

# Het assessmentcenter beoordeeld: face- en consequentiële validiteit

## Uitkomsten van een herhaald evaluatieonderzoek bij de opleiding Personeel en Arbeid

Drs. Rutger Kappe (rutger.kappe@inholland.nl) is docent bij de opleiding Personeel & Arbeid en drs. Jeroen Scholten Linde is teamleider bij de nieuwe opleiding Sport & Bewegen van de Hogeschool INHOLLAND. Beiden zijn tevens onderwijskundig adviseur bij de Dienst Onderwijs, Kwaliteit, Research & Development.

*Onderwerp van onderzoek is het assessmentcenter als toetsvorm binnen een krachtige competentiegerichte leeromgeving. Van het assessmentcenter is, aan de hand van studentevaluaties, de face- en consequentiële validiteit onderzocht. Op basis van het onderzoek blijkt dat het assessmentcenter zowel een hoge mate van face-validiteit als van consequentiële validiteit bezit. Studenten zijn van mening dat het assessmentcenter hen in realistische situaties plaatst en dat zij de uitkomsten van het assessmentcenter meenemen in hun verdere professionele ontwikkeling. Het doel waarmee de studenten aan het assessmentcenter deelnemen, blijkt geen bepalende factor te zijn wat betreft de twee validiteitsmaten en wat betreft hun 'overall' waardering van het assessmentcenter als toetsvorm. Uit het onderzoek blijkt verder dat de twee validiteitsmaten goed onderscheidbaar zijn, maar wel een duidelijke onderlinge relatie hebben. Het onderzoek onderschrijft het belang van de edumetrische benadering van kwaliteit van toetsen en de noodzaak tot een generalistische en praktische operationalisatie van consequentiële validiteit. Op basis van de resultaten wordt uiteindelijk geconcludeerd dat het assessmentcenter als toetsvorm binnen een krachtige competentiegerichte leeromgeving niet kan ontbreken.*

### INLEIDING

---

In een krachtige competentiegerichte leeromgeving neemt het leerproces en de competentieontwikkeling van studenten een belangrijke plaats in. In plaats van kennisoverdracht komen de leervragen en leerbehoeften van studenten centraal te staan. Studenten krijgen minder sturing door de opleiding en worden in grotere mate verantwoordelijk gesteld voor hun eigen professionele ontwikkeling. Inherent hieraan zijn nieuwe eisen voor het onderwijzen en de begeleiding van studenten (Biggs, 1999)

en aan de toetsen die gebruikt worden (Dochy e.a. 2002). In een krachtige competentiegerichte leeromgeving dient een toets meer te zijn dan het sluitstuk van een onderwijsseenheid. Aanvullend op certificering dient een toets het leerproces te ondersteunen. Summatieve toetsen dienen tevens formatieve informatie op te leveren. Dit laatste wordt door Black & William (1998) ook wel aangeduid als een verschuiving van 'assessment of learning' naar 'assessment for learning'. Deze verschuiving heeft gevolgen voor het definiëren en analyseren van de kwaliteit van toetsen ofwel assessments (Dierick, 2002). Om vast te kunnen stellen of de toets het leerproces ondersteunt, is het de vraag of met standaard kwaliteitscriteria als betrouwbaarheid en validiteit (begrips- en construct) kan worden volstaan. De edumetrische benadering van toetsen biedt hier wellicht uitkomst. Zo stellen Dierick e.a. (2002) voor dat bij een volledige beoordeling van de kwaliteit van nieuwe assessmentvormen naar vier aspecten moet worden gekeken; de validiteit van de taken, de validiteit (en betrouwbaarheid) van de beoordeling, de generaliseerbaarheid en de consequentiële validiteit. Binnen de validiteit van de taken staat de mate van authenticiteit van de taken centraal, waarbij wordt gestreefd naar een adequate weerspiegeling van de beroepspraktijk (kritische beroepssituatie). Deze kan vastgesteld worden door experts al dan niet in combinatie met novieten (face-validiteit). De validiteit (en betrouwbaarheid) van de beoordeling kan worden vastgesteld door de mate van overeenkomst te bepalen tussen de criteria die worden gehanteerd en hoe experts de taak beoordelen: gebruiken experts dezelfde criteria om een taak te beoordelen? Bij de generaliseerbaarheid wordt bepaald in hoeverre de taken kunnen worden veralgemeniseerd naar andere taken. De consequentiële validiteit richt zich op de consequenties van het assessment voor het onderwijs, bijvoorbeeld verandering in het leergedrag. Birenbaum (2003) maakt daarbij wel het voorbehoud dat de operationalisatie van de constructen uit de edumetrische benadering en hun bruikbaarheid nader onderzocht dienen te worden. Voor het vierde, meest vernieuwende construct geldt dat er op conceptueel niveau nog geen overeenstemming is. Momenteel vindt er wel hier en daar onderzoek plaats, op meer fundamenteel niveau binnen de KU Leuven. Het opgezette promotieonderzoek richt zich op de vraag welke kenmerken van toetsing en welke aspecten van het leren onderzocht moeten worden in het kader van de consequentiële validiteit.

In dit artikel wordt verslag gedaan van een toegepast onderzoek naar de consequentiële validiteit van het Assessmentcenter als toetsvorm binnen het Hoger Onderwijs. Door het ontbreken van conceptuele overeenstemming en daarmee een eenduidige operationalisering heeft dit onderzoek deels een experimenteel karakter.

### *Begrippenkader*

Onder een assessmentcenter wordt verstaan: een beoordelingsvorm waarbij in zo kort mogelijke tijd zo veel mogelijk relevante informatie verzameld wordt betreffende de (beroeps)competenties middels diverse praktijksimulaties, desgewenst gecombineerd met andere methoden, bijvoorbeeld schriftelijke tests en vragenlijsten (ontleend aan: Jansen & De Jongh, 1993). Meer specifiek is een assessmentcenter (AC) gericht op het in kaart brengen van het beheersniveau van relevante beroepscompetenties. In het onderwijs kan dat gebeuren met een formatieve (informatieve) doelstelling of een summatieve (beoordelende) doelstelling.

Voor het begrip competentie is hier de definitie uit het onderwijsconcept van INHOLLAND gebruikt. Een competentie is: 'Het mentale gereedschap om verschillende beroepstaken te kunnen uitvoeren. Naast kennis gaat het dan ook om vorming en ontwikkeling van vaardigheden, inzichten en houding. Gezamenlijk levert dit de afgestudeerde het vermogen om adequaat te handelen in een kritische beroepssituatie en hierop te reflecteren.' (Center of Excellence, 2003)

Eén van de in dit artikel onderzochte edumetrische validiteitsmaten betreft de validiteit van de taken. Deze wordt doorgaans door expertoordelen vastgesteld. Zij worden het best in staat geacht het realiteitsgehalte van (simulatie)taken te beoordelen. Omdat assessmentopdrachten normaliter samen met experts uit het werkveld worden ontwikkeld, is de validiteit van de taken in principe geborgd. Hoewel studenten als novieten op hun vakgebied worden beschouwd, is hun oordeel betreffende het realiteitsgehalte wellicht een relevante aanvulling. Dit geldt wellicht nog sterker wanneer de toets met een formatieve doelstelling wordt afgenomen. De vraag aan de toetsparticipanten is dan of de opdrachten van het Assessmentcenter in hun ogen relevant en realistisch zijn. Het oordeel van studenten die deelnemen aan het assessmentcenter kan vanwege hun noviteitstatus niet gelijk gesteld worden aan het oordeel van experts uit het beroepsveld. Zij kunnen echter wel hun indruk geven betreffende het realiteitsgehalte van de assessmentopdrachten. Deze indrukvaliditeit wordt door Drenth en Sijtsma (1990) ook wel 'face-validiteit' genoemd. Een begrip dat dicht in de buurt komt van wat De Groot (1970) als 'transparantheid' heeft omschreven. De Groot zag dat als een belangrijke conditie voor de 'acceptabiliteit' van een test(procedure). Drenth en Sijtsma schrijven hier verder over:

'hoe belangrijk wij dit, volgens De Groot (vaak verwaarloosde) gezichtspunt, ook achten, wij zouden "transparantheid" toch niet als een altijd geldende conditie willen zien, en zeker niet ter vervanging van validiteit. Beslissingen zijn het meest gebaat bij een juist inzicht, zelfs als dit inzicht is verkregen met een onderzoeksprocedure waarvan de leek het fijne niet snapt' (Drenth & Sijtsma, pag.182).

Het lijkt erop dat Drenth en Sijtsma in hun betoog vooral uitgaan van tests die afgenomen worden met een summatieve (beoordelende) doelstelling. Zijn doen verder geen uitspraak over de vraag hoe met 'transparantie' moet worden omgegaan bij een formatieve (informatieve) toetsdoelstelling. In de lijn van De Groot (1970) en Finucane e.a. (2002), die stellen dat assessments alleen succesvol kunnen zijn als de participanten het assessment als relevant en nuttig beschouwen, rijst de vraag of het construct face-validiteit terecht zo laag gewaardeerd wordt. Deze zou nog wel eens een voorwaarde kunnen zijn voor de meest vernieuwende edumetrische validiteitsmaat: consequentiële validiteit.

Onder consequentiële validiteit worden alle gewenste en ongewenste effecten van een assessment verstaan (onder andere Messick, 1994). Hieronder vallen zowel organisatorische als individuele gevolgen zoals aanpassingen in het curriculum dan wel wijzingen in het leergedrag van een student. Als technieken voor het onderzoeken van consequentiële validiteit worden door Gielen e.a., (2003) genoemd: observatie van instructiestrategieën, het vergelijken van de kwaliteit van leerresultaten van oude en nieuwe

vormen van assessment, uitspraken betreffende gewenste (en ongewenste) consequenties aan studenten voorleggen en het houden van semi-gestructureerde interviews.

## ONDERZOEKSVRAGEN

Focus binnen dit artikel ligt op consequentiële validiteit. Aangezien er mogelijk een relatie is met de validiteit van de taken is ook dit construct onderzocht. Bij gebrek aan een eenduidige operationalisatie van beide constructen voor gebruik van evaluatie bij assessmentcenters, ten tijde van het opstellen van de vragenlijst, zijn beide naar eigen inzicht geoperationaliseerd.

In het kort staan in dit artikel deze drie vragen centraal:

- 1 Vinden studenten het assessmentcenter bestaan uit relevante en realistische opdrachten.
- 2 Welke positieve consequenties heeft het assessmentcenter wat betreft de professionele ontwikkeling van de studenten?
- 3 Is er een relatie tussen face-validiteit en consequentiële validiteit?

Voor de volledigheid wordt in de volgende paragraaf de context waarin, en de instrumenten waarmee is gewerkt, toegelicht.

## ONDERZOEKSCONTEXT

Het onderzoek heeft plaats gevonden binnen de opleiding Personeel en Arbeid van de Hogeschool INHOLLAND, vestiging Haarlem. Deze opleiding is vanaf 1999 gestart met het aanbieden van competentiegericht onderwijs in een krachtige leeromgeving. Eén van speerpunten bij deze onderwijsinnovatie was het invoeren van assessmentcenters op diverse momenten tijdens de opleiding (zie Tabel 1). Bij het ontwikkelen van de assessmentcenters is de opleiding begeleid door een daarin gespecialiseerd bureau (Stichting NOA).

Tabel 1 Assessmentmomenten binnen de opleiding Personeel en Arbeid.

| Naam assessment        | Moment                     | Doelstelling  |
|------------------------|----------------------------|---|
| Introductie assessment | Bij aanvang opleiding      | Oriëntatie, kennismaking met assessment als beoordelingsvorm en met begrip competenties |
| Propedeuse assessment  | Eind eerste opleidingsjaar | Primair: summatief (BSA)<br>Secundair: persoonlijke ontwikkeling                        |
| Hoofdphase assessment  | Halverwege jaarstage       | Primair: formatief/persoonlijke ontwikkeling  |
| Afstuderen             | Eind opleiding             | Primair: summatief  |

Gekozen is voor assessmentcenters omdat daarmee meerdere competenties geïntegreerd en leerwegonafhankelijk gemeten kunnen worden. Daarnaast werd verondersteld dat assessmentcenters een hoge zelfreflectie waarde bezitten wat het zelfsturend vermogen van de student kan aanspreken. Andere aanname was dat het periodiek geïn-

tegreerd beoordelen het leerproces positief zou beïnvloeden. Beoogd criterium was de mate waarin de studenten de resultaten van de assessmentcenters gebruiken in hun professionele ontwikkeling. Daarbij kan gedacht worden aan diverse onderwijsleersituaties zoals trainingen, stages en afstuderen. Ter illustratie wordt in box 1 het propedeuseassessment toegelicht.

---

**Box 1: De assessmentopdrachten van het propedeuse assessment van de opleiding  
Personeel en Arbeid**

---

Het propedeuse assessment van de opleiding bestaat uit de volgende opdrachten:

- 1 Adviesgesprek met cliënt. In de een-op-een 'live' simulatie wordt de student in de rol van personeelsconsulent geplaatst. De student dient een cliënt te woord te staan en te adviseren omtrent een wervings- en selectievraagstuk.
- 2 Gespreksverslag. De opdracht betreft het op papier weergeven van wat besproken is in het adviesgesprek.
- 3 Teamvergadering. In deze groepsopdracht gaan de studenten (3 à 4) met elkaar in discussie omtrent een aantal ethische en beleidsvraagstukken. De discussie dient te leiden tot gemeenschappelijke meningsvorming.
- 4 Inwerkprogramma ontwerpen. In deze groepsopdracht dienen de studenten (3 à 4) op verzoek van het hoofd PZ van een fictieve organisatie een alternatief inwerkprogramma voor nieuwe medewerkers te ontwikkelen. Daarbij dienen zij zich aan enkele randvoorwaarden te houden. Aan het eind van het overleg dient het bedachte inwerkprogramma door de studenten gepresenteerd te worden.
- 5 Zelf-assessment. In deze schriftelijke opdracht reflecteert de student op de diverse onderdelen van het assessmentcenter en op de daarin verwerkte competenties.

In het propedeuseassessment zijn tien competenties verwerkt: adviseren, alternatieven aandragen, analyse & diagnose, effectief communiceren, leerdoelen formuleren, presenteren, samenwerken, schriftelijk uitdrukingsvermogen, verantwoorden. In de bijlage is een voorbeeld opgenomen van de competentie adviseren.

In het hoofdfaseassessment wordt eveneens een mix van 'live' simulaties en schriftelijke opdrachten gebruikt. De opdrachten van het hoofdfase assessment zijn inhoudelijk complexer en verschillend qua context. In het hoofdfase assessment komen dezelfde tien, plus drie extra (hogere orde) competenties aan de orde. Ten slotte is het gewenste beheersingsniveau van de competenties verschillend.

Binnen een jaargang doorlopen alle propedeuse- en hoofdfase-studenten een assessmentcenter. Tijdens de assessmentdag maakt elke student een afspraak met zijn assessor voor een terugkoppelingsgesprek. Deze afspraak vindt plaats binnen een week na deelname aan het assessment. Na dit gesprek krijgt de student een evaluatieformulier.

## METHODE EN RESULTATEN

Voor dit onderzoek is een vragenlijst in de vorm van een evaluatieformulier gebruikt. Evaluatie van de assessmentcenters vindt elk jaar plaats. Voor dit artikel zijn gegevens uit de jaren 2002, 2003 en 2004 gebruikt. Van alle studenten die het assessment doorlopen wordt een ingevuld evaluatieformulier ontvangen. Ondanks dit 100% responspercentage hebben een paar studenten een aantal vragen niet (duidelijk) beantwoord. Het evaluatieformulier bevat 27 vragen onderverdeeld in meerdere thema's. Alle evaluatievragen zijn geformuleerd als stellingen van het Likert-type met vijf antwoordmogelijkheden (helemaal mee oneens (1) tot en met helemaal mee eens (5)). Voor dit artikel worden alleen de vragen en resultaten vermeld, die gericht zijn op face-validiteit en consequentiële validiteit.

Allereerst wordt per jaargang het aantal respondenten waarop het onderzoek is gebaseerd gegeven (tabel 2). Daar er sprake is van zes onafhankelijke groepen is allereerst onderzocht of deze groepen bij elkaar gevoegd mogen worden. Uit variantieanalyses zijn geen significante verschillen (5%) tussen de groepen geconstateerd. Derhalve zijn de gegevens samengevoegd en verder als één groep geanalyseerd (N=232).

Tabel 2 Aantal respondenten per assessmentcenter per jaargang.

| Jaargang                  | 2002 | 2003 | 2004 | Totaal           |
|---------------------------|------|------|------|------------------|
| <b>Assessment</b>         |      |      |      |                  |
| 1 <sup>e</sup> studiejaar | 28   | 46   | 61   | 135 studenten    |
| 3 <sup>e</sup> studiejaar | 22   | 33   | 42   | 97 studenten     |
| Totaal                    | 50   | 79   | 103  | N= 232 studenten |

In de volgende tabellen staan naast de gemiddelde score per item, de standaardafwijking en de verdeling van de scores (percentage dat het niet met de stelling eens is (<3), geen uitgesproken mening (=3) en het met de stelling eens is (>3)). In de onderste tabelrij wordt de gemiddelde schaalscore weergegeven. Wat betreft face-validiteit zijn er drie items in de vragenlijst opgenomen. In tabel 3 staan de resultaten.

Tabel 3 Scores op vragenlijstitems betreffende face-validiteit.

| Items   | X   | St.dev | % < 3 | % = 3 | % > 3 | N <sup>i</sup> |
|---|-----|--------|-------|-------|-------|----------------|
| Ik vond de oefeningen die ik moest doen realistisch   | 3.9 | 0.86   | 7.5   | 16.7  | 75.5  | 227            |
| Middels het doen van de oefeningen heb ik zelf al een beeld gekregen van het beheersniveau van mijn competenties      | 3.7 | 0.89   | 9.3   | 21.6  | 69.2  | 227            |
| Middels het assessment heb ik een goed beeld gekregen welke competenties de opleiding en het beroep van mij vereisen. | 3.8 | 0.80   | 7.5   | 19,9  | 72.6  | 225            |
| Gemiddelde totaalscore  | 3.8 | 0.64   | 8.1   | 19,4  | 72,5  |                |

*i* Het aantal studenten in tabel 3 en 4 is verschillend, omdat er missing values zijn bij de beantwoording van de vragen in tabel 3 en 4.

Uit Tabel 3 blijkt dat de studenten het over het algemeen eens zijn met de uitspraken die horen bij de face-validiteit (gemiddelde totaalscore van 3.8). Voorts blijkt dat 76% van de studenten de oefeningen als realistisch beoordeelt ( $X = 3.9$ ). Daarnaast is 69% het met de stelling eens dat door het uitvoeren van de oefeningen een beeld te verkrijgen is van het beheersniveau van de eigen competenties; voor bijna 9% blijkt dat overigens niet te gelden. Dit item is net als het vorige gemiddeld met een zeer ruime 3 beoordeeld ( $X = 3.7$ ). Ruim 72% geeft aan door middel van het assessmentcenter een goed beeld van de benodigde beroepscompetenties te hebben gekregen ( $X = 3.8$ ). Ten slotte is een homogeniteitanalyse uitgevoerd. Dit heeft geresulteerd in een coefficient-alpha van 0.6.

Wat betreft consequentiële validiteit zijn er vijf items in de vragenlijst opgenomen. In de volgende tabel staan de resultaten van de items behorende bij consequentiële validiteit.

Tabel 4 Scores op vragenlijstitems betreffende consequentiële validiteit.

| Items   | X   | St.dev | % < 3 | % = 3 | % > 3 | N   |
|---|-----|--------|-------|-------|-------|-----|
| Met behulp van het assessment en het assessmentrapport ben ik beter in staat straks mijn persoonlijk ontwikkelingsplan (POP) te schrijven | 3.8 | 0.88   | 7.0   | 21.6  | 71.4  | 227 |
| Ik vond het zinvol om aan het assessment deel te nemen  | 4.1 | 0.82   | 3.5   | 12.8  | 83.7  | 227 |
| Een assessment levert extra informatie op omtrent mijn functioneren en ontwikkeling   | 3.9 | 0.96   | 9.7   | 15.4  | 74.9  | 227 |
| Middels het assessment heb ik een beter beeld gekregen van wat mij in de beroepspraktijk staat te wachten                                 | 3.4 | 0.97   | 14.9  | 32,9  | 52.3  | 222 |
| Het onderdeel assessment zou wat mij betreft niet afgeschaft mogen worden omdat het voor mij toegevoegde waarde heeft                     | 4.1 | 0.94   | 6.7   | 12.1  | 81.3  | 224 |
| Gemiddelde totaalscore  | 3.9 | 0.69   | 8.4   | 18,2  | 73,1  |     |

Uit tabel 4 blijkt dat de studenten in groten getale aangeven dat het doorlopen van het assessment meerwaarde heeft voor hun persoonlijke ontwikkeling. Uit de resultaten blijkt dat voor 71% van alle studenten extra informatie heeft verkregen omtrent zijn functioneren en beter in staat is zijn POP (persoonlijk ontwikkelingsplan) te schrijven ( $X = 3.8$ ). Ruim 83% vindt het zinvol om aan het assessment deel te nemen ( $X=4.1$ ), 3.5 % vindt dit niet. Voor 52% geldt dat zij middels het assessment een beter beeld gekregen hebben omtrent wat hen in de beroepspraktijk te wachten staat; voor 15% heeft het assessmentcenter echter nagenoeg geen toegevoegde waarde gehad voor hun beeld van de beroepspraktijk. Ruim 81% van de studenten vindt dat het assessment niet moet worden afgeschaft omdat het voor hen toegevoegde waarde heeft, 7 % vindt dit niet. De gemiddelde score op dit item is 4.1. De coëfficiënt-alpha voor de betrouwbaarheid van de schaal waarbinnen de items vallen is 0.8.

De derde onderzoeksvraag is gericht op de relatie tussen face- en consequentiële validiteit. Factoranalyse op de complete set van vragen levert bevestiging op van het bestaan van twee factoren. Zie tabel 5.

Tabel 5 Factoranalyse op alle vragenlijstitems betreffende face- en consequentiële validiteit. Extraction Method: Principal Component Analysis. Rotation Method: Varimax with Kaiser Normalization. a Rotation converged in 3 iterations.

| Items   | Componenten |    |
|---|-------------|----|
|   | 1           | 2  |
| Ik vond de oefeningen die ik moest doen realistisch   | ,1          | ,8 |
| Middels het doen van de oefeningen heb ik zelf al een beeld gekregen van het beheersniveau van mijn competenties                          | ,3          | ,4 |
| Middels het assessment heb ik een goed beeld gekregen welke competenties de opleiding en het beroep van mij vereisen.                     | ,1          | ,8 |
| Met behulp van het assessment en het assessmentrapport ben ik beter in staat straks mijn persoonlijk ontwikkelingsplan (POP) te schrijven | ,8          | ,1 |
| Ik vond het zinvol om aan het assessment deel te nemen  | ,8          | ,3 |
| Een assessment levert extra informatie op omtrent mijn functioneren en ontwikkeling   | ,8          | ,2 |
| Middels het assessment heb ik een beter beeld gekregen van wat mij in de beroepspraktijk staat te wachten                                 | ,4          | ,6 |
| Het onderdeel assessment zou wat mij betreft niet afgeschafte mogen worden omdat het voor mij toegevoegde waarde heeft                    | ,8          | ,2 |

Met uitzondering van het op een-na-laatste item, dat op beide constructen laadt, bevestigt de factoranalyse het bestaan van de bedoelde twee schalen. Met behulp van correlatieanalyse is nader gekeken naar de relatie tussen beide schalen. Er blijkt een redelijk sterk verband te zijn tussen face- en consequentiële validiteit ( $r = 0.53$ ,  $p < 0.001$ ,  $N = 226$ ). Uit een regressieanalyse met het rapportcijfer voor het assessmentcenter (zie tabel 6) als afhankelijke variabele en beide schalen als onafhankelijke variabele levert een verklaarde variantie ( $R^2$ ) van 33% op (zie box 2). Daarbij blijkt consequentiële validiteit met een beta van 0.48 factor drie zwaarder te wegen.

### Box 2: Twee resultaten regressieanalyse met het rapportcijfer als afhankelijke variabele

| Model | R       | R Square | Adjusted R Square | Std. Error of the Estimate |
|-------|---------|----------|-------------------|----------------------------|
| 1     | ,583(a) | ,340     | ,334              | ,6424                      |

a Predictors: (Constant), face-validiteit, consequentiële validiteit

| Model |              | Unstandardized Coefficients |            | Standardized Coefficients |  | t      | Sig. |
|-------|--------------|-----------------------------|------------|---------------------------|--|--------|------|
|       |              | B                           | Std. Error | Beta                      |  |        |      |
| 1     | (Constant)   | 4,432                       | ,289       |                           |  | 15,349 | ,000 |
|       | Conseq. val. | ,546                        | ,074       | ,477                      |  | 7,390  | ,000 |
|       | Face-val.    | ,207                        | ,080       | ,167                      |  | 2,592  | ,010 |

a Dependent Variable: rapportcijfer



Op het evaluatieformulier is de studenten ten slotte gevraagd het assessmentcenter in zijn totaliteit een rapportcijfer toe te kennen. Voor de totale groep heeft dit geresulteerd in een 7.3. Ondanks het ontbreken van significante verschillen tussen jaargangen wat betreft het gegeven rapportcijfer, wordt in tabel 6 per studiejaar en jaargang het rapportcijfer weergegeven.

Tabel 6 Rapportcijfer voor het assessmentcenter per studiejaar en onderwijsjaar, N=232.

|                           | <b>2001/2002</b> | <b>2002/2003</b> | <b>2003/2004</b> |
|---------------------------|------------------|------------------|------------------|
| 1 <sup>e</sup> studiejaar | 7.3              | 7.6              | 7.4              |
| 3 <sup>e</sup> studiejaar | 7.4              | 7.1              | 7.1              |

## CONCLUSIE EN DISCUSSIE

In dit onderzoek is door de 'ogen' van studenten gekeken naar het assessmentcenter als toetsvorm. Zij hebben een evaluatieformulier ingevuld met daarin opgenomen acht items betreffende face-validiteit en consequentiële validiteit. Daarnaast hebben zij het assessmentcenter als toetsvorm een rapportcijfer toegekend. Op basis van dit laatste kan geconcludeerd worden dat studenten het assessmentcenter als toetsvorm positief waarderen (tabel 6). Studenten zijn van mening dat zij in realistische en beroepsrelevante situaties worden geplaatst. Door het volgen van het assessment geven zij aan een goed beeld te krijgen van de competenties die in de beroepspraktijk van hen vereist worden. De eerste onderzoeksvraag – of studenten het assessmentcenter van de opleiding vinden bestaan uit relevante en realistische praktijksimulaties – kan bevestigend beantwoord worden. Literatuur over selectie- en loopbaanassessments gericht op toepassing in het bedrijfsleven wijst uit dat het realiteitsgehalte van praktijksimulaties bijdraagt aan de positieve waardering door kandidaten (onder andere Thornton & Byham, 1982) en de acceptatiegraad van de assessmentresultaten. De mate van face-validiteit, zoals dat naar voren gekomen is uit dit onderzoek, is in die zin eerder bevestigend dan verrassend te noemen. Hoewel de coëfficiënt-alpha van de schaal voor face-validiteit gezien het beperkte aantal items bevredigend is te noemen, achten wij het wenselijk deze verder te verhogen door enkele items aan de vragenlijst toe te voegen. Voor de opleiding Personeel en Arbeid is de positieve beantwoording van deze onderzoeksvraag overigens wel een welkome bevestiging van hun keuze voor assessmentcenters in het curriculum.

De tweede onderzoeksvraag – of het assessment gevolgen heeft voor hun verdere professionele ontwikkeling – kan eveneens bevestigend worden beantwoord. Studenten geven aan het zinvol te vinden om aan het assessment deel te nemen. Het assessment levert extra informatie omtrent het functioneren en hun ontwikkeling op. Tevens vinden zij dat ze door deelname aan het assessmentcenter beter in staat zijn om een persoonlijk ontwikkelingsplan te schrijven. De coëfficiënt-alpha van de schaal voor consequentiële validiteit is goed te noemen (0.8). In de schaal constructie levert alleen het item 'Middels het assessment heb ik een beter beeld gekregen wat mij in de beroepspraktijk

staat te wachten' qua factoranalyse een probleem op. Overwogen wordt om dit item tekstueel aan te passen of te verwijderen. Gevonden resultaten met betrekking tot de tweede onderzoeksvraag komen overeen met eerdere studies van Sambell e.a., (1997), Black en William (1998) en Smits e.a., (2005). Laatstgenoemden hebben hun onderzoek gericht op een formatief *performance assessment* van een verpleegkundeopleiding. Zij vonden eveneens dat studenten assessments nuttig vinden en dat er positieve leereffecten optreden. Dit onderzoek toont aan dat dezelfde positieve beleving ook optreedt bij een summatieve toetsdoelstelling. Ook als studenten worden 'afgerekend' op hun prestatie(s) kennen zij het assessmentcenter als toetsvorm een hoge waardering toe.

De derde onderzoeksvraag is gericht op de relatie tussen face-validiteit en consequentiële validiteit. Hoewel op basis van factoranalyse de vragenlijst twee constructen blijkt te bevatten, kunnen deze op basis van correlatieonderzoek ( $r=0.53$ ), niet als losstaand beschouwd worden. Het ligt ook voor de hand dat wanneer studenten van mening zijn dat zij onzinnige dan wel onrealistische opdrachten moeten uitvoeren, zij minder of zelfs niet van plan zijn met de uiteindelijke uitkomst(en) iets te doen. Dit pleit ons inziens voor een vernieuwde aandacht voor het construct face-validiteit bij de beoordeling van nieuwe assessmentvormen. Van de onderzochte edumetrische validiteitscriteria toont regressie-analyse overigens duidelijk aan dat voor de waardering van het assessmentcenter als toetsvorm consequentiële validiteit veel bepalender is. Het herhalende karakter en het hoge responspercentage zijn een goede garantie voor de betrouwbaarheid van de conclusies.

Wel zijn er enkele beperkingen aan het uitgevoerde onderzoek. Ten eerste is er voor beide onderzochte validiteitsmaten geen gebruikgemaakt van reeds bestaande operationaliseringen.

Dit beperkt onder meer de vergelijkbaarheid met andere studies. Daarnaast heeft de operationalisering niet tot doel gehad bruikbaar te zijn bij andere nieuwe assessmentvormen zoals de overalltoets of voortgangstoets. De bruikbaarheid van de operationalisaties is daarom zeer beperkt. Tweede beperking is gelegen in de enkelvoudige techniek die gebruikt is om consequentiële validiteit vast te stellen; de mening van studenten. Voor wat betreft consequentiële validiteit kan dit betekenen dat zij aangeven dat het assessment positieve effecten heeft ten aanzien van hun persoonlijke ontwikkeling, maar dat zij de resultaten van het assessment nooit echt gebruiken. Sambell e.a. (1997) wijzen er in dit verband op dat attitude en daadwerkelijke actie(s) in de praktijk wel eens kunnen verschillen. Uit een aanvullende evaluatie onder enkele docenten, studenten en stagebegeleiders blijkt overigens dat studenten de resultaten van het assessment regelmatig en in de volgende uiteenlopende leersituaties gebruiken:

- in hun stagevoorbereiding; voor het opstellen van stagedoelen;
- bij trainingen; onder andere bij de training persoonlijke effectiviteit;
- tijdens de stage(s); als gespreksstof met de stagebegeleider.

Op basis van deze resultaten lijkt de kloof tussen attitude en daadwerkelijk gedrag niet groot te zijn. Beperking wat betreft het construct consequentiële validiteit is gelegen in de beperkte operationalisering ervan. Er is niet in de volle breedte gekeken naar bedoelde en onbedoelde effecten van het assessmentcenter. Het onderzoek was specifiek

gericht op enkele bedoelde effecten. Daarmee is geen compleet recht gedaan aan het construct. Ten slotte is het bereik van dit onderzoek beperkt in die zin, dat het assessmentcenter slechts op twee validiteitsmaten onderzocht is. Dit levert de volgende bedenking op. Uit het onderzoek blijkt dat het assessmentcenter als een face- en consequentieel valide toetsvorm beschouwd mag worden. Volgens Drenth en Sijtsma (1990) mag men dit pas positief interpreteren als het Assessmentcenter eveneens betrouwbaar en constructvalide is. Als dat laatste niet het geval zou zijn, gaat de student zich immers richten op basis van 'onjuist inzicht'. Andersom kan men de vraag stellen wat het 'nut' (nog) is van een uiterst betrouwbaar en constructvalide assessment waarvan het resultaat geen bijdrage levert aan het ontwikkelingsproces van de student. Het is dan ook raadzaam om alle kwaliteitscriteria uit de edumetrie bij onderzoek te betrekken. Nader onderzoek naar de consequentiële validiteit van de verschillende assessmentvormen zoals de stationstoets, de casustoets, de overalltoets of de voortgangstoets verdient aanbeveling.

Op grond van de positieve resultaten van dit onderzoek mag het assessmentcenter als toetsvorm niet in een krachtige competentiegerichte leeromgeving ontbreken. Het levert een concrete meerwaarde op in het licht van de professionele ontwikkeling van studenten. Mits aan een aantal voorwaarden voldaan is, mag verwacht worden dat dezelfde resultaten met assessmentcenters bij andere opleidingen gevonden worden. Belangrijke voorwaarde bij het ontwikkelen van assessmentopdrachten is een nauwe samenwerking met het werkveld. Bij de uitvoering geldt als belangrijke factor de kwaliteit van de assessoren. Zij dienen goed onderlegd te zijn in het observeren, beoordelen en geven van feedback. Nadere richtlijnen voor training van assessoren kunnen opge- maakt worden uit een onderzoek van Percival e.a., (1994).

## REFERENTIES

---

- Biggs, J. (1999) *What the student does: teaching for quality learning at university*. Buckingham: Open University Press.
- Birenbaum, M. (2003) New insights into learning and teaching and the implications for assessment. In: M. Segers, F. Dochy & E. Cascallar (Red.), *Optimising new modes of assessment: in search of qualities and standards* (pp.13-37). Kluwer Academic Publishers.
- Black, P. & Wiliam, D. (1998) *Inside the Black Box. Raising standards through classroom assessment*. Londen: King's College London.
- Centre of Excellence (2003) *Backbone in beeld*. Hogeschool INHOLLAND. Den Haag
- Dierick, S., Watering, G. van de & Muijtjens, A. (2002) De actuele kwaliteit van assessment: ontwikkelingen in de edumetrie. In: F. Dochy, L. Heylen & H. van de Mosselaer (Red.) *Assessment in onderwijs. Nieuwe toetsvormen en examinering in studentgericht en competentiegericht onderwijs* (pp.91-122). Lemma. Utrecht.
- Drenth, P.J.D., Sijtsma, K. (1990) *Testtheorie; inleiding in de theorie van de psychologische test en zijn toepassingen*. Bohn Stafleu Van Loghum.
- Dochy, F., Heylen, L. & Mosselaer, H. van de (red.) (2002) *Assessment in onderwijs. Nieuwe toetsvormen en examinering in studentgericht en competentiegericht onderwijs*. Lemma. Utrecht.

- Finucane, P., Barron, S.R., Davies, H.A., Hadfield-Jones, R.S. & Kaigas, T.M. (2002) Towards an acceptance of performance assessment. *Medical Education*, 36, 959-965.
- Gielen, S., Dochy, F. & Dierick, S. (2003) Evaluating the consequential validity of new modes of assessment: the influence of assessment on learning, including pre-, post-, and true assessment effects. In: M. Segers, F. Dochy, E. Cascallar (Red.) *Optimising new modes of assessment: in search of qualities and standards* (pp.13-37). Kluwer Academic Publishers.
- Groot, A.D. de (1970) Some badly needed nonstatistical concepts in applied psychometrics. *Nederlands Tijdschrift voor de psychologie*, 25, 360-376.
- Jansen, P.G.W., Jongh, F.D. de (red.) (1993) *Assessmentcenters; een open boek*. Het Spectrum/Marka.
- Messick, S. (1994) The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher*, 2, 13-23.
- Percival, E., Anerson, M., & Lawson, D. (1994) Assessing beginning level competencies: the first step in continuing education. *Journal of Continuing Education in Nursing*, 25, 139-142.
- Sambell, K., McDowell, L. & Brown, S. (1997) But is it fair?: An exploratory study of student perceptions of the consequential validity of assessments. *Studies in Educational Evaluation*, 23, 349-371.
- Smits, H.S.B., Sluijsmans, D.M.A. & Jochems, W.M.G. (2005) *Performance Assessment in Nursing Education: a case study*. OTEC, Open University.
- Thornton, G.C., & Byham, W.C. (1982) *Assessmentcenters and managerial performance*. New York: Academic Press.

# Bijlage 1

*Operationalisering van de competentie Adviseren.*

## COMPETENTIE: ADVISEREN

**Definitie:** is in staat voor meerdere betrokken klantsystemen adviserend op te treden bij individuele, personele, organisatorische en maatschappelijke arbeidsvraagstukken.

| Niveau 1  | Niveau 2   | Niveau 3  | Niveau 4  | Niveau 5   |
|---|--|---|---|--|
| <p>Laat zien zich niet of nauwelijks in de situatie van het klantsysteem en zijn probleem in te kunnen leven. Stelt weinig open, verhelderende vragen en/of verdiepende vragen, geeft direct zijn mening. Cliënt is niet geholpen met advies.</p> | <p>Laat zien zich enigszins een beeld te kunnen vormen van relatie eenvoudige situatie van klantsystemen en is in staat daarbij een enkel advies te kunnen geven. Stelt enkele open en/of verhelderende vragen. Luistert naar verhaal of vraag van cliënt, maar gaat er niet diep op in. Adviezen zijn niet allemaal even relevant en zinvol. Cliënt is beperkt geholpen met advies.</p> | <p>Laat zien zich een voldoende beeld van de situatie van het klantsysteem te kunnen vormen. Formuleert meerdere, niet allemaal even relevante adviezen. Stelt open, verhelderende vragen en enkele verdiepende vragen. Luistert naar verhaal of vraag van cliënt. Cliënt is redelijk tevreden/geholpen met advies.</p> | <p>Laat zien zich een voldoende helder beeld van de situatie van het klantsysteem te kunnen vormen. Leeft zich voldoende in, formuleert het probleem of de situatie en komt met meerdere relevante adviezen. Stelt open, verhelderende en enkele verdiepende vragen. Luistert naar verhaal of vraag van cliënt en gaat daarop in. Is in staat te adviseren op het gebied van individuele, personele, organisatie en/of arbeidsvraagstukken.</p> | <p>Laat zien een adviseur te zijn. Leeft zich in, formuleert helder het probleem of de situatie, is creatief en komt met meerdere relevante en eenduidige adviezen voor individuele, personele, organisatorische en maatschappelijke arbeidsvraagstukken. Stelt open, verhelderende en verdiepende vragen. Luistert actief. Pro-actieve houding.</p> |