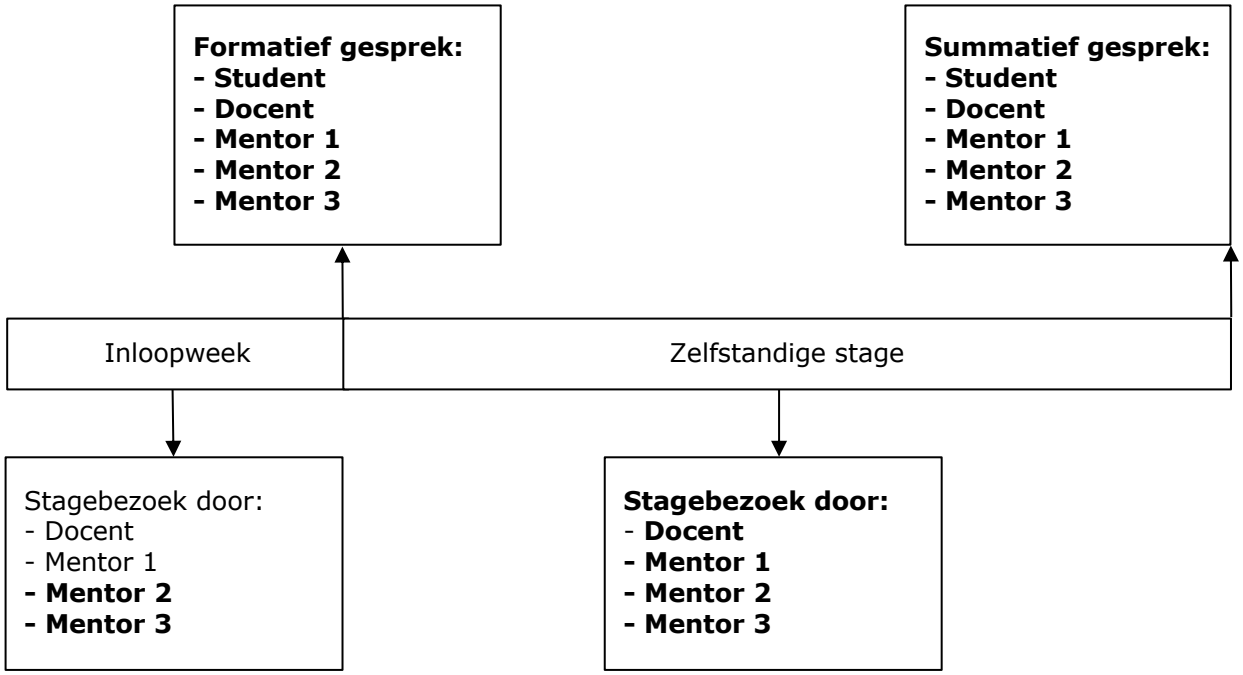
5.1 Methode

5.1.1 Procedure

De interventie werd uitgewerkt in het kader van het laatste luik van de stage in het laatste jaar: een stage van zes weken die typisch bestaat uit een inloopweek gevolgd door een zelfstandige

stage waarin de stagiair helemaal alleen een klas moet runnen. Het grootste deel van de studenten die deze stage doorliepen, doorliepen het "klassieke" traject. Deze studenten vormden de controlegroep binnen het experimentele onderzoeksdesign. De andere studenten doorliepen het "klassieke" traject met de aanpassingen die in het kader van dit onderzoek werden ingevoerd en vormden de experimentele groep. Het is niet onze bedoeling om in dit rapport het volledige verloop van het "klassieke" traject te beschrijven. Daarvoor verwijzen we door naar het stagevademecum van PBKO en PBLO die op te vragen vallen via e-mail. Het is wel onze bedoeling duidelijk te maken welke wijzigingen we in het kader van dit onderzoek hebben proberen doorvoeren. In dit geval gaat het om uitbreidingen. De uitbreidingen t.a.v. het "klassieke" traject worden weergegeven in Figuur 5.1 en hieronder beschreven.

Figuur 5.1: Het normale verloop van de stage met in het vet de uitbreidingen die in het kader van dit project werden ingevoerd in de experimentele conditie.



- Tijdens de inloopweek kregen de studenten uit de experimentele groep niet enkel een stagebezoek van de eigen klasmentor en van een docent maar ook van een tweede mentor (i.e., een leerkracht van een andere klas) en een derde mentor (= dit hoeft geen leerkracht te zijn; dit mag ook een zorgleerkracht of een directielid zijn). Het was niet de bedoeling dat hier verslagen van gemaakt werden en ook niet per se dat elk bezoek 50 minuten zou duren. Het kon ook gaan om flitsbezoeken van 10 a 15 minuten. De feedback wert nadien aan de student gegeven, die deze in zijn stageschrift opnam. Na de inloopweek volgde idealiter een groepsgesprek waarbij de student, de docent en alle mentoren aanwezig waren. Indien dit praktisch niet haalbaar bleek, overlegden de docent en mentoren voorafgaand aan het gesprek met de student (desnoods telefonisch). Het gesprek met de student werd dan aangegaan door wie beschikbaar was. Belangrijk

was wel dat de input van alle begeleiders hierin werd meegenomen. Deze stagebezoeken en het bijhorend gesprek waren formatief bedoeld, al werd hier ook beslist of de student klaar was om het zelfstandig deel van de stage aan te vatten.

- Tijdens de zelfstandige stage kregen de studenten uit de experimentele groep opnieuw bezoek van de docent en de drie mentoren. Ook deze bezoeken konden flitsbezoeken zijn. De feedback werd ook hier nadien aan de student gegeven, die deze in zijn stageschrift opnam. Na de zelfstandige stage was er idealiter een groepsgesprek waarbij de student en alle begeleiders aanwezig waren. Deze stagebezoeken en het bijhorend groepsgesprek waren summatief bedoeld. Tijdens dit gesprek werd o.a. bepaald of de student geslaagd is. Nadien maakte de docent de vertaalslag naar een cijfer. Indien het praktisch niet haalbaar bleek om een dergelijk groepsgesprek te organiseren, dan overlegden de begeleiders voorafgaand aan het gesprek met de student (desnoods telefonisch). Het gesprek met de student werd dan aangegaan door wie beschikbaar was. Belangrijk was ook hier dat de input van alle begeleiders werd meegenomen.

Na afloop van de stage kregen alle studenten en begeleiders een vragenlijst. Studenten en docenten kregen deze in de mate van het mogelijk op papier tijdens een leertrajectsbegeleidingsgesprek. Studenten en docenten die we op deze manier niet te pakken kregen, kregen een elektronische versie. Ook de mentoren kregen de vragenlijst elektronisch. De vragenlijst bestond uit drie of vier onderdelen, afhankelijk van de groep waarin iemand was ondergebracht (i.e., de controlegroep of de experimentele groep).

1. Eerst werden van iedereen een aantal persoonsgegevens (i.e. naam, voornaam en e-mailadres) en achtergrondkenmerken gevraagd (i.e., geslacht en leeftijd). Bij begeleiders werd additioneel gepolst naar hun ervaring (in termen van het aantal jaar dat ze reeds bij het begeleiden van stages betrokken zijn). Bij studenten werd additioneel gepolst naar vooropleiding en etniciteit. Aan studenten werd ook gevraagd wat volgens hen het punt is dat ze verdienen op hun stage (= een self-assessment). Tot slot werd aan studenten ook gevraagd van wie ze lesbezoek kregen en hoeveel keer.
2. Bij studenten en begeleiders uit de experimentele groep werd vervolgens gepolst naar de perceptie van de specifiek aangebrachte wijziging. Specifiek peilden de gestelde vragen naar de gepercipieerde bijdrage van deze interventie aan de kwaliteit van de algemene beoordelingsprocedure. Dit vormde een "zachte" test van onze interventie: werd de interventie op zich goed onthaald door de betrokkenen?
3. Daarna werd bij alle deelnemers gepolst naar de perceptie van de algemene beoordelingsprocedure aan de hand van de kwaliteitscriteria uit het model van Baartman. Op die manier kon de perceptie van de experimentele groep en de controlegroep vergeleken worden. Dit vormde een "harde" test van onze interventie: als die als een verbetering kon worden aanzien, dan zou de experimentele groep minstens op sommige criteria uit het model van Baartman beter moeten scoren dan de controlegroep.
4. Daarna werd bij alle deelnemers de competentiegroeimeter voor stage afgenomen ter uitdieping van het Leereffectcriterium. De competentiegroeimeter is een binnen KdG veelgebruikt instrument, gebaseerd op het werk van Braun et al. (2008; Braun & Leidner, 2009). De competentiegroeimeter laat enerzijds toe te identificeren welke competentiegroei begeleiders verwachten en anderzijds of dat zich vertaalt in de beleving van

studenten. Voor ons onderzoek betekent dit dat, als we verschillen zouden vinden tussen de scores van begeleiders uit de experimentele groep en de controlegroep, begeleiders verwachtten dat studenten andere competenties zouden ontwikkelen omwille van (het procesmatige karakter van) onze interventie. Of de competentiegroei dan ook effectief per groep verschilde, kon dan bij studenten getoetst worden.

5.1.2 Steekproef

Van de 184 ingeschreven studenten werden er 140 uitgenodigd om aan ons onderzoek deel te nemen. Studenten die op internationale stage gingen (N = 27) werden niet uitgenodigd. Hun stage duurde langer en daarop wachten was niet haalbaar in het licht van de timing van ons project. Ook studenten die stopten vóór de zelfstandige stage of niet aan deze stage mochten beginnen (N = 17) werden niet uitgenodigd. Van de 140 uitgenodigde studenten namen er 112 deel aan ons onderzoek (= 80.0%). Daarnaast werden ook 178 begeleiders uitgenodigd (31 docenten en 147 mentoren). Hiervan vulden er 93 (= 52.2%) de vragenlijst in: 29 docenten en 64 mentoren. De gemiddelde leeftijd van de deelnemende begeleiders was 40.42 jaar (SD = 10.19). Gemiddeld genomen hadden ze 12.48 jaar ervaring met stagebegeleiding (SD = 8.94). Het overgrote deel waren vrouwen (87.1%; met het oog op de analyses die nog volgen, werd deze variabele als volgt gecodeerd: 1 voor man en 2 voor vrouw). De studenten bestonden voor 92.4% uit vrouwen. Hun gemiddelde leeftijd was 22.69 jaar (SD = 3.29). Qua vooropleiding kwamen 10.3% van de studenten uit het BSO, 52.7% uit het TSO, 3.8% uit het KSO en 30.4% uit het ASO. Op basis van deze gegevens werd een variabele "educatie" geconstrueerd waarop studenten een 1 kregen als ze uit het BSO kwamen, een 2 als ze uit het TSO of het KSO kwamen of als hier geen duidelijke gegevens over te vinden waren, en een 3 als ze uit het ASO kwamen (zie Voorstudie 1). De restcategorie (= 2) vormt de baseline waartegen effecten van een lage (= BSO) en een hoge educatie (= ASO) worden afgezet. De meeste studenten hebben de Belgische nationaliteit (98.4%), zijn in België geboren (97.3%), en hebben een grootmoeder langs moederszijde die in België geboren is (93.5%). Op basis van deze gegevens werd een variabele "etniciteit" geconstrueerd waarop studenten een 0 kregen als ze de Belgische nationaliteit hebben, in België geboren zijn en een grootmoeder van Belgische nationaliteit hebben en een 1 als aan één van deze voorwaarden niet is voldaan (zie Voorstudie 1). Op deze manier werd 7.1% van de studenten gecategoriseerd als van "een andere etniciteit". De studenten vonden gemiddeld genomen dat ze 13.34 op 20 verdienden op hun stage (SD = 1.47). Geen enkele student gaf aan dat hij niet verdiende te slagen. Het gros van de studenten (69.6%) vergat zijn mentor te vernoemen als persoon die lesbezoeken deed. Op basis van de antwoorden van studenten die hun mentor wel in rekening brachten, bleek dat de invulling die mentoren aan hun rol geven erg verschillend is. Sommigen beperkten zich tot 1 lesbezoek, anderen sprongen meerdere keren per week of dagelijks binnen, en nog anderen waren constant aanwezig. Gemiddeld genomen kregen studenten lesbezoek van 2.00 personen (SD = 1.06; de mentor niet meegerekend), goed voor een gemiddelde van 3.38 lesbezoeken (SD = 3.19).

Voor studenten die niet deelnamen aan het onderzoek, werden de achtergrondgegevens uit Bamaflex gehaald. De studenten die niet deelnamen en de studenten die wel deelnamen verschilden niet significant in termen van geslacht (Mean = 1.93 en 1.92; (F (1, 182) = 0.07, ns), leeftijd (Mean = 22.88 en 22.57; (F (1, 182) = 0.37, ns), educatie (Mean = 2.26 en 2.16; (F (1, 182) = 1.27, ns) en etniciteit (Mean = 0.06 en 0.08; (F (1, 182) = 0.41, ns). Er is dus geen

aanwijzing dat de studenten die deelnamen aan het onderzoek niet representatief zouden geweest zijn voor de volledige studentenpopulatie.

De stageplaatsen werden op voorhand hetzij in de experimentele conditie hetzij in de controleconditie ondergebracht. Op die manier belandden 32 studenten (= 28.6%) en 29 begeleiders (= 31.5%) in de experimentele groep en 80 studenten (= 71.4%) en 63 begeleiders (= 68.5%) in de controlegroep. Het valt op dat de groepen niet even groot waren. Dit is te wijten aan het feit dat de stageplaatsen niet at random aan een conditie werden toegekend maar op basis van de bereidwilligheid van de stagescholen. Omdat deze werkwijze de at random verdeling in gevaar bracht, was het noodzakelijk te onderzoeken of de groepen wel vergelijkbaar zijn. Alvorens dat te doen, werd er echter eerst gekeken naar de antwoorden op de vraag naar wie op lesbezoek kwam. Die vraag, die niet door alle studenten werd beantwoord, liet ons toe te checken of de scholen zich aan ons design hielden. In de controlegroep rapporteerden de meeste studenten (60.8%) dat er, naast de mentor, slechts één persoon op lesbezoek kwam (= de docent). Daarnaast rapporteerde 36.5% lesbezoeken door twee extra personen. Die tweede persoon was ofwel een mentor-coach (in het kader van een ander lopend project) of iemand die op eigen initiatief van de school ingeschakeld werd als extra begeleider. Eén student rapporteerde drie extra bezoekers en één student rapporteerde er vier. In de experimentele groep rapporteerden 3 studenten slechts twee extra bezoekers te hebben gezien. Het merendeel van de studenten rapporteerde drie (50%) of vier (40.6%) extra bezoekers. Om overlap tussen beide condities te vermijden, werden de twee studenten die aangaven meer dan twee extra bezoekers te hebben gezien uit de controlegroep verwijderd en werden de drie studenten die aangaven minder dan drie extra bezoekers te hebben gezien uit de experimentele conditie verwijderd. Door deze ingreep werd de controlegroep beperkt tot studenten die aangaven van maximaal twee extra mensen lesbezoeken te hebben kregen en de experimentele groep beperkt tot studenten die aangaven van minimaal drie extra mensen lesbezoeken te hebben gekregen. Hierbij aansluitend werden ook twee mentoren verwijderd (één uit elke groep) omdat die de mentor waren van één van de studenten die we uit onze steekproef verwijderden. De studenten in de controlegroep rapporteerden significant minder lesbezoeken dan de studenten in de experimentele groep (Mean = 2.01 en 6.48; ($F(1, 99) = 65.48$, $p < .001$).

Om te controleren of de studenten uit de controleconditie vergelijkbaar waren met die uit de experimentele conditie, voerden we een reeks Oneway Anova's uit. Eerst onderzochten we of beide groepen vergelijkbaar waren in termen van achtergrondkenmerken. De groepen verschilden niet significant van elkaar in termen van geslacht (Mean = 1.94 en 1.86; ($F(1, 105) = 1.49$, ns), leeftijd (Mean = 22.38 en 23.03; ($F(1, 105) = 0.93$, ns), en educatie (Mean = 2.13 en 2.34; ($F(1, 105) = 2.54$, ns) maar wel in termen van etniciteit (Mean = 0.04 en 0.17; ($F(1, 105) = 5.67$, $p < .05$). Vervolgens onderzochten we of de groepen vergelijkbaar waren in termen van het punt dat ze volgens zichzelf verdienden op hun stage. De groepen verschilden niet significant van elkaar (Mean = 13.32 en 13.28; ($F(1, 105) = 0.02$, ns). Tot slot werd ook gekeken of de rol van de mentor in beide condities vergelijkbaar was. De groepen verschilden ook op dit vlak niet significant van elkaar (Mean = 6.18 en 4.63; ($F(1, 31) = 0.49$, ns). De at random verdeling over de twee condities is dus niet 100% geslaagd. De studenten in de experimentele conditie waren significant vaker van een andere etniciteit. Problematisch was dit echter niet, want latere analyses toonden aan dat etniciteit niet significant samenhing met de afhankelijke variabelen. Ook voor begeleiders onderzochten we of de groepen vergelijkbaar waren

qua achtergrondkenmerken. De groepen verschilden niet significant in termen van geslacht (Mean = 1.86 en 1.89; (F (1, 89) = 0.21, ns), leeftijd (Mean = 40.85 en 39.32; (F (1, 88) = 0.43, ns) en ervaring (Mean = 12.85 en 11.36; (F (1, 88) = 0.53, ns). Voor begeleiders is er dus geen aanwijzing dat de assumptie van een at random verdelen geschaad zou zijn.

5.1.3 Effectmetingen

5.1.3.1 Kwaliteitscriteria voor de specifieke interventie

Na afloop van de stage werd bij alle studenten en begeleiders uit de experimentele conditie gepolst naar de perceptie van de wijziging die aangebracht werd in de experimentele conditie. Dit gebeurde aan de hand van de verkorte versie van de kwaliteitscriteria-vragenlijst (zie Voorstudie 2). Deze vragenlijst werd daartoe voorzien van een andere inleiding. Alle items werden ook geherformuleerd opdat ze zouden polsen naar de gepercipieerde bijdrage van de interventie aan de kwaliteit van de algemene beoordeling. In totaal beoogde deze verkorte versie 14 criteria te meten: (1) Geschiktheid, (2) Vergelijkbaarheid, (3) Herhaalbaarheid, (4) Transparantie, (5) Acceptatie, (6) Authenticiteit, (7) Eerlijkheid, (8) Complexiteit, (9) Betekenisvolheid, (10) Tijd & Kosten, (11) Selectie, (12) Leereffect, (13) Motivatie, en (14) Onderwijseffect. Het criterium Onderwijseffect werd enkel bevraagd bij begeleiders omdat dit criterium handelt over het effect dat de beoordeling op zich kan hebben op de manier waarop de stage in de toekomst vormgegeven kan worden (iets waar studenten niet mee bezig zijn). Elk criterium werd gemeten aan de hand van drie vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 3.

De Cronbach alfas en de correlaties tussen de schalen zijn te vinden in Tabel 5.1. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen heel wat schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Die kon 57% van de totale variantie verklaren. De enige schaal met een factorlading < .40 op deze factor was Motivatie. Alle andere schalen hadden een factorlading > .40. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze score had een Cronbach alfa van .93.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 5.2. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens. De resultaten voor begeleiders zijn te vinden in Tabel 5.3. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens.

5.1.3.2 Kwaliteitscriteria voor de algemene beoordeling

Na afloop van de stage werd bij alle studenten en begeleiders uit zowel de experimentele conditie als de controleconditie ook gepolst naar de perceptie van de algemene stagebeoordeling. Dit gebeurde eveneens aan de hand van een verkorte versie van de kwaliteitscriteria-vragenlijst (zie Voorstudie 2). In totaal beoogde deze verkorte versie 14 criteria te meten: (1) Geschiktheid, (2) Vergelijkbaarheid, (3) Herhaalbaarheid, (4) Transparantie, (5) Acceptatie, (6) Authenticiteit, (7) Eerlijkheid, (8) Complexiteit, (9) Betekenisvolheid, (10) Tijd & Kosten, (11) Selectie, (12) Leereffect, (13) Motivatie, en (14) Onderwijseffect. Het criterium Onderwijseffect werd enkel bevraagd bij begeleiders omdat dit criterium handelt over het effect dat de beoordeling op zich kan hebben op de manier waarop de stage in de toekomst vormgegeven kan worden (iets waar studenten niet mee bezig zijn). Elk criterium werd gemeten aan de hand van drie vragen. Alle vragen dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 4.

De Cronbach alfa's en de correlaties tussen de schalen zijn te vinden in Tabel 5.4. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders. Met uitzondering van Onderwijseffect hadden alle schalen een goede interne consistentie (Cronbach alfa > .70). De lagere Cronbach alfa voor Onderwijseffect valt deels te verklaren door het feit dat die berekend werd op een kleinere steekproef. Gezien de substantiële correlaties tussen de schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Deze oplossing kon 53% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40 op deze factor. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze had een Cronbach alfa van .92.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 5.5. Noch de algemene score noch de onderliggende schalen, met uitzondering van Motivatie, konden significant worden voorspeld op basis van de beschikbare gegevens. Motivatie werd positief voorspeld door het resultaat dat studenten vinden dat ze verdienen en negatief door vooropleiding. Het zijn dus vooral studenten met een lage vooropleiding die vonden dat ze het heel goed deden die erg gemotiveerd waren. De resultaten voor begeleiders zijn te vinden in Tabel 5.6. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens.

5.1.3.3 Uitdieping van het Leereffectcriterium

Na afloop van de stage werd bij alle studenten en begeleiders uit zowel de experimentele conditie als de controleconditie ter uitdieping van het Leereffectcriterium de competentiegroeimeter

voor stage afgenomen. De gebruikte versie beoogde competentiegroei te meten in 12 domeinen: (1) Praktijkgericht, (2) Zelfmanagement, (3) Zelfreflectie, (4) Doelgericht werken, (5) Kansen zien en nemen, (6) Leidinggeven en activeren, (7) Vernieuwen, (8) Samenwerken, (9) Feedback geven, (10) Feedback vragen en ontvangen, (11) Klantgericht werken, en (12) Voor een groep spreken. De competentiegroeimeter voor stage bestaat in twee versies: een studenten- en een docentenversie. De docentenversie, die bij alle begeleiders werd afgenomen, laat toe te identificeren welke competentiegroei begeleiders verwachten. In deze versie wordt elke concept gemeten aan de hand van één item. De studentenversie laat toe te identificeren welke competentiegroei studenten rapporteren. In deze versie wordt elke concept gemeten aan de hand van meerdere items (van 3 tot 7 al naargelang het concept). Alle vragen dienden te worden beantwoord op een 5-punten Likertschaal ("1 = helemaal oneens" tot "5 = helemaal eens"). De docentenversie is te vinden in Appendix 8 en de studentenversie in Appendix 9.

De Cronbach alfas en de correlaties tussen de schalen van de studentenversie zijn te vinden in Tabel 5.7. De correlaties tussen de items van de docentenversie zijn te vinden in Tabel 5.8. Behalve de schalen Praktijkgericht, Vernieuwen en Feedback vragen hadden alle schalen een goede interne consistentie (Cronbach alfa > .70). De wat lagere Cronbach alfa's voor Praktijkgericht, Vernieuwen en Feedback vragen vallen deels te verklaren doordat dit vrij brede concepten zijn. Gezien de substantiële correlaties tussen heel wat concepten, besloten we via factoranalyse te onderzoeken of het aantal schalen/items gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. Voor de studentenversie wees de scree plot van een exploratorische factoranalyse op een oplossing met één factor. Deze oplossing kon 52% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40 op deze factor. Omdat zo'n factor handig is om een overzicht te krijgen van de algemene competentiegroei, berekenden we een algemene competentiegroei-score door alle schalen te middelen. Deze algemene score had een Cronbach alfa van .91. Ook voor de docentenversie wees de scree plot op een oplossing met één factor. Deze oplossing kon 50% van de totale variantie verklaren. Het enige item met een factorlading < .40 was Praktijkgerichtheid. Alle andere items hadden een factorlading > .40 op deze factor. Deze algemene score had een Cronbach alfa van .89.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen / items kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 5.9. De algemene score kon deels worden voorspeld op basis van het punt dat studenten vonden dat ze verdienden. Hetzelfde geldt voor de helft van de onderliggende schalen. Voor de meeste andere schalen gaan de resultaten in dezelfde richting. Studenten die vonden dat ze meer punten verdienden rapporteerden dus ook meer competentiegroei. In sommige gevallen (met name voor Samenwerken en Feedback vragen) werd dit nog eens versterkt door een lagere vooropleiding. Het waren vooral studenten met een lagere vooropleiding die van zichzelf vonden dat ze het heel goed deden voor hun stage die meest competentiegroei rapporteerden (voor gelijkaardige bevindingen, zie Kruger & Dunning, 1999). De resultaten voor begeleiders zijn te vinden in Tabel 5.10. De algemene score kon significant worden voorspeld op basis van de beschikbare gegevens, al was dit niet duidelijk te wijten aan één bepaalde variabelen maar eerder aan de combinatie van variabelen. Hetzelfde geldt voor het gros van de

onderliggende schalen (met uitzondering van Praktijkgericht, Zelfmanagement, Vernieuwen en Voor een groep spreken). Het waren in eerste instantie de iets oudere docenten met meer ervaring die meer competentiegroei verwachtten bij studenten.

5.1.3.4 Uitdieping van het Selectiecriterium

De studenten haalden gemiddeld 12.87 op 20 op hun stage (SD = 2.23). In totaal slaagden 8 studenten niet voor hun stage (= 7.5%). Via regressieanalyse gingen we na of het behaalde punt kon worden voorspeld op basis van de achtergrondkenmerken (i.e., leeftijd, geslacht, educatie en etniciteit). De score die studenten kregen konden niet significant voorspeld worden op basis van deze gegevens (R^2 = 0.08, F (4, 102) = 2.09, ns). De score die studenten behaalden op hun stage was sterk gecorreleerd met het punt dat ze zeggen te verdienen voor hun stage (r = .57, p < .01), wat erop wijst dat studenten hun eigen functioneren behoorlijk goed konden inschatten. Of toch in de zin dat de rangorde die ze op die manier spontaan maakten behoorlijk goed overeenkwam met de rangorde die naar voor kwam uit de stagebeoordeling. Wel was het zo dat studenten het effectieve punt dat ze kregen minder goed konden inschatten. Een paired samples t-test toont aan dat er een significant verschil is tussen het punt dat studenten vonden dat ze verdienen en het punt dat ze uiteindelijke behaalden (Mean = 13.31 en 12.87, t = 2.54, p < .05). Studenten schatten hun eigen prestatie dus positiever in dan hun beoordelaars. De score die studenten behaalden op hun stage was ook significant positief gecorreleerd met de algemene score op de kwaliteitscriteria voor de algemene beoordeling (r = .33, p < .01) en met de algemene score voor competentiegroei (r = .30, p < .01). De correlaties tussen de eindscore en de individuele kwaliteitscriteria waren .36 voor Geschiktheid, .14 voor Vergelijkbaarheid, .28 voor Herhaalbaarheid, .16 voor Transparantie, .26 voor Acceptatie, .29 voor Authenticiteit, .20 voor Eerlijkheid, .00 voor Complexiteit, .30 voor Betekenisvolheid, .25 voor Tijd & Kosten, .29 voor Selectie, .23 voor Leereffect, en .27 voor Motivatie. De correlaties tussen de eindscore en de onderliggende schalen voor competentiegroei waren .22 voor Praktijkgericht, .24 voor Zelfmanagement, .25 voor Zelfreflectie, .19 voor Doelgericht werken, .22 voor Kansen zien en nemen, .12 voor Leidinggeven en activeren, .23 voor Vernieuwen, .17 voor Samenwerken, .10 voor Feedback geven, .29 voor Feedback vragen en ontvangen, .24 voor Klantgericht werken, en .33 voor Voor een groep spreken. Alle vernoemde correlaties met een absolute waarde groter dan .20 zijn significant op niveau p < .05 en alle vernoemde correlaties met een absolute waarde groter dan .25 zijn significant op niveau p < .01.

5.2 Verwachtingen

We verwachtten dat onze interventie positief onthaald zou worden, en dat, wanneer de betrokkenen gevraagd werden of de interventie een extra bijdrage levert aan de kwaliteitscriteria, ze daar instemmend op zouden reageren. Dit vormde echter slechts een "zachte" test van onze interventie. Wanneer onze interventie echt een verbetering vormde, zou dit zich ook moeten vertalen in meer tevredenheid over bepaalde aspecten van de algemene beoordeling in de experimentele groep dan in de controlegroep. Specifiek verwachtten we dat de algemene beoordeling in de experimentele conditie beter zou scoren op de criteria Vergelijkbaarheid (niet elke beoordelaar zal op dezelfde dingen letten, maar door meer beoordelaars in te zetten, wordt de kans groter dat er bij verschillende studenten op dezelfde dingen gelet wordt), Herhaalbaarheid (de beslissingen die worden genomen over studenten zijn gebaseerd op meerdere assessmentsituaties en meerdere beoordelaars), Eerlijkheid (door meer bezoekers in te schakelen en meer

bezoeken in te lassen krijgen studenten een eerlijkere kans hun competenties te demonstreren), Betekenisvolheid (door meer bezoekers in te schakelen en meer bezoeken in te lassen krijgen studenten extra kansen op betekenisvolle feedback), Zelfsturing (de grotere kans op betekenisvolle feedback zou de zelfreflectie moeten kunnen stimuleren), Selectie (verschillende beoordelaars laten overleggen met elkaar zou de selectie van goede en slechte studenten ten goede moeten komen), en Leereffect (de grotere kans op betekenisvolle feedback zou de ontwikkeling van studenten moeten bevorderen) dan in de controleconditie. Hierbij dient te worden opgemerkt dat dit wel een "hoge" verwachting is. De ingrepen die werden doorgevoerd in het kader van onze interventie vormden immers maar een deel van het beoordelingsproces. Hun effect zou dus al erg sterk moeten zijn vooraleer het zich zo zou manifesteren.

Wat het leereffect betreft, verwachtten we dat de interventie een positieve bijdrage zou leveren aan competenties zoals Zelfreflectie en Feedback vragen. In de experimentele conditie waren er namelijk meer mensen waar men feedback aan kon vragen, en de extra feedback die studenten konden krijgen, zou kunnen leiden tot een verbeterde zelfreflectie.

Wat selectie betreft, verwachtten we dat de vernieuwde procedure de selectie van goede en slechte studenten ten goede zou komen. Dit zou zich dan moeten vertalen in de punten die studenten kregen op hun stage. In de experimentele conditie verwachtten we een grotere spreiding van de punten dan in de controleconditie.

5.3 Resultaten

5.3.1 Kwaliteitscriteria voor de specifieke interventie

Voor de deelnemers uit de experimentele conditie zijn de gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor de specifieke interventie te vinden in Tabel 5.11 en dit per rol (i.e., Student en Begeleider). T-tests wezen uit dat het algemeen gemiddelde voor alle deelnemers significant verschilt van 0 (= het neutraal middelpunt van de schaal) in positieve richting, en dit voor alle criteria. Algemeen genomen waren de deelnemers het er dus mee eens dat de interventie bijdroeg aan de kwaliteit van de algemene beoordeling.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren van elkaar verschilden betreffende hun perceptie van de interventie. Uit deze analyses bleek dat begeleiders van mening waren dat de interventie een grotere bijdrage leverde aan de kwaliteit van de algemene beoordelingsprocedure dan studenten voor wat betreft het criterium Herhaalbaarheid en Eerlijkheid. Voor wat Herhaalbaarheid betreft, week het oordeel van studenten niet significant af van 0. Begeleiders stonden ook positiever t.a.v. de haalbaarheid van de door de interventie gevergde Tijd & Kosten dan studenten. Dit is ietwat merkaardig omdat het de begeleiders waren die in het kader van deze interventie een extra inspanning dienden te leveren.

5.3.2 Kwaliteitscriteria voor de algemene beoordeling

De gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor de algemene beoordeling zijn te vinden in Tabel 5.12, en dit zowel per conditie als per rol. Zowel in de controleconditie als in de experimentele conditie wezen t-tests uit dat het algemeen gemiddelde voor alle deelnemers significant verschilt van 0 (= het neutraal middelpunt van de schaal) in positieve richting. De enige uitzonderingen op deze regel betroffen het criterium Herhaalbaarheid (in beide

condities) en het criterium Complexiteit (in de experimentele conditie). Voor deze criteria bleek het algemeen gemiddelde voor alle deelnemers niet significant van 0 te verschillen. Het criterium Herhaalbaarheid werd dus in beide condities neutraal beoordeeld. Het criterium Complexiteit werd enkel in de experimentele conditie neutraal beoordeeld.

Het gemiddelde van de deelnemers in de controleconditie verschilde enkel significant van dat van de deelnemers in de controleconditie voor Complexiteit. De deelnemers in de controlegroep vonden de beoordeling meer geschikt om de denkprocessen van een student te beoordelen als die aan een opdracht werkt dan de deelnemers in de experimentele groep. Verder waren er geen significante verschillen tussen beide condities in de beoordeling van de criteria.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren (i.e., Studenten en Begeleiders) binnen de condities van elkaar verschilden betreffende hun perceptie van de algemene beoordeling. In de controleconditie bleek uit deze analyses dat begeleiders de beoordeling significant positiever inschatten dan studenten voor Vergelijkbaarheid, Eerlijkheid, Tijd en Kosten, en Selectie. Studenten schatten de beoordeling dan weer als significant positiever in voor Betekenisvolheid en Motivatie. De experimentele conditie bevestigde alleen de sterkste verschillen, met name de positievere inschatting door begeleiders voor Vergelijkbaarheid en Eerlijkheid. Voor Herhaalbaarheid weken in beide condities noch de oordelen van studenten noch die van begeleiders significant af van 0. In de experimentele conditie week ook het oordeel van studenten Voor Vergelijkbaarheid en Complexiteit niet significant af van 0. Het oordeel van begeleiders week dan weer niet af van 0 voor Motivatie.

5.3.3 Uitdieping van het Leereffectcriterium

De gemiddelden en standaarddeviaties op de competenties uit de competentiegroeimeter zijn te vinden in Tabel 5.13, en dit zowel per conditie als per rol. In beide condities wezen t-tests uit dat zowel de algemene score als alle deelscores in positieve richting significant verschilden van 3 (= het neutraal middelpunt van de schaal), en dit zowel bij begeleiders als bij studenten. In beide condities verwachtten begeleiders competentiegroei in elk van de domeinen en de studenten rapporteerden dit ook. Een uitzondering is Klantgerichtheid in de experimentele conditie. Begeleiders gaven aan dat ze op dat vlak niet echt competentiegroei verwachtten, maar studenten gaven aan dat ze toch ook op dat vlak competentiegroei ervaarden. De competentiegroei die begeleiders verwachtten was niet significant anders in de experimentele conditie dan in de controleconditie en ook de door studenten gerapporteerde competentiegroei verschilde niet over condities heen.

In de controleconditie wezen t-tests uit dat de competentiegroei die begeleiders verwachtten op de meeste vlakken significant afweek van de competentiegroei die studenten rapporteerden. Begeleiders hadden de neiging meer competentiegroei te verwachten dan studenten rapporteerden. Dat geldt voor alle competenties, behalve Zelfreflectie, Kansen zien, Vernieuwen, Feedback vragen en Klantgericht werken. De uitzondering werd gevormd door Voor een groep spreken. Op dit vlak rapporteerden studenten meer competentiegroei dan begeleiders verwachtten. In de experimentele conditie leken de verwachtingen van de begeleiders en wat studenten aangaven minder ver uit elkaar te liggen, al kan dit ook te wijten zijn aan het feit dat de experimentele groep kleiner was, en dat het dus moeilijker is om hier significante verschillen

vast te stellen. In deze conditie zagen we enkel dat begeleiders meer competentiegroei verwachtten dan studenten rapporteerden op vlak van Praktijkgerichtheid en Zelfmanagement. Ook hier rapporteerden studenten meer competentiegroei op Voor een groep spreken dan begeleiders verwachtten. En eenzelfde patroon zien we hier ook terug voor Klantgericht werken. Ook hier rapporteerden studenten onverwacht veel competentiegroei.

5.3.4 Uitdieping van het Selectiecriterium

Teneinde na te gaan of de interventie een effect had op de selectie van studenten werden de punten die studenten behaalden op hun stage in de experimentele conditie vergeleken met de punten die studenten behaalden op hun stage in de controleconditie. Hierbij werd zowel gekeken naar de gemiddelde score als naar de spreiding van de punten (= de standaarddeviatie). Een Oneway Anova wees uit dat de controlegroep niet significant verschilde van de experimentele groep in termen van gemiddelde score (Mean = 12.92 en 12.72; F (1, 105) = 0.17, ns). Een Levene's test (Levene, 1960) wees uit dat de standaarddeviatie evenmin significant verschilde over beide groepen (SD = 2.37 en 1.85; F (1, 105) = 1.05, *ns*).

5.4 Conclusie

De "zachte" tests, waarbij we enkel binnen de experimentele groep gingen polsen naar hoe de interventie werd ervaren, leverden positieve resultaten op. Waar er specifiek gepolst werd naar de bijdrage van de interventie aan de algemene beoordeling, zien we dat het algemeen gemiddelde voor alle deelnemers in positieve richting verschilt van 0, en dit voor alle criteria. Alle betrokkenen waren het er dus gemiddeld genomen over eens dat de interventie een positieve bijdrage leverde aan de stagebeoordeling. Begeleiders schatten de bijdrage van de interventie positiever in dan studenten qua Herhaalbaarheid en Eerlijkheid. Begeleiders stonden ook positiever ten aanzien van de haalbaarheid van de door de interventie gevergde Tijd & Kosten dan studenten. Dit is enigszins merkaardig omdat het de begeleiders waren die in het kader van deze interventie een extra inspanning dienden te leveren.

De "harde" tests, waarbij we op experimentele wijze twee groepen met elkaar vergeleken, nuanceerden dit beeld. Met deze tests vonden we immers zo goed als geen effecten van de interventie. Zowel in de controlegroep als in de experimentele groep stonden de deelnemers over het algemeen positief ten aanzien van de algemene beoordeling. De enige uitzonderingen op deze regel betroffen Herhaalbaarheid (in beide groepen) en Complexiteit (in de experimentele groep). In deze uitzonderlijke gevallen werd de beoordeling als neutraal ervaren. Belangrijker is dat de experimentele groep nagenoeg niet verschilde van de controlegroep in termen van de scores op de kwaliteitscriteria. We vonden enkel een significant verschil tussen beide groepen voor wat Complexiteit betreft. In de controleconditie werd de beoordeling als beter geschikt gezien om de cognitieve vaardigheden van studenten bloot te leggen. Dit is een contra-intuïtief resultaat waar we ook niet direct een verklaring voor hebben. In de controlegroep schatten begeleiders de beoordeling significant positiever in dan studenten voor Vergelijkbaarheid, Eerlijkheid, Tijd en Kosten, en Selectie. Studenten schatten de beoordeling dan weer positiever in voor Betekenisvolheid en Motivatie. De experimentele conditie bevestigde de positievere inschatting door begeleiders voor Vergelijkbaarheid en Eerlijkheid.

Ook wanneer er ingezoomd wordt op de criteria Leereffect of Selectie vonden we geen effecten. Voor wat Selectie betreft, waren er geen significante verschillen tussen de controleconditie en

de experimentele conditie, noch in termen van gemiddelde noch in termen van spreiding (= standaarddeviatie). Op basis hiervan dienen we te concluderen dat er geen aanwijzing is voor het feit dat de interventie bijdroeg aan een beter onderscheid maken tussen goede en slechte studenten. Voor Leereffect wezen de resultaten uit dat noch de verwachtte competentiegroei noch de gerapporteerde competentiegroei verschilde over condities heen. In de controleconditie verwachtten begeleiders competentiegroei in elk domein en de studenten rapporteerden dit ook. Eenzelfde patroon vonden we in de experimentele conditie, behalve voor Klantgericht werken. Begeleiders verwachtten op dit domein geen competentiegroei, maar studenten rapporteerden die wel. Op een aantal uitzonderingen na verwachtten begeleiders te veel competentiegroei, in de zin dat hun scores significant hoger lagen dan die van studenten. De grote uitzondering werd gevormd door de dimensie Voor een groep spreken. Op dit vlak rapporteerden studenten meer competentiegroei dan begeleiders verwachtten. Binnen de experimentele conditie lagen de verwachtte en de gerapporteerde competentiegroei dichter bij elkaar, in de zin dat er minder significante verschillen waren tussen studenten en begeleiders. Dit valt allicht te verklaren door de kleinere steekproefgrootte. Enkel voor Praktijkgerichtheid en Zelfmanagement verwachtten begeleiders meer competentiegroei dan dat studenten rapporteedren. Ook hier rapporteerden studenten meer competentiegroei voor Voor een groep spreken dan begeleiders verwachtten. Hetzelfde geldt voor Klantgericht werken. Hierbij dient wel te worden opgemerkt dat het niet zo verwonderlijk is dat de interventie geen extra competentiegroei teweegbracht. Los van het feit dat er in deze studie niet expliciet werd ingezet op (andere) competenties, bleek het ook zo dat ook in de controleconditie alle competenties gemiddeld genomen een zeer hoge groeiscore vertoonden (> 4 op een schaal waar 5 het maximum is) en dit zowel bij begeleiders als bij studenten. De verbetermogelijkheden waren dan ook beperkt.

Dat de "harde" tests weinig resultaat opleverden, is op zich niet zo verwonderlijk. Ten eerste was dit een erg hoge verwachting. De ingrepen die werden doorgevoerd in het kader van onze interventie vormden maar een deel van het beoordelingsproces. Hun effect moest dus al erg sterk zijn vooraleer het zich op die manier zou manifesteren. Ten tweede is het ook zo dat we niet 100% zeker zijn dat de interventies zijn doorgevoerd zoals geconcipieerd. In de controleconditie werd de gang van zaken enigszins verstoord doordat er in sommige scholen mentorcoaches of andere personeelsleden bij de begeleiding betrokken werden. In de experimentele conditie werd die dan weer verstoord doordat er vaak minder extra begeleiders werden ingezet dan bedoeld. We probeerden dit op te vangen door een aantal studenten (en hun begeleiders) te verwijderen uit hun respectievelijke conditie, maar dan nog blijft het zo dat de gang van zaken op de werkvloer de condities meer gelijkend heeft gemaakt dan de bedoeling was, waardoor het moeilijker wordt om verschillen tussen de condities te vinden. We weten ook enkel hoeveel begeleiders er werden ingeschakeld. Naar de manier waarop mentoren hun rol invulden, hebben we het raden. Slechts 30% van de studenten gaf hier een indicatie voor. En hoewel we op basis van die indicatie geen significant verschil vonden over condities heen in de manier waarop mentoren hun rol invulden, ondermijnt het feit dat slechts een kleine groep studenten ons hier informatie over heeft verstrekt de zekerheid dat er geen significante verschillen waren. Ook naar de verdere uitvoering van onze interventie hebben we het raden. We kunnen er alleen maar van uitgaan dat er effectief een formatief en een summatief gesprek is gehouden met alle betrokkenen, of, wanneer een dergelijk gesprek toch individueel met de student zou zijn gevoerd, de begeleiders wel degelijk voorafgaandelijk contact hebben gehad, op die manier hun bevindingen hebben uitgewisseld, en die bevindingen hebben meegenomen in dat gesprek. Ten

derde ging het hier ook om relatief kleine groepen. Dat maakt het eveneens moeilijk om significante effecten te vinden. Vooral de experimentele groep viel wat klein uit. Dat was niet de bedoeling. We hadden immers een 50-50 verdeling voor ogen. Dit bleek echter praktisch niet haalbaar te zijn. Ten vierde is het ook mogelijk dat we geen verschillen vonden omdat alles erop wijst dat de betrokkenen sowieso reeds vrij tevreden waren over de gang van zaken. De algemene beoordeling werd ook in de controleconditie positief bevonden, met een gemiddelde score van 1.30 op een schaal met een maximum van 3. Ook op dit vlak kan dus betoogd worden dat de ruimte om verbeteringen aan te brengen eerder beperkt was. De algemene conclusie die uit deze gegevens dient te worden getrokken, is wel dat de interventie in de ogen van de betrokkenen een meerwaarde leek te bieden voor de stagebeoordeling. Die meerwaarde lijkt wel beperkt want ze kwam niet naar voren uit de vergelijking tussen de experimentele groep en de controlegroep.

Tabel 5.1: Cronbach alfa's van en correlaties tussen de kwaliteitscriteria voor de specifieke interventie

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.89													
02. Vergelijk	.89	.60 [2]												
03. Herhaalbaar	.70	.41 [2]	.64 [2]											
04. Transparant	.93	.53 [2]	.70 [2]	.62 [2]										
05. Acceptatie	.88	.57 [2]	.70 [2]	.55 [2]	.53 [2]									
06. Authentiek	.91	.38 [2]	.48 [2]	.31 [1]	.50 [2]	.55 [2]								
07. Eerlijk	.90	.50 [2]	.74 [2]	.55 [2]	.53 [2]	.77 [2]	.49 [2]							
08. Complex	.90	.36 [2]	.57 [2]	.49 [2]	.63 [2]	.54 [2]	.57 [2]	.58 [2]						
09. Betekenis	.96	.42 [2]	.41 [2]	.32 [1]	.23	.51 [2]	.60 [2]	.42 [2]	.47 [2]					
10. Tijd/Kosten	.97	.56 [2]	.71 [2]	.41 [2]	.45 [2]	.76 [2]	.64 [2]	.77 [2]	.45 [2]	.56 [2]				
11. Selectie	.89	.56 [2]	.72 [2]	.42 [2]	.49 [2]	.71 [2]	.58 [2]	.76 [2]	.50 [2]	.51 [2]	.75 [2]			
12. Leereffect	.91	.58 [2]	.44 [2]	.39 [2]	.38 [2]	.59 [2]	.55 [2]	.56 [2]	.53 [2]	.73 [2]	.63 [2]	.58 [2]		
13. Motivatie	.95	.14	.36 [2]	.39 [2]	.33 [1]	.38 [2]	.52 [2]	.36 [2]	.59 [2]	.38 [2]	.37 [2]	.38 [2]	.34 [1]	
14. Onderwijs	.77	.66 [2]	.69 [2]	.62 [2]	.69 [2]	.65 [2]	.44 [1]	.73 [2]	.67 [2]	.61 [2]	.58 [2]	.77 [2]	.75 [2]	.32

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 5.2: Regressie bij studenten op de kwaliteitscriteria voor de specifieke interventie

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.15	.10	-.06	-.15	.13	.33
01. Geschikt	.11	.20	-.11	.08	-.11	.12
02. Vergelijk	.31	.14	-.19	-.29	.14	.43 [1]
03. Herhaalbaar	.32	.21	.09	-.24	.40 [1]	.35
04. Transparant	.18	.02	-.04	-.24	.20	.36
05. Acceptatie	.15	.16	-.03	.08	.19	.27
06. Authentiek	.03	-.08	-.03	-.05	.12	.01
07. Eerlijk	.19	.02	.04	-.21	.27	.34
08. Complex	.17	-.01	.04	-.20	.08	.42
09. Betekenis	.17	.28	-.09	-.07	-.23	.08
10. Tijd/Kosten	.07	.08	-.02	-.17	-.04	.22
11. Selectie	.08	-.01	-.11	-.04	-.04	.25
12. Leereffect	.08	.17	-.17	-.02	-.10	-.06
13. Motivatie	.15	.02	-.06	-.09	.16	.37

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.3: Regressie bij begeleiders op de kwaliteitscriteria voor de specifieke interventie

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.08	-.17	-.08	.24
01. Geschikt	.02	.00	-.21	.26
02. Vergelijk	.11	.22	-.23	.45
03. Herhaalbaar	.08	-.01	.09	.20
04. Transparant	.10	.12	.06	.28
05. Acceptatie	.11	-.26	.03	.13
06. Authentiek	.07	-.15	.02	.16
07. Eerlijk	.02	-.01	.18	-.08
08. Complex	.10	-.18	-.13	.31
09. Betekenis	.04	-.16	-.19	.17
10. Tijd/Kosten	.10	-.29	.23	-.15
11. Selectie	.09	-.26	-.19	.16
12. Leereffect	.11	-.32	.22	-.17
13. Motivatie	.09	-.29	-.06	.00
14. Onderwijs	.01	-.01	-.03	.11

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.4: Cronbach alfa's van en correlaties tussen de kwaliteitscriteria voor de algemene beoordeling

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.81													
02. Vergelijk	.88	.38 [2]												
03. Herhaalbaar	.84	.32 [2]	.52 [2]											
04. Transparant	.82	.52 [2]	.44 [2]	.48 [2]										
05. Acceptatie	.83	.61 [2]	.51 [2]	.48 [2]	.68 [2]									
06. Authentiek	.86	.65 [2]	.38 [2]	.38 [2]	.52 [2]	.65 [2]								
07. Eerlijk	.89	.44 [2]	.57 [2]	.53 [2]	.49 [2]	.61 [2]	.58 [2]							
08. Complex	.83	.36 [2]	.47 [2]	.53 [2]	.48 [2]	.49 [2]	.46 [2]	.55 [2]						
09. Betekenis	.92	.38 [2]	.25 [2]	.30 [2]	.37 [2]	.44 [2]	.42 [2]	.40 [2]	.45 [2]					
10. Tijd/Kosten	.89	.56 [2]	.50 [2]	.50 [2]	.50 [2]	.68 [2]	.60 [2]	.74 [2]	.53 [2]	.52 [2]				
11. Selectie	.81	.49 [2]	.46 [2]	.43 [2]	.48 [2]	.58 [2]	.57 [2]	.69 [2]	.39 [2]	.43 [2]	.72 [2]			
12. Leereffect	.81	.56 [2]	.33 [2]	.33 [2]	.56 [2]	.56 [2]	.57 [2]	.54 [2]	.46 [2]	.61 [2]	.60 [2]	.62 [2]		
13. Motivatie	.92	.32 [2]	.16 [1]	.32 [2]	.36 [2]	.38 [2]	.42 [2]	.32 [2]	.45 [2]	.42 [2]	.41 [2]	.40 [2]	.49 [2]	
14. Onderwijs	.63	.40 [2]	.26 [1]	.39 [2]	.60 [2]	.43 [2]	.54 [2]	.48 [2]	.45 [2]	.40 [2]	.53 [2]	.63 [2]	.48 [2]	.55 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 5.5: Regressie bij studenten op de kwaliteitscriteria voor de algemene beoordeling

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.09	-.04	-.11	-.21 [1]	.03	.19
01. Geschikt	.10	.00	.12	-.15	.09	.20 [1]
02. Vergelijk	.06	.06	-.11	-.15	.09	.12
03. Herhaalbaar	.10	-.15	-.13	-.20 [1]	.02	.18
04. Transparant	.10	-.25 [1]	-.03	-.13	-.05	.11
05. Acceptatie	.08	-.10	-.10	-.18	-.02	.18
06. Authentiek	.07	.04	-.01	-.16	.05	.19
07. Eerlijk	.05	-.06	-.10	-.19	-.05	.04
08. Complex	.10	.09	-.05	-.21 [1]	.04	.17
09. Betekenis	.07	.15	-.08	-.14	-.06	.13
10. Tijd/Kosten	.03	.02	-.06	-.14	.02	.08
11. Selectie	.05	-.06	-.21 [1]	-.06	.07	.04
12. Leereffect	.03	-.02	-.10	-.06	.01	.14
13. Motivatie	.12 [1]	-.03	-.06	-.22 [1]	.03	.25 [1]

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat studenten vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.6: Regressie bij begeleiders op de kwaliteitscriteria voor de algemene beoordeling

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.04	-.03	.16	.03
01. Geschikt	.07	-.01	.05	.22
02. Vergelijk	.02	-.12	.02	.07
03. Herhaalbaar	.04	-.05	.16	.05
04. Transparant	.07	.01	.20	.07
05. Acceptatie	.01	-.08	-.08	.10
06. Authentiek	.01	.02	-.10	.18
07. Eerlijk	.02	-.01	.18	-.08
08. Complex	.03	-.04	-.20	.27
09. Betekenis	.07	-.00	.40 [1]	-.31
10. Tijd/Kosten	.02	-.02	.08	.08
11. Selectie	.05	.04	.34 [1]	-.20
12. Leereffect	.04	-.03	.18	.02
13. Motivatie	.01	-.01	.14	.03
14. Onderwijs	.04	.11	.25	-.15

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.7: Cronbach alfa's van en correlaties tussen de schalen voor competentiegroei bij studenten

Competenties	Alfa	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht	.66											
02. Zelfmanagement	.71	.56 [2]										
03. Zelfreflectie	.74	.53 [2]	.47 [2]									
04. Doelgericht	.81	.55 [2]	.60 [2]	.60 [2]								
05. Kansen zien	.70	.49 [2]	.59 [2]	.52 [2]	.63 [2]							
06. Leidinggeven	.80	.51 [2]	.51 [2]	.42 [2]	.51 [2]	.49 [2]						
07. Vernieuwen	.57	.49 [2]	.37 [2]	.54 [2]	.44 [2]	.41 [2]	.31 [2]					
08. Samenwerken	.81	.48 [2]	.45 [2]	.48 [2]	.70 [2]	.57 [2]	.52 [2]	.40 [2]				
09. Feedback geven	.83	.41 [2]	.51 [2]	.40 [2]	.49 [2]	.47 [2]	.47 [2]	.32 [2]	.49 [2]			
10. Feedback vragen	.63	.43 [2]	.42 [2]	.63 [2]	.50 [2]	.61 [2]	.31 [2]	.47 [2]	.48 [2]	.36 [2]		
11. Klantgericht werken	.76	.40 [2]	.52 [2]	.43 [2]	.59 [2]	.60 [2]	.49 [2]	.33 [2]	.57 [2]	.52 [2]	.45 [2]	
12. Voor groep spreken	.75	.47 [2]	.41 [2]	.38 [2]	.49 [2]	.39 [2]	.47 [2]	.43 [1]	.50 [2]	.33 [2]	.43 [2]	.45 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 5.8: Correlaties tussen de items voor competentiegroei bij begeleiders

Competenties	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht											
02. Zelfmanagement	.51 [2]										
03. Zelfreflectie	.36 [2]	.63 [2]									
04. Doelgericht	.34 [2]	.59 [2]	.64 [2]								
05. Kansen zien	.28 [2]	.54 [2]	.41 [2]	.49 [2]							
06. Leidinggeven	.19	.44 [2]	.48 [2]	.45 [2]	.51 [2]						
07. Vernieuwen	.23 [1]	.42 [2]	.30 [2]	.46 [2]	.66 [2]	.42 [2]					
08. Samenwerken	.27 [1]	.46 [2]	.58 [2]	.54 [2]	.46 [2]	.41 [2]	.57 [2]				
09. Feedback geven	.13	.36 [2]	.46 [2]	.49 [2]	.52 [2]	.44 [2]	.54 [2]	.54 [2]			
10. Feedback vragen	.14	.35 [2]	.45 [2]	.48 [2]	.47 [2]	.49 [2]	.51 [2]	.59 [2]	.72 [2]		
11. Klantgericht werken	.07	.30 [2]	.31 [2]	.44 [2]	.46 [2]	.40 [2]	.49 [2]	.44 [2]	.64 [2]	.45 [2]	
12. Voor groep spreken	-.01	.16	.15	.30 [2]	.39 [2]	.40 [2]	.39 [2]	.34 [1]	.49 [2]	.27 [2]	.56 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 5.9: Regressie bij studenten op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.17 [2]	-.02	-.02	-.18	.09	.33 [2]
01. Praktijkgericht	.09	.05	-.07	-.09	.13	.25 [1]
02. Zelfmanagement	.10 [1]	-.10	-.02	-.10	.13	.25 [1]
03. Zelfreflectie	.08	.08	-.04	-.07	.03	.26 [1]
04. Doelgericht	.17 [2]	-.05	.08	-.22	.17	.24 [1]
05. Kansen zien	.09	.06	.10	-.14	.20	-.07
06. Leidinggeven	.06	-.04	-.11	-.01	.12	.20 [1]
07. Vernieuwen	.09	-.06	-.10	-.19	.06	.19
08. Samenwerken	.13 [1]	-.10	-.08	-.25 [2]	.10	.20 [1]
09. Feedback geven	.06	-.09	.06	-.01	-.02	.19
10. Feedback vragen	.16 [2]	.16	.05	-.21 [1]	-.04	.23 [1]
11. Klantgericht werken	.15 [2]	-.04	.06	-.11	.07	.33 [2]
12. Voor groep spreken	.18 [2]	-.03	-.08	-.15	.03	.38 [2]

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.10: Regressie bij begeleiders op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.14 [2]	.08	.19	.26
01. Praktijkgericht	.03	-.04	.15	.00
02. Zelfmanagement	.08	.11	.26	.03
03. Zelfreflectie	.11 [1]	.05	.10	.25
04. Doelgericht	.11 [1]	.16	.19	.15
05. Kansen zien	.12 [1]	.10	.10	.26
06. Leidinggeven	.14 [2]	-.12	.35 [1]	-.03
07. Vernieuwen	.09	.01	.18	.13
08. Samenwerken	.12 [1]	.01	-.16	.45 [2]
09. Feedback geven	.17 [2]	.12	.19	.25
10. Feedback vragen	.14 [2]	-.03	-.08	.42 [2]
11. Klantgericht werken	.09 [1]	.11	.28	.02
12. Voor groep spreken	.03	.10	.05	.12

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 5.11: Gemiddelden per rol op de kwaliteitscriteria voor de specifieke interventie

Criteria	Iedereen	Studenten	Begeleiders	ΔBS
00. Algemeen	1.51 (0.89) [2]	1.37 (0.96) [2]	1.65 (0.80) [2]	0
01. Geschikt	1.85 (1.11) [2]	1.58 (1.12) [2]	2.12 (1.06) [2]	0
02. Vergelijk	1.50 (1.18) [2]	1.22 (1.24) [2]	1.77 (1.06) [2]	0
03. Herhaalbaar	0.79 (1.25) [2]	0.43 (1.25) [0]	1.15 (1.16) [2]	1
04. Transparant	1.06 (1.45) [2]	0.89 (1.38) [2]	1.25 (1.53) [2]	0
05. Acceptatie	1.67 (1.20) [2]	1.63 (1.36) [2]	1.71 (1.03) [2]	0
06. Authentiek	1.33 (1.35) [2]	1.31 (1.47) [2]	1.36 (1.24) [2]	0
07. Eerlijk	1.86 (1.18) [2]	1.53 (1.34) [2]	2.20 (0.88) [2]	1
08. Complex	1.00 (1.35) [2]	0.90 (1.48) [2]	1.11 (1.21) [2]	0
09. Betekenis	2.06 (0.83) [2]	2.10 (0.84) [2]	2.02 (0.84) [2]	0
10. Tijd/Kosten	1.98 (1.19) [2]	1.67 (1.36) [2]	2.30 (0.91) [2]	1
11. Selectie	1.82 (1.13) [2]	1.71 (1.17) [2]	1.93 (1.09) [2]	0
12. Leereffect	1.90 (0.92) [2]	1.83 (0.90) [2]	1.98 (0.93) [2]	0
13. Motivatie	0.82 (1.43) [2]	1.03 (1.41) [2]	0.60 (1.45) [1]	0
14. Onderwijs	1.81 (0.94) [2]		1.81 (0.94) [2]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Iedereen" is met een subscript aangegeven of het gemiddelde voor alle deelnemers verschilt van 0. Ook in de kolommen "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 5.12: Gemiddelden per conditie en per rol op de kwaliteitscriteria voor de algemene beoordeling

	Controleconditie				Experimentele conditie			
Criteria	Iedereen	Studenten	Begeleiders	ΔBS	Iedereen	Studenten	Begeleiders	ΔBS
00. [0] Algemeen	1.30 (0.78) [2]	1.20 (0.74) [2]	1.42 (0.82) [2]	0	1.22 (0.91) [2]	1.10 (1.11) [2]	1.35 (0.64) [2]	0
01. [0] Geschikt	2.14 (0.71) [2]	2.11 (0.61) [2]	2.17 (0.81) [2]	0	2.05 (0.93) [2]	1.91 (1.11) [2]	2.20 (0.69) [2]	0
02. [0] Vergelijk	0.96 (1.55) [2]	0.47 (1.55) [2]	1.56 (1.32) [2]	2	0.84 (1.58) [2]	0.43 (1.76) [0]	1.27 (1.25) [2]	1
03. [0] Herhaalbaar	-0.14 (1.43) [0]	-0.32 (1.45) [0]	0.08 (1.39) [0]	0	0.01 (1.45) [0]	-0.15 (1.96) [0]	0.18 (1.28) [0]	0
04. [0] Transparant	1.49 (0.97) [2]	1.50 (0.89) [2]	1.47 (1.08) [2]	0	1.41 (1.16) [2]	1.40 (1.36) [2]	1.42 (0.94) [2]	0
05. [0] Acceptatie	1.70 (0.95) [2]	1.62 (0.95) [2]	1.81 (0.94) [2]	0	1.60 (1.22) [2]	1.54 (1.48) [2]	1.67 (0.91) [2]	0
06. [0] Authentiek	1.82 (0.87) [2]	1.69 (0.84) [2]	1.98 (0.88) [2]	0	1.71 (1.08) [2]	1.65 (1.14) [2]	1.79 (1.02) [2]	0
07. [0] Eerlijk	1.10 (1.27) [2]	0.74 (1.28) [2]	1.54 (1.13) [2]	2	1.00 (1.36) [2]	0.62 (1.58) [1]	1.39 (0.95) [2]	1
08. [1] Complex	0.65 (1.19) [2]	0.57 (1.13) [2]	0.76 (1.25) [2]	0	0.21 (1.48) [0]	-0.05 (1.75) [0]	0.48 (1.09) [1]	0
09. [0] Betekenis	1.88 (1.06) [2]	2.05 (0.75) [2]	1.66 (1.31) [2]	1	1.82 (0.96) [2]	1.82 (1.01) [2]	1.82 (0.92) [2]	0
10. [0] Tijd/Kosten	1.35 (1.06) [2]	1.18 (0.95) [2]	1.56 (1.15) [2]	1	1.39 (1.24) [2]	1.17 (1.46) [2]	1.62 (0.92) [2]	0
11. [0] Selectie	1.32 (1.16) [2]	1.13 (1.12) [2]	1.56 (1.12) [2]	1	1.46 (1.02) [2]	1.23 (1.25) [2]	1.70 (0.69) [2]	0
12. [0] Leereffect	1.68 (0.96) [2]	1.71 (0.91) [2]	1.63 (1.04) [2]	0	1.70 (0.84) [2]	1.74 (0.88) [2]	1.67 (0.82) [2]	0
13. [0] Motivatie	0.95 (1.26) [2]	1.18 (1.19) [2]	0.65 (1.29) [2]	1	0.70 (1.55) [2]	0.99 (1.71) [2]	0.40 (1.34) [0]	0
14. [0] Onderwijs	1.30 (1.01) [2]		1.30 (1.01) [2]		1.40 (0.89) [2]		1.40 (0.89) [2]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Criteria" is met een subscript voor de naam van het criterium aangegeven of het gemiddelde van alle deelnemers verschilt over de condities heen. In de kolommen "Iedereen", "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 5.13: Gemiddelden per conditie en per rol op de schalen voor competentiegroei

			Controleconditie			Experimentele conditie		
Competenties	ΔB	ΔS	Begeleiders	Studenten	ΔBS	Begeleiders	Studenten	ΔBS
00. Algemeen	0	0	4.34 (0.58) [2]	4.14 (0.38) [2]	1	4.19 (0.52) [2]	4.12 (0.36) [2]	0
01. Praktijkgericht	0	0	4.81 (0.35) [2]	4.30 (0.37) [2]	2	4.70 (0.47) [2]	4.21 (0.48) [2]	2
02. Zelfmanagement	0	0	4.75 (0.43) [2]	4.12 (0.43) [2]	2	4.52 (0.64) [2]	4.06 (0.47) [2]	2
03. Zelfreflectie	0	0	4.57 (0.66) [2]	4.39 (0.44) [2]	0	4.47 (0.62) [2]	4.38 (0.47) [2]	0
04. Doelgericht	0	0	4.39 (0.73) [2]	4.03 (0.47) [2]	2	4.27 (0.74) [2]	4.04 (0.43) [2]	0
05. Kansen zien	0	0	4.34 (0.84) [2]	4.16 (0.59) [2]	0	4.24 (0.87) [2]	4.07 (0.59) [2]	0
06. Leidinggeven	0	0	4.44 (0.81) [2]	3.94 (0.59) [2]	2	4.27 (0.63) [2]	4.02 (0.54) [2]	0
07. Vernieuwen	0	0	4.36 (0.89) [2]	4.18 (0.50) [2]	0	4.20 (0.69) [2]	4.10 (0.58) [2]	0
08. Samenwerken	0	0	4.30 (1.02) [2]	3.98 (0.61) [2]	1	4.17 (0.82) [2]	4.11 (0.49) [2]	0
09. Feedback geven	0	0	4.19 (0.91) [2]	3.78 (0.61) [2]	2	4.04 (0.79) [2]	3.77 (0.50) [2]	0
10. Feedback vragen	0	0	4.41 (0.76) [2]	4.38 (0.49) [2]	0	4.33 (0.62) [2]	4.34 (0.58) [2]	0
11. Klantgericht werken	0	0	3.90 (1.07) [2]	4.10 (0.56) [2]	0	3.42 (1.31) [0]	4.03 (0.44) [2]	1
12. Voor groep spreken	0	0	3.62 (1.12) [2]	4.33 (0.49) [2]	2	3.67 (1.53) [1]	4.34 (0.44) [2]	1

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. Als de gemiddelden afwijken van 3 (= het neutraal middelpunt), dan is dit aangegeven met een subscript. Of de gemiddelden van begeleiders verschillen over condities heen, is weergegeven met een cijfer in de kolom "ΔB". Of de gemiddelden van studenten verschillen over condities heen, is weergegeven met een cijfer in de kolom "ΔS". Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

6 STUDIE 2

Voor academiejaar 2017-2018 werd voor het laatste jaar van de opleiding Verpleegkunde (VPK) gekeken op welke manier de stagebeoordeling verbeterd kon worden. Hierbij werd enerzijds vertrokken van ons theoretisch kader en anderzijds vanuit de noden van de opleiding zelf. In samenspraak met de opleiding werd een interventie uitgewerkt alsook een experimenteel onderzoeksdesign dat toeliet de effecten van die interventie na te gaan. Concreet was het, net als in Studie 1, de bedoeling om het effect na te gaan van het toevoegen van extra beoordelaars. Specifiek wilden we te weten komen hoe het toevoegen van extra beoordelaars (i.e., de studenten zelf, via peer assessment) onthaald zou worden door de betrokken actoren (i.e., studenten en begeleiders) en wat daar de eventuele leereffecten van zouden zijn.

Op welke wijze sluit deze interventie aan bij het theoretisch kader? Deze interventie zet, net als Studie 1, in op ratings en observaties (zie "Gebruik meer ratings en observaties"). Specifiek wordt er op deze manier geprobeerd (1) beoordelaars meer observaties te laten doen, (2) meer mensen te betrekken in het beoordelingsproces, en (3) verschillende types beoordelaars te gebruiken. In tegenstelling tot Studie 1 worden de extra beoordelaars hier echter niet gezocht onder begeleiders maar onder de studenten zelf. Concreet is het de bedoeling dat studenten elkaar beoordelen. Zijdelings zet deze interventie ook in op "Gebruik meer beoordelingsmethoden" en "Organiseer trainingen voor docenten en mentoren". Het is immers expliciet de bedoeling om op deze manier additionele bronnen van informatie aan te boren en het is ook de bedoeling om niet alleen studenten te ondersteunen bij het peer assessment maar ook de begeleiders te ondersteunen bij het in groep bespreken van de resultaten hiervan.

Omwille van de nadruk die tegenwoordig in de literatuur gelegd wordt op het procesmatige karakter van stages, werd bewaakt dat de interventie niet enkel een effect zou hebben op de kwaliteit van de beoordeling maar ook op de kwaliteit van het stageproces. Er werd dus voor gezorgd dat de interventie niet enkel zou inspelen op de stagebeoordeling zelf (= het summatief aspect) maar ook op het formatieve, procesmatige aspect. Het peer assessment werd daartoe niet enkel summatief maar ook formatief ingezet.

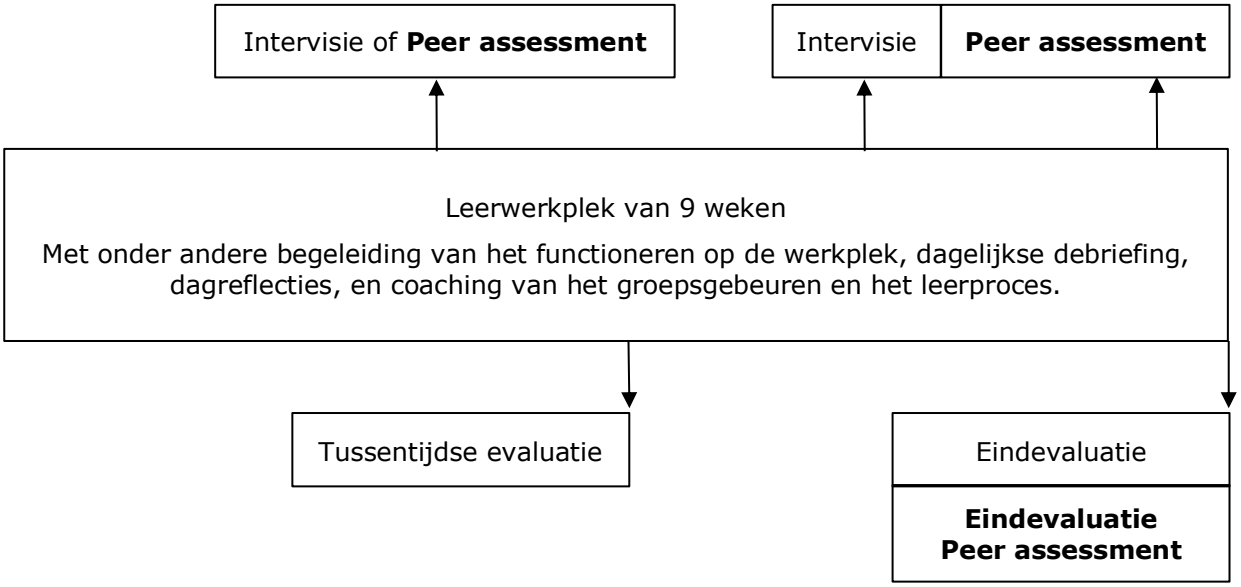
6.1 Methode

6.1.1 Procedure

De interventie werd uitgewerkt in het kader van de leerwerkplek in het laatste jaar: een stage van 9 weken die eruit bestaat dat een groep studenten een (deel van een) afdeling van een ziekenhuis overneemt. Ze nemen hierbij in team de verschillende rollen van de verpleegkundige op. Om het beroep te leren, krijgen zij de verantwoordelijkheid om alle beroepstaken zelf uit te voeren. In de praktijk betekent dit dat een groep studenten, onder begeleiding van een aantal mentoren en een docent, als zelfstandig verpleegkundig team functioneert voor een aantal kamers van de afdeling. De mentoren volgen de handelingen van dit team op en blijven eindverantwoordelijke voor de geleverde zorg. Onderwijsactiviteiten gerelateerd aan de leerwerkplek worden door de docent op de dienst of in de instelling georganiseerd.

Ongeveer de helft van de studenten die deze stage doorliepen, doorliepen het "klassiek" traject. Deze studenten vormden de controlegroep binnen het experimentele onderzoeksdesign. De andere studenten doorliepen het "klassieke traject" met de aanpassingen die in het kader van dit onderzoek werden ingevoerd en vormden de experimentele groep. Het is niet onze bedoeling om hier het volledige verloop van het "klassieke traject" te beschrijven. Daarvoor verwijzen we naar het stagevademecum dat op te vragen valt via e-mail. Het is wel onze bedoeling duidelijk te maken welke wijzigingen we in het kader van dit onderzoek doorvoerden. In dit geval gaat het om uitbreidingen. De uitbreidingen betroffen peer assessments, waarvoor gebruik werd gemaakt van het CanMEDS model (Frank, Snell, & Sherbino, 2015), waarbij CanMEDS staat voor Canadian Medical Education Directives for Specialists. Het CanMEDS model werd ontwikkeld om de vaardigheden die artsen en andere medische zorgverleners nodig hebben om aan de noden van hun patiënten tegemoet te komen te identificeren en te beschrijven. Dit model omvat zeven rollen (i.e. Medisch expert, Communicator, Teamspeler, Manager, Belangenbehartiger, Levenslang lerende, en Professional) die geassocieerd zijn met kerncompetenties en gedragsindicatoren. De studenten dienden alle medestudenten op de leerwerplek op elke rol te beoordelen aan de hand van een score, gaande van "5 = beter dan de rest van de groep" over "4 = gemiddeld van de groep", "3 = niet zo goed als de rest van de groep, "2 = geen hulp voor de groep" tot "1 = hinder voor de groep". De uitbreidingen t.a.v. het "klassieke traject" worden weergegeven in Figuur 6.1 en hieronder beschreven.

Figuur 6.1: Het normale verloop van de stage met in het vet de uitbreidingen die in het kader van dit project werden ingevoerd in de experimentele conditie.



- Tijdens Week 3 tot 5 vulden studenten via Blackboard twee CanMEDS-formulieren in, een formulier waarin ze de eigen sterke en zwakke punten moesten aangeven, en een formulier waarin ze de sterke en zwakke punten van elk van de medestudenten moesten aangeven. In Week 5 vond tijdens het geplande intervisiemoment een groepsgesprek

plaats op basis van het peer assessment. Studenten gingen er in gesprek met elkaar over de scores, en hun sterke en zwakke punten. Dit werd door de docent in goede banen geleid. Mentoren die dat wilden, konden hier ook bij aanwezig zijn. Op basis van de op die manier verkregen informatie formuleerden studenten een leerdoel. Dit leerdoel werd vervolgens opgevolgd. Dit luik was louter formatief bedoeld.

- Tijdens de laatste week vulden studenten nogmaals een CanMEDS-formulier in waarin ze de sterke en zwakke punten van elk van de medestudenten moesten aangeven. Net zoals in de controlegroep gaven de docenten scores zoals voorzien in het stagedocument maar in de experimentele groep werden de docenten ook gevraagd wat ze vonden van de scores op het peer assessment en welk gewicht die volgens hen zou moeten krijgen. Het tweede peer assessment was dus louter summatief, al is het ook pro forma, in de zin dat het peer assessment niet echt mee telde in het eindcijfer.

Na afloop van de stage kregen alle studenten en begeleiders een vragenlijst. Deze vragenlijst bestond uit drie of vier onderdelen, afhankelijk van de groep waarin iemand was ondergebracht (i.e., de experimentele groep of de controlegroep).

1. In deze vragenlijst werden voor iedereen eerst een aantal persoonsgegevens (i.e. naam, voornaam en e-mailadres) en achtergrondkenmerken bevraagd (i.e., leeftijd en geslacht). Bij de begeleiders werd additioneel gepolst naar hun ervaring (in termen van het aantal jaar dat ze reeds bij het begeleiden van stages betrokken zijn). Bij de studenten werd additioneel gepolst naar vooropleiding en etniciteit. Vervolgens werd aan studenten ook gevraagd wat volgens hen het punt was dat ze verdienden op hun stage (= een self-assessment). Ter uitdieping van het Selectiecriterium werd na de juni-zittijd aan deze gegevens het effectief verkregen punt toegevoegd.
2. Bij de studenten en begeleiders uit de experimentele groep werd vervolgens gepolst naar de perceptie van de specifiek aangebrachte wijziging. Specifiek peilden de gestelde vragen naar de gepercipieerde bijdrage van deze interventie aan de kwaliteit van de algemene beoordelingsprocedure. Dit vormde een "zachte" test van onze interventie: werd de interventie op zich goed onthaald door de betrokkenen?
3. Daarna werd bij alle deelnemers gepolst naar de perceptie van de algemene beoordelingsprocedure aan de hand van de kwaliteitscriteria uit het model van Baartman. Op die manier kon de perceptie van de experimentele groep en de controlegroep vergeleken worden. Dit vormde een "harde" test van onze interventie: als die als een verbetering kon worden aanzien, dan zou de experimentele groep minstens op sommige criteria uit het model van Baartman beter moeten scoren dan de controlegroep.
4. Tot slot werd bij alle deelnemers de competentiegroeimeter voor stage afgenomen ter uitdieping van het Leereffectcriterium. De competentiegroeimeter is een binnen KdG veelgebruikt instrument, gebaseerd op het werk van Braun et al. (2008; Braun & Leidner, 2009). De competentiegroeimeter laat enerzijds toe te identificeren welke competentiegroei begeleiders verwachten en anderzijds of dat zich vertaalt in de beleving van studenten. Voor ons onderzoek betekent dit het volgende: als we verschillen vinden tussen de scores van begeleiders uit de experimentele groep en de controlegroep, dan betekent dit dat begeleiders verwachtten dat studenten andere competenties zouden ontwikkelen omwille van (het procesmatige karakter van) onze interventie. Of dit dan effectief zo was, kon dan bij de studenten getoetst worden.

6.1.2 Steekproef

In totaal werden 104 studenten uitgenodigd om aan ons onderzoek deel te nemen. Deze studenten liepen stage in het voorjaar van 2018. Studenten die vervroegd stage liepen (= in het najaar van 2017) werden niet uitgenodigd. Van de 104 uitgenodigde studenten namen er 91 deel aan ons onderzoek (= 87.5%). Daarnaast werden ook 57 begeleiders uitgenodigd (18 docenten en 39 mentoren). Hiervan vulden er 40 (= 70.2%) de vragenlijst in: 15 docenten en 25 mentoren. De gemiddelde leeftijd van de deelnemende begeleiders was 36.68 jaar (SD = 9.66). Gemiddeld genomen hadden ze 5.69 jaar ervaring met stagebegeleiding (SD = 4.23). Het overgrote deel waren vrouwen (85.0%; met het oog op de analyses die nog volgen, werd deze variabele als volgt gecodeerd: 1 voor man en 2 voor vrouw). De studenten bestonden voor 80.8% uit vrouwen. Hun gemiddelde leeftijd was 23.80 jaar (SD = 5.57). Qua vooropleiding kwamen 11.5% van de studenten uit het BSO, 55.8% uit het TSO, en 26.9% uit het ASO. Op basis van deze gegevens werd een variabele "educatie" geconstrueerd waarop studenten een 1 kregen als ze uit het BSO kwamen, een 2 als ze uit het TSO kwamen of als hier geen duidelijke gegevens over te vinden waren, en een 3 als ze uit het ASO kwamen (zie Voorstudie 1). De restcategorie (= 2) vormt de baseline waartegen effecten van een lage (= BSO) en een hoge educatie (= ASO) worden afgezet. De meeste studenten hebben de Belgische nationaliteit (98.1%), zijn in België geboren (96.2%), en hebben een grootmoeder langs moederszijde die in België geboren is (87.5%). Op basis van deze gegevens werd een variabele "etniciteit" geconstrueerd waarop studenten een 0 kregen als ze de Belgische nationaliteit hebben, in België geboren zijn en een grootmoeder van Belgische nationaliteit hebben en een 1 als aan één van deze voorwaarden niet is voldaan (zie Voorstudie 1). Op deze manier werd 7.1% van de studenten gecategoriseerd als van "een andere etniciteit". De studenten vonden gemiddeld genomen dat ze 14.88 op 20 verdienden op hun stage (SD = 1.84). Geen enkele student gaf aan dat hij niet verdiende te slagen.

Voor studenten die niet deelnamen aan het onderzoek, werden de achtergrondgegevens uit Bamaflex gehaald. De studenten die niet deelnamen en de studenten die wel deelnamen verschilden niet significant in termen van geslacht (Mean = 1.77 en 1.81; (F (1, 102) = 0.14, ns), leeftijd (Mean = 23.08 en 23.90; (F (1, 102) = 0.25, ns), educatie (Mean = 2.23 en 2.14; (F (1, 102) = 0.24, ns) en etniciteit (Mean = 0.08 en 0.14; (F (1, 102) = 0.42, ns). Er is dus geen aanwijzing dat de studenten die deelnamen aan het onderzoek niet representatief zouden geweest zijn voor de volledige studentenpopulatie.

De stageplaatsen werden op voorhand in de experimentele conditie (6 stageplaatsen) dan wel de controleconditie (5 stageplaatsen) ondergebracht. Op die manier belandden 50 studenten (= 54%) en 21 begeleiders (= 51%) in de experimentele groep en 41 studenten (= 46%) en 19 begeleiders (= 49%) in de controlegroep. Om te controleren of de studenten uit de controleconditie vergelijkbaar waren met die uit de experimentele conditie, voerden we een reeks Oneway Anova's uit. Eerst onderzochten we of beide groepen vergelijkbaar waren qua achtergrondkenmerken. De groepen verschilden niet significant van elkaar in termen van geslacht (Mean = 1.80 en 1.82; (F (1, 89) = 0.03, ns), leeftijd (Mean = 23.93 en 23.88; (F (1, 89) = 0.00, ns), educatie (Mean = 0.65 en 0.57; (F (1, 89) = 0.98, ns) en etniciteit (Mean = 0.20 en 0.10; (F (1, 89) = 1.66, ns). Vervolgens onderzochten we of de groepen vergelijkbaar waren in termen van het punt dat ze volgens zichzelf verdienden op hun stage. De groepen verschilden

niet significant van elkaar (Mean = 14.53 en 15.12; (F (1, 86) = 2.27, ns). Er is dus geen aanwijzing dat de assumptie van een at random verdelen geschaad zou zijn. Ook voor begeleiders onderzochten we of de groepen vergelijkbaar waren qua achtergrondkenmerken. De groepen verschilden niet significant in termen van geslacht (Mean = 1.74 en 1.95; (F (1, 38) = 3.79, ns), leeftijd (Mean = 36.16 en 37.14; (F (1, 38) = 0.10, ns) en ervaring (Mean = 4.58 en 6.75; (F (1, 37) = 2.68, ns). Ook voor begeleiders is er dus geen aanwijzing dat de assumptie van een at random verdelen geschaad zou zijn.

6.1.3 Effectmetingen

6.1.3.1 Kwaliteitscriteria voor de specifieke interventie

Na afloop van de stage werd bij alle studenten en begeleiders uit de experimentele conditie gepolst naar de perceptie van de specifieke wijziging die aangebracht werd in de experimentele conditie. Ook dit gebeurde aan de hand van de verkorte versie van de kwaliteitscriteria-vragenlijst. Deze vragenlijst werd daartoe voorzien van een andere inleiding. Alle items werden ook geherformuleerd opdat ze zouden polsen naar de gepercipieerde bijdrage van de interventie aan de kwaliteit van de algemene beoordeling. In totaal beoogde deze verkorte versie 14 criteria te meten: (1) Geschiktheid, (2) Vergelijkbaarheid, (3) Herhaalbaarheid, (4) Transparantie, (5) Acceptatie, (6) Authenticiteit, (7) Eerlijkheid, (8) Complexiteit, (9) Betekenisvolheid, (10) Tijd & Kosten, (11) Selectie, (12) Leereffect, (13) Motivatie, en (14) Onderwijseffect. Het criterium Onderwijseffect werd enkel bevraagd bij begeleiders omdat dit criterium handelt over het effect dat de beoordeling op zich kan hebben op de manier waarop de stage in de toekomst vormgegeven kan worden (iets waar studenten niet mee bezig zijn). Elk criterium werd gemeten aan de hand van drie vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 5.

De Cronbach alfas en de correlaties tussen de schalen zijn te vinden in Tabel 6.1. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders.[1] Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70), behalve Herhaalbaarheid. Hiervoor was de interne consistentie aanvaardbaar (Cronbach alfa = .67). Gezien de substantiële correlaties tussen heel wat schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Die kon 67% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze had een Cronbach alfa van .96.

---

1 Zes begeleiders vulden de vragenlijst niet in voor wat de specifieke interventie betreft. Het ging telkens om mentoren die de sessies rond peer assessment niet hadden bijgewoond er dus eigenlijk niet mee in aanraking waren gekomen. Dit deel van de vragenlijst werd dus maar door 15 begeleiders ingevuld.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 6.2. Noch de algemene score noch de onderliggende schalen, met uitzondering van Tijd & Kosten, konden significant worden voorspeld op basis van de beschikbare gegevens. De haalbaarheid van de door de interventie vereiste Tijd en Kosten werd hoger ingeschat door jongere studenten die van zichzelf vonden dat ze niet zoveel punten verdienden. Misschien zagen zij hun heil in een beoordeling die meer ruimte laat voor peer assessment omdat ze dachten op die manier een beter eindoordeel te kunne krijgen. De resultaten voor begeleiders zijn te vinden in Tabel 6.3. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens.

6.1.3.2 Kwaliteitscriteria voor de algemene beoordeling

Na afloop van de stage werd bij alle studenten en begeleiders uit zowel de experimentele conditie als de controleconditie gepolst naar de perceptie van de algemene stagebeoordeling. Dit gebeurde aan de hand van een verkorte versie van de kwaliteitscriteria-vragenlijst (zie Voorstudie 2). In totaal beoogde deze verkorte versie de 14 bovenvermelde criteria te meten. Het Onderwijseffect werd enkel bevraagd bij begeleiders omdat dit criterium handelt over het effect dat de beoordeling op zich kan hebben op de manier waarop de stage in de toekomst vormgegeven kan worden (iets waar studenten niet mee bezig zijn). Elk criterium werd gemeten aan de hand van drie vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). De vragen zijn te vinden in Appendix 3.

De Cronbach alfa's en de correlaties tussen de schalen zijn te vinden in Tabel 6.4. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen de schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Deze kon 59% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40 op deze factor. Omdat zo'n factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze had een Cronbach alfa van .94.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 6.5. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens. De resultaten voor begeleiders zijn te vinden in Tabel 6.6. Ook hier konden noch de algemene score noch de onderliggende schalen significant worden voorspeld op basis van de beschikbare gegevens.

6.1.3.3 Uitdieping van het Leereffectcriterium

Na afloop van de stage werd bij alle studenten en begeleiders uit zowel de experimentele conditie als de controleconditie de competentiegroeimeter voor stage afgenomen ter uitdieping van het Leereffectcriterium. De gebruikte versie beoogde competentiegroei te meten in 12 domeinen: (1) Praktijkgericht, (2) Zelfmanagement, (3) Zelfreflectie, (4) Doelgericht werken, (5) Kansen zien en nemen, (6) Leidinggeven en activeren, (7) Vernieuwen, (8) Samenwerken, (9) Feedback geven, (10) Feedback vragen en ontvangen, (11) Klantgericht werken, en (12) Voor een groep spreken. De competentiegroeimeter voor stage bestaat in twee versies: een studenten- en een docentenversie. De docentenversie, die bij alle begeleiders werd afgenomen, laat toe te identificeren welke competentiegroei begeleiders verwachten. In deze versie wordt elke concept gemeten aan de hand van één item. De studentenversie laat toe te identificeren welke competentiegroei studenten rapporteren. In deze versie wordt elke concept gemeten aan de hand van meerdere items (van 3 tot 7 al naargelang het concept). Alle vragen dienden te worden beantwoord op een 5-punten Likertschaal ("1 = helemaal oneens" tot "5 = helemaal eens"). De docentenversie is te vinden in Appendix 8 en de studentenversie in Appendix 9.

De Cronbach alfas en de correlaties tussen de schalen van de studentenversie zijn te vinden in Tabel 6.7. De correlaties tussen de items van de docentenversie zijn te vinden in Tabel 6.8. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen heel wat concepten, besloten we via factoranalyse te onderzoeken of het aantal schalen / items gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. Voor de studentenversie wees de scree plot van een exploratorische factoranalyse op een oplossing met één factor. Deze oplossing kon 55% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40 op deze factor. Omdat zo'n factor handig is om een overzicht te krijgen van de algemene competentiegroei, berekenden we een algemene competentiegroeiscore door alle schalen te middelen. Deze algemene score had een Cronbach alfa van .92. Ook voor de docentenversie wees de scree plot op een oplossing met één factor. Deze oplossing kon 43% van de totale variantie verklaren. Alle items hadden een factorlading > .40 op deze factor. Deze algemene score had een Cronbach alfa van .87.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen / items kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 6.9. De algemene score kon deels worden voorspeld op basis van het punt dat studenten vinden dat ze verdienen. Hetzelfde geldt voor zowat de helft van de onderliggende schalen. Voor de meeste andere schalen gingen de resultaten in dezelfde richting. Studenten die vonden dat ze meer punten verdienden, rapporteerden dus ook meer competentiegroei. In sommige gevallen (met name voor Samenwerken en Feedback vragen) werd dit nog eens versterkt door een lagere vooropleiding. Het waren vooral studenten met een lagere vooropleiding die van zichzelf vonden dat ze het heel goed deden voor hun stage die meest competentiegroei rapporteerden. In eerder uitzonderlijke gevallen spelen ook geslacht, leeftijd en etniciteit een rol: competentiegroei op sommige domeinen werd meer gerapporteerd door vrouwelijke studenten (i.e., Praktijkgericht, Zelfmanagement en Kansen zien), jongere studenten (i.e. Voor een groep spreken) en autochtone studenten (i.e., Praktijkgericht). De resultaten

voor begeleiders zijn te vinden in Tabel 6.10. Noch de algemene score noch de onderliggende items konden significant worden voorspeld op basis van de beschikbare gegevens.

6.1.3.4 Uitdieping van het Selectiecriterium

De studenten haalden gemiddeld 14.27 op 20 op hun stage (SD = 2.33). In totaal slaagden 2 studenten niet voor hun stage (= 2.2%). Via regressieanalyse gingen we na of het behaalde punt kon worden voorspeld op basis van de achtergrondkenmerken (i.e., leeftijd, geslacht, educatie en etniciteit). De scores die studenten kregen konden niet significant voorspeld worden op basis van deze gegevens (R^2 = 0.05, F (4, 85) = 1.16, ns). De score die studenten behaalden op hun stage was sterk gecorreleerd met het punt dat ze zegden te verdienen voor hun stage (r = .66, p < .01), wat erop wijst dat studenten hun eigen functioneren behoorlijk goed konden inschatten. Of toch in de zin dat de rangorde die ze op die manier spontaan maakten behoorlijk goed overeenkwam met de rangorde die naar voor kwam uit de stagebeoordeling. Wel is het zo dat studenten het effectieve punt dat ze kregen minder goed konden inschatten. Een paired samples t-test toont aan dat er een significant verschil is tussen het punt dat studenten vonden dat ze verdienden en het punt dat ze uiteindelijke behaalden (Mean = 14.89 en 14.32, t = 3.05, p < .01). Studenten schatten hun eigen prestatie dus positiever in dan hun beoordelaars. De score die studenten behaalden op hun stage was niet significant positief gecorreleerd met de algemene score op de kwaliteitscriteria voor de algemene beoordeling (r = .03, ns) maar wel met de algemene score voor competentiegroei (r = .21, p < .05). De correlaties tussen de eindscore en de individuele kwaliteitscriteria waren .00 voor Geschiktheid, .16 voor Vergelijkbaarheid, .05 voor Herhaalbaarheid, .11 voor Transparantie, .11 voor Acceptatie, .11 voor Authenticiteit, .08 voor Eerlijkheid, -.01 voor Complexiteit, -.08 voor Betekenisvolheid, .06 voor Tijd & Kosten, -.02 voor Selectie, -.07 voor Leereffect, en -.13 voor Motivatie. De correlaties tussen de eindscore en de onderliggende schalen voor competentiegroei waren .20 voor Praktijkgericht, .22 voor Zelfmanagement, .09 voor Zelfreflectie, .20 voor Doelgericht werken, .14 voor Kansen zien en nemen, .23 voor Leidinggeven en activeren, .10 voor Vernieuwen, .21 voor Samenwerken, .23 voor Feedback geven, .10 voor Feedback vragen en ontvangen, .07 voor Klantgericht werken, en .14 voor Voor een groep spreken. Alle vernoemde correlaties met een absolute waarde groter dan .20 zijn significant op niveau p < .05.

6.2 Verwachtingen

We verwachtten dat onze interventie positief onthaald zou worden, en dat, wanneer de betrokkenen gevraagd werden of de interventie een extra bijdrage leverde aan de kwaliteitscriteria, ze daar instemmend op zouden reageren. Dit vormde echter slechts een "zachte" test van onze interventie. Wanneer onze interventie echt een verbetering vormde, dan zou dit zich ook moeten vertalen in meer tevredenheid over bepaalde aspecten van de algemene beoordeling in de experimentele groep dan in de controlegroep. Specifiek verwachtten we dat de algemene beoordeling in de experimentele conditie beter zou scoren op de criteria Vergelijkbaarheid (door het peer assessment wordt de kans groter dat de beoordelaar er bij verschillende studenten op dezelfde dingen gelet wordt), Herhaalbaarheid (de beslissingen die worden genomen over studenten zijn gebaseerd op meerdere assessmentsituaties en meerdere beoordelaars), Eerlijkheid (door medestudenten in te schakelen krijgen studenten een eerlijkere kans hun competenties te demonstreren), Betekenisvolheid (door medestudenten in te schakelen krijgen studenten

extra kansen op betekenisvolle feedback), Zelfsturing (de grotere kans op betekenisvolle feedback zou de zelfreflectie moeten kunnen stimuleren), Selectie (de informatie uit het peer assessment zou de selectie van goede en slechte studenten ten goede moeten komen), en Leereffect (de grotere kans op betekenisvolle feedback zou de ontwikkeling van studenten moeten bevorderen) dan in de controleconditie. Hierbij dient te worden opgemerkt dat dit wel een "hoge" verwachting is. De ingrepen die werden doorgevoerd in het kader van onze interventie vormden immers maar een deel van het beoordelingsproces. Hun effect moest dus al erg sterk zijn vooraleer het zich op die manier zou manifesteren.

Wat het leereffect betreft, verwachtten we dat de interventie een positieve bijdrage zou leveren aan competenties zoals Zelfreflectie, Samenwerken, en Feedback vragen. In de experimentele conditie waren er namelijk meer mensen waar men feedback aan kon vragen, en de extra feedback die studenten konden krijgen, kon leiden tot een verbeterde zelfreflectie. Daarnaast zou de feedback van medestudenten ook het samenwerk ten goede moeten komen.

Wat selectie betreft, verwachtten we dat de vernieuwde procedure de selectie van goede en slechte studenten ten goede zou moeten komen. Dit zou zich dan moeten vertalen in de punten die studenten krijgen op hun stage. In de experimentele conditie verwachtten we een grotere spreiding van de punten dan in de controleconditie.

6.3 Resultaten

6.3.1 Kwaliteitscriteria voor de specifieke interventie

Voor de deelnemers uit de experimentele conditie zijn de gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor de specifieke interventie te vinden in Tabel 6.11 en dit per rol (i.e., Student en Begeleider). T-tests wezen uit dat het algemeen gemiddelde voor alle deelnemers voor de meeste criteria significant verschilde van 0 (= het neutraal middelpunt van de schaal) in positieve richting. Uitzonderingen waren Herhaalbaarheid, Authenticiteit en Motivatie. Voor deze criteria verschilde het gemiddelde niet van 0. Voor de meeste criteria waren de deelnemers het er dus mee eens dat interventie bijdroeg aan de kwaliteit van de beoordeling.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren van elkaar verschilden betreffende hun perceptie van de interventie. Uit deze analyses bleek dat begeleiders over het algemeen van mening zijn dat de interventie een grotere bijdrage leverde aan de kwaliteit van de algemene beoordeling dan studenten. Dit uit zich vooral voor Geschiktheid, Betekenisvolheid en Leereffect. Begeleiders schatten ook de haalbaarheid van de door de interventie gevergde Tijd & Kosten positiever in dan studenten. Het waren dus vooral de begeleiders die een meerwaarde zagen in de interventie. Voor studenten was het zelfs zo dat de scores doorgaans niet significant verschilden van 0, met uitzondering van Acceptatie, Eerlijkheid, Complexiteit, en Selectie. De studenten stonden hier dus vrij neutraal tegenover.

6.3.2 Kwaliteitscriteria voor de algemene beoordeling

De gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor de algemene beoordeling zijn te vinden in Tabel 6.12, en dit zowel per conditie als per rol. In beide condities wezen t-tests uit dat het algemeen gemiddelde voor alle deelnemers significant verschilde van 0 (= het neutraal middelpunt van de schaal) in de positieve richting. Het gemiddelde van alle deelnemers

in de controleconditie verschilsw significant van dat van de deelnemers in de experimentele conditie voor zowel de algemene score als alle onderliggende criteria. De deelnemers in de controleconditie stonden over de hele lijn nog positiever ten aanzien van de algemene beoordeling dan de deelnemers in de experimentele conditie.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren (i.e., Studenten en Begeleiders) binnen de condities van elkaar verschilden betreffende hun perceptie van de interventie. In de controleconditie vonden we geen significante verschillen tussen studenten en begeleiders. Alle scores van zowel studenten als begeleiders verschilden ook significant van 0 in positieve richting. Ook in de experimentele conditie verschilden alle scores van zowel studenten als begeleiders verschilden ook significant van 0 in positieve richting. Maar hier was het doorgaans zo dat begeleiders nog positiever scoorden dan studenten. Enkel voor Authenticiteit en Motivatie was dit verschil niet significant. De studenten in de experimentele conditie stonden minst positief (maar nog altijd wel duidelijk positief) tegenover de algemene beoordeling.

6.3.3 Uitdieping van het Leereffectcriterium

De gemiddelden en standaarddeviaties op de competenties uit de competentiegroeimeter zijn te vinden in Tabel 6.13, en dit zowel per conditie als per rol. In beide conditie wezen t-tests uit dat zowel de algemene score als alle deelscores in positieve richting significant verschilden van 3 (= het neutraal middelpunt van de schaal), en dit zowel bij begeleiders als bij studenten. In beide condities verwachtten begeleiders competentiegroei in elk van de gerapporteerde domeinen en de studenten rapporteerden dit ook. De competentiegroei die begeleiders verwachtten was niet significant anders in de experimentele conditie dan in de controleconditie en ook de door studenten gerapporteerde competentiegroei verschilde niet over condities heen.

In beide condities wezen t-tests uit dat de competentiegroei die begeleiders verwachtten op de meeste vlakken significant afweek van de competentiegroei die studenten rapporteerden. Begeleiders hadden de neiging meer competentiegroei te verwachten dan studenten rapporteerden. Dat geldt voor alle competenties, behalve Vernieuwen (in beide condities) en Leidinggeven (in de experimentele conditie). Ook voor deze competenties verwachtten begeleiders meer competentiegroei dan studenten rapporteerden, maar dit verschil was niet significant. Net als in Studie 1 werd de uitzondering gevormd door Voor een groep spreken. Op dit vlak rapporteerden studenten in beide condities meer competentiegroei dan begeleiders verwachtten.

6.3.4 Uitdieping van het Selectiecriterium

Teneinde na te gaan of de interventie een effect had op de selectie van studenten werden de punten die studenten behaalden op hun stage in de experimentele conditie vergeleken met de punten die studenten behaalden op hun stage in de controleconditie. Hierbij werd zowel gekeken naar de gemiddelde score als naar de spreiding van de punten (= de standaarddeviatie). Een Oneway Anova wees uit dat de controlegroep niet significant verschilde van de experimentele groep in termen van gemiddelde score (Mean = 14.05 en 14.40; $F(1, 88) = 0.62$, ns). Een Levene's test (Levene, 1960) wees uit dat de standaarddeviatie evenmin significant verschilde over beide groepen (SD = 2.26 en 2.38; $F(1, 88) = 0.43$, *ns*). Hiervoor hadden we uiteraard ook geen verschillen verwacht, want het peer assessment werd niet in rekening gebracht. Interessanter is uiteraard de scores te vergelijken die studenten zouden hebben gekregen indien het peer assessment wel in rekening zou zijn gebracht.

Binnen de experimentele conditie werd daarom ook gekeken naar de punten die de studenten zouden hebben behaald als het peer assessment zou hebben meegeteld. Deze punten werden vergeleken met de punten die ze nu toegekend kregen. In het huidige systeem haalden studenten gemiddeld 14.40 (SD = 2.38). Hadden de punten van het peer assessment meegeteld, dan zouden studenten gemiddeld 14.56 (SD = 2.66) hebben behaald. Een repeated measures toonde aan dat het gemiddelde niet significant wijzigt door het peer assessment in rekening te brengen (F (1, 49) = 1.93, ns). Een Oneway Anova wees uit dat de reguliere score in de controlegroep niet significant verschilde van de aangepaste score in de experimentele groep (= de score inclusief peer assessment; Mean = 14.05 en 14.56; F (1, 88) = 0.94, ns). Een Levene's test (Levene, 1960) wees uit dat de standaarddeviatie evenmin significant verschilde over beide groepen (SD = 2.26 en 2.66; F (1, 88) = 1.73, *ns*).

6.4 Conclusie

De "zachte" tests, waarbij we enkel binnen de experimentele groep polsten naar hoe de interventie werd ervaren, toonden aan dat de interventie vrij positief ervaren werd door alle betrokkenen. Wanneer er specifiek gepolst werd naar de bijdrage van de interventie aan de algemene beoordeling, dan zien we dat het gemiddelde voor alle deelnemers in positieve richting verschilde van 0, en dit voor bijna alle criteria (uitzonderingen zijn Herhaalbaarheid, Authenticiteit en Motivatie). Meer gedetailleerde analyses wezen echter uit dat begeleiders over het algemeen van mening waren dat de interventie een grotere bijdrage leverde aan de kwaliteit van de algemene beoordeling dan studenten. Het waren dus vooral de begeleiders die positief stonden tegenover de algemene beoordeling. Voor studenten was het zo dat de scores doorgaans niet significant verschilden van 0 (met uitzondering van Acceptatie, Eerlijkheid, Complexiteit, en Selectie). Studenten vonden dus wel dat de interventie het zicht van begeleiders op hun cognitieve vaardigheden groter maakte, waardoor ze de beoordeling allicht ook als eerlijker en meer acceptabel ervaarden, maar voor de rest stonden ze eerder neutraal tegenover de interventie in kwestie.

De "harde" tests, waarbij we op experimentele wijze twee groepen vergeleken, nuanceerde het initieel positieve beeld van de interventie nog meer. Hoewel deelnemers in beide condities over het algemeen positief stonden ten aanzien van de stagebeoordeling, bleek het gemiddelde in de controleconditie significant hoger te liggen dan in de experimentele conditie, en dit voor alle criteria. De deelnemers in de controleconditie stonden dus over de hele lijn positiever ten aanzien van de stagebeoordeling dan de deelnemers in de experimentele conditie. Hieruit dienen we te concluderen dat de interventie toch niet zo positief onthaald werd: ze drukt immers de positieve houding ten aanzien van de stagebeoordeling. In de controlegroep vonden we geen significante verschillen tussen het oordeel van studenten en het oordeel van begeleiders. Alle scores van zowel studenten als begeleiders verschilden ook significant van 0 in positieve richting. In de experimentele conditie daarentegen vonden we wel significante verschillen tussen studenten en begeleiders. Alle criteria werden positiever beoordeeld door begeleiders dan door studenten. Enkel voor Authenticiteit en Motivatie was dit verschil niet significant. De studenten in de experimentele conditie stonden dus minst positief (maar nog altijd wel duidelijk positief) ten aanzien van de stagebeoordeling. De algemene conclusie die uit deze resultaten dient te worden getrokken, is dan ook dat het peer assessment weinig meerwaarde had. Begeleiders zien hier wel de meerwaarde van in (al uit zich dat niet in de "harde" tests), maar studenten

niet. Mogelijks is de minder positieve score van studenten op de algemene beoordeling in de experimentele conditie een gevolg van het feit dat het peer assessment niet meetelde, waardoor studenten het misschien veel gedoe voor niets vonden. Dat kan hun oordeel beïnvloed hebben.

Wanneer er ingezoomd werd op Leereffect en Selectie vonden we geen effecten. Voor Selectie waren er geen significante verschillen tussen de controleconditie en de experimentele conditie, noch in termen van gemiddelde noch in termen van spreiding (= standaarddeviatie). Op basis hiervan dienen we te concluderen dat de interventie niet bijdroeg aan het onderscheiden van sterke en zwakke studenten. Met andere woorden: zelfs al hadden de scores op het peer assessment wel echt mee geteld, dan had dit voor studenten niets wezenlijks veranderd. Het peer assessment bracht dus niets in rekening wat niet sowieso al in rekening werd gebracht bij het bepalen van de uiteindelijke score. Dit kan in dit geval ook wel een gevolg zijn van de specifieke situatie. Het peer assessment werd hier ingevoerd in het kader van de leerwerkplek. In dit kader is het functioneren van (de studenten in) het team essentieel. De interactie tussen studenten en hun oordeel over elkaar wordt hier dus sowieso ook nog langs andere wegen bevraagd. Voor Leereffect wijzen de resultaten uit dat noch de verwachtte noch de gerapporteerde competentiegroei verschilde over condities heen. Hierbij dient te worden opgemerkt dat het niet zo verwonderlijk is dat de interventie geen extra competentiegroei teweegbracht. In de controleconditie vertoonden alle competenties gemiddeld genomen een erg hoge groeiscore (> 4 op een schaal waar 5 het maximum is) en dit zowel bij begeleiders als bij studenten. De verbetermogelijkheden die een interventie kon veroorzaken waren dan ook beperkt. In beide condities verwachtten begeleiders competentiegroei in elk domein en studenten rapporteerden dit ook. Over het algemeen verwachtten begeleiders te veel competentiegroei, in de zin dat hun scores significant hoger lagen dan die van studenten. De grote uitzondering werd, net zoals in Studie 1, gevormd door de dimensie Voor een groep spreken. Op dit vlak rapporteerden studenten meer competentiegroei dan begeleiders verwachtten.

Tabel 6.1: Cronbach alfas van en correlaties tussen de kwaliteitscriteria voor de specifieke interventie

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.90													
02. Vergelijk	.94	.79 [2]												
03. Herhaalbaar	.67	.41 [2]	.53 [2]											
04. Transparant	.92	.70 [2]	.78 [2]	.63 [2]										
05. Acceptatie	.94	.69 [2]	.72 [2]	.59 [2]	.82 [2]									
06. Authentiek	.92	.74 [2]	.72 [2]	.50 [1]	.72 [2]	.69 [2]								
07. Eerlijk	.91	.55 [2]	.62 [2]	.50 [2]	.60 [2]	.65 [2]	.54 [2]							
08. Complex	.93	.79 [2]	.78 [2]	.40 [2]	.75 [2]	.75 [2]	.77 [2]	.59 [2]						
09. Betekenis	.97	.65 [2]	.65 [2]	.34 [2]	.55	.60 [2]	.53 [2]	.47 [2]	.64 [2]					
10. Tijd/Kosten	.93	.75 [2]	.75 [2]	.47 [2]	.67 [2]	.64 [2]	.66 [2]	.70 [2]	.69 [2]	.78 [2]				
11. Selectie	.93	.64 [2]	.70 [2]	.40 [2]	.62 [2]	.66 [2]	.51 [2]	.82 [2]	.66 [2]	.57 [2]	.77 [2]			
12. Leereffect	.94	.83 [2]	.68 [2]	.37 [2]	.64 [2]	.73 [2]	.69 [2]	.59 [2]	.84 [2]	.71 [2]	.74 [2]	.67 [2]		
13. Motivatie	.95	.74 [2]	.70 [2]	.47 [2]	.66 [2]	.63 [2]	.77 [2]	.55 [2]	.78 [2]	.61 [2]	.73 [2]	.55 [2]	.72 [2]	
14. Onderwijs	.77	.75 [2]	.68 [2]	.45	.68 [2]	.74 [2]	.91 [2]	.67 [2]	.68 [2]	.76 [2]	.92 [2]	.71 [2]	.81 [2]	.49

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 6.2: Regressie bij studenten op de kwaliteitscriteria voor de specifieke interventie

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.12	-.13	-.31 [1]	.06	.04	-.25
01. Geschikt	.10	-.21	-.28	.15	-.09	-.09
02. Vergelijk	.08	-.06	-.26	.12	-.02	-.20
03. Herhaalbaar	.07	-.10	-.19	.10	.01	-.25
04. Transparant	.07	-.04	-.19	-.08	-.08	-.14
05. Acceptatie	.08	-.11	-.19	.16	.09	-.23
06. Authentiek	.08	-.05	-.20	-.08	.00	-.19
07. Eerlijk	.20	-.01	-.39 [2]	.17	.16	-.25
08. Complex	.07	.01	-.16	-.10	-.05	-.15
09. Betekenis	.17	-.23	-.16	.03	.12	-.37 [1]
10. Tijd/Kosten	.24 [1]	-.18	-.45 [1]	.02	.06	-.30 [1]
11. Selectie	.12	.04	-.33 [1]	.06	.06	-.06
12. Leereffect	.08	-.19	-.21	.08	.04	-.19
13. Motivatie	.13	-.16	-.32 [1]	-.01	.07	-.20

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.3: Regressie bij begeleiders op de kwaliteitscriteria voor de specifieke interventie

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.06	-.05	-.30	.12
01. Geschikt	.05	-.10	-.23	.17
02. Vergelijk	.05	-.15	-.22	-.55
03. Herhaalbaar	.04	.07	.21	-.20
04. Transparant	.06	-.15	-.23	.14
05. Acceptatie	.23	.05	-.60	.31
06. Authentiek	.03	.08	-.16	.01
07. Eerlijk	.03	-.12	-.12	.14
08. Complex	.11	-.01	-.39	.35
09. Betekenis	.42	.35	-.69	.35
10. Tijd/Kosten	.23	.12	-.41	-.05
11. Selectie	.35	-.48	-.46	.02
12. Leereffect	.11	.09	-.39	.17
13. Motivatie	.06	.11	.21	.05
14. Onderwijs	.18	.09	-.52	.25

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.4: Cronbach alfas van en correlaties tussen de kwaliteitscriteria voor de algemene beoordeling

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.87													
02. Vergelijk	.82	.54 [2]												
03. Herhaalbaar	.87	.42 [2]	.59 [2]											
04. Transparant	.89	.43 [2]	.52 [2]	.63 [2]										
05. Acceptatie	.90	.52 [2]	.54 [2]	.48 [2]	.78 [2]									
06. Authentiek	.87	.60 [2]	.46 [2]	.45 [2]	.68 [2]	.69 [2]								
07. Eerlijk	.90	.49 [2]	.51 [2]	.58 [2]	.68 [2]	.75 [2]	.72 [2]							
08. Complex	.89	.40 [2]	.44 [2]	.49 [2]	.57 [2]	.59 [2]	.61 [2]	.75 [2]						
09. Betekenis	.94	.26 [2]	.20 [1]	.27 [2]	.49 [2]	.50 [2]	.50 [2]	.57 [2]	.70 [2]					
10. Tijd/Kosten	.93	.55 [2]	.56 [2]	.53 [2]	.58 [2]	.64 [2]	.66 [2]	.70 [2]	.71 [2]	.59 [2]				
11. Selectie	.91	.34 [2]	.47 [2]	.48 [2]	.55 [2]	.55 [2]	.50 [2]	.68 [2]	.70 [2]	.68 [2]	.65 [2]			
12. Leereffect	.86	.36 [2]	.36 [2]	.36 [2]	.53 [2]	.58 [2]	.61 [2]	.65 [2]	.74 [2]	.75 [2]	.60 [2]	.71 [2]		
13. Motivatie	.94	.28 [2]	.32 [2]	.49 [2]	.55 [2]	.48 [2]	.45 [2]	.59 [2]	.62 [2]	.55 [2]	.55 [2]	.61 [2]	.63 [2]	
14. Onderwijs	.89	.54 [2]	.44 [2]	.52 [2]	.73 [2]	.71 [2]	.63 [2]	.65 [2]	.62 [2]	.58 [2]	.68 [2]	.52 [2]	.64 [2]	.58 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 6.5: Regressie bij studenten op de kwaliteitscriteria voor de algemene beoordeling

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.01	.01	.06	-.05	-.03	-.05
01. Geschikt	.03	.01	.11	-.06	-.14	-.04
02. Vergelijk	.05	.10	.13	.02	-.15	.05
03. Herhaalbaar	.05	.04	.22	-.09	-.02	-.03
04. Transparant	.02	-.11	-.02	-.06	-.02	.01
05. Acceptatie	.01	.06	.05	-.03	-.04	.07
06. Authentiek	.03	.05	.03	.03	-.09	.11
07. Eerlijk	.01	.04	.04	.04	.02	-.02
08. Complex	.03	.01	.03	-.09	.02	-.12
09. Betekenis	.03	-.03	-.04	-.03	.05	-.17
10. Tijd/Kosten	.03	.08	.00	-.12	.01	-.05
11. Selectie	.01	-.04	.06	.02	-.04	-.05
12. Leereffect	.03	-.01	.04	.05	.05	-.14
13. Motivatie	.04	-.08	-.03	-.15	.05	-.08

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat studenten vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.6: Regressie bij begeleiders op de kwaliteitscriteria voor de algemene beoordeling

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.02	-.12	-.08	.08
01. Geschikt	.00	.06	.02	-.01
02. Vergelijk	.04	.01	.02	.19
03. Herhaalbaar	.04	-.13	-.18	.07
04. Transparant	.01	-.06	.06	.04
05. Acceptatie	.02	-.10	.01	.04
06. Authentiek	.05	-.20	.01	.06
07. Eerlijk	.03	-.12	-.12	.14
08. Complex	.05	-.14	-.03	.18
09. Betekenis	.03	-.05	-.15	.18
10. Tijd/Kosten	.03	-.12	-.13	-.07
11. Selectie	.07	-.11	-.24	-.05
12. Leereffect	.06	-.20	-.08	.14
13. Motivatie	.01	-.01	.02	-.10
14. Onderwijs	.03	-.14	.09	-.09

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.7: Cronbach alfas van en correlaties tussen de schalen voor competentiegroei bij studenten

Competenties	Alfa	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht	.74											
02. Zelfmanagement	.78	.66 [2]										
03. Zelfreflectie	.82	.63 [2]	.61 [2]									
04. Doelgericht	.80	.53 [2]	.68 [2]	.67 [2]								
05. Kansen zien	.77	.49 [2]	.53 [2]	.58 [2]	.61 [2]							
06. Leidinggeven	.77	.22 [1]	.40 [2]	.33 [2]	.50 [2]	.53 [2]						
07. Vernieuwen	.74	.39 [2]	.42 [2]	.60 [2]	.66 [2]	.51 [2]	.42 [2]					
08. Samenwerken	.82	.41 [2]	.63 [2]	.46 [2]	.63 [2]	.59 [2]	.61 [2]	.62 [2]				
09. Feedback geven	.79	.42 [2]	.49 [2]	.41 [2]	.50 [2]	.50 [2]	.55 [2]	.53 [2]	.64 [2]			
10. Feedback vragen	.87	.42 [2]	.40 [2]	.60 [2]	.57 [2]	.48 [2]	.18	.67 [2]	.45 [2]	.44 [2]		
11. Klantgericht werken	.86	.42 [2]	.42 [2]	.54 [2]	.53 [2]	.39 [2]	.29 [2]	.67 [2]	.55 [2]	.53 [2]	.66 [2]	
12. Voor groep spreken	.85	.29 [2]	.35 [2]	.40 [2]	.46 [2]	.51 [2]	.49 [2]	.54 [1]	.55 [2]	.51 [2]	.46 [2]	.60 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 6.8: Correlaties tussen de items voor competentiegroei bij begeleiders

Competenties	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht											
02. Zelfmanagement	.48 [2]										
03. Zelfreflectie	.33 [1]	.25									
04. Doelgericht	.39 [1]	.48 [2]	.53 [2]								
05. Kansen zien	.36 [1]	.45 [2]	.32 [1]	.80 [2]							
06. Leidinggeven	.29	.22	.11	.44 [2]	.39 [1]						
07. Vernieuwen	.18	.28	.18	.60 [2]	.57 [2]	.54 [2]					
08. Samenwerken	.19	.12	.24	.51 [2]	.35 [1]	.52 [2]	.46 [2]				
09. Feedback geven	.15	.22	.36 [2]	.37 [1]	.07	.29	.31 [1]	.26			
10. Feedback vragen	.14	.20	.25	.43 [2]	.20	.35 [1]	.25	.22	.58 [2]		
11. Klantgericht werken	.40 [2]	.24	.22	.63 [2]	.47 [2]	.59 [2]	.63 [2]	.53 [2]	.39 [1]	.52 [2]	
12. Voor groep spreken	.19	.26	.15	.37 [1]	.24	.57 [2]	.64 [2]	.53 [2]	.49 [2]	.44 [2]	.63 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 6.9: Regressie bij studenten op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.14 [1]	.15	-.12	-.20	-.01	.23 [1]
01. Praktijkgericht	.18 [2]	.26 [1]	.15	-.05	-.26 [1]	.17
02. Zelfmanagement	.10	.26 [1]	-.04	-.04	-.11	.12
03. Zelfreflectie	.01	.06	.01	-.03	-.05	-.02
04. Doelgericht	.15 [1]	.18	.05	-.29 [2]	-.13	.17
05. Kansen zien	.17 [1]	.25 [1]	-.18	-.21 [1]	.07	.12
06. Leidinggeven	.23 [2]	.02	-.17	-.29 [2]	.03	.34 [2]
07. Vernieuwen	.12	.16	-.05	-.23 [1]	.13	.14
08. Samenwerken	.10	.05	-.10	-.17	-.03	.24 [1]
09. Feedback geven	.18 [2]	-.01	-.17	-.12	-.04	.36 [2]
10. Feedback vragen	.03	.11	-.04	-.08	.06	.09
11. Klantgericht werken	.05	.06	-.16	-.08	.05	.07
12. Voor groep spreken	.16 [1]	.04	-.26 [1]	-.18	.04	.22 [1]

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.10: Regressie bij begeleiders op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.06	.13	.19	.09
01. Praktijkgericht	.05	.03	.22	-.20
02. Zelfmanagement	.16	.24	.15	-.34
03. Zelfreflectie	.15	.36 [1]	.04	.19
04. Doelgericht	.04	.15	.04	.14
05. Kansen zien	.02	.12	.07	-.00
06. Leidinggeven	.06	-.06	.23	-.03
07. Vernieuwen	.04	-.05	.02	.16
08. Samenwerken	.08	.22	-.10	.22
09. Feedback geven	.06	.09	.27	-.07
10. Feedback vragen	.07	.21	.20	.04
11. Klantgericht werken	.15	.05	.20	.26
12. Voor groep spreken	.06	-.01	.19	.08

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 6.11: Gemiddelden per rol op de kwaliteitscriteria voor de specifieke interventie

Criteria	Iedereen	Studenten	Begeleiders	ΔBS
00. Algemeen	0.53 (1.35) [2]	0.33 (1.34) [0]	1.19 (1.21) [2]	1
01. Geschikt	0.45 (1.61) [2]	0.16 (1.50) [0]	1.40 (1.64) [2]	2
02. Vergelijk	0.59 (1.57) [2]	0.40 (1.52) [0]	1.24 (1.59) [2]	0
03. Herhaalbaar	0.39 (2.29) [0]	0.33 (2.38) [0]	0.58 (2.03) [0]	0
04. Transparant	0.47 (1.57) [1]	0.39 (1.45) [0]	0.73 (1.94) [0]	0
05. Acceptatie	0.58 (1.49) [2]	0.45 (1.45) [1]	1.00 (1.60) [1]	0
06. Authentiek	0.38 (1.59) [0]	0.18 (1.52) [0]	1.04 (1.68) [1]	0
07. Eerlijk	0.58 (1.57) [2]	0.47 (1.55) [1]	0.98 (1.63) [1]	0
08. Complex	0.69 (1.63) [2]	0.48 (1.60) [1]	1.38 (1.57) [2]	0
09. Betekenis	0.84 (1.83) [2]	0.39 (1.84) [0]	2.31 (0.79) [2]	2
10. Tijd/Kosten	0.47 (1.63) [1]	0.20 (1.64) [0]	1.38 (1.28) [2]	1
11. Selectie	0.83 (1.52) [2]	0.66 (1.58) [2]	1.38 (1.19) [2]	0
12. Leereffect	0.68 (1.67) [2]	0.36 (1.68) [0]	1.76 (1.14) [2]	2
13. Motivatie	-0.03 (1.59) [0]	-0.16 (1.65) [0]	0.38 (1.35) [0]	0
14. Onderwijs	1.42 (1.15) [2]		1.81 (0.94) [2]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Iedereen" is met een subscript aangegeven of het gemiddelde voor alle deelnemers verschilt van 0. Ook in de kolommen "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 6.12: Gemiddelden per conditie en per rol op de kwaliteitscriteria voor de algemene beoordeling

	Controleconditie				Experimentele conditie			
Criteria	Iedereen	Studenten	Begeleiders	ΔBS	Iedereen	Studenten	Begeleiders	ΔBS
00. [2] Algemeen	1.98 (0.68) [2]	1.98 (0.63) [2]	1.99 (0.79) [2]	0	1.25 (0.96) [2]	1.01 (0.90) [2]	1.80 (0.88) [2]	2
01. [1] Geschikt	2.12 (0.77) [2]	2.09 (0.75) [2]	2.19 (0.83) [2]	0	1.79 (1.06) [2]	1.63 (1.07) [2]	2.17 (0.95) [2]	1
02. [2] Vergelijk	2.21 (0.80) [2]	2.20 (0.81) [2]	2.23 (0.82) [2]	0	1.54 (1.24) [2]	1.21 (1.25) [2]	2.33 (0.80) [2]	2
03. [2] Herhaalbaar	1.41 (1.25) [2]	1.54 (1.09) [2]	1.14 (1.54) [2]	0	0.63 (1.46) [2]	0.40 (1.41) [2]	1.19 (1.46) [2]	1
04. [2] Transparant	1.82 (1.16) [2]	1.86 (0.96) [2]	1.72 (1.52) [2]	0	1.20 (1.28) [2]	0.98 (1.25) [2]	1.71 (1.23) [2]	1
05. [2] Acceptatie	2.11 (0.86) [2]	2.15 (0.86) [2]	2.00 (0.88) [2]	0	1.51 (1.18) [2]	1.32 (1.19) [2]	1.97 (1.05) [2]	1
06. [2] Authentiek	2.15 (0.73) [2]	2.11 (0.73) [2]	2.23 (0.76) [2]	0	1.71 (1.08) [2]	1.49 (1.10) [2]	2.01 (1.22) [2]	0
07. [2] Eerlijk	2.04 (0.82) [2]	1.94 (0.77) [2]	2.24 (0.88) [2]	0	1.33 (1.28) [2]	1.07 (1.22) [2]	1.95 (1.23) [2]	2
08. [2] Complex	1.88 (0.80) [2]	1.87 (0.79) [2]	1.91 (0.85) [2]	0	0.86 (1.43) [2]	0.59 (1.39) [2]	1.49 (1.36) [2]	1
09. [2] Betekenis	2.28 (0.74) [2]	2.20 (0.79) [2]	2.46 (0.62) [2]	0	1.50 (1.58) [2]	1.15 (1.52) [2]	2.33 (1.41) [2]	2
10. [2] Tijd/Kosten	1.89 (0.93) [2]	1.81 (0.88) [2]	2.05 (1.03) [2]	0	1.01 (1.25) [2]	0.77 (1.25) [2]	1.59 (1.06) [2]	1
11. [2] Selectie	2.01 (0.83) [2]	1.94 (0.78) [2]	2.14 (0.94) [2]	0	1.13 (1.37) [2]	0.82 (1.44) [2]	1.87 (0.85) [2]	2
12. [2] Leereffect	2.10 (0.72) [2]	2.14 (0.65) [2]	2.01 (0.87) [2]	0	1.42 (1.23) [2]	1.23 (1.24) [2]	1.89 (1.11) [2]	1
13. [2] Motivatie	1.78 (1.08) [2]	1.93 (1.09) [2]	1.50 (1.03) [2]	0	0.63 (1.57) [2]	0.51 (1.63) [2]	0.92 (1.41) [2]	0
14. [2] Onderwijs	1.79 (0.80) [2]		1.79 (0.80) [2]		1.33 (1.20) [2]		1.33 (1.20) [2]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Criteria" is met een subscript voor de naam van het criterium aangegeven of het gemiddelde van alle deelnemers verschilt over de condities heen. In de kolommen "Iedereen", "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 6.13: Gemiddelden per conditie en per rol op de schalen voor competentiegroei

			Controleconditie			Experimentele conditie		
Competenties	ΔB	ΔS	Begeleiders	Studenten	ΔBS	Begeleiders	Studenten	ΔBS
00. Algemeen	0	0	4.41 (0.48) [2]	4.04 (0.48) [2]	2	4.43 (0.41) [2]	3.93 (0.46) [2]	2
01. Praktijkgericht	0	0	4.84 (0.38) [2]	4.30 (0.50) [2]	2	4.86 (0.36) [2]	4.30 (0.54) [2]	2
02. Zelfmanagement	0	0	4.88 (0.42) [2]	4.24 (0.53) [2]	2	4.95 (0.22) [2]	4.17 (0.51) [2]	2
03. Zelfreflectie	0	0	4.68 (0.67) [2]	4.28 (0.50) [2]	1	4.81 (0.40) [2]	4.27 (0.61) [2]	2
04. Doelgericht	0	0	4.36 (0.90) [2]	3.95 (0.48) [2]	1	4.38 (0.67) [2]	3.87 (0.52) [2]	2
05. Kansen zien	0	0	4.53 (0.61) [2]	4.05 (0.74) [2]	1	4.48 (0.60) [2]	3.89 (0.63) [2]	2
06. Leidinggeven	0	0	4.32 (0.75) [2]	3.80 (0.70) [2]	1	3.90 (0.94) [2]	3.68 (0.62) [2]	0
07. Vernieuwen	0	0	4.00 (0.88) [2]	3.89 (0.68) [2]	0	4.05 (0.86) [2]	3.83 (0.71) [2]	0
08. Samenwerken	0	0	4.58 (0.61) [2]	4.09 (0.57) [2]	2	4.57 (0.51) [2]	3.98 (0.62) [2]	2
09. Feedback geven	0	0	4.47 (0.70) [2]	3.73 (0.66) [2]	2	4.62 (0.50) [2]	3.71 (0.56) [2]	2
10. Feedback vragen	0	0	4.68 (0.48) [2]	4.22 (0.65) [2]	2	4.67 (0.58) [2]	4.14 (0.87) [2]	1
11. Klantgericht werken	0	0	4.58 (0.61) [2]	4.13 (0.55) [2]	2	4.61 (0.67) [2]	3.91 (0.77) [2]	2
12. Voor groep spreken	0	0	3.11 (1.24) [0]	3.73 (0.82) [2]	1	3.24 (1.09) [0]	3.44 (0.78) [2]	0

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. Als de gemiddelden afwijken van 3 (= het neutraal middelpunt), dan is dit aangegeven met een subscript. Of de gemiddelden van begeleiders verschillen over condities heen, is weergegeven met een cijfer in de kolom "ΔB". Of de gemiddelden van studenten verschillen over condities heen, is weergegeven met een cijfer in de kolom "ΔS". Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

7 STUDIE 3

Voor academiejaar 2017-2018 werd voor het laatste jaar van de opleiding Accountancy & Fiscaliteit (AFI) gekeken op welke manier de stagebeoordeling verbeterd kon worden. Hierbij werd enerzijds vertrokken van ons theoretisch kader en anderzijds vanuit de noden van de opleiding zelf. Gezien de noden van deze opleiding van een andere aard waren dan de noden van de andere bij ons onderzoek betrokken opleidingen, werd voor een andere aanpak gekozen. De noden van deze opleiding hadden betrekking op hoe de soft skills van studenten (i.e., zelfreflectie, het geven en ontvangen van feedback en het ontwikkelen van leerdoelen) in het eindoordeel konden worden betrokken. In samenspraak met de opleiding werden daartoe twee interventies uitgewerkt. De eerste bestond uit een aanpassing van het beoordelingsformulier. Dit formulier werd herwerkt opdat het niet meer zozeer zou polsen naar de kansen die studenten krijgen op hun stage maar naar de mate waarin ze de kansen die er zijn grijpen. Zodoende probeerden we de focus te verschuiven naar het proces dat student doorlopen teneinde de beoordeling een meer procesmatig karakter te geven. Ook soft skills werden nu hierbij in rekening gebracht. De tweede interventie bestond uit het invoeren van leergroepen die inzetten op het versterken van soft skills. Een leergroep bestond uit één KdG-docent en acht studenten. Elke leergroep kwam twee keer samen: één keer rond de kansen die een stage biedt en rond omgaan met stress en conflict, en één keer rond het eindverslag. Omwille van het geringe studentenaantal was het hier niet opportuun om te werken met een experimenteel onderzoeksdesign. Om de effecten van de interventie te kunnen nagaan, werden de studenten en docenten van voorgaand jaar in de mate van het mogelijke als vergelijkingsbasis gebruikt.

Op welke wijze sluiten deze interventies aan bij het theoretisch kader? Deze interventies zetten vooreerst in op "Gebruik meer beoordelingsmethoden". Specifiek wordt er geprobeerd om een ander soort kwalitatieve informatie (met name alles wat te maken heeft met bovengenoemde soft skills) bij de beoordeling te betrekken. Verder zetten deze interventies, en in het bijzonder de tweede interventie, in op "Gebruik meer ratings en observaties". Het is namelijk niet enkel de bedoeling andere informatie in het eindoordeel te betrekken, maar ook om (1) de frequentie van interactie met studenten te verhogen, dit zowel voor docenten (via de leergroepen) als voor mentoren (door zowel studenten als mentoren te stimuleren om vaker feedback te vragen; zie "Laat beoordelaars meer ratings doen") en (2) de input van meer mensen bij de eindbeoordeling te betrekken door studenten ook feedback aan elkaar te laten geven (zie "Laat meer mensen beoordelen" en "Gebruik verschillende types beoordelaars"). Tot slot zet deze interventie ook in op "Organiseer trainingen voor docenten en mentoren". Meer bepaald is het niet enkel de bedoeling om docenten te ondersteunen in het leiden van de leergroepen maar ook om docenten en mentoren aan te moedigen om vaker en beter feedback te geven, en dit niet alleen inzake "klassieke" technische vaardigheden maar ook inzake soft skills.

Conform de nadruk in de literatuur beoogden deze interventies niet enkel een effect te hebben op de kwaliteit van de beoordeling maar ook op de kwaliteit van het stageproces. De interventies werden niet enkel opgezet als beoordeling maar ook als begeleiding. Dit is heel expliciet het geval voor de tweede interventie, waarvan de bevindingen niet rechtstreeks meetelden in het eindoordeel, maar indirect wel zullen gespeeld hebben in de hoofden van de eindbeoordelaars omdat ze zouden moeten hebben gezorgd voor een accurater beeld van het functioneren van de te beoordelen studenten.

7.1 Methode

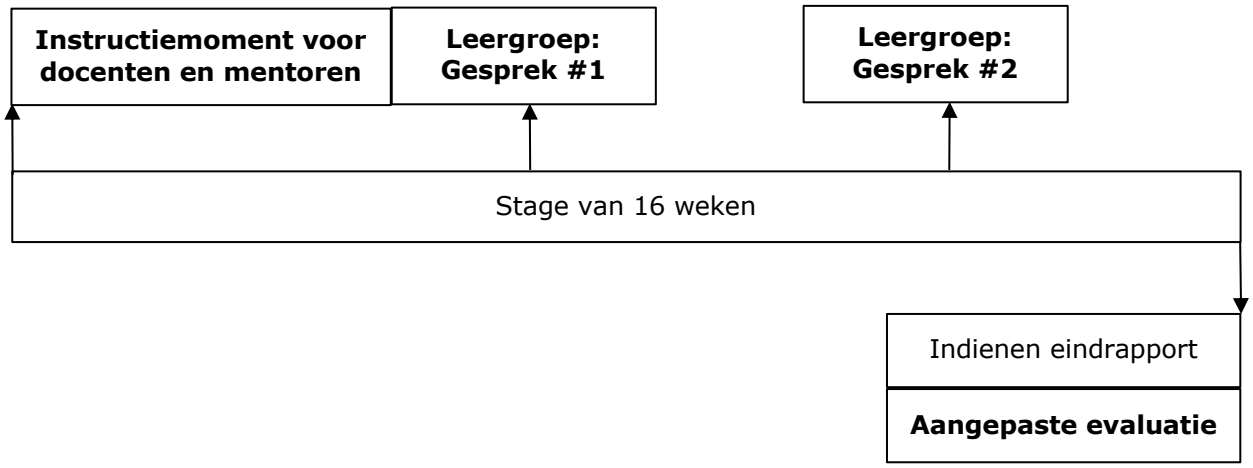
7.1.1 Procedure

De interventie werd uitgewerkt in het kader van de stage in het laatste jaar: een stage van 35 dagen van minimum 7.5 uur binnen een bedrijf, typisch gespreid over 16 weken. Het is niet onze bedoeling om in dit rapport het volledige verloop van het "klassieke traject" te beschrijven. Daarvoor verwijzen we naar het stagevademecum van AFI dat op te vragen valt via e-mail. Het is wel onze bedoeling duidelijk te maken welke wijzigingen in het kader van dit onderzoek werden doorgevoerd. In dit geval ging het om één wijziging (= een nieuw beoordelingsformulier) en uitbreidingen. De uitbreidingen hadden betrekking op de invoering van de hoger beschreven leergroepen. De wijziging en de uitbreidingen t.a.v. het "klassieke traject" worden hieronder beschreven en zijn ook weergegeven in Figuur 7.1.

- Leergroepsgesprek #1 had tot doel zicht te krijgen op de stageplaats op zich, op de kansen die studenten daar krijgen, op wat ze met die kansen doen, en op hoe ze handelen in conflict- of stresssituaties. Deze fase verliep in zes stappen (zie onder). Ter voorbereiding van dit leergroepsgesprek dienden twee verslagen gemaakt te worden. Dit leergroepsgesprek was formatief bedoeld.
    1. Voorbereiding thuis
    2. Presentaties en luistermoment
    3. Inhoudelijke feedback geven aan elkaar
    4. Reageren op deze feedback
    5. Stijlopmerkingen van studenten en docenten
    6. Verwerking thuis: Leerdoelen schriftelijk delen
- Leergroepsgesprek #2 had tot doel feedback te verzamelen op het stagerapport en een blik te werpen op de toekomst. Ook deze fase verliep in zes stappen (zie onder). Ter voorbereiding van dit gesprek diende een voorlopige versie van het eindrapport te worden gemaakt. Dit leergroepsgesprek was formatief bedoeld.
    1. Voorbereiding thuis
    2. Presentaties en luistermoment
    3. Inhoudelijke feedback geven aan elkaar
    4. Reageren op deze feedback
    5. Stijlopmerkingen van studenten en docenten
    6. Verwerking thuis: Definitieve versie eindrapport
- Het definitieve eindrapport telde uiteraard mee voor de eindevaluatie. Die gebeurde aan de hand van een nieuw beoordelingsformulier dat toespitst op het procesmatig karakter van de stage en ook soft skills in rekening brengt.[2]

[2] De instructies die studenten op voorhand op papier kregen in het kader van Stap 1, Stap 3, Stap 5 en Stap 6 van zowel Leergroepsgesprek #1 als Leergroepsgesprek #2 vallen door de geïnteresseerde lezer op te vragen bij Thomas Heynderickx. Hetzelfde geldt voor het oude en het nieuwe beoordelingsformulier.

Figuur 7.1: Het normale verloop van de stage met in het vet de uitbreidingen die in het kader van dit project werden ingevoerd in de experimentele conditie.



Na afloop van de stage kregen alle studenten en begeleiders een elektronische vragenlijst. Deze elektronische vragenlijst bestond uit vier onderdelen. Bij mentoren werd enkel het eerste en het tweede onderdeel bevraagd. Met het derde onderdeel kwamen zij immers niet in aanmerking, en daardoor verwachtten we ook geen verschillen op het vierde onderdeel. Voor het vierde onderdeel hadden we voor mentoren overigens ook geen gegevens uit het verleden.

1. In deze vragenlijst werden voor iedereen eerst een aantal persoonsgegevens (i.e. naam, voornaam en e-mailadres) en achtergrondkenmerken gevraagd (i.e., geslacht en leeftijd). Bij de begeleiders werd additioneel gepolst naar hun ervaring (in termen van het aantal jaar dat ze reeds bij het begeleiden van stages betrokken zijn). Bij de studenten werd additioneel gepolst naar vooropleiding en etniciteit. Vervolgens werd aan studenten ook gevraagd wat volgens hen het punt is dat ze verdienden op hun stage (= een self-assessment). Ter uitdieping van het Selectiecriterium werd na de juni-zittijd aan deze gegevens het effectief verkregen punt toegevoegd.
2. Daarna werd bij de studenten en begeleiders gepolst naar de perceptie van de eerste specifiek aangebrachte wijziging (met name het nieuwe beoordelingsformulier). Bij begeleiders peilden deze vragen naar de gepercipieerde bijdrage van de interventie aan de kwaliteit van de algemene beoordelingsprocedure. Omdat studenten moeilijk kunnen inschatten wat de specifieke bijdrage van een nieuw beoordelingsformulier is, werden deze vragen voor hen lichtjes aangepast en werd gepolst naar de perceptie ervan.
3. Bij de studenten en begeleiders werd vervolgens gepolst naar de perceptie van de tweede specifiek aangebrachte wijziging (met name de leergroepen). Bij begeleiders peilden deze vragen naar de gepercipieerde bijdrage van de interventie aan de kwaliteit van de algemene beoordelingsprocedure. Omdat studenten moeilijk kunnen inschatten wat de specifieke bijdrage van de leergroepen is, werden deze vragen voor hen lichtjes aangepast. Bij studenten peilden deze vragen naar de perceptie van deze interventie.

4. Tot slot werd de competentiegroeimeter voor stage afgenomen ter uitdieping van het Leereffectcriterium. De competentiegroeimeter is een binnen KdG veelgebruikt instrument, gebaseerd op het werk van Braun et al. (2008; Braun & Leidner, 2009). De competentiegroeimeter laat enerzijds toe te identificeren welke competentiegroei begeleiders verwachten en anderzijds of dat zich vertaalt in de beleving van studenten. De competentiegroeimeter werd niet alleen afgenomen bij studenten die hun laatstejaarsstage deden in academiejaar 2017-2018 en hun docenten maar ook bij studenten die hun laatstejaarsstage deden in het voorafgaand academiejaar (2016-2017) en hun docenten. Voor ons onderzoek betekent dit het volgende: als we verschillen vinden tussen de scores van docenten t.a.v. vorig jaar, dan betekent dit dat docenten verwachtten dat studenten andere competenties zouden ontwikkelen omwille van (het procesmatige karakter van) onze interventies. Of dit zo was, kon dan getoetst worden door de scores van de studenten te vergelijken met de scores van studenten van voorgaand jaar.

7.1.2 Steekproef

In totaal werden 68 studenten die hun laatstejaarsstage deden in academiejaar 2017-2018 uitgenodigd om aan ons onderzoek deel te nemen. Van de 68 uitgenodigde studenten namen er 66 deel aan ons onderzoek (= 97.1%). Daarnaast werden ook 71 begeleiders uitgenodigd (11 docenten en 60 mentoren). Hiervan vulden er 22 (= 31.0%) de vragenlijst in: 11 docenten en 11 mentoren. De gemiddelde leeftijd van de deelnemende begeleiders was 41.86 jaar (SD = 9.99). Gemiddeld genomen hadden ze 6.41 jaar ervaring met stagebegeleiding (SD = 5.17). Het overgrote deel waren mannen (63.6%; met het oog op de analyses die nog volgen, werd deze variabele als volgt gecodeerd: 1 voor man en 2 voor vrouw). De studenten bestonden voor 51.5% uit mannen. Hun gemiddelde leeftijd was 22.23 jaar (SD = 1.66). Qua vooropleiding kwamen 3.0% van de studenten uit het BSO, 59.1% uit het TSO, en 37.9% uit het ASO. Op basis van deze gegevens werd een variabele "educatie" geconstrueerd waarop studenten een 1 kregen als ze uit het BSO kwamen, een 2 als ze uit het TSO kwamen of als hier geen duidelijke gegevens over te vinden waren, en een 3 als ze uit het ASO kwamen (zie Voorstudie 1). De restcategorie (= 2) vormt de baseline waartegen effecten van een lage (= BSO) en een hoge educatie (= ASO) worden afgezet. De meeste studenten hebben de Belgische nationaliteit (95.5%), zijn in België geboren (98.5%), en hebben een grootmoeder langs moederszijde die in België geboren is (81.8%). Op basis van deze gegevens werd een variabele "etniciteit" geconstrueerd waarop studenten een 0 kregen als ze de Belgische nationaliteit hebben, in België geboren zijn en een grootmoeder van Belgische nationaliteit hebben en een 1 als aan één van deze voorwaarden niet is voldaan (zie Voorstudie 1). Op deze manier werd 18.2% van de studenten gecategoriseerd als van "een andere etniciteit". De studenten vonden gemiddeld genomen dat ze 15.03 op 20 verdienden op hun stage (SD = 1.55). Eén student gaf aan dat hij niet verdiende te slagen. Die student gaf aan een 9 te verdienen.

7.1.3 Effectmetingen

7.1.3.1 Kwaliteitscriteria voor het beoordelingsformulier

Na afloop van de stage werd bij alle studenten en begeleiders (i.e., docenten en mentoren) gepolst naar de perceptie van het vernieuwde beoordelingsformulier. Dit gebeurde aan de hand van de verkorte versie van de kwaliteitscriteria-vragenlijst. De inleiding werd aangepast aan de

interventie in kwestie. Voor begeleiders werden alle items zo geformuleerd zodat ze zouden polsen naar de gepercipieerde bijdrage van het nieuwe beoordelingsformulier aan de kwaliteit van de algemene beoordelingsprocedure. Voor studenten werden de items zo geformuleerd zodat ze zouden polsen naar de perceptie van het nieuwe beoordelingsformulier. Voor studenten had het immers weinig zin om te polsen naar de gepercipieerde meerwaarde want in tegenstelling tot Studie 1 en 2 ging het hier niet om een mooi geïsoleerde interventie. Deze interventie was veel ingrijpender en minder identificeerbaar. Voor elke interventie werd elk van de 14 hoger vermelde criteria gemeten aan de hand van drie vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). Het criterium Onderwijseffect werd enkel bevraagd bij begeleiders. De vragen voor de studenten zijn te vinden in Appendix 6, die voor begeleiders in Appendix 7.

De Cronbach alfas en de correlaties tussen de schalen zijn te vinden in Tabel 7.1. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen heel wat schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Die kon 58% van de totale variantie verklaren. Alle schalen hadden een factorlading > .40. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze had een Cronbach alfa van .93.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 7.2. Noch de algemene score noch de onderliggende schalen konden significant worden voorspeld op basis van de beschikbare gegevens. De resultaten voor begeleiders zijn te vinden in Tabel 7.3. Noch de algemene score noch de onderliggende schalen, met uitzondering van Tijd & Kosten en Selectie, konden significant worden voorspeld op basis van de beschikbare gegevens. Zowel de haalbaarheid van de door de interventie vereiste Tijd en Kosten als Selectie werd significant hoger ingeschat door jonge, vrouwelijke begeleiders, al moet deze nuance met voorzichtigheid bekeken worden omwille van het kleine deelnemersaantal.

7.1.3.2 Kwaliteitscriteria voor de leergroepen

Na afloop van de stage werd bij alle studenten en begeleiders (i.e., docenten) aan de hand van de verkorte versie van de kwaliteitscriteria-vragenlijst gepolst naar de perceptie van de leergroepen. De inleiding werd aangepast aan de interventie in kwestie. Voor studenten werden de items zo geformuleerd zodat ze zouden polsen naar de perceptie van de leergroepen. Voor begeleiders werden alle items zo geformuleerd zodat ze zouden polsen naar de gepercipieerde bijdrage van de leergroepen aan de kwaliteit van de algemene beoordelingsprocedure. Voor studenten had het immers weinig zin om te polsen naar de gepercipieerde meerwaarde want in tegenstelling tot Studie 1 en 2 ging het hier niet om een mooi geïsoleerde interventie. Deze

interventie was veel ingrijpender en minder identificeerbaar. Voor elke interventie werd elk van de 14 hoger vermelde criteria gemeten aan de hand van drie vragen die dienden te worden beantwoord op een 7-punten Likertschaal ("-3 = helemaal oneens" tot "+3 = helemaal eens"). Het criterium Onderwijseffect werd enkel bevraagd bij begeleiders. De vragen voor de studenten zijn te vinden in Appendix 6, die voor begeleiders in Appendix 7.

De Cronbach alfas en de correlaties tussen de schalen zijn te vinden in Tabel 7.4. In deze tabel werd geen onderscheid gemaakt tussen studenten en begeleiders, met uitzondering voor Onderwijseffect. Deze schaal werd immers enkel afgenomen bij begeleiders. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen heel wat schalen, besloten we via factoranalyse te onderzoeken of het aantal schalen gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. We lieten hierbij de schaal Onderwijseffect buiten beschouwing. De scree plot van een exploratorische factoranalyse wees op een oplossing met één factor. Die kon 49% van de totale variantie verklaren. Alle schalen behalve Herhaalbaarheid hadden een factorlading > .40. Omdat een dergelijke factor handig is om een overzicht te krijgen van de algemene perceptie van de interventie, berekenden we een algemene score (= Algemeen) door alle schalen te middelen. Deze algemene score had een Cronbach alfa van .90.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., geslacht, leeftijd, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Voor begeleiders was dit enkel basis van achtergrondkenmerken (i.e., geslacht, leeftijd en ervaring). De resultaten voor studenten zijn te vinden in Tabel 7.5. Noch de algemene score noch de onderliggende schalen, met uitzondering van Herhaalbaarheid en Motivatie, konden significant worden voorspeld op basis van de beschikbare gegevens. Herhaalbaarheid werd hoger ingeschat door mannelijke studenten en Motivatie door studenten die van zichzelf vonden dat ze hun stage erg goed deden. De resultaten voor begeleiders zijn te vinden in Tabel 7.6. Noch de algemene score noch de onderliggende schalen, met uitzondering van Authenticiteit en Tijd & Kosten, konden significant worden voorspeld op basis van de beschikbare gegevens. De authenticiteit werd hoger ingeschat door begeleiders met minder ervaring. De haalbaarheid van de door de interventie vereiste Tijd en Kosten werd eveneens hoger ingeschat door begeleiders met minder ervaring.

7.1.3.3 Uitdieping van het Leereffectcriterium

Na afloop van de stage werd ter uitdieping van het Leereffectcriterium bij alle studenten die hun laatstejaarsstage deden in academiejaar 2017-2018 en hun docenten de competentiegroeimeter voor stage afgenomen. De gebruikte versie van de competentiegroeimeter beoogde competentiegroei te meten in 12 domeinen: (1) Praktijkgericht, (2) Zelfmanagement, (3) Zelfreflectie, (4) Doelgericht werken, (5) Kansen zien en nemen, (6) Leidinggeven en activeren, (7) Vernieuwen, (8) Samenwerken, (9) Feedback geven, (10) Feedback vragen en ontvangen, (11) Klantgericht werken, en (12) Voor een groep spreken. De competentiegroeimeter voor stage bestaat in twee versies: een studenten- en een docentenversie. De docentenversie laat toe te identificeren welke competentiegroei begeleiders verwachten. In deze versie wordt elke concept gemeten aan de hand van één item. De studentenversie laat toe te identificeren welke competentiegroei studenten rapporteren. In deze versie wordt elke concept gemeten aan de hand van

meerdere items (van 3 tot 7 al naargelang het concept). Alle vragen dienden te worden beantwoord op een 5-punten Likertschaal ("1 = helemaal oneens" tot "5 = helemaal eens"). De docentenversie is te vinden in Appendix 8 en de studentenversie in Appendix 9.

Om te weten te komen of de nadruk op soft skills die de interventies in deze opleiding met zich mee brachten zich ook zou vertalen in (1) andere verwachtingen met betrekking tot competentiegroei door docenten en (2) andere gerapporteerde competentiegroei door studenten, werd de competentiegroeimeter ook reeds afgenomen bij studenten (N = 42) en docenten (N = 8) in het academiejaar 2016-2017. De competentie Voor een groep spreken werd toen niet gemeten omdat dit toen niet relevant was. Dit was namelijk een competentie die niet aan bod kwam. Bij mentoren werd de competentiegroeimeter toen niet afgenomen. Omdat we voor hen niet beschikten over gegevens uit het academiejaar 2016-2017, werden mentoren ook dit jaar niet bevraagd. Mentoren kwamen trouwens ook niet in aanmerking met de leergroepen. Daardoor verwachtten we ook niet dat zij nu een andere competentiegroei zouden verwachten. In totaal werd de competentiegroeimeter dus ingevuld door 108 studenten en 19 docenten.

In 2017-2018 vulden bijna alle studenten de vragenlijst in, maar in 2016-2017 was dit niet zo. Van de 69 ingeschreven studenten vulden er slechts 42 deze vragenlijst in (= 60.9%). De studenten die de vragenlijst niet invulden verschilden niet van de studenten die de vragenlijst wel invulden in termen van geslacht (1.41 en 1.57; (F (1, 67) = 1.76, ns), leeftijd (Mean = 22.11 en 21.83; (F (1, 67) = 0.50, ns), etniciteit (Mean = 2.33 en 2.29; (F (1, 67) = 0.15, ns) en educatie (0.07 en 0.10; (F (1, 67) = 0.09, ns). Er is dus geen aanwijzing dat de studenten die de vragenlijst invulden niet representatief zouden geweest zijn voor de populatie. Hieruit kunnen we afleiden dat het OK is om de deelnemende studenten van het academiejaar 2017-2018 te vergelijken met die van het academiejaar 2016-2017.

De Cronbach alfas en de correlaties tussen de schalen van de studentenversie zijn te vinden in Tabel 7.7. De correlaties tussen de items van de docentenversie zijn te vinden in Tabel 7.8. Alle schalen hadden een goede interne consistentie (Cronbach alfa > .70). Gezien de substantiële correlaties tussen heel wat concepten, besloten we via factoranalyse te onderzoeken of het aantal schalen / items gereduceerd kon worden tot een kleiner aantal betekenisvolle factoren. Voor de studentenversie kon een oplossing met één factor 46% van de totale variantie verklaren. Alle schalen behalve Feedback vragen hadden een factorlading > .40 op deze factor. Omdat zo'n factor handig is om een overzicht te krijgen van de algemene competentiegroei, berekenden we een algemene competentiegroei-score door alle schalen te middelen. Deze algemene score had een Cronbach alfa van .88. Voor de docentenversie kon een oplossing met één factor 44% van de totale variantie verklaren. Alle items behalve Vernieuwen hadden een factorlading > .40 op deze factor. Deze algemene score had een Cronbach alfa van .85.

Via regressieanalyse gingen we na of deze algemene score alsook de score op de onderliggende schalen / items kon worden voorspeld op basis van de beschikbare gegevens. Voor studenten was dit op basis van achtergrondkenmerken (i.e., leeftijd, geslacht, educatie en etniciteit) en het punt dat ze zeggen te verdienen voor hun stage. Omdat we niet over al deze gegevens beschikten voor de studenten van het jaar 2016-2017 beperkten we ons tot de studenten van het jaar 2017-2018. Voor docenten was dit enkel basis van achtergrondkenmerken (i.e., leeftijd, geslacht en ervaring). Omdat we niet over al deze gegevens beschikten voor de docenten

van het jaar 2016-2017 beperkten we ons tot de docenten van het jaar 2017-2018. De resultaten voor studenten zijn te vinden in Tabel 7.9. Enkel Vernieuwen en Feedback geven kon deels worden voorspeld op basis van de beschikbare gegevens. Competentiegroei op vlak van Vernieuwen werd vooral gerapporteerd door jonge studenten die vonden dat ze het goed deden. Competentiegroei in Feedback geven werd dan weer vooral gerapporteerd door mannelijke studenten. De resultaten voor docenten zijn te vinden in Tabel 7.10. Enkel Praktijkgericht kon deels worden voorspeld op basis van de beschikbare gegevens. De minst ervaren docenten verwachtren meest competentiegroei in de Praktijkgerichtheid van studenten.

7.1.3.4 Uitdieping van het Selectiecriterium

Over beide academiejaren heen (2016-2017 en 2017-2018) haalden de studenten gemiddeld 14.56 op 20 op hun stage (SD = 1.92). Geen enkele student slaagde niet. Via regressieanalyse gingen we bij studenten van na of het behaalde punt kon worden voorspeld op basis van de achtergrondkenmerken (i.e., leeftijd, geslacht, educatie en etniciteit). Dit was niet het geval ($F(4, 103) = 0.30$, ns). De score die studenten behaalden op hun stage was sterk gecorreleerd met de algemene score voor competentiegroei ($r = .29$, $p < .01$). De correlaties tussen het behaalde resultaat en de onderliggende schalen voor competentiegroei waren .28 voor Praktijkgericht, .17 voor Zelfmanagement, .18 voor Zelfreflectie, .20 voor Doelgericht werken, .22 voor Kansen zien en nemen, .13 voor Leidinggeven en activeren, .14 voor Vernieuwen, .31 voor Samenwerken, .10 voor Feedback geven, .20 voor Feedback vragen en ontvangen, .33 voor Klantgericht werken, en .18 voor Voor een groep spreken. Alle vernoemde correlaties met een absolute waarde groter dan .20 zijn significant op niveau $p < .05$ en alle vernoemde correlaties met een absolute waarde groter dan .25 zijn significant op niveau $p < .01$.

Uit de gegevens van studenten van 2017-2018 bleek dat de score die studenten behaalden op hun stage niet significant gecorreleerd was met het punt dat ze zegden te verdienen voor hun stage ($r = .22$, ns), wat erop wijst dat studenten hun eigen functioneren niet erg goed kunnen inschatten. Of toch in de zin dat de rangorde die ze op die manier spontaan maakten behoorlijk afwijkt van de rangorde die naar voor kwam uit de stagebeoordeling. Wel schatten ze gemiddeld genomen de hoogte van de punten goed in. Het is hier dus niet zo dat studenten hun eigen prestatie systematisch positiever inschatten dan hun beoordelaars. Een paired samples t-test toonde aan dat er geen significant verschil was tussen het gemiddeld punt dat studenten vonden dat ze verdienden en het gemiddeld punt dat ze behaalden (Mean = 15.03 en 14.71, $t = 1.19$, ns). De score die studenten behaalden op hun stage was niet significant gecorreleerd met de algemene score op de kwaliteitscriteria voor het nieuwe beoordelingsformulier ($r = .15$, ns) en ook niet met de algemene score op de kwaliteitscriteria voor de leergroepen ($r = .03$, ns). Voor het nieuwe beoordelingsformulier waren de correlaties tussen de eindscore en de individuele kwaliteitscriteria .13 voor Geschiktheid, .16 voor Vergelijkbaarheid, -.05 voor Herhaalbaarheid, .19 voor Transparantie, .25 voor Acceptatie, .19 voor Authenticiteit, .16 voor Eerlijkheid, .07 voor Complexiteit, .18 voor Betekenisvolheid, .03 voor Tijd & Kosten, .30 voor Selectie, .10 voor Leereffect, en -.17 voor Motivatie. Voor de leergroepen waren de correlaties tussen de eindscore en de individuele kwaliteitscriteria -.09 voor Geschiktheid, .23 voor Vergelijkbaarheid, -.18 voor Herhaalbaarheid, .14 voor Transparantie, .00 voor Acceptatie, -.01 voor Authenticiteit, .06 voor Eerlijkheid, .19 voor Complexiteit, -.20 voor Betekenisvolheid, .01 voor Tijd & Kosten, .04 voor Selectie, .11 voor Leereffect, en .10 voor Motivatie. Alle vernoemde correlaties met een absolute waarde van minstens .30 zijn significant op niveau $p < .05$.

7.2 Verwachtingen

We verwachtten dat onze interventies zich zouden vertalen in meer tevredenheid over bepaalde aspecten van het beoordelingsproces. We verwachtten dat studenten hier voor een aantal criteria positief tegenover zouden staan. Van begeleiders verwachtten we dan weer dat ze voor die criteria een meerwaarde zouden zien in de interventies. Wat het nieuwe beoordelingsformulier betreft, dachten we in eerste instantie aan de criteria Geschiktheid (het nieuwe formulier wordt geacht een beter beeld te geven van de competenties die studenten dienen te ontwikkelen), Authenticiteit (het nieuwe formulier wordt geacht beter overeen te stemmen met de toekomstige beroepssituatie en wat daarin van studenten verwacht zal worden), Betekenisvolheid (het nieuwe formulier wordt geacht meer waardevolle feedback op te leveren), Selectie (de leergroepen zouden de selectie van goede studenten ten goede moeten komen) en Leereffect (het nieuwe formulier wordt geacht de selectie van goede studenten te vergemakkelijken en een groter leereffect te ressorteren). Wat de leergroepen betreft, dachten we in eerste instantie aan de criteria Geschiktheid (de leergroepen zouden een beter beeld moeten geven van de competenties die studenten dienen te ontwikkelen), Vergelijkbaarheid (de leergroepen zouden de vergelijkbaarheid van het oordeel moeten vergroten), Herhaalbaarheid (de leergroepen zorgen ervoor dat er meer assessmentsituaties ontstaan), Acceptatie (de leergroepen zouden de acceptatie van de beoordelingscriteria moeten vergroten), Authenticiteit (de leergroepen zouden ervoor moeten zorgen dat de beoordeling beter overeen stemt met de toekomstige beroepssituatie), Eerlijkheid (de leergroepen zouden studenten een eerlijkere kans moeten bieden om hun competenties te tonen), Complexiteit (de leergroepen zouden meer zicht moeten verschaffen op de cognitieve vaardigheden van studenten), Betekenisvolheid (de leergroepen en de bijhorende instructies aan mentoren zouden voor extra betekenisvolle feedback moeten zorgen, niet alleen van docenten maar ook van mentoren en medestudenten), Selectie (de leergroepen zouden de selectie van goede studenten ten goede moeten komen), Leereffect (de leergroepen zouden voor een extra leereffect moeten zorgen), en Motivatie (studenten taakgericht laten interageren, zou hun motivatie ten goede moeten komen).

Wat het leereffect betreft, verwachtten we dat de interventies (en dan in eerste instantie de invoering van de leergroepen) een bijdrage zouden leveren aan alle competenties die te maken hebben met sociale interactie en zelfsturing. We dachten hierbij aan Praktijkgerichtheid, Zelfmanagement, Zelfreflectie, Doelgericht werken, Kansen zien en nemen, Samenwerken, Feedback geven, Feedback vragen, Klantgericht werken, en Voor een groep spreken. In 2017-2018 verwachtten we meer groei in bovenvermelde competenties dan in 2016-2017.

Wat selectie betreft, verwachtten we dat interventies de selectie van studenten ten goede zou komen. Dit zou zich dan moeten vertalen in de punten die studenten kregen op hun stage. In 2017-2018 verwachtten we een grotere spreiding van de punten dan in 2016-2017.

7.3 Resultaten

7.3.1 Kwaliteitscriteria voor het beoordelingsformulier

Voor de deelnemers uit 2017-2018 zijn de gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor deze specifieke interventie te vinden in Tabel 7.11 en dit per rol (i.e., Student en Begeleider). T-tests wezen uit dat het algemeen gemiddelde voor alle deelnemers voor de

meeste criteria significant verschilde van 0 (= het neutraal middelpunt van de schaal) in positieve richting. De enige uitzondering werd gevormd door het criterium Motivatie. Voor dit criterium verschilde het gemiddelde niet van 0. Voor de meeste criteria zagen betrokkenen dus gemiddeld genomen de meerwaarde van het nieuwe beoordelingsformulier.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren van elkaar verschilden betreffende hun perceptie van de interventie. Uit deze analyses bleek dat studenten voor een heel aantal criteria positiever tegenover de interventie stonden dan begeleiders. Dit uitte zich zowel voor het algemeen oordeel als voor wat betreft de criteria Vergelijkbaarheid, Transparantie, Acceptatie, Eerlijkheid, Complexiteit, Betekenisvolheid en Leereffect. Voor docenten was het zelfs zo dat de scores voor Herhaalbaarheid, Complexiteit, Betekenisvolheid en Leereffect niet significant verschilden van 0. Begeleiders staan op deze vlakken dus vrij neutraal tegenover het nieuwe beoordelingsformulier. Er waren echter ook heel wat criteria waarvoor begeleiders wel duidelijk de meerwaarde van dit nieuwe beoordelingsformulier inzagen.

7.3.2 Kwaliteitscriteria voor de leergroepen

Voor de deelnemers uit 2017-2018 zijn de gemiddelden en standaarddeviaties voor de kwaliteitscriteria voor deze specifieke interventie te vinden in Tabel 7.11 en dit per rol (i.e., Student en Begeleider). T-tests wezen uit dat het algemeen gemiddelde voor alle deelnemers voor de meeste criteria significant verschilde van 0 (= het neutraal middelpunt van de schaal) in positieve richting. De enige uitzondering werd gevormd door het criterium Herhaalbaarheid. Voor dit criterium verschilde het gemiddelde niet van 0. Voor de meeste criteria zagen betrokkenen dus gemiddeld genomen een meerwaarde in de leergroepen.

Via Oneway Anova's onderzochten we vervolgens of de verschillende actoren van elkaar verschilden betreffende hun perceptie van de interventie. Uit deze analyses bleek dat studenten voor sommige criteria positiever tegenover de interventie stonden dan begeleiders (i.e. docenten). Dit uitte zich vooral voor Transparantie, Acceptatie en Eerlijkheid. Begeleiders schatten de Transparantie, Acceptatie en Eerlijkheid van de leergroepen lager in dan studenten. Voor begeleiders was het zelfs zo dat de scores voor deze criteria niet significant verschilden van 0. De begeleiders stonden op deze vlakken dus vrij neutraal tegenover de leergroepen.

7.3.3 Uitdieping van het Leereffectcriterium

De gemiddelden en standaarddeviaties op de competenties uit de competentiegroeimeter zijn te vinden in Tabel 7.13, en dit zowel per jaar (i.e., 2016-2017 en 2017-2018) als per rol (i.e., docent en student). Voor docenten wezen t-tests uit dat de algemene score in 2016-2017 niet significant verschilde van 3 (= het neutraal middelpunt van de schaal). Docenten verwachtten enkel competentiegroei voor wat betreft Praktijkgerichtheid, Zelfmanagement, Zelfreflectie en Samenwerken. Voor deze competenties week de score in positieve richting af van het neutraal middelpunt van de schaal. Voor 2017-2018 verschilde de algemene score wel significant van het neutraal middelpunt van de schaal, en dit in positieve richting. Docenten verwachtten nu niet alleen méér compentiegroei op valk van Zelfreflectie, maar verwachtten nu ook competentiegroei op vlak van Doelgerichtheid, Kansen zien, en Feedback geven en Feedback vragen. Voor studenten wezen t-tests uit dat de algemene score in 2016-2017 significant verschilde van 3 (= het neutraal middelpunt van de schaal) en dit in positieve richting. Studenten rapporteerden competentiegroei in alle domeinen behalve Leidinggeven, Vernieuwen en Feedback geven.

In 2017-2018 rapporteerden de studenten significant meer competentiegroei dan in 2016-2017, en dit voor alle domeinen. Ook op vlak van Leidinggeven, Vernieuwen en Feedback geven rapporteerden ze nu dus een significante competentiegroei.

Voor zowel 2016-2017 als 2017-2018 wezen t-tests uit dat de competentiegroei die begeleiders verwachtten op vlak van Praktijkgerichtheid significant groter was dan de competentiegroei die studenten rapporteerden. Hetzelfde gold voor Feedback geven in 2017-2018. Op vlak van Leidinggeven, Vernieuwen en Klantgericht werken rapporteerden studenten in 2017-2018 dan weer meer competentiegroei dan de docenten verwachtten.

7.3.4 Uitdieping van het Selectiecriterium

Teneinde na te gaan of de interventie een effect had op het Selectiecriterium, werden de punten die studenten behaalden op hun stage in 2017-2018 (= na invoering van de interventies) vergeleken met de punten die studenten behaalden op hun stage in 2016-2017. Hierbij werd zowel gekeken naar het gemiddelde als naar de spreiding van de punten (= de standaarddeviatie). Een Oneway Anova wees uit dat de studenten uit 2016-2017 niet significant verschilde van de studenten uit 2017-2018 in termen van gemiddelde score (Mean = 14.13 en 14.71; $F(1, 106) = 1.13$, ns). Een Levene's test (Levene, 1960) wees uit dat de standaarddeviatie evenmin significant verschilde over beide groepen (SD = 1.97 en 1.89; $F(1, 106) = 0.40$, ns).

7.4 Conclusie

Beide interventies werden over het algemeen positief onthaald. Voor wat het beoordelingsformulier betreft, geldt dat de betrokkenen over het algemeen een meerwaarde zagen in het nieuwe beoordelingsformulier. Het algemeen gemiddelde voor alle deelnemers verschilde voor alle criteria (met uitzondering van Motivatie) significant van 0 (= het neutraal middelpunt van de schaal) in positieve richting. Studenten waren over de hele lijn lovend en bleken over het algemeen positiever te staan tegenover het nieuwe beoordelingsformulier dan begeleiders. Dit is niet zo vreemd. Studenten werden immers louter gevraagd wat ze van het beoordelingsformulier vonden, terwijl begeleiders expliciet gevraagd werd of het nieuwe formulier iets extra bijdroeg aan de beoordeling. Hun antwoord hing dus niet enkel af van het nieuwe beoordelingsformulier op zich, maar ook van wat ze vonden van het oude formulier en de rest van de stagebeoordeling. Ondanks die hoge eis, zagen ook begeleiders voor heel wat criteria een duidelijke meerwaarde in het nieuwe beoordelingsformulier. Enkel voor Herhaalbaarheid, Betekenisvolheid, Complexiteit en Leereffect zagen ze hier niet direct een meerwaarde in.

Voor wat de leergroepen betreft, geldt dat de betrokkenen over het algemeen een meerwaarde zagen in deze manier van werken. Het algemeen gemiddelde voor alle deelnemers verschilde voor alle criteria (met uitzondering van Herhaalbaarheid) significant van 0 (= het neutraal middelpunt van de schaal) in positieve richting. Studenten waren over de hele lijn lovend (behalve op vlak van Herhaalbaarheid) en bleken op sommige vlakken (met name op vlak van Transparantie, Authenticiteit en Eerlijkheid) positiever te staan tegenover de leergroepen dan docenten. Dit is ook hier weer niet zo vreemd. Studenten werden immers louter gevraagd wat ze van de leergroepen vonden, terwijl docenten expliciet gevraagd werd of de leergroepen iets extra bijdroegen aan de stagebeoordeling. Hun antwoord hing dus niet enkel af van de leergroepen op zich, maar ook van wat ze vonden van de rest van de stagebeoordeling. Ondanks die redelijk hoge eis, zagen ook docenten voor heel wat criteria duidelijk de meerwaarde van de leergroepen

in. Behalve voor Herhaalbaarheid zagen ze hier enkel voor Transparantie, Authenticiteit, Eerlijkheid en Motivatie geen duidelijke meerwaarde in. De scores op Herhaalbaarheid tonen aan dat de leergroepen misschien nog harder geapprecieerd zouden worden als ze niet door één maar door meerdere docenten zouden worden begeleid.

Wanneer er ingezoomd wordt op Leereffect, zien we duidelijk de meerwaarde van onze interventies. Gegeven dat begeleiders geen leereffect zagen van het nieuwe beoordelingsformulier en gegeven dat dit nieuwe formulier, in tegenstelling tot de leergroepen, ook niet direct een leereffect beoogde, kunnen we ervan uit gaan dat het de invoering van de leergroepen is die deze meerwaarde creëerde. Voor het jaar 2017-2018 verwachtten docenten voor zo goed als alle competenties meer competentiegroei dan voor het jaar 2016-2017, al was dit verschil enkel significant voor Zelfreflectie, Feedback geven en Feedback vragen (maar het gaat hier uiteraard ook om heel kleine groepen waardoor het effect al groot moest zijn alvorens het significant werd). In 2016-2017 verwachtten docenten enkel competentiegroei voor wat betreft Praktijkgerichtheid, Zelfmanagement, Zelfreflectie en Samenwerken. In 2017-2018 verwachtten ze niet alleen significant meer groei voor Zelfreflectie, maar verwachtten ze bovendien ook groei voor Doelgerichtheid, Kansen zien, en Feedback geven en Feedback vragen. Studenten rapporteerden in 2017-2018 dan weer voor alle domeinen significant meer competentiegroei dan in 2016-2017. Waar ze in 2016-2017 geen competentiegroei rapporteerden voor Leidinggeven, Vernieuwen en Feedback geven, deden ze dat in 2017-2018 wel. Voor zowel 2016-2017 als 2017-2018 verwachtten docenten significant meer competentiegroei op vlak van Praktijkgerichtheid dan studenten rapporteerden. Hetzelfde geldt voor Feedback geven in 2017-2018. Op vlak van Leidinggeven, Vernieuwen en Klantgericht werken rapporteerden studenten in 2017-2018 dan weer meer competentiegroei dan de docenten verwachtten. Over het algemeen kan hieruit geconcludeerd worden dat de invoering van de leergroepen een groot succes was: de leergroepen zorgden ervoor dat docenten meer overtuigd raakten van de door de stage te verwachten competentiegroei en studenten ervaarden ook daadwerkelijk meer competentiegroei.

Wanneer er ingezoomd wordt op Selectie vinden we geen effecten. Voor Selectie waren er geen significante verschillen tussen 2016-2017 en 2017-2018, noch in termen van gemiddelde noch in termen van spreiding (= standaarddeviatie). Op basis hiervan dienen we te concluderen dat er geen aanwijzing is voor het feit dat de interventie erin zou geslaagd zijn om een beter onderscheid te helpen maken tussen goede en slechte studenten. Dit kan uiteraard een hele waaier aan mogelijke oorzaken hebben. De afwezigheid van dit effect zou bijvoorbeeld te wijten kunnen zijn aan het feit dat docenten zowel qua gemiddelde als qua spreiding een standaard hanteren die ze over de jaren heen constant proberen te houden.

Tabel 7.1: Cronbach alfas van en correlaties tussen de kwaliteitscriteria voor het nieuwe beoordelingsformulier

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.82													
02. Vergelijk	.87	.59 [2]												
03. Herhaalbaar	.93	.28 [2]	.45 [2]											
04. Transparant	.84	.60 [2]	.58 [2]	.45 [2]										
05. Acceptatie	.92	.55 [2]	.66 [2]	.47 [2]	.75 [2]									
06. Authentiek	.89	.69 [2]	.62 [2]	.38 [2]	.52 [2]	.59 [2]								
07. Eerlijk	.85	.53 [2]	.65 [2]	.54 [2]	.66 [2]	.73 [2]	.61 [2]							
08. Complex	.87	.52 [2]	.55 [2]	.60 [2]	.53 [2]	.53 [2]	.49 [2]	.59 [2]						
09. Betekenis	.92	.56 [2]	.46 [1]	.35 [2]	.56 [2]	.51 [2]	.51 [2]	.49 [2]	.52 [2]					
10. Tijd/Kosten	.77	.50 [2]	.55 [2]	.49 [2]	.49 [2]	.54 [2]	.61 [2]	.58 [2]	.59 [2]	.49 [2]				
11. Selectie	.86	.62 [2]	.59 [2]	.40 [2]	.62 [2]	.66 [2]	.66 [2]	.67 [2]	.49 [2]	.56 [2]	.64 [2]			
12. Leereffect	.90	.61 [2]	.56 [2]	.38 [2]	.52 [2]	.50 [2]	.60 [2]	.58 [2]	.58 [2]	.73 [2]	.62 [2]	.68 [2]		
13. Motivatie	.93	.31 [2]	.35 [2]	.47 [2]	.39 [2]	.38 [2]	.35 [2]	.39 [2]	.63 [2]	.59 [2]	.49 [2]	.31 [2]	.59 [2]	
14. Onderwijs	.92	.79 [2]	.20	-.19	.37	.38	.25	32	-.29	.06	.77 [2]	.83 [2]	.66 [1]	.00

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 7.2: Regressie bij studenten op de kwaliteitscriteria voor het beoordelingsformulier

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.10	-.16	-.16	-.10	-.08	.14
01. Geschikt	.09	-.05	-.10	.06	-.23	.10
02. Vergelijk	.08	.00	-.22	.04	-.04	.16
03. Herhaalbaar	.14	-.12	-.19	-.27 [1]	.05	.12
04. Transparant	.05	-.06	-.02	.08	-.20	.03
05. Acceptatie	.14	.06	-.02	.00	-.30 [1]	.17
06. Authentiek	.11	.04	-.22	.07	-.04	.22
07. Eerlijk	.07	-.00	-.02	-.19	-.09	.17
08. Complex	.09	-.26	-.18	-.09	.07	-.01
09. Betekenis	.13	-.32 [1]	-.11	-.16	-.02	.03
10. Tijd/Kosten	.08	-.10	-.10	-.15	.07	.18
11. Selectie	.11	-.10	-.10	.06	-.14	.21
12. Leereffect	.12	-.31 [1]	-.26	-.03	-.01	-.02
13. Motivatie	.13	-.23	.02	-.23	.01	.09

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat studenten vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.3: Regressie bij begeleiders op de kwaliteitscriteria voor het beoordelingsformulier

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.29	.30	-.60	-.04
01. Geschikt	.19	.27	-.54	.03
02. Vergelijk	.25	.23	-70 [1]	.20
03. Herhaalbaar	.30	.05	-.69 [1]	.22
04. Transparant	.21	.25	-.61	.11
05. Acceptatie	.22	.35	-.57	.03
06. Authentiek	.29	.57 [1]	-.29	-.17
07. Eerlijk	.25	.56 [2]	-.62	.20
08. Complex	.03	.09	-.13	-.10
09. Betekenis	.28	-.11	-.51	.09
10. Tijd/Kosten	.38 [1]	.43	-.45	-.28
11. Selectie	.39 [1]	.55 [1]	-.54	-.17
12. Leereffect	.15	.17	-.28	-.20
13. Motivatie	.19	-.38	.16	-.33
14. Onderwijs	.55	.47	-.66	-.11

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.4: Cronbach alfas van en correlaties tussen de kwaliteitscriteria voor de leergroepen

Criteria	Alfa	01	02	03	04	05	06	07	08	09	10	11	12	13
01. Geschikt	.90													
02. Vergelijk	.86	.36 [2]												
03. Herhaalbaar	.88	.10	.20											
04. Transparant	.95	.35 [2]	.53 [2]	.35 [2]										
05. Acceptatie	.94	.44 [2]	.53 [2]	.31 [2]	.71 [2]									
06. Authentiek	.85	.57 [2]	.51 [2]	.11	.45 [2]	.56 [2]								
07. Eerlijk	.84	.32 [2]	.66 [2]	.31 [2]	.63 [2]	.80 [2]	.53 [2]							
08. Complex	.78	.39 [2]	.47 [2]	.20	.43 [2]	.54 [2]	.44 [2]	.53 [2]						
09. Betekenis	.94	.53 [2]	.27 [1]	.13	.36 [2]	.43 [2]	.49 [2]	.25 [1]	.48 [2]					
10. Tijd/Kosten	.95	.49 [2]	.47 [2]	.19	.36 [2]	.51 [2]	.57 [2]	.48 [2]	.49 [2]	.56 [2]				
11. Selectie	.91	.59 [2]	.20	.16	.17	.36 [2]	.45 [2]	.34 [2]	.51 [2]	.42 [2]	.68 [2]			
12. Leereffect	.77	.64 [2]	.41 [2]	.04	.32 [2]	.47 [2]	.57 [2]	.48 [2]	.64 [2]	.65 [2]	.68 [2]	.71 [2]		
13. Motivatie	.89	.39 [2]	.42 [2]	.10	.22	.45 [2]	.46 [2]	.38 [2]	.43 [2]	.55 [2]	.51 [2]	.44 [2]	.53 [2]	
14. Onderwijs	.81	.63 [2]	-.37	-.26	-.31	.31	.27	-.09	.38	.42	.26 [2]	.58	.60	.02

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 7.5: Regressie bij studenten op de kwaliteitscriteria voor de leergroepen

Criteria	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.11	-.11	-.08	-.19	-.13	.15
01. Geschikt	.07	.02	-.03	-.19	-.15	.03
02. Vergelijk	.06	-.10	.06	-.05	-.06	.17
03. Herhaalbaar	.21 [1]	-.32 [1]	-.25	-.16	-.14	.08
04. Transparant	.10	-.21	-.11	-.11	-.06	.13
05. Acceptatie	.09	-.14	-.07	-.20	-.07	.12
06. Authentiek	.05	-.03	-.06	-.15	.02	.15
07. Eerlijk	.04	-.08	-.05	-.17	.03	.05
08. Complex	.05	.07	-.06	-.18	-.08	.01
09. Betekenis	.09	-.06	-.12	-.15	-.13	.14
10. Tijd/Kosten	.07	-.11	-.02	.07	-.23	.00
11. Selectie	.07	-.10	-.02	-.19	-.17	-.05
12. Leereffect	.04	.12	-.03	-.08	-.01	.15
13. Motivatie	.10 [1]	.11	.09	-.20	-.11	.36 [2]

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat studenten vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.6: Regressie bij begeleiders op de kwaliteitscriteria voor de leergroepen

Criteria	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.27	.11	.06	-.52
01. Geschikt	.18	-.46	.22	-.28
02. Vergelijk	.44	-.06	.77	-.20
03. Herhaalbaar	.61	.58	.18	.40
04. Transparant	.11	-.26	.13	.14
05. Acceptatie	.09	-.13	-.18	-.11
06. Authentiek	.75 [1]	.11	.34	-.96 [2]
07. Eerlijk	.23	.52	-.18	.05
08. Complex	.45	-.15	-.23	-.51
09. Betekenis	.16	-.22	-.06	-.31
10. Tijd/Kosten	.76 [1]	.18	-.05	-.81 [2]
11. Selectie	.56	.44	-.27	-.44
12. Leereffect	.58	.27	-.07	-.66
13. Motivatie	.31	-.33	.13	-.58
14. Onderwijs	.03	-.00	-.12	-.08

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.7: Cronbach alfas van en correlaties tussen de schalen voor competentiegroei bij studenten

Competenties	Alfa	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht	.82											
02. Zelfmanagement	.83	.68 [2]										
03. Zelfreflectie	.81	.67 [2]	.66 [2]									
04. Doelgericht	.92	.59 [2]	.71 [2]	.64 [2]								
05. Kansen zien	.83	.54 [2]	.71 [2]	.55 [2]	.70 [2]							
06. Leidinggeven	.89	.34 [1]	.41 [2]	.32 [2]	.56 [2]	.54 [2]						
07. Vernieuwen	.83	.48 [2]	.54 [2]	.48 [2]	.65 [2]	.52 [2]	.47 [2]					
08. Samenwerken	.84	.60 [2]	.62 [2]	.52 [2]	.59 [2]	.56 [2]	.53 [2]	.61 [2]				
09. Feedback geven	.86	.41 [2]	.48 [2]	.36 [2]	.55 [2]	.51 [2]	.73 [2]	.51 [2]	.62 [2]			
10. Feedback vragen	.83	.46 [2]	.35 [2]	.41 [2]	.44 [2]	.31 [2]	.37	.42 [2]	.50 [2]	.32 [2]		
11. Klantgericht werken	.87	.55 [2]	.52 [2]	.52 [2]	.65 [2]	.54 [2]	.59 [2]	.61 [2]	.73 [2]	.64 [2]	.59 [2]	
12. Voor groep spreken	.86	.22	.47 [2]	.30 [1]	.53 [2]	.34 [2]	.70 [2]	.42 [1]	.41 [2]	.62 [2]	.33 [2]	.68 [2]

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 7.8: Correlaties tussen de items voor competentiegroei bij docenten

Competenties	01	02	03	04	05	06	07	08	09	10	11
01. Praktijkgericht											
02. Zelfmanagement	.39										
03. Zelfreflectie	.49 [1]	.36									
04. Doelgericht	.50 [1]	.74 [2]	.49 [2]								
05. Kansen zien	.36	.52 [1]	.39	.72 [2]							
06. Leidinggeven	.29	.38	.28	.51 [1]	.33						
07. Vernieuwen	.22	.15	.10	.21	.21	.65 [2]					
08. Samenwerken	.47 [1]	.64 [2]	.17	.64 [2]	.60 [2]	.30	.06				
09. Feedback geven	.49 [1]	.36	.70 [2]	.42	.49 [1]	.36	.28	.45			
10. Feedback vragen	.60 [1]	.37	.58 [2]	.39	.26	.35	.22	.46 [1]	.82 [2]		
11. Klantgericht werken	.29	.52 [1]	.14	.28	.39	.15	.29	.28	.24	.27	
12. Voor groep spreken	.32	.21	.60 [2]	.30	.32	.00	-.18	.30	.66 [2]	.55 [1]	.21

Noot: [1] $p < .05$; [2] $p < .01$.

Tabel 7.9: Regressie bij studenten op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Educatie	Etniciteit	Score 1
00. Algemeen	.04	-.11	-.10	.03	-.01	.12
01. Praktijkgericht	.09	.13	-.17	.06	.23	.07
02. Zelfmanagement	.05	-.06	.04	-.08	.11	.14
03. Zelfreflectie	.04	.08	.05	.06	-.08	-.13
04. Doelgericht	.01	.00	.08	-.03	-.07	.04
05. Kansen zien	.05	.03	-.11	.00	.17	.16
06. Leidinggeven	.08	-.20	-.15	-.04	.06	.12
07. Vernieuwen	.17 [1]	-.02	-.24	.09	-.15	.25
08. Samenwerken	.11	-.09	-.17	.08	-.05	.23
09. Feedback geven	.19 [1]	-.40 [2]	-.01	-.04	-.03	.09
10. Feedback vragen	.01	.04	.00	.09	-.04	-.04
11. Klantgericht werken	.02	-.08	-.06	.05	-.09	-.03
12. Voor groep spreken	.02	-.13	-.03	.03	-.02	.02

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken en het punt dat ze vinden dat ze verdienen (= Score 1). De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.10: Regressie bij docenten op de schalen voor competentiegroei

Competenties	R^2	Geslacht	Leeftijd	Ervaring
00. Algemeen	.15	-.18	.42	-.44
01. Praktijkgericht	.84 [2]	.29	.09	-.87 [2]
02. Zelfmanagement	.37	-.63	.37	-.51
03. Zelfreflectie	.35	.36	-.15	-.05
04. Doelgericht	.51	-.41	.66	-.87 [1]
05. Kansen zien	.10	-.29	.32	-.33
06. Leidinggeven	.34	-.40	.33	-.17
07. Vernieuwen	.59	.08	.80 [1]	-.17
08. Samenwerken	.32	-.56	.04	-.27
09. Feedback geven	.10	.26	-.36	.31
10. Feedback vragen	.26	.55	-.33	.09
11. Klantgericht werken	.21	-.45	.49	-.27
12. Voor groep spreken	.19	-.23	-.33	.06

Noot: R^2 verwijst naar de variantie in de afhankelijke variabele die verklaard wordt door de achtergrondkenmerken. De andere cijfers zijn de gestandaardiseerde beta-gewichten van de respectievelijke predictoren.

[1] $p < .05$; [2] $p < .01$.

Tabel 7.11: Gemiddelden per rol op de kwaliteitscriteria voor het beoordelingsformulier

Criteria	Iedereen	Studenten	Begeleiders	ΔBS
00. Algemeen	1.01 (0.84) [2]	1.15 (0.84) [2]	0.63 (0.78) [2]	1
01. Geschikt	1.44 (1.10) [2]	1.56 (0.92) [2]	1.35 (0.80) [2]	0
02. Vergelijk	1.16 (1.20) [2]	1.33 (1.16) [2]	0.67 (1.19) [1]	1
03. Herhaalbaar	0.48 (1.46) [2]	0.50 (1.52) [1]	0.42 (1.28) [0]	0
04. Transparant	1.30 (0.99) [2]	1.50 (0.94) [2]	0.72 (0.90) [2]	2
05. Acceptatie	1.30 (1.01) [2]	1.47 (0.95) [2]	0.80 (1.07) [2]	2
06. Authentiek	1.34 (1.34) [2]	1.41 (1.03) [2]	1.14 (0.82) [2]	0
07. Eerlijk	1.20 (1.02) [2]	1.41 (0.90) [2]	0.60 (1.14) [1]	2
08. Complex	0.61 (1.18) [2]	0.76 (1.11) [2]	0.16 (1.29) [0]	1
09. Betekenis	0.84 (1.30) [2]	1.04 (1.27) [2]	0.26 (1.24) [0]	1
10. Tijd/Kosten	1.05 (0.91) [2]	1.11 (0.91) [2]	0.88 (0.92) [2]	0
11. Selectie	1.18 (1.03) [2]	1.28 (1.04) [2]	0.88 (0.95) [2]	0
12. Leereffect	0.93 (1.25) [2]	1.13 (1.17) [2]	0.35 (1.32) [0]	1
13. Motivatie	0.28 (1.44) [0]	0.39 (1.41) [1]	-0.05 (1.51) [0]	0
14. Onderwijs	0.78 (1.02) [1]		0.78 (1.02) [1]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Iedereen" is met een subscript aangegeven of het gemiddelde voor alle deelnemers verschilt van 0. Ook in de kolommen "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 7.12: Gemiddelden per rol op de kwaliteitscriteria voor de leergroepen

Criteria	Iedereen	Studenten	Begeleiders	ΔBS
00. Algemeen	1.05 (0.81) [2]	1.06 (0.84) [2]	1.03 (0.61) [2]	0
01. Geschikt	1.44 (1.11) [2]	1.37 (1.15) [2]	1.88 (0.64) [2]	0
02. Vergelijk	1.52 (1.05) [2]	1.56 (1.56) [2]	1.28 (0.70) [2]	0
03. Herhaalbaar	0.19 (1.64) [0]	0.13 (1.66) [0]	0.61 (1.60) [0]	0
04. Transparant	0.57 (1.47) [2]	0.73 (1.41) [2]	-0.37 (1.52) [0]	1
05. Acceptatie	0.77 (1.20) [2]	0.88 (1.12) [2]	0.09 (1.15) [0]	1
06. Authentiek	0.92 (1.13) [2]	0.91 (1.15) [2]	0.94 (1.04) [1]	0
07. Eerlijk	1.04 (1.24) [2]	1.21 (1.12) [2]	0.03 (1.49) [0]	2
08. Complex	1.21 (0.92) [2]	1.19 (0.96) [2]	1.36 (0.67) [2]	0
09. Betekenis	1.23 (1.22) [2]	1.20 (1.11) [2]	1.45 (1.78) [1]	0
10. Tijd/Kosten	1.35 (1.20) [2]	1.25 (1.84) [2]	1.91 (1.23) [2]	0
11. Selectie	1.26 (1.11) [2]	1.17 (1.10) [2]	1.79 (1.05) [2]	0
12. Leereffect	1.47 (0.92) [2]	1.42 (0.94) [2]	1.73 (0.77) [2]	0
13. Motivatie	0.70 (1.35) [2]	0.70 (1.38) [2]	0.67 (1.20) [0]	0
14. Onderwijs	1.64 (0.80) [2]		1.64 (0.80) [2]	

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. In de kolom "Iedereen" is met een subscript aangegeven of het gemiddelde voor alle deelnemers verschilt van 0. Ook in de kolommen "Studenten" en "Begeleiders" is met een subscript aangegeven of het gemiddelde verschilt van 0. Of de gemiddelden van begeleiders en studenten verschillen binnen een conditie, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

Tabel 7.13: Gemiddelden per conditie en per rol op de schalen voor competentiegroei

			2016-2017			2017-2018		
Competenties	ΔB	ΔS	Begeleiders	Studenten	ΔBS	Begeleiders	Studenten	ΔBS
00. Algemeen	1	2	3.29 (0.66) [0]	3.45 (0.52) [2]	0	3.82 (0.38) [2]	4.00 (0.37) [2]	0
01. Praktijkgericht	0	2	4.63 (0.52) [2]	4.11 (0.64) [2]	1	4.94 (0.20) [2]	4.58 (0.37) [2]	2
02. Zelfmanagement	0	2	4.13 (0.64) [2]	3.69 (0.59) [2]	0	4.33 (0.60) [2]	4.31 (0.43) [2]	0
03. Zelfreflectie	1	2	3.75 (0.71) [1]	3.86 (0.56) [2]	0	4.57 (0.62) [2]	4.43 (0.41) [2]	0
04. Doelgericht	0	2	3.13 (0.83) [0]	3.23 (0.71) [1]	0	3.90 (0.80) [2]	3.98 (0.54) [2]	0
05. Kansen zien	0	2	3.25 (0.89) [0]	3.25 (0.74) [1]	0	3.66 (0.94) [1]	3.99 (0.57) [2]	0
06. Leidinggeven	0	1	1.63 (0.92) [2]	2.91 (0.81) [0]	2	2.21 (1.06) [1]	3.31 (0.75) [2]	2
07. Vernieuwen	0	2	2.88 (0.83) [0]	3.27 (0.88) [0]	0	2.82 (1.04) [0]	4.02 (0.54) [2]	2
08. Samenwerken	0	2	3.88 (0.83) [1]	3.42 (0.64) [2]	0	4.33 (0.52) [2]	3.95 (0.55) [2]	1
09. Feedback geven	1	2	3.00 (1.07) [0]	3.06 (0.68) [0]	0	4.08 (0.68) [2]	3.51 (0.63) [2]	2
10. Feedback vragen	2	2	3.63 (0.92) [0]	3.74 (0.82) [2]	0	4.57 (0.45) [2]	4.28 (0.64) [2]	0
11. Klantgericht werken	0	2	3.75 (1.38) [0]	3.33 (0.75) [2]	0	3.12 (1.48) [0]	3.96 (0.54) [2]	2
12. Voor groep spreken						3.30 (1.24) [0]	3.71 (0.65) [2]	0

Noot: Standaarddeviaties zijn weergegeven tussen haakjes. Als de gemiddelden afwijken van 3 (= het neutraal middelpunt), dan is dit aangegeven met een subscript. Of de gemiddelden van begeleiders verschillen over de jaren heen, is weergegeven met een cijfer in de kolom "ΔB". Of de gemiddelden van studenten verschillen over de jaren heen, is weergegeven met een cijfer in de kolom "ΔS". Of de gemiddelden van begeleiders en studenten verschillen binnen een jaar, is aangegeven met een cijfer in de kolommen "ΔBS".

[0] *ns*; [1] $p < .05$; [2] $p < .01$.

8 DISCUSSIE

In 2016-2017 startten we een onderzoek naar pistes ter verbetering van stagebeoordelingen. Daartoe doorzochten we de relevante literatuur. Vervolgens werden vier opleidingen geselecteerd (i.e., PBKO & PBLO, VPK, en AFI). Voor elk van deze opleidingen werd voor het laatste jaar de stagebeoordeling doorgelicht op zoek naar verbeteropties, telkens in samenspraak met de opleiding in kwestie. Voor elke opleiding werd vervolgens een onderzoeksdesign uitgewerkt voor het academiejaar 2017-2018. In de uitwerking van deze onderzoeksdesigns probeerden we de een belangrijke aanbeveling uit de literatuur voor ogen te houden: een goede stagebeoordeling dient niet enkel een summatief maar ook een formatief aspect te hebben.

Om te meten of de aangebrachte veranderingen ook verbeteringen waren, moesten we op zoek naar geschikte criteria. We kwamen hiervoor uit bij het Competentie Assessment Wiel van Baartman et al. (2006, 2007) waarin een reeks kwaliteitscriteria naar voor worden geschoven waaraan een beoordelingsprogramma moet voldoen. Omdat het niet onze bedoeling was het volledig beoordelingsprogramma van een opleiding onder de loep te nemen maar slechts één beoordeling (i.e., de stagebeoordeling in het laatste jaar), waren niet alle criteria altijd even relevant. Omdat er geen bruikbare vragenlijst voorhanden was, bestond een eerste stap erin zo'n vragenlijst op te stellen. In 2016-2017 plande een onderzoeksteam van de UAntwerpen een interventie in het Flextraject van PBLO die aansloot bij onze doelstelling. Ideaal om de vragenlijst die we opstelden om de kwaliteitscriteria te meten te testen. Die doorstond de test vrij goed, wat toeliet uitspraken te doen over de interventie zelf (i.e., D-PAC, een methode voor paarsgewijze beoordeling). Begeleiders zagen hier positieve en negatieve aspecten aan, maar de studenten zagen enkel negatieve aspecten. In Voorstudie 2 bij studenten uit het tweede jaar VPK werd onze vragenlijst op punt gesteld en ingekort. De stagebeoordeling werd er positief beoordeeld, al vonden studenten dat de beslissingen die werden genomen op onvoldoende assessmentsituaties en beoordelaars gebaseerd waren.

In Studie 1 en Studie 2, werd op deze bevinding ingespeeld door het aantal assessmentsituaties en beoordelaars op te drijven in een experimenteel onderzoeksdesign. In Studie 1 (in PBKO & PBLO) werd voor de helft van de studenten meer begeleiders ingeschakeld die ook vaker op lesbezoek gingen. Het was bovendien ook de bedoeling dat iedereen die op lesbezoek ging zowel formatief als summatief zijn stem zou laten horen en zijn bevindingen zou delen met zowel de collega-bezoekers als de studenten. In Studie 2 (in VPK) werd voor de helft van de studenten de medestudenten ingeschakeld via peer assessments. Ook hier gebeurde dit zowel formatief als summatief. In Studie 3 (in AFI) gooiden we het over een andere boeg. In deze studie werd niet zozeer ingezet op het aantal assessmentsituaties en beoordelaars maar werd er expliciet geprobeerd andere competenties in de stagebeoordeling aan bod te laten komen. Via leergroepen probeerden we studenten soft skills aan te leren en via een nieuw beoordelingsformulier probeerden we die expliciet in het eindoordeel te betrekken. Verder werden mentoren ook aangemoedigd meer en betere feedback te geven. Omwille van het geringe studentenaantal werden de kwaliteitscriteria enkel gemeten op niveau van de specifieke interventies. Voor Leereffect en Selectie werd een vergelijking gemaakt met voorgaand jaar. In wat volgt, vatten we de resultaten van deze studies samen. Vervolgens bespreken we de resultaten en staan we stil bij hoe het komt dat sommige interventies beter onthaald werden en meer effect hadden dan andere. We eindigen onze discussie met een aantal aanbevelingen en beperkingen.

8.1 Samenvatting van de resultaten

8.1.1 De perceptie van de interventies

In Studie 1 werd de interventie (= meer begeleiders vaker op lesbezoek laten gaan en hun tussentijds en voorafgaand aan het eindoordeel met elkaar laten overleggen en bevindingen uitwisselen) positief beoordeeld. Wanneer er specifiek gepolst werd naar de bijdrage van deze interventie aan de algemene stagebeoordeling, bleek het algemeen gemiddelde voor alle deelnemers in positieve richting van 0 (= het neutraal middelpunt van de schaal) te verschillen, en dit voor alle criteria. Alle betrokkenen waren het er dus gemiddeld genomen over eens dat de interventie een positieve bijdrage levert aan de stagebeoordeling. Begeleiders schatten de bijdrage van de interventie aan de stagebeoordeling positiever in dan studenten in termen van Herhaalbaarheid en Eerlijkheid. Begeleiders stonden ook positiever ten aanzien van de haalbaarheid van de door de interventie gevergde Tijd & Kosten dan studenten; een enigszins merkaardig bevinding omdat het de begeleiders zijn die in het kader van deze interventie een extra inspanning dienden te leveren. Al bij al werd de interventie dus positiever beoordeeld door begeleiders, maar ook studenten zagen hier duidelijk de meerwaarde van in.

In Studie 2 werd de interventie (= via peer assessment het oordeel van medestudenten betrekken in de stagebeoordeling) minder positief beoordeeld, of toch alleszins minder unaniem. Wanneer er specifiek gepolst werd naar de bijdrage van de interventie aan de stagebeoordeling, bleek het gemiddelde voor alle deelnemers weliswaar in positieve richting van 0 te verschillen en dit voor bijna alle criteria (uitzonderingen zijn Herhaalbaarheid, Authenticiteit en Motivatie), maar meer gedetailleerde analyses wezen op een significant verschil tussen studenten en begeleiders. Begeleiders bleken over het algemeen van mening dat de interventie een grotere bijdrage levert aan de kwaliteit van de algemene beoordeling dan studenten. Voor studenten bleek het zelfs zo dat de scores doorgaans niet significant verschillen van 0 (met uitzondering van Acceptatie, Eerlijkheid, Complexiteit, en Selectie). Studenten zagen ook positieve aspecten aan de interventie, maar staan hier over het algemeen toch vrij neutraal tegenover. Begeleiders zagen dus duidelijk de meerwaarde van deze interventie in, maar voor studenten is dat minder het geval. Dit sluit aan bij de bevindingen uit Voorstudie 1. Ook daar werd geprobeerd het oordeel van medestudenten in het eindoordeel te betrekken, en ook daar werd de interventie niet erg positief onthaald door studenten. Begeleiders zagen zowel voor- als nadelen aan de interventie in kwestie (= D-PAC; een manier van paarsgewijs vergelijken), maar studenten zagen eigenlijk enkel nadelen. Wanneer de resultaten van Voorstudie 1, Studie 1 en Studie 2 gezamenlijk bekeken worden, zou hieruit geconcludeerd kunnen worden dat het niet per se een goed idee is om meer beoordelaars bij het eindoordeel te betrekken. Of dit al dan niet een goed idee is, hangt er immers van af wie die extra beoordelaars zijn. Wanneer men die extra beoordelaars gaat zoeken onder de medestudenten, dan groeit de scepsis; vooral onder studenten, en zeker wanneer de beoordelingscriteria ook nog eens vaag gehouden worden (zoals in Voorstudie 1). Een voor de hand liggende verklaring is dat studenten onvoldoende vertrouwen hebben in (de onderbouwing van) het oordeel van hun medestudenten. Dit is echter niet per se een reden om dergelijke interventies te mijden. Het zou namelijk kunnen dat dit wantrouwen van studenten onterecht is; iets wat je zou kunnen afleiden uit het feit dat deze interventies wel positief onthaald werden door begeleiders (zie ook Falchikov & Goldfinch, 2000).

In Studie 3 werden beide interventies positief onthaald. De betrokkenen zagen een meerwaarde in *het nieuwe beoordelingsformulier*. Het algemeen gemiddelde bleek in positieve richting van 0 te verschillen voor alle criteria, uitgezonderd voor Motivatie. Studenten waren over de hele lijn lovend en bleken over het algemeen positiever te staan tegenover het nieuwe formulier dan begeleiders. Dit had vermoedelijk ook te maken met de manier waarop de vragen waren gesteld. Studenten werden louter gevraagd wat ze van het beoordelingsformulier in kwestie vonden, terwijl begeleiders expliciet gevraagd werd of het nieuwe formulier iets extra bijdraagt aan de stagebeoordeling. Hun antwoord hing dus niet enkel af van het nieuwe beoordelingsformulier op zich, maar ook van wat ze vonden van het oude formulier en de rest van de stagebeoordeling. Ondanks die hoge eis, zagen ook begeleiders voor heel wat criteria een meerwaarde in het nieuwe beoordelingsformulier. Enkel voor Herhaalbaarheid, Betekenisvolheid, Complexiteit en Leereffect was dit niet zo. Ook in *de leergroepen* zagen de betrokkenen een meerwaarde. Het algemeen gemiddelde voor alle deelnemers verschilde in positieve richting van 0 voor alle criteria, uitgezonderd Herhaalbaarheid. Studenten waren over de hele lijn lovend, behalve op vlak van Herhaalbaarheid, en bleken op sommige vlakken (met name op vlak van Transparantie, Authenticiteit en Eerlijkheid) positiever te staan tegenover de leergroepen dan docenten. Ook hier had dit vermoedelijk te maken met de vraagstelling. Studenten werden louter gevraagd wat ze van de leergroepen in kwestie vonden, terwijl docenten expliciet gevraagd werd of de leergroepen iets extra bijdragen aan de stagebeoordeling. Ondanks die hoge eis, zagen ook docenten voor heel wat criteria de meerwaarde van de leergroepen in. Behalve voor Herhaalbaarheid zagen ze hier enkel voor Transparantie, Authenticiteit, Eerlijkheid en Motivatie geen duidelijke meerwaarde in. De scores voor Herhaalbaarheid tonen aan dat de leergroepen misschien nog harder geapprecieerd zoudeb worden door zowel studenten als docenten als ze niet door één maar door meerdere docenten zouden worden begeleid.

8.1.2 Het globaal effect van de interventies

Daar waar mogelijk (= in Studie 1 en Studie 2) beperkten we ons niet enkel tot onderzoeken hoe de interventies werden ervaren. Vanuit wetenschappelijk oogpunt is dat immers een vrij "zachte" test die zich situeert op het niveau van de beleving. Een "hardere" test is kijken of de mensen die bij de interventie betrokken waren effectief positiever stonden ten aanzien van de stage en de stagebeoordeling dan mensen die niet bij de interventie betrokken waren. Daarom werden, daar waar mogelijk, zowel studenten als begeleiders in twee groepen verdeeld, en werd de interventie slechts bij 1 van de 2 groepen doorgevoerd. In Studie 3 was dit helaas niet mogelijk omwille van het relatief kleine studentenaantal.

In Studie 1 bleek uit de "hardere" tests, waarbij we op experimentele wijze de twee groepen met elkaar vergeleken, dat onze interventie zo goed als geen effect had op de beoordeling van de stagebeoordeling. Zowel in de controlegroep als in de experimentele groep stonden de deelnemers over het algemeen positief ten aanzien van de stagebeoordeling en dit voor alle criteria met uitzondering van Herhaalbaarheid (in beide groepen) en Complexiteit (in de experimentele groep). We vonden geen verschillen tussen de experimentele groep en de controlegroep behalve voor Complexiteit. In de controleconditie werd de beoordeling als beter geschikt gezien om de cognitieve vaardigheden van studenten bloot te leggen. Dit is een contra-intuïtief resultaat waar we ook niet direct een verklaring voor hebben. In de controlegroep schatten begelei-

ders de stagebeoordeling significant positiever in dan studenten voor Vergelijkbaarheid, Eerlijkheid, Tijd en Kosten, en Selectie. Studenten schatten de beoordeling dan weer positiever in voor Betekenisvolheid en Motivatie. De experimentele conditie bevestigde de positievere inschatting door begeleiders voor Vergelijkbaarheid en Eerlijkheid. Voor Studie 1 kan dus gesteld worden dat de "hardere" tests de eerder gevonden tevredenheid over de interventie niet ondersteunen, maar ze spreken die tevredenheid ook niet tegen.

In Studie 2 bleek uit de "hardere" tests dat onze interventie een licht negatief effect had op de beoordeling van de stagebeoordeling. Zowel in de controlegroep als in de experimentele groep stonden de deelnemers over het algemeen positief ten aanzien van de stagebeoordeling en dit voor alle criteria met uitzondering van, wederom, Herhaalbaarheid (in beide groepen), maar de deelnemers in de controleconditie stonden positiever ten aanzien van de stagebeoordeling dan de deelnemers in de experimentele conditie. Hieruit dienen we te concluderen dat de interventie eerder slecht onthaald werd: ze drukte de positieve houding ten aanzien van de stagebeoordeling. In de controlegroep vonden we geen verschillen tussen het oordeel van studenten en het oordeel van begeleiders. Alle scores van zowel studenten als begeleiders verschilden ook significant van 0 in positieve richting. In de experimentele conditie daarentegen vonden we significante verschillen tussen studenten en begeleiders voor alle criteria behalve Authenticiteit en Motivatie. Alle criteria werden positiever beoordeeld door begeleiders dan door studenten. Enkel voor Motivatie was dit verschil niet significant. De studenten in de experimentele conditie stonden dus minst positief ten aanzien van de stagebeoordeling. Het was dus vooral bij studenten dat de interventie de tevredenheid over de stagebeoordeling naar beneden trok.

8.1.3 Het leereffect van de interventies

In alle studies werd ingezoomd op het leereffectcriterium: wat is het effect van de interventies op het leren van studenten? Daartoe werd er zowel gekeken naar de competentiegroei die begeleiders verwachtten van studenten als naar de competentiegroei die studenten rapporteerden. In Studie 1 en Studie 2 gebeurde dit op basis van een vergelijking tussen de experimentele groep en de controlegroep. Leereffecten van de interventie zouden zich immers moeten vertalen in een grotere groei op minstens sommige competenties in de experimentele conditie. In Studie 3 gebeurde dit op basis van een vergelijking met de studenten van het jaar ervoor. Leereffecten van de interventies zouden zich hier moeten vertalen in een grotere groei op minstens sommige competenties in het jaar waarin de interventies werden ingevoerd.

In Studie 1 wezen de resultaten uit dat noch de verwachtte noch de gerapporteerde competentiegroei verschilde over condities heen. In beide condities verwachtten begeleiders competentiegroei in bijna elk domein en de studenten rapporteerden dit ook. In de controleconditie verwachtten begeleiders doorgaans te veel competentiegroei, in de zin dat hun scores significant hoger lagen dan die van studenten. De grote uitzondering werd gevormd door de dimensie Voor een groep spreken. Op dit vlak rapporteerden studenten meer competentiegroei dan verwacht. In de experimentele conditie waren er minder significante verschillen tussen studenten en begeleiders, wat deels te verklaren valt door de kleinere steekproefgrootte. Enkel voor Praktijkgerichtheid en Zelfmanagement verwachtten begeleiders meer competentiegroei dan studenten rapporteerden. Ook hier rapporteerden studenten meer competentiegroei voor Voor een groep spreken dan verwacht. Idem voor Klantgericht werken.

In Studie 2 wezen de resultaten uit dat noch de verwachtte competentiegroei noch de gerapporteerde competentiegroei verschilt over condities heen. In beide condities verwachtten begeleiders competentiegroei in elk domein en de studenten rapporteerden dit ook. Over het algemeen verwachtten begeleiders meer competentiegroei dan studenten rapporteerden. De grote uitzondering werd, net als in Studie 1, gevormd door de dimensie Voor een groep spreken. Op dit vlak rapporteerden studenten meer competentiegroei dan verwacht.

In studie 3 zien we op dit vlak wel een duidelijke de meerwaarde van de interventies. Gegeven dat begeleiders geen leereffect zagen van het nieuwe beoordelingsformulier en gegeven dat dit nieuwe formulier, in tegenstelling tot de leergroepen, ook niet direct een leereffect beoogde, kunnen we ervan uit gaan dat het de invoering van de leergroepen is die deze meerwaarde creëerde. In 2017-2018 verwachtten docenten voor zo goed als alle competenties meer competentiegroei dan in 2016-2017, al was dit verschil omwille van de beperkte steekproefgrootte enkel significant voor Zelfreflectie, Feedback geven en Feedback vragen. Studenten rapporteerden in 2017-2018 ook voor alle domeinen meer competentiegroei dan in 2016-2017. In beide jaren verwachtten docenten meer competentiegroei op vlak van Praktijkgerichtheid dan studenten rapporteerden. Hetzelfde geldt voor Feedback geven in 2017-2018. Op vlak van Leidinggeven, Vernieuwen en Klantgericht werken rapporteerden studenten in 2017-2018 dan weer meer competentiegroei dan verwacht. Over het algemeen kan hieruit geconcludeerd worden dat de invoering van de leergroepen een succes was: de leergroepen zorgden ervoor dat docenten meer overtuigd raakten van de door de stage te verwachten competentiegroei en studenten ervaarden ook daadwerkelijk meer competentiegroei.

8.1.4 Het selectie-effect van de interventies

In alle studies werd vervolgens ook ingezoomd op het selectiecriterium: wat is het effect van de interventies op de punten die studenten krijgen? Daartoe werd zowel gekeken naar het gemiddelde als de spreiding. In Studie 1 en Studie 2 gebeurde dit op basis van een vergelijking tussen de experimentele groep en de controlegroep. Effecten van de interventie op de selectie van studenten zouden zich immers moeten vertalen in een ander gemiddelde of een andere spreiding in de experimentele conditie. In Studie 3 gebeurde dit op basis van een vergelijking met de studenten van het jaar ervoor. Effecten van de interventies op de selectie van studenten zouden zich immers moeten vertalen in een ander gemiddelde of een andere spreiding in het jaar waarin de interventies werden ingevoerd.

In Studie 1 waren er geen significante verschillen tussen de controleconditie en de experimentele conditie, noch in termen van gemiddelde noch in termen van spreiding. Hetzelfde kan worden gezegd voor Studie 2. Ook in Studie 3 waren er geen significante verschillen tussen 2016-2017 en 2017-2018, noch in termen van gemiddelde noch in termen van spreiding. Op basis hiervan dienen we te concluderen dat er geen aanwijzing is voor het feit dat de interventies toelieten om een beter onderscheid te maken tussen goede en slechte studenten. Dit kan uiteraard een hele waaier aan mogelijke oorzaken hebben. De afwezigheid van dit effect zou bijvoorbeeld te wijten kunnen zijn aan het feit dat docenten zowel qua gemiddelde als qua spreiding een standaard hanteren die ze over de jaren heen constant proberen te houden. In dit geval zouden ze zo eventuele effecten van de interventies uitvlakken. Of mogelijks is het ook gewoon zo dat er door de interventies niets in rekening werd gebracht dat al niet sowieso in rekening werd gebracht. Het duidelijkste voorbeeld daarvan wordt misschien gevormd door

Studie 2. Mogelijks bracht het peer assessment niets in rekening wat niet sowieso al in rekening werd gebracht bij het bepalen van de uiteindelijke score omwille van de specifieke context. Het peer assessment werd hier ingevoerd in het kader van de leerwerkplek, en in dit kader is het functioneren van (de studenten in) het team essentieel. De interactie tussen studenten en hun oordeel over elkaar werd daarom ook nog langs andere wegen bevraagd.

8.2 Bespreking van de resultaten

Uit Voorstudie 2 bleek dat studenten vonden dat de beslissingen die worden genomen op onvoldoende assessmentsituaties en beoordelaars gebaseerd waren. Een klassiek probleem dat ook in de literatuur naar voor komt (e.g., Govaerts et al., 2007; Govaerts & van der Vleuten, 2013, Tigelaar & van der Vleuten, 2014; van der Vleuten & Schuwirth, 2005; van Lohuizen et al., 2010). Aan dit probleem probeerden we in Studie 1 en Studie 2 tegemoet te komen en ook de interventie in Voorstudie 1 was hierop gericht. Wanneer de resultaten van Voorstudie 1, Studie 1 en Studie 2 naast elkaar gelegd worden, zou geconcludeerd kunnen worden dat beoordelaars toevoegen niet altijd unaniem positief onthaald wordt. Hoe dit onthaald wordt, lijkt af te hangen van wie als beoordelaars worden ingezet en hoe die worden ingeschakeld. In Studie 1 werden extra begeleiders ingeschakeld die vaker op stagebezoek gingen, meer feedback gaven, en onderling overlegden. Ook al had deze interventie geen direct aantoonbare impact op de beoordeling van de stagebeoordeling (in de zin dat de experimentele groep niet positiever stond ten aanzien van de algemene stagebeoordeling dan de controlegroep), en ook al bracht ze geen competentiegroei teweeg, toch werd de interventie op zich zowel door begeleiders als door studenten positief onthaald. Zowel begeleiders als studenten zagen hier een meerwaarde in voor zo goed als alle kwaliteitscriteria. In Studie 2 daarentegen werden medestudenten ingeschakeld als extra beoordelaars. Deze interventie werd minder positief onthaald. Begeleiders zagen hier wel een meerwaarde in voor zo goed als alle kwaliteitscriteria, maar onder studenten was dit minder het geval. Studenten vonden wel dat de interventie het zicht van begeleiders op hun cognitieve vaardigheden groter maakte, waardoor ze de beoordeling allicht ook als eerlijker en meer acceptabel ervaarden, maar voor de rest stonden ze eerder neutraal ten aanzien van de interventie. In Voorstudie 2, waar eveneens medestudenten werden ingeschakeld als extra beoordelaars, sloeg die relatief neutrale houding bij studenten om in een negatieve houding. De reden hiervoor is allicht dat studenten deze keer niet getraind werden in de beoordelingsmethode en ook geen duidelijk kader of houvast meekregen om hun medestudenten te beoordelen. Begeleiders vinden dat blijkbaar minder noodzakelijk (zij zien nog steeds positieve aspecten aan deze interventie), maar studenten lijken dat louche te vinden. Een voor de hand liggende verklaring is dat studenten onvoldoende vertrouwen hebben in (de onderbouwing van) het (spontane) oordeel van hun medestudenten. Dit is echter geen reden om peer assessment te mijden. Veel hangt er immers van af hoe je dit vormgeeft. In de literatuur worden een aantal factoren geïdentificeerd die mee bepalen hoe tevreden studenten zullen zijn met peer assessment. Zo speelt niet enkel het type peer assessment dat men gebruikt hierin een rol (e.g., Xiao & Lucking, 2008) maar ook (de kwaliteit van) de training van studenten die hieraan voorafgaat, want die is belangrijk om ervoor te zorgen dat de feedback kwaliteitsvol is (e.g., Hovardas, Tsivitanidou, & Zacharia, 2014) en niet te hard beïnvloed wordt door bijvoorbeeld bestaande vriendschapsbanden (e.g., Dochy, Segers, & Sluijsmans, 1999).

Zoals reeds aangestipt, vonden we noch in Studie 1 noch in Studie 2 een direct aantoonbare impact van de interventie op het algemeen oordeel over de stagebeoordeling (in de zin dat de experimentele groep niet positiever stond ten aanzien van de algemene stagebeoordeling dan de controlegroep). Hierbij dient te worden opgemerkt dat het niet zo verwonderlijk is dat de interventies geen effect hadden op het algemeen oordeel over de stagebeoordeling. Zowel in Studie 1 als in Studie 2 werd de algemene beoordeling ook reeds in de controleconditie positief bevonden, met gemiddelde scores van respectievelijk 1.30 en 1.98 op een schaal met 3 als maximum. De meerwaarde die interventies konden teweegbrengen was dus beperkt. In Studie 2 bleken studenten uit de experimentele conditie minder positief te staan ten aanzien van de algemene beoordeling dan studenten uit de controleconditie. Mogelijks is deze minder positieve score een gevolg van het feit het peer assessment niet echt meetelde, waardoor studenten het misschien veel gedoe voor niets vonden. Dat kan hun oordeel beïnvloed hebben.

Ook een directe impact op de competentiegroei van studenten bleef in Studie 1 en 2 uit. In Studie 3 beschikten we niet over de nodige data om na te gaan of de algemene stagebeoordeling erop vooruitgegaan was door de interventies, maar konden we wel vaststellen dat studenten meer competentiegroei rapporteerden naar aanleiding van de invoering van de interventies. Wil men competentiegroei bewerkstelligen, dan lijkt het nodig om hier actief op in te zetten. Iets wat niet gebeurde in Studie 1 en 2 (hier werden enkel indirecte effecten op competentiegroei verwacht door de toename in feedback die gepaard ging met de interventie) maar wel in Studie 3, waar expliciet ingezet werd op soft skills. Studie 3 toont dat een brede aanpak, waarin zowel competenties worden aangeleerd als methoden worden ingevoerd om die ook expliciet in het eindoordeel te betrekken, loont. Dit sluit ook aan bij de aanbevelingen uit de literatuur (e.g., McNamara, 2013). Ook hierbij dient wel te worden opgemerkt dat het niet zo verwonderlijk is dat de interventie geen extra competentiegroei teweegbracht in Studie 1 en 2. Los van het feit dat er in deze studies niet expliciet werd ingezet op (andere) competenties, bleek het ook zo dat ook in de controleconditie alle competenties gemiddeld genomen een erg hoge groeiscore vertoonden (> 4 op een schaal waar 5 het maximum is) en dit zowel bij begeleiders (4.34 in Studie 1 en 4.41 in Studie 2) als bij studenten (4.14 in Studie 1 en 4.04 in Studie 2). De verbetermogelijkheden waren in Studie 3 dus ook groter dan in Studie 1 en 2, waar het al bij voorbaat quasi onmogelijk bleek om op dit vlak nog verbeteringen aan te brengen.

8.3 Aanbevelingen

Wil men de stagebeoordeling verbeteren, dan doet men er, zoals ook uit de literatuur naar voor komt, goed aan om de stagebeoordeling niet los te zien van de stage als leerproces (Dochy et al. 2007; Higgs, 2014; Hodges et al., 2014; McNamara, 2013). Studie 1 en 2 suggereren dat, wanneer je enkel inzet op de beoordeling zelf, zelfs al heb je hierbij oog voor de formatieve aspecten van de beoordelingswijze, je in het beste geval wel een breder draagvlak creëert voor de beoordeling, maar dat je op die manier het leerproces van studenten geen extra boost geeft. Het lijkt dus niet voldoende om enkel in te spelen op de in de literatuur aangereikte verbeteringsmogelijkheden van de beoordeling zelf (= gebruik meer beoordelingsmethoden, meer ratings en observaties, en organiseer trainingen voor docenten en mentoren), maar het lijkt, zoals eveneens in de literatuur gesteld wordt, ook nodig om de stage zelf grondig onder de loep te nemen. Op welke competenties werkt de stage in? Zijn dit wel alle competenties die je in studenten wil zien ontwikkelen? Als dat niet het geval is, dan is het nodig om de stage te hervormen

zodat ook die competenties expliciet aan bod komen. Niet enkel in de stage maar ook in de uiteindelijke beoordeling. Dit was het opzet van Studie 3. In Studie 3 werden onderontwikkelde competenties geïdentificeerd en werd nagedacht over hoe die konden worden versterkt, en dit zowel vanuit de instelling (= door de invoering van leergroepen) als vanop de stageplek zelf (= door mentoren te wijzen op het belang van goede feedback). Teneinde die competenties ook in het hoofd van studenten aan belang te doen winnen, werden deze competenties in overeenstemming met het principe van constructive alignment (Biggs, 1996; 2007) ook expliciet betrokken in het eindoordeel (= het nieuwe beoordelingsformulier). Een dergelijke brede interventie, die niet alleen inspeelt op elk van de in de literatuur aangereikte verbeteringsmogelijkheden, maar tevens expliciet inspeelt op de pijnpunten van de stage, lijkt meest effectief om te bekomen wat men beoogt: het opleiden van studenten tot competente professionals die in staat zijn levenslang zelfstandig te leren.

Voor opleidingen die zich afvragen of hun stage(beoordeling) kan of dient te worden verbeterd, lijkt het een goed idee om bij alle betrokkenen (i.e., studenten en begeleiders) een voorafname te doen van zowel de kwaliteitscriteria als de competentiegroeimeter. Op basis daarvan kan men dan zien of er pijnpunten of lacunes zitten in de stage(beoordeling). Zijn er kwaliteitscriteria die zwak scoren? Zijn er competenties die onvoldoende groei kennen doorheen de stage? Eens dergelijke lacunes geïdentificeerd zijn, kan men nadenken over remediërende interventies. Daartoe kan men zijn heil zoeken in de literatuur. Die biedt handvaten en inspiratie ter verbetering van de stage(beoordeling). Uiteraard is het niet per definitie zo dat er problematische aspecten zijn aan een stage(beoordeling). Een voorafname kan vanzelfsprekend ook uitwijzen dat men "goed bezig" is. Iets wat zich ook toont in de huidige studies. Zo zagen we in Studie 1 en 2 bijvoorbeeld dat alle competenties ook reeds in de controleconditie erg hoge groeiscores vertoonden, en dit zowel bij begeleiders als bij studenten.

8.4 Beperkingen

Elke studie die we uitvoerden had zijn eigen beperkingen. Studie 1 en Studie 2 hadden een onvoldoende brede scope doordat ze niet echt expliciet inzetten op de stage als proces (ook al werd er wel voor gezorgd dat de interventie ook een formatief aspect kreeg), in Studie 1 was de verdeling van studenten over de experimentele groep en de controlegroep niet optimaal en werd er onvoldoende toegezien op de uitvoering van ons onderzoeksdesign (deden docenten en mentoren wel wat van hen verwacht werd?), Studie 3 werd doorgevoerd in een opleiding die te klein was om met een experimenteel design te werken, … Deze beperkingen kwamen reeds eerder aan bod en we wensen daar hier niet dieper op in te gaan. We wensen ons hier te beperken tot een algemene beperking, eigen aan al onze studies.

Om de impact van de interventies op het leren van studenten na te gaan, werd gebruik gemaakt van de competentiegroeimeter (Braun et al. 2008; Braun & Leidner, 2009). Die meet zelfgerapporteerde competentiegroei. Toekomstige studies zouden er goed aan doen om niet alleen te kijken naar zelfgerapporteerde competentiegroei maar om een dergelijke competentiegroei ook te proberen kwantificeren. Op dit vlak was de enige kwantificeerbare maat in ons onderzoek het behaalde eindresultaat. In geen enkele studie vonden we effecten van de interventie op het eindresultaat, maar we betoogden reeds dat dit een hele reeks oorzaken kan hebben. Zo zou de afwezigheid van dit effect te wijten kunnen zijn aan het feit dat docenten zowel qua gemiddelde als qua spreiding een standaard hanteren die ze over de jaren heen constant proberen te

houden, waardoor een eventueel effect van de interventies uitgevlakt zou worden. Vanuit die optiek zou toekomstig onderzoek er dan ook goed aan doen om niet alleen naar het eindcijfer te gaan kijken, maar ook op zoek te gaan naar methoden om de vooropgestelde competentiegroei meer in detail in objectieve cijfers te gieten (cfr. het principe van constructive alignment; Biggs, 1996, 2007). Studenten mogen dan wel competentiegroei rapporteren, maar is die ook zichtbaar voor externe beoordelaars (docenten, mentoren, medestudenten, …)? Mocht de zelfgerapporteerde competentiegroei gestaafd kunnen worden door externe beoordelaars, dan zou dit het vertrouwen in de meerwaarde van een interventie alleen maar versterken.

9 REFERENTIES

Anema, M. G., & McCoy, J. (2010). Competency-based nursing education: Guide to achieving outstanding learner outcomes. Springer: New York.

Aronson, L. (2011). Twelve tips for teaching reflection at all levels of medical education. Medical Teacher, 33, 200-205.

Atkinson, D., & Murray, M. (1987). Improving interrater reliability. Paper presented at the Annual Meeting of the Conference on College Composition and Communication, March 19-21, Atlanta.

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2006). The wheel of competency assessment: Presenting quality criteria for Competency Assessment Programmes. Studies in Educational Evaluation, 32, 153-170.

Baartman, L. K. J., Bastiaens, T. J., Kirschner, P. A., & Van der Vleuten, C. P. M. (2007). Teachers' opinions on quality criteria for Competency Assessment Programs. Teaching and Teacher Education, 23, 857–867.

Baartman, L., Kloppenburg, R., & Prins, F. (2014). Kwaliteit van toetsprogramma's. In H. Van Berkel, A. Bax, & D. Joosten-ten Brink (Eds.), Toetsen in het hoger onderwijs. Bohn Stafleu van Loghum.

Biggs, J. (1996). Enhancing teaching through constructive alignment. Higher Education, 32, 347–64.

Biggs, J. (2003). Teaching for quality learning at university. United Kingdom: The Society for Research into Higher Education and Open University Press.

Boud, D., & Falchikov, N. (2006). Aligning assessment with long-term learning. Assessment and Evaluation in Higher Education, 3, 399–413.

Bradley, R. A., & Terry, M. E. (1952). Rank analysis of incomplete block designs. The method of paired comparisons. Biometrika, 39, 324-345.

Braun, E., Gusy, B., Leidner, B. & Hannover, B. (2008). Kompetenzorientierte Lehrevaluation – Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). Diagnostica, 54, 30-42.

Braun, E., & Leidner, B. (2009). Academic Course Evaluation. Theoretical and Empirical Distinctions Between Self-Rated Gain in Competences and Satisfaction with Teaching Behavior. European Psychologist, 14, 297–306.

Carracio, C., Wolfsthal, S. D., Englander, R., Ferentz, K., & Martin, C. (2002). Shifting paradigms: From Flexner to competencies. Academic Medicine, 77, 361-367.

Comfort, J. & Ferns, S. (2014). ePortfolios as evidence of standards and outcomes in work-integrated learning. Asia-Pacific Journal of Cooperative Education, 15, 267-278.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika. 16, 297-334.

Dannefer, E. F. (2013). Beyond assessment of learning toward assessment for learning: Educating tomorrow's physicians. Medical Teacher, 35, 560-563.

Deci, E. L., & Ryan, R. M. (2000). The "what" and the "why" of goal pursuits: Human needs and the self-determination of behavior. Psychological Inquiry, 11, 227–268.

Dochy, F., Segers, M., Gifbels, D., & Struyven, K. (1997). Assessment engineering: breaking down barriers between teaching and learning, and assessment. In D. Boud & N. Falchikov (Eds.), Rethinking assessment in higher education: Learning for the longer term, London / New York: Routledge.

Dochy, F., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and coassessment in higher education: A review. Studies in Higher Education, 24, 331-350.

Downing, S. M. (2004). Reliability: On the reproducibility of assessment data. Medical Education, 38, 1006-1012.

Duriez, B., Kummu, M., Sebreghts, E., & Smis, D. (2016). Beginassessments: Een onderzoek naar de effectiviteit van beginassessments binnen de KdG-context. Karel De Grote Hogeschool, Antwerpen: Intern Rapport.

Duquet, N., Gloriaux, I., Laurijssen, I., & Van Dorsselaer, Y. (2006). Wit krijt schrijft beter. Schoolloopbanen van allochtone jongeren in beeld. Antwerpen/Apeldoorn: Garant.

Embo, M., Driessen, E. W., Valcke, M., & van der Vleuten, C. P. M. (2015). Integrating learning assessment and supervision in a competency framework for clinical workplace education. Nurse Education Today, 15, 341–346.

Epstein, R. M., & Hundert, E. M. (2002). Defining and assessing professional competence. JAMA, 287, 226-235.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. Review of educational research, 70, 287-322.

Fastré, G. M. J., van der Klink, M. R., Sluijsmans, D., & van Merriënboer, J. J. G. (2013). Towards an integrated model for developing sustainable assessment skills. Assessment & Evaluation in Higher Education, 38, 611-630.

Frank, J. R., Snell, L. S., Ten Cate, O., Holmboe, E. S., Carracio, C., Swing, S. R., Harris, P., Glasgow, N. J., Campbell, C., Dath, D., Harden, R. M., Iobst, W., Long, D. M., Mungroo, R., Richardson, D. L., Sherbino, J., Silver, I., Taber, S., Talbot, M, & Harris, K. A. (2010). Competency-based medical education: Theory to practice. Medical Teacher, 32, 638-645.

Frank, J.R., Snell L., & Sherbino J. (Eds) (2015). CanMEDS 2015 Physician Competency Framework. Ottawa: Royal College of Physicians and Surgeons of Canada.

Govaerts, M., & van der Vleuten, C. P. M. (2013). Validity in work-based assessments: Expanding our horizons. Medical Education, 47, 1164-1174.

Govaerts, M. J. B., van der Vleuten, C. P. M., Schuwirth, L. W. T., & Muijtjens, A. M. M. (2007). Broadening perspectives on clinical performance assessment: Rethinking the nature of in-training assessment. Advances in Health Sciences Education, 12, 239-260.

Govaerts, M. J. B., Van de Wiel, M. W. J., Schuwirth, L. W. T., van der Vleuten, C. P. M., & Muijtjens, A. M. M. (2013). Workplace-based assessment: Raters' performance theories and constructs. Advances in Health Sciences Education, 18, 375-396.

Gray, J. D. (1996). Global rating scales in residency education. Academic Medicine, 71, 856-863.

Hattie, J. A., & Timperley, H. (2007). The power of feedback. Review of Educational Research, 77, 81-112.

Higgs, J. (2014). Assessing the immeasurables of practice. Asia-Pacific Journal of Cooperative Education (Special Issue), 15, 253-267.

Hodges, D., Eames, C., & Coll, R. K. (2014). Theoretical perspectives on assessment in cooperative education placements. Asia-Pacific Journal of Cooperative Education (Special Issue), 15, 189-207.

Hovardas, T., Tsivitanidou, O. E., & Zacharia, Z. C. (2014). Peer versus expert feedback: An investigation of the quality of peer feedback among secondary school students. Computers & Education, 71, 133-152.

Hunt, D. D. (1992). Functional and dysfunctional characteristics of the prevailing model of clinical evaluation systems in North American medical schools. Academic Medicine, 67, 254-259.

Kicken, W., S. Brand-Gruwel, J.J.G., Van Merrienboer, & Slot, W. (2009). Design and evaluation of a development portfolio: How to improve students' self-directed learning skills. Instructional Science, 37, 453–73.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. Journal of Personality and social Psychology, 77, 1121-1134.

Laming, D. (2003). Human judgment: The eye of the beholder. Andover: Cengage Learing EMEA.

Lesterhuis, M., Verhavert, S., Coertjens, L., Donche, V., & De Mayer, S. (2017). Comparative judgement as a promising alternative to score competences. In E. Cano, & G. Ion (Eds.), Innovative practices for higher education assessment and measurement. IGI Global: Hershey (PA).

Levene, H. (1960). Robust tests for equality of variances. Contributions to probability and statistics, 1, 278-292.

Luce, R. D. (1959). On the possible psychophysical laws. Psychological Review, 66, 81-95.

McNamara, J. (2013) The challenge of assessing professional competence in work integrated learning. Assessment & Evaluation in Higher Education, 38, 183-197.

Norcini, J., Anderson, B., Bollela, V., Burch, V., Costa, M. J., Duvivier, R., Galbraith, R., Hays, R., Kent, A., Perrott, V., & Roberts, T. (2011). Criteria for good assessment: Consensus statement and recommendations from the Ottawa 2010 Conference. Medical Teacher, 33, 206-214.

Norcini, J., & Burch, V. (2007). Workplace-based assessment as an educational tool: AMEE Guide No. 31. Medical Teacher, 29, 855-871.

Pijl-Zieber, M. (2014). Mind the gap: Quantifying the performance gap between observed and required clinical competencies in undergraduate nursing students. Nurse Education Today, 34, 667-669.

Prescott-Clements, L., van der Vleuten, C. P. M., Schuwirth, L. W. T., Hurst, Y., & Rennie, J. S. (2008). Evidence for validity within workplace assessment: The Longitudinal Evaluation of Performance (LEP). Medical Education, 42, 488-495.

Sanches, J. I., & DeLa Torre, P. (1996). A second look at the relationship between rating and behavioral accuracy in performance appraisal. Journal of Applied Psychology, 81, 3-10.

Short, J. P. (1993). The importance of strong evaluation standards and procedures in training residents. Academic Medicine, 68, 522-525.

Thurstone, L. L. (1927). A law of comparative judgment. Psychological Review, 34, 273-286.

Tigelaar, D. E. H., & van der Vleuten, C. P. M. (2014. Assessment of professional competence. In S. Billett, C. Harteis, & H. Gruber (Eds.), International handbook of research in professional and practice-based Learning (pp. 1237-1269). Springer: New York.

Turnbull, J., Gray, J., & MacFadyen, J. (1998). Improving in-training evaluation programs. Journal of General Internal Medicine, 13, 317-323.

Turnbull, J., MacFadyen, J., van Barneveld, C., & Norman, G. (2000). Clinical Work Sampling: A new approach to the problem of in-training evaluation. Journal of General Internal Medicine, 15, 556-561.

van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. Medical Education, 39, 309-317.

van der Vleuten, C. P. M., Schuwirth, L. W., Scheele, F., Driessen, E. W., & Hodges, B. (2010). The assessment of professional competence: Building blocks for theory development. Best Practice & Research Clinical Obstetrics and Gynaecology, 24, 703-719.

van Lohuizen, M., Kuks, J. B. M., van Hell, E. A., Raat, A.N., Stewart, R. E., & Cohen-Schotanus, J. (2010). The reliability of in-training assessment when performance improvement is taken into account. Advances in Health Science Education, 15, 659-669.

Verney, T. P., Holoviak, S. J., & Winter, A. S. (2009). Enhancing the reliability of internship evaluations. Journal of Applied Business and Economics, 9, 22-33.

Walsh, A. (2007). An exploration of Biggs' constructive alignment in the context of workbased learning. Assessment and Evaluation in Higher Education, 32, 79-87.

Xiao, Y., & Lucking, R. (2008). The impact of two types of peer assessment on students' performance and satisfaction within a Wiki environment. Internet and Higher Education, 11, 186,193.

Young. J. Q., Lieu, S., O'Sullivan, P., & Tong, L. (2011). Development and initial testing of a structured clinical observation tool to assess pharmacotherapy competence. Academic Psychiatry, 35, 27-24.

Zegwaard, K., Coll. R. K., & Hodges, D. (2003). Assessment of Workplace Learning: A Framework. Asia-Pacific Journal of Cooperative Education, 4, 10-18.

10 APPENDICES

10.1 Kwaliteitscriteria zoals gebruikt in Voorstudie 1

Deze appendix bevat de vragenlijst inzake kwaliteitscriteria die werd gebruikt in Voorstudie 1. De items werden voorafgegaan door de introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept. Negatief geformuleerde items worden aangegeven met "(-)".

Onderstaande vragen polsen naar hoe je het gebruik van deze beoordelingswijze hebt ervaren. Wat vind je van deze manier van beoordelen? Geef aan de hand een cijfer aan in welke mate je het met onderstaande uitspraken eens bent. Ben je het helemaal oneens met een bepaalde uitspraak, omcirkel dan -3. Ben je het helemaal eens, omcirkel dan +3. De tussenliggende cijfers kun je gebruiken om een meer genuanceerd antwoord te geven. Deze gegevens worden verzameld in het kader van een onderzoekproject rond stagebeoordelingen. De gegevens die je verstrekt zullen volledig anoniem verwerkt worden.

Geschiktheid

1. Deze beoordeling is geschikt om de bekwaamheid van studenten te beoordelen.
2. Deze beoordeling sluit aan bij de competenties die studenten moeten ontwikkelen.
3. Deze beoordeling sluit aan bij de competenties die de opleiding vereist.
4. Deze beoordeling geeft een goed beeld van studenten hun competenties.
5. Deze beoordeling laat toe in te schatten of studenten klaar zijn voor het werkveld.

Vergelijkbaarheid

6. De criteria zijn voor alle studenten vergelijkbaar.
7. De opdrachten zijn voor alle studenten vergelijkbaar.
8. Studenten worden op een vergelijkbare wijze beoordeeld.
9. De criteria worden door alle beoordelaars op dezelfde wijze gehanteerd.
10. De verschillende beoordelaars hanteren elk hun eigen criteria. (-)

Herhaalbaarheid

11. Het eindoordeel is gebaseerd op meerdere observaties en situaties.
12. Het eindoordeel is gestoeld op het oordeel van meerdere beoordelaars.
13. Op basis van deze info zouden andere beoordelaars hetzelfde oordeel vellen.
14. Het eindoordeel zou anders zijn mocht iemand anders beoordeeld hebben. (-)
15. Het eindoordeel is subjectief en afhankelijk van de beoordelaar. (-)

Transparantie

16. Het is duidelijk wat er wordt beoordeeld.
17. Het is duidelijk op welke manier er wordt beoordeeld.
18. Het is duidelijk hoe een oordeel tot stand komt.
19. Het is duidelijk wat de criteria concreet betekenen in de praktijk.
20. Het is duidelijk hoe verschillende oordelen tot een eindoordeel leiden.

Acceptatie

21. Wat er beoordeeld wordt, is voor mij aanvaardbaar.
22. De manier waarop er beoordeeld wordt, is voor mij aanvaardbaar.
23. Ik kan me vinden in de manier waarop een oordeel tot stand komt.
24. Ik kan me vinden in de criteria die moeten gehaald worden.
25. Ik kan me vinden in hoe de verschillende oordelen tot een eindoordeel leiden.

Authenticiteit

26. Deze beoordeling heeft een duidelijke link met de professionele context.
27. Deze beoordeling gebruikt situaties die ook in een professionele context voorkomen.
28. Deze beoordeling weerspiegelt wat professioneel van studenten verwacht wordt.
29. Deze beoordeling sluit aan bij wat het werkveld van pas afgestudeerden verwacht.
30. Deze beoordeling sluit aan bij wat het werkveld van laatstejaars verwacht.

Eerlijkheid

31. Deze beoordeling is een eerlijke manier van beoordelen.
32. Deze beoordelingswijze laat toe een gefundeerd en eerlijk oordeel te vellen.
33. De beoordelaars hebben voldoende zicht op het functioneren van studenten.
34. Studenten krijgen zo voldoende kansen om hun competentie te tonen.
35. Deze beoordeling geeft een eerlijk beeld van studenten.

Complexiteit

36. Deze beoordeling houdt rekening met de redeneringen die studenten maken.
37. Deze beoordeling vereist dat studenten hun beslissingen kunnen motiveren.
38. Deze beoordeling weerspiegelt het cognitief vermogen van studenten.
39. Deze beoordeling legt het denkproces van studenten tijdens hun stage bloot.
40. Deze beoordeling peilt naar de aanpak van studenten en niet enkel naar resultaat.

Betekenisvolheid

41. Deze beoordeling is ook een leermoment.
42. Deze beoordeling biedt studenten leerkansen.
43. Deze beoordeling verschaft studenten waardevolle feedback.
44. Deze beoordeling geeft studenten de kans om dingen bij te leren.
45. Deze beoordeling geeft studenten de kans om inzichten te verwerven.

Zelfsturing

46. Deze beoordeling leert studenten reflecteren.
47. Deze beoordeling leert studenten zichzelf beoordelen.
48. Deze beoordeling leert studenten hun eigen leerdoelen formuleren.
49. Deze beoordeling helpt studenten hun eigen leerproces in handen te nemen.
50. Deze beoordeling toont studenten hoe ze aan zwakke punten kunnen werken.

Tijd en kosten

51. Deze beoordelingswijze is haalbaar qua tijdsinvestering.
52. De voordelen van deze beoordelingswijze wegen op tegen de nadelen.
53. Ondanks de gevraagde inspanningen loont deze beoordelingswijze de moeite.
54. Het is een goed idee om te investeren in deze manier van beoordelen.
55. De meerwaarde van deze beoordelingswijze weegt niet op tegen de inspanning. (-)

Consequenties: Selectie

56. Deze beoordeling selecteert studenten op beroepsgeschiktheid.
57. Deze beoordeling laat toe goed presterende studenten te identificeren.
58. Deze beoordeling laat toe slecht presterende studenten te identificeren.
59. Deze beoordeling laat toe het functioneren van studenten te vergelijken.
60. Deze beoordeling laat toe goede van slechte studenten te onderscheiden.

Consequenties: Leereffect

61. Deze beoordeling stimuleert de gewenste leerprocessen.
62. Deze beoordeling geeft studenten rijke feedback over hun prestatie.
63. Deze beoordeling zet studenten aan na te denken over hun eigen ontwikkeling.
64. Deze beoordeling stimuleert het zelfstandig ontwikkelvermogen van studenten.
65. Deze beoordeling helpt studenten hun eigen kennen en kunnen inschatten.

Consequenties: Motivatie

66. Deze beoordeling kwam mijn motivatie ten goede.
67. Deze beoordeling heeft mij zin gegeven om verder te gaan in mijn vakgebied.
68. Deze beoordeling heeft mijn interesse in mijn vakgebied vergroot.
69. Deze beoordeling heeft ervoor gezorgd dat ik mezelf verder wil ontwikkelen.
70. Deze beoordeling heeft mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

71. Deze beoordeling verschaft info over het gemiddeld niveau van studenten.
72. Deze beoordeling verschaft rijke informatie over de kwaliteit van de stage.
73. Deze beoordeling verschaft rijke informatie over de kwaliteit van de begeleiding.
74. Deze beoordeling kan gebruikt worden om de begeleiding te verbeteren.
75. Deze beoordeling kan gebruikt worden om de stage te verbeteren.

10.2 Kwaliteitscriteria zoals gebruikt in Voorstudie 2

Deze appendix bevat de vragenlijst inzake kwaliteitscriteria die werd gebruikt in Voorstudie 2. De items werden voorafgegaan door de introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept. Negatief geformuleerde items worden aangegeven met "(-)". De items die doorstreept zijn, maken geen deel uit van de verkorte schalen die werden geconstrueerd.

Aan de hand van onderstaande stellingen willen wij polsen naar hoe je de stagebeoordeling voor de voorbije stageperiode hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. ~~Deze beoordeling is geschikt om de bekwaamheid van studenten te beoordelen~~.
2. Deze beoordeling sluit aan bij de competenties die studenten moeten ontwikkelen.
3. Deze beoordeling sluit aan bij de competenties die de opleiding vereist.
4. Deze beoordeling geeft een goed beeld van studenten hun competenties.
5. ~~Deze beoordeling laat toe in te schatten of studenten klaar zijn voor het werkveld~~.

Vergelijkbaarheid

6. ~~De criteria zijn voor alle studenten vergelijkbaar~~.
7. ~~De opdrachten zijn voor alle studenten vergelijkbaar~~.
8. Studenten worden op een vergelijkbare wijze beoordeeld.
9. Wat verwacht wordt, is voor alle studenten vergelijkbaar van aard.
10. Het beoordelingsproces verloopt voor alle studenten of dezelfde manier.

Herhaalbaarheid

11. Voor deze beoordeling maakt het niet veel uit wie de beoordelaar is.
12. Verschillende beoordelaars komen op deze manier al snel tot dezelfde conclusie.
13. Op basis van deze info zouden andere beoordelaars hetzelfde oordeel vellen.
14. ~~Het eindoordeel zou anders zijn mocht iemand anders beoordeeld hebben~~. (-)
15. ~~Het eindoordeel is subjectief en afhankelijk van de beoordelaar~~. (-)

Transparantie

16. Het is duidelijk wat er wordt beoordeeld.
17. Het is duidelijk op welke manier er wordt beoordeeld.
18. Het is duidelijk hoe een oordeel tot stand komt.
19. ~~Het is duidelijk wat de criteria concreet betekenen in de praktijk~~.
20. ~~Het is duidelijk hoe verschillende oordelen tot een eindoordeel leiden~~.

Acceptatie

21. Wat er beoordeeld wordt, is voor mij aanvaardbaar.
22. De manier waarop er beoordeeld wordt, is voor mij aanvaardbaar.
23. Ik kan me vinden in de manier waarop een oordeel tot stand komt.

24. ~~Ik kan me vinden in de criteria die moeten gehaald worden~~.
25. ~~Ik kan me vinden in hoe de verschillende oordelen tot een eindoordeel leiden~~.

Authenticiteit

26. Deze beoordeling heeft een duidelijke link met de professionele context.
27. Deze beoordeling gebruikt situaties die ook in een professionele context voorkomen.
28. Deze beoordeling weerspiegelt wat professioneel van studenten verwacht wordt.
29. ~~Deze beoordeling sluit aan bij wat het werkveld van pas afgestudeerden verwacht~~.
30. ~~Deze beoordeling sluit aan bij wat het werkveld van laatstejaars verwacht~~.

Eerlijkheid

31. Deze beoordeling is een eerlijke manier van beoordelen.
32. Deze beoordelingswijze laat toe een gefundeerd en eerlijk oordeel te vellen.
33. ~~De beoordelaars hebben voldoende zicht op het functioneren van studenten~~.
34. ~~Studenten krijgen zo voldoende kansen om hun competentie te tonen~~.
35. Deze beoordeling geeft een eerlijk beeld van studenten.

Complexiteit

36. Deze beoordeling houdt rekening met de redeneringen die studenten maken.
37. ~~Deze beoordeling vereist dat studenten hun beslissingen kunnen motiveren~~.
38. Deze beoordeling weerspiegelt het cognitief vermogen van studenten.
39. Deze beoordeling legt het denkproces van studenten tijdens hun stage bloot.
40. ~~Deze beoordeling peilt naar de aanpak van studenten en niet enkel naar resultaat~~.

Betekenisvolheid

41. Deze beoordeling is ook een leermoment.
42. Deze beoordeling biedt studenten leerkansen.
43. ~~Deze beoordeling verschaft studenten waardevolle feedback~~.
44. Deze beoordeling geeft studenten de kans om dingen bij te leren.
45. ~~Deze beoordeling geeft studenten de kans om inzichten te verwerven~~.

Tijd en kosten

46. ~~Deze beoordelingswijze is haalbaar qua tijdsinvestering~~.
47. De voordelen van deze beoordelingswijze wegen op tegen de nadelen.
48. Ondanks de gevraagde inspanningen loont deze beoordelingswijze de moeite.
49. Het is een goed idee om te investeren in deze manier van beoordelen.
50. ~~De meerwaarde van deze beoordelingswijze weegt niet op tegen de inspanning~~. (-)

Consequenties: Selectie

51. ~~Deze beoordeling selecteert studenten op beroepsgeschiktheid~~.
52. Deze beoordeling laat toe goed presterende studenten te identificeren.
53. Deze beoordeling laat toe slecht presterende studenten te identificeren.
54. Deze beoordeling laat toe het functioneren van studenten te vergelijken.
55. ~~Deze beoordeling laat toe goede van slechte studenten te onderscheiden~~.

Consequenties: Leereffect

56. ~~Deze beoordeling stimuleert de gewenste leerprocessen~~.
57. Deze beoordeling geeft studenten rijke feedback over hun prestatie.

58. Deze beoordeling zet studenten aan na te denken over hun eigen ontwikkeling.
59. Deze beoordeling stimuleert het zelfstandig ontwikkelvermogen van studenten.
60. ~~Deze beoordeling helpt studenten hun eigen kennen en kunnen inschatten~~.

Consequenties: Motivatie

61. ~~Deze beoordeling kwam mijn motivatie ten goede~~.
62. Deze beoordeling heeft mij zin gegeven om verder te gaan in mijn vakgebied.
63. Deze beoordeling heeft mijn interesse in mijn vakgebied vergroot.
64. ~~Deze beoordeling heeft ervoor gezorgd dat ik mezelf verder wil ontwikkelen~~.
65. Deze beoordeling heeft mij nieuwe energie gegeven.

10.3 Studie 1: Kwaliteitscriteria voor de specifieke interventie

---

Deze appendix bevat de vragenlijst inzake kwaliteitscriteria voor de specifieke interventie zoals gebruikt in Studie 1. Items werden afgestemd op de specifieke context en voorafgegaan door een aan de studie aangepaste introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

Dit jaar werd er in het kader van een onderzoek naar stagebeoordelingen een nieuw onderdeel toegevoegd aan de stagebeoordeling. In het kader van dit onderzoek zijn we geïnteresseerd in hoe je dit hebt ervaren. Specifiek gaat het over ***extra lesbezoeken*** (= er werden meer lesbezoeken gedaan door meer mensen die daar ook onderling over overlegden). Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. Extra bezoeken geven meer zicht op de competenties die studenten moeten ontwikkelen.
2. Extra bezoeken verbeteren het zicht op de competenties die de opleiding vereist.
3. Extra bezoeken zorgen voor een beter beeld van studenten hun competenties.

Vergelijkbaarheid

4. Extra bezoeken helpen studenten op vergelijkbare wijze te beoordelen.
5. Extra bezoeken helpen de verwachtingen voor alle studenten vergelijkbaar te maken.
6. Extra bezoeken helpen de beoordeling voor alle studenten gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Door de extra bezoeken maakt het minder uit wie de beoordelaars zijn.
8. Door de extra bezoeken komen beoordelaars sneller tot dezelfde conclusie.
9. Door de extra bezoeken vellen beoordelaars gemakkelijker hetzelfde oordeel.

Transparantie

10. Door de extra bezoeken wordt het duidelijker wat er wordt beoordeeld.
11. Door de extra bezoeken wordt het duidelijker op welke manier er wordt beoordeeld.
12. Door de extra bezoeken wordt het duidelijker hoe een oordeel tot stand komt.

Acceptatie

13. Extra bezoeken maken wat er beoordeeld wordt voor mij meer aanvaardbaar.
14. Extra bezoeken maken de manier waarop er beoordeeld wordt meer aanvaardbaar.
15. Extra bezoeken maken dat ik me beter kan vinden in hoe een oordeel tot stand.

Authenticiteit

16. Extra bezoeken verduidelijken de link met de professionele context.
17. Extra bezoeken gebruiken extra situaties die ook in professionele context voorkomen.
18. Extra bezoeken weerspiegelen beter wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. Extra bezoeken zorgen voor een eerlijkere manier van beoordelen
20. Extra bezoeken maken het makkelijker om een gefundeerd en eerlijk oordeel te vellen.
21. Extra bezoeken geven een eerlijker beeld van studenten.

Complexiteit

22. Extra bezoeken helpen rekening te houden met de redeneringen die studenten maken.
23. Extra bezoeken helpen het cognitief vermogen van studenten te weerspiegelen.
24. Extra bezoeken helpen het denkproces van studenten tijdens hun stage blootleggen.

Betekenisvolheid

25. De extra bezoeken bieden bijkomende leermomenten.
26. De extra bezoeken bieden studenten extra leerkansen.
27. De extra bezoeken geven studenten extra kansen om dingen bij te leren.

Tijd en kosten

28. De voordelen van de extra bezoeken wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen lonen de extra bezoeken de moeite.
30. Het is een goed idee om te investeren in extra bezoeken.

Consequenties: Selectie

31. Extra bezoeken maken het makkelijker om goed presterende studenten te identificeren.
32. Extra bezoeken maken het makkelijker om slecht presterende studenten te identificeren.
33. Extra bezoeken zorgen dat het functioneren van studenten beter vergeleken kan worden.

Consequenties: Leereffect

34. Extra bezoeken geven studenten rijkere feedback over hun prestatie.
35. Extra bezoeken zetten studenten extra aan na te denken over hun eigen ontwikkeling.
36. Extra bezoeken stimuleren het zelfstandig ontwikkelvermogen van studenten nog meer.

Consequenties: Motivatie

37. De extra bezoeken hebben mij zin gegeven om verder te gaan in mijn vakgebied.
38. De extra bezoeken hebben mijn interesse in mijn vakgebied vergroot.
39. De extra bezoeken hebben mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

40. Extra bezoeken verschaffen bijkomende info over het gemiddeld niveau van studenten.
41. Extra bezoeken verschaffen rijkere informatie over de kwaliteit van de stage.
42. Extra bezoeken geven extra info die kan gebruikt worden om de stage te verbeteren.

10.4 Kwaliteitscriteria voor de algemene beoordeling

Deze appendix bevat de vragenlijst inzake kwaliteitscriteria voor de algemene beoordeling zoals gebruikt in Studie 1 en 2. De items werden voorafgegaan een aan de studie in kwestie aangepaste variant van door de introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

Aan de hand van onderstaande stellingen willen wij polsen naar hoe je de stagebeoordeling voor *{opleidingsonderdeel}* hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. Deze beoordeling sluit aan bij de competenties die studenten moeten ontwikkelen.
2. Deze beoordeling sluit aan bij de competenties die de opleiding vereist.
3. Deze beoordeling geeft een goed beeld van studenten hun competenties.

Vergelijkbaarheid

4. Studenten worden op een vergelijkbare wijze beoordeeld.
5. Wat verwacht wordt, is voor alle studenten vergelijkbaar van aard.
6. Het beoordelingsproces verloopt voor alle studenten of dezelfde manier.

Herhaalbaarheid

7. Voor deze beoordeling maakt het niet veel uit wie de beoordelaar is.
8. Verschillende beoordelaars komen op deze manier al snel tot dezelfde conclusie.
9. Op basis van deze info zouden andere beoordelaars hetzelfde oordeel vellen.

Transparantie

10. Het is duidelijk wat er wordt beoordeeld.
11. Het is duidelijk op welke manier er wordt beoordeeld.
12. Het is duidelijk hoe een oordeel tot stand komt.

Acceptatie

13. Wat er beoordeeld wordt, is voor mij aanvaardbaar.
14. De manier waarop er beoordeeld wordt, is voor mij aanvaardbaar.
15. Ik kan me vinden in de manier waarop een oordeel tot stand komt.

Authenticiteit

16. Deze beoordeling heeft een duidelijke link met de professionele context.
17. Deze beoordeling gebruikt situaties die ook in een professionele context voorkomen.
18. Deze beoordeling weerspiegelt wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. Deze beoordeling is een eerlijke manier van beoordelen.
20. Deze beoordelingswijze laat toe een gefundeerd en eerlijk oordeel te vellen.
21. Deze beoordeling geeft een eerlijk beeld van studenten.

Complexiteit

22. Deze beoordeling houdt rekening met de redeneringen die studenten maken.
23. Deze beoordeling weerspiegelt het cognitief vermogen van studenten.
24. Deze beoordeling legt het denkproces van studenten tijdens hun stage bloot.

Betekenisvolheid

25. Deze beoordeling is ook een leermoment.
26. Deze beoordeling biedt studenten leerkansen.
27. Deze beoordeling geeft studenten de kans om dingen bij te leren.

Tijd en kosten

28. De voordelen van deze beoordelingswijze wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen loont deze beoordelingswijze de moeite.
30. Het is een goed idee om te investeren in deze manier van beoordelen.

Consequenties: Selectie

31. Deze beoordeling laat toe goed presterende studenten te identificeren.
32. Deze beoordeling laat toe slecht presterende studenten te identificeren.
33. Deze beoordeling laat toe het functioneren van studenten te vergelijken.

Consequenties: Leereffect

34. Deze beoordeling geeft studenten rijke feedback over hun prestatie.
35. Deze beoordeling zet studenten aan na te denken over hun eigen ontwikkeling.
36. Deze beoordeling stimuleert het zelfstandig ontwikkelvermogen van studenten.

Consequenties: Motivatie

37. Deze beoordeling heeft mij zin gegeven om verder te gaan in mijn vakgebied.
38. Deze beoordeling heeft mijn interesse in mijn vakgebied vergroot.
39. Deze beoordeling heeft mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

40. Deze beoordeling verschaft info over het gemiddeld niveau van studenten.
41. Deze beoordeling verschaft rijke informatie over de kwaliteit van de stage.
42. Deze beoordeling kan gebruikt worden om de stage te verbeteren.

10.5 Studie 2: Kwaliteitscriteria voor de specifieke interventie

---

Deze appendix bevat de vragenlijst inzake kwaliteitscriteria voor de specifieke interventie zoals gebruikt in Studie 2. Items werden afgestemd op de specifieke context en voorafgegaan door een aan de studie aangepaste introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

In het kader van een onderzoek rond de stages binnen KdG zijn we geïnteresseerd in hoe je het ***peer assessment*** (= het inzetten van medestudenten in het beoordelingsproces) hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. Peer assessment geeft meer zicht op de competenties die studenten moeten ontwikkelen.
2. Het peer assessment verbetert het zicht op de competenties die de opleiding vereist.
3. Het peer assessment zorgt voor een beter beeld van studenten hun competenties.

Vergelijkbaarheid

4. Het peer assessment helpt studenten op vergelijkbare wijze te beoordelen.
5. Het peer assessment helpt de verwachtingen voor alle studenten vergelijkbaar te maken.
6. Het peer assessment helpt de beoordeling voor iedereen gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Door het peer assessment maakt het minder uit wie de beoordelaars zijn.
8. Door het peer assessment komen beoordelaars sneller tot dezelfde conclusie.
9. Door het peer assessment vellen beoordelaars gemakkelijker hetzelfde oordeel.

Transparantie

10. Door het peer assessment wordt het duidelijker wat er wordt beoordeeld.
11. Door het peer assessment wordt het duidelijker op welke manier er wordt beoordeeld.
12. Door het peer assessment wordt het duidelijker hoe een oordeel tot stand komt.

Acceptatie

13. Het peer assessment maakt wat er beoordeeld wordt voor mij meer aanvaardbaar.
14. Het peer assessment maakt de manier waarop er beoordeeld wordt meer aanvaardbaar.
15. Het peer assessment maakt dat ik me beter kan vinden in hoe een oordeel tot stand.

Authenticiteit

16. Het peer assessment verduidelijkt de link met de professionele context.
17. Het peer assessment gebruikt extra situaties die ook in professionele context voorkomen.
18. Het peer assessment weerspiegelt beter wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. Het peer assessment zorgt voor een eerlijkere manier van beoordelen.
20. Peer assessment maakt het makkelijker om een gefundeerd en eerlijk oordeel te vellen.
21. Het peer assessment geeft een eerlijker beeld van studenten.

Complexiteit

22. Peer assessment helpt rekening te houden met de redeneringen die studenten maken.
23. Het peer assessment helpt het cognitief vermogen van studenten te weerspiegelen.
24. Het peer assessment helpt het denkproces van studenten tijdens hun stage blootleggen.

Betekenisvolheid

25. Het peer assessment biedt een bijkomend leermoment.
26. Het peer assessment biedt studenten een extra leerkans.
27. Het peer assessment geeft studenten een extra kans om dingen bij te leren.

Tijd en kosten

28. De voordelen van het peer assessment wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen loont het peer assessment de moeite.
30. Het is een goed idee om te investeren in het peer assessment.

Consequenties: Selectie

31. Het peer assessment vergemakkelijkt het identificeren van goed presterende studenten.
32. Het peer assessment vergemakkelijkt het identificeren van slecht presterende studenten.
33. Het peer assessment verhoogt de vergelijkbaarheid van studenten hun functioneren.

Consequenties: Leereffect

34. Het peer assessment geeft studenten rijkere feedback over hun prestatie.
35. Het peer assessment zet studenten extra aan na te denken over hun eigen ontwikkeling.
36. Het peer assessment stimuleert het zelfstandig ontwikkelvermogen van studenten extra.

Consequenties: Motivatie

37. Het peer assessment heeft mij zin gegeven om verder te gaan in mijn vakgebied.
38. Het peer assessment heeft mijn interesse in mijn vakgebied vergroot.
39. Het peer assessment heeft mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

40. Het peer assessment verschaft bijkomende info over het gemiddeld niveau van studenten.
41. Het peer assessment verschaft rijkere informatie over de kwaliteit van de stage.
42. Het peer assessment geeft extra info die kan gebruikt worden om de stage te verbeteren.

10.6 Studie 3 (Studenten): Kwaliteitscriteria voor de specifieke interventies

In Studie 3 ging het niet om één maar om twee interventies. Deze appendix bevat de vragenlijst inzake kwaliteitscriteria voor de twee specifieke interventies zoals gebruikt bij studenten. De items werden afgestemd op de specifieke context en voorafgegaan door een aan de studie aangepaste introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

Interventie 1

Dit jaar werd voor de stage ook ***een nieuw beoordelingsformulier*** ingevoerd. In het kader van dit onderzoek zijn we eveneens geïnteresseerd in hoe je dit hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. Het nieuwe formulier sluit aan bij de competenties die studenten moeten ontwikkelen.
2. Het nieuwe formulier sluit aan bij de competenties die de opleiding vereist.
3. Het nieuwe formulier geeft een goed beeld van studenten hun competenties.

Vergelijkbaarheid

4. Het nieuwe formulier helpt studenten op vergelijkbare wijze te beoordelen.
5. Het nieuwe formulier helpt de verwachtingen voor alle studenten vergelijkbaar te maken.
6. Het nieuwe formulier helpt de beoordeling voor iedereen gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Door het nieuwe formulier maakt het niet veel uit wie de beoordelaar is.
8. Door dit formulier komen verschillende beoordelaars al snel tot dezelfde conclusie.
9. Door het nieuwe formulier zouden andere beoordelaars hetzelfde oordeel vellen.

Transparantie

10. Door het nieuwe formulier is het duidelijk wat er wordt beoordeeld.
11. Door het nieuwe formulier is het duidelijk op welke manier er wordt beoordeeld.
12. Door het nieuwe formulier is het duidelijk hoe een oordeel tot stand komt.

Acceptatie

13. Wat er beoordeeld wordt, is voor mij aanvaardbaar met het nieuwe formulier.
14. De manier waarop beoordeeld wordt, is voor mij aanvaardbaar met het nieuwe formulier.
15. Ik kan me vinden in de wijze waarop een oordeel tot stand komt met het nieuwe formulier.

Authenticiteit

16. Het nieuwe formulier heeft een duidelijke link met de professionele context.
17. Het nieuwe formulier gebruikt situaties die ook in een professionele context voorkomen.
18. Het nieuwe formulier weerspiegelt wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. Het nieuwe formulier zorgt voor een eerlijke manier van beoordelen.
20. Het nieuwe formulier laat toe een gefundeerd en eerlijk oordeel te vellen.
21. Het nieuwe formulier geeft een eerlijk beeld van studenten.

Complexiteit

22. Het nieuwe formulier houdt rekening met de redeneringen die studenten maken.
23. Het nieuwe formulier weerspiegelt het cognitief vermogen van studenten.
24. Het nieuwe formulier legt het denkproces van studenten tijdens hun stage bloot.

Betekenisvolheid

25. Het nieuwe formulier biedt ook een leermoment.
26. Het nieuwe formulier biedt studenten leerkansen.
27. Het nieuwe formulier geeft studenten de kans om dingen bij te leren.

Tijd en kosten

28. De voordelen van het nieuwe formulier wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen loont het nieuwe formulier de moeite.
30. Het is een goed idee om te investeren in het nieuwe formulier.

Consequenties: Selectie

31. Het nieuwe formulier laat toe goed presterende studenten te identificeren.
32. Het nieuwe formulier laat toe slecht presterende studenten te identificeren.
33. Het nieuwe formulier laat het functioneren van studenten te vergelijken.

Consequenties: Leereffect

34. Het nieuwe formulier geeft studenten rijke feedback over hun prestatie.
35. Het nieuwe formulier zet studenten aan na te denken over hun eigen ontwikkeling.
36. Het nieuwe formulier stimuleert het zelfstandig ontwikkelvermogen van studenten.

Consequenties: Motivatie

37. Het nieuwe formulier heeft mij zin gegeven om verder te gaan in mijn vakgebied.
38. Het nieuwe formulier heeft mijn interesse in mijn vakgebied vergroot.
39. Het nieuwe formulier heeft mij nieuwe energie gegeven.

Interventie 2

Dit jaar werd er in het kader van een onderzoek een nieuw onderdeel toegevoegd aan de stage. In het kader van dit onderzoek zijn we geïnteresseerd in hoe je dit hebt ervaren. Specifiek gaat het hier om het werken met ***leergroepen*** en leergroepgesprekken ter bevordering van de ontwikkeling van soft skills. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. De leergroepen geven zicht op de soft skills die studenten moeten ontwikkelen.

2. De leergroepen geven zicht op de soft skills die de opleiding vereist.
3. De leergroepen zorgen voor een goed beeld van studenten hun soft skills.

Vergelijkbaarheid

4. De leergroepen helpen studenten op vergelijkbare wijze te beoordelen.
5. De leergroepen helpen de verwachtingen voor alle studenten vergelijkbaar te maken.
6. De leergroepen helpen de beoordeling voor alle studenten gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Voor de leergroepen maakt het niet uit wie de docent is.
8. Bij de leergroepen zou een andere docent al snel tot dezelfde conclusie komen.
9. Bij de leergroepen zou een andere docent gemakkelijk hetzelfde oordeel vellen.

Transparantie

10. Door de leergroepen wordt het duidelijker wat er wordt beoordeeld.
11. Door de leergroepen wordt het duidelijker op welke manier er wordt beoordeeld.
12. Door de leergroepen wordt het duidelijker hoe een oordeel tot stand komt.

Acceptatie

13. De leergroepen maken wat er beoordeeld wordt voor mij meer aanvaardbaar.
14. De leergroepen maken de manier waarop er beoordeeld wordt meer aanvaardbaar.
15. De leergroepen maken dat ik me beter kan vinden in hoe een oordeel tot stand.

Authenticiteit

16. De leergroepen verduidelijken de link met de professionele context.
17. De leergroepen gebruiken situaties die ook in professionele context voorkomen.
18. De leergroepen weerspiegelen wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. De leergroepen zorgen voor een eerlijke manier van beoordelen.
20. De leergroepen laten toe een gefundeerd en eerlijk oordeel te vellen.
21. De leergroepen geven een eerlijk beeld van studenten.

Complexiteit

22. De leergroepen helpen rekening te houden met de redeneringen die studenten maken.
23. De leergroepen helpen het cognitief vermogen van studenten te weerspiegelen.
24. De leergroepen helpen het denkproces van studenten tijdens hun stage blootleggen.

Betekenisvolheid

25. De leergroepen zijn ook leermomenten.
26. De leergroepen bieden studenten leerkansen.
27. De leergroepen geven studenten kansen om dingen bij te leren.

Tijd en kosten

28. De voordelen van de leergroepen wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen lonen de leergroepen de moeite.
30. Het is een goed idee om te investeren in de leergroepen.

Consequenties: Selectie

31. De leergroepen laten toe studenten met goede soft skills te identificeren.
32. De leergroepen laten toe studenten met slechte soft skills te identificeren.
33. De leergroepen laten toe de soft skills van studenten te vergelijken.

Consequenties: Leereffect

34. De leergroepen geven studenten rijke feedback over hun prestatie.
35. De leergroepen zetten studenten aan na te denken over hun eigen ontwikkeling.
36. De leergroepen stimuleren het zelfstandig ontwikkelvermogen van studenten.

Consequenties: Motivatie

37. De leergroepen hebben mij zin gegeven om verder te gaan in mijn vakgebied.
38. De leergroepen hebben mijn interesse in mijn vakgebied vergroot.
39. De leergroepen hebben mij nieuwe energie gegeven.

10.7 Studie 3 (Begeleiders): Kwaliteitscriteria voor de specifieke interventies

In Studie 3 ging het niet om één maar om twee interventies. Deze appendix bevat de vragenlijst inzake kwaliteitscriteria voor de twee specifieke interventies zoals gebruikt bij begeleiders. De items werden afgestemd op de specifieke context en voorafgegaan door een aan de studie aangepaste introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

Interventie 1

Dit jaar werd voor de stage ook ***een nieuw beoordelingsformulier*** ingevoerd. In het kader van dit onderzoek zijn we eveneens geïnteresseerd in hoe je dit hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. Dit formulier geeft meer zicht op competenties die studenten moeten ontwikkelen.
2. Het nieuwe formulier verbetert het zicht op de competenties die de opleiding vereist.
3. Het nieuwe formulier zorgt voor een beter beeld van studenten hun competenties.

Vergelijkbaarheid

4. Het nieuwe formulier helpt studenten op vergelijkbare wijze te beoordelen.
5. Het nieuwe formulier helpt de verwachtingen voor alle studenten vergelijkbaar te maken.
6. Het nieuwe formulier helpt de beoordeling voor iedereen gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Door het nieuwe formulier maakt het minder uit wie de beoordelaars zijn.
8. Door dit formulier komen verschillende beoordelaars sneller tot dezelfde conclusie.
9. Door het nieuwe formulier vellen beoordelaars gemakkelijker hetzelfde oordeel.

Transparantie

10. Door het nieuwe formulier wordt het duidelijker wat er wordt beoordeeld.
11. Door het nieuwe formulier wordt het duidelijker op welke manier er wordt beoordeeld.
12. Door het nieuwe formulier wordt het duidelijker hoe een oordeel tot stand komt.

Acceptatie

13. Het nieuwe formulier maakt wat er beoordeeld wordt voor mij meer aanvaardbaar.
14. Het nieuwe formulier maakt de manier waarop er beoordeeld wordt meer aanvaardbaar.
15. Het nieuwe formulier maakt dat ik me beter kan vinden in hoe een oordeel tot stand komt.

Authenticiteit

16. Het nieuwe formulier verduidelijkt de link met de professionele context.
17. Het nieuwe formulier gebruikt extra situaties die ook in professionele context voorkomen.
18. Het nieuwe formulier weerspiegelt beter wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. Het nieuwe formulier zorgt voor een eerlijkere manier van beoordelen.
20. Het nieuwe formulier maakt het makkelijker een gefundeerd en eerlijk oordeel te vellen.
21. Het nieuwe formulier geeft een eerlijker beeld van studenten.

Complexiteit

22. Dit formulier helpt rekening te houden met de redeneringen die studenten maken.
23. Het nieuwe formulier helpt het cognitief vermogen van studenten te weerspiegelen.
24. Het nieuwe formulier helpt het denkproces van studenten tijdens hun stage blootleggen.

Betekenisvolheid

25. Het nieuwe formulier biedt een bijkomend leermoment.
26. Het nieuwe formulier biedt studenten extra leerkansen.
27. Het nieuwe formulier geeft studenten een extra kans om dingen bij te leren.

Tijd en kosten

28. De voordelen van het nieuwe formulier wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen loont het nieuwe formulier de moeite.
30. Het is een goed idee om te investeren in het nieuwe formulier.

Consequenties: Selectie

31. Het nieuwe formulier maakt het makkelijker goed presterende studenten te identificeren.
32. Het nieuwe formulier maakt het makkelijker slecht presterende studenten te identificeren.
33. Dit formulier zorgt dat het functioneren van studenten beter vergeleken kan worden.

Consequenties: Leereffect

34. Het nieuwe formulier geeft studenten rijkere feedback over hun prestatie.
35. Het nieuwe formulier zet studenten extra aan na te denken over hun eigen ontwikkeling.
36. Dit formulier stimuleert het zelfstandig ontwikkelvermogen van studenten nog meer.

Consequenties: Motivatie

37. Het nieuwe formulier heeft mij zin gegeven om verder te gaan in mijn vakgebied.
38. Het nieuwe formulier heeft mijn interesse in mijn vakgebied vergroot.
39. Het nieuwe formulier heeft mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

40. Het nieuwe formulier verschaft bijkomende info over het gemiddeld niveau van studenten.
41. Het nieuwe formulier verschaft rijkere informatie over de kwaliteit van de stage.
42. Het nieuwe formulier geeft extra info die kan gebruikt worden om de stage te verbeteren.

Interventie 2

Dit jaar werd er in het kader van een onderzoek een nieuw onderdeel toegevoegd aan de stage. In het kader van dit onderzoek zijn we geïnteresseerd in hoe je dit hebt ervaren. Specifiek gaat het hier om het werken met ***leergroepen*** en leergroepgesprekken ter bevordering van de ontwikkeling van soft skills. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan -3. Ben je het helemaal eens met een

stelling, omcirkel dan +3. Gebruik de tussenliggende cijfers om een meer genuanceerd antwoord te geven.

Geschiktheid

1. De leergroepen geven meer zicht op de soft skills die studenten moeten ontwikkelen.
2. De leergroepen verbeteren het zicht op de soft skills die de opleiding vereist.
3. De leergroepen zorgen voor een beter beeld van studenten hun soft skills.

Vergelijkbaarheid

4. De leergroepen helpen studenten op vergelijkbare wijze te beoordelen.
5. De leergroepen helpen de verwachtingen voor alle studenten vergelijkbaar te maken.
6. De leergroepen helpen de beoordeling voor alle studenten gelijkaardig te laten verlopen.

Herhaalbaarheid

7. Voor de leergroepen maakt het niet uit wie de docent is.
8. Bij de leergroepen zou een andere docent al snel tot dezelfde conclusie komen.
9. Bij de leergroepen zou een andere docent gemakkelijk hetzelfde oordeel vellen.

Transparantie

10. Door de leergroepen wordt het duidelijker wat er wordt beoordeeld.
11. Door de leergroepen wordt het duidelijker op welke manier er wordt beoordeeld.
12. Door de leergroepen wordt het duidelijker hoe een oordeel tot stand komt.

Acceptatie

13. De leergroepen maken wat er beoordeeld wordt voor mij meer aanvaardbaar.
14. De leergroepen maken de manier waarop er beoordeeld wordt meer aanvaardbaar.
15. De leergroepen maken dat ik me beter kan vinden in hoe een oordeel tot stand.

Authenticiteit

16. De leergroepen verduidelijken de link met de professionele context.
17. De leergroepen gebruiken extra situaties die ook in professionele context voorkomen.
18. De leergroepen weerspiegelen beter wat professioneel van studenten verwacht wordt.

Eerlijkheid

19. De leergroepen zorgen voor een eerlijkere manier van beoordelen.
20. De leergroepen maken het makkelijker om een gefundeerd en eerlijk oordeel te vellen.
21. De leergroepen geven een eerlijker beeld van studenten.

Complexiteit

22. De leergroepen helpen rekening te houden met de redeneringen die studenten maken.
23. De leergroepen helpen het cognitief vermogen van studenten te weerspiegelen.
24. De leergroepen helpen het denkproces van studenten tijdens hun stage blootleggen.

Betekenisvolheid

25. De leergroepen bieden bijkomende leermomenten.
26. De leergroepen bieden studenten extra leerkansen.
27. De leergroepen geven studenten extra kansen om dingen bij te leren.

Tijd en kosten

28. De voordelen van de leergroepen wegen op tegen de nadelen.
29. Ondanks de gevraagde inspanningen lonen de leergroepen de moeite.
30. Het is een goed idee om te investeren in de leergroepen.

Consequenties: Selectie

31. De leergroepen maken het identificeren van studenten met goede soft skills makkelijker.
32. De leergroepen maken het identificeren van studenten met slechte soft skills makkelijker.
33. De leergroepen zorgen dat de soft skills van studenten beter vergeleken kunnen worden.

Consequenties: Leereffect

34. De leergroepen geven studenten rijkere feedback over hun prestatie.
35. De leergroepen zetten studenten extra aan na te denken over hun eigen ontwikkeling.
36. De leergroepen stimuleren het zelfstandig ontwikkelvermogen van studenten nog meer.

Consequenties: Motivatie

37. De leergroepen hebben mij zin gegeven om verder te gaan in mijn vakgebied.
38. De leergroepen hebben mijn interesse in mijn vakgebied vergroot.
39. De leergroepen hebben mij nieuwe energie gegeven.

Consequenties: Onderwijseffect

40. De leergroepen verschaffen bijkomende info over het gemiddeld niveau van studenten.
41. De leergroepen verschaffen rijkere informatie over de kwaliteit van de stage.
42. De leergroepen geven extra info die kan gebruikt worden om de stage te verbeteren.

10.8 De competentiegroeimeter voor stage voor docenten

Deze appendix bevat de competentiegroeimeter voor stage voor docenten zoals gebruikt in deze studies. De items werden voorafgegaan door een aan de studie aangepaste variant van de introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen.

In welke mate waren onderstaande zaken volgens jou een doel van de stage 2017-2018 voor *{opleidingsonderdeel}*. Geef voor elke stelling aan in welke mate je ermee akkoord gaat dat dit één van de doelen was. Ben je het helemaal niet eens met een stelling, omcirkel dan 1. Ben je het helemaal eens met een stelling, omcirkel dan 5. De tussenliggende cijfers kun je gebruiken om een meer genuanceerd antwoord te geven. Een doel van deze stage was studenten…

1. Voorbereiden op de beroepspraktijk en theorie leren toepassen in de praktijk.
2. Leren hun werk te organiseren en zelfstandig problemen aan te pakken.
3. Beter zicht laten krijgen op de eigen sterke en zwakke punten.
4. Leren doelen bepalen, risico's inschatten en planmatig doelen bereiken.
5. Leren verbeterkansen in de omgeving zien en initiatief nemen.
6. Leren leidinggeven en een (doel)groep te activeren.
7. Leren openstaan voor nieuwe ideeën en leren om te durven experimenteren.
8. Leren bijdragen aan en zich houden aan afspraken binnen een team.
9. Leren feedback geven en zichzelf uitdrukken in een gesprek of discussie.
10. Leren openstaan voor feedback van anderen en daar actief naar leren vragen.
11. Leren luisteren naar de klant en leren professioneel reageren op diens vragen.
12. Gestructureerde presentaties leren brengen, aangepast aan de doelgroep.

10.9 De competentiegroeimeter voor stage voor studenten

Deze appendix bevat de competentiegroeimeter voor stage voor studenten zoals gebruikt in deze studies. De items werden voorafgegaan door een aan de studie aangepaste variant van de introductietekst en worden gepresenteerd in de volgorde waarin ze werden afgenomen. Voor de helderheid zijn ze gegroepeerd per concept.

Aan de hand van onderstaande stellingen willen wij polsen naar hoe je de stage voor *{opleidingsonderdeel}* hebt ervaren. Geef voor elke stelling aan in welke mate je ermee akkoord gaat. Ben je het helemaal niet eens met een stelling, omcirkel dan 1. Ben je het helemaal eens met een stelling, omcirkel dan 5. De tussenliggende cijfers kun je gebruiken om een meer genuanceerd antwoord te geven. Dankzij mijn stage voor *{opleidingsonderdeel}*...

Praktijkgericht

1. Heb ik vaardigheden verworven die voor mijn later beroep relevant zijn.
2. Zie ik duidelijker de linken tussen de behandelde leerinhouden en de beroepspraktijk.
3. Heb ik een duidelijkere voorstelling van de toekomstige beroepsmogelijkheden.
4. Heb ik kennis en methoden leren toepassen op de praktijk.
5. Weet ik beter welk gedrag een werkgever verwacht van mij als professional.

Zelfmanagement

6. Kan ik effectiever naar informatie zoeken om een specifiek probleem aan te pakken.
7. Kan ik de juiste hulpmiddelen selecteren om een specifiek probleem aan te pakken.
8. Kan ik mijn werk beter zelfstandig organiseren.
9. Heb ik mijn manier van werken verbeterd.
10. Kan ik gemakkelijker mijn omgeving efficiënt inzetten om mijn doel te bereiken.

Zelfreflectie

11. Zie ik sneller kansen om bij te leren.
12. Heb ik een beter zicht op mijn eigen werkpunten.
13. Weet ik op welke vlakken ik nog kan groeien als professional.
14. Heb ik een beter zicht op mijn sterke punten.
15. Kan ik vlotter aangeven welke fouten ik maak en waarom ik ze maak.

Doelgericht werken

16. Kan ik flexibeler omgaan met tegenslagen.
17. Vind ik het nu gemakkelijker om voor mezelf doelen te bepalen.
18. Kan ik nu beter prioriteiten bepalen.
19. Weet ik beter welke factoren invloed kunnen hebben op mijn planning.
20. Kan ik nu beter risico's inschatten.
21. Kan ik beter bepalen welke stappen ik moet zetten om mijn doelen te bereiken.
22. Weet ik beter wanneer ik een bepaald resultaat behaald heb.

Kansen zien en nemen

23. Vind ik het eenvoudiger om initiatief te nemen.
24. Zie ik sneller kansen voor verbetering van mijn werkomgeving.

25. Vind ik het eenvoudiger om een probleem als een kans te zien.

Leidinggeven en activeren

26. Vind ik het eenvoudiger om bij een tegenslag anderen gemotiveerd te houden.
27. Vind ik het leuker om leiding te geven aan anderen.
28. Vind ik het eenvoudiger om leiding te geven aan anderen.
29. Ben ik nu beter in het enthousiasmeren van anderen voor een doel of project.
30. Vind ik het eenvoudiger om mijn doelgroep te activeren.

Vernieuwen

31. Sta ik nu meer open voor nieuwe ideeën.
32. Zal ik nu sneller nieuwe dingen proberen.
33. Vind ik het eenvoudiger om 'out of the box' te denken.

Samenwerken

34. Kan ik mij nu beter inzetten voor een constructieve werksfeer binnen een team.
35. Kan ik me identificeren met het bereikte resultaat van het team.
36. Heb ik geleerd om mij te houden aan teamafspraken.
37. Kan ik mijn verantwoordelijkheid beter opnemen voor de resultaten van mijn team.
38. Kan ik binnen een team beter de juiste taken aan de juiste personen toewijzen.

Feedback geven en reageren

39. Kan ik nu beter constructieve feedback geven aan anderen.
40. Vind ik het gemakkelijker om constructieve feedback te formuleren.
41. Slaag ik er beter in om aan anderen iets begrijpelijk uit te leggen.
42. Lukt het me beter om in discussies een helder standpunt te formuleren.
43. Kan ik discussies beter leiden.

Feedback ontvangen en vragen

44. Sta ik meer open voor feedback van anderen op mijn prestaties.
45. Vraag ik nu sneller om feedback op het door mij geleverde werk.
46. Vraag ik nu gemakkelijker om uitleg wanneer ik iets niet begrepen heb.

Klantgericht werken

47. Kan ik beter luisteren naar de vraag van mijn gesprekspartner(s).
48. Vind ik het gemakkelijker om in gesprek te gaan met anderen.
49. Begrijp ik sneller wat mijn gesprekspartner nodig heeft van mij.
50. Kan ik professioneler reageren op negatieve reacties van mijn doelgroep/klanten.
51. Weet ik beter hoe ik moet omgaan met mijn doelgroep/klanten.

Voor een groep spreken

52. Voel ik me zelfzekerder voor een publiek.
53. Kan ik de tijdsduur van mijn presentaties/lessen beter inschatten.
54. Kan ik in het algemeen sterker gestructureerd presenteren/les geven.
55. Kan ik de inhoud van mijn presentatie/les beter afstemmen op mijn publiek.
56. Kan ik mijn taalgebruik beter afstemmen op mijn doelgroep.