

Kwaliteitsstandaarden basiskwaliteit toetsing

Een methodische handreiking

Fontys Hogescholen

Dienst Onderwijs en Onderzoek

Eindhoven, juni 2017

Els van Zijl, Mieke Jaspers, Claudia Loijens

Contactpersoon: Claudia Loijens, Dienst Onderwijs & Onderzoek, c.loijens@fontys.nl

Voorwoord

Toetsing is een complex proces met grote belangen voor de studenten waaraan, gezien het civiel effect van het getuigschrift, hoge maatschappelijke eisen gesteld worden. Binnen Fontys is de laatste jaren, onder andere ingegeven door externe ontwikkelingen, sterke (na)druk komen te liggen op toetsing en examinering en wordt binnen de interne kwaliteits- en verantwoordingssystematiek (MACON/MARAP-cyclus) gefocust op de standaarden toetsing (standaard 3) en gerealiseerde eindkwalificaties (standaard 4). Ongewenst neveneffect van de sterke focus op toetsing kan overigens zijn dat in het primaire proces toetsing het doel van handelen wordt in plaats van het middel dat faciliterend is voor het leerproces van studenten. Het kan tevens leiden tot een overvol toetsprogramma dat zorgt voor veel werkdruk onder docenten en administratieve medewerkers waardoor de toetskwaliteit onder druk komt te staan.

In feite moeten alle opleidingen altijd 'in control' zijn en de toetskwaliteit altijd aan de maat. Om opleidingen te ondersteunen in het continu verbeteren van de toetskwaliteit zijn in opdracht van het College van Bestuur kwaliteitsstandaarden ontwikkeld die onafhankelijk van een didactisch concept en de visie op leren en opleiden, de basiskwaliteit aangeven.

Bij de ontwikkeling van de standaarden en bijbehorend diagnose-instrument zijn medewerkers van Audit en de afdeling Juridische Zaken van de dienst Onderwijs & Onderzoek (O&O) betrokken evenals een externe toetsdeskundige. Met het diagnose-instrument is een pilot uitgevoerd bij de bacheloropleiding 'Toegepaste Psychologie' (TP) van 'Fontys Hogeschool voor HRM en Toegepaste psychologie' en de bacheloropleiding 'Docent Algemene Beeldende Vorming' van 'Fontys Hogeschool voor de Kunsten'. De belangrijkste bevindingen:

- Voorafgaand aan de zelfdiagnose is het belangrijk om samen te bepalen wat het doel is en welke groepen stakeholders worden uitgenodigd deel te nemen.
- Het is erg waardevol gebleken om stakeholders in verschillende functies en rollen binnen de opleiding te laten participeren zodat zij vanuit hun eigen perspectief een bijdrage kunnen leveren.
- Studenten zijn onmisbaar om een compleet beeld van de basiskwaliteit te krijgen. Vooral op het onderdeel 'toets- en beoordelingsinstrument' is hun inbreng zeer waardevol. Bovendien vinden zij het prettig om betrokken en gehoord te worden.
- Het groepsgesprek is een belangrijk onderdeel van de zelfdiagnose; enkel het individueel scoren van de standaarden levert weinig zinvolle informatie op. Tijdens het groepsgesprek wordt er diepgaand gediscussieerd en worden perspectieven uitgewisseld.

Tot slot hebben de groepsgesprekken en de discussie over de kwaliteit van het instrument geleid tot nieuwe ideeën voor de verbetering ervan. De aandachtspunten uit de pilots zijn meegenomen in de finale versie van het instrumentarium en worden bij de doorontwikkeling van het diagnose-instrument gebruikt.

Dienst Onderwijs & Onderzoek
Els van Zijl, Mieke Jaspers, Claudia Loijens
juni 2017

Inhoudsopgave

1. Inleiding	4
2. Conceptueel model	5
2.1 Kwaliteitspiramide van toetsing	5
1. Professionele cultuur van samenwerken	5
2. Toetsorganisatie.....	5
3. Toetsbeleid.....	5
4. Toetsprogramma	5
5. Toets- en beoordelingsinstrumenten	5
2.2 Gehanteerde kwaliteitscriteria	6
3. Kwaliteitsstandaarden basiskwaliteit toetsing	7
4. Praktische toepassing	8
4.1 Werkwijze	8
Literatuur	10

1. Inleiding

De afgelopen jaren is de kwaliteit van toetsing in het hoger onderwijs aan een kritisch oog onderworpen. De behoefte aan hogere toetskwaliteit is inmiddels vanuit overheidswege omgezet in een aantal concrete acties. Zo trad in september 2010 de wet Versterking Besturing in werking en werd in januari 2011 het nieuwe accreditatiestelsel geïntroduceerd, waarin 'Toetsing en gerealiseerde eindkwalificaties' als aparte standaard is opgenomen. Voorts is het rapport 'Vreemde ogen' gepubliceerd onder leiding van de commissie Bruijn, die een aantal adviezen geformuleerd heeft over de externe validering examenkwaliteit. In 2014 is het beoordelingskader voor opleidingen wederom aangepast (NVAO, 2014). 'Toetsing' en 'Gerealiseerde eindkwalificaties' zijn twee aparte standaarden waardoor de lat voor de toetsingspraktijk nog hoger wordt gelegd, omdat voor de standaard Toetsing elke standaard voldoende moet zijn en geen compensatie meer mogelijk is.

Dat allerm minst duidelijk is wat onder kwaliteit verstaan wordt, maakt onderstaand citaat uit het advies 'Kwaliteit in het hoger onderwijs van de Onderwijsraad' glashelder (Onderwijsraad, 2015).

"The question of what quality is, is difficult to answer, for quality is relative and "in the eye of the beholder". The key perspectives on the definition of quality can be found in the literature that appeared since Harvey and Green (1993), distinguishing quality as: exceptional, perfection, fitness-for-purpose, value for money, and transformation. Although these quality perspectives are connected to different objectives, rationales, value systems and theories, and therefore suggest connotations of intangibility and volatility, quality perspectives can be made operational by asking: to whom, by whom and by which standards (and against which value)?" (Onderwijsraad, 2015).

Kwaliteit en kwaliteitsstandaarden komen in samenspel en dialoog tot stand. Jaspers & van Zijl (2014) stellen dat in een professionele kwaliteitscultuur samenwerking vanuit een gemeenschappelijk gedragen visie en standaarden cruciaal is. Om blijvend toetskwaliteit te realiseren zijn gezamenlijk gedragen standaarden nodig als handvat voor (zelf)evaluatie en verbetering van toetskwaliteit. De minimale eisen van toetsing en examinering (de kwaliteitsstandaarden) vormen de basiskwaliteit en komen in dialoog met stakeholders en experts tot stand. Basiskwaliteit omvat inhoudelijke, procesmatige en procedurele standaarden ten aanzien van toetsing en examinering die in elk geval moeten worden gerespecteerd.

Bij de kwaliteitsstandaarden horen een diagnose-instrument en een website met uitleg en voorbeelden¹. Het ontwikkelde instrumentarium is een handreiking voor iedereen die betrokken is bij toetsing en examinering en die vanuit bundeling van krachten een bijdrage wil leveren aan een duurzaam en breed gedragen vertrouwen in de kwaliteit van het getuigschrift in het hoger onderwijs. Het complete instrumentarium kan door opleidingen als zelfdiagnose-instrument worden gebruikt:

- a. voor de realisatie van de 'basis op orde';
- b. voor ontwikkeling naar een hoger ambitieniveau;
- c. als basis voor toetsing en examinering in een nieuw kwaliteitsstelsel.

In deze handreiking staan de kwaliteitstandaarden voorafgegaan door een beschrijving van het conceptuele model dat is gebruikt om de standaarden te herleiden en te ordenen. In deze handreiking is een hoofdstuk gewijd aan hoe het bijbehorend diagnose-instrument bestaande uit een digitale scan, een analyse-instrument en een groepsgeprek, kan worden toegepast.

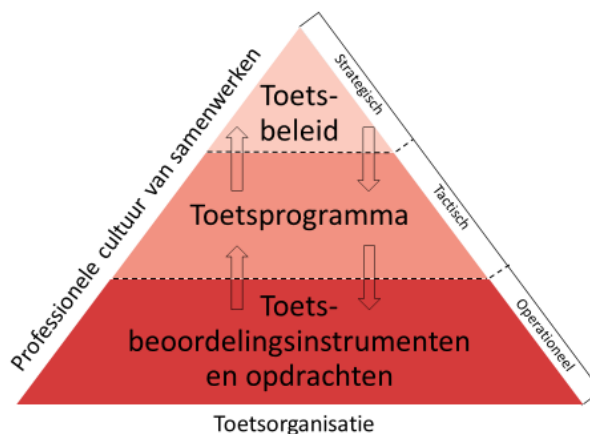
¹ Najaar 2017 operationeel

2. Conceptueel model

Als onderliggend model is gebruik gemaakt van de kwaliteitspiramide van toetsen en beoordelen door Sluijsmans, Peeters, Jakobs & Weijzen (2012), aangepast door Jaspers & van Zijl (2014). De toetskwaliteit is geoperationaliseerd in acht kwaliteitscriteria, afgeleid van bestaande inzichten en modellen.

2.1 Kwaliteitspiramide van toetsing

De kwaliteitspiramide van toetsen en beoordelen is geïntroduceerd om invulling te geven aan het begrip toetskwaliteit. In één figuur wordt alles qua toetsing weergegeven en met elkaar in verband gebracht. Onder de piramide een korte omschrijving van elke laag.



Figuur 1: Kwaliteitspiramide 1 van toetsen en beoordelen Sluijsmans, Peeters, Jakobs & Weijzen (2012), aangepast door Jaspers & van Zijl (2014)

1. *Professionele cultuur van samenwerken*

Een professionele cultuur van samenwerken is een cultuur waarin alle betrokkenen zich continu richten op het definiëren en realiseren van de gewenste toetskwaliteit door middel van een kritisch-reflectieve en onderzoekende houding (Rapport Onderwijsraad, Kwaliteit in het Hoger Onderwijs, 2015).

2. *Toetsorganisatie*

De toetsorganisatie heeft betrekking op de verdeling van rollen, taken en verantwoordelijkheden van alle betrokken actoren (structuur) én de afstemming daartussen zodat er blijvend toetskwaliteit geleverd kan worden (processen). (Sluijsmans, Joosten-Ten Brinke & van Schilt-Mol, 2015); (Jaspers & van Zijl, 2014). Tot de toetsorganisatie behoort ook de examencommissie.

3. *Toetsbeleid*

Het toetsbeleid is een samenhangend stelsel van visie en uitgangspunten en is sturend voor de inrichting van toetsprogramma, toets- en beoordelingsinstrumenten en toetsorganisatie (Sluijsmans, Joosten-Ten Brinke & van Schilt-Mol, 2015); (Jaspers & van Zijl, 2014).

4. *Toetsprogramma*

Het toetsprogramma is een zorgvuldig doordacht, evenwichtig programma van beoordelingsmomenten dat informatie oplevert over de prestaties van de student, voldoende feedback oplevert ten behoeve van het leerproces en garant kan staan voor de borging van de opleidingskwalificaties (Sluijsmans, Joosten-Ten Brinke & van Schilt-Mol, 2015); (Jaspers & van Zijl, 2014).

5. *Toets- en beoordelingsinstrumenten*

De toets- en beoordelingsinstrumenten (in de standaarden 'toetsen') omvatten het instrumentarium waarmee de prestaties van de student gemeten, gewaardeerd en beoordeeld worden (Sluijsmans, Joosten-Ten Brinke & van Schilt-Mol, 2015); (Jaspers & van Zijl, 2014).

2.2 Gehanteerde kwaliteitscriteria

Om de kwaliteit van toetsing zichtbaar te maken, worden acht kwaliteitscriteria gebruikt. Deze zijn afgeleid van het KIT-model (Baartman, 2013), het model 'De toetsing getoetst' (Sluijsmans, Joosten-Ten Brinke & van Schilt-Mol), het model van der Vleuten (van der Vleuten, Schuwirth, Driessen, Dijkstra, Tiggelaar, Baartman & van Tartwijk, (2012) en de in het boek 'Toetsen in het hoger onderwijs' (van Berkel, 2015) gebruikte kwaliteitscriteria. Het kwaliteitscriterium 'Interne consistentie' wordt in de literatuur als een aspect van validiteit beschreven. Het is in deze methodische handreiking als separaat criterium benoemd om de samenhang tussen en binnen de lagen van de piramide te kunnen duiden.

Tabel 1: Kwaliteitscriteria

Kwaliteitscriteria	Definitie
Betrouwbaarheid	mate waarin de scores consistent, nauwkeurig en reproduceerbaar zijn en dus staat kan maken op meetresultaten
Deskundigheid	mate van deskundigheid (kennis, vaardigheden attitude) van betrokkenen
Interne consistentie	mate van consistentie tussen de kaders en de vertaling naar de uitvoering
Leereffect	mate waarin het toetsprogramma inzicht geeft in de studievoortgang en een motiverend effect heeft op het leerproces van de student (verwevenheid leer- en toetsproces)
Organiseerbaarheid	de mate van efficiëntie waarmee de toetsing kan worden uitgevoerd
Studeerbaarheid	de mate waarin de opleiding is afgestemd op de aanleg, voorkennis, leervermogen, interesse en eigen doelen van studenten en externe omstandigheden (Wijnen,1992)
Transparantie	mate waarin de toetsen en de bijbehorende procedures duidelijk en begrijpelijk zijn voor alle betrokkenen en het voor alle betrokkenen helder is wat er wordt verwacht
Validiteit	mate waarin de toets betekenisvol en bruikbaar is en de conclusies die worden getrokken zijn gerechtvaardigd (interne validiteit) en ook opgaan in andere omstandigheden (generaliseerbaarheid, externe validiteit)

De definities zijn gebaseerd op genoemde bronnen en de literatuurreview 'Kwaliteit van toetsing binnen handbereik' van Cito (Maassen, den Otter, Wools, Hemker, Straetmans & Eggen, 2014). Voor de definitie van Studeerbaarheid is gebruik gemaakt van het rapport 'Te doen of niet te doen' van Prof. Wijnen uit 1992 (Wijnen, 1992).

3. Kwaliteitsstandaarden basiskwaliteit toetsing

Per laag van de kwaliteitspiramide is een aantal standaarden benoemd die bijdragen aan de toetskwaliteit gerelateerd aan acht onderscheiden kwaliteitscriteria.

Tabel 2: Kwaliteitsstandaarden basiskwaliteit toetsing

Nr.	Kwaliteitsstandaard	Kwaliteitscriterium
A. Professionele cultuur van samenwerken		
1.	Samenwerking tussen actoren vindt plaats vanuit gezamenlijk gedefinieerde kwaliteitsstandaarden.	deskundigheid
2.	Sturing vindt plaats op samenhang tussen commissies en actoren.	deskundigheid
3.	Actoren leveren vanuit de eigen rol en expertise een bijdrage aan de gewenste toetskwaliteit.	deskundigheid
4.	Actoren vragen zich continue af of de geldende kaders werken en toereikend zijn.	interne consistentie
5.	Actuele, theoretische inzichten zijn de basis voor professioneel handelen.	deskundigheid
B. Toetsorganisatie		
1.	Rollen, taken en verantwoordelijkheden zijn op elkaar afgestemd.	interne consistentie
2.	Professionalisering is afgestemd op het realiseren van toetskwaliteit.	interne consistentie
3.	Middelen zijn afgestemd op de uitvoering van het toetsbeleid.	organiseerbaarheid
4.	Monitoring van de toetskwaliteit vindt plaats op basis van indicatoren.	interne consistentie
5.	Professionalisering vindt effectief plaats door gebruik te maken van een beperkt aantal toetsvormen.	organiseerbaarheid
C. Toetsbeleid		
1.	De visie op toetsing is consistent met het onderwijsconcept.	interne consistentie
2.	(Wettelijke) kaders en regelgeving zijn richtinggevend voor het toetsbeleid.	interne consistentie
3.	Uitgangspunten voor de inrichting van de toetsorganisatie zijn geformuleerd vanuit de visie op toetsing.	interne consistentie
4.	Uitgangspunten voor de inrichting van het toetsprogramma zijn geformuleerd vanuit de visie op toetsing.	interne consistentie
5.	Uitgangspunten voor het ontwerp van toets- en beoordelingsinstrumenten zijn geformuleerd vanuit de visie op toetsing.	interne consistentie
D. Toetsprogramma		
1.	Opleidingskwalificaties zijn uitgewerkt in meetbare leeruitkomsten	validiteit
2.	De combinatie van toetsen is dekkend voor de leeruitkomsten.	validiteit
3.	Belangrijke beslissingen worden gebaseerd op uitkomsten van meerdere beoordelingsmomenten in vergelijkbare situaties.	betrouwbaarheid
4.	Er is een beperkte set van verschillende toetsvormen.	organiseerbaarheid
5.	De combinatie van toetsen levert voldoende rijkheid aan informatie op ten behoeve van het leerproces van de student.	leereffect
6.	De toetsvormen zijn passend bij het niveau van de opleiding.	validiteit
7.	Het aantal toetsen is beperkt.	studeerbaarheid
8.	Toetsen zijn evenwichtig gespreid over het studiejaar.	studeerbaarheid
9.	Actoren hebben zicht op het totale toetsprogramma (alle toetsen).	transparantie
10.	Actoren hebben zicht op de toetsprogrammering.	transparantie
E. Toets- en beoordelingsinstrumenten		
1.	Toets- en beoordelingsinstrumenten zijn geconstrueerd vanuit vormeisen.	betrouwbaarheid
2.	De inhoud van de toetsen is passend bij de leeruitkomsten.	validiteit
3.	Beoordelingscriteria en cesuur zijn passend bij de leeruitkomsten.	interne consistentie
4.	Bij de constructie van de toets zijn meerdere examinatoren betrokken.	betrouwbaarheid
5.	Bij het beoordelen van studenten zijn meerdere examinatoren betrokken.	betrouwbaarheid
6.	Examinatoren hebben de expertise om een kwalitatief betrouwbare beoordeling te geven.	betrouwbaarheid
7.	Toetsafname vindt onder gecontroleerde omstandigheden plaats.	betrouwbaarheid
8.	Beoordelingscriteria en cesuurmethode zijn vooraf bij alle actoren bekend.	transparantie
9.	Toetsen geven de student rijkheid aan informatie voor het leerproces.	leereffect
10.	Studenten krijgen effectieve feedback.	leereffect
11.	Studenten vinden de toets betekenisvol.	leereffect

4. Praktische toepassing

In dit hoofdstuk wordt uiteengezet hoe het instrument kan worden toegepast.

Voorafgaand aan de zelfdiagnose is het belangrijk dat duidelijk wordt gedefinieerd voor welk doel de diagnose wordt uitgevoerd, welke groepen stakeholders van binnen en buiten de organisatie die betrokken zijn bij toetsing en examinering deel zullen nemen, en een minimaal aantal respondenten. Als aan het minimale aantal respondenten niet voldaan wordt, zijn de resultaten uit de scan minder goed bruikbaar en kunnen deze slechts voorzichtig als indicatie gebruikt worden.

4.1 Werkwijze

De werkwijze bestaat uit een individuele fase waarin een steekproef van alle stakeholders (studenten, docenten, management, werkveld, toets- en examencommissie) de basiskwaliteit evalueren vanuit hun eigen perspectief, gevolgd door een groepsinterview of groeps gesprek waarin alle individuele diagnoses worden samengenomen en besproken.

Fase 1: individuele evaluaties van Basiskwaliteit

De individuele fase is bedoeld om reflectie over de kwaliteit op te roepen en om aanknopingspunten te bieden voor het gesprek over basiskwaliteit. De zelfdiagnose heeft dus een formatief doel. In deze fase wordt een combinatie van verschillende perspectieven op kwaliteit binnen de opleiding (maar wellicht ook van stakeholders buiten de opleiding) verzameld. Stakeholders hebben bij voorkeur verschillende functies binnen de opleiding. Denk bijvoorbeeld aan directeurs, opleidingsmanagers, studenten, docenten, commissieleden van examencommissie en toetscommissie.

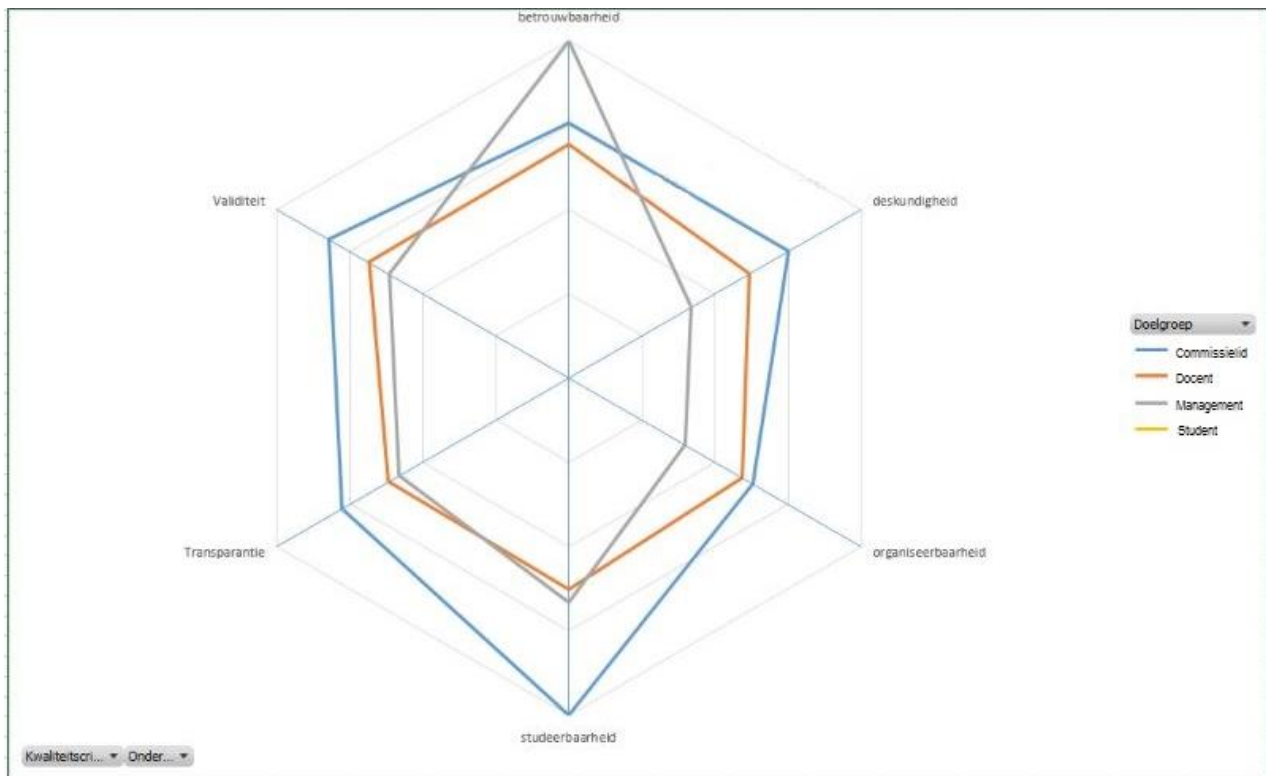
Stakeholders vullen zelfstandig de digitale vragenlijst in, waarin zij een waardering geven voor de standaarden vanuit hun eigen perspectief (driepuntsschaal); alle begrippen genoemd in de standaarden zijn voorzien van een uitleg.

Deze fase kan ook individueel door één docent uitgevoerd worden om inzicht te krijgen in de kwaliteit van de eigen toetsing in relatie tot het grotere geheel van toetsing binnen de opleiding.

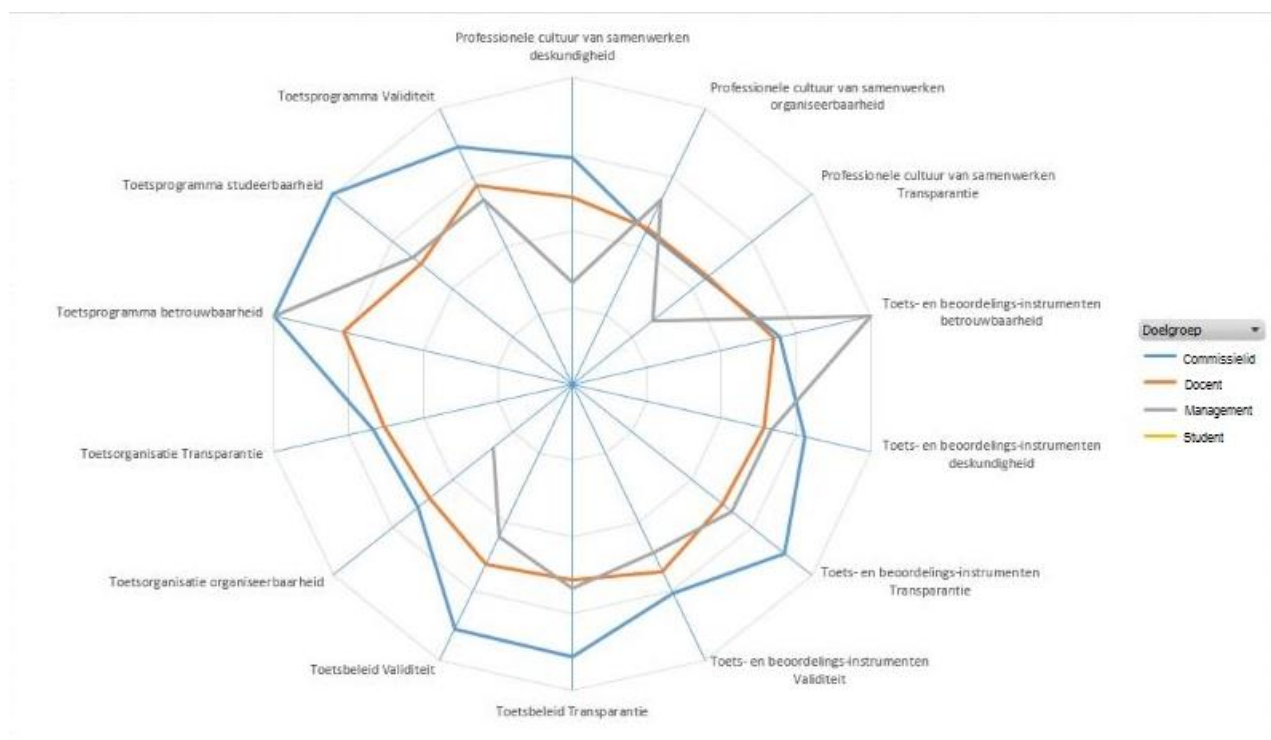
Fase 2: groepsinterview

Na de eerste fase worden de resultaten uit de vragenlijst geanalyseerd en in een spindigram geplaatst. Hierin wordt zichtbaar wat de gemiddelde score is voor alle standaarden, afgebeeld per groep stakeholders. Het diagram geeft een beeld van de kwaliteitscriteria (laatste kolom van tabel 2) waarbij doorgeklikt kan worden naar de onderliggende kwaliteitsstandaarden (tweede kolom van tabel 2) en lagen van de kwaliteitspiramide. Het voorbeeld in Figuur 2 laat het gemiddelde zien per kwaliteitscriterium, afgebeeld per groep stakeholders. Figuur 3 is een weergave van dezelfde resultaten gesorteerd op lagen van de piramide, uitgesplitst in kwaliteitscriteria. In het groepsinterview of groeps gesprek worden de standaarden uitgediept zodat iedereen daarvan een helder (gedeeld) beeld heeft. Daarna worden aan de hand van de verschillende spindigrammen die standaarden besproken waarover verschil van inzicht bestaat of waarvan de deelnemers aangeven dat deze verder uitgediept moet worden. Als afsluiting van het groeps gesprek wordt besproken wat de ontwikkelpunten zijn voor de opleiding. Er wordt een ontwikkelplan opgesteld. De resultaten uit het diagnose-instrument vormen een mooie aanvulling in het kwaliteitsportfolio.

Advies is dat het management samen met de medewerker met kwaliteitszorg in de portefeuille en andere medewerkers betrokken bij de opleiding het doel bepaalt en –na de diagnose- het ontwikkelplan vormgeeft. Het is raadzaam de analyse en het voeren van het gesprek te beleggen bij de medewerker belast met kwaliteitszorg (onafhankelijke voorzitter) en een toetsdeskundige van binnen en/of buiten de organisatie.



Figuur 2: voorbeeld van een spindigram, gesorteerd op kwaliteitscriteria.



Figuur 3: voorbeeld van een spindigram, gesorteerd op lagen van de piramide, uitgesplitst in kwaliteitscriteria.

Literatuur

Baartman, L.K.J., Kloppenburg, R. & Prins, F. Kwaliteit van toetsprogramma's. Verschenen in H. Van Berkel, A. Bax, & D. Joosten-ten Brinke (2013). *Toetsen in het Hoger Onderwijs*. Houten: Bohn Stafleu Van Loghum.

Berkel, H. J. M. van. (2015). *Toetsen in het hoger onderwijs*. Houten, Nederland: Bohn Stafleu van Loghum.

Cito RCEC, Maassen, N., Otter, D. den, Wools, S., Hemker, B., Straetmans, G., & Eggen, T. (2014). *Kwaliteit van toetsen binnen handbereik*. Geraadpleegd van http://www.cito.nl/~media/cito_nl/files/onderzoek%20en%20wetenschap/cito-eindrapportage-reviewstudie-kwaliteit-van-toetsen.ashx

Hogeschool van Amsterdam, Bruijns, V., & Kok, M. (2014). *Leidraad toetsen en beoordelen*. Geraadpleegd van <https://score.hva.nl/Bronnen/HvA%20Leidraad%20Toetsen%20en%20Beoordelen.pdf>

Jaspers, M., & Zijl, E. van. (2014). *Samenwerken aan toetskwaliteit in het Hoger Onderwijs*. Eindhoven, Nederland: Fontys Hogescholen.

Maassen, N. den Otter, D., Wools, S., Hemker, B., Straetmans, G. & Eggen, T. *Kwaliteit van toetsing binnen handbereik*. Geraadpleegd via www.cito.nl.

Sluijsmans, D., Peeters, A., Jakobs, L., & Weijzen, S. (2012). *De kwaliteit van toetsing onder de loep*. Tijdschrift OnderwijsInnovatie, 4, 17-25.

Sluijsmans, D., Joosten-Ten Brinke, D. & van Schilt-Mol, T.(2015). *Kwaliteit van toetsing onder de loep. Antwerpen: MAKLU uitgevers*.

NVAO (2014). *Beoordelingskader accreditatiestelsel hoger onderwijs Nederland*. Geraadpleegd via www.nvao.net.
Onderwijsraad (2015). *Kwaliteit in het Hoger Onderwijs*. Den Haag. Geraadpleegd via <https://www.rijksoverheid.nl/documenten/rapporten/2015/08/01/kwaliteit-in-het-hoger-onderwijs>.

Van der Vleuten, C.P.M., Schuwirth, L.W.T., Driessen, E.W., Dijkstra, J, Tggelaar, D. Baartman, L.K.J. & van Tartwijk, J. (2012). A model for pragmatic assesement fit for purpose. *Medical Teacher* , 34, 205-214

Wijnen, W., e.a. (1992). Te doen of niet te doen, Advies over studeerbaarheid van onderwijsprogramma's in het hoger onderwijs. Ministerie OC&W

Zuyd Hogeschool. (2015). *Toetskompas*. Geraadpleegd op 01-03-2017 via www.zuyd.nl