

De perceptie van hbo-studenten op de kwaliteit van **toetsing**

□ Karin Gerritsen-van Leeuwenkamp en Desirée Joosten-ten Brinke

Studententevredenheid wordt door hogescholen gebruikt als een indicator voor de kwaliteit van toetsing en heeft dan ook een plaats in de prestatieafspraken tussen het ministerie van OCW en de hogescholen. Sturing op basis van studententevredenheid is echter moeilijk. Een student geeft weliswaar een gevoelsreactie op de kwaliteit, maar daarmee is niet altijd direct duidelijk welk specifiek aspect verbeterd moet worden. Bovendien is vanuit marketingonderzoek bekend dat het vervullen van een kwaliteitsaspect niet altijd automatisch leidt tot tevredenheid. Hoe kan een hogeschool gericht de perceptie van hbo-studenten op de kwaliteit van toetsing gebruiken?

Perceptie van studenten

Het betrekken van de perceptie van studenten bij kwaliteitsborging en -verbetering van het hbo-onderwijs is vereist, omdat de student een belanghebbende is van een onderwijsorganisatie. Een student betaalt immers een onderwijsorganisatie voor diensten die leiden tot een diploma van waarde. Als de kwaliteit van toetsing onder de maat is, merken studenten het effect daarvan: de kwaliteit van toetsing beïnvloedt hun tevredenheid. Studenten worden veelal betrokken bij kwaliteitsverbetering en -borging van toetsing door middel van evaluaties, bijvoorbeeld de Nationale Studenten Enquête. Een onderwijsinstelling kent daardoor de mening van studenten over een aantal aspecten. Maar wat een onderwijsinstelling

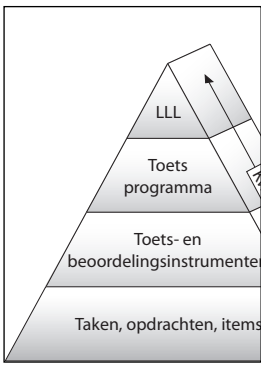
dan nog niet weet, is of alle essentiële kenmerken die de studententevredenheid beïnvloeden wel in de evaluatie zijn meegenomen. Ook is onduidelijk of alle kwaliteitskenmerken belangrijk of even belangrijk zijn voor de studententevredenheid. In dit artikel wordt het relatieve belang van vijftig kwaliteitskenmerken van toetsing voor de studententevredenheid beschreven.

Kwaliteitskenmerken van toetsing

De kwaliteitskenmerken van toetsing geven de eisen aan waaraan toetskwaliteit minimaal moet voldoen. De literatuur over de kwaliteit van toetsing onderscheidt diverse kwaliteitskenmerken (Boud, 2000; Dijkstra & Baartman, 2011; Evers, et al., 2009; Gulikers, Bastiaens & Kirschner, 2004; Joosten-ten Brinke, 2011; Linn, Bakker & Dunbar, 1991; Messick, 1995; Ploegh, Tillema & Segers, 2009; Sluijsmans, 2008; Tillema, Leenknecht & Segers, 2011; Van Berkel & Bax, 2006). Joosten-ten Brinke en Sluijsmans (in Joosten-ten Brinke, 2011) hebben op basis hiervan een kwaliteitspiramide voor eigentijds toetsen en beoordelen ontwikkeld (zie figuur 1). In deze piramide wordt de kwaliteit van toetsing onderverdeeld in vier niveaus, de kwaliteit van: 1) taken, opdrachten en items; 2) toets- en beoordelingsinstrumenten; 3) het toetsprogramma en 4) beoordelen in de context van leven lang leren.



Figuur 1. Kwaliteitspiramide voor eigentijds toetsen en beoordelen

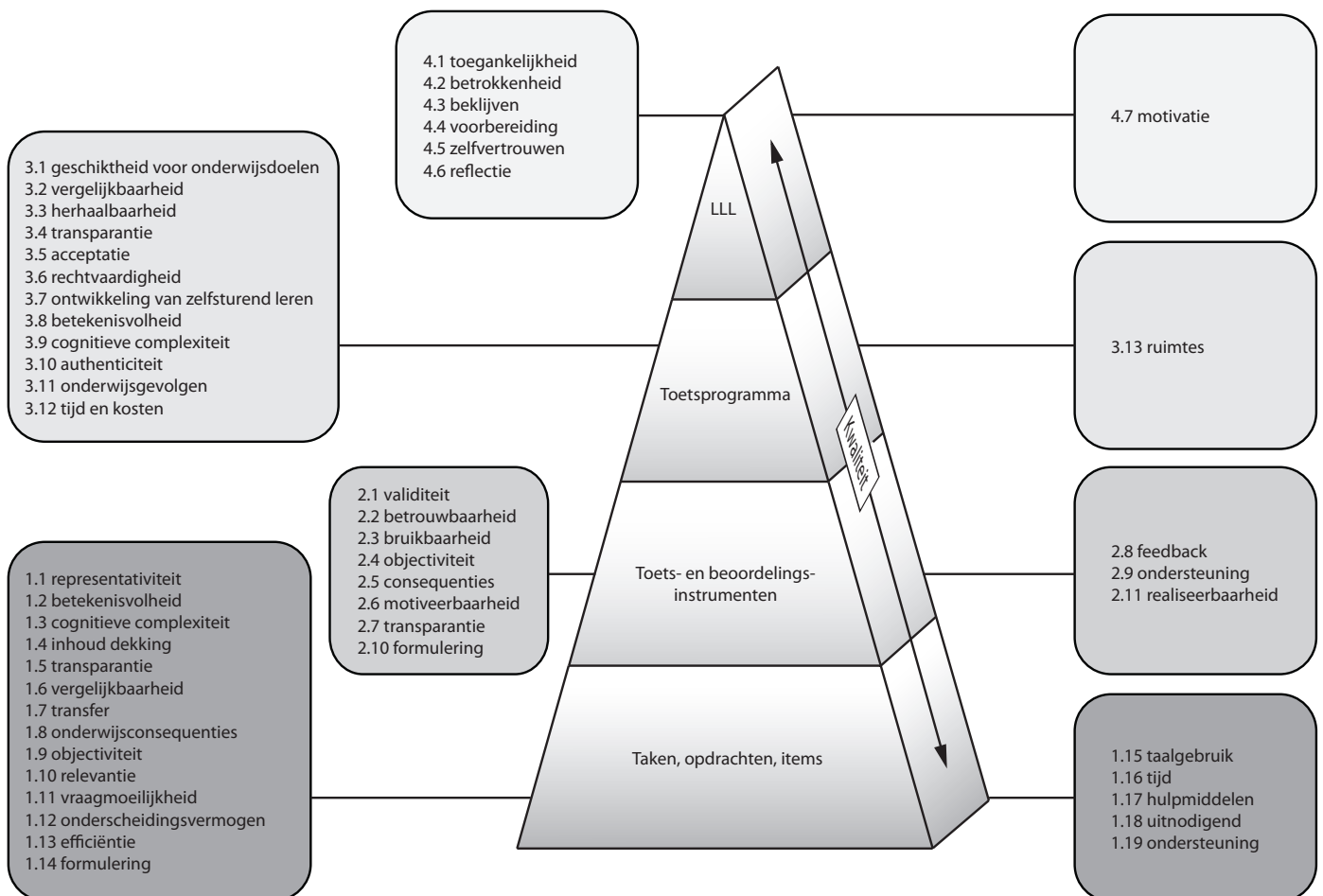


De onderwijsorganisatie moet zich in de eerste plaats richten op de instandhouding van basiskwaliteitskenmerken

In figuur 2 zijn links veertig kwaliteitskenmerken gerangschikt die uit literatuurstudie zijn verkregen. Daarnaast zijn op basis van groepsinterviews met studenten van de opleidingen fysiotherapie, podotherapie en verpleegkunde van alle leerjaren, tien nieuwe kwaliteitskenmerken naar voren gekomen die rechts in figuur 2 staan. Deze kwaliteitskenmerken werden niet in de bestudeerde literatuur over toetskwaliteit genoemd. Het zijn: motivatie, ruimtes, feedback, ondersteuning, realiseerbaarheid, taalgebruik, tijd, hulpmiddelen, uitnodigend, ondersteuning. Enkele

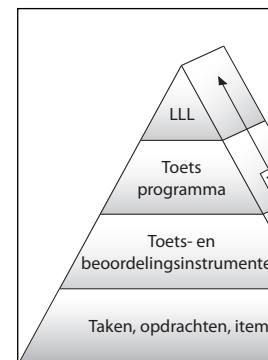
uitspraken van studenten over deze kwaliteitskenmerken: "Ik vind het belangrijk om een rustig lokaal te hebben. We hebben vaak genoeg een lokaal aan de weg gehad, nou dat vind ik bloedirritant." en "Als je een opdracht uit moet voeren dat je daar voldoende hulp bij krijgt, als je om duidelijkheid vraagt, dat je die ook krijgt."

Opgemerkt moet worden dat het wenselijk is om in de toekomst de vijftig kwaliteitskenmerken te clusteren en in aantal terug te brengen. Daarnaast is echter nog geen onderzoek gedaan.



Figuur 2. Vijftig kwaliteitskenmerken gerelateerd aan de niveaus van de kwaliteitspiramide voor eigentijds toetsen en beoordelen

Vervulling en verbetering van de lineaire kwaliteitskenmerken zorgt voor studententevredenheid



Het relatieve belang van kwaliteitskenmerken voor studententevredenheid

Er zijn relaties tussen kwaliteit en tevredenheid. Het gevoel van tevredenheid wordt veroorzaakt doordat kwaliteitskenmerken die volgens iemand vervuld moeten zijn, voldoende dan wel niet of onvoldoende vervuld zijn (Matzler & Hinterhuber, 1998). Bijvoorbeeld als de moeilijkheid van de toetsvraag niet aansluit bij het kennisniveau van de student, kan een student ontevreden zijn. Vanuit marketingonderzoek is bekend dat het vervullen van een kwaliteitskenmerk niet altijd zorgt voor meer tevredenheid. Dit wordt veroorzaakt door de niet altijd lineaire relatie tussen kwaliteit en tevredenheid (Kano et al., 1984). Het relatieve belang van een kwaliteitskenmerk voor tevredenheid kan verschillen. Dit is afhankelijk van of het kwaliteitskenmerk vanuit de perceptie van de student wel of niet vervuld moet zijn (Kano et al., 1984). De perceptie is het beeld dat een student heeft van de werkelijkheid om hem heen en dit kan veranderen door de tijd heen, op basis van ervaringen of nieuwe kennis. Maar ook het geslacht en de professionele identiteit van het toekomstige beroep beïnvloeden de perceptie (Lizzio & Wilson, 2004). Het relatieve belang van kwaliteitskenmerken voor de tevredenheid is vastgesteld met het Kano-model. In het onderzoek zijn 221 hbo-studenten in fysiotherapie, podotherapie en verpleegkunde van Saxion gevraagd naar het al dan niet vervullen van elk kwaliteitskenmerk. Elk kwaliteitskenmerk is dus twee keer bevraagd. De student gebruikte de volgende antwoordcategorieën: 1) prettig; 2) noodzakelijk; 3) neutraal; 4) acceptabel of 5) onprettig. De antwoorden op beide vragen zijn voor elke student afzonderlijk gecombineerd conform een daarvoor beschikbare matrix. Vervolgens zijn de frequenties op elke classificatie, per kwaliteitskenmerk bij elkaar opgeteld. De categorie met de hoogste frequentie is de dominante categorie van het kwaliteitskenmerk (Kano et al., 1984). Dit heeft geleid tot de classificatie van een kwaliteitskenmerk in één van de volgende vier categorieën:

1. basiskwaliteitskenmerken
2. lineaire kwaliteitskenmerken
3. aantrekkelijke kwaliteitskenmerken
4. invloedloze kwaliteitskenmerken

Kano et al. (1984) onderscheiden ook nog een vijfde categorie, de omgekeerde kwaliteitskenmerken, deze

wordt in dit artikel niet beschreven, omdat er door de studenten geen kwaliteitskenmerken als omgekeerde kwaliteitskenmerken zijn geïdentificeerd.

De geïdentificeerde kwaliteitskenmerken zijn als Kano model weergegeven in figuur 3. De twee dimensies in het figuur zijn 'vervulling van het kwaliteitskenmerk' (horizontaal) en de 'mate van tevredenheid' (verticaal). Achtereenvolgens worden de vier categorieën met kwaliteitskenmerken beschreven.

Basiskwaliteitskenmerken

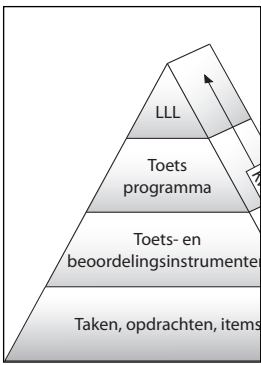
Kwaliteitskenmerken die het meest van belang zijn, zijn de basiskwaliteitskenmerken. Deze kenmerken zijn vanzelfsprekend wanneer zij vervuld zijn, maar leiden bij vervulling niet tot een gevoel van tevredenheid en resulteren in ontevredenheid wanneer zij niet vervuld zijn. De studenten classificeerden vijftien van de vijftig kwaliteitskenmerken als basiskwaliteitskenmerken (zie figuur 3). De drie basiskwaliteitskenmerken met de hoogste frequentie zijn:

- 3.12 tijd en kosten (58%): de uitvoering van een toetsprogramma is voor studenten en beoordelaars haalbaar.
- 2.11 realiseerbaarheid (57%): de standaarden en criteria zijn realiseerbaar voor gekwalificeerde studenten op het betreffende competentieniveau.
- 3.6 rechtvaardigheid (57%): alle studenten krijgen een eerlijke kans om hun competenties te laten zien. Studenten ervaren de beoordeling als eerlijk.

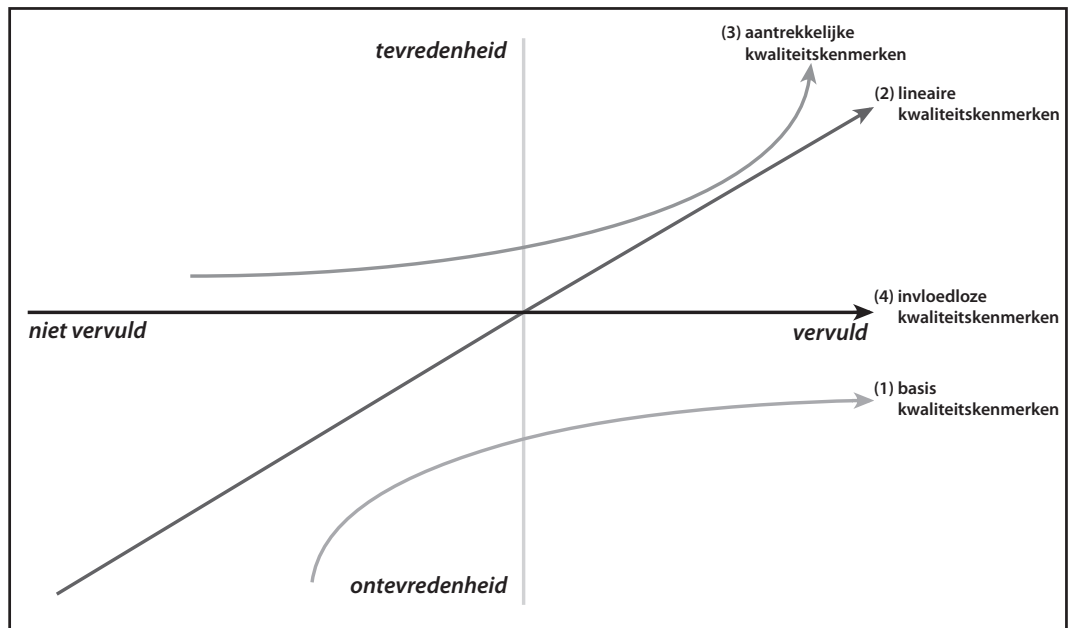
Ten aanzien van kwaliteitsverbetering en -borging is het wenselijk dat de onderwijsorganisatie zich eerst richt op investering in de instandhouding van basiskwaliteitskenmerken. Investeren in de verbetering van al vervulde basiskwaliteitskenmerken rendert niet, omdat deze niet leiden tot tevredenheid van studenten.

Lineaire kwaliteitskenmerken

De op een na belangrijkste kwaliteitskenmerken zijn de lineaire kwaliteitskenmerken. Deze negentien kenmerken (zie figuur 3) resulteren in tevredenheid wanneer zij vervuld zijn en resulteren in ontevredenheid wanneer zij niet vervuld zijn. De drie lineaire kwaliteitskenmerken met de hoogste frequentie zijn:



Het is niet eenvoudig om studenten tevreden te krijgen over de kwaliteit van toetsing



(1) basis kwaliteitskenmerken

- 1.1 representativiteit
- 1.5 transparantie
- 1.6 vergelijkbaarheid
- 1.9 objectiviteit
- 1.17 hulpmiddelen
- 1.19 ondersteuning
- 2.1 validiteit
- 2.2 betrouwbaarheid
- 2.3 bruikbaarheid
- 2.4 objectiviteit
- 2.6 motiveerbaarheid
- 2.7 transparantie
- 2.10 formulering
- 2.11 realiseerbaarheid
- 3.1 geschiktheid voor onderwijsdoelen
- 3.2 vergelijkbaarheid
- 3.4 transparantie

(2) lineaire kwaliteitskenmerken

- 1.7 transfer
- 1.11 vraagmoeilijkheid
- 1.13 efficiëntie
- 1.14 formulering
- 1.15 taalgebruik
- 1.16 tijd
- 1.18 uitnodigend
- 2.5 consequenties
- 2.8 feedback
- 2.9 ondersteuning
- 3.3 herhaalbaarheid
- 3.7 ontwikkeling van zelfsturend leren
- 3.8 betekenisvolheid

- 3.5 acceptatie
- 3.6 rechtvaardigheid
- 3.10 authenticiteit
- 3.11 onderwijsgevolgen
- 3.12 tijd en kosten
- 3.13 ruimtes
- 4.3 bekliven
- 4.4 voorbereiding

(3) aantrekkelijke kwaliteitskenmerken

- 1.2 betekenisvolheid
- 1.10 relevantie

(4) invloedloze kwaliteitskenmerken

- 1.3 cognitieve complexiteit
- 1.4 inhoud dekking
- 1.8 onderwijsconsequenties
- 1.12 onderscheidingsvermogen

Figuur 3. Het relatieve belang van vijftig kwaliteitskenmerken voor de tevredenheid van studenten

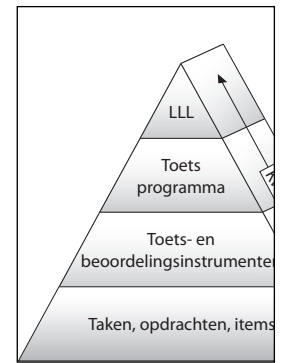
- 4.5 zelfvertrouwen (51%): het assessment geeft zelfvertrouwen om toekomstige leeractiviteiten of toetsen te ondernemen.
- 4.7 motivatie (48%): het assessment geeft motivatie om toekomstige leeractiviteiten of toetsen te ondernemen.
- 1.18 uitnodigend (48%): de taak of opdracht nodigt de student uit om deze te volbrengen. Het is wenselijk dat de onderwijsorganisatie zich

richt op de vervulling en verbetering van de lineaire kwaliteitskenmerken, omdat dit zorgt voor studententevredenheid.

Aantrekkelijke kwaliteitskenmerken

Een minder belangrijke categorie met kwaliteitskenmerken is die met de aantrekkelijke kwaliteitskenmerken. De aantrekkelijke kwaliteitskenmerken (zie figuur 3) geven tevredenheid wanneer zij vervuld zijn, maar

Niet elk kwaliteitskenmerk is even belangrijk



het is acceptabel als zij niet vervuld zijn. Deze twee kwaliteitskenmerken zijn:

- 1.2 betekenisvolheid (33%): de taak is een weergave van een levensecht probleem van het domein dat getoetst wordt.
- 1.10 relevantie (32%): de toetsvraag heeft een directe relatie met het leerdoel.

De onderwijsorganisatie kan investeren in de aantrekkelijke kwaliteitskenmerken. Dit heeft echter minder prioriteit dan de basis- en lineaire kwaliteitskenmerken omdat aantrekkelijke kwaliteitskenmerken alleen van belang zijn voor de tevredenheid en niet voor de ontevredenheid van studenten.

Invloedloze kwaliteitskenmerken

Vier van de vijftig kwaliteitskenmerken resulteren niet in tevredenheid maar ook niet in ontevredenheid, of zij nu wel of niet vervuld zijn (zie figuur 3). Deze kenmerken zijn:

- 1.8 onderwijsconsequenties (44%): de taak heeft impact op leren en instructie.
- 1.12 onderscheidingsvermogen (41%): de toetsvraag maakt een onderscheid tussen studenten met een hoge en lage eindscore of vaardigheid/kennisiniveau.
- 1.3 cognitieve complexiteit (34%): aan de hand van de taak kunnen de denkprocessen zichtbaar gemaakt en beoordeeld worden die een student doorloopt tijdens het werken aan een taak.
- 1.4 inhoud dekking (30%): de taak dekt de breedte van de lesstof die een student heeft ontvangen.

Investeren in invloedloze kwaliteitskenmerken is vanuit het studentenperspectief niet zinvol, omdat deze kwaliteitskenmerken noch van belang zijn voor de tevredenheid noch voor de ontevredenheid.

Uit de categorisering van de vijftig kwaliteitskenmerken blijkt dat het relatieve belang van de vijftig kwaliteitskenmerken van toetsing voor de tevredenheid van studenten niet voor elk kwaliteitskenmerk hetzelfde is. Ook blijkt dat meer kwaliteitskenmerken kunnen resulteren in ontevredenheid (vierenveertig) dan in tevredenheid van studenten (eenentwintig). Dit komt omdat het niet vervuld zijn van basis- en lineaire kwaliteitskenmerken (vierenveertig) leidt

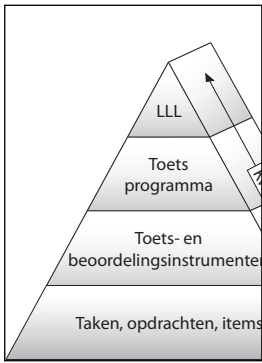
tot ontevredenheid. Voor studenten is het immers vanzelfsprekend dat de basiskwaliteitskenmerken vervuld zijn (zie figuur 2). Alleen de aantrekkelijke en de lineaire kwaliteitskenmerken kunnen resulteren in tevredenheid (eenentwintig). In de dagelijkse praktijk is het dus niet eenvoudig om studenten tevreden te krijgen over de kwaliteit van toetsing.

Verskil in het relatieve belang tussen groepen

Het relatieve belang van kwaliteitskenmerken voor tevredenheid blijkt te verschillen per soort opleiding (fysiotherapie, podotherapie en verpleegkunde) en per leerjaar (leerjaar 1, 2 en 3-4). Ten aanzien van het soort opleiding is gebleken dat vergeleken met verpleegkunde de opleidingen fysiotherapie en podotherapie onderling meer overeenkomsten hebben tussen de geclassificeerde kwaliteitskenmerken en de wijze van classificeren.

Eenzijds is een verschil in professionele identiteit een mogelijke verklaring die zorgt voor het verschil in perceptie (Lizzio & Wilson, 2004). Er is mogelijk meer overeenkomst in professionele identiteit tussen de beroepen van fysiotherapeut en podotherapeut. Binnen de opleidingen fysiotherapie en podotherapie, die opleiden tot een zelfstandige professie binnen de gezondheidszorg, neemt de competentie 'ondernemen' een prominentere plaats in dan in de opleiding tot verpleegkundige.

Anderzijds is een mogelijke verklaring dat het verschil in geslacht zorgt voor een verschil in perceptie (Lizzio & Wilson, 2004). Het aandeel mannen in de onderzoeksgroepen van fysiotherapie (18%) en podotherapie (28%) was aanzienlijk groter dan in de onderzoeksgroep van verpleegkunde (8%). Om dit met zekerheid vast te kunnen stellen is vervolgonderzoek wenselijk. Ten aanzien van de invloed van het leerjaar is gebleken dat de leerjaren 2 en 3-4 de meeste overeenkomst hebben tussen de geclassificeerde kwaliteitskenmerken en de wijze van classificeren in vergelijking met leerjaar 1. De eerstejaars studenten hebben ook meer kwaliteitskenmerken als invloedloos geclassificeerd dan de tweede en derde-vierdejaars studenten. Een verschil in ervaring, dat zorgt voor een verschil in perceptie, is een mogelijke verklaring hiervoor (Lizzio & Wilson, 2004).



De noodzaak van het betrekken van de perceptie van hbo-studenten bij de kwaliteit van toetsing wordt algemeen erkend

Tot slot

De noodzaak van het betrekken van de perceptie van hbo-studenten bij de kwaliteit van toetsing wordt algemeen erkend. In onderwijsorganisaties wordt geëvalueerd of studenten tevreden zijn. Een kwaliteitsverbetering zorgt echter niet automatisch voor tevredenheid van studenten. Voor gerichte kwaliteitsverbetering is het van belang eerst te analyseren welke kwaliteitskenmerken nu wel of niet in meer of mindere mate belangrijk zijn voor die tevredenheid. Vervolgens kan bij de verdeling van beschikbare middelen geïnvesteerd worden in juist die kwaliteitskenmerken die het grootste relatieve belang hebben voor de tevredenheid. Daarbij is het wenselijk dat er rekening gehouden wordt met verschillen die er zijn tussen leerjaren en opleidingstypen. Uit dit onderzoek zijn kwaliteitskenmerken naar voren gekomen die vanuit een studentenperspectief van belang zijn voor de kwaliteit. Voor toekomstig onderzoek is het relevant om te kijken welke van de vijftig kenmerken het grootste effect hebben op de kwaliteitsbeleving, daarmee kunnen de vijftig kenmerken eventueel in aantal teruggebracht worden. Kwaliteitsborging en -verbetering van toetsing kan zo, vanuit de perceptie van de student, door een onderwijsinstelling gericht worden uitgevoerd.

Literatuur

Dit artikel is gebaseerd op een masterthesis. De volledige thesis is in te zien in: dspace.ou.nl/bits-tream/1820/4295/1/OWKGerritsen-09052012.pdf.

- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151-167.
- Dijkstra, A. & Baartman, L.K.J. (2011). Zelfevaluatie van de kwaliteit van assessment. *Onderwijsinnovatie*, 13(1), 17-25.
- Evers, A., Lucassen, W., Meijer, R. et al. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests*. Amsterdam: NIP/COTAN.
- Gulikers, J., Bastiaens, T.J. & Kirschner, P.A. (2004). A five-dimensional framework for authentic assessment. *Educational Technology Research and Development*, 52(3), 67-86.
- Joosten-ten Brinke, D. (2011). *Eigentijds toetsen en beoor-*

- delen* (lectorale rede). Tilburg: Fontys Hogescholen.
- Kano, N., Seraku, N., Takahashi, F. et al. (1984). Attractive quality and must-be quality. *The Journal of the Japanese Society for Quality Control*, 14(2), 39-48.
- Linn, R.L., Bakker, E.L. & Dunbar, S.B. (1991). *Complex, performance-based assessment: Expectations and validation criteria*. Los Angeles: Center for Research on Evaluation, Standards and Student Testing.
- Lizzio, A. & Wilson, K. (2004). First-year students' perception of capability. *Studies in Higher Education*, 29(1), 110-128. doi: 10/1080.1234567032000165903.
- Matzler, K. & Hinterhuber, H.H. (1998). How to make product development projects more successful by integrating Kano's model of customer satisfaction into quality function deployment. *Technovation*, 18(1), 25-38.
- Messick, S. (1995). Validity of psychological assessment. Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741-749.
- Ploegh, K., Tillema, H.H. & Segers, M.S.R. (2009). In search of quality criteria in peer assessment practices. *Studies in Educational Evaluation*, 35, 102-109. doi: 10.1016/j.stueduc.2009.05.001
- Sluijsmans, D. (2008). *Betrokken bij beoordelen* (lectorale rede). Nijmegen: Hogeschool van Arnhem en Nijmegen.
- Tillema, H., Leenknecht, M. & Segers, M. (2011). Assessing assessment quality: Criteria for quality assurance in design of (peer) assessment for learning - a review of research studies. *Studies in Educational Evaluation*, 37, 25-34. doi: 10.1016/j.stueduc.2011.03.004
- Van Berkel, H. & Bax, A. (2006). *Toetsen in het hoger onderwijs*. Houten: Bohn Stafleu van Loghum.
- Mw. K.J. Gerritsen-van Leeuwenkamp MSc is onderwijskundig adviseur bij de Academie gezondheidszorg van Saxion en kenniskringlid bij het lectoraat Assessment van het Kenniscentrum Onderwijsinnovatie van Saxion. E-mail: kj.vanleeuwenkamp@saxion.nl.
Mw.dr. D. Joosten-ten Brinke is universitair docent bij de Open Universiteit en lector Eigentijds toetsen en beoordelen bij Fontys lerarenopleiding Tilburg.