

**Assessing the assessment: het praktisch examen verpleegkundige
vaardigheden onder de loep**

Een voorstel tot optimalisatie van de huidige evaluatievorm

Productverslag Interdisciplinair Project – Opleidings- en
onderwijswetenschappen

Projectgroep: Joachim Bonny, Katrien
Cornelis, Casper Fleerackers ,
Kimberley Gregoor, Danny Majzik,
Anne Mariën en Sofie van Ostaeyen.

Procesbegeleider: Marije Lesterhuis
Externe opdrachtgever: Thomas More
Mechelen, Carl Vanspauwen en Lien
Proost

Dankwoord

Dit Interdisciplinair Project kwam tot stand dankzij de inzet en het doorzettingsvermogen van zeven teamleden. Door ieders teamlid zijn inzet en betrokkenheid hebben we een succesvol leertraject afgelegd. Aan het einde van dit traject willen we graag onze dankbaarheid betuigen aan de mensen die ons steeds hebben bijgestaan en betrokken waren bij de uitwerking van dit project.

Aan Thomas More Mechelen, zonder de steun en medewerking van onze externe opdrachtgever was dit project niet mogelijk. We hopen dat de resultaten voor jullie even waardevol zijn als het (leer)proces voor ons was.

Aan Lien Proost en Carl Vanspauwen, voor hun openheid en bereidwillige medewerking gedurende het hele project.

Aan de docenten die deelnamen aan het focusgroepgesprek en de studenten die de tijd namen om de survey in te vullen. De boeiende dialoog en insteek die jullie gaven waren noodzakelijk voor de uitwerking van dit project.

Aan onze tweede lezer, Roos Van Gasse, voor de nuttige feedback en kritische blik op het tussentijds rapport.

Aan onze procesbegeleider, Marije Lesterhuis, voor de waardevolle feedback, haar enthousiasme, de uren van persoonlijke begeleiding en haar interesse in het onderwerp.

Anne, Casper, Danny, Joachim, Katrien, Kimberley en Sofie

Figuren

Figuur 1: visie van de Thomas More hogescholen op onderwijs en studenten

Figuur 2: Miller's pyramid of assessment

Figuur 3: wheel of competence

Figuur 4: constructive alignment, Biggs

Tabellen

Tabel 1: aantal geslaagde studenten voor het theoretisch en praktisch examen verpleegkundige vaardigheden, januari 2019.

Tabel 2: aantal deelnemers voor het theoretisch en praktisch examen verpleegkundige vaardigheden, januari 2019.

Tabel 3: extremen van de psychometrische en edumetrische benaderingen

Tabel 4: correlatiematrix voor de variabelen transparantie, feedback en stress

Tabel 5: beschrijving uitkomsten studenten verpleegkunde

Inhoudsopgave

| | |
|---|----|
| 1 Inleiding: situering opdrachtgever, probleemstelling, onderzoeksvragen en doelstellingen | 6 |
| 1.1 Inleiding | 6 |
| 1.2 Situering opdrachtgever | 7 |
| 1.3 Probleemstelling, onderzoeksvragen en doelstellingen | 9 |
| 2 Theoretisch kader | 12 |
| 2.1 Competenties | 12 |
| 2.2 Assessment van competenties | 13 |
| 2.2.1 Competence assessment programs | 16 |
| 2.3 Kwaliteit van assessments | 18 |
| 2.3.1 Wheel of competence | 20 |
| 2.3.2 Focus IP | 22 |
| 2.3.2.1 Vergelijkbaarheid | 23 |
| 2.3.2.2 Reproduceerbaarheid van beslissingen | 24 |
| 2.3.2.3 Transparantie | 25 |
| 2.3.2.4 Betekenisvolheid | 26 |
| 3 Opzet - methodologie van het project | 29 |
| 3.1 Onderzoeksmethode | 29 |
| 3.2 Methodologie survey | 31 |
| 3.2.1 Respondenten | 31 |
| 3.2.2 Dataverzameling | 32 |
| 3.2.3 Materiaal | 32 |
| 3.2.4 Data-analyse | 33 |
| 3.2.4.1 Samenstelling schalen | 34 |
| 3.2.4.2 Beantwoorden onderzoeksvragen | 35 |
| 3.3 Methodologie observatie praktisch examen | 36 |
| 3.4 Methodologie focusgroep | 36 |
| 3.4.1 Respondenten | 36 |
| 3.4.2 Dataverzameling | 37 |
| 3.4.3 Data-analyse | 37 |
| 3.5 Methodologie documentanalyse | 38 |
| 4 Producten - resultaten | 39 |
| 4.1 Rapportage onderzoeksresultaten survey | 39 |
| 4.1.1 Samenstelling schalen | 39 |
| 4.1.2 Beantwoorden onderzoeksvragen | 41 |
| 4.1.2.1 Beschrijven uitkomsten | 41 |
| 4.1.2.2 Invloed van transparantie op stress | 42 |
| 4.1.2.3 Invloed van feedback op stress | 43 |
| 4.2 Rapportage onderzoeksresultaten observatie praktisch examen | 43 |
| 4.3 Rapportage onderzoeksresultaten documentanalyse | 46 |
| 4.3.1 Generieke beoordelingsschema | 46 |
| 4.3.2 Toetsprocedures | 46 |
| 4.3.3 Videofragmenten | 47 |

| | |
|--|-----------|
| 4.4 Rapportage onderzoeksresultaten focusgroep | 47 |
| 5 Conclusies en discussie | 52 |
| 5.1 Conclusie | 52 |
| 5.2 Discussie | 58 |
| 5.2.1 Theoretisch kader | 58 |
| 5.2.2 Survey | 58 |
| 5.2.3 Focusgroep | 59 |
| 6 Aanbevelingen | 61 |
| 6.1 Vergelijkbaarheid | 61 |
| 6.2 Reproduceerbaarheid van beslissingen | 62 |
| 6.3 Transparantie | 63 |
| 6.4 Betekenisvolheid | 66 |
| 6.5 Vijf punten plan | 70 |
| 7 Literatuur | 72 |
| 8 Bijlagen | 82 |
| 8.1 Bijlage 1: vragenlijst studenten | 82 |
| 8.2 Bijlage 2: samenvatting focusgroep | 90 |
| 8.3 Bijlage 3: geteste modellen met bijhorende fit indices (EFA) | 98 |
| 8.4 Bijlage 4: generiek beoordelingsschema Thomas More | 100 |
| 8.5 Bijlage 5: voorbeeld van een toetsprocedure Thomas More | 102 |
| 8.6 Bijlage 6: rubric voor rubrics | 105 |

Deel 1: Beschrijving van het project

1 Inleiding: situering opdrachtgever, probleemstelling, onderzoeksvragen en doelstellingen

1.1 Inleiding

Op 11 september 2015 verspreidde het kabinet van de Vlaamse minister van Onderwijs, Hilde Crevits, een persbericht in verband met de hervorming van de bacheloropleiding verpleegkunde. Crevits (2015) verwoordde het als volgt:

De zorgsector wordt steeds complexer en de verwachtingen van verpleegkundigen liggen hoog. Onder impuls van de Europese regelgeving grijpen we de kans om de bacheloropleidingen verpleegkunde grondig te vernieuwen. Studenten zullen vier jaar studeren, maar zullen een bredere opleiding krijgen, minder vervolgoopleidingen nodig hebben, meer ervaring opdoen en nog steeds na drie jaar meedraaien op de werkvloer, binnen het nieuwe concept van de contractstage. Verpleegkundigen genieten terecht een groot vertrouwen van de burgers. Dat willen we in de toekomst versterken.

De hervorming kwam er naar aanleiding van de nieuwe Europese richtlijnen over gereguleerde beroepen die aanstuurden op een modernisering van de opleiding. Deze brachten zowel kwalitatieve (te behalen competenties) als kwantitatieve vereisten (duur van de opleiding) met zich mee. Bovendien was het ook de bedoeling om het inhoudelijke aspect van de opleiding te versterken en zo innovatieve vormen van praktijkonderwijs (vb. simulaties...) te benutten. Niet alleen de stages werden dus vernieuwd, ook het theoretisch en klinisch onderwijs werden in de hervorming betrokken.

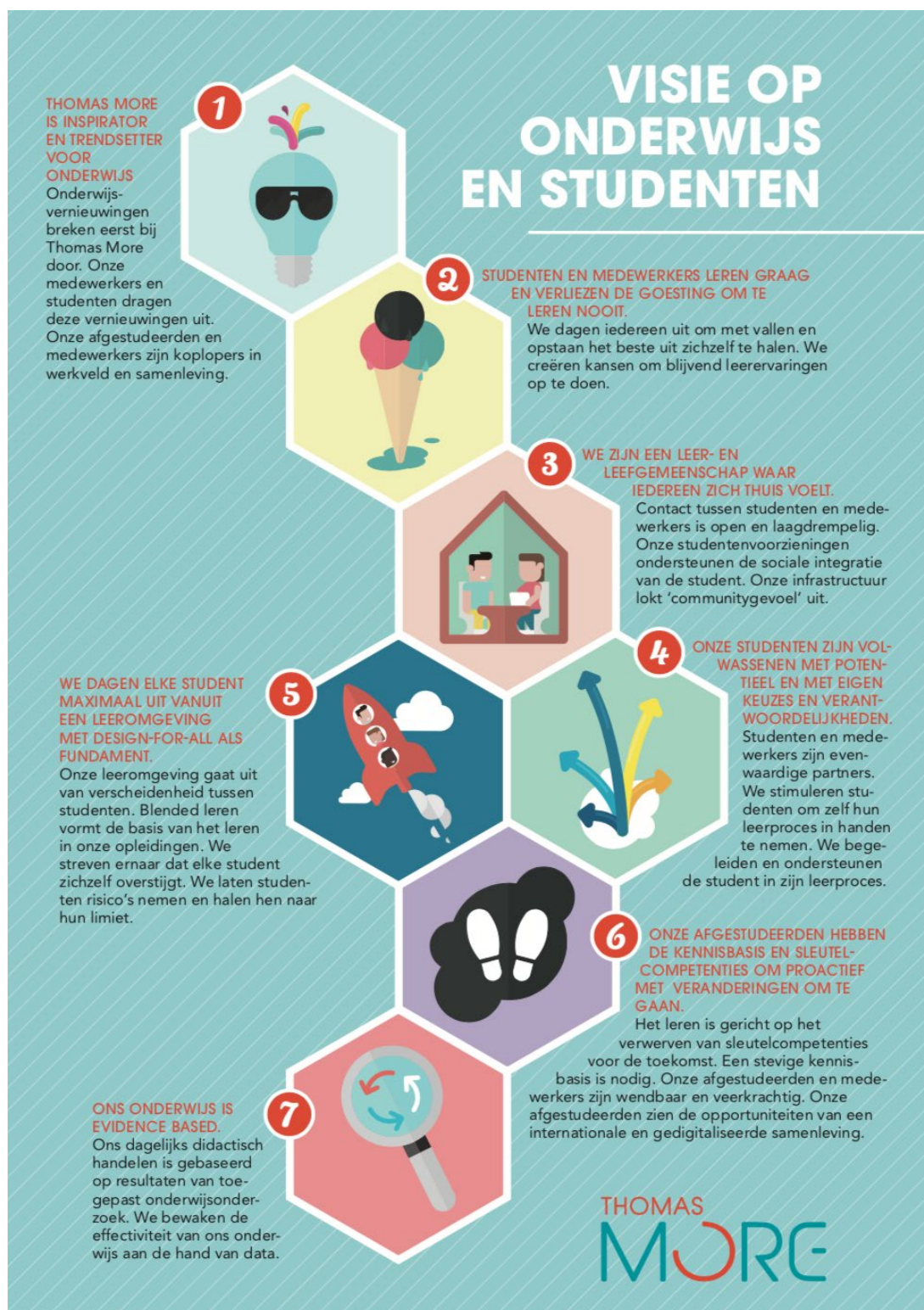
Thomas More is één van de hogescholen die in het academiejaar 2015-2016 zijn opleiding verpleegkunde hervormde. In hun streven naar een kwaliteitsvolle opleiding die studenten opleidt tot gekwalificeerde verpleegkundigen ervaren zij ook de toegenomen druk en complexiteit binnen de zorgsector. De kwaliteitseisen liggen hoger en dat weerspiegelt zich in de resultaten van de studenten voor bepaalde opleidingsonderdelen. Deze tendens komt bij het vak 'verpleegkundige

vaardigheden' sterk naar voor, dat georganiseerd wordt in het eerste jaar van deze professionele bachelor.

Dit rapport beschrijft het Interdisciplinair Project (IP) dat in samenwerking met de hogeschool Thomas More Mechelen in het academiejaar 2018-2019 werd uitgevoerd. Meer specifiek wordt in dit project het praktijkexamen verpleegkundige vaardigheden in het eerste jaar van de professionele bachelor verpleegkunde onder de loep genomen. Door inzicht te bieden in de belangrijkste factoren die van invloed zijn op de scores van de studenten en door de docenten mogelijkheden aan te bieden voor de optimalisatie van hun competentie-assessment, hoopt dit IP een vertrekpunt te zijn voor het verhogen van de slaagkansen van de studenten.

1.2 Situering opdrachtgever

Thomas More is een unie van hogescholen binnen de Associatie KU Leuven. Thomas More biedt studenten innovatief professioneel hoger onderwijs aan in nauwe samenwerking met het werkveld. De visie van Thomas More op hun onderwijs en studenten wordt weergegeven in een infographic, zie figuur 1. De Thomas More Hogescholen zijn gevestigd in Antwerpen, Duffel, Geel, Herentals, Lier, Mechelen, Sint Katelijne Waver, Sint Niklaas, Turnhout en Vorselaar. De professionele bachelor verpleegkunde wordt aangeboden in de vestigingen in Lier, Mechelen en Turnhout.



Figuur 1: visie van de Thomas More hogescholen op onderwijs en studenten *Noot*. Herdrukt van "Missie en visie", door Thomas More (2018). Geraadpleegd van https://www.thomasmore.be/sites/www.thomasmore.be/files/onderwijsvisie_thomas_more_infographic.pdf

Dit IP wordt in samenwerking met Thomas More Mechelen, Campus De Vest, tot stand gebracht. De professionele bachelor verpleegkunde bestaat uit 240

studiepunten. De reden hiervoor is dat de gezondheidszorg steeds complexer wordt. De technologie en wetenschap staan niet stil. Ook de patiënt evolueert, de bevolking wordt ouder en de patiënt wordt mondiger. De zorg van de toekomst is innovatief en vraagt daarom om een andere aanpak van de verpleegkunde waarbij kwaliteit, veiligheid en teamwork centraal staan (Thomas More, 2018a). Het opleidingsonderdeel verpleegkundige vaardigheden, waarbij wij ons specifiek gaan richten op de summatieve evaluatie van het praktijkexamen van dit opleidingsonderdeel, bestaat uit zeven studiepunten.

1.3 Probleemstelling, onderzoeksvragen en doelstellingen

De hogeschool Thomas More Mechelen is op zoek naar een optimalisatie van de huidige summatieve evaluatie voor het praktijkexamen van het opleidingsonderdeel 'verpleegkundige vaardigheden'. De doelgroep zijn de studenten uit het eerste jaar van de professionele bachelor verpleegkunde. De docenten van het opleidingsonderdeel verpleegkundige vaardigheden stellen vast dat studenten beter scoren op het theoretisch examen dan op het praktisch examen, ze hebben bovendien het gevoel dat veel studenten afhaken voor het praktisch examen.

De voornaamste problemen die zich voordoen binnen het competentie-assessment voor het vak 'verpleegkundige vaardigheden', volgens de docenten, zijn enerzijds dat de studenten hun slaagkansen laag liggen (tabel 1) en anderzijds dat zij stress ervaren tijdens het afleggen van het examen. We stellen daarnaast vast dat de deelname aan het praktisch examen lager ligt dan bij het theoretisch examen (tabel 2). We trachten inzicht te bieden in de belangrijkste factoren die van invloed zijn op de slaagkansen van de studenten en de stress die gepaard gaat bij het afleggen van het examen.

Tabel 1: aantal geslaagde studenten voor het theoretisch en praktisch examen verpleegkundige vaardigheden, januari 2019

| | Geslaagden | Niet geslaagden | Niet deelnemers | Totaal |
|---------------------------|-------------------|------------------------|------------------------|---------------|
| Theoretisch examen | 60 | 19 | 5 | 84 |
| Praktisch examen | 39 | 25 | 20 | 84 |

Tabel 2: aantal deelnemers voor het theoretisch en praktisch examen verpleegkundige vaardigheden, januari 2019

| | Deelnemers | Niet deelnemers | Totaal |
|---------------------------|-------------------|------------------------|---------------|
| Theoretisch examen | 79 | 5 | 84 |
| Praktisch examen | 64 | 20 | 84 |

In dit Interdisciplinair Project willen we op een wetenschappelijk onderbouwde manier en in samenwerking met de hogeschool Thomas More Mechelen op zoek gaan naar antwoorden op de volgende onderzoeksvragen:

1. a. In hoeverre ervaren studenten het praktisch examen verpleegkundige vaardigheden als transparant?
 - 1.b. In welke mate heeft dit invloed op stress?
2. a. Hoe ervaren studenten feedback in de huidige vorm?
 - 2.b. In welke mate heeft dit invloed op stress?
3. In hoeverre ervaart Thomas More problemen in verband met de betrouwbaarheid en de validiteit, en dus de kwaliteit, van het praktisch examen verpleegkundige vaardigheden?

Met bovenstaande onderzoeksvragen trachten de onderzoekers een analyse te doen van het huidige competentie-assessment met betrekking tot betrouwbaarheid, validiteit, transparantie, feedback en stress. Om te illustreren wat we willen bereiken met dit project kunnen onze onderzoeksvragen vertaald worden in volgende doelstellingen:

- Doelstelling 1: beschrijven wat bepaalt of een assessment een goed assessment is.
- Doelstelling 2: nagaan of het huidige competentie-assessment in Thomas More een goed assessment is aangaande betrouwbaarheid, validiteit, transparantie en feedback, en hoe dit de stress en/of uitval bij studenten kan verklaren.
- Doelstelling 3: aanbevelingen doen om het huidige assessment te verbeteren.

2 Theoretisch kader

Om een antwoord te kunnen bieden op de onderzoeksvragen en om de projectdoelstellingen te kunnen realiseren is het van belang om een duidelijk beeld te schetsen van wat een goed competentie-assessment is. Dit theoretisch kader is grotendeels gebaseerd op een werk van Baartman (2008). Zij deed onderzoek naar competentie-assessment en bracht criteria in kaart die nodig zijn om de kwaliteit van dit soort assessments te garanderen. Zo kwam ze tot een kader, 'the wheel of competence assessment' genoemd, bestaande uit criteria gebaseerd op de klassieke kwaliteitsnormen validiteit en betrouwbaarheid en op nieuwere ideeën rond de kwaliteit van assessment. De criteria kunnen toegepast worden op formatieve en summatieve vormen van assessment. Bovendien werd het kader gevalideerd door experts en leerkrachten (Jonsson, Baartman & Lennung, 2009).

Eerst wordt dieper ingegaan op wat competenties zijn en hoe ze het best geëvalueerd kunnen worden. Daarna wordt beschreven wat criteria zijn om de kwaliteit van assessments te borgen. Hierbij wordt 'the wheel of competence' van Baartman (2008) toegelicht. Vervolgens wordt de selectie van de kwaliteitscriteria die onderzocht werden bij de opleiding verpleegkunde van Thomas More Mechelen in de diepte besproken. Stress kan een gevolg zijn wanneer onvoldoende voldaan wordt aan deze criteria. Dit wordt aansluitend op elk van de door ons geselecteerde criteria van 'the wheel of competence' verder verduidelijkt.

2.1 Competenties

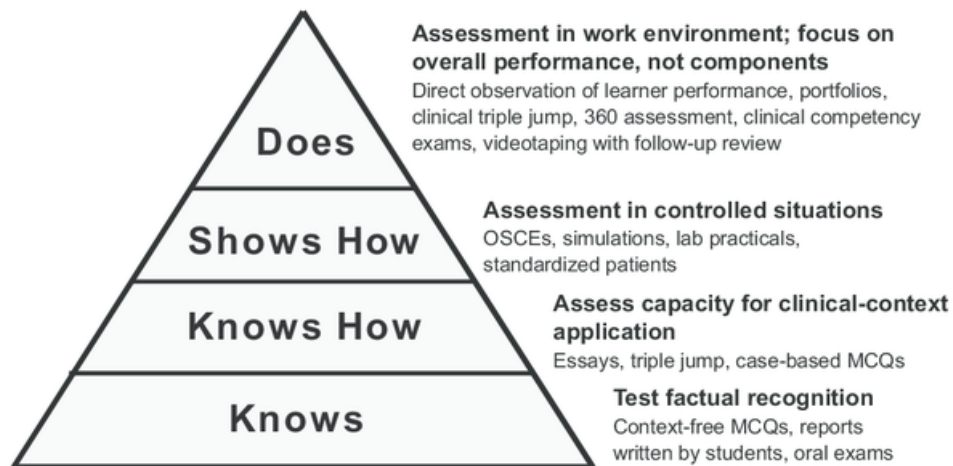
In het verleden was het hoofdobjectief van onderwijs de kennisoverdracht in een bepaald vakdomein. In de huidige samenleving wordt het hebben van een ruime kennisbasis niet meer als voldoende gepercipieerd om goed gewapend in het professioneel leven te staan. In het onderwijs wordt er daarom, naast de kennisoverdracht, ook meer aandacht gegeven aan probleemoplossend werken en het kunnen handelen in authentieke situaties. Dit gebeurt ook in de opleiding verpleegkunde. Bengtsson, Dahlquist en Carlson (2016) veronderstellen immers dat studenten verpleegkunde naast de achterliggende theorie moeten kunnen aantonen hoe ze in verschillende levensechte situaties presteren en handelen. Birenbaum (1996) stelt daarbij dat het om succesvol te zijn in de huidige

samenleving noodzakelijk is om als individu flexibel en autonoom te zijn en de nodige kennis van zaken te hebben. Daarnaast is dit individu in staat om zijn leren zelf te reguleren en samen te werken met andere.

Samenvattend kunnen we stellen dat in moderne organisaties mensen in staat moeten zijn om informatie te analyseren, problemen op te lossen, te communiceren en moeten kunnen reflecteren over hun eigen leerproces. Zo zijn ze ook in staat om zelfstandig kennis te verwerven en hun verworven kennis te gebruiken in onvoorziene probleemsituaties (Dochy, Segers, & Sluijsmans, 1999). In de verpleegkundige opleiding bij Thomas More Mechelen is het ontwikkelen van de juiste competenties bij studenten verpleegkunde één van de doelstellingen. De docenten proberen immers om de studenten zodanig voor te bereiden dat ze in hun beroepsleven in staat zijn om de nodige competenties te hanteren.

2.2 Assessment van competenties

Het verwerven van de juiste competenties wordt dus steeds belangrijker. Daarom is het nodig dat deze ook geëvalueerd worden. Onderstaande piramide van Miller (2009) presenteert hierbij een bepaalde volgorde van toetsing:



Figuur 2: Miller's pyramid of assessment

Noot. Herdrukt van 'Assessing Dental Students' Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education", door Albino, J., et al. (2009), *Journal of dental education*, 72, p. 1405-1435.

Dit model start met de beoordeling van kennis en eindigt met de beoordeling van gedrag in de praktijk. Op het laagste niveau van de piramide vindt men het terugroepen van feitelijke kennis (weten - knows) terug dat kan worden beoordeeld door schriftelijke examens, gevolgd door toepassing van kennis (weten hoe - knows how). Het beoordelen van praktische vaardigheden kan worden uitgevoerd (laten zien hoe - shows how) door OSCE (Objective Structured Clinical Examination)¹ of simulaties in kunstmatige settings. Als laatste stap in de piramide worden er best examens afgenomen (doen - does) in een omgeving waarbij echte patiënten betrokken zijn. In dit geval spreekt men van 'bedside' examens waarbij

¹ OSCE is een assessment methode gebaseerd op de prestaties van studenten om hun klinische competenties te beoordelen. Met behulp van een stationsproef worden de studenten getoetst in verschillende situaties.

studenten zorgen voor een patiënt en vervolgens hun bevindingen dienen samen te vatten en bespreken met één of twee examinatoren.

Ramani en Leinster (2008) en Al-Wardy (2010) wijzen erop dat het beoordelen van de prestaties in het dagelijks leven tijdens de dagelijkse patiëntenzorg het enige examen is dat het hoogste niveau van de piramide bereikt.

Er is in het onderwijs dus een verschuiving van een kennisgerichte naar een deels competentiegerichte aanpak waardoor ook er ook in de evaluatievormen een bijsturing doorgevoerd dient te worden. In een onderwijsvorm die kennisgericht is, wordt het evalueren door Segers (2004) omschreven als een testcultuur met een aantal specifieke kenmerken. Zo zijn dit type evaluaties vooral gericht op kennisreproductie, zijn ze gedecontextualiseerd en geïsoleerd van het leerproces, de evaluatie is summatief en gaat uit van de docent. Zoals eerder werd gesteld beslaan competenties een ruimere variatie aan kenmerken die geëvalueerd dienen te worden. Segers (2004) spreekt in deze context over een assessment dat dient uitgevoerd te worden. Belangrijk daarbij is dat het brede pallet aan elementen - cognitieve, metacognitieve, sociale en affectieve - dat onderdeel is van de competenties tijdens de evaluatiemomenten getoetst kunnen worden. Het toetsen van deze verschillende variaties aan competenties gebeurt bij voorkeur geïntegreerd en is authentiek of gecontextualiseerd.

Invloed van het assessment op stress

Studenten beschouwen examens meestal als stressvolle gebeurtenissen waarvan men de uitkomst vreest, omdat hun toekomst afhankelijk is van de resultaten. Bengtsson et al. (2016) beschrijven dat de studenten hun vrees voor de beoordeling zelf en het feit beoordeeld te worden leidde tot angst, nervositeit en stress. Het wordt algemeen ook aangenomen dat angst voor het afleggen van examens toeneemt naarmate het examen dichterbij komt (Lotz, C., & Sparfeldt J.R., 2016). Deze angst voor een examen heeft meestal de neiging om de te evalueren vaardigheden van studenten te verzwakken. Ook neemt hun vermogen om kennis over de inhoud van de te evalueren cursus af, wat leidt tot slechte prestaties tijdens de evaluatie die op hun beurt de testcores en algehele cijfers kunnen beïnvloeden (Markman, Balik, Bercovitz & Ehrenfeld, 2010). De meeste studenten verpleegkunde ervaren een hoge mate van angst tijdens het

eindexamen, hoewel hun cijfers gedurende het hele semester goed zijn (Akbar-Boorang & Aminyazdi, 2009).

Door de studenten te ondersteunen bij het ontwikkelen van hun bewustzijn van hun eigen theoretische kennis, evenals hun emotionele, praktische en interpersoonlijke vaardigheden door het gebruik van verschillende lesmethoden, kunnen de studenten de stressvolle aard van het assessment beter aan. Daarnaast gaat de voorkeur uit naar een formatieve evaluatie en is die leerlinggericht. Hierbij stellen McKeachie & Svinicki (2006) het gebruik van niet-beoordeelde assessments voor als aanvulling. Wanneer studenten dezelfde procedure voor het examen uitvoeren en feedback krijgen van andere studenten of docenten, kunnen ze zich meer bewust zijn van hun mogelijkheden waarbij het zelfvertrouwen positief geprikkeld kan worden. Het mixen en trianguleren van verschillende evaluatiemethodes verdient tot slot ook de nodige aandacht.

2.2.1 Competence assessment programs

Door de accentverschuiving van het evalueren van kennis naar het evalueren van competenties dient de manier waarop deze evaluaties vorm krijgen ook kritisch bekeken te worden. Baartman (2008) deed in haar onderzoek een aantal aanbevelingen om de betrouwbaarheid van de evaluatie van competenties te versterken. Ze stelt dat voor de betrouwbare uitbouw van de evaluatie van competenties een 'Competence Assessment Program' ontwikkeld moet worden. Deze Competence Assessment Programs (CAP) hebben tot doel om de brede variëteit aan competenties die afgetoetst worden bij een student op een volledige en betrouwbare manier tegen het licht te houden. Baartman geeft aan dat een CAP bij voorkeur aan de volgende 12 kwaliteitscriteria voldoet: acceptatie, authenticiteit, betekenisvolheid, cognitieve complexiteit, eerlijkheid, geschiktheid voor onderwijsdoelen, onderwijsgevolgen, herhaalbaarheid van beslissingen, ontwikkeling van zelfsturend leren, tijd en kosten, transparantie en vergelijkbaarheid.

Acceptatie: alle betrokkenen staan achter de assessmentcriteria en de manier waarop het assessment zal uitgevoerd worden. Ze zijn ook overtuigd van de kwaliteit van het assessment dat zal uitgevoerd worden.

Authenticiteit: de uit te voeren taak leunt qua fysieke uitvoering en sociale context zo sterk mogelijk aan bij de taken die in het professioneel leven.

Betekenisvolheid: voor de stakeholders (in eerste instantie studenten en docenten) moet het assessment ook een meerwaarde bieden. Voor de studenten moet er een leerervaring aan gekoppeld kunnen worden en heeft het een betekenis in het leerproces zelf. Voor de docenten moet de link met de competenties die verworven dienen te zijn voor de toekomstige professionele activiteiten ook helder te zijn.

Cognitieve complexiteit: in een CAP moeten de door de student verworven kenniselementen weerspiegeld worden en is er zicht op het denkproces zelf.

Eerlijkheid: studenten moeten een eerlijke kans krijgen hun competenties te demonstreren op verschillende manieren, zo wordt de kans op een vertekening van de beoordeling kleiner.

Geschiktheid voor de onderwijsdoelen: de assessmentdoelen moeten afgemeten worden aan de onderwijsdoelen. Het assessment moet in de lijn liggen van de standaarden, instructies en het curriculum dat in de onderwijsfase werd gehanteerd.

Onderwijsgevolgen: de mate waarin een CAP de positieve effecten van leren ondersteunt en anderzijds de negatieve effecten zo beperkt mogelijk houdt.

Herhaalbaarheid van beslissingen: het resultaat van een beoordeling in een CAP mag niet afhankelijk zijn van de specifieke evaluator of van de specifieke situatie waarin het assessment plaatsvindt. Daarom gebeurt een CAP bij voorkeur door verschillende evaluatoren en worden verschillende evaluatiemomenten gecombineerd of getrianguleerd.

Ontwikkeling van zelfsturend leren: een CAP ondersteunt het best het zelfsturend leren van de student. Dit type assessment bevat dan ook bij voorkeur specifieke methodes die het leren ondersteunen zoals het oefenen van zelfassessment en het geven en ontvangen van feedback.

Tijd en kosten: de investeringen op het vlak van geld en tijd dienen zowel voor de student als de evaluator in verhouding te zijn tot de voordelen die uit de aanpak gegenereerd kunnen worden.

Transparantie: het CAP moet voor alle betrokkenen helder en begrijpelijk zijn. Studenten en evaluatoren moeten het doel van het assessment kennen en de criteria om te kunnen slagen. Ook voor externen dient het duidelijk te zijn hoe het CAP ontwikkeld is en hoe het uitgevoerd wordt.

Vergelijkbaarheid: het assessment moet op een consequente manier opgezet en uitgevoerd worden. De taken, criteria en omstandigheden moeten in verhouding staan tot de sleutelcomponenten die afgetoetst worden. Ook het geven van scores aan de studenten dient consequent te gebeuren.

Belangrijk om aan te stippen is dat het introduceren van assessments in het onderwijs niet hoeft te betekenen dat de eerder beschreven testcultuur volledig dient uitgevlakt te worden (Baartman, 2008; Jonsson et al., 2009). Beide evaluatiemethodes vullen elkaar aan omdat de student ook nog altijd eerst de nodige basiskennis moet verwerven om een assessment van competenties tot een goed einde te brengen. Daarom is een evaluatie van kenniselementen ook aangewezen. Een CAP is met andere woorden geen vervanging van de klassieke kennisevaluatie - zoals een schriftelijk examen -, maar biedt er een aanvulling op.

Omdat competenties dus een complex geheel vormen, en omdat het CAP zo kwaliteitsvol mogelijk moet zijn voor alle betrokken partijen, is het aangewezen dat verschillende types van testen en assessment met elkaar gecombineerd worden. Concreet kan dit door het afleggen van een theoretisch examen voor een opleidingsonderdeel en aanvullend een vaardigheidsexamen te organiseren (Baartman, 2008; Jonsson et al., 2009). Zo wordt ook vermeden dat de kennisonderdelen onderbelicht blijven. Waar het zwaartepunt van de evaluatie dient te liggen is afhankelijk van de criteria die in het opleidingsonderdeel gelden om ervoor te kunnen slagen.

2.3 Kwaliteit van assessments

De uitkomsten van een CAP worden vaak als basis gebruikt voor het nemen van high-stake beslissingen over de lerende. De kwaliteit van CAP's mag dus niet genegeerd worden en vormt een sleutelement in de verschuiving van een test-naar een assessmentcultuur (Baartman, 2008).

Naast de accentverschuiving op vlak van wat geëvalueerd wordt, zijn er ook veranderingen in de visie over de kwaliteit van assessments. Het betreft hier een verschuiving van de psychometrische traditie naar een edumetrische benadering (Dierick & Dochy, 2001). De psychometrische visie sluit eerder aan bij de beschrijving van kwaliteit binnen de testcultuur, terwijl de edumetrische benadering meer geschikt is voor de verschillende kenmerken van de assessmentcultuur (Moss, 1994). Deze twee visies over kwaliteit worden vaak gezien als uiteinden van een continuüm.

Er zijn vier grote verschillen tussen de twee benaderingen. Ten eerste is er een verschil in het onderwerp van de meting: in de psychometrie worden onveranderlijke kenmerken gemeten (intelligentie, persoonlijkheid), terwijl in de edumetrische benadering competenties worden geëvalueerd die dynamisch zijn en kunnen veranderen doorheen de tijd. Ten tweede is het referentiepunt dat genomen wordt om tot een oordeel te komen verschillend. In de psychometrische benadering worden lerenden met elkaar vergeleken om ze te beoordelen. Vanuit de edumetrie wordt hier anders naar gekeken. Er zijn binnen deze benadering twee mogelijke uitgangspunten: enerzijds kunnen criteria die beschrijven wat geleerd moet worden als maatstaf gebruikt worden of anderzijds kunnen lerenden vergeleken worden met hun eerdere prestaties (Martin, 1997; Sadler, 1987). Ten derde is er in de psychometrische traditie een andere visie op betrouwbaarheid dan binnen de edumetrische. Binnen de psychometrie wordt betrouwbaarheid nagestreefd door standaardisatie en objectiviteit (Birenbaum, 1996). De edumetrische benadering gaat uit van de idee dat CAP's gedeeltelijk afhankelijk zijn van menselijke observaties (Cronbach, Linn, Brennan & Haertel, 1997). Zo verschuift de focus dus van standaardisatie naar het gebruik van meerdere metingen en de generaliseerbaarheid over beoordelaars en

taken. Ten vierde wordt in de edumetrie de nadruk gelegd op het formatieve van het assessment en minder op het summatieve, zoals in de psychometrie.

Tabel 3 geeft een overzicht van de belangrijkste verschillen tussen de twee benaderingen.

Tabel 3

Extremen van de psychometrische en edumetrische benaderingen

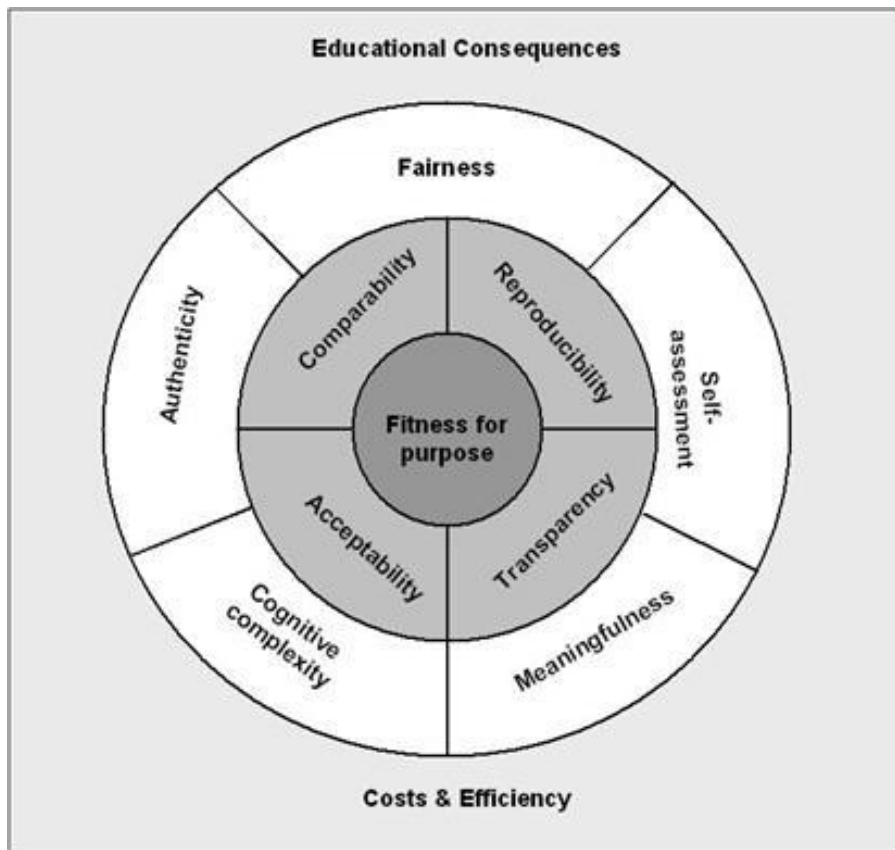
| | Testcultuur | Assessmentcultuur |
|-----------------|---|--|
| Onderwerp | Vaste kenmerken | Competentie-ontwikkeling |
| Referentie | Normen | Criteria/zichzelf |
| Betrouwbaarheid | Objectiviteit | Menselijke observatie |
| Meting | Standaardisatie | Meerdere metingen en generaliseerbaarheid |
| Functie | Summatief | Formatief |
| Kwaliteit | Betrouwbaarheid is voorwaarde voor validiteit | Complementaire kwaliteitscriteria (betekenisvol, feedback, leerproces) |

Noot. Herdrukt van 'Assessing the assessment. Development and use of quality criteria for Competence Assessment Programmes', door Baartman, L., (2008), Doctoral dissertation, Universiteit Utrecht, Nederland.

Deze verschuiving van een psychometrische benadering naar een edumetrische benadering brengt nieuwe perspectieven met zich mee om te kijken naar de kwaliteit van assessments.

2.3.1 Wheel of competence

Zoals beschreven in 2.2.1 definieerde Baartman (2008) 12 kwaliteitscriteria waaraan een CAP moet voldoen. Ze besprak deze criteria met leerkrachten en experts in assessments. Op die manier kwam ze tot 'the wheel of competence' (competentiewiel). Dit is een kader waarin naast de kwaliteitscriteria ook de verwevenheid ervan wordt benadrukt. Figuur 3 toont het competentiewiel.



Figuur 3. Wheel of competence

Noot. Herdrukt van "Assessing the assessment. Development and use of quality criteria for Competence Assessment Programmes", door Baartman, L., (2008), Doctoral dissertation, Universiteit Utrecht, Nederland.

Het wiel bestaat uit 3 lagen om te benadrukken dat de criteria met elkaar verband houden. In de kern zit de geschiktheid voor de onderwijsdoelen. Dit criterium vormt de basis voor de ontwikkeling van CAP's. Rond de kern bevindt zich de eerste laag. Hierin zitten de basiscriteria: vergelijkbaarheid, reproduceerbaarheid van beslissingen, acceptatie en transparantie. De buitenste laag bestaat uit de nieuwere criteria die gebonden zijn aan de assessmentcultuur: eerlijkheid, authenticiteit, cognitieve complexiteit, betekenisvolheid en geschiktheid voor de ontwikkeling van zelfsturend leren. Het vierkant rond het wiel staat voor de bredere educatieve context waarin het assessment plaatsvindt. Hierbinnen bevinden zich twee criteria: ten eerste zijn er tijd en kosten en ten tweede de onderwijsgevolgen.

De criteria in de binnenste laag zijn een voorwaarde voor de criteria in de laag er rond. Wanneer aan een CAP gewerkt wordt, kan best eerst gewerkt worden aan de criteria in de binnenste laag, alvorens te werken aan die in de buitenste laag.

Naast deze nieuwere criteria zitten ook de meer traditionele kwaliteitscriteria voor evaluatie, namelijk betrouwbaarheid en validiteit, verweven in het competentiewiel. Het directe gebruik ervan is immers problematisch voor het evalueren van competenties (Baartman, 2008).

2.3.2 Focus IP

Binnen dit IP-project werd gekozen om te focussen op een selectie van vier criteria uit het kader van Baartman (2008). Deze selectie gebeurde op basis van de probleemstelling die in hoofdstuk 1 geschetst werd. De focus ligt hierbij op volgende criteria: vergelijkbaarheid, reproduceerbaarheid van beslissingen, transparantie en betekenisvolheid.

Vanuit de criteria vergelijkbaarheid en reproduceerbaarheid van beslissingen willen we kijken naar de betrouwbaarheid van het examen. Daarnaast willen we de transparantie ervan onderzoeken. Bij het criterium zinvolheid spitsen we ons toe op de zinvolheid van de feedback die gegeven wordt door de docenten en hoe die ervaren wordt door de studenten. In deze vier criteria zitten ook vormen van validiteit vervat. Vanuit deze criteria werpen we een blik op de validiteit van het examen.

In wat volgt worden de criteria die de IP-groep selecteerde verder toegelicht.

2.3.2.1 Vergelijkbaarheid

Vergelijkbaarheid is een eerste kwaliteitscriterium dat onder de betrouwbaarheid van evaluaties van competenties valt. Een CAP kan maar vergelijkbaar zijn volgens Baartman (2008) als ze wordt uitgevoerd op een consistente en betrouwbare manier. CAP's zijn over het algemeen minder gestandaardiseerd dan de traditionele toetsen en hangen sterk af van een menselijk oordeel. Daarom is er nood aan consistentie bij de beoordeling (Haertel, 1991). Daarnaast moeten ook de omstandigheden waarin het assessment wordt afgenomen zo veel mogelijk hetzelfde zijn voor alle leerders.

Vervolgens geeft Benett (1993) aan dat assessments die worden uitgevoerd in minder gecontroleerde en gestandaardiseerde situaties, zoals het assessment op de werkplaats, evengoed gebaseerd kunnen zijn op een reeks van taken. Deze taken zouden dan niet identiek zijn, maar de assessments zijn wel consistent met de belangrijkste functies in het achterhoofd. Voorbeelden hiervan zijn een gemeenschappelijke beoordelingsprocedure of een thema en doel. Benett (1993) concludeert dus dat het niet noodzakelijk is om volledige standaardisatie te waarborgen om te komen tot voldoende vergelijkbaarheid. Het vergelijkend onderzoek toont aan dat dit op verschillende manieren bereikt kan worden en dat CAP's in competentiegericht onderwijs vergelijkbaar kunnen zijn als er metingen gebeuren om vergelijkbaarheid te garanderen.

Naast betrouwbaarheid toont Baartman (2008) ook aan dat vergelijkbaarheid gerelateerd kan worden aan het externe aspect van validiteit (Messick, 1994, 1995). Externe validiteit is een voorwaarde voor generaliseerbaarheid. Dit is mogelijk wanneer relaties tussen assessments en andere metingen de competentie weerspiegelen. Hiernaast maakt vergelijkbaarheid het ook makkelijker om te komen tot reproduceerbaarheid van beslissingen. Vergelijkbaarheid komt tot stand wanneer twee assessments van dezelfde competentie afgenomen worden op een verschillend moment of door verschillende observatoren en deze sterk vergelijkbaar zijn en een hoge correlatie vertonen. In dit geval kan men ervan uitgaan dat de beslissingen gebaseerd op de resultaten van de twee metingen reproduceerbaar zullen zijn door verschillende observatoren en in verschillende situaties. Bijgevolg kan er gesproken worden van generaliseerbaarheid. Anderzijds is vergelijkbaarheid niet helemaal hetzelfde als het externe aspect van Messick

(1994, 1995). Vergelijkbaarheid focust op de condities bij het uitvoeren van een assessment, bijvoorbeeld als de taken, beoordelingscriteria en omstandigheden vergelijkbaar zijn over verschillende assessments. Messick (1994, 1995) daarentegen vergelijkt vooral verschillende testen die hetzelfde meten en verschillende constructies.

Invloed van het assessment op stress

Wass , Van der Vleuten, Shatzer en Jones (2001) hebben een praktisch examen beschreven als een echte uitdaging voor opvoeders. Het is moeilijk om de validiteit en betrouwbaarheid te garanderen door het gebruik van meerdere examinatoren in verschillende gevallen. De condities en zorgsituaties van de betrokken patiënten in een medisch praktisch examen variëren, dus is het problematisch om enige standaardisatie te creëren. Daarom is het belangrijk om studenten te laten handelen en te beoordelen op basis van een goed gedefinieerd protocol om stress te vermijden.

2.3.2.2 Reproduceerbaarheid van beslissingen

Reproduceerbaarheid van beslissingen is een tweede kwaliteitscriterium dat, naast vergelijkbaarheid, sterk aansluit bij betrouwbaarheid. Dit criterium slaat volgens Baartman (2008) op het feit dat beslissingen die gemaakt worden over studenten gebaseerd moeten zijn op meerdere assessments die bovendien worden afgenomen door verschillende assessoren en in verschillende omstandigheden. Het doel van een assessment is niet om een prestatie te beoordelen in één bepaalde situatie, geobserveerd door één beoordelaar, maar om een algemene conclusie te kunnen trekken over de competenties van de leerder.

De reproduceerbaarheid biedt bovendien een mate van betrouwbaarheid die zeer geschikt is voor het beoordelen van competenties en vaardigheden. Ondanks dit voordeel blijkt uit het onderzoek van Baartman (2008) dat wanneer docenten verschillende kwaliteitsaspecten gaan scoren naargelang hun belangrijkheid, de reproduceerbaarheid van beslissingen relatief laag scoort ten opzichte van de andere criteria. Een reden hiervoor zou kunnen zijn dat docenten gesteld zijn op hun autonomie en niet het gewoon zijn om samen met andere docenten de studenten te beoordelen. Een tweede, eerder praktische, reden die wordt

aangehaald is dat gebruik van meerdere beoordelaars wordt gezien als niet haalbaar, te duur en bovendien tijdrovend. Nochtans is het inschakelen van meerdere beoordelaars een effectieve manier om aan reproduceerbaarheid te werken (Moss, 1994; Van der Vleuten & Schuwirth, 2005).

2.3.2.3 Transparantie

Volgens Baartman (2008) betekent transparantie dat een CAP begrijpbaar moet zijn voor alle participanten. Zo moeten enerzijds studenten de criteria kennen waarop ze gescoord zullen worden, wie de beoordelaars zullen zijn en wat het doel van het assessment is (Bagheri, Forogeth & Fallah, 2012). Het moet voor hen duidelijk zijn wat van hen verwacht wordt zodat ze zich kunnen voorbereiden op het assessment. Sluijsmans, Peeters, Jakobs en Weijzen (2012) verstaan hieronder dat alle informatie voorhanden is die de best denkbare voorbereidings- en antwoordstrategie van studenten mogelijk maakt. Anderzijds moeten de docenten weten wat het CAP inhoudt zodat zij zich kunnen voorbereiden op hun rol als beoordelaar. Sluijsmans et al. (2012) benadrukken hierbij ook het belang van de toetsbekwaamheid van docenten. De kwaliteit van toetsing wordt namelijk in grote mate bepaald door de kwaliteit van de beoordelaar. Elke docent moet immers in staat zijn om toetsinformatie te interpreteren en te gebruiken om vast te stellen waar hun studenten staan en hoe zij verder kunnen bijdragen aan het leren van de student (Absolum, Flockton, Hattie, Hipkins & Reid, 2009). Sluijsmans et al. (2012) hebben toetscompetenties van docenten beschreven. Docenten moeten:

- in staat zijn om toegankelijke en meetbare leerdoelen te formuleren die aansluiten bij de gewenste inhoud en beheersingsniveau volgens de curriculum doelen en visies.
- beschikken over verschillende strategieën om met studenten te communiceren over de leerdoelen en hoe deze bereikt kunnen worden.
- effectieve en bruikbare feedback op studentenwerk kunnen geven.
- in staat zijn om beoordelingsmodellen te construeren die leiden tot bruikbare informatie om beslissingen te kunnen nemen over studenten en/of klassen. Deze beslissingen moeten leiden tot verbeteringen van het leren, groeien of ontwikkelen van de student.

Transparantie is ook gelinkt aan twee vormen van validiteit (Baartman, 2008). In de eerste plaats hangt het samen met het structurele aspect van constructvaliditeit. Dit betekent dat de manier van scoren consistent moet zijn met wat belangrijk en minder belangrijk is bij de competentie. Bovendien moet het voor de leerder duidelijk zijn wat de criteria zijn en hoe zwaar ze doorwegen bij de beoordeling (Messick, 1994). Ten tweede is er ook een verband tussen transparantie en het consequentiële aspect van constructvaliditeit (Messick, 1995). Wanneer de criteria gekend zijn en begrepen worden door de leerders beïnvloedt dit immers het leren (Dochy & McDowel, 1997). Jonsson et al. (2009) geven aan dat het tot betere leerprocessen leidt. Daarom moeten product- en proceskwaliteitscriteria worden geïdentificeerd, gedefinieerd en gecommuniceerd aan kandidaten en beoordelaars (Dochy & Moerkerke, 1997).

Invloed van het assessment op stress

Een manier om aan de transparantie van een assessment te werken is door het gebruik van vooraf gekende evaluatielijsten. Dit wordt door examinatoren over het algemeen als positief ervaren: het leidt voor hen tot minder twijfel op het evaluatiemoment zelf. Voor studenten zijn de effecten van zulke evaluatielijsten echter niet eenduidig positief. Enerzijds voelen ze dit aan als een fair systeem, maar anderzijds levert een vooraf gekende evaluatielijst extra stress op bij de studenten (Bagheri et al., 2012).

2.3.2.4 Betekenisvolheid

Alle stakeholders moeten het assessment als zinvol beschouwen. Een assessment zou studenten moeten laten omgaan met betekenisvolle problemen die ervoor zorgen dat ze waardevolle ervaringen opdoen. Hierbij moet een link zijn tussen de assessmenttaak en de persoonlijke interesses van de leerder (McDowell, 1995). Een CAP moet betekenisvol zijn en moet met andere woorden een significante waarde hebben voor zowel de docenten als de leerders, mede door de feedback die gegeven en verkregen wordt (Linn, Baker & Dunbar, 1991). Baartman (2008) voegt hieraan toe dat de leerders nood hebben aan betekenisvolle feedback en assessment criteria om hun leerproces te sturen. Voor formatieve assessments is het belangrijk dat er vaak effectieve feedback gegeven wordt aan de leerder. Voor summatieve assessments is het belangrijk dat feedback gerelateerde criteria zoals betekenisvolheid opgenomen worden (Baartman, 2008).

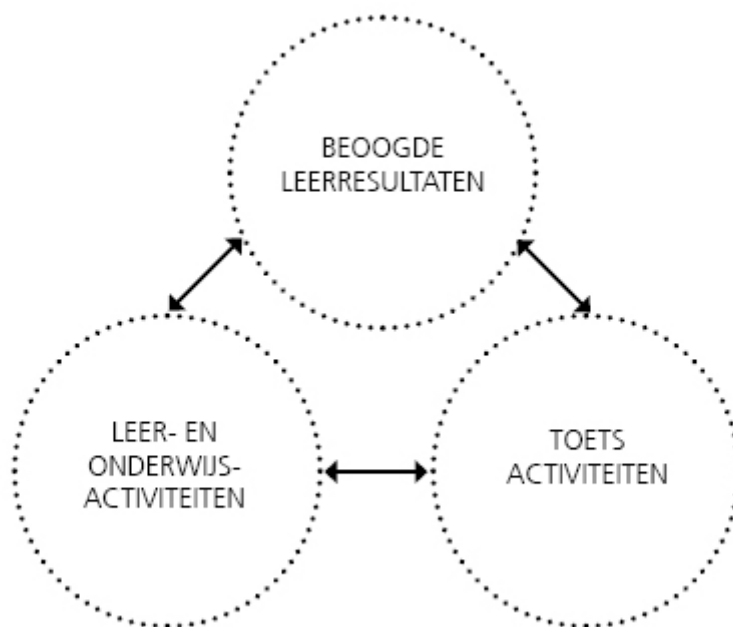
Bovendien is het criterium van betekenisvolheid gelinkt aan drie aspecten van validiteit: het inhoudelijke, het substantieve en het consequentiële aspect (Baartman, 2008). Onder het inhoudelijke aspect wordt verstaan dat de taken van het assessment herkenbaar moeten zijn voor de leerders en ze deze taken als waardevol moeten beschouwen. Het substantieve aspect betekent dat het assessment moet bestaan uit herkenbaar gedrag en processen die nodig zijn op de werkvloer. Het derde en laatste aspect, consequentiële validiteit, houdt de positieve en negatieve consequenties van het assessment in. Het is namelijk zo dat als een assessment een waardevolle leerervaring en een gids voor leerprocessen wil zijn, hij herkenbaar moet zijn voor de leerder en door hem als waardevol beschouwd moet worden. Als dit niet het geval is, zou het kunnen dat vooropgestelde doelen niet bereikt worden. Volgens McDowell (1995) moet een assessmenttaak inhoudelijk betekenisvol zijn voor leerders om een leereffect en dus een positieve consequentie te kunnen bereiken. Er kunnen dus analogieën gevonden worden tussen betekenisvolheid en inhoudelijke- en consequentiële validiteit.

Betekenisvolheid is op te delen in drie criteria: geschiktheid voor het doel (*fitness for purpose*), geschiktheid voor self-assessment (*fitness for self-assessment*) en aannemelijkheid (*acceptability*). De invloed van geschiktheid voor het doel en aannemelijkheid op stress wordt voor deze criteria's aansluitend beknopt besproken.

Geschiktheid voor het doel

Geschiktheid voor het doel is vergelijkbaar met de idee van constructive alignment, figuur 4 (Biggs, 1996; Biggs, 1999). Dit omschrijft dat alle CAP's afgestemd moeten zijn op het doel van het leerproces en het gegeven onderwijs (Baartmann, 2008). Biggs (1996; 1999) geeft aan dat constructive alignment geen specifiek type van onderwijs, leren en assessment voorschrijft, maar dat de drie afgestemd moeten zijn op elkaar. Deze afstemming wordt bedreigd, aangezien onderwijs en leren meer en meer competentiegericht worden en de ontwikkeling van adequate beoordelingsmethoden achter lijkt te blijven. Als het onderwijs en het leren gebaseerd is op het verwerven van competenties, impliceert constructive alignment dat het assessment ook competentiegericht moet zijn. Het is belangrijk

dat studenten hun leeromgeving (leer- en onderwijsactiviteiten) ervaren als één die gebaseerd is op wederzijds vertrouwen tussen docenten en studenten, omdat studenten vastgesteld hebben dat wanneer zij frequente feedback krijgen van docenten, zij zich meer gemotiveerd voelen en effectiever leren werken (Rowbatham, 2010; Tiberius & Billson, 1991). Daarbij is het belangrijk dat docenten verpleegkunde begrijpen hoe studenten redeneren en zich voorbereiden op het competentiegericht examen. Zo kunnen de docenten leer- en onderwijsactiviteiten aanbieden die de voorbereiding van de studenten ondersteunen en om zo nervositeit van de studenten te voorkomen (Bengtsson et al., 2016).



Figuur 4: Constructive alignment, Biggs.

Noot. Herdrukt van "Toetscyclus ontwerpen", door Hogeschool van Amsterdam, (2016). Geraadpleegd van

<https://score.hva.nl/docent/toetscyclus/Paginas/Ontwerpen.aspx>

Geschiktheid voor self-assessment

Dit is gerelateerd aan betekenisvolheid omdat een CAP als betekenisvoller wordt ervaren als het zelfregulerend leren stimuleert, een kwaliteit die van competente professionals wordt verwacht (Baartman, 2008). Zelfregulerend leren is het

zelfstandig, en met verantwoordelijkheid, in handen nemen van de sturing voor de eigen leerprocessen (Meusen-Beekman, Joosten-Ten Brinke & Boshuizen, 2016). Een hoge mate van zelfregulatie leidt tot de bijstelling en controle van de eigen inzet, concentratie, motivatie, werkwijze en het resultaat. Feedback op het leerproces en de behaalde prestatie is van groot belang om zelfregulatie te ontwikkelen bij de leerder. Assessment kan een rol spelen in het proces om meer zelfregulatie te ontwikkelen door duidelijk te maken wat de criteria zijn, door het aantonen van zwaktes (door feedback) en door het stimuleren van reflectie op het eigen leerproces (Baartman, 2008). Moorthy, Munz, Sarker en Darzi (2003) geven aan dat zonder objectieve feedback het moeilijk is om prestaties te verbeteren van de leerder, het gebruik van gevalideerde methoden zoals checklists of beoordelingsschalen is daarom aangewezen. Bovendien heeft constructieve feedback een grote invloed op het leren en is feedback vaak belangrijker voor de student dan zijn of haar examenresultaat (Bengtsson et al., 2016).

Aannemelijkheid

Een CAP wordt gemakkelijker geaccepteerd als het betekenisvol is, daarom dat aannemelijkheid gerelateerd is aan betekenisvolheid (Baartman, 2008). Bengtsson et al. (2016) beschrijven dat de beoordeling van bekwaamheid, ook al is het stressvol, van essentieel belang is voor studenten die verpleegkunde studeren. Om op die manier ervoor te zorgen dat veilige zorg van goede kwaliteit aan alle patiënten wordt gegeven. Het in relatie brengen van kwaliteitsvolle zorg en het afleggen van een assessment kan de aannemelijkheid dus ten goede ondersteunen. Aannemelijkheid voegt aan het criterium 'transparantie' toe dat stakeholders (docenten en studenten) de gebruikte procedures en criteria zouden moeten accepteren. Het heeft te maken met de attitudes en de kijk ten opzichte van het assessment. Ook zouden de stakeholders vertrouwen moeten hebben in de kwaliteit van het CAP (Stokking, Van der Schaaf, Jaspers & Erkens, 2004).

3 Opzet - methodologie van het project

3.1 Onderzoeksmethode

We verdeelden het onderzoek in een theoretisch en een praktisch luik. In het theoretisch luik wilden we nagaan of het huidig competentie assessment van

Thomas More een goed assessment is, rekening houdend met de items uit 'the wheel of competence'. Dit onderzochten we zowel kwantitatief als kwalitatief. Bijgevolg spreken we van een mixed method methode en wordt er voldaan aan methodologische triangulatie om de interne validiteit van ons onderzoek te verhogen (Mortelmans, 2013). Het kwantitatieve gedeelte werd ingevuld door een survey af te nemen bij de studenten vlak na het afleggen van het praktisch examen verpleegkundige vaardigheden. De resultaten van deze vragenlijst dienden om inzicht te krijgen in de belangrijkste factoren die van invloed zijn op de slaagkansen van de studenten. Studenten die niet deelnamen aan het praktisch examen kregen ook een survey toegestuurd om te onderzoeken welke factoren een rol spelen bij hun afwezigheid op het praktisch examen verpleegkundige vaardigheden. Aan de hand van de resultaten die voortkwamen uit de survey proberen we een antwoord te bieden op volgende onderzoeksvragen:

1. a. In hoeverre ervaren studenten het praktisch examen verpleegkundige vaardigheden als transparant?
 - 1.b. In welke mate heeft dit invloed op stress?
2. a. Hoe ervaren studenten feedback in de huidige vorm?
 - 2.b. In welke mate heeft dit invloed op stress?

Voor het kwalitatief gedeelte observeerden de projectleden het praktisch examen verpleegkundige vaardigheden, voerden ze een documentenanalyse uit en organiseerden ze een focusgroepgesprek met de docenten van Thomas More die het praktisch examen verpleegkundige vaardigheden evalueren. De observatie en documentenanalyse dienden om zicht te krijgen op de werkelijke situatie. Met het focusgroepgesprek wilden we aftoetsen hoe de docenten tegenover de huidige aanpak staan en hoe zij de lagere slaagkansen van de studenten verklaren. Aan de hand van de resultaten die voortkwamen uit de observatie, documentenanalyse en de focusgroep proberen we specifiek een antwoord te bieden op volgende onderzoeksvraag:

3. In hoeverre ervaart Thomas More problemen in verband met de betrouwbaarheid en de validiteit en dus de kwaliteit van het praktisch examen verpleegkundige vaardigheden.

In het praktisch luik (hoofdstuk 6) geven we suggesties aan Thomas More om een kwaliteitsvol assessment uit te werken om het praktisch examen verpleegkundige vaardigheden te optimaliseren.

3.2 Methodologie survey

Om de eerste twee onderzoeksvragen en hun deelvragen (OV1a, OV1b, OV2a en OV2b) te beantwoorden, wilden we een zicht krijgen op de mening van de studenten over de transparantie van het examen en over de manier waarop ze de feedback die ze krijgen ervaren. We kozen ervoor om hen te bevragen aan de hand van een survey omwille van drie redenen. Ten eerste omdat een survey een geschikt instrument is om informatie te verzamelen over een groot aantal respondenten (Axinn & Pearce, 2006). Voor ons onderzoek was het immers opportuun om zo veel mogelijk studenten te bevragen over het examen. Ten tweede stelde het gebruik van een survey de onderzoekers in staat om zich te distantiëren van de studenten, wat wenselijk was omdat deze anonimiteit ervoor kon zorgen dat de studenten op een zo eerlijk mogelijke manier de vragen zouden beantwoorden. Zo werden vertekeningen geminimaliseerd en konden de studenten op een objectieve manier bevraagd worden over verschillende aspecten van het vak verpleegkundige vaardigheden en het praktische examen ervan (Wilson, 2019). Ten derde werden door het gebruik van een survey data verzameld die met behulp van statistische technieken geanalyseerd konden worden. Zo kon gezocht worden naar verbanden tussen de variabelen en kon nagegaan worden of de verbanden gegeneraliseerd konden worden. Dit leidde tot een nauwkeurige beschrijving van de resultaten (Wilson, 2019).

3.2.1 Respondenten

De populatie voor dit onderzoek bestond uit de studenten die het eerste bachelorjaar verpleegkunde volgen aan Thomas More Mechelen. In dit onderzoek wensten de onderzoekers zoveel mogelijk studenten uit het eerste jaar verpleegkunde Thomas More Mechelen te bevragen, dus was het nodig de vragenlijst gestructureerd te houden met vooral numerieke en gesloten vragen (Cohen, Manion & Morrison, 2007). In totaal zijn er 45 vragenlijsten ingevuld door de studenten die deelnamen aan het praktisch examen verpleegkundige vaardigheden. De vragenlijst die verstuurd werd naar de niet-deelnemers werd door vijf studenten ingevuld.

Onder de respondenten (deelnemers en niet-deelnemers samen) waren 40 vrouwen en 10 mannen. De jongste respondent was 18 jaar oud en de oudste 45. De gemiddelde leeftijd van de deelnemers was 21,18 jaar en van de niet-deelnemers 18,80 jaar. Het aantal keer dat de deelnemers het examen reeds aflegden varieerde van één tot en met vijf keer.

3.2.2 Dataverzameling

Op 21 januari 2019 vond het praktisch examen voor het opleidingsonderdeel verpleegkundige vaardigheden plaats. Na afloop van het examen bevroegen we de studenten aan de hand van een online survey in Qualtrics. Ook de studenten die niet deelgenomen hebben aan het praktisch examen hebben we een survey toegestuurd via onze contactpersoon Lien Proost. Zo konden we een beeld krijgen van de factoren die gerelateerd waren aan deze studenten.

3.2.3 Materiaal

De vragen uit de survey baseerden we op verschillende gevalideerde vragenlijsten. Om te onderzoeken in hoeverre de studenten het examen als transparant ervaren (OV1a) zochten we in gevalideerde vragenlijsten naar vragen die het concept transparantie meten. Ten eerste haalden we enkele vragen uit de 'Dundee Ready Education Environment Measure' (DREEM). Dit meetinstrument biedt de kans om de educatieve omgeving in hogescholen te evalueren en wordt gezien als de meest bruikbare tool voor dit doel. Bovendien kunnen aan de hand van dit instrument problemen worden gediagnosticeerd (Safari-Moradabadi et al., 2018). Ten tweede

vonden we bruikbare vragen in de 'Educational Practices Questionnaire (Student Version)' en de 'Student Satisfaction and Self-confidence in Learning' van de National League for Nursing (NLN). Deze gevalideerde vragenlijsten zijn instrumenten waarmee opleidingen aan zelfevaluatie kunnen doen. Ten derde gebruikten we ook enkele items uit een vragenlijst van Solvik & Struksnes (2018). Zij deden onderzoek naar de mate van tevredenheid onder de studenten over de organisatie van de lestijden en de relevantie ervan. Voor hun onderzoek baseerden zij zich ook op de vragenlijsten van de NLN. De vragen uit deze gevalideerde vragenlijsten werden aangevuld met een aantal vragen die we zelf opstelden.

Om na te gaan hoe de studenten de feedback in de huidige vorm ervaren (OV2a) baseerden we ons op twee gevalideerde vragenlijsten. We selecteerden er items uit die peilden naar het concept 'feedback'. Enerzijds haalden we hiervoor enkele items uit de DREEM. Anderzijds kozen we hiervoor enkele geschikte vragen uit de 'Educational Practices Questionnaire (Student Version)' van NLN.

Teneinde de invloed van de concepten transparantie en feedback op stress te kunnen onderzoeken (OV1b en OV2b) selecteerden we ook enkele vragen die het stressniveau van de studenten konden meten. Hiervoor vonden we enkele geschikte vragen in de DREEM. Daarnaast was ook de verkorte versie van de 'Spielberger State-Trait Anxiety Inventory' (STAI) een belangrijk bron waaruit enkele vragen haalden. De STAI is immers één van de meeste gebruikte meetinstrumenten om angst te onderzoeken en bovendien is de verkorte versie voldoende betrouwbaar² ($\alpha = 0.82$) (Marteau & Bekker, 1992).

De volledige vragenlijst met bronverwijzing per item werd opgenomen in bijlage 1.

² Een schaal is voldoende betrouwbaar als de Chronbach's alpha gelijk aan of groter is dan 0,60 (De Maeyer, Donche & van Daal, 2017).

3.2.4 Data-analyse

De analyse van de data gebeurde in drie stappen. In de eerste plaats werden de gebruikte schalen geanalyseerd met behulp van respectievelijk een exploratieve factoranalyse (EFA) en een confirmatieve factoranalyse (CFA). Als tweede stap in het analyseproces werden een aantal onafhankelijke variabelen beschreven. In de laatste stap werd er aan de hand van bivariate regressieanalyses gezocht naar verbanden tussen verschillende variabelen.

3.2.4.1 Samenstelling schalen

Op basis van een EFA werd nagegaan welke dimensies er schuilgaan in de gemeten indicatoren, meer specifiek het vinden van groepen samenhangende variabelen in de gegevens. Er werd gewerkt met een oblieke rotatie omdat we verwachtten dat de latente variabelen zouden correleren. Als schattingsmethode pasten we de 'maximum likelihood'-methode toe. Om te bepalen hoeveel factoren meegenomen werden in de EFA werd gekeken naar de screeplot en de eigenwaarden. Bij de screeplot werd de breuk in de lijn als referentiepunt genomen: het aantal factoren dat zich voor de breuk bevond werd weerhouden. Daarnaast werd ook naar de eigenwaarden gekeken: het aantal factoren waarvan de eigenwaarde hoger was dan één werd meegenomen. De betrouwbaarheid van de drie weerhouden factoren werd in een volgende stap getest. De waarde van de Cronbach's alpha moet daarbij gelijk zijn aan of boven 0.80 liggen om te spreken van een goede interne consistentie (De Maeyer et al., 2017)

Vervolgens werd ook een CFA uitgevoerd. Hiervoor werden de drie factoren gebruikt die uit de EFA kwamen en ondersteund werden door de theorie. Op basis hiervan werd een factorstructuur opgesteld en werd getoetst of dit model ook bij de data past. Zo kon de constructvaliditeit van de survey nagegaan worden (De Maeyer & van Daal, 2018). Aan de hand van twee fit-indexes werd nagegaan of het model bij de data past: de Comparative fit-index (CFI) en de Root-Mean-Square-Error-of-Approximation (RMSEA). Het model werd als aanvaardbaar beschouwd wanneer de CFI-waarde groter of gelijk was aan 0,95 en de RMSEA-waarde tussen 0,05 en 0,08 lag (De Maeyer & Van Daal, 2018).

Na de uitgevoerde EFA en CFA werden de items die verband hielden met elkaar, samengevoegd tot nieuwe variabelen. Voor elk concept werden de waarden van de verschillende items per respondent opgeteld en gedeeld door het aantal items. Zo werd een gemiddelde waarde berekend voor elk concept. Om de interpretatie van de resultaten later te vergemakkelijken, werden van deze waarden z-scores berekend (De Maeyer, Coertjens & Ardies, 2012). Deze z-scores werden gebruikt bij de regressieanalyses.

3.2.4.2 Beantwoorden onderzoeksvragen

Om de hoofdvragen van de eerste (OV1a) en tweede (OV2a) onderzoeksvraag te beantwoorden, werden de data geanalyseerd aan de hand van technieken uit de beschrijvende statistiek. Dit om de verzamelde data zo nauwkeurig en bondig mogelijk te beschrijven (Howitt & Cramer, 2007).

Vooraleer we de deelvragen konden beantwoorden, bekeken we de correlaties tussen de verschillende variabelen. Op deze manier kregen we een zicht op welke variabelen samenhangen, hoe sterk de correlatie is en of dit significant is. Aan de hand van de Pearson correlatiecoëfficiënt doen we uitspraken over de sterkte van de correlaties. Een correlatie is significant wanneer de p-waarde kleiner is dan 0,05 (De Maeyer et al., 2012).

Voor het beantwoorden van de deelvragen van de eerste (OV1b) en tweede (OV2b) onderzoeksvraag werden bivariate regressieanalyses uitgevoerd. Deze analysetechniek laat het toe om op zoek te gaan naar causale verbanden tussen twee kwantitatieve variabelen (De Maeyer et al., 2012). Er werden twee bivariate regressieanalyses uitgevoerd: één om het effect na te gaan van transparantie op stress (OV1b) en één om het effect na te gaan van feedback op stress (OV2b). Hierbij werd telkens gekeken naar de model fit en de statistische significantie van de regressiecoëfficiënten en het verband zelf. Om uitspraken te doen over hoe goed het regressiemodel bij de data past (model fit) werd gekeken naar de R²-waarde³. Aan de hand van de vuistregels van Cohen⁴ (1988) kon worden bepaald

³ De R²-waarde geeft aan hoeveel procent van de variantie in de afhankelijke variabele kan worden verklaard aan de hand van het regressiemodel (De Maeyer et al., 2012).

⁴ R² < 0,01 = verwaarloosbaar effect; 0,01 < R² < 0,059 = klein effect; 0,059 < R² < 0,138 = medium effect en 0,138 < R² = groot effect (De Maeyer et al., 2012).

hoe sterk het effect was van de onafhankelijke variabele op de afhankelijke. Met behulp van de F-toets kon nagegaan worden of het verband doorgetrokken kon worden naar de populatie. Hierbij werd een p-waarde kleiner of gelijk aan 0.05 als statistisch significant beschouwd (De Maeyer et al., 2012)

Alle analyses gebeurden met behulp van het statistische softwareprogramma R-studio.

3.3 Methodologie observatie praktisch examen

Naast het afnemen van de vragenlijst observeerden we het verloop van het praktisch examen en trachtten we hiermee een antwoord te formuleren op de derde (OV3) onderzoeksvraag. Dit deden we aan de hand van een ongestructureerde observatie. We gebruikten hierbij geen observatieschema om het gedrag van zowel de docenten als de studenten zo veel mogelijk in detail te kunnen vastleggen. Bovendien was de studie niet-participerend waardoor we duidelijk de rol van observator innamen (De Maeyer et al., 2017).

We benadrukten bij de studenten dat we zelf niets van de vaardigheden kenden om ervoor te zorgen dat we hen geen extra stress bezorgden.

3.4 Methodologie focusgroep

Om eveneens de derde (OV3) onderzoeksvraag te beantwoorden hebben we ervoor gekozen om een focusgroepgesprek te organiseren. Focusgroepgesprekken worden georganiseerd om mensen die een gemeenschappelijk kenmerk delen samen te brengen en onder leiding van een moderator te praten over een bepaald onderzoeksthema (Mortelmans, 2013). In dit geval zijn de respondenten docenten binnen de opleiding verpleegkunde die elk een vaardigheidsexamen afnemen en hierbij verschillende problemen ervaren. Hierbij streven we twee doelen na: enerzijds het aftoetsen van de antwoorden van de studenten uit de eerder afgenomen survey, anderzijds om het doel van ons onderzoek te concretiseren en af te toetsen bij de andere docenten binnen Thomas More (Mortelmans, 2013).

3.4.1 Respondenten

Aan de focusgroep namen 11 docenten deel, waaronder 5 mannen en 6 vrouwen. Enkel de docenten die een vaardigheidsexamen verpleegkundige vaardigheden afnemen binnen de opleiding verpleegkunde konden deelnemen. Verder was hun deelname aan het focusgroepgesprek geheel vrijwillig. We hebben de deelnemers gecontacteerd via onze contactpersonen bij Thomas More, Lien Proost en Carl Vanspauwen. We stuurden de contactpersonen een mail met de verduidelijking van het opzet van de focusgroep en stelden de vraag of ze dit aan hun collega's konden doorgeven. Aan de hand van een Doodle hebben we op het moment waarop de meeste docenten beschikbaar waren de focusgroep georganiseerd.

3.4.2 Dataverzameling

We hebben de data verzameld aan de hand van een leidraad met open vragen en stellingen, zie bijlage 2 (Bloor, Frankland, Thomas, & Robson, 2001; Billiet & Waege, 2001). Bij een aantal vragen van deze vragenlijst hebben we gevraagd om individueel en anoniem een antwoord te noteren. Het grootste deel van de vragen hebben we via de methode van een focusgroep bevroegd. De focusgroep vond plaats op 14 februari 2019 van 14u. tot 15.30u. in de lokalen van Thomas More Mechelen. Tijdens het gesprek maakten we zowel aantekeningen als een audio-opname van het gesprek. Aan de hand van bijbehorende notities en de audio-opname hebben we het focusgesprek uitgeschreven. De transcriptie werd nagelezen en goedgekeurd door Lien Proost, één van de deelnemende docenten en opdrachtgevers.

3.4.3 Data-analyse

We hebben de analyse uitgevoerd op basis van de aantekeningen waarbij de audio-opname als geheugensteun werd gebruikt. We pasten in de analyse het theoretisch kader van het onderzoek toe en hebben zo de antwoorden op deductieve wijze via open coding in verschillende categorieën kunnen onderverdelen. Deze onderverdeling hebben we gevisualiseerd door middel van twee tabellen: een tabel met de schriftelijke antwoorden en een tabel met de mondelinge antwoorden van de deelnemende docenten (bijlage 2). We hebben er voor gekozen om dit op te splitsen om zo bij de rapportering van de resultaten de mogelijkheid te hebben om eventuele discrepanties aan te tonen tussen de individuele en (anonieme)

schriftelijke en de mondelijke antwoorden. Bovendien kunnen we met dergelijke tabel eenvoudig aantonen welke onderwerpen vaak aan bod zijn gekomen en onder welke categorie de onderzoekers ze plaatsten. Gezien we met verschillende onderzoekers te werk zijn gegaan in het analyseren van de data, kunnen we spreken van onderzoeker triangulatie om de interne validiteit te waarborgen (Mortelmans, 2013).

3.5 Methodologie documentanalyse

Met de analyse van de verkregen relevante interne documenten trachten we mede de derde (OV3) onderzoeksvraag te beantwoorden. Deze documenten leverden waardevolle informatie op over de inhoud en evaluatie van het opleidingsonderdeel verpleegkundige vaardigheden. Dit was een meerwaarde omdat we op deze manier inzicht kregen in hoe de studenten voorbereid werden op het praktisch examen tijdens de lessen. Verder kregen we gedurende het onderzoek inzage in de toetsprocedures die de docenten gebruikten als evaluatie instrument tijdens het praktisch examen. Ten slotte werden de generieke beoordelingsschema's die gebruikt werden door Thomas More bij het opstellen van evaluatie instrumenten aangereikt.

Meer specifiek kregen we van Thomas More toegang tot Toledo, een online leerplatform. Op dit online leerplatform werden de toetsprocedures van elke verpleegkundige vaardigheid ter beschikking gesteld alsook de bestaande videofragmenten van de meeste verpleegkundige vaardigheden.

4 Producten - resultaten

In wat volgt worden de onderzoeksresultaten beschreven. Eerst wordt dieper ingegaan op de resultaten van de survey. Hierbij wordt een opsplitsing gemaakt tussen de manier waarop de schalen werden samengesteld en de beschrijving van de uitkomsten. Vervolgens worden de observatie van het praktisch examen en de resultaten van de focusgroep besproken. Tot slot volgen de bevindingen over de documentanalyse.

4.1 Rapportage onderzoeksresultaten survey

4.1.1 Samenstelling schalen

Voor het uitvoeren van de EFA werd eerst een nieuwe dataset gemaakt met alle items die volgens de theorie feedback, transparantie en stress bevroegen. Er werd een correlatiematrix aangemaakt met de werkelijke observaties en aan de hand hiervan werd een screeplot opgevraagd. De screeplot gaf een breuk aan tussen de factoren vier en vijf (hetgeen wees op het bestaan van vier factoren). Daarnaast werden ook de eigenwaarden opgevraagd. Vier van deze factoren hadden een eigenwaarde hoger dan één. Er werden dus vier factoren weerhouden voor het uitvoeren van de EFA.

Vervolgens werd de EFA uitgevoerd. Na rotatie verklaren de vier factoren samen 47% van de variantie. Echter weerhouden we een oplossing met drie factoren omdat de vierde factor slechts drie variabelen bevat waarvan de lading groter is dan 0.40 en deze moeilijk onder te brengen zijn onder één van de concepten uit de theorie.

De eerste factor beschrijft de latente variabele feedback. Hiertoe behoren de items Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10. De items die in theorie ook feedback bevroegen, nl. Fb2, Fb5, Fb6, werden niet opgenomen omdat hun factorladingen kleiner waren dan 0.40 (resp. -0.11; 0.13; 0.21). De betrouwbaarheid van deze schaal bedraagt 0.9, wat wijst op een goede interne consistentie.

De tweede gevonden factor geeft de latente variabele transparantie weer. Deze schaal bestaat uit de items T2, T3, T5, T6, Fb5, Fb6, Fb9. De betrouwbaarheid van deze schaal bedraagt 0.84. Wat wijst op een goede consistentie.

De derde factor beschrijft de latente variabele stress. Tot deze schaal behoren de items S2, S3, S4, S5, S6, S8. De andere items die stress bevroegen, nl. S1, S7 en S9, werden niet opgenomen. De factorladingen van deze items waren kleiner dan 0.40 (resp. 0.02; -0.13; -0.37). De betrouwbaarheid van de schaal 'stress' was 0.87. Dit betekent dat ook deze schaal consistent is.

De factoren die uit de EFA naar voor kwamen werden als vertrekpunt genomen voor het uitvoeren van een CFA. Dit model blijkt echter niet bij de data te passen (CFI=0,898 en RMSEA=0,086). Vervolgens werd naar de data gekeken om het model verder te verbeteren. Dit gebeurde met behulp van de modification indices. Zo werden verschillende errorcovarianties opgenomen in het model totdat de fit-indexes aanvaardbare waardes bereikt hadden. Bijlage 3 geeft een overzicht van de geteste modellen met bijbehorende fit indices.

Voor we de deelvragen afzonderlijk beantwoorden geven we de verbanden tussen de variabelen weer. In tabel 4 zijn de correlaties aangegeven tussen de verschillende variabelen.

Tabel 4: correlatiematrix voor de variabelen transparantie, feedback en stress

| Variabele | Transparantie | Feedback | Stress |
|------------------|---------------|--------------|-----------|
| Transparantie | 1,00 | 3,104084e-05 | 0,3635365 |
| Feedback | 0,6022222 | 1,00 | 0,1332127 |
| Stress | -0,1456524 | -0,2384799 | 1.00 |

De correlaties in de tabel zijn voor een gestandaardiseerd model met $M=0$ en $SD=1$.

De correlatie tussen variabelen transparantie en stress bedraagt -0.15 . Er is dus een zwak negatief verband tussen beide variabelen. We kunnen daarom stellen dat hoe meer transparantie er is, hoe minder stress er ervaren wordt. De correlatie tussen de variabelen feedback en stress bedraagt -0.24 . Dit is eveneens een zwak negatief verband tussen deze variabelen. Hieruit kunnen we afleiden dat hoe hoger de ervaren feedback is, hoe minder stress studenten ervaren. Echter moeten we wel rekening houden met de significantietoets. Beide verbanden zijn niet statistisch significant aangezien de p -waarde de kritische waarde $0,05$ overschrijdt. De kans is dus groter dan 5% dat er in de populatie geen samenhang is tussen deze variabelen. Maar zoals bij alle significantietoetsen is deze toets, ook hier, sterk afhankelijk van het aantal observaties (De Maeyer et al., 2012). Aangezien de analyses uitgevoerd zijn op een beperkte set van complete data, moeten we bijgevolg voorzichtig zijn met het interpreteren van de resultaten.

4.1.2 Beantwoorden onderzoeksvragen

4.1.2.1 Beschrijven uitkomsten

Binnen dit onderdeel beschrijven we in hoeverre de studenten het praktisch examen verpleegkundige vaardigheden als transparant ervaren (OV1a) en hoe ze feedback in de huidige vorm ervaren (OV2a). De uitkomsten zijn te lezen in tabel 5. Hieruit blijkt dat de studenten eerder akkoord zijn met de mate van transparantie ($M=3,97$; $SD=0,58$). Feedback ($M=3,59$; $SD=0,75$) daarentegen wordt door de studenten tussen de schalen neutraal en akkoord gescoord. De studenten beschrijven het praktisch examen verpleegkundige vaardigheden dus als (voldoende) transparant. Verder geven zij aan dat de feedback die zij ontvangen matig is en bijgevolg voor verbetering vatbaar is.

Tabel 5: beschrijving uitkomsten studenten verpleegkunde

| Variabele | <i>M</i> | <i>SD</i> |
|------------------|----------|-----------|
| Transparantie | 3,97 | 0,58 |
| Feedback | 3,59 | 0,75 |
| Stress | 3,53 | 0,82 |

1 = helemaal niet akkoord; 2 = niet akkoord; 3 = neutraal (akkoord noch niet akkoord); 4 = akkoord; 5 = helemaal akkoord.

4.1.2.2 Invloed van transparantie op stress

Onderzoeksvraag 1b gaat na welke invloed transparantie op stress heeft. De invloed van transparantie op stress is geanalyseerd via een bivariate regressieanalyse. Zo kan worden nagegaan of een stijging van de ervaren transparantie gepaard gaat met een daling van stress.

Uit deze analyse blijkt dat 2,12% ($R^2=0,02121$) van verschillen in stress die studenten ervaren toe te schrijven zijn aan verschillen in transparantie. Het gaat dus om een klein effect. Op basis van de F-test ($F(1,39) = 0,8453$, $p>0,05$) leren we bovendien dat de mate waarin studenten het examen als transparant ervaren enkel een deel van de variantie in stress verklaart in de steekproef. In dit specifieke geval is dat niet verwonderlijk gegeven de beperkte dataset waar we gebruik van maken (De Maeyer et al., 2012). Een student met een gemiddelde score op de schaal 'Transparantie', zou $-2,403e^{-16}$ standaardafwijkingen lager scoren op de schaal 'Stress'. Per standaardafwijking dat een student hoger scoort op transparantie, daalt stress met $-0,1457$ standaardafwijkingen. Zowel het intercept als de hellingsgraad zijn niet statistisch significant ($p>0,05$).

We kunnen er dus van uitgaan dat er een klein effect waarneembaar is binnen de steekproef. Hoe hoger de studenten het praktisch examen verpleegkundige vaardigheden als transparant scoren, hoe minder stress ze ervaren.

4.1.2.3 Invloed van feedback op stress

Onderzoeksvraag 2b gaat na welke invloed feedback op stress heeft. De invloed van feedback op stress is eveneens geanalyseerd via een bivariate regressieanalyse. Er werd nagegaan of een stijging van de ervaren feedback gepaard gaat met een daling van stress.

Uit deze analyse blijkt dat 5,69% ($R^2=0,05687$) van verschillen in stress die studenten ervaren toe te schrijven zijn aan verschillen in het ervaren van feedback. Het gaat dus wederom om een klein effect. Op basis van de F-test ($F(1,39) = 2,352$; $p>0,05$) leren we bovendien dat de mate waarin feedback ervaren wordt door de studenten enkel een deel van de variantie in stress verklaart in de steekproef. In dit specifieke geval is dat niet verwonderlijk gegeven de beperkte dataset waar we gebruik van maken (De Maeyer et al., 2012). Een student met een gemiddelde score op de schaal 'Feedback', zou $-2,403e^{-16}$ standaardafwijkingen lager scoren op de schaal 'Stress'. Per standaardafwijking dat een student hoger scoort op feedback, daalt de score van stress met $-0,2385$ standaardafwijkingen. Ook hier zijn zowel het intercept als de hellingsgraad zijn niet statistisch significant ($p>0,05$).

We kunnen er dus van uitgaan dat er een klein effect waarneembaar is binnen de steekproef. Hoe meer feedback de studenten ervaren, hoe minder last ze hebben van stress.

4.2 Rapportage onderzoeksresultaten observatie praktisch examen

Op 21 januari 2019 zijn drie leden van de projectgroep het praktisch examen verpleegkundige vaardigheden gaan observeren. De observatie werd enerzijds uitgevoerd om een beter zicht te krijgen op het verloop van het vaardigheidsexamen en anderzijds om meer inzicht te krijgen in de validiteit en betrouwbaarheid (OV3).

Organisatie en planning

Het praktisch examen ging door in de drie skillslabs die Thomas More Mechelen ter beschikking heeft. Deze drie lokalen liggen op eenzelfde gang waar ook het voorbereidingslokaal gelegen is. De studenten werden in groepjes van acht

verdeeld over de drie skillslabs. De planning voor het afleggen van het examen was op voorhand vastgelegd met de studenten. Het examen zelf nam +/- 25 minuten in beslag.

Vorbereiding

De studenten kwamen op het afgesproken tijdstip naar de gang waar de skillslabs zich bevinden. Op deze gang was een lokaal voorzien waar zij zich konden omkleden en hun spullen in een locker konden plaatsen. Hierna begaven de studenten zich naar een persoon die de aanwezigheden opnam. Bij deze persoon werden twee vaardigheden getrokken die de studenten moesten afleggen. Na het trekken van de vaardigheden kregen de studenten 10 minuten de tijd om in het voorbereidingslokaal het nodige materiaal te verzamelen, zodat ze hun vaardigheden correct konden uitvoeren.

Relatie student-docent

Wanneer de voorbereidingstijd verstreken was, kwamen de evaluatoren de studenten afhalen en werden ze meegenomen naar hun werkplaats in het skillslab. De docenten begroetten de studenten en er heerste een gemoedelijke sfeer. De relatie student-docent verliep bij de ene docent informeler dan bij de andere. In het skillslab legde de evaluator kort het verloop van het examen uit en werd er gevraagd aan de student of alles duidelijk was. Hierna begon de student met het afleggen van het examen. Tijdens het examen viel het echter op dat sommige evaluatoren bijkomstige vragen stelden aan de studenten of op eigen initiatief extra moeilijkheden toevoegden aan het praktisch examen, terwijl anderen dit niet deden. Zo kruiste een evaluator bijvoorbeeld zijn benen tijdens het nemen van de bloeddruk, wat de moeilijkheidsgraad verhoogde voor deze techniek. Bepaalde docenten lieten merken wanneer een student foutief bezig was, anderen lieten de studenten gewoon begaan.

Stress tijdens het examen

De meeste studenten waren zichtbaar gespannen en wisten eerst niet goed hoe te beginnen. Ze wisten niet of er verwacht werd dat ze hun handelingen uitlegden of dat ze tegen de patiënt (een pop of een docent, afhankelijk van de vaardigheid) moesten praten. De evaluator greep in en stuurde bij wanneer hij/zij merkte dat de student niet goed wist wat er verwacht werd. De onderzoekers merkten op dat

sommige studenten moeilijk uit hun woorden kwamen en minder vlot met het materiaal uit de voeten konden dan anderen. We merkten dat studenten moeite hadden met helder nadenken wanneer er eenvoudige vragen gesteld werden. Bijvoorbeeld op de vraag wat ze zouden doen wanneer ze iemand zagen liggen op de grond in een ziekenhuiskamer en vaststelden dat er geen hartslag meer was, antwoordden meerdere studenten dat ze een ziekenwagen zouden bellen.

Na het examen

Na het examen kregen de studenten nog vijf minuten de tijd om het gebruikte materiaal op te ruimen. Het opruimen had geen invloed op de evaluatie. Aan de lockers vroegen de studenten aan elkaar hoe het was geweest. Studenten die het examen nog moesten afleggen stelden vragen aan medestudenten, die reeds het examen hadden afgelegd, over het verloop van het examen en welke vaardigheden er geëvalueerd werden. Het was opvallend dat sommige studenten na afloop van het examen eerder het gevoel hadden dat ze een heel goed examen hadden afgelegd en dat hun slaagkansen eerder gematigd zijn. Hier is dus een belangrijke discrepantie. Tabel 1 geeft aan dat er slechts 39 geslaagden waren voor het praktisch examen en 25 niet geslaagden ondanks dat de meerderheid van de studenten aangaf dat ze een goed gevoel hadden na het afleggen van het examen. In het totaal waren er 84 studenten die het vak verpleegkundige vaardigheden hebben opgenomen. De overige 20 studenten hebben niet deelgenomen aan het examen.

Evaluatie van het examen

Wanneer de studenten het skillslab verlaten hadden, werden de beoordelingsdocumenten ingevuld. Evaluatoren die zelf geen beslissing konden nemen over het al dan niet slagen van hun student overlegden met Lien Proost. Verschillende evaluatoren gaven aan dat ze zich er bewust van waren dat studenten bij een andere evaluator mogelijk een ander resultaat zouden hebben behaald en dat dit in principe niet zou mogen. Ze vonden het moeilijk om een concreet cijfer op een uitgevoerde vaardigheid te plakken. Soms voerde een student de verpleegkundige vaardigheid grotendeels correct uit, maar vergat hij of zij een cruciale handeling, zoals bijvoorbeeld het wisselen van steriele handschoenen. Voor docenten was het dan moeilijk om te bepalen of een student geslaagd was of niet. Bovendien zou dit bij de bekendmaking van de resultaten

voor discussie kunnen zorgen: de docent zou kunnen zeggen dat de student iets niet deed, maar de student zou het omgekeerde kunnen beweren. Enkele docenten wezen ons hier op de mogelijkheid van het gebruik van video-opnames. Zo zouden ze kunnen bewijzen wat er op het examen gebeurde en zou bovendien gericht feedback gegeven kunnen worden.

4.3 Rapportage onderzoeksresultaten documentanalyse

Door een analyse te maken van de verschillende documenten die aangeboden werden door Thomas More Mechelen en het digitaal leerplatform Toledo, trachtten we tot inzicht te komen om OV3 te kunnen beantwoorden.

4.3.1 Generieke beoordelingsschema

Het generieke beoordelingsschema (bijlage 4) is een document waarin staat beschreven wat studenten moeten presteren tijdens het vaardigheidsexamen om een bepaalde score te behalen. Binnen Thomas More Mechelen wordt dit type document standaard onder alle - nieuwe - docenten verspreid overheen de richtingen. Alle startende docenten krijgen dit document zodat ze een ijkpunt hebben bij het vormgeven van hun lessen en evaluaties. Dit type document heeft dus tot doel duidelijkheid te scheppen en de transparantie te vergroten zowel voor de docent als voor de student.

4.3.2 Toetsprocedures

De toetsprocedure (bijlage 5) is een document dat aangereikt wordt aan de studenten via het digitaal leerplatform Toledo. De toetsprocedure is een stappenplan, per vaardigheid, dat ontwikkeld is door de docenten. De toetsprocedure bestaat uit drie onderdelen: de inleiding van de zorg, de uitvoering van de zorg en de nazorg. De stappen die beschreven worden binnen het onderdeel 'uitvoering', staan niet altijd in chronologische volgorde. De toetsprocedure is niet concreet omschreven aan de hand van criteria, maar eerder algemeen geformuleerd. Docenten gebruikten deze toetsprocedures als beoordelingsinstrument om het praktisch examen te evalueren. Er werd geen gebruik gemaakt van concrete criterialijsten, checklists of rubrics tijdens de beoordeling.

4.3.3 Videofragmenten

Voor de meeste vaardigheden zijn er videofragmenten beschikbaar via het digitaal leerplatform Toledo. In deze videofragmenten wordt telkens één vaardigheid gedemonstreerd die de student te allen tijde kan raadplegen. Elke video is zonder geluid opgenomen. Concreet ziet de student dus enkel de demonstratie van de vaardigheid in beeld, maar wordt er geen mondelinge of schriftelijke toelichting gegeven bij de handelingen.

4.4 Rapportage onderzoeksresultaten focusgroep

Met het focusgroepgesprek trachten we tot inzichten te komen om mede OV3 te beantwoorden en na te gaan in hoeverre de docenten verbonden aan het examen bij Thomas More problemen ervaren in verband met de betrouwbaarheid en de validiteit en dus de kwaliteit van het praktisch examen verpleegkundige vaardigheden. De rapportage volgt hiervoor het kader van Baartman (2008) met de vier criteria waarop wij dieper zijn ingegaan met dit project.

De focusgroep richtte zich op de verdere exploratie van de problematiek die ervaren werd bij de organisatie van het praktisch examen. Het doel van de bijeenkomst was om te komen tot een overzicht van de ervaren knelpunten en welke bijdrage bepaalde aanbevelingen in de toekomst kunnen leveren bij het wegwerken ervan. De rol van de examinerator tijdens het assessment maakten we via de focusgroep bespreekbaar. Gedurende de gehele focusgroep merkten we op dat de sfeer niet aanvallend of verwijtend aanvoelde. De voltallige groep was oprecht geïnteresseerd naar een optimale inrichting van het praktisch examen. Hierbij werd ieders mening gerespecteerd. Er kwam duidelijk een verschil in opinies naar voor, maar dit op een respectvolle manier naar elkaar toe. We merkten op dat de meningen van de docenten niet beïnvloed werden door bijvoorbeeld groepsdruk. Alle docenten voelden zich vrij om hun mening uit te drukken en aanvaardden dat er verschillende inzichten leefden in de groep.

Wanneer we op het einde van de focusgroep de resultaten uit de survey, en dus de bevindingen van de studenten, deelden met de docenten was het opvallend dat deze resultaten niet overeenstemden met de verwachtingen die de docenten daarrond hadden gevormd. De docenten reageerden hierbij verrast.

Vergelijkbaarheid

Voor elke te kennen techniek bestaat er binnen Thomas More Mechelen een toetsprocedure die voor elke student toegankelijk is alvorens het praktisch examen plaatsvindt. Deze aanpak ondersteunt al gedeeltelijk de vergelijkbaarheid van het praktisch examen. Docenten gaven echter aan dat per techniek de focus waaraan men aandacht geeft, verschuift. Alsook dat men niet elk detail kan opnemen in de toetsprocedures. Naast de techniek op zich zijn er veel basisvaardigheden die bij elke handeling belangrijk zijn zoals communicatie met de patiënt, veiligheid, enz. Het is onmogelijk deze allemaal te verwerken in een toetsprocedure. Over het verleggen van de focus per techniek heerst er een consensus binnen de groep. De deelnemers stelden dat dit onvermijdelijk is binnen hun sector en dat dit bijgevolg nodig was. Ze spraken vanuit expertise dat bij bepaalde technieken bijvoorbeeld hygiëne belangrijker was dan bij andere technieken. Echter geraakte men het niet eens over de transparantie hieromtrent.

Quote 1: "Er zijn geen vooropgestelde criteria naar de studenten toe gecommuniceerd per techniek, die een breekpunt kunnen zijn."

Quote 2: "Ik heb ooit een steriele techniek gehad, een inspuiting geven. Die student had die inspuiting correct uitgevoerd maar die moest de patiënt draaien. Die pop viel uit het bed want de student had haar bedhekken niet omhoog gedaan. Zij was niet geslaagd. Die techniek was perfect maar de patiënt was wel uit het bed gevallen. Dat kan ik daar toch niet op voorhand inzetten dat dit gaat gebeuren? Let op dat de patiënt niet uit het bed valt."

Reproduceerbaarheid van beslissingen

Tijdens het focusgroepgesprek gaven veel docenten aan dat de inzet van meerdere examinatoren op een examenmoment hen zorgen baarde. Meerdere docenten trokken de haalbaarheid hiervan in twijfel en ze vreesden dat dit een onhaalbare werklast met zich mee zou brengen. Elke docent hanteert op dit ogenblik dezelfde toetsprocedure, toch gaven de docenten aan dat het praktisch examen per examiner anders kan verlopen. Als voorbeeld gaf een docent aan dat een examiner bij het bepalen van de bloeddruk de benen kruist. Hierbij verhoogde de moeilijkheidsgraad van het praktisch examen bij deze specifieke examiner.

Binnen de focusgroep werd de vraag gesteld of dit type aanpak de validiteit van de evaluatie naar andere studenten toe niet in het gedrang brengt. Niet elke examiner verhoogde op eigen initiatief de moeilijkheidsgraad of stelde bijkomende vragen aan de student. Iedere docent had dan ook zijn of haar werkwijze bij het afnemen van het examen. Sommigen maakten geen notities tijdens het examen uit vrees om details tijdens de handeling over het hoofd te zien. Andere docenten gaven dan weer aan dat als ze het niet opschrijven, ze vrezen details te vergeten. Tijdens het onderzoek merkten wij een duidelijk knelpunt in de interbeoordelaarsbetrouwbaarheid.

Quote 3: "Ik ben er zeker van dat als een student elke keer bij iemand verschillend zijn examen doet, hij wel 5 of 6 keer een ander punt krijgt. De collega's beamen maar vullen aan met: maar met minimale verschillen."

Quote 4: "Ik was super streng en doorheen de tijd heb ik mij gewoon aangepast. Dat is eigenlijk raar, dat zou eigenlijk niet mogen. Ik heb mij, misschien is dat foutief gedrag van mij, aangepast aan de studenten hier en ben milder geworden. Omdat er geen één student geslaagd was."

Transparantie

Tijdens de observatie van het praktisch examen merkten wij een duidelijk knelpunt in de transparantie. De meningen waren verdeeld tijdens de focusgroep. Velen wilden meer transparantie, andere groepsleden vreesden dat hoe transparanter men gaat werken, hoe gericht de studenten gaan studeren. De groep stelde zich de vraag: "of een verpleegkundige die zijn techniek vanbuiten leert een goede verpleegkundige is?". Dit bracht de groep tot het knelpunt in verband met generieke beoordelingsschema's. Er zijn tot nu toe geen vooropgestelde criteria beschikbaar voor studenten omtrent bepaalde aspecten van de uit te voeren techniek, die bijvoorbeeld breekpunten konden zijn voor het slagen of niet slagen. Deze breekpunten kunnen verschillen per techniek. De docenten vrezen dat als ze hierin de transparantie verhogen, de studenten gericht gaan studeren. Indien een verpleegkundige enkel de cruciale handelingen kan uitvoeren maar zich niet focust op de kleine dingen, leiden we dan goede verpleegkundige studenten op? Tijdens het onderzoek komt naar voor dat niet elke docent deze generieke beoordelingsschema's daadwerkelijk gebruikte tijdens het examen. Enerzijds

omdat sommigen zich het bestaan ervan niet onmiddellijk konden herinneren, anderzijds omdat men soms aangaf meer belang te hechten aan overleg met collega's om tot een examenresultaat te komen. Nochtans werd er tijdens het focusgroepgesprek aangegeven dat de docenten wel akkoord waren met de inhoud van deze beoordelingsschema's.

Quote 5: "Die generieke (generieke beoordelingsschema's, cfr) die leven hier wel maar dat is niet iets dat standaard opgelegd wordt in ieder evaluatiemoment dat wij hier hebben."

Quote 6: "Er is voor elke techniek een toetsprocedure waarin alle stappen staan die de studenten moet doen. Zowel bij de voorbereiding, tijdens de zorg als de nazorg."

Betekenisvolheid

De nadruk in de gehele groep lag sterk op het opleiden van kwaliteitsvolle verpleegkundigen. Docenten gaven aan dat ze historisch gezien al wel enkele maatregelen geïmplementeerd hebben, maar dat het moeilijk is om nieuwe ideeën in te voeren. Veel docenten gaan uit van hun eigen opgebouwde expertise als maatstaf in hun handelen. Een punt waar dit erg duidelijk naar voor kwam was het al dan niet geven van meer gerichte feedback. Er werd een consensus bereikt dat feedback essentieel kon zijn om het leerrendement te verhogen. De meningen waren echter verdeeld over de manier waarop men feedback kon geven. Het voorstel om onmiddellijk feedback te geven na het praktisch examen werd niet door alle docenten als wenselijk beschouwd. De grootste bedenking hieromtrent was de extra werkbelasting dat dit met zich mee zou brengen. Hiernaast bieden de oefenmomenten in het skillslab een opportuniteit om feedback te geven. Docenten gaven hierbij aan dat er altijd een docent aanwezig was bij de verplichte skillslabs. Feedback werd gegeven als de studenten hierom vroegen of als de docent een fout bij het inoefenen opmerkte. Docenten verklaarden dat zij rondliepen en hierdoor niet alles konden opmerken. Ideeën omtrent coaching, peerfeedback van meer ervaren studenten en het opstellen van een groeiportfolio beschouwden de docenten wel als een positieve leermogelijkheid die zou kunnen bijdragen tot het zich eigen maken van de praktische vaardigheden. Bij het voorstel om proefexamens in te richten kregen we onmiddellijk weerstand omtrent

de praktische kant van dit alternatief. Er zijn te weinig uren en personeel om dit in te richten en ook de vrees voor extra werkdruk vormde hier wederom een belangrijk knelpunt. Wanneer studenten praktisch meer zouden oefenen, zou het stressniveau voor het examen significant dalen volgens de docenten.

Quote 7: "Wat willen we bereiken? We willen allereerst dat meer studenten slagen op de vaardigheidstoetsen en dat dus de techniek correct wordt uitgevoerd."

Quote 8: "Ik zie studenten die zeer nauwgezet de toetsprocedure hanteren (bij het oefenen in het skillslab, cfr) en elkaar daar wel in corrigeren."

5 Conclusies en discussie

5.1 Conclusie

Dit Interdisciplinair Project had drie doelstellingen die beantwoorden aan de probleemstelling die naar voor werd geschoven door de docenten van de opleiding verpleegkunde in Thomas More Mechelen.

Het eerste doel hield in dat we wilden beschrijven wat een assessment tot een goed assessment maakt. Dit deden we in de literatuurstudie aan hand van the wheel of competence van Baartman (2008).

Het tweede doel was gericht op de bestaande situatie in Thomas More. We wilden namelijk nagaan of het huidige competentie-assessment een goed assessment is aangaande betrouwbaarheid, validiteit, transparantie en feedback en hoe dit stress en/of uitval bij studenten kan verklaren. Om dit doel te bereiken stelden we drie onderzoeksvragen op die leidend waren bij ons onderzoek:

1. a. In hoeverre ervaren studenten het praktisch examen verpleegkundige vaardigheden als transparant?
 - 1.b. In welke mate heeft dit invloed op stress?
2. a. Hoe ervaren studenten feedback in de huidige vorm?
 - 2.b. In welke mate heeft dit invloed op stress?
3. In hoeverre ervaart Thomas More problemen in verband met de betrouwbaarheid en de validiteit, en dus de kwaliteit, van het praktisch examen verpleegkundige vaardigheden?

Om de eerste en tweede onderzoeksvraag en hun deelvragen te kunnen beantwoorden, namen we een survey af bij de studenten uit het eerste bachelorjaar verpleegkunde. We bevroegen zowel studenten die deelnamen aan het praktisch examen als studenten die niet deelnamen. Uit deze bevraging konden we afleiden dat de studenten een matige hoeveelheid stress ervaarden ($M=3,53$). Dit stressniveau is lager dan wordt beschreven in de literatuur. Volgens Akbaryboorang & Aminyazdi (2009) ervaren de meeste studenten verpleegkunde namelijk een hoge mate van angst tijdens het eindexamen.

Aan de hand van de antwoorden van de studenten konden we in kaart brengen in hoeverre de studenten het praktisch examen verpleegkundige vaardigheden als

transparant ervaren (OV1a). Uit de resultaten konden we afleiden dat deze het examen als voldoende transparant beschouwden ($M=3,97$). Dit impliceert dat de studenten het examen begrijpbaar vinden (Baartman, 2008). Echter is er nog ruimte om te verbeteren. De studenten gaven aan eerder akkoord te zijn met de mate van transparantie, maar ze waren niet helemaal akkoord. Verder onderzochten we in welke mate de ervaren transparantie een invloed had op stress (OV1b). We constateerden dat er een klein, negatief effect is van transparantie op stress binnen de steekproef. Naarmate de studenten het praktisch examen verpleegkundige vaardigheden hoger scoorden op transparantie, hoe minder stress ze ervaarden.

Daarnaast konden we ook aan de hand van de survey nagaan hoe de studenten de huidige feedback ervaren (OV2a). Volgens Baartman (2008) hebben leerders nood aan betekenisvolle feedback om hun leerproces te sturen. De resultaten van het onderzoek tonen aan dat de studenten de huidige vorm van feedback als matig ervaren ($M=3,59$). De nood van de leerders aan betekenisvolle feedback wordt dus slechts gedeeltelijk ingelost. Ook hier is er ruimte tot verbetering. De invloed hiervan op stress (OV2b) was klein in de steekproef. Bovendien ging het ook hier om een negatief effect: hoe hoger de studenten de ervaren feedback scoorden, hoe minder last ze hadden van stress.

Gezien een hogere mate van transparantie en feedback ervoor zorgen dat studenten minder stress ervaren, is het belangrijk om hierop in te zetten. Verschillende studies tonen immers een negatieve correlatie aan tussen stress voor een assessment en prestaties van studenten. Een hogere mate van stress beperkt de aandacht en concentratie van studenten, wat kan leiden tot mindere resultaten en uitval tijdens het assessment (Chapell et al., 2005; Iroegbu, 2013; Rezazadeh & Tavakoli, 2009; Trifoni & Shahini, 2011). Mogelijk zullen de examenresultaten dus verhogen wanneer de docenten van Thomas More inzetten op het verhogen van de transparantie en feedback aangaande het praktisch examen verpleegkundige vaardigheden.

Met de derde onderzoeksvraag wilden we nagaan in hoeverre Thomas More problemen ervaart in verband met de betrouwbaarheid en de validiteit en dus de kwaliteit van het praktisch examen verpleegkundige vaardigheden. Om hierover

uitspraken te kunnen doen, organiseerden we een focusgroep met de docenten, voerden we een documentanalyse uit en observeerden we het praktisch examen. Vermits betrouwbaarheid vervat zit in de kwaliteitscriteria 'vergelijkbaarheid' en 'reproduceerbaarheid van beslissingen' en validiteit verweven is in alle kwaliteitscriteria die in dit IP bestudeerd werden, wordt het antwoord op deze onderzoeksvraag hieronder besproken per onderzocht kwaliteitscriterium.

Vergelijkbaarheid

Uit onderzoek blijkt dat een CAP maar vergelijkbaar kan zijn als ze wordt uitgevoerd op een consistente en betrouwbare manier (Baartman, 2008). Daarnaast moeten ook de omstandigheden waarin het assessment wordt afgenomen zo veel mogelijk hetzelfde zijn voor alle leerders. De docenten in de focusgroep gaven aan dat het soms onvermijdelijk is om een andere nadruk te leggen op criteria afhankelijk van hun te beoordelen techniek. Dit komt omdat er meerdere factoren, zoals bijvoorbeeld de communicatie met de patiënt, bij komen kijken. Ze hanteren hierbij een toetsprocedure die verschilt per techniek, maar wel consistent is over alle beoordelingen heen. Dit komt overeen met de bevinding van Bennett (1993) die stelt dat tot het komen van voldoende vergelijkbaarheid, het niet noodzakelijk is om volledige standaardisatie te garanderen. Hierbij valt wel op te merken dat de docenten per techniek geen concrete criterialijsten hanteren tijdens de beoordeling om te bepalen of een student al dan niet slaagt. De toetsprocedures die voorhanden zijn hebben bovendien niet de vorm van een checklist of rubric om uit te maken wanneer een student slaagt voor het examen. De huidige aanpak kan er dan ook voor zorgen dat eenzelfde examen door verschillende examinatoren niet gelijk beoordeeld wordt. Uit de literatuur (Moorthy et al., 2003) blijkt dat dit type aanpak de objectiviteit van het evaluatiemoment in het gedrang brengt. Het resultaat van de evaluatie wordt op die manier immers een stuk afhankelijk van de examinerende docent. Voor studenten is het bijgevolg niet altijd duidelijk wanneer men geslaagd is en kan dit stress veroorzaken. Wass et al. (2001) stellen daarom dat het wel belangrijk is om te werken aan de hand van een goed gedefinieerd protocol om stress bij studenten te vermijden.

Uit de documentenanalyse konden we afleiden dat er op voldoende wijze wordt tegemoet gekomen aan de condities van vergelijkbaarheid die Messick (1994; 1995) beschrijft om zo tot externe validiteit te komen. We stelden vast dat de

toetsprocedures gehanteerd worden zowel bij het doceren en demonstreren van de technieken als bij het inoefenen in het skillslab en het eigenlijke examen. Ook de omstandigheden waarin het examen wordt afgenomen blijven identiek. Dit hebben we kunnen vaststellen door middel van onze observatie.

Reproduceerbaarheid van beslissingen

Baartman (2008) geeft aan dat wanneer docenten verschillende kwaliteitsaspecten gaan scoren naargelang hun belangrijkheid, de reproduceerbaarheid van beslissingen relatief laag scoort ten opzichte van de andere criteria. Een reden hiervoor zou kunnen zijn dat docenten gesteld zijn op hun autonomie en het niet gewoon zijn om samen met andere docenten studenten te beoordelen. Een tweede, eerder praktische reden die wordt aangehaald is dat gebruik van meerdere beoordelaars wordt gezien als niet haalbaar, te duur en bovendien tijdrovend. Nochtans is het inschakelen van meerdere beoordelaars een effectieve manier om aan reproduceerbaarheid te werken (Moss, 1994; Van der Vleuten & Schuwirth, 2005). Tijdens het onderzoek kwam naar voor dat docenten enkel in overleg gaan bij het beoordelen van studenten indien men twijfelt. Algemeen wordt de beoordeling door de docent dus individueel uitgevoerd. Alsook bevestigden docenten dat het praktisch moeilijk haalbaar is om meerdere beoordelaars in te schakelen. Zij vreesden dat dit een enorme stijging van de werklast met zich mee zou brengen. Uit de resultaten van het onderzoek merkten wij een duidelijk knelpunt in de interbeoordelaarsbetrouwbaarheid. Elke docent heeft namelijk zijn of haar persoonlijke werkwijze in het afnemen van het assessment, wat de reproduceerbaarheid van beslissingen in gevaar brengt. Bovendien wordt er geen gebruik gemaakt van verschillende meetmomenten. Baartman (2008) geeft aan dat het resultaat van een beoordeling in een CAP niet afhankelijk mag zijn van de specifieke evaluator of van de specifieke situatie waarin het assessment plaatsvindt.

Het gebruik van generieke beoordelingsschema's, om na te gaan wat studenten dienen te presteren om een bepaalde score te behalen is, in tegenstelling tot de toetsprocedures, een document dat voor alle docenten hetzelfde oogt. Hoewel niet alle docenten dit hanteren, zou dit de beoordelingen meer reproduceerbaar kunnen maken. Echter zagen we hier dat de gradaties tussen de verschillende scores eveneens voor interpretatie vatbaar zijn. Hiermee bedoelen we dat de generieke

beoordelingsschema's zijn opgedeeld in zes categorieën. Elke categorie staat voor een range van scores met per categorie deskundig uitgewerkte verwachtingen die de studenten moeten tonen op hun examen. Op deze manier kan je de student plaatsen in een categorie maar heb je als docent nog geen exacte score vastgelegd. Bijvoorbeeld categorie één bevat de scores 20,19,18 en 17/20. Wat bepaalt nu of de student binnen deze categorie een 19 of een 17 zal krijgen? Dit is afhankelijk van de docent die het assessment afneemt. En dat verhoogt de subjectiviteit en verlaagt de reproduceerbaarheid van beslissingen.

Transparantie

Volgens Baartman (2008) moeten enerzijds studenten de criteria kennen waarop ze gescoord zullen worden, wie de beoordelaars zullen zijn en wat het doel van het assessment is (Bagheri et al., 2012). De opleiding biedt zowel theoretische ondersteuning als praktische oefenmomenten aan. Echter zijn studenten niet altijd bekend met de docent die het assessment afneemt. Uit de resultaten van het onderzoek merkten wij een duidelijk knelpunt in de transparantie. Niet alle betrokkenen staan immers achter de assessmentcriteria en de manier waarop het assessment uitgevoerd wordt. Thomas More is door gebrek aan acceptatie dan ook op zoek naar alternatieven om de kwaliteit van het assessment te verhogen.

Transparantie is gelinkt aan validiteit (Baartman, 2008). Transparantie hangt samen met het structurele aspect van constructvaliditeit. Dit betekent dat de manier van scoren consistent moet zijn met wat belangrijk en minder belangrijk is bij de competentie. De manier van scoren bij Thomas More is op dit moment niet consistent. Praktisch is het voor de docenten moeilijk haalbaar om alle details uit te schrijven en hier gewichten aan toe te kennen. Bijvoorbeeld een patiënt uit bed laten vallen spreekt voor zich als een grove fout, desondanks een foutloos uitgevoerde techniek. De toetsprocedures zijn voor de studenten beschikbaar per techniek. Deze geven de belangrijkste criteria weer en worden gehanteerd tijdens het assessment. Echter is het niet duidelijk hoe zwaar de criteria doorwegen per techniek. Volgens Messick (1994) zouden deze criteria voor de leerder duidelijk moeten zijn en hoe zwaar ze doorwegen bij de beoordeling. Door de transparantie hierin te verhogen kan stress bij de studenten verlaagd worden. Wat betreft de generieke beoordelingsschema's lieten de resultaten zien dat niet alle docenten deze gebruiken om punten toe te kennen. De manier van scoren is bijgevolg

minder consistent, aangezien de schema's met vooropgestelde richtlijnen door sommige docenten niet worden gehanteerd.

Betekenisvolheid

Een assessment zou studenten moeten laten omgaan met betekenisvolle problemen die ervoor zorgen dat ze waardevolle ervaringen opdoen. In een opleiding verpleegkunde is het praktisch examen een waardevolle ervaring om mee te nemen in de praktijk. Volgens McDowell (1995) moet een CAP betekenisvol zijn en moet deze met andere woorden een significante waarde hebben voor zowel de docenten als de leerders, mede door de feedback die gegeven en verkregen wordt (Linn et al, 1991). Baartman (2008) voegt hieraan toe dat de leerders nood hebben aan betekenisvolle feedback en assessmentcriteria om hun leerproces te sturen. Een CAP ondersteunt het best het zelfsturend leren van de student. Het assessment bevat dan bij voorkeur specifieke methodes die het leren ondersteunen zoals het oefenen van zelfassessment en het geven en ontvangen van feedback. Het skillslab biedt mogelijkheden om feedback te verhogen. Docenten gaven aan dat het verhogen van feedback tijdens praktische oefenmomenten een manier kan zijn voor studenten om zich beter voorbereid te voelen voor het assessment. Hierbij stelden ze vast dat deze feedback vooral een meerwaarde zou kunnen betekenen bij het ontwikkelen van praktische vaardigheden. De feedback zou zich dan bijvoorbeeld kunnen uiten in de vorm van een peerassessment, groeiportfolio of coaching waarbij er vooral op een constructieve manier feedback gegeven wordt. De docenten gaven aan dat ze niet op een actieve manier feedback kunnen geven gedurende de oefenmomenten in het skillslab. Wanneer studenten regelmatig feedback ontvangen van hun docenten voelen ze zich meer gemotiveerd en leren ze effectiever werken (Rowbatham, 2010; Tiberius & Billson, 1991).

De docenten erkenden wel dat feedback het leerrendement kan verhogen, maar ze stellen het voorstel om feedback te verlenen meteen na het examen in vraag. Het doorslaggevend tegenargument is de grotere werkbelasting voor de docenten. Hiernaast werd ook de praktische kant van het invoeren van proefexamens in vraag gesteld, terwijl McKeachie & Svinicki (2006) beweren dat het gebruik van niet-beoordeelde assessments stress bij studenten kan verlagen. Zij stellen dat wanneer studenten dezelfde procedure voor het examen uitvoeren en hierbij

feedback krijgen van andere studenten of docenten, ze zich meer bewust kunnen zijn van hun mogelijkheden. Hierdoor kan mogelijk het zelfvertrouwen van de studenten verhoogd worden. Het mixen en trianguleren van verschillende evaluatiemethodes kan hierbij stressverlagend werken.

5.2 Discussie

Het onderzoek dat gevoerd werd door ons als IP-groep reikt inzichten aan over enkele aspecten van het praktisch examen van het vak verpleegkundige vaardigheden. Desondanks heeft het gevoerde onderzoek enkele beperkingen. Deze beperkingen vormen mogelijke invalshoeken voor toekomstig onderzoek.

5.2.1 Theoretisch kader

Voor de uitwerking van ons theoretisch kader baseerden we ons grotendeels op the wheel of competence van Baartman (2008). We onderzochten bij Thomas More echter maar vier kwaliteitscriteria die Baartman (2008) beschrijft. Dit is een beperking van ons onderzoek. Het is namelijk mogelijk dat ook andere criteria - al dan niet - voor verbetering vatbaar zijn. Indien blijkt dat onze aanbevelingen niet het verwachte effect zouden hebben, raden we Thomas More aan om verder onderzoek te doen naar de andere criteria uit '*the wheel of competence*'.

5.2.2 Survey

Een eerste belangrijke beperking van de uitgevoerde survey was het klein aantal respondenten. Hierdoor konden bepaalde analyses, zoals bijvoorbeeld een logistische regressie of een T-test, niet uitgevoerd worden. Dit had mogelijk gedeeltelijk te maken met de lengte van de survey. De bevraging was uitgebreid waardoor het veel tijd kostte om de survey helemaal in te vullen. Bovendien hebben zijn in de analyses niet de antwoorden op alle vragen meegenomen. We hebben ons beperkt tot de vragen die de latente concepten transparantie, feedback en stress bevroegen. Indien we ons voor de survey hadden beperkt tot de items die we zouden meenemen in verdere analyses hadden we wellicht meer respondenten kunnen overtuigen om onze survey in te vullen.

Een tweede beperking was dat we bij het afnemen van onze survey niet konden bevragen of de studenten geslaagd waren of niet. Op het moment dat we de survey afnamen, wisten de studenten immers hun resultaten nog niet. Daarnaast was de bevraging anoniem, dus konden de resultaten achteraf ook niet gekoppeld worden aan de antwoorden van de respondenten. Door het gebrek aan deze informatie konden we niet op zoek gaan naar verbanden tussen de kwaliteitscriteria, de resultaten en het ervaren stressniveau. Inzichten in deze verbanden zouden immers kunnen leiden tot gerichte acties. Dit vormt dus een mogelijke opportuniteit voor vervolgonderzoek.

Ten derde was het moment waarop we het stressniveau van de studenten bevroegen niet optimaal. We wilden met deze vragen peilen naar de hoeveelheid stress die ze ervaren voor en tijdens het examen. De resultaten van de afgenomen survey wijken echter af van wat we lezen in ander wetenschappelijk onderzoek: de studenten ervaren een matig stressniveau, terwijl in andere bronnen gesproken wordt van een hoog stressniveau (Akbari-boorang & Aminyazdi, 2009). Dit kan te wijten zijn aan het meetmoment, aangezien we tijdens de observatie van het praktisch examen duidelijk zagen dat de studenten erg gespannen waren. Om de validiteit van het onderzoek te vergroten hadden we twee meetmomenten kunnen voorzien. Tijdens het eerste meetmoment hadden we het stressniveau kunnen meten net voor de studenten het examen effectief zouden afleggen, het tweede moment zou dan na het praktisch examen plaatsvinden om te peilen naar de ervaren stress tijdens het examen.

5.2.3 Focusgroep

Voor het focusgroepgesprek hebben we ons enkel gericht op de docenten om een zicht te krijgen op de problemen die worden ervaren met betrekking tot de kwaliteit van het praktisch examen verpleegkundige vaardigheden. Hierbij wilden we ook peilen in welke mate de bevindingen van de studenten, die uit de vragenlijst naar boven kwamen, overeenstemden met de verwachtingen van de docenten. We hebben dit bijgevolg afgetoetst bij de docenten waarbij we ondervonden hebben dat hun verwachtingen niet aansloten bij de percepties van de studenten. Bijgevolg kunnen we het als een gemiste kans beschouwen dat we geen studenten betrokken hebben bij het focusgroepgesprek om interactie uit te lokken tussen beide partijen, om zo hopelijk tot nieuwe inzichten te komen. Een

mogelijk alternatief zouden twee verschillende focusgroepen kunnen zijn geweest, waarbij enerzijds de docenten samengenomen werden in een groep (zoals in de huidige vorm) en anderzijds enkele studenten. Om de beschreven fenomenen meer in de diepte te onderzoeken zou het bevragen van de studenten in een focusgroepgesprek, na het afnemen van de survey, een meerwaarde kunnen vormen voor dit onderzoek.

Aangezien meerdere docenten een verschillende kijk hadden op het praktisch examen en ook telkens instaan voor het afnemen van een andere techniek op het examenmoment, hadden we ook aan de hand van interviews meer diepgaand tot verschillende inzichten kunnen komen. Vervolgens zouden we de interviews met elkaar kunnen vergelijken om verschillen in de antwoorden meer te benadrukken. Daarbij werd er regelmatig door elkaar gesproken waardoor niet alle respondenten volledig hun eigen mening konden uiten. Bijgevolg werd het moeilijk om de opname letterlijk te transcriberen. Deze beperkingen hadden we met interviews wel kunnen opvangen.

6 Aanbevelingen

Dit laatste hoofdstuk vormt het praktische luik van het onderzoek. De derde doelstelling van dit IP was om aanbevelingen te doen om het huidige assessment te verbeteren. Hieronder bespreken we enkele aanbevelingen gebaseerd op de onderzochte elementen uit het kader van Baartman (2008).

6.1 Vergelijkbaarheid

Gebruik goed gedefinieerde protocols voor de verschillende vaardigheden om de vergelijkbaarheid te verhogen.

De vergelijkbaarheid versterken van het praktisch examen verdient ons inziens de nodige aandacht. Wass et al. (2001) concluderen dat het gebruik van uitgewerkte protocollen de vergelijkbaarheid vergroot. Volgens de onderzoekers is het noodzakelijk dat in de eerste plaats de elementen die manipuleerbaar zijn in de omgeving voor alle studenten gelijk gemaakt worden. Als de docenten de mogelijkheid zien om voor alle te examineren vaardigheden een basissituatie uit te werken - met andere woorden een goed gedefinieerd protocol - dat consequent wordt gehanteerd (bv. de benen van de patiënt wel of niet kruisen bij een behandeling), komt dit de vergelijkbaarheid ten goede. De invloed van de examinerator wordt zo teniet gedaan. Ook voor de student kan dit stressverlagend werken omdat ze allemaal met dezelfde beginsituatie het examen aanvatten.

Om de studenten te beoordelen gebruiken de docenten toetsprocedures. Deze toetsprocedures zijn niet neergeschreven onder de vorm van een checklist of rubric om uit te maken wanneer een student slaagt voor het examen. Wanneer de toetsprocedures omgevormd zouden worden tot een checklist of rubric, zou dit de vergelijkbaarheid verhogen en eveneens de betrouwbaarheid ten goede komen (Masterproef Westenberg-Wiersum, Open Universiteit Nederland, Nederland). Een rubric kan op verschillende momenten in het curriculum worden ingezet en geeft weer wat er geleerd moet worden en hoe dat wordt beoordeeld (Wessel, 2018). Een rubric kan een oplossing bieden voor het organiseren van peerfeedback, om feedback te geven of een beoordeling te formuleren. Het docentenplatform

'Vernieuwonderwijs'⁵ biedt een tool 'Rubric voor Rubrics'⁶ aan om na te gaan hoe rubrics het best worden opgesteld. Deze tool biedt hulp om een (bestaande) rubric te beoordelen op onder andere validiteit en inzet (bijlage 6).

6.2 Reproduceerbaarheid van beslissingen

Verhoog de reproduceerbaarheid van beslissingen door gebruik te maken van comparatieve beoordelingen met meerdere beoordelaars.

De docenten gaven tijdens de focusgroep aan dat ze enkel in overleg gaan bij het beoordelen van studenten indien ze twijfelen, over het algemeen wordt de student dus door één docent beoordeeld. Comparatief beoordelen kan hier een oplossing bieden.

Het fundamenteel principe van beoordelen steunt op het maken van vergelijkingen. De vergelijking gebeurt tussen twee objecten, prestaties of tussen een prestatie en beoordelingscriteria (Tarricone & Newhouse, 2016). Comparatief beoordelen is een methode waarbij op basis van specifieke taakcriteria conclusies worden getrokken over de prestaties van een student in vergelijking met die van een andere. Mensen zijn veel betrouwbaarder wanneer ze een prestatie vergelijken met een andere prestatie, dan wanneer ze een absoluut oordeel vellen (Jones & Alcock, 2012). Pollit (2012) voegt hieraan toe dat comparatief beoordelen extreem betrouwbare scores oplevert en het beoordelaars helpt in de hogere lagen van competentiemeting (cfr. figuur 2: Miller's pyramid of assessment, p. 14) (D-PAC, 2019), waardoor competenties kwaliteitsvol beoordeeld kunnen worden. Bovendien zorgt dit ervoor dat de individuele beoordeling van een beoordelaar overwonnen wordt omdat het beoordelingsproces gericht is op de waarde van elke prestatie en niet de individuele beoordelingsstandaard van een docent (Bramley, 2007; Pollit, 2012). Zelfs wanneer beoordelaars verschillende beslissingen nemen over de kwaliteit van de prestatie, kan het proces van comparatief beoordelen, na meerdere beoordelingsrondes, de kwaliteit van de prestaties nauwkeurig rangschikken (Pollit, 2012).

⁵ <https://www.vernieuwonderwijs.nl/>

⁶ <https://www.vernieuwonderwijs.nl/rubric-voor-rubrics/>

Om in de praktijk aan de slag te gaan met comparatief beoordelen willen we de tool D-PAC in combinatie met video-opnames aanbevelen. D-PAC⁷ is namelijk een tool die het comparatief beoordelen kan faciliteren. Men hoeft de verschillende werken niet meer te vergelijken met abstracte kwaliteitscriteria in bijvoorbeeld een rubric, maar direct met elkaar, in willekeurige samengestelde paren. Bij elk paar hoeft men enkel aan te geven welk werk van hogere kwaliteit is en waarom (Bouwer, Goossens, Mortier, Lesterhuis & De Maeyer, 2018). Bovendien zorgt deze tool voor een goede betrouwbaarheid omdat het gebruik maakt van verschillende assessoren. In principe zijn minimaal twee beoordelaars nodig voor een betrouwbare rangorde, maar het laagst aantal tijdens de projectjaren van D-PAC waren vier beoordelaars (D-PAC team, persoonlijke communicatie, 14 december 2018).

Eveneens geeft het D-PAC team (persoonlijke communicatie, 14 december 2018) aan dat het aantal keer dat een product⁸ (tekst, prestatie, video-opname...) wordt geëvalueerd belangrijk is, hoe vaker producten beoordeeld worden in vergelijkingen, des te hoger de kans op een betrouwbare rangorde. Om een betrouwbaarheidsniveau te behalen van 0.70 (70%) zijn er 10 tot 14 vergelijkingen per prestatie nodig. Om een betrouwbaarheidsniveau van 0.90 (90%) te behalen zijn er 26 tot 37 vergelijkingen per prestatie nodig (Verhavert, Bouwer, Donche & De Maeyer, 2019). Thomas More kan dus door een relatief klein aantal vergelijkingen een relatief goed (70%) betrouwbaarheidsniveau behalen, bovendien zijn deze scores dan ook niet meer beoordelaarsafhankelijk.

⁷ *Digital Platform for the Assessment of Competences* (www.d-pac.be).

⁸ Via volgende weblink kan u cases terugvinden waarvoor D-PAC al gebruikt werd: <https://www.d-pac.be/cases/>

6.3 Transparantie

Ken gewichten toe aan de verschillende criteria waarop studenten beoordeeld worden en leg breekpunten vast om de transparantie te verhogen.

De studenten scoorden de transparantie van het praktisch examen als voldoende. We constateerden echter een klein negatief effect van de mate van transparantie die ervaren werd op stress. Wanneer de docenten het stressniveau van de studenten willen doen dalen, kunnen ze dus acties ondernemen om de transparantie van het examen te doen stijgen.

Maak gebruik van vooraf gekende evaluatielijsten om de transparantie te verhogen. In de eerste plaats kan de transparantie nog meer verhoogd worden door middel van het opstellen van gedetailleerde overzichten die aangeven welke criteria juist geëvalueerd worden en hoe zwaar deze doorwegen bij de beoordeling (Messick, 1994). Ook 'breekpunten' die ontegensprekelijk leiden tot een onvoldoende kunnen hierbij geïdentificeerd worden (bv. onhygiënisch werken). Deze gedetailleerde overzichten moeten eveneens gecommuniceerd worden naar de studenten (Dochy & Moerkerke, 1997). Onderzoek toont immers aan dat dit leerders en hun leerprocessen op een positieve manier beïnvloedt (Dochy & McDowel, 1997; Jonsson et al., 2009). Hierbij willen we wel opmerken dat het gebruik van zulke vooraf gekende evaluatielijsten extra stress oplevert voor de studenten (Bagheri et al., 2012). Evaluatielijsten zijn vaak lege begrippen voor studenten. De studenten actief laten werken met de criteria zorgt ervoor dat ze de criteria beter gaan begrijpen en het verhoogt de transparantie. Onderzoek van Carless en Chan (2016) stelt dat, wanneer studenten voorbeelden analyseren, het hen kan helpen bij de kennismaking met academische maatstaven en hun vaardigheden in het maken van academische oordelen ondersteund kunnen worden. Om de studenten van Thomas More bijgevolg op een creatieve manier te laten kennismaken met de evaluatielijsten voorafgaand aan het praktisch examen, kan er aan de hand van onderstaande methode gewerkt worden aan de transparantie van de te kennen vaardigheden. Momenteel zijn er instructievideo's per techniek beschikbaar voor alle studenten. Echter bevatten deze bestaande video's geen bijhorende instructies in audiovorm. Het kan een meerwaarde bieden om deze video's met de hulp van de studenten

opnieuw vorm te geven. De transparantie in verband met de leerdoelen kan hierdoor verhoogd worden voor de studenten. Door co-creatie van nieuwe instructievideo's krijgen studenten een dieper inzicht in de verpleegkundige handelingen. Docenten geven aan dat verpleegkundige vaardigheden uniform maken niet eenvoudig is. Om de transparantie hier in te verhogen kan men meerdere filmpjes maken per techniek. Dit kan als voordeel meebrengen dat het voor studenten duidelijk wordt dat er meerdere goede aanpakken zijn. Sommige protocollen zijn misschien heel strikt, maar voor sommige zaken is het misschien niet erg om wat af te wijken. Dit wordt duidelijk als je meerdere goede (en misschien ook minder goede) voorbeelden laat zien. Deze manier van co-constructief werken biedt zowel voor de studenten als voor de docenten voordelen (Carless & Chan, 2016). Wanneer de studenten in dialoog gaan met elkaar voorziet dit kansen om hun oordelen te verklaren en te herzien. De dialoog met de docenten die hierop volgt, met betrekking tot de voorbeelden, biedt mogelijkheden voor de studenten om meer inzicht te krijgen in de impliciete kennis van de docenten over de normen en criteria die studenten moeten verwerven. Een tweede mogelijkheid betreft het uitvoeren van een focusgroep waarbij docenten aangevuld met studenten samen het assessment kunnen evalueren. Deze samenwerking verhoogt niet alleen de betrokkenheid van docenten bij het assessment, maar kan ook leiden tot meer transparantie bij studenten over de leerdoelen die zij moeten behalen (Sluijsmans, Brinke, & Schilt-Mol, 2015).

Creëer bewustwording van de toetscompetenties van docenten om de transparantie te verhogen.

Als laatste zou er ook gewerkt kunnen worden aan de toetsbekwaamheid van de docenten. Volgens Sluijsmans et al. (2012) wordt de kwaliteit van toetsing namelijk in grote mate bepaald door de kwaliteit van de beoordelaar. Hiermee willen we als IP-groep de competenties van de docenten niet in vraag stellen. We willen hen enkel bewust maken van het belang ervan. Een manier om dit te realiseren is door het organiseren van workshops waarin verschillende toetscompetenties besproken worden en waarin de docenten concreet aan deze competenties werken. De toetscompetenties die beschreven worden door Sluijsmans et al. (2012) kunnen hiervoor een kader bieden

(cfr. 2.3.2.3 transparantie, p. 25). Toetscompetenties die kunnen samenhangen met de ervaren stress door studenten zijn de volgende:

- In welke mate zijn docenten in staat om toegankelijke en meetbare leerdoelen te formuleren?
- Welke strategieën worden gehanteerd om met de studenten te communiceren over de leerdoelen?
- Hoe worden deze leerdoelen bereikt?

Hierbij kan men zich meer specifiek richten naar docentengedrag dat invloed heeft op de uniformiteit van het examen. Door bijvoorbeeld een workshop 'hoe stel ik mij op tijdens een examen' te organiseren kan men in alle openheid discussiëren over hoe docentengedrag een impact kan hebben op de stressbeleving van studenten. Tijdens de workshop kunnen docenten zich bewust proberen te worden van hun eigen pedagogisch handelen tijdens het assessment. Men kan reflecteren over communicatie, lichaamstaal en expliciete handelingen tijdens het assessment. Deze reflectieve houding kan vervolgens benut worden om te komen tot een beeld over hoe zij in de ogen van de studenten overkomen. Men kan ervaringen delen van wat men als positief of negatief ervaart in kader van het docentengedrag. Hieruit kan men trachten gezamenlijk afspraken te maken waardoor het assessment meer uniform verloopt. Alsook kan dit een positieve impact hebben op de stressbeleving van studenten.

6.4 Betekenisvolheid

Stimuleer kwaliteitsvolle feedback om zo de betekenisvolheid te verhogen.

Bij het onderzoek naar het kwaliteitscriterium 'betekenisvolheid' focusten we ons op de betekenisvolheid van de feedback die de studenten kregen. Het is immers belangrijk dat bij summatieve assessments feedback gerelateerde criteria zoals betekenisvolheid opgenomen worden (Baartman, 2008). Deze feedback werd door de studenten als matig beoordeeld. Hieruit leiden we af dat de feedback die gegeven wordt betekenisvol is, maar dat er nog mogelijkheden zijn tot verbetering. Bovendien stelden we een klein negatief effect vast van feedback op stress waardoor het stressniveau van de studenten zal afnemen wanneer er gewerkt wordt aan feedback.

Maak gebruik van (online) tools en peerfeedback, om het geven en krijgen van kwaliteitsvolle feedback te stimuleren en dus de betekenisvolheid te verhogen. Niet enkel de docenten moeten aan de studenten feedback geven, ook studenten kunnen elkaar feedback geven (peerfeedback). Peerfeedback is duurzaam wanneer het de studenten ondersteunt in het zelf monitoren van hun eigen werk, onafhankelijk van de docent(en) (Carless, Salter, Yang & Lam, 2011). Duurzame feedback ondersteunt en informeert de studenten over de huidige taak en ontwikkelt tegelijkertijd het vermogen om zelf de prestaties van toekomstige taken te reguleren. Het geven van feedback aan elkaar helpt hen de verwachte criteria eigen te maken (cfr. 6.3 transparantie, p. 64), alsook krijgen andere studenten feedback. Dit levert een win-win situatie op. Het is belangrijk dat studenten hierin worden begeleid en dat het voor hen duidelijk is wat de leerdoelen, de beoordelingscriteria en de verwachte prestatie zijn (Arts & Jaspers, 2018). Om studenten hierin te ondersteunen kunnen docenten samen met studenten beoordeeld materiaal bestuderen, om zo te kijken of beoordelingscriteria consistent worden gehanteerd. Het gebruik van een voorbeeld kan een hulpmiddel zijn, belangrijk hierbij is dat de feedback en de beoordeling van het voorbeeld beschikbaar zijn (Handley & Williams, 2011) zodat studenten kunnen zien wat er concreet verwacht wordt. We bespreken enkele voorbeelden van (online) tools, namelijk Flipgrid en D-PAC. Deze (online) tools kunnen dit proces ondersteunen en zorgen ervoor dat docenten kunnen monitoren of de kwaliteit van de (peer)feedback en de prestaties van de studenten goed zijn.

De website www.flipgrid.com is een platform waarop, in afgesloten community's, filmpjes gedeeld kunnen worden. Bovendien kunnen de leden van de community er op eenvoudige wijze op elkaars filmpjes reageren. Wij zien hier interessante mogelijkheden in voor het vak 'verpleegkundige vaardigheden'. Zo kunnen de docenten de opdracht geven aan hun studenten om in de loop van het semester filmpjes te maken van de uitvoering van verschillende verpleegkundige vaardigheden en deze te posten op Flipgrid. De andere studenten kunnen deze dan via het platform bekijken en aan elkaar feedback geven. Ook docenten hebben zo de mogelijkheid om op een eenvoudige manier terug te koppelen naar de studenten over de door hen uitgevoerde vaardigheid. Daarnaast hebben de docenten een beter zicht op wie al dan niet oefent gedurende het semester.

Indien het geven van feedback door de studenten moeizaam verloopt en de docenten deze als minder betekenisvol ervaren, kan er op Toledo een leidraad gepost worden met tips die hen helpt bij het geven van feedback. Zo krijgen de studenten ook inzicht in de manier waarop ze bij het praktisch examen beoordeeld zullen worden. Wat op zijn beurt ook de transparantie ten goede komt. Een handvat dat hierbij gebruikt kan worden zijn de drie meest voorkomende types van feedback om leren te ondersteunen volgens Kirschner & Neelen (2018), namelijk correctieve (*corrective*), directieve (*directive*) en epistemische (*epistemic*) feedback.

Correctieve feedback geeft informatie aan de student of de taak die hij uitvoert goed of niet goed is, waarbij de feedback bovendien een correcte oplossing biedt over de mogelijke fout. Directieve feedback geeft de student informatie over hoe dingen beter kunnen gedaan worden, deze feedback geeft informatie over hoe de student iets moet doen zodat hij/zij de taak naar behoren en/of beter kan uitvoeren. Tot besluit stimuleert de epistemische feedback de student om na te denken over het 'waarom' in relatie tot het uitvoeren van de taak. Deze feedback geeft geen informatie over of het louter juist of fout is (correctieve feedback), noch hoe ze het beter zouden kunnen doen (directieve feedback), maar geeft de studenten een tip of stelt een vraag aan de student om zo hun zelf te laten ontdekken hoe ze het anders of beter zouden kunnen doen. Belangrijk hierbij is dat epistemische feedback het meest effectief is, maar meer tijd in beslag neemt om te geven, gevolgd door directieve feedback (Guasch, Esposa, Alvarez & Kirschner, 2013).

Tot slot is D-PAC eveneens een tool die ingezet kan worden om de studenten kwaliteitsvolle feedback te laten geven aan elkaar (Bouwer et al., 2018). Ze kunnen op deze manier na elke vergelijking feedback geven op de plus- en minpunten van elk product. Wanneer een reeks paarsgewijze vergelijkingen in D-PAC voltooid zijn, ontvangen de studenten feedback over hun eigen werk. D-PAC hanteert twee verschillende vormen van feedback namelijk feedback over het huidige prestatieniveau door een positie in de rangorde (zo krijgt de student inzicht in het niveau van zijn/haar prestatie ten opzichte van die van de medestudenten en/of ten opzichte van het gewenste doel) en feedback over specifieke plus- en verbeterpunten (de student krijgt geschreven commentaar, dit wordt steeds anoniem weergegeven). Deze combinatie van feedbackvormen zijn cruciaal voor

het leren van feedback (Hattie & Timperley, 2007). Studenten dienen op deze manier zelf na te gaan welke feedback ze bruikbaar vinden om te gebruiken in het verbeteren van hun werk. Belangrijk hierbij is dat de docent extra ondersteuning biedt aan de studenten bij het begrijpen en implementeren van de feedback. Docenten kunnen door gebruik te maken van D-PAC kostbare tijd besparen, die ze bijgevolg dan kunnen gebruiken voor het ondersteunen van studenten tijdens het geven en krijgen van feedback aan elkaar (Bouwer et al., 2018). Daarenboven is het waardevol dat docenten in D-PAC de kwaliteit van de peerfeedback kunnen monitoren, hierbij zijn volgens Bouwer et al. (2018, p. 103-104) vooral de volgende informatiebronnen van waarde:

1. De betrouwbaarheid van de rangorde:
dit geeft inzicht in hoeverre studenten het eens zijn over welk product van hogere kwaliteit is. Als de betrouwbaarheid laag is, betekent dit dat niet iedereen hetzelfde werk goed of slecht vindt. Met behulp van extra analyses is het mogelijk om er achter te komen welke studenten afwijken van de gemiddelde groepsconsensus. Zijn dit de studenten die zelf ook minder goed presteren? Hebben zij wellicht extra uitleg nodig?
2. De rangorde
deze geeft informatie over welke producten studenten beter of minder goed vinden. Komt dit overeen met de mening van docenten?
3. De specifieke feedback over plus- en minpunten:
dit geeft informatie over aspecten waar studenten tijdens het vergelijken op letten. De docent kan deze aspecten vergelijken met de vooropgestelde kwaliteitscriteria: komen deze overeen, of zijn er aspecten waar studenten nog onvoldoende naar kijken? Ook kan de docent de gegeven feedback gebruiken om een indruk te krijgen van wat studenten volgens henzelf al wel goed doen (pluspunten) en wat nog beter kan (verbeterpunten).

4. De feedback gegroepeerd per student:
door na te gaan welke student welke feedback heeft ontvangen en gegeven, krijgt de docent inzicht in het prestatieniveau van individuele studenten en de aspecten waar zij op letten tijdens het vergelijken van elkaar werk. Deze informatie kunnen docenten gebruiken om individuele studenten extra uitleg of begeleiding te geven.

6.5 Vijf punten plan

1. Gebruik goed gedefinieerde protocols voor de verschillende vaardigheden om de vergelijkbaarheid te verhogen.
2. Verhoog de reproduceerbaarheid van beslissingen door gebruik te maken van comparatieve beoordelingen met meerdere beoordelaars.
3. Ken gewichten toe aan de verschillende criteria waarop studenten beoordeeld worden en leg breekpunten vast om de transparantie te verhogen.
4. Creëer bewustwording van de toetscompetenties van docenten om de transparantie te verhogen.
5. Stimuleer kwaliteitsvolle feedback om zo de betekenisvolheid te verhogen.

How to CAP*?

*Competence Assessment Program

Vergelijkbaarheid

Gebruik goed gedefinieerde protocols voor de verschillende vaardigheden om de vergelijkbaarheid te verhogen.

Reproduceerbaarheid

Verhoog de reproduceerbaarheid van beslissingen door gebruik te maken van comparatieve beoordelingen met meerdere beoordelaars.



Betekenisvolheid

Stimuleer kwaliteitsvolle feedback om zo de betekenisvolheid te verhogen.

Transparantie

Creëer bewustwording van de toetscompetenties van docenten om de transparantie te verhogen.

Ken gewichten toe aan de verschillende criteria waarop studenten beoordeeld worden en leg breekpunten vast om de transparantie te verhogen.

7 Literatuur

Absolum, M., Flockton, L., Hattie, J., Hipkins R., & Reid, I. (2009) *Directions for Assessment in New Zealand (DANZ)*. New Zealand: Ministry of Education.

Akbary-boorang M. & Aminyazdi A. (2009). Test-Anxiety and Self-Efficacy. *Horizon of Medical Sciences Journal*, 2(15), 70-7.

Albino, J., Young, S., Neumann, L., Kramer, G., Andrieu, S., Hensen, L (2009) Assessing Dental Students' Competence: Best Practice Recommendations in the Performance Assessment Literature and Investigation of Current Practices in Predoctoral Dental Education. *Journal of dental education*, 72, 1405-1435.

Al-Wardy N. (2010). Assessment of methods in undergraduate medical education. *Sultan Qaboos University Medical Journal*, 10(2), 203-209.

Arts, J., & Jaspers, M. (2018). Feedback als integraal onderdeel van leren en opleiden. In D. Sluijsmans & M. Segers (Red.), *Toetsrevolutie: naar een feedbackcultuur in het hoger onderwijs* (p. 54-69). Culemborg, Nederland: Phronese.

Axinn, W., & Pearce, L. (2006). *Mixed Method Data Collection Strategies*. New York: Cambridge University Press.

Baartman, L. (2008). *Assessing the assessment: Development and use of quality criteria for Competence Assessment Programmes* (Doctoral dissertation, Universiteit Utrecht, Nederland). Geraadpleegd van <https://dspace.library.uu.nl/handle/1874/27155>

Bagheri, M., Forogeth, M.S., Fallah M.S. (2012) The Comparison of Stressors in the Assessment of Basic Clinical Skills with Traditional Method and OSCE in nursing Students. *Life Science Journal Acta Zhengzhou University Overseas Edition*, 9, 1748-1752.

Benett, Y. (1993). The validity and reliability of assessments and self-assessments of work-based learning. *Assessment & Evaluation in Higher Education*, 18, 83-95.

Bengtsson, M., Dahlquist, B., & Carlson, E. (2016). Expectations of nursing students prior to a skills-based exam performed in clinical practice. *Journal of nursing education and practice*, 6(8), 22-29.

Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher education*, 32, 347-364

Biggs, J. (1999). *Teaching for quality learning at university*. Buckingham, UK: SRHE and Open University Press

Birenbaum, M. (1996). Assessment 2000: Towards a pluralistic approach to assessment. In M. Birenbaum & F. J. R. C. Dochy (Eds.), *Alternatives in assessment of achievement, learning processes and prior knowledge* (pp. 3-29). Boston, MA: Kluwer Academic Publishers.

Billiet, J. & Waeye, H. (Eds.) (2001) *Een samenleving onderzocht: methoden van sociaal-wetenschappelijk onderzoek*. Antwerpen: Standaard Uitgeverij.

Bloor, M., Frankland, J., Thomas, M. & Robson, K. (2001) *Focus groups in social research*. London: Sage.

Bouwer, R., Goossens, M., Mortier, A.V., Lesterhuis M., & de Maeyer, S. (2018). Een comparatieve aanpak voor peer assessment: leren door te vergelijken. In D. Sluijsmans & M. Segers (Red.), *Toetsrevolutie: naar een feedbackcultuur in het hoger onderwijs* (p. 92-105). Culemborg, Nederland: Phronese.

Bramley, T. (2007). Paired comparison methods. In: P. Newton, J-A. Baird, H. Goldstein, H. Patrick, & P. Tymms (Eds.), *Techniques for monitoring the comparability of examination standards* (p. 264-294). QCA, London.

Carless, D. & Chan, K. K. H., (2016): Managing dialogic use of exemplars. *Assessment & Evaluation in Higher Education*, DOI: 10.1080/02602938.2016.1211246

Carless, D., Salter, D., Yang, M., & Lam. J. (2011). Developing sustainable feedback practices. *Studies in higher education*, 36(4), 395-407.

Chan, D. (2003). Validation of the Clinical Learning Environment Inventory. *Western Journal of Nursing Research*, 25(5), 519–532.

Chan, D. (2004). Nursing Students' Perceptions of Hospital Learning Environments - an Australian Perspective. *International Journal of Nursing Education Scholarship*, doi:10.2202/1548-923X.1002

Chapell, M.S., Blanding, Z.B., Silverstein, M.E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. *Journal Educational Psychology*, 97(2), 268–274.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd edition). New Jersey: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (zesde editie), New York:Routledge.

Cronbach, L. J., Linn, R. L., Brennan, R. L., & Haertel, E. H. (1997). Generalisability analysis for performance assessments of student achievement or school effectiveness. *Educational and Psychological Measurement*, 57, 373-399.

Dawood, E., Al Ghadeer, H., Mitsu, R., Almutary, N., & Alenezi, B. (2016) Relationship between Test Anxiety and Academic Achievement among Undergraduate Nursing Students. *Journal of Education and Practice*, 7-2, 57-65.

De Maeyer S., Coertjens, L., & Ardies, J. (2012). *Bivariate en multivariate statistiek met R. Een openleerpakket*. Gent: Academia Press.

De Maeyer, S., Donche, V., & van Daal, T. (2017). *Hoofdstuk 2: Observeren* [cursustekst], Universiteit Antwerpen, Antwerpen.

De Maeyer, S., Donche, V., & van Daal, T. (2017). *Hoofdstuk 4: Vragenlijsten* [cursustekst], Universiteit Antwerpen, Antwerpen.

De Maeyer, S., & van Daal, T. (2018). *Structurele vergelijkingsmodellen* [cursustekst], Universiteit Antwerpen, Antwerpen.

Dierick, S. & Dochy, F. J. R. C. (2001). New lines in edumetrics: new forms of assessment lead to new assessment criteria. *Studies in Educational Evaluation*, 27, 307-329.

Dochy, F. J. R. C., & McDowell, L. (1997). Introduction: Assessment as a tool for learning. *Studies in Educational Evaluation*, 23, 279-298.

Dochy, F., & Moerkerke, G. (1997). Assessment as a major influence on learning and instruction. *International journal of educational research*, 27(5), 415-432.

Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer- and co-assessment in higher education. *Studies in Higher Education*, 24, 331-350.

Donche, V. (2015). *Inleiding in methoden en technieken* [cursustekst]. Universiteit Antwerpen, Antwerpen.

D-PAC. [dpaccj]. (2019, 4 april). Waarom comparatief beoordelen? Omdat het beoordelaars helpt in de hogere lagen van competentiemeting [Tweet]. Geraadpleegd van <https://twitter.com/i/status/1128457922864762880>

Guasch, T., Esposa, A., Alvarez, I. M., & Kirschner, P. A. (2013). Effects of feedback on collaborative writing in an online learning environment: Type of feedback and the feedback-giver. *Distance Education, 34*, 324–338.

Haertel, E.H. (1991) New forms of teacher assessment. *Review of research in education, 17*, 3-29.

Handley, K., & Williams, L. (2011). From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education, 36*(1), 95-108.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Howitt, D., & Cramer, D. (2007). *Statistiek in de sociale wetenschappen* (3e editie). Amsterdam: Pearson Education.

Iroegbu, M.N. (2013). Effect of test anxiety, gender and perceived self-concept on academic performance of Nigerian students. *International Journal Psychology, 5*(7), 143–146.

Jones, I., & Alcock, L. (2012). Summative peer assessment of undergraduate calculus using adaptive comparative judgement. In: P. Iannone, & A. Simpson (Eds.). *Mapping university mathematics assessment practices* (p. 63-74). University of East Anglia, East Anglia.

Jonsson, A., Baartman, L.K.J., & Lennung, S. (2009). Estimating the quality of performance assessments: the case of an 'interactive examination' for teacher competencies. *Learning environ res, 12*, 225-241.

Kabinet van Jo Vandeuren, Vlaams minister van Welzijn, Volksgezondheid en Gezin. (2015). *Werk maken van werk in de zorg- en welzijnssector. Actieplan 3.0*. Geraadpleegd op 2 februari 2019, <https://www.vlaanderen.be/nl/publicaties/detail/werk-maken-van-werk-in-de-zorg-en-welzijnssector-actieplan-3-0>

Kabinet Vlaams minister van Onderwijs (2015) *Bacheloropleiding verpleegkunde wordt vierjarige opleiding vanaf academiejaar 2016-2017*. Geraadpleegd op 2 februari 2019, <https://onderwijs.vlaanderen.be/nl/bacheloropleiding-verpleegkunde-wordt-vierjarige-opleiding-vanaf-academiejaar-2016-2017>

Kirschner, P.A., & Neelen, M. (5 juni, 2018). *No feedback, no learning*.

Geraadpleegd op 13 mei 2019,

<https://3starlearningexperiences.wordpress.com/tag/effective-feedback/>

Lotz, C., Sparfeldt J.R. (2016). Does test anxiety increase as the exam draws naer?-Students' state test anxiety recorded over the course of one semester. *Personality and individual differences*, 104(2017), 397-400.

Linn, R. L., Baker, J., & Dunbar, S. B. (1991). Complex, performance based assessment: expectations and validation criteria. *Educational Researcher*, 20, 15-21.

Markman, U., Balik, C., Bercovitz, H., & Ehrenfeld, M. (2010) The Effect of Nursing Students' Health Beliefs on Their Willingness to Seek Treatment for Test Anxiety. *Journal of Nursing Education*, 50(5), 248-52

Marteau, T., & Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British Journal of Clinical Psychology*, 31(3), 301-306.

Martin, S. (1997). Two models of educational assessment: A response from initial teacher education: If the cap fits. *Assessment & Evaluation in Higher Education*, 22, 337-342.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Education and Training International*, 32, 302-313.

McKeachie WJ, Svinicki M. *McKeachie's teaching tips*. Boston: Houghton Mifflin company, 2006.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 23, 13-23.

Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50, 741-749.

Meusen-Beekman, K., Joosten-Ten Brinke, D., & Boshuizen H.P.A. (2016). Beter schrijven door self-en peer-assessment? De effecten van formatieve toetsvormen op de kwaliteit van schrijfproducten. *Levende talen tijdschrift*, 17(4), 15-25.

Miller G. The assessment of clinical skills/competence/performance. *Academic Medicine*. 1990; 65(Suppl 9): 63-67.

Moorthy, K., Munz, Y., Sarker, S. K., & Darzi, A. (2003). Objective assessment of technical skills in surgery. *British Medical Journal*, 327(7422), 1032-1037.

Mortelmans, D. (2013). *Handboek kwalitatieve onderzoeksmethoden*. Leuven: Acco.

Moss, P. M. (1994). Can there be validity without reliability? *Educational Research*, 23, 5-12.

Papathanasiou, I., Tsaras, K., & Sarafis, P. (2014). Views and perceptions of nursing students on their clinical learning environment: teaching and learning. *Nurse Education Today*, 34(1), 57-60.

Pollitt, A. (2012). Comparative judgement for assessment. *International Journal of Design Education*, 22, 157-170.

Ramani S, Leinster S. AMEE Guide no. 34: teaching in the clinical environment. *Medical Teacher*. 2008; 30(4): 347-364.

Rezazadeh, M., & Tavakoli, M. (2009). Investigating the relationship among test anxiety, gender, academic achievement and years of study: a case of Iranian EFL university students. *English Language Teaching, 2*(4), 68–74.

Rowbatham, M., (2010). Teacher perspectives and psychosocial climate of the classroom in a traditional BSN program. *International Journal of Nursing Education Scholarship 7*(1), 1–14.

Safari-Moradabadi, A., Kasmaei, P., Khalife Nilsaz, M., Moradi, M., Ahmady, F., Mohammadloo, A., et al. (2018). Evaluation of the learning environment based on the dundee ready education environment measure model from the perspective of primary school students in Roudsar city. *Journal of research in medical and dental science, 6*(1), 140-145.

Sadler, R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*, 191-209. Evaluation of the Learning Environment based on the Dundee Ready Education Environment Measure Model from the Perspective of Primary School Students in Roudsar City. *Journal of Research in Medical and Dental Science, 6*(1), 140-145.

Segers, M. (2004). Assessment en leren als tweede eenheid: onderzoek naar de impact van assessment op leren [the dyad of assessment and learning: a study of the impact of assessment on learning]. *Tijdschrift voor Hoger Onderwijs, 22*, 188-220.

Serena, P., & Anna, B. (2009). Italian nursing students' perception of their clinical learning environment as measured with the CLEI tool. *Nurse Education Today, 29* (8), 886–890.

Shivers, E., Hasson, F., & Slater, P. (2017). Pre-registration nursing student's quality of practice learning: Clinical learning environment inventory (actual) questionnaire. *Nurse Education Today, 5*, 58-64.

Sluijsmans, D., Peeters, A., Jakobs, L., & Weijzen, S. (2012). De kwaliteit van toetsing onder de loep. *Onderwijsinnovatie*, nummer 4, december 2012, 17-25. Geraadpleegd op 5 februari 2019 via <https://docentenportal-farmacie.sites.uu.nl/wp-content/uploads/sites/133/2017/11/DeKwaliteitVanToetsenOnderDeLoep.pdf>

Sluijsmans, D., Brinke, D.J., & Schilt-Mol, T., (2015). *Kwaliteit van toetsing onder de loep: Handvatten om de kwaliteit van toetsing in het hoger onderwijs te analyseren, verbeteren en borgen*.

Solvik, E., & Struksnes, S. (2018). Training Nursing Skills: A Quantitative Study of Nursing Students' Experiences before and after Clinical Practice. *Nursing Research and Practice*, 2018, 1-9.

Stokking, K., Van der Schaaf, M., Jaspers, J., & Erkens, G. (2004). Teachers' assessment of students' research skills. *British Educational Research Journal*, 30, 93-116.

Tarricone, P., & Newhouse, C.P. (2016). Using comparative judgement and online technologies in the assessment and measurement of creative performance and capability. *International journal of educational technology in higher education*, 13(16), 1-11.

Thomas More. (2018). *Missie en visie*. Geraadpleegd op 29 februari 2019 van https://www.thomasmore.be/sites/www.thomasmore.be/files/onderwijsvisie_thomas_more_infographic.pdf

Thomas More. (2018a). *Professionele bachelor Verpleegkunde*. Geraadpleegd op 29 februari 2019 van <https://www.thomasmore.be/opleidingen/professionele-bachelor/verpleegkunde>

Tiberius, R.G., Billson, J.M. (1991). The social context of teaching and learning. *New Directions for Teaching and Learning*, 45, 67-86.

Toetscyclus ontwerpen. (5 juli, 2017). *Hogeschool van Amsterdam*.

Geraadpleegd van

<https://score.hva.nl/docent/toetscyclus/Paginas/Ontwerpen.aspx>

Trifoni, A., & Shahini, M. (2011). How does exam anxiety affect the achievement of university students? *Mediterranean Journal of Social Sciences*, 2(2), 93–100.

Van der Vleuten, C., & Schuwirth, L. (2005) Assessing professional competence: from methods to programmes. *Medical Education*, 39 (3), 309-317.

Verhavert, S., Bouwer, R., Donche, V., & De Maeyer, S. (2019). A meta-analysis on the reliability of comparative judgement. *Assessment in education: principles, policy & practice*, DOI: 10.1080/0969594X.2019.1602027

Wass V, Van der Vleuten C, Shatzer J., & Jones, R. (2001) *Assessment of clinical competence*. *Lancet*. 357(9260): 945-949.

Westenburg-Wiersum, W. (2016). *Verantwoord Toetsen en Beoordelen met een Rubric: een Onderzoek naar het Mediërend Effect van interbeoordelaarsbetrouwbaarheid op de Relatie tussen het Toepassen van een Rubric en Studenttevredenheid* (Masterproef, Open Universiteit Nederland). Geraadpleegd van <http://hdl.handle.net/1820/7369>

Wessel, P. (2018, 22 november). Rubric voor Rubrics. Geraadpleegd van <https://www.vernieuwonderwijs.nl/rubric-voor-rubrics/>

Wilson, L. (2019). Quantitative Research. In P. Liamputtong (Ed.), *Handbook of Research Methods in Health Social Sciences* (pp. 27-49). Singapore: Springer.

8 Bijlagen

8.1 Bijlage 1: vragenlijst studenten

De vragenlijst geldt zowel voor de studenten die bevroegd zijn vlak na het afleggen van het praktisch examen verpleegkundige vaardigheden als voor de studenten die niet hebben deelgenomen aan het praktisch examen. De vragen die rood aangeduid zijn in de vragenlijst, zijn de vragen die weggelaten zijn bij de bevraging van de studenten die niet hebben deelgenomen aan het praktisch examen.

Vragenlijst studenten

Inleiding

Dit onderzoek kadert binnen het opleidingsonderdeel 'interdisciplinair project' dat deel uitmaakt van de master opleidings- en onderwijswetenschappen aan de Universiteit Antwerpen. Het onderzoek heeft als doel het praktisch examen verpleegkundige vaardigheden te evalueren dat studenten in hun eerste jaar van de professionele bachelor verpleegkunde dienen af te leggen. Zo kan in kaart gebracht worden wat goed loopt en kan er advies gegeven worden over wat anders zou kunnen. Het opleidingsonderdeel verpleegkundige vaardigheden telt mee voor zeven studiepunten.

Deze bevraging gebeurt online en zal maximum 10 minuten duren. De verwerking van de gegevens gebeurt volledig anoniem. De rapportering naar Thomas More Mechelen wordt enkel geaggregeerd over de studenten en is niet te herleiden naar personen. Het al dan niet deelnemen aan deze bevraging heeft geen enkele impact op het examen van verpleegkundige vaardigheden of andere opleidingsonderdelen.

Titel onderzoek: Evaluatie van het examen verpleegkundige vaardigheden

Verantwoordelijke onderzoekers: Joachim Bonny, Katrien Cornelis, Casper Fleerackers, Kimberley Gregoor, Danny Majzik, Anne Mariën en Sofie van Ostaeyen.

Projectbegeleider Universiteit Antwerpen: Marije Lesterhuis

Consent

In te vullen door de deelnemer

- * Ik verklaar op een voor mij duidelijke wijze te zijn ingelicht over de aard, methode, doel en belasting van het onderzoek.
- * Ik weet dat de gegevens van het onderzoek beschikbaar zullen zijn voor de onderzoeksleider en de onderzoeksuitvoerders. De data zullen alleen anoniem aan derden bekend gemaakt worden. Wanneer de data of over de data gepubliceerd wordt, is dit telkens anoniem.
- * Mijn vragen zijn naar tevredenheid beantwoord.

Ik stem geheel vrijwillig in met deelname aan dit onderzoek. Ik behoud me daarbij het recht voor om op elk moment mijn deelname aan dit onderzoek te beëindigen, zonder dat ik hiervoor een reden moet opgeven.

In qualtrics de keuzemogelijkheid aanmaken: ja of nee

Inleiding vragenlijst

Een aantal stellingen worden gegeven. Lees elke stelling en duid het meest toepasselijke antwoord voor elke stelling aan. Er zijn geen goede of foute antwoorden. Spendeer niet te veel tijd aan een stelling, maar beantwoord ze zoals jouw gevoel de stelling het best omschrijft.

- 1: Helemaal niet akkoord met de stelling
- 2: Niet akkoord met de stelling
- 3: Neutraal: akkoord noch niet akkoord met de stelling
- 4: Akkoord met de stelling
- 5: Helemaal akkoord met de stelling
- NA: De stelling is niet van toepassing op mij.

| | Vraag | Wijze van antwoorden | Bron |
|-------------------------------|--|----------------------|------|
| | Algemene vragen | | |
| Achtergrondinfo | Geslacht | - man - vrouw | Zelf |
| | Leeftijd | Open vraag (getal) | Zelf |
| | Hoeveel keer legde u het examen verpleegkundige vaardigheden reeds af? (deze keer erbij geteld) | Open vraag (getal) | zelf |
| | Het vak 'verpleegkundige vaardigheden' De volgende stellingen gaan over uw perceptie op de lessen verpleegkundige vaardigheden uit semester 1. | | |
| Timing Opbouw Materiaal | De lessen verpleegkundige vaardigheden hebben mij geholpen bij de ontwikkeling van mijn competenties. | Likertschaal | 3 |
| | De lessen verpleegkundige vaardigheden waren interessant. | Likertschaal | 4 |
| | De lestijden werden goed benut. | Likertschaal | 3 |
| | De kwaliteit van het aangeboden cursusmateriaal was voldoende om op het praktisch examen verpleegkundige vaardigheden te kunnen slagen. | Likertschaal | 2 |
| | De opbouw van de lessen van het opleidingsonderdeel verpleegkundige vaardigheden zit goed in elkaar. | Likertschaal | 3 |

| | | | |
|--|--|--|------|
| | De opbouw van het vak verpleegkundige vaardigheden is goed: eerst het theoretisch examen verpleegkundige vaardigheden (november), dan de stage (december) en nadien het praktisch examen verpleegkundige vaardigheden (januari). | Likertschaal | Zelf |
| Transparantie | Het was mij duidelijk wat de leerdoelen waren van het opleidingsonderdeel verpleegkundige vaardigheden. | Likertschaal | 3 |
| | Mijn docenten communiceerden de doelen en verwachtingen voor de verpleegkundige vaardigheden die ik diende uit te voeren. | Likertschaal | 5 |
| | De doelen en verwachtingen voor de verpleegkundige vaardigheden die ik diende uit te voeren waren duidelijk. | Likertschaal | 5 |
| | Er werd een duidelijk, chronologisch stappenplan aan mij bezorgd voor elke verpleegkundige vaardigheid. | Likertschaal | Zelf |
| | Ik had een goed beeld van hoe ik elke verpleegkundige vaardigheid diende uit te voeren. | Likertschaal | Zelf |
| | Ik wist waar ik terecht kon voor hulp wanneer ik de verpleegkundige vaardigheden niet begreep. | Likertschaal | 6 |
| Instructiefase: voorbereiding op de lessen | | | |
| | Voor de aanvang van de lessen 'verpleegkundige vaardigheden' bereidde ik me goed voor. | Likertschaal | 2 |
| | Tijdens de lessen merkte ik dat ik mij beter had moeten voorbereiden. | Likertschaal | 2 |
| | Voor de aanvang van de lessen 'verpleegkundige vaardigheden' nam ik de toetsprocedures op Toledo door. | Likertschaal | 2 |
| Instructiefase: voorbereiding op het examen - skillslab | | | |
| Oefenmomenten | Ik werd aangemoedigd om deel te nemen aan de oefenkansen in het skillslab. | Likertschaal | 3 |
| | De aangeboden oefenmomenten in het skillslab werden effectief gebruikt door mijzelf. | Likertschaal | 2 |
| | De aangeboden oefenmomenten in het skillslab werden effectief gebruikt door mijn medestudenten. | Likertschaal | 2 |
| Frequentie | Hoe vaak heb je gebruik gemaakt van de vrije oefenkansen die aangeboden werden in het skillslab? | <ul style="list-style-type: none"> - meer dan 10 keer - 9 tot 10 keer - 7 tot 8 keer - 5 tot 6 keer - 3 tot 4 keer - 1 tot 2 keer - nooit | 2 |
| | Ik had graag meer tijd gekregen om te oefenen in het skillslab. | Likertschaal | 2 |

| | | | |
|--|--|--------------|---|
| Mening | Het is mijn verantwoordelijkheid als student om de vaardigheden in te oefenen totdat ik ze beheers. | Likertschaal | 2 |
| | Ik steek veel moeite in het goed uitvoeren van de verpleegkundige vaardigheden tijdens de oefenkansen | Likertschaal | 4 |
| | Het oefenen in het skillslab is een goede manier om mij voor te bereiden op het praktisch examen verpleegkundige vaardigheden. | Likertschaal | 2 |
| | Ik vind de ervaring die ik opdoe tijdens mijn deelname aan de oefenkansen in het skillslab teleurstellend. | Likertschaal | 3 |
| | Ik heb het gevoel dat mijn deelname aan de oefenkansen in het skillslab tijdverspilling is. | Likertschaal | 4 |
| | Het inoefenen van verpleegkundige vaardigheden tijdens de oefenkansen in het skillslab is interessant. | Likertschaal | 4 |
| | Het gebruik van simulatie activiteiten hebben mijn leertijd productiever gemaakt. | Likertschaal | 5 |
| | Ik heb vertrouwen dat de oefenmomenten in het skillslab bijdragen aan belangrijke inhoud die noodzakelijk is voor het beheersen van de verpleegkundige vaardigheden. | Likertschaal | 6 |
| Stress | De sfeer is ontspannen tijdens het oefenen binnen de lessen en het skillslab. | Likertschaal | 3 |
| | Het plezier om deel te nemen aan het opleidingsonderdeel verpleegkundige vaardigheden overtreft mijn gevoelens van stress. | Likertschaal | 3 |
| Feedback docenten | Ik had de kans om tijdens het inoefenen in het skillslab de vaardigheden te bespreken met mijn docent. | Likertschaal | 5 |
| | Ik had de kans om tijdens het inoefenen in het skillslab de vaardigheden te bespreken met medestudenten. | Likertschaal | 5 |
| | De docenten geven bruikbare feedback aan studenten tijdens de lessen. | Likertschaal | 3 |
| | De docenten geven duidelijke feedback aan studenten tijdens de lessen. | Likertschaal | |
| | De docenten geven bruikbare feedback aan studenten tijdens de oefenkansen in het skillslab. | Likertschaal | 3 |
| | De docenten geven duidelijke feedback aan studenten tijdens de oefenkansen in het skillslab. | Likertschaal | |
| | De docenten geven opbouwende kritiek aan studenten tijdens de lessen. | Likertschaal | 3 |
| | De docenten geven opbouwende kritiek aan studenten tijdens de oefenkansen in het skillslab. | Likertschaal | 3 |
| | De docenten wandelen rond om de studenten feedback te geven tijdens de oefenmomenten. | Likertschaal | 3 |
| Ik heb geleerd uit de feedback die de docent mij heeft gegeven voor, tijdens en na de oefenkansen. | Likertschaal | 5 | |

| | Examenmoment | | |
|--------------------|---|---|------|
| | Volgende stellingen hebben betrekking tot uw gevoel na het examen verpleegkundige vaardigheden te hebben afgelegd. | | |
| voor het examen | Voor de aanvang van het praktisch examen voelde ik mij kalm | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen voelde ik mij gespannen | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen voelde ik mij van streek | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen voelde ik mij ontspannen | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen voelde ik mij tevreden | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen voelde ik mij bezorgd | Likertschaal | 1 |
| | Voor de aanvang van het praktisch examen was het voor mij duidelijk hoe het examen zou verlopen. | Likertschaal | Zelf |
| | Voor de aanvang van het examen wist ik op welke criteria ik beoordeeld zou worden. | Likertschaal | Zelf |
| tijdens het examen | Ik was tevreden over mijn voorbereiding van de verpleegkundige vaardigheden om het praktisch examen af te kunnen leggen. | Likertschaal | 2 |
| | De sfeer was ontspannen tijdens het praktisch examen. | Likertschaal | 3 |
| | Het uitvoeren van de verpleegkundige vaardigheden tijdens het praktisch examen was... | <ul style="list-style-type: none"> - zoals ik had verwacht. - makkelijker dan ik had verwacht. - moeilijker dan ik had verwacht. | 2 |
| | Het uitvoeren van de verpleegkundige vaardigheden tijdens het praktisch examen was gelijkaardig aan het uitvoeren van de verpleegkundige vaardigheden tijdens de aangeboden oefenkansen in de lessen. | Likertschaal | 2 |
| | Het uitvoeren van de verpleegkundige vaardigheden tijdens het praktisch examen was gelijkaardig aan het uitvoeren van de verpleegkundige vaardigheden tijdens de aangeboden oefenkansen in het skillslab. | Likertschaal | 2 |
| | Er was een goede ondersteuning voor studenten die stress ervaren voor en tijdens het praktisch examen verpleegkundige vaardigheden. | Likertschaal | 3 |
| | Naar toekomst toe | | |

| | | | |
|--|---|--------------|------|
| | Ik heb er vertrouwen in dat ik de aangeleerde verpleegkundige vaardigheden in de praktijk beheers en kan toepassen op een echte patiënt. | Likertschaal | 2 |
| | Ik heb het gevoel dat ik na het volgen van het opleidingsonderdeel verpleegkundige vaardigheden goed voorbereid ben op het uitvoeren van verpleegkundige vaardigheden in de toekomst. | Likertschaal | 3 |
| | Slotvraag | | |
| | Zijn er nog zaken over het opleidingsonderdeel verpleegkundige vaardigheden en/of het examen van het opleidingsonderdeel die niet bevraagd werden en die je nog graag zou vermelden? | Open vraag | zelf |

Bronnen

Bron 1:

Marteau, T.M. en Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *British journal of clinical psychology*, 31, 301-306.

Bron 2:

Solvik, E. en Struksnes, S. (2018). Training nursing skills: a quantitative study of nursing students' experiences before and after clinical practice. *Nursing research and practice*, 2018, 1-9.

Bron 3:

Safari-Moradabadi, A., Kasmaei, P., Khalfe Nilsaz, M., Moradi, M., Ahmady, F. Mohammadloo, A. et al. (2018). Evaluation of the learning environment based on the dundee ready education environment measure model from the perspective of primary school students in roudsar city. *Journal of research in medical and dental science*, 6 (1), 140-145.

Bron 4:

Shivers, E., Hasson, F. En Slater, P. (2017). Pre-registration nursing student's quality of practice learning: clinical learning environment inventory (actual) questionnaire. *Nurse education today*, 55, 58-64.

Bron 5:

National league for nursing (2018), *Educational practices questionnaire (student version)*, (online) http://www.nln.org/docs/default-source/default-document-library/instrument-1_educational-practices-questionnaire.pdf?sfvrsn=0, gelezen op 30.12.2018

Bron 6:

National league for nursing (2018), *Student satisfaction and self-confidence in learning*, (online) http://www.nln.org/docs/default-source/default-document-library/instrument-2_satisfaction-and-self-confidence-in-learning.pdf?sfvrsn=0, gelezen op 30.12.2018

8.2 Bijlage 2: samenvatting focusgroep

Fase 1: DOEL VAN DE FOCUSGROEP

- **Wie zijn wij?** De onderzoekers stellen zich voor.
- **Waarom zijn wij hier?** Doel van het de focusgroep bekendmaken = onderzoek naar het praktisch examen verpleegkundige vaardigheden.
- In het kader van de betrouwbaarheid zal dit verslag gemaild worden aan één van de geïnterviewden, opdat hij/zij kan aangeven of het verslag een volledige, correcte en nauwkeurige weergave is van hetgeen gezegd is. We vragen één van de aanwezigen of hij/zij deze taak op zich wil nemen.
- **Consent:** Gebruik van geluidsopname: vertrouwelijk. Het gesprek wordt opgenomen. De opname wordt alleen door de interviewer zelf gebruikt ten behoeve van het gespreksverslag.
- Opbouw van de focusgroep: open vragen + stellingen
 - Hoe zouden jullie het praktisch examen verpleegkundige vaardigheden graag georganiseerd zien?
 - Op welke manier kan je de participatie aan het praktisch examen verhogen?
 - Op welke manier kan de validiteit en betrouwbaarheid van het praktisch examen 'verpleegkundige vaardigheden' verbeterd worden?

Timing: +- 15min

Fase 2: INLEIDING (post it's)

1. Kunnen jullie 2 of 3 elementen noteren die jullie beschouwen als een absoluut minimum om te kunnen slagen voor het praktisch examen?

Timing: +- 15min

Fase 3: TRANSITVRAAG (post it's)

- Benoem wat je goed vindt gaan en wat je zou willen verbeteren i.v.m. het praktisch examen.
 - peilen naar waarom ze iets hebben opgeschreven
 - informeren naar alternatief video-materiaal.
- Kunnen jullie elk een werkpunt en een positief punt noteren i.v.m. jullie eigen werkwijze tijdens het praktisch examen? (Zij die wensen mogen dit toelichten, echter is niemand verplicht.)

Timing: +- 20min

Fase 4: SLEUTELVRAAG combinatie open vragen en stellingen

Participatie:

- Welke bedenkingen maken jullie bij de deelname of niet deelname aan het praktisch examen?
 - Welke redenen zien jullie achter het niet deelnemen van de studenten?
 - Ligt het afwezigheidscijfer in dezelfde lijn van andere vakken?
 - Welke mogelijkheden zien jullie in het verhogen van de deelname aan het praktisch examen?

Timing/opbouw/materiaal:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
 - A: De lessen bereiden de student goed voor op het examen
 - B: De lessen bereiden de student niet voldoende voor op het examen
 - Waarom antwoordt u A of B?*
- In welke mate vinden jullie dat de studenten hier voldoende gebruik van maken?
→ skillslab

Transparantie:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
 - A: Het is duidelijk wat er van de student verwacht wordt tijdens het examen
 - B: Het is niet duidelijk voor studenten wat er nu precies van hen verwacht wordt tijdens het examen
 - Waarom antwoordt u A of B?*
- Welke mogelijkheden zien jullie om op het praktisch examen de kwaliteit van toetsing over de studenten heen te bevorderen (eenvormigheid, transparantie, over de docenten heen, betrouwbaarheid, **video-opnames**, ...)?

Feedback:

- Vanuit de literatuur wordt sterk benadrukt dat feedback een cruciale factor is om het leereffect te vergroten. Op welke manier maken jullie gebruik van feedback?
- In welke mate wordt er feedback gegeven tijdens de vrije oefenmomenten?
- Op welke concrete punten kan er in de vrije momenten worden bijgestuurd om het leerrendement van deze momenten te verhogen?

Timing: +- 30min

Tot slot:

- Of er vanuit de docenten nog vragen/opmerkingen zijn?
- Eventueel samenvatten wat er naar boven gekomen is in het gesprek. Om op die manier af te sluiten met een soort van consensus in hoe de groep het ziet?
- Bedanken voor hun tijd en openheid.

Timing: +- 10min

Taakverdeling:

- Anne: moderator
- Katrien: adjunct-moderator (timing in het oog houden, eventueel bijsturen en nuanceren waar nodig, ...)
- Casper: techniek (dubbele geluidsopname) en assisteert ook in de sturing van het gesprek

Onderstaand alternatief voor fase 4 (louter stellingen) werd niet weerhouden

Stellingen eventueel grote papieren maken? Kunnen we deze ophangen op een bord? zodat dit zichtbaar is? We weten namelijk niet of PPT mogelijk is.

Participatie:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
A: Studenten zijn geneigd om steeds deel te nemen aan het praktisch examen
B: Studenten nemen niet altijd deel aan het praktisch examen
→ stress? ondernemen docenten iets om stress hiervoor te verlagen?

Timing/opbouw/materiaal:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
A: De lessen bereiden de student goed voor op het examen
B: De lessen bereiden de student niet voldoende voor op het examen

Transparantie:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
A: Het is duidelijk wat er van de student verwacht wordt tijdens het examen
B: Het is niet duidelijk voor studenten wat er nu precies van hen verwacht wordt tijdens het examen

Skillslab:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
A: Oefenen in het skillslab geeft een meerwaarde voor het praktisch examen
B: Oefenen in het skillslab geeft geen meerwaarde voor het praktisch examen
→ doorvragen naar gebruik skillslab, hoe stimuleren? wat zien ze graag anders? zijn oefenkansen skillslab gelijkaardig aan het praktisch examen?

Feedback:

- Geef aan of u eerder denkt dat uitspraak A of uitspraak B van toepassing is
A: Feedback is belangrijk bij het aanleren van praktische vaardigheden
B: Feedback is niet zo belangrijk bij het aanleren van praktische vaardigheden
→ op welke manier wordt feedback gegeven en wanneer? wat willen zij anders zien? wat vinden zij haalbaar?

De bevindingen uit de vragenlijst achteraf kort bekendmaken en aftoetsen aan de antwoorden van de docenten.

Tabel mondelinge antwoorden

| Criteria | Deel v. criteria | Quote | Σ |
|-------------------|------------------|--|---|
| Vergelijkbaarheid | | | 0 |
| | Betrouwbaarheid | <p>1. Voor elke techniek is er een toetsprocedure waarin alle stappen staan die de studenten moeten doen (voorbereiding, tijdens de zorg, nazorg) ⇒ worden ook gebruikt bij het examineren van de examens</p> <p>2. De student zou telkens een ander resultaat krijgen indien hij/zij zijn examen bij een andere docent uitvoert (minimale verschillen)</p> <p>3. Het is duidelijk dat er individuele verschillen zijn, maar deze zal je nooit 100% kunnen uitsluiten ⇒ wel de vraag stellen hoe deze kunnen worden geminimaliseerd</p> <p>4. Het generieke (cfr generieke beoordelingsschemas) leeft maar wordt niet standaard opgelegd aan de docenten.</p> <p>5. Ene docent kruist benen en andere docenten doen dit niet? Is examen van die student dan moeilijker?</p> | 5 |
| | Validiteit | 1. Met praktijkexamen kan je meer testen. Bv: benen kruisen bij bloeddruk. Dit kan je niet testen via filmpjes. | 1 |
| Transparantie | | <p>1. Er zijn geen vooropgestelde criteria naar studenten toe gecommuniceerd omtrent bepaalde aspecten van de uit te voeren techniek die bv. breekpunten kunnen zijn voor het slagen of niet slagen. Per jaar, per techniek ligt accent anders. Hoe dit duidelijk communiceren naar studenten toe?</p> <p>2. 1 deelnemer geeft aan dat voor hem het grootste probleem is dat er te weinig transparantie is.</p> <p>3. Eén situatie week zo hard af van hoe de toetsprocedure dit vermeldde, dat het voor</p> | 4 |

| Criteria | Deel v. criteria | Quote | Σ |
|------------------|------------------|--|---|
| | | <p>studenten moeilijk was om in te schatten wat ze moesten doen om te kunnen slagen.</p> <p>4. Examensituatie week af van oefensituatie/toetsprocedure</p> | |
| Betekenisvolheid | Feedback | <p>1. Consensus: om het leerrendement te verhogen zou dit (cfr: feedback) wel ideaal zijn!</p> <p>2. Voorstel om onmiddellijk feedback te geven na het examen: de meningen zijn hierover verdeeld. Docenten zijn ook bang van extra werkbelasting en van discussie.</p> <p>3. Als oefening (cfr videomateriaal) : docenten zien wel voordeel in coaching, peerfeedback van meer getrainde studenten (bv 2de jaars)</p> <p>4. In het kader van een groeiportfolio (cfr videomateriaal) zien meerdere docenten dit als een positieve leermogelijkheid dat kan bijdragen tot het eigen maken van de praktische vaardigheden.</p> <p>5. Organisatie van een proefexamen: dit brengt echter extra werkdruk met zich mee</p> <p>6. Organisatie skillslab: Wanneer ingeroosterd is er altijd een docent aanwezig. Feedback wordt gegeven als het gevraagd wordt of als de docent een fout bij het oefenen opmerkt. Maar docenten lopen rond en hebben niet alles gezien. Als een student extra gaat oefenen is er geen begeleiding. Voorstel om coaching aan te bieden door de meer ervaren studenten. Combinatie van 1ste en 2de jaars kan productief zijn. 2 e jaars zijn volwassener en nemen hun opleiding meer aux serieux.</p> <p>7. Een docent geeft het voorbeeld uit andere hogeschool waarbij ene student meevolgt met andere en feedback geeft. Een deel van de punten</p> | 9 |

| Criteria | Deel v. criteria | Quote | Σ |
|---------------------|------------------|---|---|
| | | <p>staat op feedback. Stressniveau kan zo naar beneden en je toont toch dat je leerstof beheerst.</p> <p>8. Wat als 3de jaars studenten feedback geven aan 1ste jaars studenten? Is men zeker dat de 3de jaars studenten alle technieken correct onder de knie heeft? Peerfeedback vinden de meesten een goed idee.</p> <p>9. Feedback lijkt een belangrijk punt te worden, samen met coaching en verbetersleutel.</p> | |
| Reproduceerbaarheid | | <p>1. Benoem wat je zou willen verbeteren i.v.m. het praktisch examen. Antwoord: de objectiviteit van het examen</p> <p>2. Er zit verschil tussen docenten. Onderscheid in handelingen en waar leg je cesuur.</p> <p>3. Docenten stellen zich open op naar studenten om stressfactor te verlagen, sommigen stellen hen ook vooraf gerust.</p> <p>4. Het examen wordt als niet objectief ervaren.</p> <p>5. Subjectiviteit tussen verschillende docenten is groot, soms zelfs tot op het niveau van het persoonlijke. Studenten hebben letterlijk "schrik" van sommige docenten</p> <p>Bij het afnemen van een techniek was mijn benadering totaal anders dan dezelfde techniek bij een andere collega. Vb. collega deed met opzet iets, waardoor als een student dat niet opmerkte, deze niet geslaagd was</p> <p>6. Er is een groot verschil tussen studenten die men in het begin van de dag een examen afnam en het laatste examen van die dag</p> | 6 |

Tabel schriftelijke antwoorden

| Criteria | Deel v. criteria | Quote | Σ |
|-------------------|------------------|--|---|
| Vergelijkbaarheid | | | 0 |
| | Betrouwbaarheid | <p>1. Bij het afnemen van een techniek was mijn benadering totaal anders dan dezelfde techniek bij een andere collega. Vb: collega deed met opzet iets, waardoor als een student dat niet opmerkte, deze niet geslaagd was</p> <p>2. +: ik geef student na elke techniek de kans om aanvullingen/correcties te doen + stel mezelf voor</p> <p>3. -: geen criteria, wanneer geef je 12-14-16?</p> <p>4. -: Ik moet fouten van studenten direct noteren anders zou ik zaken kunnen vergeten, door dit continu te noteren raken sommige studenten afgeleid</p> | 4 |
| | Validiteit | 1. Met praktijkexamen kan je meer testen. Bv: 1 benen kruisen bij bloeddruk. Dit kan je niet testen via filmpjes. | 1 |
| Transparantie | | <p>1. Eén situatie week zo hard af van hoe de toetsprocedure dit vermeldde, dat het voor studenten moeilijk was om in te schatten wat ze moesten doen om te kunnen slagen.</p> <p>examensituatie week af van oefensituatie/toetsprocedure</p> <p>2. +: Focus op toetsprocedure en technische uitvoering</p> <p>3. +: toelichting kort voor aanvang, student geruststellen. Zenuwen wat reduceren zodat student wat kalmer kan nadenken en minder wordt afgeleid als ik iets noteer of wat er nog in het lokaal gebeurt.</p> <p>4. +: ik probeer studenten op hun gemak te stellen</p> <p>5. +: student vooraf geruststellen door voldoende uitleg te geven, dit lukt meestal</p> | 6 |

| Criteria | Deel v. criteria | Quote | Σ |
|---------------------|------------------|--|----|
| | | 6. +: ik probeer studenten gerust te stellen voor ze aan examens beginnen | |
| Betekenisvolheid | Feedback | 1. men (cfr: de docenten) geeft niet meteen feedback 2. +: argumentatie niet slagen | 2 |
| Reproduceerbaarheid | | 1. Subjectiviteit tussen verschillende docenten is groot, soms zelfs tot op het niveau van het persoonlijke. Studenten hebben letterlijk "schrik" van sommige docenten 2. Zelf is er een groot verschil tussen studenten die ik een 1e keer afnam en het laatste examen van die dag 3. objectiviteit van het examen 4. +: neutraliteit – geen zichtbare reacties die een richting kunnen geven over de prestatie tijdens het examen 5. +: studenten kennen mij van andere lessen waardoor ze zich minder geïntimideerd voelen bij ondervraging 6. -: moeilijk soms om een non-verbale pokerface op te zetten 7. -: heb ik alles gezien tijdens de vaardigheid? 8. -: nog meer noteren ook als het goed is. Nog meer pokerface. 9. -: duidelijker noteren i.v.m. techniek op blad toetsprocedure 10. -: misschien kan de student aan mijn mimiek al zien wat ik denk over het examen 11. -: vaak twijfel bij puntentoekening waardoor vaak afstemmen bij collega's uiteindelijk niet nodig want zelfde idee | 11 |

8.3 Bijlage 3: overzicht van de geteste modellen met bijhorende fit indices (EFA)

| | | Fit indices | |
|---------------------|---|--------------------|--------------|
| Getest model | | CFI | RMSEA |
| 1 | <p>Feedback Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10</p> <p>Transparantie T2, T3, T5, T6, Fb5, Fb6, Fb9</p> <p>Stress S2, S3, S4, S5, S6, S8</p> | 0.898 | 0.086 |
| 2 | <p>Feedback Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10</p> <p>Transparantie T2, T3, T5, T6, Fb5, Fb6, Fb9</p> <p>Stress S2, S3, S4, S5, S6, S8</p> <p>Errorcovariantie(s) Fb9~~Fb6</p> | 0.913 | 0.080 |
| 3 | <p>Feedback Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10</p> <p>Transparantie T2, T3, T5, T6, Fb5, Fb6, Fb9</p> <p>Stress S2, S3, S4, S5, S6, S8</p> <p>Errorcovariantie(s) Fb9~~Fb6 Fb9~~S8</p> | 0.929 | 0.072 |
| 4 | <p>Feedback Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10</p> <p>Transparantie T2, T3, T5, T6, Fb5, Fb6, Fb9</p> <p>Stress S2, S3, S4, S5, S6, S8</p> <p>Errorcovariantie(s) Fb9~~Fb6 Fb9~~S8 Fb8~~S8</p> | 0.949 | 0.061 |

| | | | |
|---|--|-------|-------|
| 5 | <p>Feedback Fb1, Fb3, Fb4, Fb7, Fb8, Fb9, Fb10, T7, T10</p> <p>Transparantie T2, T3, T5, T6, Fb5, Fb6, Fb9</p> <p>Stress S2, S3, S4, S5, S6, S8</p> <p>Errorcovariantie(s) Fb9~~Fb6 Fb9~~S8 Fb8~~S8 T2~~T3</p> | 0.965 | 0.051 |
|---|--|-------|-------|

8.4 Bijlage 4: generiek beoordelingsschema Thomas More



Generieke beoordelingsniveaus

| Cijfer | Omschrijving | ECTS grade |
|----------------------------------|--|------------|
| 20/20 19/20 18/20 17/20 | <p>De student heeft de competenties van het opleidingsonderdeel verworven. Zijn prestatie situeert zich op een hoger competentieniveau.</p> <p>Dit betekent dat de student de basiskennis beheerst en daarbij kenniselementen van andere disciplines integreert. De student is vaardig, waardoor hij routines en standaardprocedures aanpast in afwijkende situaties en praktijkgevallen. Hij getuigt ook van de bijhorende attitudes.</p> <p>De student slaagt erin met de nodige creativiteit en probleemoplossend vermogen bovenstaande elementen succesvol te integreren in ongekende (probleem)situaties van een hogere complexiteit.</p> <p>Het succesvol handelen van de student is van een uitzonderlijke kwaliteit en getuigt van een hoger competentieniveau.</p> | A |
| 16/20 15/20 | <p>De student heeft de competenties van het opleidingsonderdeel verworven. Zijn prestatie vertoont enkele kenmerken van een hoger competentieniveau.</p> <p>Dit betekent dat de student de basiskennis beheerst en daarbij kenniselementen van andere disciplines integreert. De student is vaardig, waardoor hij routines en standaardprocedures aanpast in afwijkende situaties en praktijkgevallen. Hij getuigt ook van de bijhorende attitudes.</p> <p>De student slaagt erin met de nodige creativiteit en probleemoplossend vermogen bovenstaande elementen succesvol te integreren in ongekende (probleem)situaties.</p> <p>Het succesvol handelen van de student vertoont kenmerken van een hoger competentieniveau.</p> | B |
| 14/20 13/20 | <p>De student heeft het vereiste niveau van de competenties (minimumcriteria) van het opleidingsonderdeel verworven.</p> <p>Dit betekent dat de student de basiskennis beheerst. De student is vaardig, waardoor hij routines en standaardprocedures aanpast in afwijkende situaties en praktijkgevallen. Hij getuigt ook van de bijhorende attitudes. Het geïntegreerd gebruiken van deze elementen in nieuwe (probleem)situaties van eenzelfde complexiteit, gelijkend op de gekende (probleem)situaties, wordt gerealiseerd.</p> <p>De student vertoont geen lacunes meer. De student zal succesvol handelen in de toekomst.</p> | C |

| | | |
|------------------------------|---|----|
| 12/20 11/20 | <p>De student heeft het vereiste niveau van de competenties (minimumcriteria) van het opleidingsonderdeel verworven.</p> <p>Dit betekent dat de student de basiskennis, vaardigheden beheerst en van de bijhorende attitudes getuigt. Het geïntegreerd gebruiken van deze elementen in nieuwe (probleem)situaties van eenzelfde complexiteit, gelijkend op de gekende (probleem)situaties wordt gerealiseerd.</p> <p>De student zal succesvol handelen in de toekomst, ondanks eventueel vastgestelde lacunes op de competenties.</p> | D |
| 10/20 | <p>De student heeft het vereiste niveau van het merendeel van de competenties (het merendeel van de minimumcriteria) van het opleidingsonderdeel verworven. Dit betekent:</p> <ul style="list-style-type: none"> • De student beheerst de basiskennis, met andere woorden hij heeft inzicht in de kennisgehele en kan onderscheid in begrippen en principes maken. • De student beheerst de omschreven vaardigheden, met andere woorden de student voert routines en standaardprocedures. • De student getuigt van de bijhorende attitudes. • De student integreert bovenstaande elementen in nieuwe (probleem)situaties van eenzelfde complexiteit, gelijkend op de gekende (probleem)situaties. <p>Ondanks vastgestelde tekorten op competenties, zijn in het handelen van de student aanwijzingen dat de student over de nodige groeipotentieel beschikt om deze tekorten in de toekomst ongedaan te maken.</p> <p>Succesvol handelen is in de toekomst mogelijk.</p> | E |
| 9/20 8/20 | <p>De student heeft het vereiste niveau van het merendeel van de competenties (het merendeel van de minimumcriteria) van het opleidingsonderdeel verworven.</p> <p>Dit betekent dat de student voor het merendeel van de competenties in beperkte mate de basiskennis en de vaardigheden beheerst en van de bijhorende attitudes getuigt. Het geïntegreerd gebruiken van deze elementen gebeurt onvoldoende in nieuwe (probleem)situaties van eenzelfde complexiteit.</p> <p>De student vertoont fundamentele tekorten op enkele competenties. In het handelen van de student zijn geen aanwijzingen dat de student over de nodige groeipotentieel beschikt om deze tekorten in de toekomst ongedaan te maken. Op basis van de prestatie van de student kan er geen uitspraak over het succesvol handelen in de toekomst worden gedaan.</p> <p>Deze tekorten zijn tolereerbaar.</p> | FX |
| 7/20 6/20 5/20 | <p>De student heeft het minimaal vereiste niveau van de competenties (minimumcriteria) van het opleidingsonderdeel niet verworven.</p> <p>Dit betekent dat de student onvoldoende de basiskennis, de vaardigheden beheerst en van de bijhorende attitudes getuigt. Het geïntegreerd gebruiken van deze elementen in nieuwe (probleem)situaties is daardoor niet mogelijk.</p> <p>De student vertoont fundamentele tekorten op zijn totale competentiebeheersing. Er zijn enkele (deel)competenties op het vereiste niveau verworven. In het handelen van de student zijn geen aanwijzingen dat de student over de nodige groeipotentieel beschikt om deze tekorten in de toekomst ongedaan te maken. Succesvol handelen is in de toekomst uitgesloten.</p> <p>Deze tekorten zijn niet tolereerbaar.</p> | F |
| 4/20 | <p>De student heeft het minimaal vereiste niveau van de competenties (minimumcriteria) van het opleidingsonderdeel niet verworven.</p> <p>Dit betekent dat de student de basiskennis en vaardigheden niet beheerst en ook niet getuigt van de bijhorende attitudes. Het geïntegreerd gebruiken van deze elementen in nieuwe (probleem)situaties is bijgevolg niet mogelijk.</p> <p>De student zal niet succesvol handelen in de toekomst.</p> | |
| 3/20 2/20 1/20 0/20 | De student presteert/antwoordt niet. | |

8.5 Bijlage 5: voorbeeld van een toetsprocedure Thomas More

Toetsprocedure intradermaal injecteren

VOOR

| | Goed | Niet | Fout + bemerkingen |
|--|------|------|--------------------|
| Verpleegkundige | | | |
| <ul style="list-style-type: none"> • Voer handhygiëne uit • Raadpleeg het voorschrift en controleer nogmaals <ul style="list-style-type: none"> <input type="checkbox"/> Juiste geneesmiddel <input type="checkbox"/> Juiste patiënt <input type="checkbox"/> Juiste dosis <input type="checkbox"/> Juiste toedieningswijze <input type="checkbox"/> Juiste tijdstip • Ken de plaatsen voor ID inspuiting | | | |
| Materiaal | | | |
| <ul style="list-style-type: none"> • Voorgeschreven geneesmiddel • Plateau/nierbekken • Ontsmettingsstof op basis van alcohol (controleer vervaldatum) • 2 deppers • Naaldcontainer | | | |
| Patiënt | | | |
| <ul style="list-style-type: none"> • Stel jezelf voor • Controleer identiteit van de patiënt aan de hand van het identificatiebandje • Leg de procedure uit aan de patiënt, beantwoord eventuele vragen en vraag naar toestemming om de handeling uit te voeren • Informeer naar bijzonderheden | | | |
| Omgeving | | | |

| | | | |
|---|--|--|--|
| <ul style="list-style-type: none"> • Doe het aanwezigheidslichtje aan • Zorg voor privacy • Controleer de omgevingstemperatuur en pas aan indien nodig • Creëer werkruimte • Doe ergonomische aanpassingen zoals bed op werkhoogte brengen | | | |
|---|--|--|--|

TIJDENS

| | Goed | Niet | Fout + bemerkingen |
|---|------|------|--------------------|
| <ul style="list-style-type: none"> • Ontbloom de huid (meestal binnenkant onderarm) • Palpeer en controleer de injectieplaats • Ontsmet de insteekplaats en respecteer contacttijd • Neem spuit en naald en verwijder het hoesje • Span de huid met de duim en wijsvinger van uw vrije hand • Verwittig de patiënt • Heb aandacht voor eventuele reacties van de patiënt • Steek de naald, met de opening naar boven gekeerd, evenwijdig aan het huidoppervlak in de bovenste huidlaag; de punt van de naald moet zich onder de huid aftekenen • Laat de huid los • Spuit de vloeistof langzaam en regelmatig in, er moet een kleine onderhuidse verhevenheid ontstaan • Trek de naald en spuit eruit (houd een gaasje voor de opvang van een eventuele vloeistof- of bloeddruppel gereed; niet afdrukken of masseren!) • Deponeer spuit en naald in de naaldcontainer • Markeer zo nodig het bobbeltje (onderhuidse | | | |

| | | | |
|---|--|--|--|
| verhevenheid) met een balpen of watervaste stift. | | | |
|---|--|--|--|

NA

| | Goed | Niet | Fout + bemerkingen |
|--|------|------|--------------------|
| Verpleegkundige | | | |
| <ul style="list-style-type: none"> • Voer handhygiëne uit • Rapporteer • Teken medicatie af | | | |
| Materiaal | | | |
| <ul style="list-style-type: none"> • Ruim met zorg het gebruikte materiaal op • Verwijder afval op een correcte manier | | | |
| Patiënt | | | |
| <ul style="list-style-type: none"> • Breng de patiënt in een comfortabele houding • Plaats oproepsysteem en ander materiaal binnen handbereik • Informeer naar de beleving van de patiënt | | | |
| Omgeving | | | |
| <ul style="list-style-type: none"> • Zorg voor een ordelijke kamer • Doe het aanwezigheidslichtje uit en regel verlichting | | | |

8.6 Bijlage 6: rubric voor rubrics

Rubric 1 (Wessel, paragraaftitel: rubric voor rubrics, rubric 1, 2018):

Een rubric die je er makkelijk bij pakt en waarmee je snel kan evalueren of het ontwerp van de rubric goed is en of deze op een goede manier wordt ingezet. Een handige rubric voor een snelle blik of als uitgangspunt bij het ontwerpen van een nieuwe.

Rubric voor Rubrics

| Criteria | Niveau 1: Onvoldoende | Niveau 2: Matig | Niveau 3: Goed |
|--|--|---|---|
| <i>Ontwerp</i> | | | |
| <i>Keuze & duidelijkheid in criteria</i> | De gekozen criteria zijn onduidelijk, hebben veel overlap, of zijn niet te herleiden tot de juiste doelen voor de taak. | De criteria zijn duidelijk maar niet altijd dekkend en/of de doelen zijn nog niet helder. | Alle criteria zijn duidelijk, er is geen overlap en de doelen zijn helder voor de taak. |
| <i>Verskil tussen niveaus</i> | Er is nauwelijks tot geen verschil te zien tussen de verschillende niveaus | Sommige verschillen zijn duidelijk, maar bij andere niveaus is het verschil te groot/klein. | Elk niveau is duidelijk verschillend en er is een logische opbouw. |
| <i>Duidelijkheid van beschrijving</i> | De tekst is niet voor iedereen te begrijpen. Het taalgebruik is vaag en/of onduidelijk waardoor er veel discussie mogelijk is. | Over het algemeen is de tekst te begrijpen. Sommige onderdelen leiden nog tot discussie. | De tekst is voor iedereen te begrijpen en is helder over wat behaald moet worden. |
| <i>Gebruik</i> | | | |
| <i>Invloed van leerlingen</i> | Leerlingen zijn niet betrokken bij het ontwikkelen. | Leerlingen hebben feedback kunnen geven op de tekst. | Leerlingen hebben echt meegedacht met docenten bij het ontwikkelen. |
| <i>Gebruik van rubric voor feedup, feedback en feedforward</i> | De rubric is niet gedeeld met leerlingen. | De rubric is gedeeld met leerlingen bij het nakijken. | De rubric is het eerste referentiepunt van een nieuwe taak. |

Dr. Bonnie B. Mullinix, Monmouth University, NJ. (2012)

Rubric 2 (Wessel, paragraaftitel: rubric voor rubrics, rubric 2, 2018):

Een rubric die zeer gedetailleerd ingaat op alle ontwerpaspecten van een rubric. Deze rubric is vooral goed te gebruiken om een bestaande rubric helemaal door te lichten op de criteria en niveaus. Een goede rubric als je er zeker van wilt zijn dat je een valide, duidelijk rubric hebt.

Rubric voor Rubrics

| Criteria | Zwak | Gemiddeld | Sterk |
|---------------------------------------|--|--|---|
| <i>Validiteit</i> | | | |
| <i>Dekkingsgraad</i> | De inhoud weerspiegelt niet wat er van leerlingen mag worden verwacht gezien hun niveau en de leerdoelen. De inhoud van de rubric geeft geen goede weergave van wat er nodig is om een goede prestatie te leveren, waardoor niet duidelijk is wat de rubric beoogt te meten. | De inhoud weerspiegelt wat er van leerlingen realistisch gezien mag worden verwacht gezien hun niveau en de leerdoelen, al geldt dat niet voor alle criteria. De inhoud van de rubric geeft grotendeels een goede weergave van wat er nodig is om een goede prestatie te leveren, maar er zijn punten waarop verbetering nodig is. | De inhoud weerspiegelt correct wat er van leerlingen realistisch gezien mag worden verwacht gezien hun niveau en de leerdoelen. De inhoud van de rubric geeft de beste mogelijke weergave van wat er nodig is om een goede prestatie te leveren op de vaardigheden die of het product dat wordt beoordeeld. |
| <i>Prestatie-Niveau</i> | De eisen die worden gesteld om het hoogste niveau te bereiken zijn onrealistisch en/of leerlingen kunnen door zwakke prestatie op meerdere criteria reeds een voldoende of hoger scoren. | Leerlingen mogen verondersteld worden in staat te zijn het hoogste niveau te bereiken op alle criteria. Op enkele criteria kan met een zwakke prestatie reeds een voldoende of hoger worden behaald. | Leerlingen mogen verondersteld worden in staat te zijn het hoogste niveau te bereiken op alle criteria. Om welk criterium op 'voldoende' niveau te beheersen moeten leerlingen een acceptabele prestatie neerzetten. |
| <i>Criteria</i> | | | |
| <i>Aantal, samenhang, opbouw</i> | De rubric bestaat uit een lange lijst criteria die niet of nauwelijks met elkaar lijken samen te hangen en waarin geen keuzes lijken te zijn gemaakt. De opbouw is niet logisch. | Het aantal criteria vereist enige aanpassing, zoals het opsplitsen van een afzonderlijk criterium in twee aparte criteria, of het samenvoegen van twee criteria. De criteria kennen verder een goede samenhang, maar de opbouw is niet altijd logisch. | De rubric omvat precies het juiste aantal criteria, zodat de complexiteit van de te leveren prestatie wordt weerspiegeld. De criteria kennen een goede samenhang en logisch opbouw. |
| <i>Weging</i> | Het is niet duidelijk of de rubric bedoeld is voor formatieve en of summatieve doeleinden. In de rubric wordt niet expliciet (ook niet in de begeleidende tekst) aangegeven hoe zwaar elk criterium meetelt en welke criteria belangrijker zijn dan andere. De criteria krijgen hierdoor niet de nadruk die zij verdienen. | Het is duidelijk of de rubric bedoeld is voor de formatieve en/of summatieve doeleinden. Indien summatieve doeleinden, dan is in de rubric weliswaar expliciet (eventueel in de begeleidende tekst) aangegeven hoe zwaar elk criterium meetelt en welke criteria belangrijker zijn dan andere, maar sommige criteria krijgen te veel of te weinig nadruk. | Het is duidelijk of de rubric bedoeld is voor de formatieve en/of summatieve doeleinden. Indien summatieve doeleinden, dan is in de rubric expliciet aangegeven hoe zwaar elk criterium meetelt, waardoor is af te leiden welke criteria belangrijker zijn dan andere (eventueel in begeleidende tekst). Alle criteria krijgen de nadruk die zij verdienen. |
| <i>Onafhankelijkheid</i> | De criteria zijn niet onafhankelijk. Er zijn veel overlappen tussen verschillende indicatoren. Verschillende indicatoren omvatten hetzelfde, waardoor de criteria niet verschillende aspecten meten. | De criteria zijn grotendeels onafhankelijk. Er is enige overlap tussen verschillende indicatoren, maar de rubric is desondanks in staat om verschillende aspecten te meten. | De criteria zijn onafhankelijk van elkaar, waardoor zij verschillende aspecten meten. Er is geen overlap tussen verschillende indicatoren. Wat hoort bij de ene indicator komt nergens anders terug. |
| <i>Onderscheiden d' vermogen</i> | Het aantal beoordelingsniveaus is veel te groot of juist te klein om zinvol en betrouwbaar onderscheid te kunnen maken tussen leerlingen. | Het aantal beoordelingsniveaus per criterium is niet overal toereikend om adequaat onderscheid te kunnen maken tussen goede en minder goede leerlingen en om de voortgang te kunnen meten, maar het aantal is redelijk eenvoudig aan te passen door één niveau toe te voegen dan wel twee niveaus samen te voegen. | Het aantal beoordelingsniveaus per criterium is logisch en toereikend. Er zijn voldoende niveaus om adequaat onderscheid te kunnen maken tussen goede en minder goede leerlingen en om de voortgang te kunnen meten. |
| <i>Indicatoren</i> | | | |
| <i>Aanwezig</i> | Geen van de beoordelingsniveaus is voorzien van indicatoren. | Niet alle beoordelingsniveaus zijn voorzien van indicatoren. | Elk beoordelingsniveau is voorzien van indicatoren. |
| <i>Transparantie en paralleliteit</i> | De indicatoren zijn vaag en weinig concreet beschreven, waardoor niet duidelijk wordt wat er van de leerlingen wordt verwacht. Of er wordt alleen onderscheid gemaakt tussen beoordelingsniveaus met woorden als 'zeer', 'erg', 'enige' of ze zijn volledig kwantitatief. De paralleliteit ontbreekt meestal. | Er is een poging ondernomen om indicatoren gedetailleerd en concreet te formuleren, maar sommige bevatten nog enkele vage beschrijvingen. Ze zijn wel grotendeels kwalitatief (en niet kwantitatief). Er is veelal sprake van paralleliteit, maar bij een aantal criteria is de opbouw niet logisch. | De indicatoren zijn gedetailleerd en concreet genoeg zodat duidelijk is wat de leerlingen moeten laten zien om een goede prestatie te leveren (ze zijn kwalitatief i.p.v. kwantitatief) en de beoordelaar een juiste beoordeling kan geven. Er is een logische opbouw van indicatoren over de niveaus (paralleliteit). |