

Handboek toetsen en tentamens bij de Open Universiteit

1 juni 2011

Auteurs

Huub Breukers

Monique Doorten

Desirée Joosten-ten Brinke

Francisca Loth

George Moerkerke

Fred de Roode

Kathleen Schlusmans

©Open universiteit

Inhoudsopgave

<i>Inhoudsopgave</i>	3
Inleiding	1
I. Toetsen en tentamens aan de Open Universiteit	I-1
I-1. De functies van tentamens	I-1
I-2. Eisen te stellen aan toetsen en tentamens.....	I-2
I-3. Open versus gesloten vragen in toetsen en tentamens.....	I-3
I-4. De bijzondere verplichting	I-5
II. Actoren en afdelingen betrokken bij tentaminering	II-1
II-1. Speciale categorieën studenten	II-1
II-2. Tentamenlocaties.....	II-1
II-3. Overleg.....	II-1
II-4. Basisprocedures voor tentaminering.....	II-1
III. Het vaststellen van de vorm en inhoud van een tentamen	III-1
III-1. Het opstellen van een tentamenprofiel.....	III-1
III-2. Het opstellen van een toetsmatrijs.....	III-3
IV. Het construeren en screenen van gesloten vragen	IV-1
IV-1. Eisen te stellen aan gesloten vragen	IV-1
IV-2. Het kiezen van een vraagvorm	IV-8
IV-3. Vormen van gesloten vragen.....	IV-9
IV-4. Het construeren van gesloten vragen.....	IV-15
IV-5. Regels voor de lay-out van gesloten vragen.....	IV-17
V. Het construeren en screenen van open vragen	V-1
V-1. Eisen te stellen aan open vragen	V-1
V-2. Vormen van open vragen.....	V-4
V-3. Constructieregels voor standaard open vragen	V-7
V-4. Het gebruik van casussen.....	V-8
V-4.1. Wat is een casus?	V-9
V-4.2. Wat is het verschil tussen casussen op papier en met de computer?	V-9
V-4.3. Wat voor typen casussen zijn er ?	V-10
V-4.4. Wanneer wordt een casus gebruikt bij een tentamen?	V-10
V-4.5. Welke aanwijzingen zijn er voor het gebruik van casussen bij tentamens?	V-11
V-5. Richtlijnen voor het nakijken van open vragen	V-11
V-6. Het opstellen van een antwoordmodel	V-12
V-7. Screeningsregels voor open vragen	V-13

V-8.	Checklists	V-13
V-9.	Richtlijnen voor de lay-out van open vragen	V-15
VI.	Het samenstellen van een tentamen en het bepalen van de voorlopige cesuur	VI-1
VI-1.	Stappen in het samenstellingsproces van tentamens	VI-1
VI-2.	Screeningsregels voor tentamens	VI-6
VII.	Analyse en definitieve cesuur.....	VII-1
VII-1.	Inleiding	VII-1
VII-2.	Informatie over het gehele tentamen.....	VII-2
VII-3.	Informatie over de afzonderlijke vragen	VII-5
VII-4.	De vraaganalyse.....	VII-6
VII-5.	Het vaststellen van de definitieve cesuur.....	VII-11
VIII.	CBI-tentaminering	VIII-1
VIII-1.	Ontstaan en systematiek van Computer gebaseerd Individueel tentamineren (CBI) ..	VIII-1
VIII-2.	De fasen in het CBI-tentamentraject.....	VIII-2
IX.	Mondeling tentamineren	IX-1
IX-1.	Wat is mondeling tentamineren?.....	IX-1
IX-2.	De voorbereiding	IX-3
IX-3.	De feitelijke afname en beoordeling	IX-7
IX-4.	De uitslagbepaling	IX-9
IX-5.	Online mondeling tentamen: Teletentamen	IX-10
X.	Bijzondere verplichtingen.....	X-1
X-1.	Organisatie van bijzondere verplichtingen	X-2
XI.	Competentiegericht toetsen.....	XI-1
XI-1.	Inleiding	XI-1
XI-2.	Het toetsspecificatieplan	XI-2
XI-3.	Het opdrachtontwerpschema.....	XI-3
XI-4.	De opdrachten	XI-7
XI-5.	De beoordelingsformulieren	XI-11
XI-6.	Het registreren van toetsinformatie.....	XI-13
XII.	Index	15

Inleiding

Kwaliteit van toetsen en beoordelen heeft bij de Open Universiteit altijd hoog op de agenda gestaan. Al vanaf de oprichting van de OU in 1984 behoort tentaminering, certificering en diplomering tot een van de belangrijkste primaire processen. Ook bij de accreditatie van opleidingen is de kwaliteit van tentamens een belangrijk speerpunt.

Dit handboek 'Toetsen en tentamens bij de Open Universiteit' is een vernieuwde uitgave van de zogenoemde katernenreeks die voor het eerst het licht zag in 1988. Huub Breukers, Francisca Loth en Kathleen Schlusmans schreven in het project 'Toetsworkshop' de eerste zeven katernen. In 1991 werd de eerste aanvulling geschreven door Desirée Joosten-ten Brinke. Deze katern had betrekking op de sys-tentaminering. In 1992 volgde de tweede aanvulling die betrekking had op mondeling tentamineren, geschreven door Monique Doorten. De laatste inhoudelijk katern die er aan toegevoegd werd ging over Competentiegericht toetsen, geschreven door George Moerkerke en Monique Doorten.

Het domein van toetsen en beoordelen staat echter niet stil en vanaf 2008 is de Open universiteit de weg ingeslagen van computergebaseerd toetsen (CBT). Was sys al grotendeels in het voor- en natraject een computergebaseerd systeem, met CBT kunnen de studenten hun tentamens digitaal, direct achter de computer afleggen. Deze verandering bood het CBT-projectteam de mogelijkheid om alle katernen door te lopen en aan te passen aan de huidige ontwikkelingen. De grootste aanpassingen betroffen de katern over sys-tentaminering, die daardoor niet alleen van inhoud, maar ook van naam gewijzigd is: CBT-tentaminering. Het invoeren van CBT werkt ook door in de andere secties. Daarnaast is door George Moerkerke een bijdrage toegevoegd over het werken met casussen. Tot slot is er een nieuwe katern toegevoegd onder de naam Bijzondere verplichtingen.

We hopen met dit nieuwe handboek antwoorden te geven op toetskundige vragen die leven bij docenten, examinatoren, leden van de examencommissie en cursusontwikkelaars. De uitgave is opgesplitst in aparte secties zodat deze afzonderlijk onderhouden kunnen worden.

Voor de totstandkoming van deze versie wil ik met name Fred de Roode bedanken die deze uitgave heeft geredigeerd, de leden van het CBT-project en deelprojectleider en leden van het deelproject Toetskwaliteit die feedback hebben gegeven op eerdere versies.

Desirée Joosten-ten Brinke, Heerlen, juni 2011

Dit handboek is geactualiseerd op 1 april 2016

George Moerkerke, Heerlen, mei 2016

Dit handboek is geschreven door de toetsdeskundigen van de Open universiteit. Zoveel mogelijk is daarbij geprobeerd aan te sluiten bij de regelingen die door de CvE zijn vastgesteld en die na te lezen zijn in de OER en op de website www.ou.nl.

Bij verschillen in regels tussen dit handboek en de officiële regelingen gelden uiteraard de officiële regels.

I. Toetsen en tentamens aan de Open Universiteit

Deze publicatie is bedoeld voor een ieder die zich bezig houdt met of belangstelling heeft voor toetsen (en in het bijzonder tentamens) aan de Open Universiteit (OU), dus zowel de cursusteamleiders als de examinatoren, zowel de auteurs als de vraagconstructeurs en zowel de beleidsmakers als de beleidsuitvoerders.

Het doel van deze sectie is een inleiding te geven op de problematiek rondom tentamens. In deze sectie komen een aantal aspecten aan de orde, die in de daarop volgende secties worden uitgewerkt, zoals soorten toetsen en soorten vragen en eisen te stellen aan toetsen, gesloten vragen en open vragen.

I-1. De functies van tentamens

Toetsen kunnen zowel summatief als formatief ingezet worden in het onderwijs.

Summatieve functie

Summatieve toetsing heeft primair ten doel na te gaan of studenten de leerdoelen van een bepaalde cursus voldoende beheersen om het certificaat te verkrijgen. Soms is het slagen voor het tentamen echter niet voldoende voor het behalen van het certificaat en moet ook voldaan zijn aan een zogenaamde bijzondere verplichting. Bij bepaalde cursussen is het mogelijk dat geen tentamensessie wordt georganiseerd, maar dat het tentamen geheel uit een bijzondere verplichting bestaat.

Het tentamen en (eventueel) de bijzondere verplichting vormen dus de formele, afsluitende evaluatie van studenten.

Bij de OU worden verschillende toetsvormen voor summatieve doeleinden gebruikt:

- Schriftelijke groepstentamens
- Computergebaseerde tentamens, individueel (CBI) of groepsgewijs (CBG)
- Mondelinge tentamens
- Bijzondere verplichtingen (scripties, opdrachten, practica)

Tentamens van de OU kunnen zowel schriftelijk als digitaal afgelegd worden.

Bij enkele cursussen kunnen studenten bonuspunten halen die mee tellen bij het eindresultaat. Ook kan een tentamen uit deelttoetsen bestaan. De deelttoetsen kunnen compensatoir zijn of niet.

Formatieve functie

In het (elektronisch) cursusmateriaal is doorgaans ook een aantal toetsen opgenomen dat tot doel heeft om het studeergedrag van de studenten bij te sturen, hen bewust te maken van hun vorderingen en hen voor te bereiden op het tentamen. Deze toetsen hebben vooral een didactische functie en ondersteunen het leerproces.

In het OU-cursusmateriaal kunnen verschillende soorten formatieve toetsen worden onderscheiden: zelftoetsen, herhalingstoetsen en eindtoetsen, alle drie met als doel voortgangstoetsing. Dit onderscheid tussen de verschillende toetsen is afgeleid uit het zgn. leereenheden-cursusmodel; in andere cursusmodellen kan het onderscheid tussen zelftoetsen en herhalingstoetsen vervagen.

Zelftoetsen

Zelftoetsen zijn relatief korte toetsen die zijn opgenomen aan het einde van elke leereenheid. Ze geven studenten onmiddellijk na bestudering van een afgerond studiegeheel de gelegenheid na te gaan of zij de leerdoelen bereikt hebben: zelftoetsen geven dus een indicatie of de studie effectief is geweest. Een student moet een zelftoets onmiddellijk na het bestuderen kunnen beantwoorden en moet ook direct feedback kunnen krijgen. De (bespreking van de) antwoorden op de zelftoets worden dan ook altijd in het cursusmateriaal opgenomen. Omdat een zelftoets zeer nauw aansluit bij de leereenheid, ligt het voor de hand dat de auteurs van die leereenheid ook de zelftoets ontwikkelen.

Herhalingstoetsen

Herhalingstoetsen bestrijken een groter gebied dan zelftoetsen. Ze hebben meestal betrekking op meer onderdelen van de cursus en geven de studenten de kans om na te gaan of zij ook de leereenheid-overstijgende doelen bereikt hebben. Op deze wijze kunnen studenten regelmatig inzicht krijgen in hun vorderingen. De herhalingstoets heeft meestal dezelfde vorm en ongeveer dezelfde moeilijkheidsgraad als het tentamen, zij het dat de herhalingstoets een kleiner domein aan leerdoelen bestrijkt. Omdat herhalingstoetsen uiteraard aan het eind van een herhalingseenheid worden opgenomen, worden ze vaak ontwikkeld door de auteur van die herhalingseenheid. Het is echter ook mogelijk herhalingstoetsen samen te stellen uit hetzelfde vraagbestand waaruit ook de eindtoets en de tentamens worden gegenereerd.

Eindtoetsen

De eindtoets ('proeftentamen'), opgenomen aan het eind van de cursus, behoort zoveel mogelijk gelijkwaardig te zijn aan het (schriftelijk) tentamen. De eindtoets vormt een soort generale repetitie voor het eigenlijke tentamen.

I-2. Eisen te stellen aan toetsen en tentamens

De belangrijkste functie van toetsen en tentamens is om op een zo valide en betrouwbaar mogelijke wijze te bepalen of de student de leerdoelen van de cursus (of een onderdeel daarvan) bereikt heeft. Daarom worden aan toetsen en tentamens een aantal algemene eisen gesteld. Afhankelijk van de specifieke functie van de toets zal een bepaalde eis meer of minder stringent zijn. Achtereenvolgens zullen worden behandeld de eisen van:

1. Objectiviteit
2. Betrouwbaarheid
3. Validiteit
4. Efficiëntie
5. Transparantie

1. Objectiviteit

Onder objectiviteit verstaat men dat de bepaling van de toetsscore onafhankelijk is van de persoon van de beoordelaar. Bij een toets, waarbij het antwoordmodel in het materiaal is opgenomen en die dus door de student zelf beoordeeld wordt, mag het niet zo zijn dat eenzelfde antwoord door de ene student wel en door de andere student niet juist gerekend wordt.

Ingeval van een tentamen betekent dit dat alle correctoren eenzelfde respons op een vraag of opdracht op dezelfde wijze moeten scoren. Dit impliceert dat volledig duidelijk moet zijn wat wel en wat niet als een juist antwoord beschouwd moet worden en dat vakinhoudelijk deskundigen het eens moeten zijn over het juiste antwoord of, in geval van vragen waarbij meerdere juiste oplossingen mogelijk zijn, over de criteria waaraan een juist antwoord moet voldoen.

2. Validiteit

Onder validiteit verstaat men de mate waarin de vragen in een toets of tentamen representatief zijn voor de leerdoelen die in de cursus of onderdeel daarvan worden nagestreefd. Dat wil zeggen dat alle vragen betrekking moeten hebben op de leerdoelen en dat de vragen in de toets of het tentamen evenwichtig gespreid moeten zijn over de verschillende (groepen van) leerdoelen.

3. Betrouwbaarheid

Onder de betrouwbaarheid van een toets verstaat men de mate waarin de meting vrij is van niet bedoelde, toevallige invloeden. Concreet wil dit zeggen in hoeverre een student bij het opnieuw afleggen van een vergelijkbare toets of tentamen onder gelijke omstandigheden eenzelfde score zou

behalen. De behaalde score mag dus niet afhangen van toevalligheden zoals de keuze van de onderwerpen of een 'black out' van de student tijdens de afname.

De betrouwbaarheid van een toets of een tentamen hangt ondermeer samen met de lengte ervan. Een langere toets of tentamen is over het algemeen betrouwbaarder omdat de toets dan een grotere steekproef uit het vraagdomein vormt. Het feit dat een student nèt dat ene onderwerp niet goed beheerst of nèt een 'black out' heeft bij één vraag speelt dan een kleinere rol dan bij een kortere toets. Studenten krijgen ingeval van een langere toets meer kansen om te demonstreren dat ze de desbetreffende leerdoelen gerealiseerd hebben.

4. Efficiëntie

Onder efficiëntie verstaat men dat de kwaliteit van de informatie die een toets of een tentamen oplevert in verhouding moet staan tot de inspanning die geleverd is bij het ontwikkelen, afnemen en corrigeren ervan. Ingeval van tentamens aan de OU betekent dit onder meer dat de benodigde correctietijd in een redelijke verhouding moet staan tot de meerwaarde van de informatie die dit oplevert. Gebruik in een tentamen dus geen open vragen als hetzelfde ook met gesloten vragen getoetst kan worden.

5. Transparantie

Transparantie in toetsing betekent dat studenten voor aanvang van de toets weten waar zij aan toe zijn, zodat zij zich zo goed mogelijk kunnen voorbereiden. Studenten zijn voor aanvang van de toets in ieder geval op de hoogte van:

- Het tijdstip waarop de toets plaatsvindt;
- De vorm en omvang van de toets;
- De beschikbare tijd
- De toegestane hulpmiddelen;
- De normering en hoe de voldoende/onvoldoende grens (cesuur) bepaald wordt;
- De beoordelingscriteria.
- Bij bijzondere verplichtingen: de beoordelingsformulieren.

I-3. Open versus gesloten vragen in toetsen en tentamens

Bij de constructie van toetsen en tentamens heeft men de keuze uit verschillende soorten vragen. De belangrijkste keuze waarvoor men zich gesteld ziet is: kies ik voor open of gesloten vragen of voor een combinatie van beide?

De literatuur geeft weinig houvast bij deze keuze. Het staat niet op voorhand vast dat de ene vorm geschikter zou zijn dan de andere om bepaalde leerdoelen te toetsen. Wij geven op deze plaats dan ook alleen enkele overwegingen bij het maken van deze keuze.

Open vragen zijn vragen waarbij studenten zelf een antwoord moeten formuleren: een redenering, argumentatie of een verklaring.

Gesloten vragen zijn vragen waarbij studenten niet zelf een antwoord moeten formuleren, maar waarbij ze uit tenminste twee alternatieven het juiste alternatief moeten kiezen. Binnen de categorie gesloten vragen kan men een onderscheid maken tussen *meerkeuzevragen* en *juist-onjuist vragen*. Deze kan men beschouwen als basisvormen. Varianten hierop zijn mogelijk in het kader van *Computer gebaseerd toetsen, CBG en CBI* (zie sectie IV).

We zetten de sterke en zwakke punten van open en gesloten vragen op een rij en geven enkele richtlijnen die men kan gebruiken bij het maken van een keuze voor open of gesloten vragen.

Bij de keuze van de tentamenvorm spelen naast inhoudelijke overwegingen ook *procedurele* en *financiële* overwegingen een rol. Zo zal men zich bij kleine groepen studenten bijvoorbeeld moeten afvragen of het de moeite en de kosten loont om een tentamen met gesloten vragen te ontwikkelen.

Kiest men voor open vragen dan zal men rekening moeten houden met correctietijd, organisatie van de correctie en de tijdige beschikbaarheid van correctoren. Verder is het bij gebruik van open vragen wenselijk om een *dubbele correctie* uit te voeren. Dit is van belang om tot een zo groot mogelijke objectivering te komen. Daarnaast garandeert het de continuïteit van de onderwijsuitvoering omdat het de kwetsbaarheid van het systeem vermindert en de overdraagbaarheid van taken vergroot.

Maar: dit alles kost extra tijd en geld en vereist speciale computerprogrammatuur. Kortom: in de tentamenpraktijk zullen altijd afwegingen en keuzes gemaakt moeten worden.

Of men bij eind- en herhalingstoetsen open of gesloten vragen (dan wel een combinatie van beide) hanteert, hangt voor een belangrijk deel samen met de gekozen tentamenvorm. De eindtoets en in grote lijnen ook de herhalingstoetsen, bereiden voor op het tentamen en dienen qua samenstelling gelijkwaardig te zijn. Hiervoor gelden dus dezelfde regels als bij het tentamen: alleen gesloten vragen tenzij de leerdoelen dit niet toelaten. Bij de zelftoets is de relatie met het tentamen wat losser. Men zou in de zelftoets natuurlijk zoveel mogelijk gesloten vragen kunnen gebruiken. Men zou echter ook kunnen opteren voor het gebruik van open vragen in de zelftoets omdat deze toets vooral een sturende functie heeft en omdat er bij een zelftoets toch een uitgebreid antwoordmodel ontwikkeld moet worden aan de hand waarvan de studenten gerichte feedback krijgen.

Gesloten vragen

We zetten eerst de voor- en nadelen van gesloten vragen op een rij.

Voordelen van gesloten vragen:

- omdat het juiste antwoord op voorhand vastligt en voor alle studenten gelijk is, is de scoring volledig objectief
- de scoring kan geautomatiseerd verlopen: dit levert tijdswinst op ingeval van relatief grote groepen studenten
- In vergelijking met open vragen kunnen studenten meer vragen oplossen in eenzelfde tijdsbestek en dit verhoogt in het algemeen de betrouwbaarheid van de toets of het tentamen.
- Daarnaast kan met een groter aantal vragen een betere afspiegeling van de leerdoelen gerealiseerd worden en neemt daardoor ook de inhoudsvaliditeit toe.

Nadelen van gesloten vragen:

- bepaalde leerdoelen zoals het produceren van argumenten, oplossingen of verklaringen, lenen zich niet of minder makkelijk voor toetsing met gesloten vragen
- het vraagt veel tijd en creativiteit van de ontwikkelaars om relevante gesloten vragen te maken, vooral voor het toetsen van leerdoelen die betrekking hebben op redeneervaardigheid (inzicht en toepassing) en daarom is het gevaar groot dat men in een toets of tentamen met gesloten vragen minder vragen opneemt ten aanzien van deze leerdoelen en het grootste deel toespitst op de makkelijker te toetsen leerdoelen die betrekking hebben op de reproductie van feiten, begrippen en regels.

Open vragen

Voordelen van open vragen:

- bijna alle leerdoelen kunnen met open vragen getoetst worden, vooral leerdoelen die creatieve productie vergen
- open vragen sluiten in het algemeen beter aan bij de wijze waarop studenten een cursus bestuderen en bij de opgaven en studeeropdrachten die in het cursusmateriaal verwerkt zijn
- studenten voelen zich vaak meer 'serieus' genomen met dit type vragen, omdat ze dan het gevoel hebben een eigen inbreng te hebben.

Nadelen van open vragen:

- vergen relatief veel correctietijd en de correctie kan niet geautomatiseerd worden
- laten zich moeilijker objectief scoren
- een (uitgebreid) antwoordmodel is nodig, het opstellen ervan kost tijd en zelfs dan kunnen correctoren nog in de beoordeling van de antwoorden verschillen
- bij correctie en scoring van lang-antwoord-vragen bestaat het gevaar dat correctoren beïnvloed worden door factoren als schrijfstijl en handschrift
- het kost relatief veel tijd om lang-antwoord-vragen te beantwoorden; in een toets of tentamen kunnen daarom slechts een beperkt aantal vragen worden opgenomen en dit komt de betrouwbaarheid en de inhoudsvaliditeit doorgaans niet ten goede.

Wanneer zijn open vragen echt noodzakelijk?

- indien het erom gaat dat studenten zelf moeten kunnen formuleren en de formulering op zich wordt getoetst
- indien de creativiteit of inventiviteit van studenten bij het vinden van oplossingen moet worden getoetst
- indien studenten informatiebronnen moeten raadplegen en als het weten van de weg in de informatiebronnen van belang is (bijvoorbeeld het kunnen opzoeken van een toepasselijk wetsartikel)
- indien men wil toetsen of studenten een sluitende argumentatie kunnen opbouwen of als men wil nagaan of en hoe zij een bepaalde methode toepassen.

I-4. De bijzondere verplichting

Soms is het niet mogelijk om in een gestandaardiseerde tentamensessie van drie uur alle leerdoelen te toetsen, die men zou willen toetsen. Vandaar dat de mogelijkheid bestaat bijzondere verplichtingen op te nemen naast en soms in plaats van schriftelijke of gedigitaliseerde tentamens. Bijzondere verplichtingen zijn activiteiten die een student met goed gevolg moet verrichten en die, samen met het slagen voor het schriftelijke of digitale tentamen of op zich zelf staand, een voorwaarde vormen voor het behalen van het certificaat.

In sectie X wordt de bijzondere verplichting verder beschreven. In sectie XI wordt de rol van toetsen in competentiegericht onderwijs beschreven.

II. Actoren en afdelingen betrokken bij tentaminering

Voor een vlot en correct verloop van het tentamenproces aan de OU zijn heldere procedures en richtlijnen voor alle betrokkenen van groot belang. Hier wordt in het kort aangegeven welke afdelingen binnen de OU een rol spelen bij de tentaminering en welke richtlijnen examinatoren moeten naleven. Voor verschillende soorten toetsen en bepaalde doelgroepen studenten zijn specifieke procedures. Deze zijn na te lezen op www.ou.nl/procedures. De examenreglementen worden jaarlijks vastgesteld door de CvE's. De tentamen- en examenprocedures worden jaarlijks vastgesteld door de GSO (Gemeenschappelijke Service Organisatie). Het Centraal kader voor de Commissie voor de Examens wordt jaarlijks vastgesteld door het CvB (College van Bestuur). De Onderwijs en Examenregeling worden vastgesteld door het CvB.

II-1. Speciale categorieën studenten

Naast de 'reguliere' OU-studenten worden speciale categorieën studenten onderscheiden, namelijk

- gedetineerden
- studenten buiten Nederland en Vlaanderen
- studenten met een functiebeperking
- studenten die onder afspraken met een contractpartij gelijke of gelijkwaardige tentamens afleggen

II-2. Tentamenlocaties

- studiecetra (zie voor een overzicht <https://www.ou.nl/web/studiecentra>)
- penitentiaire inrichtingen (voornamelijk in Nederland en België)
- thuislocatie in geval van studenten met een functiebeperking

Buiten Europa:

- ambassades/consulaten
- penitentiaire inrichtingen
- Aruba, Curaçao, St. Maarten en de BES-gemeenten (Bonaire, St. Eustatius en Saba)
- Militaire bases
- aan boord van marineschepen (niet bij de koopvaardij)

II-3. Overleg

Tijdens het academisch jaar vindt overleg plaats tussen de CvE en de decaan van de betreffende faculteit om afspraken te maken rondom tentaminering en het afgelopen jaar te evalueren.

Voordat een cursus in productie gaat vindt er tussen het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek en de faculteiten overleg plaats inzake de tentamenvorm, tentamendata en toegestane hulpmiddelen bij tentamens. Ook wordt in dit overleg de laatste tentamenmogelijkheid van een uitlopende cursus bepaald.

II-4. Basisprocedures voor tentaminering

Afhankelijk van toetsvormen zijn er specifieke procedures voor aanlevering, correctie en het vaststellen van de uitslag. De basisprocedure staat hieronder beschreven. Informatie over termijnen is ook terug te vinden op de website: www.ou.nl/tentamen.

Schriftelijke tentamens en CBG-tentamens

Aanleveren van schriftelijke tentamens. Schriftelijke tentamens dienen tijdig te worden aangeleverd,

d.w.z. drie maanden vóór iedere tentamenronde ten behoeve van de opmaak door de faculteiten en de definitieve versie zes weken vóór iedere tentamenronde bij het cluster Tentaminering van Administratie en onderwijslogistiek. Tevens dient het antwoordmodel (ook van 'geheime' tentamens) en het vaststellingsverslag van de inhoud zes weken vóór de tentamenafname bij het cluster Tentaminering van Administratie en onderwijslogistiek te worden aangeleverd.

Correctie en vaststelling uitslag van schriftelijke tentamens/CBG-tentamens. Nadat de tentamenzittingen van de schriftelijke tentamens hebben plaatsgevonden, moeten de open vragen binnen vijf werkdagen na ontvangst van het correctiepakket gecorrigeerd zijn en moeten de behaalde scores via Optisch Leesbare Formulieren (OLF's) aangeleverd worden.

De CBG-tentamens in open vraagvorm kunnen direct na afname door de corrector via de computer binnen vijf werkdagen worden gescoord.

De examinerator dient het vaststellingsverslag van de tentamens zowel in open- als in gesloten vraagvorm binnen vier weken na de tentamendatum in te leveren bij het cluster Tentaminering van Administratie en onderwijslogistiek.

Bijzondere verplichtingen

De resultaten van bijzondere verplichtingen (scripties, afstudeeropdrachten, practica) mogen alleen worden afgegeven binnen de periode waarin de cursus in exploitatie is en binnen de inschrijftermijn van de student. De resultaten dienen op een resultaatformulier doorgegeven te worden aan het cluster Tentaminering van Administratie en onderwijslogistiek. Dit resultaatformulier moet volledig ingevuld zijn, d.w.z. voorzien van studentnummer, studentnaam, cursuscode, cursusomschrijving, adres, inleverdatum, naam examinerator, datum ondertekening door examinerator, en resultaat.

Mondelinge tentamens

Aanmeldingen voor mondelinge tentamens met van te voren vastgestelde data mogen uitsluitend verlopen via de het cluster Tentaminering van Administratie en onderwijslogistiek. Voor mondelinge tentamens zonder van te voren vastgestelde data worden rechtstreeks afspraken gemaakt tussen student en examinerator.

De afname van mondelinge tentamens ten behoeve van speciale categorieën studenten is uitsluitend toegestaan na goedkeuring door de CvE van de opleiding.

Ook bij mondelinge tentamens dient het eerder genoemde resultaatformulier te worden ingevuld en geldt dezelfde procedure als beschreven bij de bijzondere verplichtingen.

CBI-tentamens

Met betrekking tot de tentaminering via CBI gelden de volgende algemene regels:

- tijdig aanleveren van het vragenbestand volgens vastgesteld traject, na overleg met de het cluster Tentaminering van Administratie en onderwijslogistiek ;
- aanleveren van het vragenbestand uitsluitend via vastgestelde redactionele richtlijnen en kwaliteitseisen;
- beoordelen van de noodzaak van acute/periodieke mutaties, binnen de gestelde termijn, (mede) op grond van statistische analyses van het vragenbestand en op basis van eventuele reacties van studenten.

De eerste analyse vindt plaats drie maanden na het moment dat het desbetreffende vraagbestand operationeel is geworden en daarna eens per jaar. Bij deze analyses wordt een toelichting op de statistische gegevens meegeleverd (voor analyse zie sectie VII.). Mocht de analyse aanleiding geven tot het aanbrengen van wijzigingen in het vragenbestand, dan dient daartoe het formulier 'Verzoek herziening itembank' te worden ingediend bij de het cluster Tentaminering van Administratie en onderwijslogistiek.

III. Het vaststellen van de vorm en inhoud van een tentamen

Deze sectie is bedoeld voor degenen die de vorm en inhoud van een tentamen moeten bepalen, namelijk de examinatoren. Het doel van deze sectie is hun richtlijnen te geven voor het bepalen van de vorm en inhoud van het tentamen.

In een vroeg stadium van de cursusontwikkeling zal de . examiner - alvorens tentamenvragen of - opdrachten te (laten) ontwikkelen - moeten bepalen hoe het tentamen er in grote lijnen uit zal zien:

- Welke leerstofonderwerpen moeten in het tentamen aan bod komen?
- Welke leerdoelen zullen getoetst worden in het tentamen?
- Is sprake van alleen gesloten vragen, alleen open vragen of een combinatie van beide?
- Hoeveel gesloten vragen/open vragen moeten in het tentamen opgenomen worden?
- Is sprake van een bijzondere verplichting?
- Is sprake van multimedieagebruik?

In deze sectie zullen we aangeven hoe men op verantwoorde wijze kan komen tot beantwoording van de bovenstaande vragen en daarmee tot het bepalen van de vorm en inhoud van het tentamen.

Bij het afnemen van tentamens is het de bedoeling de prestaties van studenten te toetsen aan een criterium om een uitspraak te kunnen doen over het al dan niet beheersen van de leerdoelen en op basis daarvan al dan niet een certificaat toe te kennen. Het gehanteerde criterium is dus de mate waarin de student de vooropgestelde leerdoelen heeft gerealiseerd. De relatie tussen leerdoelen en tentamenvragen is zeer belangrijk: de leerdoelen geven aan wat de student na afloop van de cursus moet kennen en kunnen en het tentamen is het instrument dat meet in welke mate die leerdoelen gerealiseerd zijn. Anders gezegd: om de (inhouds)validiteit van een tentamen te waarborgen dienen inhoud en niveau van de tentamenvragen een adequate weerspiegeling van de voor de cursus geformuleerde leerdoelen te zijn. Dit betekent enerzijds dat de te meten vakinhoud (dus de leerstofonderdelen) en het gewenste beheersingsniveau tot uitdrukking dienen te komen in de tentamenvragen en anderzijds dat in het totaal van de tentamenvragen de leerdoelen evenwichtig vertegenwoordigd moeten zijn. Omdat een tentamen wordt afgenomen over een gehele cursus is het soms niet mogelijk om alle leerdoelen te toetsen c.q. over elk leerdoel in het tentamen een vraag te stellen. Een tentamen zal dan ook slechts een deel van alle leerdoelen kunnen toetsen. Men zal dus keuzes moeten maken, zodanig dat een representatief en evenwichtig tentamen ontstaat dat zo goed mogelijk voldoet aan de eisen van (inhouds)validiteit en betrouwbaarheid.

In 3.1 worden richtlijnen gegeven voor het opstellen van een tentamenprofiel en in 3.2 van een toetsmatrijs.

III-1. Het opstellen van een tentamenprofiel

Voor elke cursus moet door de examiner een tentamenprofiel worden opgesteld dat op hoofdlijnen aangeeft hoe de cursus zal worden getentamineerd. Indien een examiner de wijze van tentamineren tussentijds wil veranderen, moet het tentamenprofiel worden geactualiseerd. Wijzigen kan pas na toestemming van de CvE.

Het tentamenprofiel begint met een samenvatting van de leerdoelen: het cursusdoel. Daarna volgt een keuze van de toetsvorm of combinatie van toetsvormen. Deze keuze dient uiteraard gebaseerd te zijn op de geschiktheid van de toetsmethode(n) voor het beoordelen van de studentprestaties met betrekking tot de verschillende leerdoelen. (In sectie I zijn de voor- en nadelen van de verschillende vraag- en vraagvormen besproken en de examiner vindt daar argumenten om zijn keuze te onderbouwen.)

Hieronder volgt een voorbeeld van een tentamenprofiel:

<p>Tentamenprofiel Cursus XXXXXX</p> <p>(moet eenmalig worden ingevuld bij de start van de exploitatie van een cursus en bij revisies of andere wijzigingen en bij de go-beslissing aangeleverd)</p>
<p>Cursusnaam:</p> <p>Cursuscode:</p> <p>Datum start exploitatie:</p>
<p>1. De afsluiting van de cursus bestaat uit:</p> <p>A. Tentamentype: (kruis aan)</p> <ul style="list-style-type: none">• computer gebaseerd tentamen (CBG of CBI)• regulier schriftelijk tentamen• mondeling tentamen• bijzondere verplichting• een combinatie van bovenstaande, namelijk: Verplichte volgorde (indien aanwezig): <p>B. Vraagtype (kruis aan)</p> <p>1. Open</p> <ul style="list-style-type: none">• Standaard open vraag• Invulvraag• Open vraag numeriek <p>2. Gesloten</p> <ul style="list-style-type: none">• Standaard meerkeuzevraag (hoeveel alternatieven?)• Stellingvraag• Meerkeuzevraag met pull down menu• Matrixvraag• Matchingvraag• Standaard meerkeuzevraag met videofragment• Standaard meerkeuzevraag met audiofragment• Meerkeuzevraag met afbeeldingen• Drag and drop• Rangorde-vraag• Tekst match-vraag• Meerdere antwoorden juist-vraag <p>C. Compensatorische regeling</p> <ul style="list-style-type: none">• Is de compensatorische regeling van toepassing op basis van de samenstellende delen? ja/nee <p>D. Bonuspunten</p> <ul style="list-style-type: none">• Worden bonuspunten toegekend? Ja/nee Zo ja, hoe tellen deze mee in de eindwaardering?
<ul style="list-style-type: none">○ 2. Toegestane hulpmiddelen: (N.b.: standaard is een 'schoon' Nederlands woordenboek toegestaan) Ja, namelijk:

<p>3. Correctie door:</p> <ul style="list-style-type: none"> • examiner • correctoren onder verantwoordelijkheid van de examiner
<p>4. Bijzonderheden (indien van toepassing):</p>
<p>Naam examiner:</p> <p>Handtekening examiner:</p> <p>Naam plaatsvervangend examiner:</p> <p>Datum:</p>

Eerder is gesteld dat een examiner keuzes moet maken om op een verantwoorde wijze inhoud en vorm van een tentamen te bepalen. Het hulpmiddel hierbij is het opstellen van een toetsmatrijs.

III-2. Het opstellen van een toetsmatrijs

Onder een toetsmatrijs verstaat men een tweedimensionale tabel waarin de inhoud van de te toetsen cursus wordt afgezet tegen het beheersingsniveau. De inhoud en het gewenste beheersingsniveau van die inhoud zijn neergelegd in de leerdoelen van de cursus en moeten tot uitdrukking komen in de tentamenopgaven. Hoe zo'n toetsmatrijs er uit kan zien en hoe deze opgesteld moet worden, beschrijven we aan de hand van de volgende drie kernvragen:

1. Welke leerstofonderdelen en beheersingsniveaus moeten in het tentamen aan bod komen?
2. Hoeveel procent van de opgaven in het tentamen moet betrekking hebben op de opgesomde leerstofonderdelen en beheersingsniveaus?
3. Hoeveel opgaven van een bepaalde categorie moeten per leerstofonderdeel en beheersingsniveau in het tentamen voorkomen?

Kernvraag 1: Welke leerstofonderdelen moeten in het tentamen aan bod komen?

Voorbeeld bij een cursus met vier leereenheden

Vakinhoud	
Leereenheid 1: Theorie van Piaget	
Leereenheid 2: Theorie van Freud	
Leereenheid 3: Sociale leertheorie	
Leereenheid 4: Informatieverwerkende benadering	

Zoals eerder aangegeven kunnen we aan de leerdoelen van een cursus een vakinhoudelijke component onderscheiden. Deze vakinhoudelijke component laat zich meestal eenvoudig benoemen: de concrete leerstof ofwel de behandelde leerstofonderdelen. Vanuit de opsomming van leerdoelen bij elke leereenheid van een cursus kan men zo komen tot een overzicht van leerstofonderdelen voor de totale cursus. Op de verticale as van de toetsmatrijs moeten die leerstofonderdelen vermeld worden die in het tentamen aan bod moeten komen. Dit impliceert dat een keuze gemaakt moet zijn omtrent welke onderdelen wel en welke niet in het tentamen aan de orde moeten komen. Afhankelijk van de wens om

gedetailleerd of meer globaal te werk te gaan kunnen de verschillende leerstofonderdelen ook geclusterd worden tot grotere gehelen: op de 'vakinhoudelijke'-as van de toetsmatrijs worden bijvoorbeeld blokken of leereenheden dan wel hoofdstukken/paragrafen opgesomd. Echter: hoe nauwkeuriger de vakinhoud gespecificeerd is, des te beter het overzicht van de te toetsen onderdelen.

Kernvraag 2: Hoeveel procent van de vragen in het tentamen moet betrekking hebben op de opgesomde leerstofonderdelen?

Een tentamen bestaat uit een x-aantal vragen (100%). Per leerstofonderdeel moet nu worden aangegeven hoeveel vragen van het tentamen (uitgedrukt in een percentage) hierop betrekking moeten hebben. Deze percentages geven dus de verhouding weer van het aantal vragen waaruit het tentamen per leerstofonderdeel moet worden samengesteld. Het moge duidelijk zijn dat wanneer bepaalde leerstofonderdelen van een cursus belangrijker en/of omvangrijker zijn, dit ook tot uitdrukking moet komen in het aantal vragen dat over die onderdelen in het tentamen wordt opgenomen

Voorbeeld bij een cursus met vier leereenheden

Vakinhoud	Aantal
Leereenheid 1: Theorie van Piaget	25%
Leereenheid 2: Theorie van Freud	10%
Leereenheid 3: Sociale leertheorie	25%
Leereenheid 4: Informatieverwerkende benadering	40%
Totaal aantal vragen	100%

Kernvraag 3: Hoeveel vragen van een bepaalde categorie moeten per leerstofonderdeel in het tentamen voorkomen?

Elk leerdoel kent naast een vakinhoudelijke component ook een gedragscomponent. Deze gedragscomponent (denkgedrag/oplosgedrag) is in tegenstelling tot de inhoudscomponent minder eenvoudig te benoemen. In de literatuur zijn allerlei classificatieschema's in omloop voor het indelen van leerdoelen naar verschillende gedragsniveaus. Eén van die schema's is de 'Taxonomy of Educational Objectives' van Bloom uit 1956. Dit schema bestaat uit een hiërarchische ordening van belangrijk geachte vaardigheden, namelijk kennis, begrip, toepassing, synthese en evaluatie. Uit deze taxonomie stamt de vaak gehanteerde indeling naar kennis, inzicht en toepassing. Op dit moment lijkt echter een ander schema, ontwikkeld door De Corte (1973), het meest adequaat om leerdoelen in te delen. Dit schema omvat in zijn vereenvoudigde vorm drie gedragscategorieën: reproductie van informatie, redeneren met informatie en productie van informatie. Op basis hiervan kunnen we vragen naar drie typen oplosgedrag/niveaus van beheersing categoriseren.

Reproduceer vragen: Het gaat hierbij om een gebruikelijke vraagstelling, waarbij letterlijk of geparafraseerd gevraagd wordt zaken uit de cursustekst te reproduceren.

Voorbeeld 1

Wat verstaat men onder de term 'differentiële drempel' uit de psycho-fysica?

- de minimale intensiteit die een prikkel moet hebben om waargenomen te worden
- de minimale intensiteit die twee stimuli nodig hebben om tot de waarneming van een verschil te komen
- het minimale verschil in intensiteit tussen twee stimuli dat nodig is om tot de waarneming van een verschil te komen

Toelichting: Het begrip 'differentiële drempel' is in de leerstof letterlijk omschreven.

Redeneer vragen: Hierbij gaat het om een vraagstelling waarbij het antwoord niet gevonden kan worden door het enkel reproduceren of parafraseren, maar door het bewerken van deze informatie. De student kan relaties leggen tussen feiten en begrippen (inzicht) en regels toepassen.

Voorbeeld 2

In het dictum van een uitspraak in een administratief-rechtelijk geschil leest u het volgende: 'Verklaart de bestreden beschikking nietig; bepaalt dat verweerder opnieuw zal beslissen met inachtneming van het in deze uitspraak overwogene.'

Welke instantie zou gezien dit citaat de uitspraak gedaan kunnen hebben?

- de Kroon
- Gedeputeerde staten
- de Afdeling Geschillen van Bestuur van de Raad van State
- de Afdeling Rechtspraak van de Raad van State

Toelichting: De student wordt hier geconfronteerd met iets nieuws: uit een onbekend citaat moet de student afleiden welke instantie deze uitspraak gedaan zou kunnen hebben. (Maar als het antwoord letterlijk in de cursus zou staan, zou het gaan om reproduceren in plaats van redeneren!)

Produceervragen: Hierbij gaat het om een probleem waarvan de oplossing creatief denken vraagt: een nieuwe probleemsituatie, een ongebruikelijke oplossingsmethode, evaluatie volgens zelf te kiezen criteria. In tegenstelling tot de categorieën 1 en 2, is het bij vragen van de categorie 3 zo goed als onmogelijk een model-antwoord op te stellen omdat meerdere goede antwoorden mogelijk zijn.

Voorbeeld 3

Het opstellen van een business- of behandelplan, het oplossen van een ingewikkeld meetkundevraagstuk, het schrijven van een opstel, enzovoort.

Toelichting: Het oplossen van dergelijke problemen vergt een originele, aan de specifieke situatie aangepaste combinatie van redeneervaardigheden in tegenstelling tot het toepassen van een 'kookboekmethode'.

Op grond van de leerdoelen die betrekking hebben op een bepaald leerstofonderdeel moet bepaald worden of, en zo ja, hoeveel vragen van een bepaalde categorie over dat onderdeel in het tentamen moeten voorkomen. Op de horizontale as van de toetsmatrijs worden de drie gedragscategorieën vermeld. Per leerstofonderdeel moet door een bepaald breukgetal (2,11, enz.), in te vullen in de corresponderende cel van de toetsmatrijs, worden aangegeven hoeveel vragen van een bepaalde categorie in het tentamen moeten worden opgenomen.

Voorbeeld bij een cursus met vier leereenheden:

Vakinhoud	reproductie	redeneer	productie	Aantal
Leereenheid 1: Theorie van Piaget	2	1/2	-	25%
Leereenheid 2: Theorie van Freud	2	1/2	-	10%
Leereenheid 3: Sociale leertheorie	1	-	-	25%
Leereenheid 4: Informatieverwerkende benadering	2	1/2	-	40%
Totaal aantal vragen				100%

Nadat in de toetsmatrijs globaal is vastgelegd hoeveel vragen van een bepaalde categorie over de verschillende leerstofonderdelen in het tentamen moeten worden opgenomen, dient vervolgens het absolute aantal en de vorm van de vragen (gesloten vragen; open vragen) te worden vastgesteld, evenals eventuele audio- en/of videofragmenten.

Tentamenvorm

Allereerst moet worden nagegaan of het noodzakelijk is om open vragen in het tentamen op te nemen. Mede op basis van de ingevulde toetsmatrijs kan men nu nagaan of het noodzakelijk is om over bepaalde leerstofonderdelen open vragen te formuleren of zelfs een bijzondere verplichting deel van het tentamen te laten uitmaken.

Optimale lengte

Nadat een beslissing is genomen over het al of niet opnemen van open vragen in het tentamen moet vervolgens worden bepaald uit hoeveel vragen het gehele tentamen gaat bestaan.

Het totaal aantal mogelijk op te nemen vragen is afhankelijk van de volgende factoren:

- het wel of niet opnemen van een x-aantal open vragen in het tentamen
- de complexiteit van de vragen: moet een complex probleem worden opgelost dat veel denk- en/of rekenwerk vergt of gaat het om een eenvoudige vraagstelling?
- de omvang van de vragen: bevat een vraag veel inleidende tekst/gegevens of veel alternatieven zodat er nogal wat tijd aan lezen 'verloren' gaat?
- de omvang van het gewenste antwoord: gaat het om het aanstrepen van het juiste alternatief of moet de student zelf een (uitgebreid) antwoord formuleren?
- de beschikbare afnametijd: de afnametijd voor OU-tentamens is gesteld op maximaal drie uur
- de raadkans bij meerkeuzevragen: naargelang vragen minder antwoordalternatieven bevatten is het noodzakelijk terwille van de betrouwbaarheid om meer vragen op te nemen.

In de loop van de tijd heeft de OU-tentamenpraktijk een aantal ervaringsgegevens opgeleverd omtrent de 'optimale' lengte van een tentamen met alleen gesloten vragen: voor een tentamen van drie uur ongeveer 40 vierkeuze- of 60 driekeuze- of 80 tweekeuzevragen. Indien het tentamen tevens open vragen bevat, zal men bij het bepalen van het totaal op te nemen vragen moeten schatten hoeveel tijd de beantwoording van deze open vragen zal vergen.

Uiteindelijk moet men, uitgaande van de gewenste lengte van het tentamen en gebruik makend van de gegevens uit de toetsmatrijs (breuken, percentages), de definitieve toetsmatrijs opstellen.

Voorbeeld van een toetsmatrijs voor een tentamen met alleen gesloten vragen:

Vakinhoud	reproductie	redeneer	productie	Aantal
Leereenheid 1: Theorie van Piaget		, waarvan 1 met video	-	10
Leereenheid 2: Theorie van Freud	2	2	-	4
Leereenheid 3: Sociale leertheorie	10, waarvan 1 audio	-	-	10
Leereenheid 4: Informatieverwerkende benadering	8	8	-	16
Totaal aantal vragen				40

Uiteindelijk ontstaat door het opstellen van een toetsmatrijs een blauwdruk van het te construeren tentamen. Zo wordt zichtbaar hoe de vragen over de leerstof verdeeld zullen zijn en in welke mate het tentamen vraagt naar reproduceren, redeneren en produceren. Nadat op deze wijze inhoud en vorm van het tentamen bepaald zijn, kan men beginnen met de constructie van de tentamenvragen en kan men in een latere fase een tentamen samenstellen. Na de samenstelling van het tentamen moet terugkoppeling naar de toetsmatrijs plaatsvinden: ziet het tentamen eruit zoals bedoeld was?

IV. Het construeren en screenen van gesloten vragen

Deze sectie is bedoeld voor degenen die ten behoeve van de OU gesloten vragen voor toetsen en tentamens construeren, de vraagconstructeurs. Tevens wordt examinatoren aangeraden deze sectie door te nemen, aangezien zij degenen zijn die de toetsen en tentamens moeten samenstellen c.q. screenen.

De navolgende onderwerpen zullen achtereenvolgens aan bod komen:

- eisen te stellen aan gesloten vragen
- vormen van gesloten vragen
- constructie- en screeningsregels voor gesloten vragen en
- regels voor de opmaak van gesloten items.

IV-1. Eisen te stellen aan gesloten vragen

We definiëren een vraag als een vraag of een opgave, tezamen met het antwoordmodel.

Een gesloten vraag bestaat uit een vraag of opgave waarbij de student niet zelf het antwoord hoeft te construeren maar bijvoorbeeld uit meerdere alternatieven het juiste moet kiezen. Het antwoordmodel bestaat uit een aanduiding van het juiste antwoord. Voordat vragen in een toets of tentamen opgenomen kunnen worden, moeten zij voldoen aan de eisen van relevantie, objectiviteit, efficiëntie en moeilijkheidsniveau.

Relevantie

De relevantie van een vraag heeft betrekking op de mate waarin een vraag toetst wat het beoogt te toetsen, met andere woorden op de mate waarin de vraag een directe relatie heeft met het leerdoel.

Een vraag kan *niet* als relevant worden beschouwd indien:

1. het kennis en vaardigheden meet, die niet in de leerdoelen omschreven zijn, bijvoorbeeld:
 - a. de vraag toetst tekstbegrip, terwijl het kennis dient te toetsen;
 - b. de vraag toetst algemene kennis en niet kennis van de inhoud van de cursus;
 - c. de studenten kunnen de vraag oplossen door het principe uit het hoofd te leren (en dit komt niet overeen met het te toetsen leerdoel);
 - d. de vraag is triviaal (bijvoorbeeld vragen naar detailkennis in voetnoten);
 - e. er worden namen van personen gevraagd en dat is niet van belang voor het kennen van de stof.
2. het verkeerd beantwoorden van de vraag veroorzaakt wordt door andere aspecten dan het niet bereikt zijn van het leerdoel. Bijvoorbeeld als een vraag zo moeilijk of onduidelijk is dat studenten eerst de vraag moeten ontcijferen alvorens zij deze kunnen beantwoorden.
3. er een andere activiteit wordt gevraagd dan in het leerdoel beoogd wordt. Als een leerdoel luidt: 'De relatieve cumulatieve frequentie kunnen berekenen', dan is een vraag waarbij de vraag is: 'Wat is een relatieve cumulatieve frequentie?' niet relevant omdat dan een andere activiteit dan het kunnen berekenen wordt getoetst.
4. geen probleem gesteld wordt in de stam (bijvoorbeeld door het ontbreken van een werkwoord);
5. het met gezond verstand kan worden opgelost;
6. het een strikvraag is;

7. woorden als 'nooit', 'alleen', 'altijd', 'alle' gebruikt worden en deze een aanwijzing kunnen geven voor het juiste alternatief;
8. het juiste alternatief opvallend langer of juist korter is dan de overige alternatieven;
9. enkele alternatieven hetzelfde zeggen;
10. de onjuiste alternatieven er 'met de haren bijgesleept zijn';
11. het juiste alternatief systematisch op een bepaalde plaats staat;
12. de onjuiste alternatieven elkaar overlappen;
13. niet alle alternatieven aansluiten op de stam (logisch of grammaticaal);
14. een woord uit de stam terugkomt in het juiste alternatief;
15. de onjuiste alternatieven te heteroog en daardoor niet geloofwaardig zijn.

Hieronder volgen enkele voorbeelden van vragen die niet voldoen aan de eis van relevantie:

Voorbeeld 1

Compatibiliteit

- is een toegepaste vorm van emulatie.
- heeft betrekking op de leer van de overeenkomsten tussen computersystemen.
- houdt de leer van overeenkomsten tussen programmeertalen in.
- is de leer van de overeenkomsten tussen computersystemen.

Toelichting: Het hoofdidee van een meerkeuzevraag behoort in de stam tot uitdrukking te worden gebracht. In de stam moet een werkwoord voorkomen, zodat er een probleem in de stam staat (vergelijk punt 4).

Voorbeeld 2

H. Roland Holst was de schrijver van 'Voorbij de wegen'.

Deze stelling is

- juist
- onjuist

Toelichting: De voorletter is in feite A, niet H. Dit is een bedrieglijke presentatie, te meer omdat H. Roland Holst ook schreef (vergelijk punt 2).

Voorbeeld 3

Waarvan moet werkloosheid het gevolg zijn om te spreken van conjuncturele werkloosheid?

- een teveel aan ongeschoolde arbeid
 - seizoensinvloeden in bepaalde bedrijfstakken
 - wijzigingen in de conjunctuur
 - wijzigingen in de productiestructuur
-

Toelichting: De woorden 'conjuncturele' en 'conjunctuur' zijn zo duidelijk verwant, dat het goede antwoord sterk op de voorgrond treedt (vergelijk punt 14).

Voorbeeld 4

Wat vormt een rechtsbron in formele zin?

- het geweten
 - het christendom
 - het verdrag tot bescherming van de rechten van de mens en de fundamentele vrijheden (het Verdrag van Rome)
 - het geloof
-

Toelichting: Doordat het juiste alternatief door preciezer formuleren veel langer is dan de overige alternatieven, kan een student dit vraag met gezond verstand oplossen (vergelijk punt 5 en 8). Tevens is er een overlap tussen de onjuiste alternatieven b en d (vergelijk punt 12).

Voorbeeld 5

Welk mechanisme speelt volgens Freud in de mobiliteit van energie en de variabiliteit van driftobjecten een belangrijke rol?

- overdracht
 - catharsis
 - verdringing en sublimatie
-

Toelichting: Niet alle alternatieven sluiten grammaticaal aan bij de stam. Alternatief c geeft twee mechanismen, terwijl er naar één wordt gevraagd (vergelijk punt 13).

Voorbeeld 6

Columbus maakte vijf ontdekkingsreizen naar het westelijk halfrond.

Deze stelling is

- juist.
 - onjuist.
-

Toelichting: Deze stelling is onjuist, want het waren er maar vier. Dit soort 'subtiliteiten' maakt het juist-onjuist vraag triviaal (vergelijk punt 6).

Objectiviteit

Een vraag is objectief indien de score van de vraag onafhankelijk is van degene die de scoring uitvoert. Dit impliceert dat

- de procedure voor het beoordelen van het antwoord op een vraag zo is geregeld dat geen beslissingen aan het oordeel van de beoordelaar worden overgelaten en het antwoordmodel geen ruimte voor interpretatie openlaat
- deskundigen het eens zijn over het juiste antwoord of over de criteria waaraan een juist antwoord moet voldoen.

Een vraag kan *niet* objectief worden geacht indien

1. het vraagt naar de mening van de student;
2. het niet éénduidig waar is (of bij stelling juist of onjuist);
3. niet duidelijk is wat met een bepaald woord in de vraag wordt bedoeld;
4. onvoldoende informatie in de stam is opgenomen om het juiste alternatief te kunnen aanstrepen;
5. meer dan één alternatief juist is (behalve bij het 'meerdere antwoorden juist-type!);
6. meerdere deskundigen een verschillende opvatting hebben over het juiste antwoord;
7. een ander alternatief ook te verdedigen is, maar dat staat niet in de cursus vermeld.

Enkele voorbeelden van vragen die niet voldoen aan de eis van objectiviteit:

Voorbeeld 1

Bij welke onderwijskundige staat volgens u het kind het meest centraal?

- 0 Boeke
 - 0 Goodlad
 - 0 Montessori
 - 0 Steiner
-

Toelichting: De vraag is niet objectief, omdat de vraag naar de mening van de student en omdat er niet voldoende informatie in de stam staat om het juiste alternatief aan te strepen (vergelijk punt 1 en punt 4).

Voorbeeld 2

Wie van de onderstaande personen behoort tot de Amerikaanse behavioristische stroming?

- 0 Bell
 - 0 Einstein
 - 0 Skinner
 - 0 Watson
-

Toelichting: Zowel Watson als Skinner horen bij deze behavioristische stroming, dus beide alternatieven zijn juist (vergelijk punt 5).

Voorbeeld 3

Cognitief gedrag heeft motorische voorlopers.

Deze stelling is

0 juist.

0 onjuist.

Toelichting: Deze stelling kan al naar gelang de opvatting van cognitie die men hanteert met juist of onjuist worden beantwoord (vergelijk punt 4). Beter is: Piaget stelt dat cognitief gedrag motorische voorlopers heeft.

Voorbeeld 4

Regenwolken zijn licht in gewicht.

Deze stelling is

0 juist.

0 onjuist.

Toelichting: 'Licht in gewicht' is een relatieve term. Het is maar waar men vanuit gaat en juist deze vergelijkingsterm ontbreekt. Beter is: Lucht met een hoge vochtigheidsgraad is minder dicht dan lucht met een lage vochtigheidsgraad.

Efficiëntie

Wanneer men bij het ontwikkelen van een vraag kan kiezen uit verschillende soorten vragen die in alle opzichten gelijkwaardig zijn, dan verdient de vraag die het minste tijd kost in termen van afname en correctie de voorkeur.

Een vraag kan *niet* efficiënt worden geacht indien

1. het niet duidelijk is geformuleerd;
2. er irrelevante inleidingen in de stam, stelling of uitspraak staan;
3. er een dubbele ontkenning in voorkomt;
4. het onnodig negatief is gesteld;
5. het grammaticaal niet juist geformuleerd is (bijv. enkelvoud in de stam, meervoud in één van de alternatieven);
6. er ingewikkelde zinsconstructies en woorden in voorkomen;
7. de zinnen te lang zijn (het streven is zinnen van maximaal vijftien woorden te gebruiken);
8. er een (voor de student) onbekend woord in staat;
9. alternatieven in decimalen zijn gegeven, terwijl het juiste antwoord een geheel getal is;
10. tussenvoegsels de leesbaarheid verlagen;
11. er onnodige verwijzingsstekens ingevoerd zijn;
12. er in de alternatieven steeds één zinsdeel uit de stam herhaald wordt.

Hieronder volgen voorbeelden van vragen die *niet* voldoen aan de eis van efficiëntie.

Voorbeeld 1

De Russische leider Stalin, onder wiens leiding zoveel slachtoffers zijn gevallen, had *niet* op de Duitse aanval in Rusland in 1941 gerekend, omdat hij in 1939 een niet-aanvalsverdrag met Hitler had gesloten.

Deze stelling is

juist.

onjuist.

Toelichting: De mededeling over de slachtoffers is in dit verband niet ter zake doende en doet dus afbreuk aan de efficiëntie (vergelijk punt 2).

Voorbeeld 2

Het is *niet* juist om te beweren, dat de klassieke psychoanalyse zich *niet* bezighoudt met

geconditioneerde reflexen.

mathematische gedragsmodellen.

testconstructie.

theorievorming.

Toelichting: Dubbele ontkenningen roepen verwarring op bij de studenten en introduceren een extra probleem in de vraag (vergelijk punt 3).

Voorbeeld 3a

Welke functie hebben getallen in een nominale schaal?

De getallen hebben de functie van kengetal.

De getallen hebben de functie van kengetal en rangordegetal.

De getallen hebben de functie van rangordegetal.

De getallen hebben de functie van kardinaal getal.

Toelichting: De leestijd van deze vraag kan bekort worden en de leesbaarheid verhoogd door de volgende formulering (vergelijk punt 12):

Voorbeeld 3b

Welke functie hebben getallen in een nominale schaal?

kengetal

kengetal en rangordegetal

rangordegetal

kardinaal getal

Moeilijkheidsgraad

De moeilijkheidsgraad van een vraag moet aanvaardbaar zijn, gegeven het niveau van de studenten. Met name bij het tentamineren zijn we geïnteresseerd in het verdelen van de studenten over de categorieën 'voldoende beheersing van de leerstof' en 'onvoldoende beheersing van de leerstof'. Daarom willen we vooral de zogenaamde grensstudenten nauwkeurig van elkaar onderscheiden. Dit bereiken we door deze studenten opgaven voor te leggen die goed bij hun niveau aansluiten. Als we deze grensstudenten te moeilijke vragen geven is geen scherp onderscheid mogelijk, omdat de hele groep (nagenoeg) alle opgaven fout zal beantwoorden. Als de opgaven te gemakkelijk zijn, beantwoordt de hele groep grensstudenten (nagenoeg) alle vragen juist en is ook geen scherp onderscheid mogelijk.

Een vraag kan te moeilijk zijn of worden geacht indien

1. er meer dan één probleem aan de orde wordt gesteld;
2. de alternatieven vrij homogeen zijn.

Een vraag kan te gemakkelijk zijn of worden geacht indien

3. de afleiders uit een heel ander of verschillend (vak)gebied afkomstig zijn;
4. onware uitspraken te kunstmatig zijn;
5. één of meer alternatieven te duidelijk incorrect zijn;
6. in de stelling elementen gecombineerd worden die niets met elkaar te maken hebben.

Een voorbeeld van een vraag die te gemakkelijk is:

Voorbeeld 1

Welke politieke partij heeft *geen* minister-president geleverd?

- 0 CDA
 - 0 PvdA
 - 0 Partij voor de dieren
 - 0 VVD
-

Toelichting: Het foute alternatief ligt erg voor de hand (kleine, nog vrij nieuwe partij). Hierdoor is dit alternatief duidelijk incorrect (vergelijk punt 5).

Voorbeeld 2

Martin Luther King was tegen de inval van Amerika in Grenada.

Deze stelling is

- 0 juist.
- 0 onjuist.

Toelichting: Deze twee elementen (King en de inval in Grenada) hebben niets met elkaar te maken, alleen al chronologisch gezien. Het levert een absurde uitspraak op. (Vergelijk punt 6).

De gezichtspunten waaronder de uitspraken gecategoriseerd zijn, kunnen niet beschouwd worden als elkaar uitsluitende categorieën. Zo kan men bijvoorbeeld een uitspraak als 'de vraag vraagt naar de mening van de student' niet alleen vangen onder 'objectiviteit' (er is immers geen juist antwoord mogelijk), maar ook onder 'relevantie' (indien meningsvragen niet omschreven zijn in de te toetsen leerdoelen). De zin van deze gezichtspunten is dat zij vraagconstructeurs een 'checklist' kunnen verschaffen, waardoor zij makkelijker en systematischer kritiek op de vragen kunnen leveren.

IV-2. Het kiezen van een vraagvorm

Door de invoering van computergebaseerd toetsen is het mogelijk om veel meer vraagvormen te gebruiken dan bovengenoemde basistypen. We onderscheiden open vragen met kort antwoord, open vragen met lang antwoord, meerkeuzevragen (3-, 4-, meer-alternatieven), meerkeuzevragen met behulp van eenpulldownmenu, ja/nee-vragen, juist/onjuist vragen, meerdere-antwoorden-juist-vragen, aanwijsvragen (hotspot), matrixvragen, matchingvragen, ordeningsvragen, drag-and-drop-vragen, invulvragen en numerieke vragen.

Voor examinatoren zijn meerdere vraagvormen geschikt om te gebruiken voor het beoordelen of de leerdoelen behaald zijn. Op grond van technische eisen en efficiëntie is er echter een voorkeur voor bepaalde vraagvormen.

De 'meerdere antwoorden juist' en de matchingvraag zijn varianten op de standaard meerkeuzevraag en is minder geschikt omdat dit type vraag een ingewikkelder scoring vereist waardoor de transparantie voor de studenten wordt aangetast en de psychometrische analyse en cesuurbepaling lastiger worden.

De meerkeuzevraag met pull-down-menu verschilt alleen qua uiterlijk van een standaard meerkeuzevraag. De matrixvraag is op te splitsen in vier afzonderlijke standaard meerkeuze-vragen.

De invulvraag is te beschouwen als een kort-antwoordvraag, maar als examinator moet je bij het gebruik van dit vraagtype anticiperen op alle mogelijke typefouten en synoniemen.

Het gebruik van audio en videofragmenten bij de gesloten vragen biedt goede mogelijkheden om aan te sluiten bij bepaalde leerdoelen, maar kan alleen gebruikt worden als er op de studiecetra kop telefoons beschikbaar zijn (en dat is momenteel (lees: 2016) niet het geval).

Bij het kiezen van een vraagvorm is het van belang rekening te houden met:

- o de technische mogelijkheden;
- o de inrichtingsvoorwaarden voor de studiecetra
- o een ondersteuningsstructuur voor de examinatoren, zowel toetstechnisch als automatiseringstechnisch.

Op basis van deze analyse is het advies om voorlopig alleen gebruik te maken van de volgende vraagvormen:

- o Standaard open vraag (met of zonder afbeeldingen)
- o Standaard meerkeuzevraag
- o Meerkeuzevraag met afbeeldingen
- o Stellingvraag
- o Meerkeuzevraag met pull down menu

De volgende itemvormen mogen pas gebruikt worden na een duidelijke instructie aan de examinatoren en een testfase van de technische afhandeling van deze itemvormen:

- o invulvraag
- o matchingvraag

Voor de volgende itemvormen was de ingeschatte bruikbaarheid door examinatoren lager:

- o tekst-matchvragen
- o drag-and-drop-vragen
- o open vragen numeriek

Het gebruik van deze itemvormen zal altijd pas kunnen nadat de afhandeling van deze itemvormen technisch getest is en als de examiner hiervoor specifieke instructies ontvangen heeft van een toetsdeskundige.

De volgende itemvormen worden voorlopig om toetskundige redenen nog niet gebruikt in summatieve toetsen:

1. meerdere-antwoorden-juist-vraag
2. rangordevraag.

IV-3. Vormen van gesloten vragen

In paragraaf IV-1 hebben we ons beperkt tot voorbeelden uit twee basistypen binnen de groep van gesloten vragen:

1. stellingvragen (juist-onjuist)
2. meerkeuzevragen (multiple choice)

In deze paragraaf gaan we in op de verschillende varianten van deze basistypen.

1. Stellingvragen

Bij een stellingvraag gaat het om het als juist of onjuist beoordelen van een uitspraak.

Voorbeeld juist-onjuist-vraag

Beoordeel de juistheid van de volgende stelling.

Een variantie-covariantiematrix is positief semi-definiet.

Deze stelling is

- 0 juist.
 - 0 onjuist.
-

Verschillende varianten zijn mogelijk op dit type. Zo kunnen woordparen beoordeeld worden op de vraag of ze wel of geen synoniemen zijn, of ze hetzelfde dan wel het tegengestelde betekenen.

Uitspraken

kunnen beoordeeld worden naar de vraag of het feiten of opinies zijn. Of conclusies kunnen beoordeeld worden op geldigheid of ongeldigheid.

Voorbeeld stellingvraag met extra informatie in het 'gegeven'

Gegeven:

In het Civil Conservation Corps speelde het leger een belangrijke rol.

Gevraagd:

Beoordeel de juistheid van de volgende stelling.

Roosevelt streefde ernaar deze instelling zo gauw mogelijk op te heffen.

Deze stelling is

- 0 juist.
0 onjuist.
-

Een reeks van dergelijke vragen kan bijvoorbeeld gesteld worden naar aanleiding van een grafiek of een stuk tekst waarin een mening wordt gegeven. Kortom, juist-onjuist vragen bieden ruime toepassingsmogelijkheden.

Als men snel achter elkaar een stel feiten wil toetsen is een stellingvraag boven een vierkeuzevraag te verkiezen. Dit type vragen is eenvoudig te construeren, omdat er niet moeizaam alternatieven bedacht hoeven te worden. Tevens zijn zij door de studenten snel te beantwoorden. Het is wel lastig om ervoor te zorgen dat de stellingen slechts voor één interpretatie vatbaar zijn. De juist-onjuist vragen zijn niet alleen geschikt voor het toetsen van 'feitjes'. Wat een vraag meet, of bedoeld is te meten, wordt in de eerste plaats bepaald door de inhoud (dat wil zeggen de leerstof en de vraagstelling) en niet door de vorm.

2. Meerkeuzevragen

Het meest gangbare meerkeuzevraag bestaat uit een stam en vier alternatieven, waarvan er één juist dient te zijn, de overige onjuist. Deze onjuiste alternatieven worden afleiders genoemd. De vraag (of het probleem) kan op verschillende manieren in de stam geformuleerd worden.

We onderscheiden:

- een positief geformuleerde stam.
- een negatief geformuleerde stam.

Daarnaast kunnen we een indeling maken naar vraagtype:

- een directe vraag.
- een zin met 'weglating'.

Hieronder volgen enkele voorbeelden van verschillende vraagvormen.

Voorbeeld 1 meerkeuzevraag met vier alternatieven

Bij het definiëren van de 'betrouwbaarheid van een bedrijf' wordt een aantal indicatoren betrokken. Welke daarvan is *geen* goede indicator?

- de leeftijd van het bedrijf
 - het uitgekeerde winstpercentage
 - het percentage orders waarover klachten binnenkomt
 - het percentage klanten van vorig jaar dat nu nog klant is
-

Voorbeeld 2

Wat is de opeenvolging van geheugentechnieken, in de tijd gezien?

- elektronenbuis - transistor - integrated circuit - microprocessorchip
 - transistor - elektronenbuis - microprocessorchip - integrated circuit
 - elektronenbuis - transistor - microprocessorchip - integrated circuit
 - transistor - elektronenbuis - integrated circuit - microprocessorchip
-

Voorbeeld 3

Voor de opwekking van elektriciteit kan warmte en elektriciteit gecombineerd geproduceerd worden.

Hoe wordt dit genoemd?

- total energysystem
 - warmte-kracht-koppelingssysteem
 - elektriciteit-warmte-koppelingssysteem
 - total production system
-

De variaties in meerkeuzevragen worden bepaald door de opdracht die in de vraagstelling vervat is. Meestal is de opdracht het juiste of het beste antwoord te kiezen, maar de opdracht kan ook zijn het onjuiste, het slechtste, het meest omvattende of het meest ongelijke antwoord te kiezen. Ook kan de opdracht zijn een onderdeel van de stam te vervangen door iets anders (substitutievraag). Kortom: er zijn mogelijkheden tot variatie te over.

De driekeuze en de vijfkeuzevraag verschillen slechts in één opzicht van de gewone vierkeuzevraag namelijk in het aantal alternatieven.

Voorbeeld driekeuzevraag

Wat is de hoofdstad van Frankrijk?

- Parijs
 - Madrid
 - Rome
-

Voorbeeld vijfkeuzevraag

Wat is de hoofdstad van Spanje?

- Parijs
 - Madrid
 - Rome
 - Amsterdam
 - Brussel
-

Het kunnen beschikken over de vraagvorm met drie alternatieven is een aanwinst. Eén van de moeilijkheden bij het construeren van vier- en vijfkeuzevragen is telkens weer het bedenken van plausibele afleiders. Vaak is een alternatief al direct te doorzien en functioneert dan niet meer als reële afleider. Soms zijn er op een vraag ook maar drie antwoorden mogelijk. Dit geldt voor antwoorden als groter, kleiner en gelijk aan. Het is wenselijk dat men zich bij de keuze van het aantal alternatieven laat leiden door inhoudelijke overwegingen. Zijn vier of meer alternatieven zinvol, dan maakt men een vier- of vijfkeuze- vraag. Als slechts drie zinvolle alternatieven te construeren zijn is het onverstandig een vier- of vijfkeuzevraag op te stellen.

Een mogelijk bezwaar is dat de raadkans groter is bij driekeuzevragen dan bij vierkeuzevragen. Door meer vragen in de toets op te nemen, kan aan dit bezwaar tegemoet worden gekomen. Bijkomend voordeel van een groter aantal vragen is de betere dekking van de leerdoelen. Dit betekent echter niet dat elk vierkeuzevraag straffeloos in een driekeuzevraag is om te zetten. Het onderstaande voorbeeld illustreert dat hier het weglaten van een alternatief de student op het spoor van het juiste antwoord kan zetten:

Voorbeeld vierkeuzevraag die niet goed om te zetten is naar een driekeuzevraag

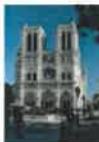
Wat is de juiste vergelijking van een cirkel met middelpunt (3, -2) en straal 5?

- $(x+3)^2 + (y-2)^2=5$
 - $(x-3)^2 + (y+2)^2=5$
 - $(x+3)^2 + (y-2)^2=25$
 - $(x-3)^2 + (y+2)^2=25$
-

Een variant op het meerkeuzevraag is het tonen van afbeeldingen, figuren of symbolen op de plaats van de alternatieven.

Voorbeeld meerkeuzevraag met afbeelding

Op welke foto staat de Dom van Keulen afgebeeld? (klik op de juiste foto)



Een andere variant op het meerkeuzevraag is de zogenaamde matrixvraag.

Voorbeeld matrixvraag

Koppel elke stad met het land waar hij bij hoort.

Kathmandu: Nigeria Finland Senegal Nepal

Helsinki: Nigeria Finland Senegal Nepal

Dakar: Nigeria Finland Senegal Nepal

Lagos: Nigeria Finland Senegal Nepal

Gebruikelijk is dat meerkeuzevragen slechts één juist antwoord kennen, maar tot de mogelijkheden binnen CBT behoort ook het 'meerdere antwoorden juist'-vraag.

Voorbeeld meerdere-antwoorden-goed-vraag

Aan welke eisen moet een goede toets voldoen?

validiteit
betrouwbaarheid
herhaalbaarheid
uitvoerbaarheid
moeilijkheid
bruikbaarheid

Tot slot behoort in het kader van CBT ook een rangordevraag tot de mogelijkheden:

Voorbeeld rangordevraag

In welke volgorde staan de kleuren in een regenboog? (Begin bij rood)

Rood 1

Indigo .

Blauw .

Groen .

Geel .

Oranje .

Violet .

IV-4. Het construeren van gesloten vragen

In deze paragraaf gaan we in op de constructie van gesloten vragen. Bij de beschrijving van constructieregels maken we ook hier onderscheid tussen de standaard meerkeuze- en de stellingvragen. We doen dit omdat het construeren en screenen van stellingvragen op een aantal punten afwijkt van dat van standaard meerkeuzevragen. We wijzen er echter op dat de bovengenoemde eisen voor beide typen vragen gelden en dat er een grote mate van overlap is in constructie- en screeningsregels.

Constructie- en screeningsregels gesloten vragen

In deze paragraaf zullen we eerst een aantal algemene constructieregels geven. Daarna zullen we ingaan op de screening van meerkeuzevragen en tenslotte geven we een checklist waarmee we meerkeuzevragen op hun waarde kunnen beoordelen.

Algemene constructieregels

Tijdens het construeren van standaard meerkeuzevragen moet op de volgende, algemene regels worden gelet:

1. Indien de stam of een uitspraak negatief geformuleerd is, verdient het aanbeveling de student hier door middel van cursivering op te wijzen. Een negatief geformuleerde stam zou anders wel eens een strikvraag kunnen zijn. In het algemeen is een positieve vraagvorm te verkiezen.
2. Stellingen moeten eenduidig juist of onjuist zijn, maar het gebruik van woorden als 'nooit', 'alleen', 'altijd', en 'alle' is in het algemeen niet wenselijk. Er zullen altijd wel uitzonderingsgevallen te vinden zijn, waardoor een dergelijke uitspraak op voorhand niet juist kan zijn. De uitspraak mag geen halve of gedeeltelijke waarheden bevatten.
3. Neem geen letterlijke uitspraken uit het cursusmateriaal. Door oppervlakkig bestuderen kan de student de uitspraak als juist herkennen.
4. Indien extra informatie nodig is om te beoordelen of een uitspraak of stelling juist is of om een meerkeuzevraag goed te beantwoorden, dan is het noodzakelijk duidelijk aan te geven wat de extra informatie is en wat de te beoordelen stelling.
5. Slechts één van de alternatieven mag juist zijn (behalve bij het 'meerdere antwoorden juist'-vraag!).
6. De alternatieven moeten zo helder geformuleerd worden dat de essentiële verschillen ertussen duidelijk tot uitdrukking komen.
7. Stel altijd eerst het (de) juiste alternatief(ven) op en pas daarna de afleiders.
8. De afleiders moeten geloofwaardig lijken voor degenen die de stof niet of niet goed hebben bestudeerd.
9. Gebruik als afleiders ook veel voorkomende vergissingen. Na afname van de toets kunnen dan de kennistekorten bij de student(en) eenvoudig en snel worden vastgesteld.
10. Rangschik de alternatieven in een logische volgorde zoals opklimmende getal grootte en alfabetische volgorde.
11. Alle alternatieven bij één vraag moeten zoveel mogelijk overeenkomen in grammaticale structuur en verschijningsvorm (lengte!).
12. Het alternatief 'geen van bovengenoemde' of 'alle bovengenoemde' is in het algemeen niet wenselijk.
13. Vermijd dat twee afleiders elkaar uitsluiten.

14. Indien de verschillen tussen de alternatieven groot zijn (heterogene alternatieven), dan is de vraag meestal gemakkelijk. De vraag kan men moeilijker maken door de verschillen tussen de alternatieven te verfijnen (homogeniseren). Let er echter wel op dat het dan géén lees oefening wordt!

Screeningsregels voor gesloten vragen

Wanneer de vraagconstructeur, rekening houdend met de hiervoor genoemde richtlijnen en opmerkingen, een conceptvraag heeft geschreven dan kan het toch voorkomen dat het conceptvraag op grond van die richtlijnen revisie behoeft. Om dit snel te controleren, moet eerst worden nagegaan of de volgende vier vragen met 'ja' kunnen worden beantwoord:

1. Is de vraag inhoudelijk correct?
2. Staat de vraag in relatie tot een leerdoel?
3. Is er slechts één juist alternatief?
4. Zijn alle afleiders plausibel?

Moet één van deze vragen met 'nee' worden beantwoord, dan dient het concept-vraag gereviseerd te worden.

Indien alle vragen met 'ja' kunnen worden beantwoord, dan is het aan te bevelen de vraag nog eens te screenen aan de hand van de checklist. Als op alle vragen 'ja' respectievelijk 'nee' kan worden geantwoord, dan is de vraag correct. Zo niet, dan dienen de tekorten alsnog te worden aangevuld.

Checklist voor gesloten vragen			
criterium		Ja*	Nee*
Relevantie	1. Is de inhoud van de vraag in overeenstemming met het leerdoel?	*	
	2. Is er een probleem in de stam gesteld?	*	
	3. Is de vraag vakinhoudelijk juist?	*	
	4. Zijn de alternatieven plausibel?	*	
	5. Zijn de alternatieven in alfabetische of logische volgorde gerangschikt?	*	
	6. Sluiten de alternatieven aan op de stam?	*	
	7. Is de vraag geschikt voor het niveau van de studenten die deze cursus volgen?	*	
	8. Is de vraag met gezond verstand op te lossen?		*
	9. Is het een strikvraag?		*
	10. Sluit het juiste alternatief qua woordkeuze en stijl te sterk aan bij de stam in relatie tot de afleiders?		*
	11. Is er een overlap in de betekenis van de alternatieven?		*
	12. Zijn de alternatieven te homogeen dan wel te heterogeen?		*
Objectiviteit	1. Is de stam ondubbelzinnig (eenduidig) geformuleerd?	*	
	2. Zijn deskundigen het eens over het goede antwoord?	*	

	3. Bevat de stam/stelling voldoende informatie om het juiste alternatief te kunnen aanstrepen?	*	
	4. Vraagt de vraag naar de mening van de student?		*
	5. Is meer dan één alternatief juist?		*
Efficiëntie	1. Is het taalgebruik aangepast aan het niveau van de kandidaat/cursus?	*	
	2. Is de vraag grammaticaal goed en logisch geformuleerd?	*	
	3. Is de vorm van de vraag de meest geschikte toetsvorm?	*	
	4. Is de vraag duidelijk, kort en bondig geformuleerd?	*	
	5. Bevat de stam van de vraag meer informatie dan nodig is?		*
	6. Bevat de vraag een dubbele ontkenning?		*
	7. Is de vraag onnodig negatief gesteld?		*
	8. Zijn de zinnen in de vraag of in het probleem te ingewikkeld?		*
	9. Bevat de vraag onnodig vakjargon?		*
Moeilijkheids graad	1. Bevat de vraag meer dan één vraag of probleem?		*
	2. Is de stelling of zijn de alternatieven te moeilijk of te gemakkelijk?		*
* Sommige vragen moeten beantwoord worden met 'ja' en sommige met 'nee'. De vakken met de wenselijke antwoorden zijn grijs en bevatten een *.			

IV-5. Regels voor de lay-out van gesloten vragen

Achtereenvolgens worden de redactionele richtlijnen weergegeven van:

1. Drie- en vierkeuzevragen
2. Juist-onjuistvragen

Na elk vraagtype wordt aangegeven hoe de feedback eruit moet zien.

1. Drie- en vierkeuzevraag

De zin van de vraag loopt grammaticaal gezien niet door in de (drie of vier) alternatieven.

Regels:

- *witregel* tussen inleiding en vraag
- *witregel* tussen vraag en (drie of vier) alternatieven
- *vraag* wordt afgesloten met een vraagteken
- *alternatief* begint met een kleine letter (want de zin is grammaticaal onvolledig) (uitzondering: alternatief begint met een eigennaam)
- *alternatief* heeft geen punt ter afsluiting (want de zin is grammaticaal onvolledig)

Voorbeeld

Geen staat heeft het recht op haar grondgebied zodanige activiteiten te (laten)ontplooiën dat daardoor in een andere staat (of staten) ernstige en bewezen schade ontstaat.

Hoe staat dit uitgangspunt bekend?

- als de Trail Smelter-doctrine
 - als het territorialiteitsbeginsel
 - als het wederkerigheidsbeginsel
 - als de Franse-kalimijnen-doctrine
-

Let op! Bij gebruik van jaartallen, cijfers, aantallen en dergelijke de alternatieven van laag naar hoog aanbieden, dus nooit in willekeurige volgorde zetten.

Voorbeeld

Welk jaartal geldt als het geboortejaar van de academische psychologie?

- 1789
 - 1856
 - 1879
 - 1907
-

Let op! In onderstaand voorbeeld wijkt het laatste antwoordalternatief af. Het is een volledige zin: hoofdletter én punt.

Voorbeeld

Van welke hedendaagse psychologische stroming wordt of stromingen worden het experimenteel onderzoek van Wundt beschouwd als de voorloper?

- cognitieve psychologie
 - experimentele sociale psychologie
 - Zowel de cognitieve psychologie als de experimentele sociale psychologie.
-

De alternatieven bestaan alle uit grammaticaal volledig geformuleerde zinnen.

Regel

- *witregel* tussen vraag en (drie of vier) alternatieven
- *alternatief* begint met een hoofdletter
- *alternatief* heeft een punt ter afsluiting

Voorbeeld

Welke opvatting heeft de filosoof Plato (429-348 v.Chr.) over literatuur?

- Plato vindt literatuur bij uitstek geschikt om het nuttige met het aangename te verenigen.
 - Plato vindt dat literatuur via het bijzondere het algemene uitbeeldt.
 - Plato vindt literatuur nutteloos, derderangs en vaak leugenachtig.
-

In de vraag is een ontkenning opgenomen (woorden als 'niet', 'geen' of 'onjuist')

Regel

- het woord voor de ontkenning wordt *cursief* weergegeven

Voorbeeld

Wat is een *onjuist* voorbeeld van een procedé van naturalisatie?

- het sprookje
 - de briefvorm
 - de dagboekvorm
 - de manuscriptfictie
-

NB.

- Als in de vraag veel cursiveringen staan (bijv. vanwege cursivering bij boektitels), mogen bij uitzondering voor de ontkenning hoofdletters worden gebruikt: NIET, GEEN.
- Gebruik geen dubbele ontkenningen in de vraagstelling.

Opmaak en formulering feedback bij drie- en vierkeuzevragen

Bij een correct antwoord wordt de feedback weergegeven zoals in het volgende voorbeeld:

Voorbeeld

Juist.

[zachte return] [zachte
return]

Zie tekstboek p. (of pp. bij meerdere pagina's)

Bij een foutief antwoord wordt de feedback weergegeven zoals in het volgende voorbeeld:

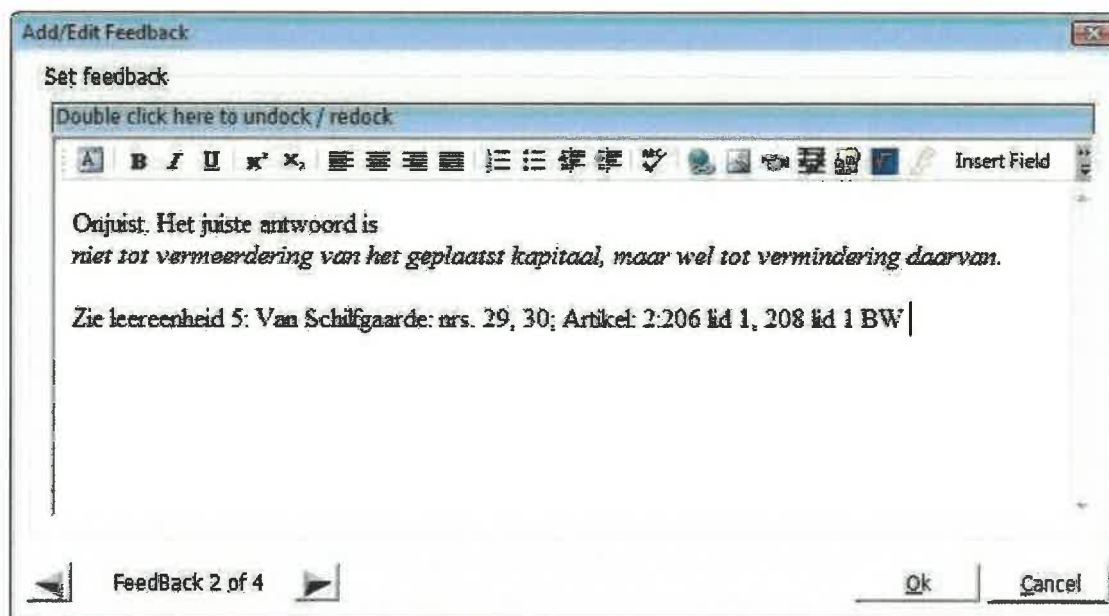
Voorbeeld

Onjuist. Het juiste antwoord is [zachte return]
het antwoord [cursief].

[zachte return] [zachte
return]

Zie tekstboek p. (of pp. bij meerdere pagina's) [dit is de minimum feedback, bij
voorkeur iets uitgebreider, d.w.z. gelijk aan antwoord dat je opneemt bij foutieve
antwoorden]

In Questionmark Perception ziet dat er dan als volgt uit:



De auteur kan besluiten het antwoord/de bron (zowel bij juist als bij onjuist) toe te lichten met een tekst, afbeelding of anders. Dit gebeurt dan met een witregel onder de bronvermelding.

2. Juist-onjuistvragen

Er wordt één stelling vermeld, daarna wordt gevraagd of deze stelling juist of onjuist is. Regel

- *witregel* tussen de stelling en de zin 'Deze stelling is'
- *witregel* tussen 'Deze stelling is' en de twee alternatieven ('juist' of 'onjuist')
- *de woorden 'juist' en 'onjuist'* worden met een kleine letter weergegeven (want woorden maken deel uit van de zin 'Deze stelling is')
- *de woorden 'juist' en 'onjuist'* krijgen ter afsluiting een punt (want zijn afsluiting van een grammaticaal doorlopende zin)

Voorbeeld

De islam, het Jodendom en het christendom zijn monotheïstische godsdiensten.

Deze stelling is

0 juist.

0 onjuist.

Eerst wordt een gegeven vermeld, dan volgt één stelling, en vervolgens wordt gevraagd of deze stelling juist of onjuist is.

Regel

- vermelding *Gegeven* (met hoofdletter en cursief), daaronder zonder witregel de uitwerking van het 'gegeven'
- *witregel* tussen Gegeven en Stelling
- vermelding *Stelling* (met hoofdletter en cursief), daaronder zonder witregel de uitwerking van de 'stelling'
- *witregel* tussen Stelling en de zin 'Deze stelling is'
- *de woorden 'juist' en 'onjuist'* worden met een kleine letter weergegeven (want woorden maken deel uit van de zin 'Deze stelling is')
- *de woorden 'juist' en 'onjuist'* krijgen ter afsluiting een punt (want zijn afsluiting van een grammaticaal doorlopende zin)

Voorbeeld

Gegeven:

Het beginsel van rechtvaardigheid wordt door de moslimethicus Al-Isfahani beschouwd als eerste hoofdprincipe van orde en harmonie in menselijke relaties.

Stelling:

Hij stelt hierbij dat rechtvaardigheid nog boven de liefde staat in de hiërarchie van de deugden.

Deze stelling is

0 juist.

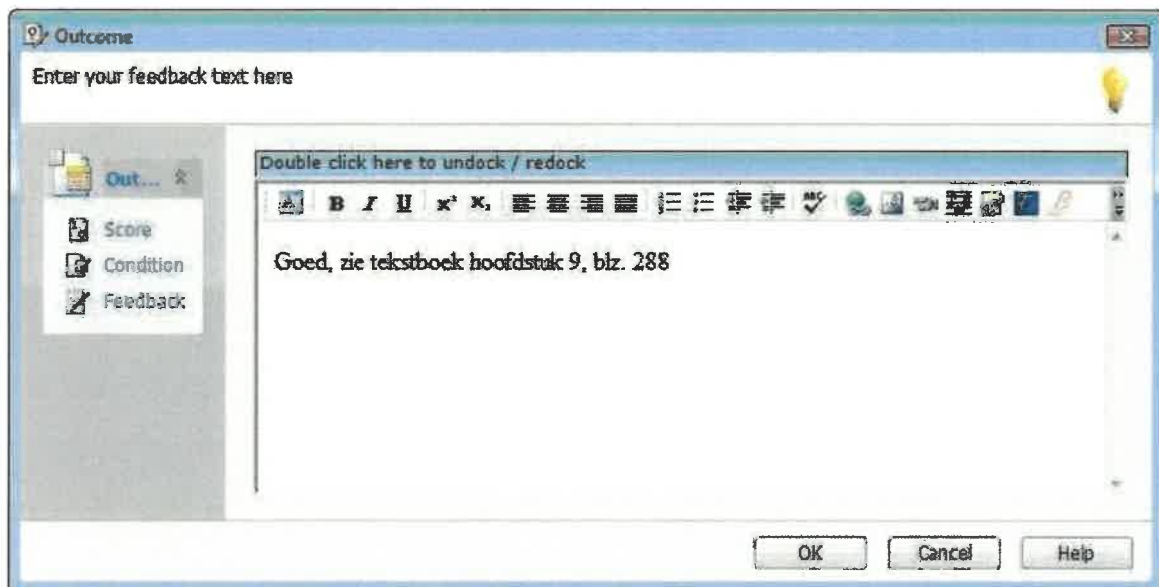
0 onjuist.

Opmaak en formulering feedback bij stellingvragen

Bij een foutief antwoord volgt de volgende feedback:



Bij een juist antwoord volgt de volgende feedback:



Het heeft geen zin om het juiste alternatief te herhalen dat wekt alleen verwarring.

V. Het construeren en screenen van open vragen

Deze sectie is met name bedoeld voor degenen die voor de OU open vragen ten behoeve van toetsen en tentamens construeren, de vraagconstructeurs. Ook examinatoren wordt aangeraden deze sectie door te nemen, aangezien zij belast zijn met het samenstellen en screenen van de eindtoets respectievelijk het tentamen.

Hier zullen de volgende drie onderwerpen aan bod komen:

- eisen die aan open vragen moeten worden gesteld
- vormen van open vragen
- constructie- en screeningsregels voor open vragen en
- richtlijnen voor de lay-out van open vragen.

V-1. Eisen te stellen aan open vragen

We definiëren een open vraag als een vraag of opgave, tezamen met het antwoordmodel.

Een open vraag is een vraag of opgave waarbij de student het antwoord zelf moet formuleren. De beoordelingsinstructie bevat (de onderdelen van) het ideale antwoord (in geval van reproductie- en redeneervragen) of criteria waaraan het goede antwoord moet voldoen (in het geval van productievragen).

Voordat vragen voor een toets of tentamen bruikbaar zijn, moeten zij voldoen aan de eisen van

- validiteit
- objectiviteit
- efficiëntie
- moeilijkheidsgraad

Validiteit

De validiteit van een open vraag heeft betrekking op de mate waarin die vraag toetst wat hij beoogt te toetsen, met andere woorden: op de mate waarin de vraag een directe relatie heeft met het leerdoel.

Een vraag kan *niet* valide zijn of geacht worden, indien:

1. de inhoud van de vraag *niet* in overeenstemming is met het leerdoel;
2. er geen probleem wordt gesteld in de vraag;
3. de vraag triviaal is;
4. de vraag een strikvraag is;
5. de vraag met gezond verstand opgelost kan worden;
6. de vraag correct kan worden beantwoord, ook al is het leerdoel niet werkelijk bereikt;
7. het verkeerd beantwoorden van de vraag veroorzaakt wordt door andere factoren dan het niet bereikt zijn van het leerdoel. Als een vraag bijvoorbeeld zo moeilijk is of onduidelijk gesteld dat studenten eerst de vraag moeten ontcijferen alvorens zij deze kunnen beantwoorden, is deze vraag niet valide.

Hieronder volgen enkele voorbeelden:

Voorbeeld 1

Als een leerdoel luidt: 'De relatieve cumulatieve frequentie kunnen berekenen', dan is de vraag: 'Wat is een relatieve cumulatieve frequentie?' geen valide vraag, omdat hiermee een andere activiteit dan het kunnen berekenen wordt getoetst.

Toelichting: De vraag sluit niet aan bij het leerdoel (vergelijk punt 1).

Voorbeeld 2

niet:

Wat zijn de drie belangrijkste kenmerken van een parlementaire democratie en van een volksdemocratie?

wel:

Wat is het essentiële verschil tussen een parlementaire democratie en een volksdemocratie met betrekking tot het beginsel van evenredige vertegenwoordiging in het parlement?

Toelichting: Indien open vragen zijn gekozen omdat deze bij uitstek het redeneervermogen kunnen toetsen, moet men er geen reproductie mee toetsen (vergelijk punt 1)

Voorbeeld 3

Niet:

Wat is een relatieve cumulatieve frequentie?

wel:

Bereken de relatieve cumulatieve frequentie aan de hand van de volgende gegevens.

Objectiviteit

Een vraag is objectief indien de score op die vraag onafhankelijk is van de corrector die de scoring uitvoert. Dit impliceert dat

1. de procedure voor het beoordelen van het antwoord zo is geregeld dat geen beslissingen aan het oordeel van de corrector worden overgelaten en het antwoordmodel geen ruimte voor subjectieve interpretatie openlaat en
2. deskundigen het eens zijn over het juiste antwoord of over de criteria waaraan een juist antwoord moet voldoen.

Een vraag kan *niet* objectief zijn, indien

1. beslissingen met betrekking tot de kwaliteit van het antwoord aan het oordeel van de corrector worden overgelaten en het antwoordmodel ruimte open laat voor subjectieve interpretatie;
2. deskundigen het niet met elkaar eens zijn over het juiste antwoord of over de criteria waaraan een juist antwoord moet voldoen;
3. deze vraag naar de mening van de student;
4. er niet voldoende informatie in de vraag is opgenomen om het juiste antwoord te kunnen geven;

5. de vraag onvoldoende informatie geeft over de gewenste lengte en de vorm van het antwoord;
6. de kandidaat niet weet dat het antwoord gemotiveerd moet worden;
7. er geen antwoordmodel is.

Voorbeeld 1

Karel de Grote is geboren in .

Toelichting: De kandidaat heeft niet genoeg informatie om het juiste antwoord te kunnen geven, want bij deze vraag kan zowel een datum als een plaats worden ingevuld. (vergelijk punt 4)

Voorbeeld 2

Niet: Geef in het kort de filosofie van Descartes weer.

Beter: Geef in ongeveer 200 woorden de filosofie van Descartes weer.

Toelichting: Bij deze vraag is niet duidelijk wat de student moet verstaan onder 'in het kort'. (vergelijk punt 5).

Efficiëntie

Wanneer men bij het ontwikkelen van een open vraag kan kiezen uit verschillende soorten vragen die in alle opzichten gelijkwaardig zijn, dan verdient de vraag die het minste tijd kost voor het afnemen en corrigeren de voorkeur.

Een vraag kan *niet* efficiënt zijn of worden geacht, indien

1. deze onduidelijk en/of dubbelzinnig is geformuleerd;
2. er ingewikkelde zinsconstructies en woorden in voorkomen;
3. de zinnen te lang zijn;
4. er irrelevante informatie in de vraag staat;
5. er een dubbele ontkenning in staat;
6. tussenvoegsels de leesbaarheid verlagen;
7. deze onnodig negatief is gesteld;
8. het referentiekader van de vraag niet duidelijk is (hierdoor weet de kandidaat niet op welk standpunt hij zich moet stellen of vanuit welke invalshoek hij de probleemstelling moet aanpakken);
9. de vraag een te omvangrijk probleem omvat;
10. de vraag grammaticaal onjuist is geformuleerd;
11. de vraag door klemtoonverschuiving duidelijk van betekenis kan veranderen;
12. er teveel onnodig vakjargon in de vraag voorkomt;
13. bij de vraag figuren worden gebruikt die onduidelijk zijn of niet goed zijn afgestemd op de tekst/vraagstelling;

14. het betreffende leerdoel beter met een andere vraagvorm getoetst kan worden, met name met een meerkeuzevraag.

Voorbeeld 1

Niet: Geef de houding van Amerika tijdens de eerste jaren van de Tweede Wereldoorlog weer.

Wel: Geef in ongeveer 100 woorden aan welke houding Amerika innam tijdens de eerste jaren van de Tweede Wereldoorlog en illustreer deze houding aan de hand van de volgende gebeurtenissen (en deze gebeurtenissen vermelden).

Toelichting: vergelijk punt 9.

Voorbeeld 2

Niet: Beschrijf de rol van Amerika in het Midden-Oosten.

Wel:

In welk jaar werd Amerika daadwerkelijk betrokken in het Midden- Oosten conflict?

Waardoor werd Amerika daadwerkelijk betrokken in het Midden- Oosten conflict?

Door welke ontwikkeling nam het Amerikaanse belang in het Midden-Oosten toe?

Met welke middelen kan Amerika zijn invloed laten gelden in de betrokken landen?

Toelichting: vergelijk 1

Moeilijkheidsgraad

De moeilijkheidsgraad van een vraag moet aanvaardbaar zijn voor het niveau van de studenten. Met name bij het tentamineren zijn we geïnteresseerd in het verdelen van studenten over de twee categorieën 'voldoende beheersing van de leerstof' en 'onvoldoende beheersing van de leerstof'. Daarom willen we vooral de grensstudenten nauwkeurig van elkaar onderscheiden. Dit bereiken we door studenten vragen voor te leggen die zo goed mogelijk aansluiten bij hun niveau. Als we deze grensstudenten namelijk allemaal te moeilijke vragen voorleggen, is er geen scherp onderscheid tussen deze studenten mogelijk omdat de hele groep (nagenoeg) alle opgaven fout zal beantwoorden. Als de opgaven te gemakkelijk zijn, beantwoordt de hele groep grensstudenten (nagenoeg) alle vragen correct en is ook geen scherp onderscheid meer mogelijk.

Een vraag kan te moeilijk of te gemakkelijk zijn, indien

1. de vraag niet de aanvaardbare moeilijkheidsgraad voor het beoogde niveau van de studenten heeft;
2. de moeilijkheidsgraad verhoogd wordt door irrelevante gegevens aan de studenten te verstrekken.

V-2. Vormen van open vragen

Binnen de categorie open vragen onderscheiden we twee vormen:

- vragen die een kort antwoord behoeven: invul- en aanvulvragen
- vragen die een lang antwoord behoeven

Van beide soorten vragen geven we een korte omschrijving, enkele voorbeelden en de voor- en nadelen.

Invul- en aanvulvragen

Bij dit type vragen moet de student zelf een kort antwoord formuleren: één woord of een getal.

Voorbeeld 1

Wat is een ander woord voor 'vraag'?

Voorbeeld 2

Maak de volgende logische reeks af.

2 4 6 8 10 .

Voorbeeld 3

. is een periode van snelle fysieke volwassenwording met hormonale veranderingen die plaatsvindt aan het begin van de adolescentieperiode.

Voorbeeld 4

Gegeven: Op de grafiek van de functie $f: x \rightarrow -2+3$ ligt het punt $(a, -5)$

Opgave: Bereken a.

Antwoordmodel: 4

Voordelen van invul- en aanvulvragen:

- vaak laat een lang-antwoord-vraag zich opsplitsen in een aantal invul- en aanvulvragen
- in tegenstelling tot een lang-antwoord-vraag is de correctie hiervan niet zo tijdrovend
- omdat deze vragen minder afnametijd vragen dan lang-antwoord-vragen, kunnen er meer vragen opgenomen worden en dit verhoogt de betrouwbaarheid van de toets of het tentamen
- in vergelijking met gesloten vragen kunnen dit type vragen bij het toetsen van kennis meer een beroep doen op herinneren dan op herkennen
- in tegenstelling tot gesloten vragen is hier de kans op raden nagenoeg uitgesloten
- dit type vragen kan betrekkelijk objectief gescoord worden en met QMP volledig objectief.

Nadelen van invul- en aanvulvragen:

- het ontwikkelen van een antwoordmodel dat voldoende gedetailleerd is en automatisch nagekeken kan worden is moeilijk en arbeidsintensief
- ten behoeve van de correctie moet soms toch nog een corrector worden ingeschakeld
- de training van correctoren en het uitvoeren van de correctie is arbeidsintensief
- de correctie van dit type vragen vraagt meer tijd en financiële investering dan gesloten vraagvormen
- afhankelijk van de inhoud van de te toetsen leerstof kan het noodzakelijk zijn dat de corrector vakinhoudelijk deskundig is en dit heeft nog meer financiële consequenties.

Open vragen

Open vragen bestaan uit vragen of opdrachten die een uiteenzetting, een verklaring, een opsomming, een schets, een argumentatie of een samenvatting vragen en het hierbij behorende antwoordmodel. We geven enkele voorbeelden:

Voorbeeld 1

Gegeven:

In 1971 stortte het systeem van Bretton Woods ineen. Dit betekende dat de valutakoersen gingen zweven. Binnen bepaalde grenzen is dit nog steeds het geval.

Gevraagd:

Geef in het kort aan wat het verschil en de overeenkomsten zijn tussen vrij en beheerst zweven ('managed floating') van de koers.

Antwoordmodel: Bij het beheerst zweven is sprake van vrije koersvorming via vraag en aanbod op de valutamarkten (zoals bij vrij zweven), maar zullen de centrale banken/overheden ingrijpen/interveniëren bij al te grote fluctuaties op korte termijn of bij een ongewenste ontwikkeling op lange termijn.

Voorbeeld 2

Gegeven:

De ruim 9 miljoen werklozen van 1938 waren geen visitekaartje voor de New Deal. Toch is het niet juist om op grond hiervan de New Deal het etiket 'geheel mislukt' op te plakken.

Gevraagd:

Geef aan wat de mening van Van Rossem hierover is.

Antwoordmodel: Van Rossem meent dat de New Deal zowel mislukkingen als successen kende. Als mislukking noemt hij de bestrijding van de crisis. Als succes noemt hij onder andere: het afbreken van de (bekrompen) ethisch-politieke sfeer van de Progressive Movement, de stemmingsverandering ten goede onder het Amerikaanse volk en de verandering in de aard van het presidentschap door toedoen van het optreden van Roosevelt.

Voorbeeld 3

Aan welke vijf vereisten moet voldaan zijn, wil een inroeping van nietigheid op grond van art. 1377 BW kunnen slagen?

Antwoordmodel: 1. De schuldenaar moet een rechtshandeling hebben verricht. 2. Deze rechtshandeling moet onverplicht zijn verricht. 3. Hierdoor moet een schuldeiser in zijn verhaalsrecht zijn benadeeld. 4. De schuldenaar moet geweten of vermoed hebben dat benadeling van schuldeisers door de rechtshandeling zou optreden (wetenschap van benadeling). 5. Tenzij het een handeling om niet betreft moet ook sprake zijn van de wetenschap van benadeling bij degene met wie de schuldenaar heeft gehandeld.

Voordelen van standaard open vragen:

- Ze zijn geschikt voor het toetsen van leerdoelen met betrekking tot de compositie of structuur in een antwoord of werkstuk, oorspronkelijkheid en nauwkeurigheid van werken
- Het is de vraagvorm bij uitstek wanneer men wil toetsen hoe studenten gestructureerd omgaan met (relatief) open problemen
- Ze bieden de kans het niveau van persoonlijke expressie in de toetsing te betrekken
- Ze bieden de kans de schrijfvaardigheid en het uitdrukkingsvermogen van de studenten te toetsen
- Ze bieden de student de gelegenheid een 'eigen' redenering en argumentatie op te bouwen.

Nadelen van standaard open vragen:

- Ontwikkelen van een antwoordmodel en beoordelingsinstructie die voldoende exact en gedetailleerd zijn, is arbeidsintensief.
- Zelfs bij een voldoende gedetailleerd antwoordmodel blijft de objectiviteit van de beoordeling aanvechtbaar; zaken als redactionele vaardigheid van de student kunnen hier een rol spelen.
- Omdat dit type vragen ook meer afnametijd vergen, kunnen slechts een beperkt aantal van deze vragen in een toets of tentamen worden opgenomen en dit kan een negatief effect hebben op de betrouwbaarheid en validiteit.
- Bij tentamens moet voor dit type vragen een corrector worden ingeschakeld. De training van correctoren en de uitvoering van de correctie is arbeidsintensief.
- De correctie van dit soort vragen vergt meer tijd dan gesloten vraagvormen.
- Bij dit soort of tentamens is het noodzakelijk dat de corrector vakdeskundig is en dit feit heeft nog meer financiële consequenties.

V-3. Constructieregels voor standaard open vragen

In deze paragraaf geven we een aantal algemene constructieregels.

Maar eerst enkele opmerkingen vooraf.

In de eerste plaats geldt dat het aantal uitspraken niet uitputtend is. Het nut van deze verzameling is dat een aantal fouten aangegeven wordt, die in standaard open vragen kunnen voorkomen.

In de tweede plaats kunnen de gezichtspunten waaronder de uitspraken gecategoriseerd zijn niet beschouwd worden als elkaar uitsluitende categorieën. Zo kan men bijvoorbeeld een uitspraak als 'de vraag vraagt naar de mening van de student' niet alleen vangen onder gebrek aan objectiviteit (er is immers geen juist antwoord mogelijk), maar ook onder gebrek aan validiteit (wanneer meningsvragen niet omschreven zijn in de te toetsen leerdoelen). De reden voor deze invoering van gezichtspunten is

dat zij vraagconstructeurs een checklist (zie paragraaf 6) kunnen verschaffen, waardoor zij gemakkelijker en systematischer kritiek op (concept)vragen kunnen leveren.

Algemene constructieregels

Tijdens het construeren van open vragen moet op de volgende, algemene regels worden gelet:

1. Stel bij elke vraag de antwoordmogelijkheden op. Formuleer bij elke vraagstelling het gewenste goede antwoord of de goede antwoordmogelijkheden. Bedenk welke voor de hand liggende foute antwoorden door studenten gegeven kunnen worden.
2. Het formuleren van antwoordmogelijkheden kan leiden tot het aangeven van noodzakelijke responsrestricties of eventueel tot revisie van de vraagstelling.
3. Geef in de vraagstelling duidelijk aan wat van de student verlangd wordt en op welke criteria hij/zij beoordeeld wordt. Geef bovendien, indien nodig, de omvang van het verwachte antwoord aan.
4. Toets een omvangrijk probleem bij voorkeur door een serie vragen of samengestelde vragen en niet door een enkelvoudige vraag.

Deze voorkeur is primair gebaseerd op het gegeven dat het stellen van meer vragen of vraagonderdelen met betrekking tot één probleem beter duidelijk maakt of de student dit in al zijn facetten kan oplossen. De te meten doelstelling wordt beter gedekt door meer vragen dan door een enkele vraag.

Deze voorkeur is bovendien gebaseerd op de volgende overwegingen:

- een uitgesplitste vraagstelling is vaak minder dubbelzinnig dan een samengestelde vraag
- tijdens de beoordeling is het gemakkelijker telkens een deel van de beschikbare punten per onderdeel toe te kennen dan te bepalen voor welk deel een antwoord op een enkelvoudige vraag juist is.

V-4. Het gebruik van casussen

Een casus bestaat uit een beschrijving van een informatierijke context en een serie *vragen* over deze context. Bij tentamens worden casussen gebruikt om leerdoelen te toetsen die gericht zijn op hogere orde vaardigheden, zoals bijvoorbeeld het analyseren van situaties of het nemen van beslissingen. In de praktijk van het hoger onderwijs is de kennis over het gebruik van casussen bij tentamens vooral ontwikkeld in het medisch domein.

Onderstaande informatie kan tevens gebruikt worden voor het gebruik van casussen bij gesloten vragen.

De volgende richtlijnen worden aanbevolen voor het gebruik van casussen in QMP:

1. Gebruik casussen voor de toetsing van hogere orde vaardigheden.
2. Volg de algemene aanwijzingen voor het ontwikkelen en gebruiken van casussen van Schuwirth (2002).
3. Volg bij het gebruik van QMP voor het aanbieden van casussen het stroomschema "Hoe realiseer ik casussen in QMP (p. 8), waarbij de volgende regels gelden:
 - a. bij casussen met een stimulus van beperkte omvang en 1 vraag bij die stimulus kan ieder question type van QMP gebruikt worden.
 - b. bij casussen met een stimulus van een uitgebreide omvang met 1 of meer vragen waarbij studenten mogen (terug-) bladeren tijdens de toetsafname dient het QxQIntroFixed template van QMP gebruikt te worden.

- c. bij casussen met een stimulus met meer vragen waarbij studenten niet mogen (terug-) bladeren tijdens de toetsafname moet elke vraag of set van vragen in een apart Question Block worden gezet.
4. Beperk het aantal vragen bij een casus, zodat je meer casussen op 1 tentamen kunt aanbieden.
5. Gebruik op dit moment nog geen audio of video.

In het vervolg van dit document worden de achtergronden geschetst die hebben geleid tot deze richtlijnen. In paragraaf 2 wordt ingegaan op de onderwijskundige aspecten.

V-4.1. Wat is een casus?

In een casus wordt een situatie beschreven of informatie gegeven die de kandidaat moet gebruiken om een probleem op te lossen. Soms wordt het probleem in de casustekst gepresenteerd, soms in de vragen. Casussen bij tentamens bestaan uit twee onderdelen: de casus en een of meer vragen naar aanleiding van de casus (Schuwirth, 2002).

V-4.2. Wat is het verschil tussen casussen op papier en met de computer?

De verschillen tussen casussen op papier en met de computer kunnen liggen in een verschil in het aanbieden van de stimulus of een verschil in het beantwoorden van de vragen.

De verschillen in het aanbieden van de stimulus vloeien voort uit verschil in technologie. Bij casussen op papier kunnen tekst, foto's, grafieken en tabellen worden gepresenteerd. Bij casussen met de computer kan dit worden uitgebreid met geluid, video, animatie, documenten en eventueel links naar het web. In de summatieve toetsen (de tentamens) kan dit alleen als de locatie voor tentamenafname daartoe is ingericht. Dat is momenteel (lees: 2016) niet het geval.

Een verschil bij het beantwoorden van vragen is dat bij papieren afname de kandidaten kunnen bladeren. Bij een standaard schriftelijk tentamen kan een student nog eens terug naar eerder gemaakte vragen. Voor de student heeft dit als voordeel dat deze het werk kan controleren. Voor de maker van de toets is het nadeel van "bladeren" dat deze er voor moet zorgen dat in de vraagteksten geen aanwijzingen komen te staan waaruit het antwoord op eerdere vragen kan worden afgeleid. Bij afname met de computer heeft de ontwerper van de toets een extra optie: de technische mogelijkheid om het "terugbladeren in de toets" te blokkeren. Het voordeel voor de toetsontwerper is dat de casussen zo gemaakt kunnen worden dat de vragen logisch op elkaar volgen zonder dat er een zorg is dat de informatie bij de ene (latere) vraag aanwijzingen geeft voor het antwoord op een eerdere vraag. Deze optie maakt het mogelijk om casussen realistisch en praktijkgericht te maken.

Als het bladeren geblokkeerd wordt, dan biedt dat ook de mogelijkheid om tussentijds feedback te geven op de antwoorden van studenten. Bijvoorbeeld na elke vraag kan het correcte antwoord worden gepresenteerd. Dit heeft als voordeel dat elke student bij een volgende vraag dezelfde uitgangspositie in de casus heeft. Een nadeel kan zijn dat studenten zenuwachtig worden door gemaakte fouten.

Het blokkeren van terugbladeren is vooral relevant in domeinen waarin het handelen van de professional in relatie tot de casus belangrijk is. Bijvoorbeeld in een medische casus waarin een kandidaat eerst wordt gevraagd wat de diagnose is bij een ziektebeeld, welk onderzoek in het laboratorium uitgevoerd moet worden, en hoe de onderzoeksgegevens te interpreteren. De vragen "hangen" in elkaar, in de zin dat uit de vraagtekst bij een vraag het antwoord op een eerdere vraag kan worden afgeleid. In vakgebieden waarbij terugredeneren niet aan de orde is heeft het blokkeren van bladeren minder zin.

V-4.3. Wat voor typen casussen zijn er ?

In de internationale handboeken over toetsconstructie komt het woord *casustoets* niet voor. Haladyna (2004) gebruikt de volgende termen die betrekking hebben op het presenteren van een informatierijke context, met daaraan gekoppeld meerdere vragen:

- o Vignette or scenario item set
- o Problem solving item set
- o Pictorial item set.

Onder een *scenario* of *vignette* verstaat Haladyna een kort verhaal met daarin relevante informatie met het oog op het oplossen van een probleem. Een *problem solving item set* wordt door Haladyna vooral genoemd in relatie tot natuurwetenschappelijke problemen. Bij een *pictorial item set* neemt hij een complexe tabel als voorbeeld waarover vragen worden gesteld. Een uitbreiding van de classificatie van Haladyna zou kunnen zijn de *multimedia item set*. Dit is niet zozeer een systematische uitbreiding, maar legt de nadruk op stimuli die alleen via computerafname aangeboden kunnen worden.

In Nederland heeft Schuwirth (2002) over toetsen met casussen geschreven. Hij baseert zich vooral op toetsen in het medische domein. De casus zoals die beschreven wordt door Schuwirth komt overeen met de *vignette or scenario item set* van Haladyna.

Een onderscheid dat Schuwirth ook nog maakt is het verschil tussen een lange en een korte casus. Dit onderscheid is vooral relevant voor het samenstellen van toetsen. Als je wilt dat er voldoende variatie is in onderwerpen dan dienen casussen (gegeven een beperkte afnametijd voor de toets) niet te omvangrijk te zijn. De representativiteit van de toets naar het gehele domein waarbinnen de leerdoelen gedefinieerd zijn vraagt om een variatie naar leerstof-elementen, problematieken en contexten. Gegeven de beperkte afnametijd kan het dan zinvol zijn om wat meer kleinere casussen aan te bieden, dan 1 of 2 omvangrijke. In de praktijk wordt wel een optimum gevonden bij vijf vragen per casus. Ook in die situaties waar bijvoorbeeld wel 10 vragen over de casus te bedenken zijn. Ook wordt in de praktijk vaak gekozen voor ongeveer een zelfde hoeveelheid vragen per casus.

De verdeling van het aantal casussen in een tentamen wordt vastgelegd in het toetsprofiel. In het toetsprofiel wordt de balans gezocht tussen de representativiteit van de gehele toets ten opzichte van het domein van doelstellingen en de relevantie van elk van de afzonderlijke casussen ten opzichte van de doelstellingen.

V-4.4. Wanneer wordt een casus gebruikt bij een tentamen?

Haladyna (2004) stelt dat casussen vooral gebruikt kunnen worden bij het meten van *ability*. Wat een *ability* precies is wordt niet scherp gedefinieerd. Haladyna noemt *ability* in de tegenstelling tot *knowledge* en *cognitive skills*.

Zonder in een uitgebreide definitie discussie terecht te komen lijken de lijnen als volgt te lopen:

1. Multiple-choice vragen zijn geschikt voor leerdoelen die betrekking hebben op kennis van feiten (zoals herkennen of het geven van een definitie) en het uitvoeren van geïsoleerde cognitieve activiteiten (zoals classificeren, berekenen, het geven van een voorbeeld).
2. Casussen zijn geschikt voor leerdoelen die betrekking hebben op het herkennen van informatie in een rijke context (zoals het analyseren van een casus en herkennen van relevante informatie, het herkennen van een fout, het interpreteren van een tabel) of het nemen van een beslissing (kiezen van een geschikte onderzoeksmethode, procedure of treatment, het afsluiten van een behandeling, het doorverwijzen naar een andere discipline).

Als het pedagogisch kader van een opleiding de uitdrukkelijke eis stelt dat toetsen binnen het competentiegerichte onderwijs bruikbaar moeten zijn, dan kan een additionele afweging zijn dat het stellen van vragen in contextrijke situaties altijd het uitgangspunt moet zijn bij het tentamineren. In dat geval ligt het voor de hand om altijd casussen te gebruiken.

V-4.5. Welke aanwijzingen zijn er voor het gebruik van casussen bij tentamens?

Schuwirth geeft een aantal aanwijzingen voor het gebruik van casussen. Schuwirth werkt vooral in het medische domein. Hier is een lange traditie in het gebruik van casussen. De aanwijzingen (zie tabel X-1) zijn toch zo algemeen dat ze ook in andere vakgebieden bruikbaar zijn.

Tabel X-1. Aanwijzingen voor het gebruik van casussen (nagenoeg rechtstreeks overgenomen van Schuwirth (2002)).

Tips voor het toetsen met casussen

Tips voor casustoetsen in het algemeen

- Maak bij het construeren van een casustoets gebruik van een blauwdruk.
- Gebruik de ervaringen van mensen uit de praktijk voor het verzamelen van materiaal voor casussen.
- Leg de casus vóór afname onder studenten ter controle voor aan collega's of aan panels van experts.
- Verander de casus en/of vraag wanneer er onvoldoende overeenstemming over de juiste beslissing tussen collega's bestaat.
- Houd het scoringsmechanisme simpel.
- Werk voor het maken van casustoetsen samen met andere opleidingen of instituten.

Tips ten aanzien van het schrijven van de casus

- Gebruik de representatie van werkelijke situaties.
- Zorg ervoor dat de casusbeschrijving zo helder en concreet mogelijk is.
- Zorg voor voldoende inhoudelijke en contextuele informatie.
- Zorg voor voldoende 'negatieve' informatie.
- Presenteer de informatie ongeïnterpreteerd (ruw).
- Wees voorzichtig met het geven van afleidende informatie.
- Vermijd problemen die ook niet in de dagelijkse praktijk bestaan.

Tips ten aanzien van het schrijven van de vragen

- Relateer de vraag direct aan de casus.
- Richt de vraag zo veel mogelijk op essentiële beslissingen.
- Beperk het aantal beslissingen dat per casus wordt bevraagd.
- Formuleer de vraag zo helder mogelijk.
- Richt de vraag precies op de te bevragen beslissing.
- Laat de inhoud van de vraag de vorm definiëren in plaats van andersom.
- Zorg ervoor dat het bedoelde antwoord verdedigbaar correct is, en dat 'foute' antwoorden ook echt fout zijn.

V-5. Richtlijnen voor het nakijken van open vragen

De betrouwbaarheid van de toetsuitslagen neemt toe als er gewerkt wordt met meerder beoordelaars. Uit praktische overwegingen is dit echter niet altijd haalbaar. Schakel indien mogelijk twee correctoren in bij het nakijken van lang antwoordvragen bij cursussen met 50 kandidaten of meer en zorg voor een controle op de correctie bij minder dan 50 kandidaten door altijd 10% van de tentamens dubbel te scoren.

Bij het werken met twee correctoren kan voor de uitvoering van het nakijken gekozen worden om (1) de na te kijken toetsen te verdelen over twee (of meer) correctoren, of (2) een dubbele correctie van (een deel van) de items. Bij keuze (1) kunnen de gemiddelde scores vergeleken worden tussen correctoren.

Deze mogen niet significant afwijken. Is dat wel het geval, dan dient er overleg te zijn tussen de correctoren over het hanteren van het antwoordmodel of de criteria. Bij keuze (2) dient de interbeoordelaars-betrouwbaarheid nagegaan te worden (Cohen's Kappa). Deze kappa dient minimaal 0.60 te zijn en bij voorkeur boven de 0.80 te liggen. Is de kappa lager dan 0.60, dan dient er overleg te zijn tussen de correctoren over het hanteren van het antwoordmodel of de criteria.

V-6. Het opstellen van een antwoordmodel

Bij langantwoord vragen dient er sprake te zijn van een modelantwoord of van beoordelingscriteria (als er meer goede antwoorden mogelijk zijn). Bij modelantwoord en beoordelingscriteria kan een puntenverdeling aanwezig zijn voor de onderdelen van het antwoord. Er is geen puntenverdeling per criterium omdat de criteria elkaar (kunnen) beïnvloeden. Een modelantwoord wordt bekendgemaakt bij de toetsuitslag. De beoordelingscriteria dienen van tevoren bekendgemaakt te worden in oefentamens en te worden herhaald bij de toetsuitslag. Het antwoordmodel omvat modelantwoorden en een puntenverdeling.

We zullen de constructieregels ervan hieronder bespreken.

1. Geef, indien mogelijk, in het antwoordmodel zo precies en zo volledig mogelijk aan wat het juiste antwoord moet zijn. Indien er niet één juist antwoord is geef dan een opsomming van alle elementen die in een antwoord zouden moeten voorkomen. Geef ook aan welk antwoord zeker fout is.
2. Geef beoordelingscriteria indien het onmogelijk is bij een vraag een ideaalantwoord te formuleren.
3. Soms moet er nog bijkomende informatie aan de beoordelaars gegeven worden. Zo zou men hier kunnen ingaan op aspecten als:
 - a. de wiskundige bewerking werd weliswaar juist uitgevoerd, maar er werd een verkeerde maateenheid gebruikt
 - b. de redenering was juist, maar een rekenfout leidde tot een verkeerde oplossing
 - c. een student in de eerste opgave al een verkeerd numeriek antwoord heeft gegeven, dan moet de volgende opgave waarin met dit foute antwoord doorgerekend moet worden, als juist beschouwd worden, als de student wel de juiste rekenkundige operatie heeft uitgevoerd maar een verkeerd antwoord heeft verkregen.

Probeer zoveel mogelijk dit soort aspecten te voorzien en geef duidelijk aan hoe een antwoord in dit geval beoordeeld moet worden. We geven twee voorbeelden:

Voorbeeld 11

Een student krijgt de opdracht een recensie van een boek te schrijven.

Bij deze vraag is het onmogelijk een modelantwoord op te stellen (zie punt 2). Derhalve dienen beoordelingscriteria te worden opgesteld. Bij dit voorbeeld zouden we de volgende beoordelingscriteria kunnen geven:

- 1 Opbouw van recensie
 - Is het voor een niet-lezer duidelijk wat de inhoud van het boek is?
 - Maakt de student onderscheid tussen de inhoud van het boek en zijn eigen mening?
 - Is zijn mening onderbouwd met een steekhoudende argumentatie?
 - Komt de student tot een conclusie?

 - 2 Verifieerbaarheid
 - Is de weergave van feiten en meningen juist en betrouwbaar?

 - 3 Verantwoord taalgebruik
 - Is het taalgebruik formeel correct? (spelling, zinsbouw en interpunctie)
-

Voorbeeld 2

Een student krijgt de opdracht aan de hand van bepaalde prestatiegegevens op een toets te bepalen of de experimentele groep qua prestatie van de controlegroep verschilt. De vraag luidt nu: Maak de berekening en geef aan welke conclusie op basis van deze berekening getrokken moet worden.

Toelichting: In deze vraag kunnen we twee elementen onderscheiden, namelijk 1. de berekening en 2. de te trekken conclusie. In het antwoordmodel moeten we dus beide elementen opvoeren (zie punt 3).

V-7. Screeningsregels voor open vragen

Wanneer een vraagconstructeur, rekening houdend met de hiervoor genoemde richtlijnen en opmerkingen, een concept-vraag heeft geschreven, kan het toch voorkomen, dat het concept op grond van die richtlijnen een revisie behoeft. Om dit snel te kunnen nagaan moeten de volgende drie vragen met 'ja' kunnen worden beantwoord:

1. Is de vraag gebaseerd op een leerdoel?
2. Is de vraag inhoudelijk correct?
3. Is een antwoordmodel aanwezig?

Is dit niet het geval dan vertoont de concept-vraag zodanige tekorten dat revisie noodzakelijk is.

Indien de vragen wel met 'ja' kunnen worden beantwoord, dan is het goed de concept-vraag nog eens te screenen aan de hand van de twee gedetailleerde checklists (zie hierna). Als op alle vragen met 'ja', respectievelijk 'nee', kan worden geantwoord, voldoet de vraag aan de gestelde eisen.

V-8. Checklists

Hieronder vind je een checklist voor open vragen en een checklist voor het antwoordmodel bij open vragen.

Checklist voor open vragen

criterium		Ja*	Nee*
Validiteit	1. Is de inhoud van de vraag in overeenstemming met het leerdoel?	*	
	2. Is de vraag vakinhoudelijk juist?	*	
	3. Is de vraagvorm geschikt voor de te toetsen doelstelling?	*	
	4. Is er een duidelijk probleem in de vraag gesteld?	*	
	5. Is de vraag met gezond verstand op te lossen?		*
Objectiviteit	1. Zijn de deskundigen het eens over het (de) goede antwoord(en)?	*	
	2. Bevat de vraag voldoende informatie om het juiste antwoord te kunnen geven?	*	
	3. Is er een antwoordmodel?	*	
	4. Geeft de vraag voldoende informatie over de lengte en vorm van het gewenste antwoord?	*	
	5. Vraagt de vraag naar de mening van de student?		*
Efficiëntie	1. Is de vraag duidelijk en ondubbelzinnig geformuleerd?	*	
	2. Is de vraag grammaticaal juist geformuleerd?	*	
	3. Zijn eventuele figuren e.d. functioneel?	*	
	4. Zijn eventuele figuren bij de vraag duidelijk?	*	
	5. Zijn eventuele figuren bij de vraag correct?	*	
	6. Is het referentiekader van de vraag duidelijk?	*	
	7. Is de vorm van de vraag de meest geschikte vraagvorm?	*	
	8. Bevat de vraag meer informatie dan nodig is?		*
	9. Bevat de vraag een dubbele ontkenning?		*
	10. Is de vraag onnodig negatief gesteld?		*
	11. Is de formulering van de vraag te ingewikkeld?		*
	12. Bevat de vraag onnodig vakjargon of onnodig moeilijke woorden?		*
	13. Omvat de vraag een te omvangrijk probleem?		*
Moeilijkheidsgraad	1. Heeft de vraag een aanvaardbare moeilijkheidsgraad voor het bedoelde niveau?	*	
	2. Wordt de moeilijkheidsgraad verhoogd door irrelevante gegevens?		*
* Sommige vragen moeten beantwoord worden met 'ja' en sommige met 'nee'. De vakken met de wenselijke antwoorden zijn grijs en bevatten een *.			

Checklist voor antwoordmodel			
	Criteria	Ja*	Nee*
	1. Is voor elke vraag een antwoordmodel opgesteld?	*	
	2. Zijn de antwoorden in het antwoordmodel realistisch?	*	
	3. Is, waar nodig, duidelijk aangegeven wat er in de antwoorden van kandidaten als 'niet juist' of 'niet geheel juist' beoordeeld moeten worden?	*	
	4. Zijn in het antwoordmodel duidelijke beoordelingsinstructies opgenomen over de wijze van beoordelen?	*	
	5. Indien het onmogelijk is bij een of meer vragen een modelantwoord te formuleren, zijn er dan beoordelingscriteria in het antwoordmodel opgenomen?	*	
	6. Is de lay-out van het antwoordmodel zodanig dat de beoordelaar snel inzicht kan krijgen in zijn of haar beoordelaarstaak?	*	
	7. Is het antwoordmodel te algemeen, te globaal, waardoor van een uniforme beoordeling weinig terecht komt?		*
	8. Laat het antwoordmodel ruimte open voor subjectieve interpretatie?		*
* Sommige vragen moeten beantwoord worden met 'ja' en sommige met 'nee'. De vakken met de wenselijke antwoorden zijn grijs en bevatten een *.			

V-9. Richtlijnen voor de lay-out van open vragen

Regel

- *kleine tab* achter a en b
- *puntenaanduiding* staat achter de sluitpunt, tussen haakjes en is cursief

Voorbeeld

-
1. a. Aan welke eis moet een theorie volgens Popper voldoen om falsifieerbaar te zijn? (3 punten)
 - b. Geef een voorbeeld van een theorie die volgens Popper niet falsifieerbaar is. (1 punt)
 - c. Waartoe dient het begrip empirische inhoud volgens Popper? (2 punten)
-

Opmaak en formulering feedback bij open vragen

De feedback bij de open vragen moet bestaan uit de volgende zinnen:

Voorbeeld

De open vragen moeten nog nagekeken worden. Daarom heeft u nu nog geen punten voor deze vraag ontvangen.

Het juiste antwoord bevat de volgende inhoud:

[...toevoegen juiste antwoord of antwoordrichting]

VI. Het samenstellen van een tentamen en het bepalen van de voorlopige cesuur

Deze sectie is bedoeld voor degenen die een tentamen moeten samenstellen en screenen, de examinatoren. Het doel van deze sectie is de examinatoren handvatten te bieden bij het samenstellen en screenen van een tentamen.

De volgende onderwerpen komen aan de orde:

1. Stappen van het samenstellingsproces van tentamens
2. Screeningsregels voor een tentamen

VI-1. Stappen in het samenstellingsproces van tentamens

Na het construeren en screenen van vragen moet de examinator uit de verzameling vragen een tentamen samenstellen. Het samenstellen van een tentamen dient te gebeuren conform de eerder opgestelde toetmatrijs en het tentamenprofiel (zie sectie III).

Het samenstellen van een tentamen verloopt in een vijftal stappen:

1. selecteren van vragen
2. bepalen van de volgorde van de vragen
3. bepalen van de puntenwaardering
4. vaststellen van de voorlopige cesuur
5. opstellen beoordelingsinstructie
6. opstellen van een gedetailleerde instructie voor studenten

Deze stappen dienen te resulteren in een tentamen bestaande uit een serie vragen (de stappen 1 en 2), een bijbehorend antwoordmodel/correctievoorschrift (stap 3), de voorlopige cesuur (stap 4) en een instructie voor de studenten (stap 5).

Stap 1: selecteren van vragen

Zoals eerder opgemerkt dient de samenstelling van het tentamen overeen te komen met het eerder opgestelde tentamenprofiel. Als bijvoorbeeld besloten is dat een bepaald leerstofonderdeel in een tentamen met uitsluitend gesloten vragen belangrijker is dan andere onderdelen, dan moeten over dat onderdeel ook meer vragen worden geselecteerd.

Nadat de vragen geselecteerd zijn, dient men na te gaan of het tentamen een acceptabele moeilijkheidsgraad bezit, dat wil zeggen of de moeilijkheidsgraad van het tentamen aansluit bij het niveau dat met de cursus beoogd is. Deze laatste eis wil concreet zeggen dat we de studenten niet moeten tentamineren met opgaven die te moeilijk of te makkelijk voor hen zijn. Wij zijn immers bij het tentamineren vooral geïnteresseerd in het verdelen van de studenten over de categorieën 'voldoende' en 'onvoldoende' beheersing van de leerstof. Daarom willen we vooral de zogenaamde grensstudenten zo nauwkeurig mogelijk van elkaar onderscheiden. Dit bereiken we door deze studenten opgaven voor te leggen die zo goed mogelijk bij hun vaardigheidsniveau aansluiten. Als we deze grensstudenten namelijk te moeilijke vragen geven, is vrijwel geen onderscheid mogelijk, omdat de hele groep (nagenoeg) alle opgaven fout zal beantwoorden. Als de opgaven te gemakkelijk zijn, maakt de hele groep grensstudenten (vrijwel) alle opgaven goed en is ook nagenoeg geen onderscheid tussen studenten meer mogelijk.

Stap 2: bepalen van de volgorde van de vragen

Na de selectie van vragen moet de volgorde van die vragen binnen een tentamen worden vastgesteld. Hierbij valt te denken aan (een combinatie van) de volgende opties:

- een opbouw van relatief gemakkelijke vragen (gelegenheid om te wennen aan de tentamensituatie) via relatief moeilijke naar relatief gemakkelijke vragen
- een volgorde van vragen in het tentamen geclusterd naar leerstofonderwerp/leereenheid/blok/hoofdstuk
- een tentamen bestaande uit eerst een deel met de gesloten vragen en vervolgens een deel met alleen open vragen

Deze stap resulteert in een tentamen met een x-aantal vragen in een bepaalde volgorde. Daarbij hoort ook een antwoordmodel. Bij gesloten vragen bestaat zo'n antwoordmodel uit een opsomming van de juiste alternatieven, bij open vragen uit een opsomming van goede, minder goede en foute antwoorden en eventueel voorzien van een beoordelingsinstructie per vraag (zie secties IV en V).

Stap 3: bepalen van de puntenwaardering

Na bovengenoemde stap hebben we nu een tentamen bestaande uit een x-aantal vragen met een bijbehorend antwoordmodel. Dit is voldoende om na afname van het tentamen te kunnen nagaan welke vragen studenten goed en fout hebben beantwoord. Om na afloop van het tentamen de uitslag te kunnen bepalen (is de student gezakt of geslaagd; wat is zijn/haar tentamencijfer?) moeten punten toegekend worden aan de antwoorden. Dit volgens bepaalde regels toekennen van punten aan waargenomen prestaties noemen we scoren. Dit betekent dat voorafgaand aan de afname van een tentamen -dus in de fase van vraag- en tentamenconstructie- reeds moet zijn bepaald hoe de scoring dient plaats te vinden. In deze stap gaat het erom te bepalen en vast te leggen hoeveel punten een goed, gedeeltelijk goed of fout beantwoord vraag kan opleveren en hoeveel punten maximaal te behalen zijn voor het gehele tentamen.

Bij het bepalen van de puntenwaardering lopen twee zaken door elkaar, die moeilijk van elkaar te scheiden zijn: enerzijds de scoring van de vragen en anderzijds de weging van vragen en/of van tentamenonderdelen. Onder 'weging' verstaan we in dit verband het toekennen van verschillende/variërende aantallen maximaal te behalen punten per vraag of tentamenonderdeel. De redenen die men kan hebben om vragen of tentamenonderdelen te wegen zijn te herleiden tot (vermeende) intrinsieke kwaliteitsverschillen tussen vragen of onderdelen zoals de belangrijkheid van de te toetsen doelstelling, de geschatte moeilijkheidsgraad, de tijd die nodig is voor beantwoording en de aard van het te toetsen gedrag. Indien men besluit bepaalde vragen of tentamenonderdelen te wegen, dan dient men rekening te houden met de opgestelde toetsmatrijs en het tentamenprofiel, aangezien ook hierin al is aangegeven welke leereenheden (of secties, blokken) belangrijk worden geacht.

Hoewel het lastig is scoring en weging van elkaar te scheiden, zullen we toch afzonderlijk van elkaar de scoring van gesloten en open vragen en de mogelijkheden van weging binnen een tentamen behandelen.

Scoring

Scoring van gesloten vragen. Binnen de OU is bij gesloten vragen gekozen voor het 'gedwongen raden'-scoringssysteem. Bij scoring volgens dit systeem moet de student bij elk vraag een alternatief aanstrepen en dus geen vragen openlaten. De scoringsregel luidt: Goed antwoord- 1 punt/2 punten/3 punten, etc. geen of fout antwoord 0 punten.

Scoring van open vragen. Ook bij open vragen moet vooraf worden bepaald en vastgelegd hoe naderhand gescoord gaat worden. Per vraag of vraagonderdeel moet worden bepaald hoeveel punten maximaal te behalen zijn. Met uitzondering van zogenaamde invul-/aanvulvragen, die in veel gevallen een dichotome scoring kennen (slechts twee mogelijkheden in het antwoord worden onderscheiden:

goed of fout), is het bij open vragen gebruikelijk per vraag een scoreschaal van meerdere waarden te kunnen aanbrengen tussen een volledig en een gedeeltelijk juist antwoord: bijv. een 4-puntsschaal waarbij 0,1,2 of 3 punten voor het antwoord op een vraag kunnen worden toegekend. Als scoreregels zal duidelijk moeten worden aangegeven welke antwoorden in bovengenoemd voorbeeld 0,1,2 of 3 punten opleveren. Is sprake van bonus- en/of aftrekpunten dan zal dit ook duidelijk tot uitdrukking moeten komen.

Weging

Wegen per vraag. Bij tentamens met gesloten vragen komt weging per vraag momenteel nauwelijks voor; meestal worden alle vragen als even belangrijk beschouwd en levert elk juist beantwoordde vraag evenveel punten op. Onderzoek heeft uitgewezen dat het van tevoren afzonderlijk wegen van gesloten vragen nauwelijks invloed heeft op de rangorde van studenten op de scoreschaal en dat het zelfs kan leiden tot betrouwbaarheidsverlies.

Een reëler alternatief is om af te zien van weging en het belang van bepaalde leerdoelen tot uitdrukking te laten komen door het naar rato verhogen van het aantal vragen in het tentamen dat op die leerdoelen betrekking heeft.

Weging van open vragen komt in tentamens veelvuldig voor. Bij open vragen is er vaak sprake van meer antwoordmogelijkheden, waarvan er een of meer als juist te kenschetsen zijn en een aantal als gedeeltelijk juist. De scoringsmogelijkheden zijn dan polytoom. Een open vraag met meer antwoordmogelijkheden kan dus meer gewicht krijgen dan een open vraag die minder antwoordmogelijkheden of gradaties kent.

Wegen per tentamenonderdeel. Als een tentamen zowel uit een open- als een meerkeuzedeel bestaat, is het vaak nodig de puntentelling van beide delen bij elkaar aan te passen. Als het open- en meerkeuzedeel van een tentamen qua belangrijkheid en qua maximaal te behalen aantal punten identiek aan elkaar zijn, dan hoeven de scores op beide delen slechts bij elkaar opgeteld te worden. Een weging is dan niet nodig. Als het maximum aantal te behalen punten op een van de twee delen niet in overeenstemming is met de belangrijkheid van dat deel, moet er gewogen worden. Stel dat X_g de score van een student op het gesloten gedeelte is en X_o de score van een student op het open gedeelte, dan wordt de eindscore X_e van de student als volgt bepaald:

$$X_e = X_o + w X_g$$

Als we $w = 1$ kiezen, betekent dit dus dat een punt dat de student op een gesloten vraag behaalt, even zwaar weegt als een punt dat via een open vraag verdiend wordt. Zo kunnen we punten van gesloten vragen zwaarder laten wegen door $w > 1$ te kiezen.

Het opstellen van een antwoordmodel

We zullen hier niet ingaan op het antwoordmodel voor gesloten vragen, aangezien dit eenduidig is en niet relevant voor correctoren. Wel zullen we uitvoerig stilstaan bij het antwoordmodel voor open vragen. Nadat de volgorde van de vragen is bepaald moet de scoring (en dus ook de weging) per vraag worden vastgelegd in het antwoordmodel.

We geven een aantal regels voor het opstellen van een dergelijk antwoordmodel:

1. Ga per vraag na welke elementen we in het antwoord kunnen onderscheiden.
2. Probeer de elementen van een antwoord te wegen, bijvoorbeeld door aan te geven welk element belangrijker is en wat de puntenverhouding is tussen de elementen.
3. Geef voor elk element van een vraag een maximale score van die vraag in relatie tot het hele tentamen.

4. Hanteer de volgende lay-out per vraag voor het scoringsvoorschrift:

	Totaal aantal punten voor alle elementen (subvragen)	. punten
A	Element (subvraag) 1	. punten
B	Element (subvraag) 1	. punten
c	Element (subvraag) 1	. punten

5. Geef het maximaal aantal te behalen punten voor het gehele tentamen aan.

We geven twee voorbeelden:

Voorbeeld 1

Een student krijgt de opdracht aan de hand van bepaalde prestatiegegevens op een toets te bepalen of de experimentele groep qua prestatie van de controle-groep verschilt. De vraag luidt nu: Maak de berekening en geef aan welke conclusie op basis van deze berekening moet worden getrokken.

Toelichting: In deze vraag kunnen we twee elementen onderscheiden, namelijk de berekening en de te trekken conclusie. In het antwoordmodel moeten we dus deze twee elementen opvoeren (zie regel 1.)

Voorbeeld 2

In het bovenstaande voorbeeld zou de examinerator kunnen besluiten dat de berekening belangrijker is dan de conclusie en dat aan de totale vraag 10 punten worden toegekend.

Scoringsvoorschrift:

Juiste berekening	7 punten
Juiste conclusie	3 punten
Totaal	10 punten

Toelichting: zie regels 2. en 4.

Stap 4: vaststellen van de voorlopige cesuur

Om na afloop van een tentamen te kunnen bepalen of een student gezakt dan wel geslaagd is, moet de grens tussen voldoende en onvoldoende bekend zijn. Deze grens tussen de hoogste tentamenscore waaraan nog een onvoldoende en de laagste tentamenscore waaraan een voldoende wordt toegekend, noemen we de cesuur. Er bestaan verschillende methoden om de cesuur te bepalen. We zullen hier slechts de methode behandelen die binnen de OU wordt gehanteerd. Bij regulier schriftelijke tentamens en CBG wordt de cesuur voorafgaand aan de afname van het tentamen, *voorlopig* vastgesteld. Nadat het tentamen is afgenomen kan de cesuur, mede op grond van de toets- en vraaganalyse en eventueel reacties van studenten, worden bijgesteld.

Bij het vaststellen van de voorlopige cesuur wordt bij de Open Universiteit de volgende algemene beslisregel gehanteerd: cesuur = (raadkans) + 55% (van het restant) van het maximaal aantal te behalen punten. De 55% is het gewenste beheersingsniveau om een cursus met een voldoende af te sluiten.

De berekening is afhankelijk van de samenstelling van het tentamen:

Tentamensamenstelling	cesuurberekening
1. tentamen met alleen gesloten vragen	raadkans + 55% van het restant
2. tentamen met alleen open vragen	55% van het maximaal aantal te behalen punten voor het gehele tentamen
3. tentamen met gesloten en open vragen	raadkans + 55% van het restant (gesloten vragen) + 55% van het maximaal aantal te behalen punten voor het open gedeelte

De uitkomst van bovenstaande cesuurberekening levert slechts één getal op, namelijk de laagste tentamenscore waaraan nog een voldoende wordt toegekend (de zogenaamde grensscore). Levert de berekening het getal 27 op, dan ligt de cesuur op 26/27 punten, waarbij 26 de hoogste onvoldoende tentamenscore is en 27 de laagste voldoende tentamenscore.

Alleen bij gesloten vragen kan het effect van 'raden' optreden. Als een student het juiste antwoord niet weet, kan hij daarnaar raden. Afhankelijk van het aantal alternatieven heeft de student meer of minder kans het juiste alternatief aan te strepen: bij tweekeuzevragen 50%, bij vierkeuzevragen 25%. Binnen de OU wordt deze raadkans verdisconteerd in de cesuurbepaling. De raadkans is het aantal punten dat de student alleen al door blind raden kan behalen: maximaal aantal te behalen punten voor alle gesloten vragen samen maal de kans dat bij een gesloten vraag het juiste alternatief wordt aangestreept (bij tweekeuzevragen: 50%, bij vierkeuzevragen: 25%).

55% van het restant: met 'restant' wordt bedoeld het maximaal aantal punten te behalen voor de gesloten vragen tezamen minus de raadkans, bijvoorbeeld: bestaat een tentamen uit 40 vierkeuzevragen die elk 1 punt opleveren, dan is de raadkans $1/4 * 40 = 10$ punten. Het restant bedraagt dan 30 punten ($40p - 10p = 30p$).

Stap 5: opstellen beoordelingsinstructie

Eén van de belangrijke aandachtspunten bij een tentamen dient de rechtvaardige behandeling van de student te zijn. Dit houdt in dat studenten niet alleen dezelfde vragen krijgen voorgelegd, maar ook volgens dezelfde standaard beoordeeld worden. Grondslag van een rechtvaardige beoordeling is dat gelijke prestaties ook gelijk beoordeeld worden. Naast het aandachtspunt van 'rechtvaardige beoordeling' zijn validiteit en betrouwbaarheid van een tentamen voor een belangrijk deel afhankelijk van de beoordeling. Het verschil tussen open en gesloten vragen is, wat betreft de beoordeling, de mate waarin die beoordeling objectief tot stand komt. Onder volledig objectieve beoordeling moet men verstaan dat een zodanig geheel van voorschriften voorhanden is dat een niet-deskundige in het vakgebied de score zou kunnen bepalen.

Bij gesloten vragen is zo'n set van voorschriften aanwezig in de vorm van een antwoordmodel (opsomming van de juiste alternatieven per vraag) en een bepaalde scoringsregel. De correctie en scoring verlopen dan objectief, dat wil zeggen onafhankelijk van degene die deze uitvoert en kan daardoor aan de computer worden overgelaten.

Open vragen zijn in mindere mate objectief scorebaar, enerzijds omdat de set van voorschriften vaak minder eenduidig en volledig kan zijn, anderzijds omdat er in tegenstelling tot gesloten vragen een (inhoudelijke) beoordeling moet plaatsvinden door correctoren (meestal studiebegeleiders). Hier doet zich dan het gevaar voor dat bij de correctie en scoring subjectieve elementen een rol gaan spelen. Het feit dat een corrector niet altijd een 'eerlijke' beoordeling produceert, houdt geen verwijt in aan de correctoren (deze werken meestal oprecht en consciëntieus), maar kan verklaard worden uit het feit dat zijn/haar oordeel door allerlei factoren onbewust kan worden beïnvloed.

Voorbeelden hiervan zijn:

- correctoren hebben een verschillende opvatting over het te beoordelen criterium
- andere kenmerken van het te beoordelen antwoord kunnen invloed hebben op de beoordeling (bijvoorbeeld: een vloeiende schrijfstijl kan een hogere waardering krijgen dan een houderige stijl)
- een middelmatig antwoord na enkele uitstekende antwoorden kan een lagere beoordeling krijgen dan datzelfde middelmatige antwoord beoordeeld na enkele slechte antwoorden.

Het is dus zaak deze invloeden zoveel mogelijk onder controle te houden. Eén van de maatregelen die zeer nuttig is gebleken om dit soort invloeden te minimaliseren, is gebruik te maken van een beoordelingsinstructie voor het tentamen in zijn totaliteit naast het al eerder opgestelde antwoordmodel. In deze beoordelingsinstructie zijn algemene richtlijnen voor de beoordeling vermeld.

Elke corrector ontvangt samen met de afgelegde tentamens van de studenten een formulier met standaardinstructies. Hierin staat vermeld wat de corrector allemaal krijgt toegezonden, hoe hij/zij de tentamens moet beoordelen en hoe de scoreformulieren (Optische Leesbare Formulieren, OLF) moeten worden ingevuld. Deze informatie is zeer algemeen en niet toegesneden op elk afzonderlijk tentamen. Indien nodig moet de examinerator dus aanvullende instructies voor de corrector opstellen, bijvoorbeeld:

- beoordeel de tentamens per vraag. Men kan zo beter de antwoorden vergelijken en relateren aan het antwoordmodel
- spel-, taal- en grammaticale fouten mogen de student niet worden aangerekend, tenzij hierdoor het antwoord inhoudelijk niet correct is.

Stap 6: opstellen van een gedetailleerde instructie voor studenten

Elke student krijgt bij de tentamenafname een formulier met standaardinstructies uitgereikt, de 'Instructie voor de kandidaat'. Hierin staat vermeld welke gegevens studenten moeten invullen, hoe de tentamenopgaven beantwoord dienen te worden, hoe bij schriftelijke tentamens het antwoordblad (OLF) moet worden ingevuld en wat studenten moeten inleveren dan wel mee mogen nemen. Deze informatie kan echter niet toegesneden zijn op elk afzonderlijk tentamen. Daarom dient de examinerator, voor zover nodig, een aanvullende 'gedetailleerde instructie voor studenten' op te stellen.

Deze instructie omvat tentamenspecifieke informatie zoals:

- afwijkende tijdsduur van het tentamen;
- toelichting op de samenstelling van het tentamen (uit hoeveel en welk soort vragen bestaat het tentamen?);
- toelichting op de puntenwaardering (wat is het maximaal aantal te behalen punten voor het gehele tentamen en per vraag?);
- toegestane hulpmiddelen;
- vermelden van de opmerking dat 'de vragen van dit tentamen gebaseerd zijn op het cursusmateriaal dat hoort bij de cursus. De antwoorden behoren in overeenstemming te zijn met de inhoud van de cursus';
- advies om eerst de gesloten vragen te beantwoorden en dan pas de open vragen.

Deze instructie vormt het tweede blad van het tentamenopgavenformulier.

VI-2. Screeningsregels voor tentamens

Zodra alle stappen uit het bovenstaande zijn doorlopen, is een tentamen samengesteld met bijbehorend correctievoorschrift (dus inclusief antwoordmodel, scoringsvoorschrift en beoordelingsinstructie) en instructie voor de student.

Voordat het tentamen aan het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek wordt overgedragen is het noodzakelijk om het tentamen in zijn geheel te screenen. Om na te gaan of het tentamen de gewenste vorm en inhoud heeft, moeten de vragen van de volgende checklist

met ja (met uitzondering van vraag 8 bij het correctievoorschrift) kunnen worden beantwoord.

Checklist voor tentamen, correctievoorschrift en instructie voor studenten			
criterium		Ja*	Nee*
t.a.v. tentamen	1. Is het tentamen een redelijke representatie van de te toetsen leerdoelen?	*	
	2. Is het tentamen in overeenstemming met de opgestelde toetsmatrijs en het tentamenprofiel?	*	
	3. Is de moeilijkheidsgraad van het tentamen aanvaardbaar voor het niveau van de cursus?	*	
	4. Staat de tijd nodig voor het beantwoorden van elk vraag afzonderlijk in redelijke verhouding tot de relevantie van de vraag?	*	
	5. Kan een 'gemiddelde' student het tentamen binnen de daarvoor beschikbare drie uur afleggen?	*	
t.a.v. correctie- voorschrift	1. Is het tentamen objectief scorebaar?	*	
	2. Is het correctievoorschrift nauwkeurig?	*	
	3. Is waar nodig duidelijk aangegeven wat er in de antwoorden van de studenten als 'niet geheel juist' of 'niet juist' beoordeeld moeten worden?	*	
	4. Is duidelijk aangegeven hoeveel punten aan de juiste elementen uit het antwoord toegekend kunnen worden?	*	
	5. Zijn in het correctievoorschrift richtlijnen van algemene aard opgenomen?	*	
	6. Is in het correctievoorschrift een duidelijke beoordelingsinstructie opgenomen over de wijze van beoordelen?	*	
	7. Indien het onmogelijk is bij een of meer vragen een antwoordmodel te formuleren, zijn er dan beoordelingscriteria in het correctievoorschrift opgenomen?	*	
	8. Is het correctievoorschrift niet te gedetailleerd zodat de beoordelaars het door de omvang moeilijk kunnen hanteren?		*
	9. Is de lay-out van het correctievoorschrift zodanig dat de beoordelaar snel inzicht kan krijgen in zijn taak?	*	
	10. Is bij het correctievoorschrift rekening gehouden met de restricties van het OU-tentamenverwerkend systeem ?	*	
t.a.v instructie voor student	1. Is de instructie voor de student helder en eenduidig?	*	
	2. Weet de student wat van hem/haar verwacht wordt?	*	

* De vragen moeten beantwoord worden met 'ja' (m.u.v. vraag 8). De vakken met de wenselijke antwoorden zijn grijs en bevatten een *.

VII. Analyse en definitieve cesuur

Deze sectie is met name bedoeld voor degenen die de definitieve cesuur moeten vaststellen, namelijk de examinatoren. Het doel van deze sectie is de examinatoren te leren omgaan met tentamen- en vraaganalyses die dienen als hulpmiddel om inzicht te krijgen in de prestaties van de studenten enerzijds en de kwaliteit van het tentamen en de afzonderlijke vragen anderzijds.

(Opmerking: Deze sectie is geschreven tijdens de overgang naar Computergebaseerd toetsen. In de uitwerking wordt telkens eerst de praktijk van de schriftelijke groepstentamens beschreven en daarna de praktijk van computergebaseerde groepstentamens (CBG))

VII-1. Inleiding

Nadat een schriftelijk of computergebaseerd groepstentamen is afgenomen en alle tentamengegevens centraal in Heerlen zijn verzameld en verwerkt, moet in het tentamentraject de uitslag bepaald worden. Daarna kunnen de uitslagbrieven met het behaalde tentamencijfer en - bij voldoende resultaat - de certificaten naar de individuele studenten gestuurd worden.

Concreet betekent dit dat de examinator in deze fase de definitieve cesuur dient vast te stellen omdat de transformatie van scores naar cijfers pas kan plaatsvinden nadat de examinator de kwaliteit van de toets en toetsvragen beoordeeld heeft.

Voorafgaand aan de tentamenafname, dus bij het samenstellen van het tentamen, is de voorlopige cesuur bepaald (zie sectie VI). Voordat de definitieve cesuur kan worden vastgesteld moet de kwaliteit van de afzonderlijke vragen worden onderzocht: hoe groter de kwaliteit van het tentamen, hoe meer sprake is van een verantwoorde beoordeling van de student op dat tentamen. Bij het beoordelen van de kwaliteit van een tentamen en de samenstellende vragen zijn de moeilijkheidsgraad en het discriminerend vermogen van groot belang. Naast inzicht in de kwaliteit van de vragen willen we na afloop van een tentamen ook informatie hebben over hoe de studenten gepresteerd hebben op het tentamen in zijn geheel en op de afzonderlijke vragen. Als hulpmiddel om inzicht te krijgen in de kwaliteit van en informatie over de prestatie van de studenten op het gehele tentamen en op de afzonderlijke vragen is een analysetechniek ontwikkeld, die aangeduid wordt met de term tentamen- en vraaganalyse.

Een tentamen- en vraaganalyse wordt gevormd door een geheel van statistische indices waarmee we inzicht kunnen krijgen in de prestaties van een groep studenten en waarmee we de kwaliteit van het tentamen en van de vragen kunnen beoordelen op basis van de resultaten van die groep studenten.

De tentamen- en vraaganalyse bestaat uit drie componenten:

- een tentamen- en vraaganalyse voor alleen gesloten vragen
- een tentamen- en vraaganalyse voor alleen open vragen
- een tentamen- en vraaganalyse voor gesloten en open vragen tezamen

Voor het gehele tentamen en de afzonderlijke vragen moet de examinator het volgende nagaan:

- welke kernvragen moeten gesteld worden om inzicht te krijgen in de kwaliteit van het gehele tentamen en de vragen afzonderlijk en in de prestaties van de studenten?
- welke indices in de tentamen- en vraaganalyse geven antwoord op de geformuleerde kernvragen?
- welke waarden moeten deze indices normaliter aannemen en wat moet de examinator doen indien deze hiervan afwijken?

Door een inspectie van de gegevens uit de tentamen- en vraaganalyse kan de examinator nagaan of er mogelijk fouten zijn gemaakt tijdens de constructie van het tentamen. Zo kan het noodzakelijk blijken

alsnog het tentamen op bepaalde punten bij te stellen en/of de cesuur aan te passen. In de laatste paragraaf gaan we in op de procedure voor het vaststellen van de definitieve cesuur.

VII-2. Informatie over het gehele tentamen

Een tentamen moet aan een aantal kwaliteitseisen voldoen, waarvan validiteit en betrouwbaarheid de belangrijkste zijn. Als we na afname van een tentamen willen nagaan of het tentamen aan deze eisen voldoet, moeten we antwoord krijgen op de volgende twee kernvragen:

A1. Meet het tentamen wat ermee bedoeld wordt te meten?

A2. Meet het tentamen consistent, ongeacht het doel?

Om inzicht te krijgen in de prestaties van de studenten op een tentamen moeten we antwoord krijgen op de derde kernvraag:

A3. Hoe hebben de studenten op het tentamen gescoord?

Kernvraag A1: Meet het tentamen wat bedoeld wordt te meten?

Hiermee is de vraag gesteld naar de validiteit van een tentamen, namelijk de representativiteit en evenwichtigheid van het tentamen en van de afzonderlijke vragen ten opzichte van de te toetsen leerdoelen, naar vakinhoud en vraagniveau. Een tentamen is te beschouwen als een steekproef uit alle mogelijke vragen die op een cursus betrekking hebben. Naarmate de vragen een betere operationalisatie zijn van de te toetsen leerdoelen en naarmate de steekproef van vragen representatiever is, zal een tentamen meer (inhouds)valide zijn.

Kernvraag A2: Meet het tentamen consistent, ongeacht het doel?

Hiermee wordt gedoeld op een ander belangrijk kenmerk van een tentamen: de betrouwbaarheid. De vragen/vragen in een tentamen behoren beantwoord te worden op grond van beheersing van de stof. De score moet hoger uitvallen naarmate de beheersing van die stof beter is. In het algemeen zal echter de totstandkoming van een tentamenscore niet foutloos zijn. Er zijn namelijk factoren die kunnen veroorzaken dat de vragen niet uitsluitend een beroep doen op de beheersing van de stof. Dit soort factoren veroorzaakt dat scores niet nauwkeurig weergeven in welke mate de getentamineerde studenten de stof beheersen: een deel van de score behaald door een individuele student moeten we dan toeschrijven aan toevallige fouten in de meetprocedure.

Dit kunnen we als volgt in een gelijkheid uitdrukken:

$$X_i = T_i + E_i, \text{ waarin}$$

X_i = de waargenomen of geobserveerde tentamenscore van student i ; T_i = de ware tentamenscore, dat wil zeggen de score die verkregen zou zijn indien geen meetfouten waren gemaakt; E_i = de toevallige meetfout bij deze meting voor student i .

Het probleem van de betrouwbaarheid nu is om te bepalen hoe ernstig deze meetfouten zijn.

Betrouwbaarheid is een noodzakelijke voorwaarde voor validiteit. Anders gezegd: een onbetrouwbare meting kan niet valide zijn, een betrouwbare meting kan ook een valide meting zijn, maar hoeft dat niet te zijn.

Voor het schatten van de betrouwbaarheid zijn een aantal methoden ontwikkeld, waarvan we er één in het kort zullen bespreken, nl. de interne consistentie-methode. Deze is gebaseerd op de onderlinge samenhang van de vragen waaruit het tentamen bestaat. Dit komt er op neer dat de afzonderlijke vragen opgevat worden als aparte toetsjes, die allemaal dezelfde eigenschap meten. De gemiddelde correlatie tussen de vragen geeft dan een indicatie van de betrouwbaarheid. Omdat het in feite over één

tentamen gaat spreekt men over de interne consistentie van een tentamen. Een op deze (gemiddelde) correlatie tussen de vragen gebaseerde betrouwbaarheidscoëfficiënt is de zogenaamde alpha-coëfficiënt dan wel de KR-20. Deze coëfficiënt is standaard in de tentamen- en vraaganalyses opgenomen.

Hierbij gaat het om de betrouwbaarheid van de toets. De betrouwbaarheid is een noodzakelijke voorwaarde voor validiteit. Anders gezegd: een onbetrouwbare meting kan niet valide zijn, een betrouwbare wel valide. Dit betekent echter niet dat een betrouwbare meting ook altijd valide is.

De bedoeling van een toets is dat de vragen beantwoord worden op grond van beheersing van de stof. De score moet hoger zijn naarmate de beheersing van die stof beter is. In het algemeen zal echter het vaststellen van een toetscore niet foutloos geschieden. Er zijn namelijk verschillende factoren die veroorzaken dat de toets en de vragen niet uitsluitend een beroep doen op de beheersing van de stof (bijvoorbeeld door de ingewikkeld geformuleerde vragen). Dit soort factoren maakt dat scores niet nauwkeurig weergeven in welke mate de getentamineerde studenten de stof beheersen: een gedeelte van de score behaald door een individuele student op een bepaald toets moet toegeschreven worden aan fouten in de meetprocedure. Een maat voor de betrouwbaarheid is de Cronbach's Alpha (zie figuur VII-1). Dit is een maat voor de statistische betrouwbaarheid van de toets, de consistentie van de resultaten van de studenten over de vragen van de toets heen. De waarde geeft weer in hoeverre de vragen in het assessment samenhangen en hetzelfde construct meten. Cronbach's alpha ligt tussen de 0 en 1 en hoe dichter bij 1 hoe hoger de betrouwbaarheid.

Table of Test Statistics

Number of examinees	626	Mean	73.42/140 (52.44%)	Standard error of mean	0.67/140 (0.48%)
Number of items	140	Median	73/140 (52.14%)	Standard error of measurement	5.11/140 (3.65%)
Maximum possible score	140	Mode	72/140 (51.43%)	Skew	0.062
Minimum achieved score	3/140 (2.14%)	Standard deviation	16.69/140 (11.92%)	Kurtosis	0.982
Maximum achieved score	133/140 (95%)	Variance	278.69/140 (199.06%)	Test reliability (Cronbach's Alpha)	0.906

Reliability is most meaningful if all items cover the same subject area.

Figuur VII-1 Informatie over de toets. Rechtsonder staat de waarde voor de betrouwbaarheid (Cronbach's alpha)

Voor toetsen met een hoog belang (high-stakes tests) is een betrouwbaarheid van boven de 0,9 gewenst. Bij studietoetsen wordt een ondergrens van 0,7 gehonoreerd.

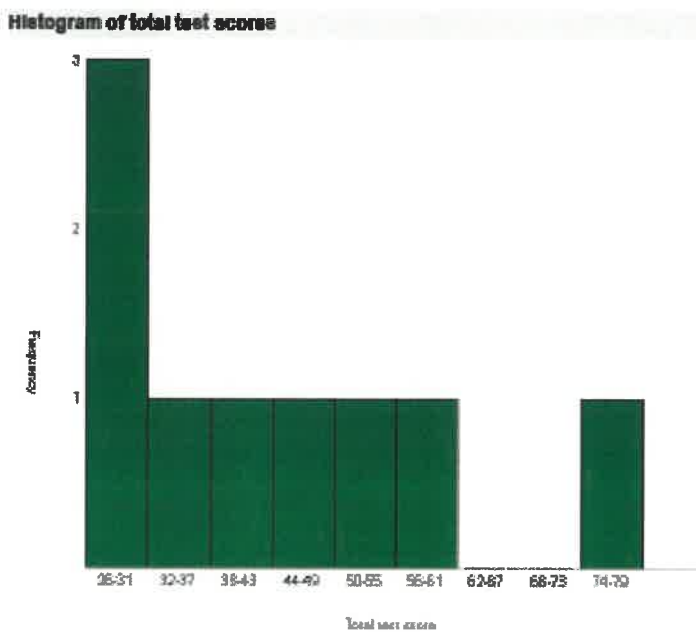
Wanneer uit de analyse blijkt dat er sprake is van een lage betrouwbaarheidscoëfficiënt kan dat verschillende redenen hebben:

- onvoldoende toetslengte: uitbreiding van het aantal vragen levert meestal een verhoging van de betrouwbaarheid op (meer meetpunten betekent minder toeval);
- te korte afnametijd: de betrouwbaarheidsschatting is weinig zinvol indien studenten te weinig tijd krijgen voor het maken van de toets;
- kenmerken van de afzonderlijke vragen: constructiefouten, moeilijkheidsgraad en aantal alternatieven (bij gesloten vragen) hebben invloed op de betrouwbaarheid.

Kernvraag A3: Hoe hebben de studenten op het tentamen gescoord?

Om te kunnen beoordelen hoe een groep studenten op het tentamen heeft gepresteerd, moeten we informatie verzamelen over:

- Om verder te kunnen beoordelen hoe een groep studenten op het tentamen heeft gepresteerd, kijken we naar:
- het maximaal aantal te behalen punten op het tentamen ofwel de maximaal haalbare tentamenscore (in Figuur VII-3 is dat 128 punten);
- de scores die feitelijk door de studenten behaald zijn op het tentamen (in Figuur VII-3 af te lezen in de kolom 'raw scores');
- het aantal of percentage studenten dat een bepaalde score of een lagere score heeft behaald (in Figuur VII-3 af te lezen in de kolom 'percentage' and 'cumulative percentage');
- de gemiddelde score en de mate van spreiding van de scores (in Figuur VII-3 af te lezen in de bovenste tabel);
- het aantal of percentage studenten dat een voldoende resultaat behaald heeft (te berekenen door 100o minus het cumulatieve percentage van de hoogste zakwaarde).



Figuur VII-2. Histogram van de scoreverdeling

Test Analysis Report

Assessment name	BKD CBT MW1 assessment	Assessment description	
Assessment author	DU/Schoor	Assessment ID	5741884342638234
Assessment test modified	Jul 24 2009 00:00:00	Report date & time	Dec 08 2009 16:08:30

Filters
 All dates
 Participants who finished

MaxPossibleScoreText

Table of Test Statistics

Number of assessments	9	Mean	43.56/128 (34.03%)	Standard error of mean	5.32/128 (4.16%)
Number of items	44	Median	41/128 (32.03%)	Standard error of measurement	8.06/128 (6.3%)
Maximum possible score	128	*Mode	26/128 (20.31%)	Skew	0.958
Minimum achieved score	26/128 (20.31%)	Standard deviation	15.95/128 (12.46%)	Kurtosis	0.363
Maximum achieved score	75/128 (58.59%)	Variance	254.26/128 (198.66%)	Test reliability (Cronbach's Alpha)	0.745

Reliability is most meaningful if all items cover the same subject area.
 *There are multiple modes

Reliability (Topic Level)

Topic	Number of Items	Mean	Standard deviation	Reliability
BKD cursus CBT MW	34	38.67/111 (34.84%)	16.18/111 (14.58%)	0.772
BKD cursus CBT MWLVV Huiswerkopdracht 1	5	2.11/11 (19.18%)	2.42/11 (22%)	0.681
BKD cursus CBT MWVWZ Huiswerkopdracht	5	2.78/6 (46.33%)	0.83/6 (13.83%)	-0.36

Frequency distribution of total test scores

Raw Score	Number	Percentage	Cumulative percentage
26	1	11.11	11.11
30	1	11.11	22.22
31	1	11.11	33.33
34	1	11.11	44.44
41	1	11.11	55.56
44	1	11.11	66.67
53	1	11.11	77.78
58	1	11.11	88.89
75	1	11.11	100

Figuur VII-3. Toetsanalyse

VII-3. Informatie over de afzonderlijke vragen

Na inspectie van het tentamen als geheel bekijken we de analysegegevens per afzonderlijk vraag. Informatie over de kwaliteit van een vraag en de resultaten van de groep studenten op een vraag verkrijgen we door beantwoording van de volgende drie kernvragen:

Kernvraag 8I: Wat is de moeilijkheidsgraad van de vraag?

Om na te gaan wat de moeilijkheidsgraad van een vraag is moeten we informatie verkrijgen over het relatieve aantal (of het percentage) studenten dat de vraag juist beantwoord heeft, het percentage

studenten dat de vraag werkelijk beheerst en welke scores op de vraag (gemiddeld) behaald zijn. Op grond van deze informatie kunnen we dan beoordelen of een vraag te moeilijk of te makkelijk is geweest voor de onderhavige groep studenten.

Kernvraag 82: Hoe functioneren de afleiders?

Bij gesloten vragen kunnen we per vraag nagaan hoe de afzonderlijke afleiders gefunctioneerd hebben, bijvoorbeeld welke keuze ten grondslag lag aan een foutieve beantwoording: het percentage studenten dat een onjuist alternatief koos, geeft aan of (en welke) afleiders naar behoren gefunctioneerd hebben.

Kernvraag 83: Heeft de vraag discriminerend vermogen?

Een tentamen moet onderscheid maken ('discrimineren') tussen studenten die de leerstof (goed) beheersen en studenten die de leerstof niet of slecht beheersen. We gaan er hierbij van uit dat de goede beheersers degenen zijn die een hoge totaalscore behalen en de slechte beheersers degenen die een lage totaalscore behalen. Elk afzonderlijk vraag levert een bijdrage in de totaalscores van de studenten. De bijdrage van elk afzonderlijk vraag moet uit het oogpunt van dit onderscheid positief zijn: de vraag moet dus een onderscheid maken tussen 'goede' en 'slechte' studenten. Om na te gaan in hoeverre elk afzonderlijk vraag hierin geslaagd is, kan per vraag een discriminatie-index berekend worden. De meest geschikte maat daarvoor is de correlatiecoëfficiënt.

VII-4. De vraaganalyse

In de tentamen- en vraaganalyse worden een aantal indices berekend die betrekking hebben op de verschillende kernvragen. Per kernvraag zullen we aangeven welke indices hierbij een rol spelen, hoe deze geïnterpreteerd moeten worden en zo mogelijk wat we kunnen doen indien de waarden van deze indices afwijken van bepaalde normen.

Kernvraag 81: Wat is de moeilijkheidsgraad van de vraag?

In de tentamen- en vraaganalyse van alleen gesloten vragen worden de volgende indices berekend die informatie opleveren over de prestaties van studenten op en de moeilijkheidsgraad van de afzonderlijke vragen:

p-waarde (= proportie goed)

De p-waarde is een getal tussen 0.0 en 1.0, dat de moeilijkheidsgraad van een vraag weergeeft en iets zegt over de prestaties van de studenten op dat vraag. Als een p-waarde relatief hoog is, is er sprake van een relatief gemakkelijk vraag en/of de studenten hebben relatief weinig moeite gehad met het beantwoorden van dat vraag.

Voorbeeld

Als een p-waarde 0.78 bedraagt betekent dit dat 78% van de studenten de vraag juist heeft beantwoord.

In de tentamen- en vraaganalyse voor alleen open vragen en gesloten en open vragen tezamen worden de volgende indices berekend:

- Maximale score (max.): Het maximale aantal punten dat voor elk vraag of vraagonderdeel behaald kan worden.
- Gemiddelde score (gem.): Dit zijn alle voorkomende scores per vraag of vraagonderdeel bij elkaar opgeteld en gedeeld door het aantal voorgekomen scores. Hier is sprake van twee soorten informatie in één getal. De gemiddelde score zegt zowel iets over de moeilijkheidsgraad van de vraag (hoe hoger de gemiddelde score, hoe gemakkelijker de vraag) als over de prestaties van de studenten op dit vraag.

Voorbeeld

Als in de analyse staat

Vraagnr.	Max. score	Gem.score
1	6	4,73
2	6	2.08

dan kunnen we hieruit afleiden dat vraag 1 gemiddeld beter is gescoord (gemiddeld 4.73 bij maximaal 6 te behalen punten) dan vraag 2.

Een beter hanteerbare maat voor de moeilijkheidsgraad van een vraag vormt de p'-waarde (= gemiddelde score gedeeld door de maximaal haalbare score); deze p'-waarden zijn genormeerd en kunnen zonder meer met elkaar vergeleken worden.

- Scoreverdeling per vraag: Per vraag wordt steeds in percentages aangegeven hoe frequent elke mogelijke score is toegekend. Alleen bij open vragen met een polytome scoring levert deze verdeling extra informatie op in vergelijking met p-/p'-waarden: een 'vreemde' scoreverdeling kan mogelijk een indicatie zijn voor onduidelijkheden in het correctievoorschrift (antwoordmodel, scoringsvoorschrift, beoordelaarsinstructie) .

Voorbeeld

Als in de analyse staat

Vraagnr.	code	0	1	2	3	4	5
1	ov 1	0	0	0	0	2	8
2	ov 2	0	0	0	0	4	10
3	ov 3	0	0	0	1	2	29

dan houdt dat in dat bij vraag 1 in 0o van de gevallen een score van 0, 1, 2 of 3 is gegeven, in 2o van de gevallen een score van 4, in 8o een score van 5, enzovoort.

- p'-waarde (p'): De p'-waarde is een getal tussen 0.0 en 1.0, dat de moeilijkheidsgraad van een vraag weergeeft. Bij open vragen wordt de p'-waarde berekend door de gemiddelde score per vraag of vraagonderdeel te delen door de maximaal haalbare score op dit vraag of dat vraagonderdeel. Bij gesloten vragen is deze gelijk aan de p-waarde. Net als bij de gemiddelde score is ook hier sprake van twee soorten informatie in één getal.

Voorbeeld

Als een p-waarde voor een gesloten vraag 0.78 bedraagt betekent dit dat 78o van de studenten de vraag juist beantwoord heeft. Als een p'-waarde voor een open vraag 0.78 bedraagt betekent dit dat de studenten gemiddeld 78o van het aantal te behalen punten hebben behaald.

De normen voor deze verschillende indices zijn afhankelijk van de de functie van het tentamen. Het tentamen moet bepalen of een individuele student de stof beheerst. Daarmee maakt de toets onderscheid tussen studenten die de leerstof beheersen en niet beheersen. Elk afzonderlijk vraag levert

een bijdrage in de tentamenscores van de studenten. Die bijdrage kan positief zijn uit het gezichtspunt van differentiatie van 'goede' en 'slechte' studenten: 'goede' studenten beantwoorden de vraag in het algemeen goed en 'slechte' studenten beantwoorden de vraag fout. Maar die bijdrage kan ook nihil of zelfs negatief zijn: zowel 'goede' als 'slechte' studenten beantwoorden de vraag juist resp. 'goede' studenten maken de vraag fout en 'slechte' studenten beantwoorden de vraag goed. Een afzonderlijk vraag heeft een maximale selectieve bijdrage in de tentamenscore bij een p-waarde van 0.50: dan immers splitst de vraag de groep studenten op in twee gelijke delen: de beheersers en de niet-beheersers.

Bij gesloten vragen gaat dit niet op omdat daarbij een raadeffect kan optreden. In dit geval moet men dan ook niet trachten de p-waarde gelijk aan 0.50 te krijgen. Met een bepaalde formule kan de optimale p-waarde en de ondergrens en bovengrens worden berekend:

	Optimale p-waarde	Ondergrens	Bovengrens
Bij tweekeuze vragen	0,75	0,61	0,90
Bij driekeuze vragen	0,67	0,50	0,90
Bij vierkeuze vragen	0,62	0,44	0,90
Bij open vragen p'	0,50	0,25	0,90

Te hoge en te lage p-/p'-waarden moeten bij selectieve toetsen worden vermeden; immers dergelijke vragen leveren geen of weinig bijdrage aan de rangorde van studenten op de scoreschaal. In elk geval moet bij vragen met een lage p-/p'-waarde (zeker wanneer $p-/p' < 0.30$) een inspectie plaatsvinden van indices die bij de volgende twee kernvragen behandeld worden.

Figuur VII-4 Voorbeeld van een vraaganalyse uit het Question statistics report

Deze gegevens zijn af te leiden uit de Question statistics analysis en vraag analyses report (zie figuur VII-4 als voorbeeld van een Question statistic analysis).

Een combinatie van deze gegevens geven informatie over de vraag. Op basis hiervan moet gekeken worden of de vraag aangepast zou moeten worden of dat de vraag inhoudelijk volledig helder en correct is.

Vraagnr	05-01		
vraagtekst			
vraagtype	Multiple Choice		
aantal keer gepresenteerd	13		
aantal afnames	13		
max. score	1		
p-waarde	0.615		
correlatie	0.452		
analyse per alternatief	alternatief	# keer gekozen	percentage
	* 0 Medewerkers die werken op basis van uur	8	62%
	1 In beide gevallen gingen medewerkers mee	1	8%
	2 In beide gevallen gingen medewerkers min	1	8%
	3 Medewerkers die werken op basis van uur	3	23%

Kernvraag 82: Hoe functioneren de afleiders?

Analoog aan de berekening van de p-waarde kan men bij gesloten vragen de proportie studenten bepalen die een fout alternatief (= een afleider) hebben aangestreept als antwoord: de a-waarde. Aan de a-waarden kan men zien of en hoe de afleiders 'werken'. De a-waarden van een vraag moeten normaliter kleiner zijn dan de p-waarde van dat vraag.

Wanneer a-waarden hoger blijken te zijn dan de bijbehorende p-waarde kan dit duiden op fouten in de vraag (bijvoorbeeld er zijn meerdere juiste alternatieven: de vraag is niet 'objectief') of een fout in het antwoordmodel. Om een betere indicatie te krijgen of er iets aan de hand is dient de waarde van de discriminatie-index (kernvraag B3) te worden beschouwd en moet het desbetreffende vraag inhoudelijk nog eens grondig gecontroleerd worden.

De optimale a-waarde van een vierkeuzevraag is ongeveer 0.12. Bij een lage a-waarde, bijvoorbeeld kleiner dan 0.05, kan men stellen dat de afleider niet als zodanig heeft gefunctioneerd. In geval van hergebruik van dat vraag moet dan getracht worden betere afleiders te formuleren.

Kernvraag 83: Heeft de vraag discriminerend vermogen?

Met behulp van de discriminatie-index kan men nagaan in hoeverre een vraag differentieert tussen 'goede' en 'slechte' studenten c.q. tussen studenten met een hoge en lage tentamenscore. De meest geschikte maat hiervoor is de correlatiecoëfficiënt.

In de tentamen- en vraaganalyse kunnen verschillende discriminatie-indices berekend worden.

- Bij zowel de analyses van de schriftelijke toetsen als de computergebaseerde toetsen wordt de vraag-totaalcorrelatie (R_{it}) gegeven. Dit is de correlatie tussen vragencore en totaalscore van de toets.
- Bij de analyses van de schriftelijke toetsen wordt tevens de vraag-restcorrelatie (R_{ir}) gegeven. Dit is hetzelfde als R_{it} , alleen is van de totaalscore de score op het betreffende vraag afgetrokken (aangezien de score op de vraag uiteraard perfect met zichzelf correleert, levert de R_{it} een overschatting op).

- Bij de analyses van de computergebaseerde toetsen wordt daarnaast een discrimination index gegeven. Hierbij wordt het verschil weergegeven tussen de 250 best scorende studenten en de 250 zwakst scorende studenten op de hele toets.

De waarden van deze correlatiecoëfficiënten kunnen variëren tussen -1.0 en +1.0. Normaliter moet elk vraag een goed onderscheid maken tussen 'goede' en 'slechte' studenten: dit betekent dat studenten met een hoge tentamenscore de vraag goed moeten hebben beantwoord en studenten met een lage tentamenscore de vraag fout beantwoord moeten hebben. In dit geval hebben de Rit en Rir een waarde die positief is en boven de 0.10 resp. 0.20 ligt.

In het algemeen geldt verder dat hoe hoger de correlatie van een vraag is, hoe meer de vraag samenhangt met het gehele tentamen en hoe meer het bijdraagt aan de betrouwbaarheid van het tentamen. De bijdrage van een vraag aan een goed onderscheid tussen studenten en aan de betrouwbaarheid kan ook gering zijn: zowel 'goede' als 'slechte' studenten beantwoorden de vraag correct: de correlaties zijn laag positief (tussen 0.0 en 0.10 resp. 0.20) of zelfs negatief: 'goede' studenten beantwoorden de vraag fout en 'slechte' studenten beantwoorden de vraag goed: de correlaties zijn negatief.

Globaal kan men de volgende normen aangeven voor de correlaties en de "discrimination index":

	Rit	Rir	Discrimination
Slecht	< 0 (negatief)	< 0,10	< 0 (negatief)
Matig	0 < - < 0,10	0,10 < - < 0,20	0 < - < 0,10
Redelijk	0,10 < - < 0,20	0,20 < - < 0,30	0,10 < - < 0,20
Goed	> 0,20	> 0,30	> 0,20

Lage of negatieve waarden geven dus aan dat de vraag niet (of weinig) bijdraagt tot een betrouwbare meting en onvoldoende onderscheid maakt tussen 'goede' en 'slechte' studenten. Dit kan een indicatie zijn voor ambiguïteit of constructiefouten van de vraag.

Toch kan men niet stellen dat vragen met een lage correlatie zonder meer uit het tentamen moeten worden geschrapt. Indien alle studenten relatief hoge scores op het tentamen hebben behaald (duidend op een hoog vaardigheidsniveau), zal in het algemeen het discriminerend vermogen van de vragen afnemen. Een slecht discriminerend vraag hoeft dus niet per se een 'slecht' vraag te zijn, maar het kan een aanwijzing zijn dat er inhoudelijk iets niet deugt.

Een andere reden om niet alléén op grond van bepaalde waarden van indices (p-/p' -waarde, Rit/Rir) te beslissen omtrent het handhaven of verwijderen van vragen, is gelegen in het feit dat de betrouwbaarheid van deze indices afhankelijk is van het aantal tentaminandi. Hoe kleiner het aantal personen dat het tentamen gemaakt heeft, hoe onbetrouwbaarder deze indices worden. P-waarden zijn redelijk betrouwbaar bij een groep van 25 studenten, maar correlaties pas bij groepen boven de 100 studenten.

Alleen indien blijkt dat een vraag inhoudelijk niet te verdedigen is (bijvoorbeeld niet objectief, valt buiten de leerstof), is er reden de vraag te verwijderen en niet te betrekken in de bepaling van de uitslag.

Soms behoort een aanpassing van het antwoordmodel tot de mogelijkheden, zodat de vraag blijft meetellen in het tentamen.

In dit kader is het dringend aan te raden eventuele reacties (brieven of forumberichten) met inhoudelijk commentaar van studenten en studiebegeleiders, die binnenkomen na afloop van het tentamen, naast

de gegevens uit de tentamen- en vraaganalyse te leggen, daar deze een mogelijke verklaring leveren waardoor de kwaliteit van sommige vragen, psychometrisch gezien, te wensen overlaat.

Bij de computergebaseerde toetsen kunnen examinatoren in het systeem kijken of de studenten opmerkingen hebben ten aanzien van de inhoud van de toets.

VII-5. Het vaststellen van de definitieve cesuur

Indien na bestudering van de tentamen- en vraaganalyse en eventuele reacties van studenten geen aanleiding bestaat om wijzigingen aan te brengen in de cesuur, is de definitieve cesuur gelijk aan de voorlopige.

De examinerator kan besluiten om bepaalde vragen uit het tentamen te verwijderen, of een andere sleutel te formuleren. De examinerator dient op het mutatieformulier (dat wordt verstrekt bij de initiële analyse) aan te geven welke vragen verwijderd dienen te worden of bij welke vragen de sleutel gewijzigd moet worden. Vervolgens wordt een nieuwe tentamen- en vraaganalyse aangemaakt. Het effect van de verandering kan dan vastgelegd worden in het definitieve vaststellingsverslag van het tentamen.

Bij computergebaseerd toetsen worden vragen niet verwijderd, maar worden deze vragen voor alle studenten goedgekeurd. Hierdoor hoeft de cesuur niet aangepast te worden.

Bij de schriftelijke tentamens moet de cesuur worden aangepast als de examinerator besluit om vragen niet te betrekken in de uitslagbepaling. De voorlopige cesuur moet dan verlaagd worden met het maximaal aantal te behalen punten voor de verwijderde vragen tezamen.

Voorbeeld cesuurverlaging bij schriftelijke toetsen

De voorlopige cesuur van een tentamen met 40 vierkeuzevragen ligt bij 26/27. De examinerator besluit drie vragen (die elk 1 punt kunnen opleveren), te verwijderen op grond van gebleken onvolkomenheden. De definitieve cesuur komt dan te liggen bij 23/24 punten.

Indien alleen een wijziging in het antwoordmodel moet worden aangebracht hoeft de cesuur niet aangepast te worden. Wel zal in zo'n geval een nieuwe analyse uitgevoerd moeten worden.

Als het volgens de examinerator wenselijk is om de cesuur om andere redenen (een te hoog percentage gezakten, te lage betrouwbaarheid) aan te passen of extra te verlagen dan zal hij/zij daartoe (bij voorkeur in overleg met een ECOP-toetsdeskundige) een met redenen omkleed voorstel moeten indienen bij het Dagelijks Bestuur van de CvE van de opleiding.

Nadat de definitieve cesuur is bepaald worden de scores automatisch, via lineaire transformatie, omgezet naar een cijferschaal, variërend van 1 tot en met 10.

VIII. CBI-tentaminering

Deze sectie is primair bedoeld voor de examinatoren, die te maken hebben of krijgen met CBI-tentaminering.

Het doel van deze sectie is aan te geven wat CBI-tentaminering betekent voor de examinatoren. Het CBI-tentamentraject wordt besproken en de verschillen met de tentaminering in het 'reguliere' tentamentraject worden aangegeven.

In deze sectie wordt het ontstaan van de CBI-tentaminering beschreven. Tevens wordt een overzicht van de fasen en activiteiten van het CBI-tentamentraject gegeven, met daarbij vermeld de betrokken afdelingen en functionarissen.

VIII-1. Ontstaan en systematiek van Computer gebaseerd Individueel tentamineren (CBI)

Eén van de hoofdkenmerken van de OU is dat studenten in eigen tempo een cursus kunnen bestuderen. Een logisch vervolg van de daaruit voortvloeiende tempodifferentiatie is dat de student op een door hem of haar zelf te bepalen moment tentamen moet kunnen afleggen.

De OU is daarom eind 2008 gestart met de ontwikkeling en implementatie van een geïndividualiseerd en geautomatiseerd tentamenvorm onder de naam CBI. Bij CBI kunnen studenten zich digitaal aanmelden binnen vastgestelde tentamenweken voor een individueel af te leggen tentamen. Zij bepalen binnen deze grenzen datum, tijdstip en studiecentrum van afname dat hun goed uitkomt.

In geval van CBI-tentaminering kan een student in principe, afhankelijk van de openingstijden van de CBI-tentamenlocaties op de studiecentra, tentamen afleggen wanneer hij/zij dat wil. Aanmelding gebeurt van te voren via mijnOU. Het aanmeldsysteem laat zien welke mogelijkheden er zijn. De student ontvangt een oproep met de datum en tijd waarop hij/zij het gewenste tentamen kan afleggen en ook de locatie en de toegestane hulpmiddelen op het tentamen.

Op basis van de aanmelding wordt er een sessie ingeroosterd en wordt een tentamen gegenereerd uit een itembank als de student op het studiecentrum inlogt voor zijn tentamenafname.

Een itembank (of vraagbestand) is een verzameling vragen met bijbehorend antwoordmodel en clusterindeling. Het genereren van een tentamen houdt in dat de computer op basis van de specificaties in de toetsmatrijs een tentamen uit de itembank samenstelt. Op deze wijze zijn alle te genereren tentamens per cursus qua samenstelling zo veel mogelijk gelijkwaardig.

Voor het afnemen van de computergebaseerde toetsen wordt gebruik gemaakt van Questionmark Perception (QMP).

Zodra de student het tentamen heeft afgelegd wordt direct de (voorlopige) uitslag bepaald. Deze uitslag verschijnt op het scherm. Daarna krijgt de student nog tijd om op het beeldscherm de foutief beantwoorde vragen samen met feedback op de vraag te bekijken. De definitieve uitslag krijgt de kandidaat uiterlijk zes weken later vanuit Heerlen toegezonden en bij positief resultaat uiteraard tevens het certificaat.

VIII-2. De fasen in het CBI-tentamentraject

In het navolgende wordt het CBI-tentamentraject in drie fasen weergegeven: de itembankconstructie, de formele vaststelling van de inhoud van het vragenbestand en het onderhoud van een vraagbestand. Bij elke fase is tevens aangegeven welke afdelingen resp. functionarissen daarbij betrokken zijn.

Fase 1: Itembankconstructie

Tot de fase van de itembankconstructie behoren alle activiteiten vanaf het bepalen van de vorm en de inhoud van het CBI-tentamen tot en met het samenstellen van de voorbeeldtentamens.

a. Bepalen van vorm en inhoud van het tentamen

Net als bij het construeren van reguliere schriftelijke tentamens moet vooraf bepaald worden welke leerstofonderdelen aan de orde moeten komen en welk type oplosgedrag van de studenten verwacht wordt. Dit betekent dat een toetsmatrijs en een tentamenprofiel ontwikkeld moeten worden. In sectie III wordt besproken hoe deze opgesteld moeten worden. Een verschil tussen de toetsmatrijs, zoals die in sectie III besproken is en de CBI-toetsmatrijs is dat bij de laatste de verdeling van vragen over clusters meer expliciet gemaakt dient te worden (zie voorbeeld hieronder). Deze clustering kan gebeuren op grond van bijvoorbeeld leerinhoudelijke-, beheersingsniveau- of gedragscriteria en is nodig om bij een selectie van vragen een evenwichtige spreiding over de stof te verkrijgen. Het aantal clusters dat onderscheiden kan worden is afhankelijk van de toetsmatrijs voor de desbetreffende tentamens.

Een belangrijke voorwaarde is dat het maximaal aantal clusters nooit groter mag zijn dan het aantal vragen in één tentamen.

In voorbeeld 1 (zie hieronder) betekent dit dat het maximum aantal clusters gelijk is aan 60. Het is dus niet mogelijk uit bepaalde clusters geen vragen in het tentamen op te nemen. Van belang is voorts dat het tentamenprofiel en de clusterindeling reeds opgesteld is voordat het bestand opgemaakt en ingevoerd gaat worden.

Een ander punt van aandacht is dat een vraag uit een cluster geen informatie mag bevatten, waardoor een vraag uit een ander cluster opgelost kan worden. Hierop moet dus op voorhand gecontroleerd worden.

De minimale omvang van een itembank.

De minimale omvang van een itembank is afhankelijk van de het aantal vragen in de toets. Bij 80 en 60 vragen wordt de factor dan 6x, bij 40 vragen 7x en bij 30 vragen 8x het aantal vragen. Voor het berekenen van de optimale omvang van itembanken is een instrument ontwikkeld die inzicht geeft in de relatie tussen de verhouding van de grootte van de itembank t.o.v. de grootte van het tentamen enerzijds (de zgn. multiplicatiefactor) en de te verwachten overlap tussen twee willekeurige trekkingen uit een itembank anderzijds. Gekeken is naar de consequenties van een lagere multiplicatiefactor voor de kans dat twee tentamens, gegenereerd uit een vraagbestand, één of meer vragen gemeenschappelijk hebben. De vraag die de CvE's zich daarbij gesteld hebben is wat de maximale overlap mag zijn tussen tentamens uit een bank en welke maximale overschrijding bij de gekozen maximale overlap acceptabel is. Een gegeven hierbij is dat pas vanaf een maximale overlap van 25% winst geboekt kan worden in vergelijking met de huidige norm. De CvE's vinden een maximale overlap van 25% acceptabel.

Voorbeeld van clusterindeling

CBI-cluster	aantal opgaven	aantal selectie	herkomst leereenheden
01	15	1	1.1
02	15	1	1.2 - 1.4
03	15	1	1.5 - 1.6
04	30	2	2.1 - 2.2
05	30	2	2.3, 2.5, 2.6
06	30	2	2.4
07	75	5	2.7 - 2.9
08	75	5	3.1 - 3.4
09	75	5	3.5, 3.9
10	75	5	3.6 - 3.8
11	75	5	4.1
12	75	5	4.3
13	75	5	4.3
14	75	5	4.4
15	30	2	4.5 - 4.6
16	30	2	5.1 - 5.3
17	30	2	6.1
18	30	2	6.2 - 6.5
19	30	2	6.6
20	15	1	6.7
totaal	900	60	

Het effect van de huidige regel op de minimale omvang van itembanken - een variabele multiplicatiefactor met een maximale overlap van 25% tussen tentamen en hertentamen - is in Tabel VIII-1. in kaart gebracht. Hierbij is uitgegaan van vier gangbare tentamenlengten, namelijk 30, 40, 60 en 80 vragen met als parameters het aantal te genereren vragen per cluster (= 1), het aantal clusters (= resp. 30, 40, 60 en 80) en een maximaal toegestane overlap tussen tentamens (= 25%) en de daaruit resulterende multiplicatiefactor en minimale omvang van het vragenbestand.

Tabel VIII-1.

aantal gegenereerde vragen per cluster	aantal clusters	maximale overlap	multiplicatiefactor	minimale omvang vraagbestand
1	30	25%	8	240
1	40	25%	7	280
1	60	25%	6	360
1	80	25%	6	480

Het effect van de 'tentamengeschiedenis' op de minimale omvang van vraagbestanden

In Tabel VIII-2. is doorgerekend wat het effect zal zijn op de minimale omvang van vraagbestanden indien het te kiezen pakket de mogelijkheid biedt de tentamengeschiedenis van (gezakte) studenten bij te houden. Hierbij gelden dezelfde parameters als bovengenoemd, maar bedraagt de maximale overlap 0% en wordt verondersteld dat gezakte studenten *maximaal vier keer* herkansen. De multiplicatiefactor wordt daarmee 1 tentamen + 4 herkansingen = 5. Verder gelden als randvoorwaarden dat vragen niet openbaar worden gemaakt en dat het gaat om de tot nu toe gangbare gesloten vragen.

Tabel VIII-2.

aantal gegenereerde vragen per cluster	aantal clusters	maximale overlap	multiplicatiefactor	minimale omvang vraagbestand
1	30	0%	5	150
1	40	0%	5	200
1	60	0%	5	300
1	80	0%	5	400

Het bijhouden van de tentamengeschiedenis van gezakte studenten leidt in alle gevallen tot een kleinere minimale omvang van *nieuw* te ontwikkelen vraagbestanden dan tot nu toe.

Gaat het om *bestaande*, te converteren, bestanden, blijft de winst dat geen herkansende student een bij hem eerder afgenomen vraag opnieuw onder ogen zal krijgen.

b. (Laten) construeren van tentamenvragen

Nadat de toetsmatrijs en het tentamenprofiel zijn opgesteld kunnen de vragen geconstrueerd worden. Voorafgaande aan de constructie moet de examinator bepalen hoeveel vragen er geconstrueerd moeten worden. Bij het bepalen van het aantal te construeren vragen moet men

- er rekening mee houden dat een groot aantal vragen geconstrueerd moet worden, doordat er telkens een selectie van de vragen gemaakt wordt voor een tentamen.
- rekening houden met de toetsmatrijs, zodat gerichte schrijfp opdrachten aan de vraagconstructeurs gegeven kunnen worden
- rekening houden met de vorm van het tentamen: bij tweekeuze vragen moeten er in het ideale geval 80 vragen per tentamen te selecteren zijn en bij vierkeuze vragen ongeveer veertig.

De tentamenvragen kunnen rechtstreeks door examinatoren in QMP ingevoerd worden of hiervoor kan ondersteuning aangevraagd worden bij de faculteit.

c. Screenen van de tentamenvragen

De vragen die de examiner aangeleverd krijgt van de vraagconstructeurs, moeten gescreend worden. Dit wil zeggen dat de vragen beoordeeld moeten worden aan de hand van taalkundige, toetstechnische, inhoudelijke en lay-out technische criteria. De screening gebeurt door de examiner (en, zo mogelijk, door andere vakinhoudelijke deskundigen en/of toetsdeskundigen).

d. Samenstellen clusters en antwoordmodel en opmaak van het bestand

Nadat de vragen goedgekeurd zijn stelt de examiner een overzicht op van de verdeling van de vragen over de clusters. Daarnaast worden de vragen en de juiste antwoorden gecombineerd tot het antwoordmodel. Als de examiner niet zelf de items invoert in QMP, dan levert hij de CBI-bank aan bij de medewerker(s) die verantwoordelijk is (zijn) voor de invoer.

e. Definitief maken van het tentamenprofiel en bepalen van de cesuur

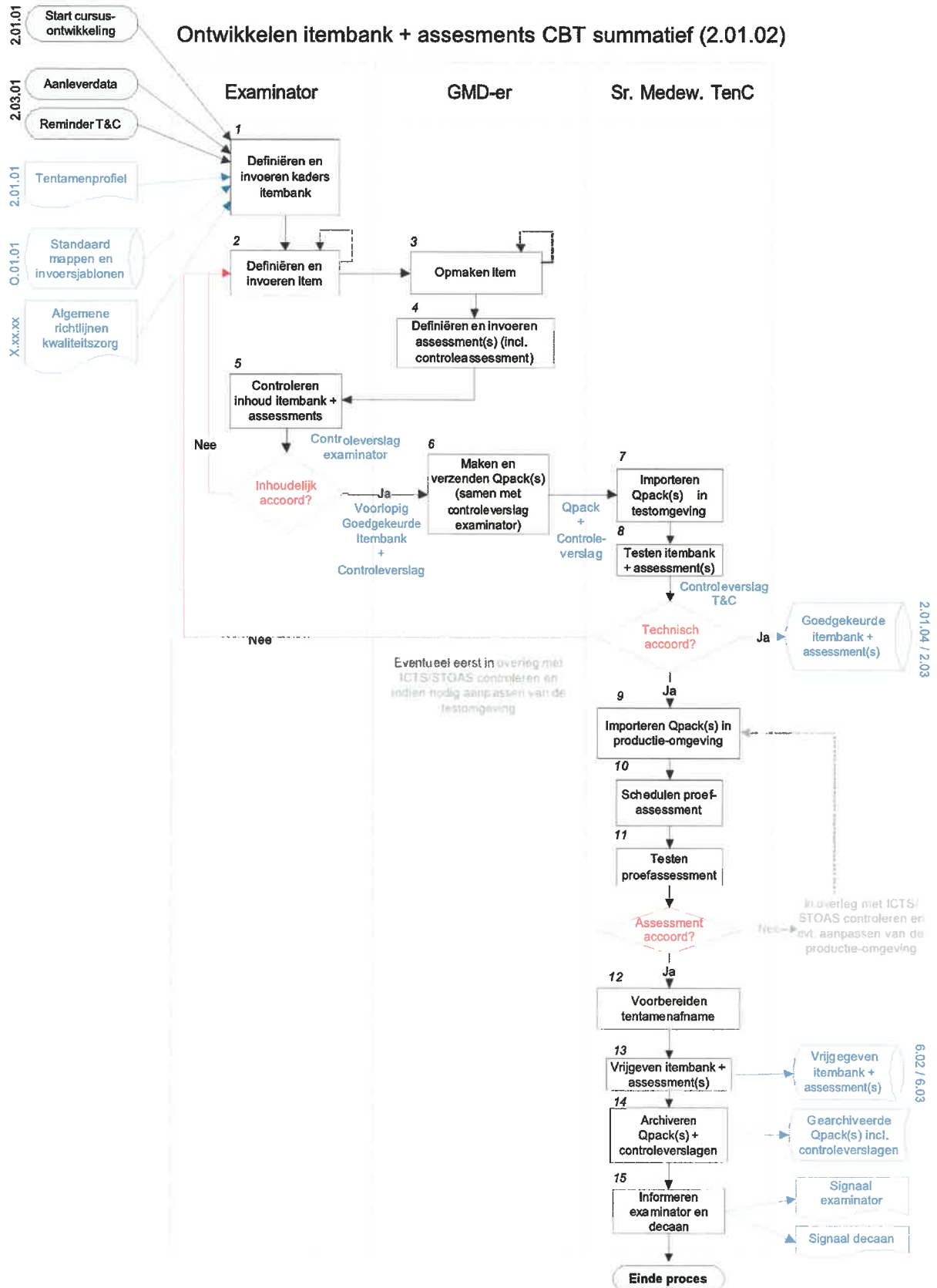
De ingevoerde items moeten gecontroleerd worden door de examiner. Op basis van de toetsmatrijs wordt dan het 'assessment' in QMP vast gesteld. Hierin staat hoe de vragen geselecteerd worden, wat de cesuur en puntenverdeling is (zie sectie VI, Vaststellen van de voorlopige cesuur).

Als de itembank gereed is en het assessment is beschreven dan wordt de bank overgedragen aan het *cluster Tentaminering van de afdeling Administratie en onderwijslogistiek*. Voor de richtlijnen voor de aanlevering van CBI-itembanken verwijzen wij naar sectie II.

f. Beschikbaar stellen voorbeeldtentamens

Aan de studenten worden via de elektronische leeromgeving twee voorbeeldtentamens beschikbaar gesteld. Met behulp van deze voorbeeldtentamens kunnen studenten zich voorbereiden op het 'echte' tentamen. Dit tentamen is qua vorm en inhoud vergelijkbaar met de echte toets.

Het itembankconstructieproces staat weergegeven in figuur VIII-1.



Figuur VIII-1. Ontwikkelen itembank (de afdeling T&C heet in 2016 cluster Tentaminering van de afdeling Administratie onderwijslogistiek)

Fase 2: De formele vaststelling van de itembank

Nadat de CBI-itembank is geconstrueerd en opgemaakt, moet de inhoud van de itembank formeel worden vastgesteld door de examinator. Het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek stuurt daartoe een vaststellingsverslag naar de examinator.

Het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek geeft de itembank vervolgens vrij voor tentaminering.

Fase 3: Onderhoud van de itembank

Nadat een CBI-itembank in gebruik genomen is, is het nodig deze itembank te 'onderhouden' met als doel de kwaliteit van de bank te optimaliseren. Onderhoud kan het gevolg zijn van bijvoorbeeld reacties van studenten en uitspraken in beroepszaken (a.), wijzigingen in het cursusmateriaal (b.) en resultaten van vraaganalyses (c.).

In het vervolg van deze paragraaf worden deze verschillende aanleidingen tot onderhoud beschreven.

a. Mondelinge en schriftelijke reacties op tentamens

Mondelinge reacties over kleine fouten in een afgenomen CBI-tentamen komen direct na de tentamenafname binnen bij het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek. Dit kunnen zowel reacties zijn van studenten als van de tentamenbeheerders (de laatstgenoemden houden van elke tentamensessie protocollen bij waarop onregelmatigheden genoteerd dienen te worden). Deze reacties worden doorgegeven aan de examinator en aan de contactpersoon van de desbetreffende faculteit. De examinator geeft aan het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek door welke fouten gecorrigeerd moeten worden. Wil een student inzage krijgen in het tentamen, dan kan hij/zij gedurende vier weken nadat de definitieve uitslag bekend is gemaakt, schriftelijk of per email verzoeken inzage te krijgen in het beoordeelde werk. De student kan dan een afspraak maken met de tentamenbeheerder van het desbetreffende studiecentrum, die ook een kopie van het overzicht ontvangen heeft. Onder toezicht van de tentamenbeheerder kan de student inzage plegen in de door hem of haar fout beantwoorde vragen. Hierbij wordt geen inhoudelijke toelichting gegeven en mogen de vragen niet worden gekopieerd. Ook aantekeningen zijn niet toegestaan. Naar aanleiding van deze inzage kan de student beslissen of hij/zij beroep wil aantekenen.

Beroep en klacht

Een student kan beroep aantekenen tegen de behandeling tijdens de tentamenzitting en tegen de tentamenuitslag bij het College van beroep voor de examens. Als een student beroep wil aantekenen dan mag hij hiervoor tijdens de inzage de desbetreffende vraagnummers opschrijven. Nadat het beroep is ingediend worden de betreffende vragen aan student toegezonden zodat hij het beroep beter kan motiveren.

Voor meer informatie omtrent de beroepsprocedure zie de regeling van het College van beroep op www.ou.nl bij studie-informatie (regelingen en procedures).

Klachten

Studenten kunnen klachten hebben over de gang van zaken tijdens het tentamen. Indien zich dit voordoet of als zich bijzondere situaties voordoen zoals geluidsoverlast, wordt dit vermeld op het protocol van de tentamenzitting en wordt dit gemeld bij de CvE van de opleiding.

Een student kan zich ook rechtstreeks tot de CvE van de opleiding melden met klachten of de klachtenprocedure volgen via het 1-loket.

b. Leerstofinhoudelijke wijzigingen

Als in de loop van de tijd veranderingen optreden in de inhoud van de cursus (revisie van de cursus), zal de inhoud van de tentamens aangepast moeten worden. De vragen die hiervoor geconstrueerd worden zullen weer gescreend en in de itembank opgenomen moeten worden.

c. Vraaganalyse

Drie maanden nadat een CBI-bestand operationeel is geworden, wordt standaard een statistische analyse van de vraaggegevens gemaakt en ter beschikking van de examinator gesteld; daarna gebeurt dit jaarlijks (en, indien nodig, tussentijds op verzoek van de examinator). Ook deze analysegegevens kunnen aanleiding geven tot revisie van het vragenbestand (zie voor de interpretatie van de analysegegevens sectie VII).

Belangrijk is dat de examinator op basis van deze signalen initiatief moet nemen!

Revisie van de itembank betekent dat vragen en/of clusters op in-actief (in QMP krijgt het item de status 'retired') geplaatst kunnen worden en/of dat sleutelwijzigingen aangebracht kunnen worden. Van revisie is ook sprake indien vragen en/of clusters van vragen toegevoegd moeten worden, als tekstuele wijzigingen in vragen c.q. terugkoppelingsteksten noodzakelijk zijn en als het tentamenprofiel of de puntentoekening gewijzigd moeten worden.

Binnen de bovengenoemde mutaties kan een onderscheid worden gemaakt naar acute en periodieke mutaties. Van een acute mutatie is sprake als een sleutfout of een ernstig gebrek aan een vraag is geconstateerd: situaties die direct ingrijpen rechtvaardigen. Van een periodieke mutatie is sprake indien bijvoorbeeld nieuwe vragen aan het vragenbestand moeten worden toegevoegd.

In beide gevallen maakt de examinator een kopie van het te muteren vraag. De mutatie wordt doorgevoerd op de kopie en dit vraag krijgt een verhoogd versienummer. Het originele vraag wordt op 'retired' gezet. Nadat de itembank gereviseerd is moet de gereviseerde bank opnieuw van de ontwikkelomgeving naar de productieomgeving van het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek gezet worden.

d. Andere overwegingen

De verantwoordelijkheid van de faculteit om de inhoudelijke kwaliteit van haar onderwijs te borgen. De verantwoordelijkheid ligt bij de decaan en aan hem/haar is het dan ook om nadere instructies te geven aan de faculteitsmedewerkers (o.a. examinatoren).

De kwaliteit van de tentaminering is een onderdeel van het geheel dat periodiek aan bod kan komen tijdens de accreditatie van de opleiding. Daar is dus ook de reden om voor de verantwoordelijke decaan te bepalen in hoeverre de kwaliteit van de itembank moet worden gewaarborgd.

Wel kan worden overwogen of er een advies voor de decanen kan worden ontwikkeld, die men al dan niet kan opvolgen. Hierbij kan gedacht worden aan het risico dat de items bekend raken bij de andere studenten. Het aantal keer dat een item is gebruikt kan worden geregistreerd en bijgehouden; dit heet de exposure van een item. Men kan denken aan een bepaalde waarde (stel 50). Een rapportage met items met een exposurewaarde van 50 of meer kan aanleiding zijn voor de faculteit om deze items te veranderen of te vervangen. De beslissing om dit door te voeren moet echter bij de faculteit liggen (zie voorgaande argumentatie).

Ook kunnen de gegevens over de exposurewaarden van alle cursussen en bij alle faculteiten een onderwerp zijn voor periodiek overleg (zie advies verder in deze notitie) tussen betrokkenen. Zo zou uit dit overleg een common practice kunnen ontstaan, als leidraad bij dat overleg. De CvE van de opleiding zou in deze situatie de positie kunnen innemen door periodiek bij de faculteiten na te vragen wat de uitkomsten zijn t.a.v. het onderwerp verversing items en het al dan niet naleven van de afspraken die zijn gemaakt bij dat overleg.

Of faculteiten rapporteren uit eigen beweging de opvattingen over het verversen van items en geven aan hoe een en ander is uitgevoerd.

X. Bijzondere verplichtingen

Soms is het niet mogelijk om in een gestandaardiseerde tentamensessie van drie uur alle leerdoelen te toetsen, die men zou willen toetsen. Vandaar dat de mogelijkheid bestaat bijzondere verplichtingen op te nemen naast en soms in plaats van schriftelijke of gedigitaliseerde tentamens. Bijzondere verplichtingen zijn activiteiten die een student met goed gevolg moet verrichten en die, samen met het slagen voor het schriftelijke of digitale tentamen of op zich zelf staand, een voorwaarde vormen voor het behalen van het certificaat. Een bijzondere verplichting is dus noodzakelijk als bepaalde (of alle) leerdoelen niet adequaat getoetst kunnen worden in een schriftelijk of mondeling tentamen. Zo is een gangbare tentamensessie van drie uur niet geschikt om te toetsen of studenten zelfstandig onderzoek kunnen doen en daarover kunnen rapporteren. Een OU-tentamensessie is evenmin geschikt om mondelinge vaardigheden te toetsen.

Voorbeelden van bijzondere verplichtingen zijn:

- uitvoeren van een scheikundepracticum
- analyseren van een computerprogramma
- maken van een werkstuk aan de hand van zelf verzamelde literatuur
- schrijven van een filosofisch essay
- schrijven van een juridisch advies

Anders dan bij gesloten en open vragen zijn in het geval van bijzondere verplichtingen geen antwoordmodellen mogelijk (zie sectie 4 en 5), maar zijn beoordelingsschema's aan de orde. In een dergelijk schema zijn criteria opgesomd op grond waarvan het 'product' moet worden beoordeeld.

Is dat product een schriftelijk werkstuk (essay, recensie, juridisch advies), dan zou men een beoordelingsschema als onderstaand kunnen gebruiken:

Beoordelingsformulier tentamenopdracht score (0 punten = voldoet niet aan het criterium; 1 punt = voldoet gedeeltelijk aan het criterium; 2 punten = voldoet goed aan het criterium).	
• Het gekozen thema is voldoende duidelijk beschreven zodat degene die het materiaal niet kent zich toch een voorstelling ervan kan maken.	0-1-2
• De keuze voor het materiaal is voldoende toegelicht.	0-1-2
• Literatuuropgave is volgens APA-norm.	0-1-2
• De aannames en veronderstellingen zijn voldoende geëxpliciteerd en onderbouwd.	0-1-2
• De redenering is duidelijk.	0-1-2
• Maximale omvang: 4000 woorden.	0 of 2
• Er zijn voldoende verwijzingen naar het bronnenmateriaal (zowel uit deze cursus als ander materiaal) opgenomen en hierbij zijn de officiële verwijzingsvormen in acht genomen.	0-1-2
Cesuur = 7/8, d.w.z. bij een score van 8 of meer heeft men een voldoende behaald	

Leerproces en toetsmoment

Bijzondere verplichtingen leveren specifieke uitdagingen op omdat ze zich op de grens bevinden van leerproces en toetsmoment. De student leert bepaalde vaardigheden door het uitvoeren van een bijzondere verplichting, maar tegelijkertijd wordt hij of zij erop beoordeeld en wordt deze bijzondere verplichting (mede) een voorwaarde tot het behalen van het certificaat. Juist door deze speciale plaats

die bijzondere verplichtingen innemen, moet er speciale aandacht aan worden geschonken .

Omdat bijzondere verplichtingen een toetsmoment in zich dragen, is het van belang dat hierbij ook rekening wordt gehouden met de eisen die aan schriftelijke en digitale toetsen en tentamens gesteld worden. Zo is het noodzakelijk dat ook bijzondere verplichtingen valide zijn. Zij moeten duidelijk betrekking hebben op leerdoelen die in een cursus gesteld worden. Bijzondere verplichtingen moeten eveneens zo objectief mogelijk beoordeeld kunnen worden. Dit betekent dat de eisen die aan een bijzondere verplichting gesteld worden duidelijk vast moeten liggen en dat correctoren goed geïnstrueerd moeten worden. Het beoordelen van bijzondere verplichtingen brengt extra problemen met zich mee omdat toetsmoment en leermoment *verstrengeld* zijn. Het is belangrijk dat leren en toetsen goed van elkaar onderscheiden worden en dat men ervoor zorgt dat het leermoment ook voldoende aandacht krijgt. Studenten moeten de te toetsen vaardigheden tenslotte eerst leren, ze moeten aanwijzingen krijgen en ze moeten kunnen oefenen. Men moet er dus voor waken dat men studenten niet beoordeelt op aspecten die ze nog aan het leren of oefenen zijn.

Een specifiek probleem dat zich bij bijzondere verplichtingen kan voordoen is het fraudeprobleem: hoe garandeert men dat individuele prestaties van studenten beoordeeld worden?

(Wie geïnteresseerd is in mogelijke oplossingen van deze fraudeproblematiek, kan het volgende rapport raadplegen: De Volder, M., Moerkerke, G., & De Roode, F. (2007). Een 5D strategie tegen werkstukfraude, Document 14/2007, , Open Universiteit Nederland. Te vinden op het Intranet, of op te vragen bij het secretariaat van ECOP.)

X-1. Organisatie van bijzondere verplichtingen

De resultaten van bijzondere verplichtingen (scripties, afstudeeropdrachten, practica) mogen alleen worden afgegeven binnen de periode waarin de cursus in exploitatie is en binnen de inschrijftermijn van de student. De resultaten dienen op een beoordelingsformulier doorgegeven te worden aan het cluster Tentaminering van de afdeling Administratie en onderwijslogistiek. Dit resultaatformulier moet volledig ingevuld zijn, d.w.z. voorzien van studentnummer, studentnaam, cursuscode, cursusomschrijving, adres, inleverdatum, naam examinator, datum ondertekening door examinator, en resultaat.

XI. Competentiegericht toetsen

Binnen het hoger onderwijs is toenemende aandacht te constateren voor de ontwikkeling van vaardigheden, persoonlijke groei en competenties. Nieuwe vormen van instructie vragen ook om andere dan de gangbare toetsen, een andere kijk op toetsen of een ander gebruik van toetsresultaten en toetsinformatie.

XI-1. Inleiding

In deze sectie wordt een methode gepresenteerd om toetsen te ontwikkelen voor competentiegerichte curricula. In dit verband worden zowel formele als informele beoordelingsprocedures bedoeld. In competentiegerichte curricula wordt immers om twee redenen beoordeeld:

1. om feedback te geven op het leren: dit dient ter ondersteuning van de inhoudelijke ontwikkeling van de student, gericht op het opstellen en bijstellen van leerplannen en op het verbeteren van prestaties;
2. om feedback te geven naar het formele registratiesysteem: dit dient ter ondersteuning van formele beslissingen die de start, voortgang en afsluiting van het leren bepalen. Voldoende (leer-)prestaties worden vastgelegd in certificaten en diploma's.

Ook in competentiegerichte curricula dienen toetsen uiteraard van hoge kwaliteit te zijn. Die kwaliteit is vanuit verschillend perspectief te analyseren.

Ten eerste dienen de toetsinstrumenten in engere zin (opdrachten en beoordelingsformulieren) van hoge kwaliteit te zijn. Hierbij moet men denken aan representativiteit, authenticiteit en beoordelaarsbetrouwbaarheid.

Ten tweede dienen de toetsprocedures van hoge kwaliteit te zijn. Hierbij moet men denken aan documentatie van toetsontwikkeling, gestandaardiseerde samenwerking tussen studenten, voor het curriculum geldig verklaarde (en dus vastgestelde) toetsinstrumenten en voor het curriculum geldig verklaarde (en dus vastgestelde) beoordelingsprocedures.

Ten derde dient het toetsprogramma van een curriculum een logisch, consistent en op de eindtermen afgestemd geheel te zijn. Afzonderlijke tentamens en beoordelingen dienen op elkaar afgestemd te zijn. De kwaliteiten waaraan men dan moet denken zijn onderhoudbaarheid, consistentie, transparantie, structuur in de vorm van rode (vaardigheids-)draden door het curriculum, informatiebeheer in studentdossiers en hergebruik van beoordelingsformulieren.

De kwaliteit van competentiegerichte curricula zal in belangrijke mate afhankelijk zijn van de bovengenoemde kwaliteiten. Binnen competentiegerichte opleidingen dient een samenhangend toetsprogramma gerealiseerd te worden. Dat toetsprogramma verzamelt informatie over het competentieniveau van de student (via beoordelingen) en 'monitort' de ontwikkeling met een skills management systeem. Het toetsprogramma dient niet alleen informatie op te leveren voor certificeringsbeslissingen, maar ook over het inrichten van het verdere leerproces.

Toetsinstrumenten kunnen via de zogenaamde vierstappenmethode worden ontwikkeld. Elke stap levert een document (toetsspecificatieplan en opdrachtontwerpschema) of een toetsinstrument (opdracht en beoordelingsformulier) op. De vierstappenmethode leidt niet alleen tot kwalitatief hoogstaande instrumenten, maar ook tot documentatie van de genomen beslissingen. De methode is ontwikkeld voor een opleiding waar de tentamens en de instructie niet het 'bezit' zijn van een individuele docent. Opdrachten en beoordelingsformulieren worden in groepen ontwikkeld. Een groep ontwikkelt toetsen voor (aspecten van) competenties en plant de inzet van de opdrachten en beoordelingen in het curriculum. Bij een dergelijke werkwijze past dat de bevindingen van een projectgroep-toetsconstructie

worden vastgelegd in rapporten, opdrachten en beoordelingsformulieren. Een dergelijke werkwijze maakt het onderwijs transparant voor de verschillende belanghebbenden, zoals studenten, (intredende) docenten, leden van accreditatiecommissies en commissies voor de examens. We benadrukken hier de gerichtheid op de producten vanwege de bijdrage aan de transparantie, doelgerichtheid en efficiëntie van het leren en de instructie.

De vierstappenmethode levert de volgende producten op:

1. Een toetsspecificatieplan: een beschrijving van functie en context van de toets.
2. Een opdrachtontwerpschema: een beschrijving van de aspecten van opdrachten door een uitwerking van handelingen, contexten van handelingen, noodzakelijk instrumentarium, informatiebronnen, kwaliteitscriteria en probleemtypologie.
3. Opdrachten die dienen om studenten uit te nodigen tot zinvol en doelgericht handelen.
4. Beoordelingsformulieren die dienen om beoordelaars expliciet, motiverend en objectief te doen beoordelen.

XI-2. Het toetsspecificatieplan

Het toetsspecificatieplan legt vast wat de onderwijskundige context van de toets is, de functie van de toets, welke kenmerken de studenten hebben (relevante kennis en ervaring, taalvaardigheid, mogelijke belemmeringen, e.d.) en wie de gebruikers van de toetsresultaten zijn. Het toetsspecificatieplan levert een pakket van eisen en specificaties en is onderdeel van een cursus- of curriculumplan. Het toetsspecificatieplan dient als leidraad bij de constructie van toetsen, maar ook bij de evaluatie van de kwaliteit van toetsen. Het toetsspecificatieplan geeft antwoord op vragen als:

- Wat is het doel/functie van de toets? (bijvoorbeeld: diagnosticeren, certificeren, selecteren?)
- Wie neemt op basis van de beoordeling/toetsscore een beslissing? (de docent, examencommissie of de student zelf ingeval van een diagnostische toets)
- Wat gebeurt er met het resultaat van de beoordeling? (gaat het om het vaststellen van het niveau van een individuele student met als doel een formele beoordeling of gaat het om het bijsturen van het leerproces, individueel of op opleidingsniveau?)
- Wat is het toetsbeleid/visie op toetsing waarbinnen beoordeling, tentamen of toets moet worden vormgegeven? (beleid instelling/opleiding, randvoorwaarden, uitgangspunten)
- Wie zijn de studenten? (opleiding, fase opleiding, aantal en eventueel bijzondere kenmerken)
- Welke middelen zijn nodig? (apparatuur, ICT, wie is aanspreekpunt, tijd van studenten en docenten)

In een toetsspecificatieplan wordt op een aantal punten aandacht besteed aan de rol van de studenten. Met de introductie van het projectonderwijs is aandacht gekomen voor de betrokkenheid van de studenten bij beoordelingen. Voor goede samenwerking dienen studenten elkaar aan te kunnen spreken op de kwaliteit van de prestaties. Het vermogen om het eigen handelen en het handelen van anderen te beoordelen en te waarderen is één van de vaardigheden die geoefend en getoetst wordt. Ingeval van self assessment leert de student een kwalitatieve beoordeling te geven van de eigen werkzaamheden en producten in relatie tot criteria die gelden in het werkveld. Dit sluit nauw aan bij conclusies uit onderzoek naar expertgedrag: experts zijn in staat het eigen gedrag te 'monitoren' en te sturen aan de hand van vakinhoudelijke criteria en protocollen. In de beroepsuitoefening is het beoordelen van voorstellen en verslagen van collega's, vakgenoten, projectmedewerkers en gepubliceerd werk een veelvoorkomend onderdeel van het functioneren.

Een voorbeeld van een toetsspecificatieplan is hieronder gegeven:

Voorbeeld 1: een toetsspecificatieplan voor een bestuurskundige cursus

1. De beoordeling moet leiden tot een formele beslissing; bepaald dient te worden of de student een specifieke vaardigheid beheerst.
 2. De beoordeling moet informatief zijn naar het product, d.w.z. voldoende inzicht geven in zwakke en sterke kanten van het product of de prestatie.
 3. De beoordeling moet informatief zijn naar het niveau van de vaardigheid, d.w.z. voldoende inzicht geven in zwakke en sterke kanten van de vaardigheid van een individuele student ten opzichte van het gewenste eindresultaat.
 4. De studenten kunnen zich voorbereiden door oefenopdrachten welke van hetzelfde niveau zijn als de toetsopdrachten.
 5. De oefen- en toetsopdrachten zijn openbaar.
 6. De beoordelingsformulieren zijn openbaar.
 7. De toetsen moeten kunnen worden afgenomen op individuele basis, maar ook groepsgewijs.
 8. De studenten hebben veelal een functie in het openbaar bestuur.
 9. De studenten dienen toegang te hebben tot Internet.
-

Bij peer assessment beoordeelt de student het gedrag en de prestaties van zijn medestudenten.

In het toetsspecificatieplan in Voorbeeld 1 wordt de betrokkenheid geregeld door de beoordelingsformulieren onderdeel van de instructie te maken. Daarnaast wordt het recht van de student op een gemotiveerde beoordeling concreet gemaakt door vast te leggen dat de beoordeling informatief moet zijn naar de geleverde prestatie (waardering van gedrag) en naar de te beoordelen vaardigheid (waardering van de ontwikkeling van een persoon). Het recht op een individuele beoordeling wordt concreet gemaakt door vast te leggen dat beoordeling en terugkoppeling individueel zijn. In het examenreglement (en dus niet noodzakelijk in elk van de toetsspecificatieplannen) dient te zijn vastgelegd dat een student recht heeft op een beoordeling door een lid van de staf of een andere beroepsbeoefenaar die daartoe is erkend. Dit betekent dat een examinator bijvoorbeeld in het geval van peer-assessment eindverantwoordelijk is voor de beoordeling en het oordeel van peers overneemt.

XI-3. Het opdrachtontwerpschema

Het opdrachtontwerpschema geeft een beschrijving van de vaardigheid en de aspecten die een rol spelen bij het construeren van opdrachten. Het opdrachtontwerpschema koppelt het object van de toetsing (kennis, vaardigheid, competentie) aan de inhoud en de vorm van de opdrachten. In voorbeeld 2 (zie hieronder) wordt voor de vaardigheid Contacten onderhouden voor bestuurskundigen een beschrijving gegeven van de aspecten van deze vaardigheid en de aspecten van de omstandigheden waarin deze worden aangewend door een bestuurskundige.

Voorbeeld 2: opdrachtontwerpschema contacten onderhouden voor bestuurskundigen

Handelingen

Informeren, argumenteren, overtuigen, onderhouden van relaties, dossiervorming, corresponderen

Prestaties

Producten

Dossier, procedures, projectplan, plan van aanpak, nota, correspondentie, presentatie, web-pagina

Demonstratie

Lezing, voorlichtingsgesprek, vergadering, productverdediging

Managementactiviteiten

Relatiemanagement, projectmanagement, informatiemanagement

Opdrachtgever

Overheid, semi-overheid, bedrijfsleven, non-gouvernementeel

Rollen

Bestuurder (wethouder, gedeputeerde)

Medewerker (mandaathouder, projectleider)

Methoden en technieken

Analyse

Actor proces element methode, configuratie analyse, reconstructie beleidstheorie, framing met heuristieken

Ontwerp

Actor proces element methode, inconsistentie model Snellen, wenken van Hoppe

Evaluatie

Effectrapportage, praktische effectmeting, methode Fischer, ex ante evaluatie, wetsevaluatie

Tools

Digitale informatie, projectmanagement, data-analyse, tekstverwerking, presentatie, ICT, literatuur (boek, artikel, rapport)

Probleemsituatie

Ketenbeheer, verzelfstandiging, regulering

Ingrediënten van het probleem

Milieu, veiligheid, economische ontwikkeling, werkgelegenheid, Europese samenwerking, betrokkenheid burgers, overheid als werkgever, specifieke taken overheid, politieke besluitvorming

Criteria

Handelen (algemene beginselen van behoorlijk bestuur, systematisch, gedocumenteerd)

Product (argumentatieleer, leesbaarheid, volledigheid, functionaliteit, specificiteit, voldoet aan standaard werkveld en kennisdomein, correctheid)

Toetsspecifieke aspecten

Is beoordelingsmoment gecreëerd of onderdeel van lopende activiteiten?

Wordt veel informatie gegeven of weinig over de casus?

In welke mate is de opdracht sturend?

Veronderstelt de casus veel of weinig (voor-)kennis van de problematiek?

10.3.1 Competenties en vaardigheden

Het object van toetsing is een competentie of onderdelen van een competentie. Bij het uitwerken van het competentiebegrip maken wij gebruik van de toelichting van Kessels (1997). Kessels biedt een beschrijving van competenties die bruikbaar is bij het classificeren van eindtermen. Hij schetst de competent functionerende mens in een informatiemaatschappij waarin kenniscreativiteit noodzakelijk is voor innovatie. De student zal na het hoger onderwijs een arbeidsplaats gaan bezetten in een wereld die gekenmerkt wordt door globalisering, een explosieve groei van kennis en technologie en verandering van werkprocessen. Naast inhoudelijke veranderingen in de werkvelden en kennisdomeinen verandert ook de vraagfactor: de nieuwe 'klanten' van het hoger onderwijs combineren en integreren werken en studeren, zowel bij de start van het werkzaam leven als tijdens het werkzaam leven. De vele nieuwe vormen van duaal onderwijs illustreren dit. Ook bij de 'reguliere' programma's wordt meer en meer contact gezocht met de werkvelden om instructie en beoordeling actueel te houden. Op die manier kunnen studenten goed voorbereid worden op een functie na hun opleiding. Competentie wordt geleerd in onderwijsvormen waarbij complexe problemen worden aangepakt, zowel bij individueel als bij gezamenlijk leren. De ontwikkelingen in de werkvelden gebiedt de inzet van informatie- en communicatietechnologie bij het aanpakken van problemen. Het gebruik van ICT en boeken vormt een betere afspiegeling van de beroepspraktijk dan groepstentamens met rekenmachines en kernbetrekkingskaarten.

Kessels onderscheidt vijf aspecten aan de competentie die nodig is in de snel veranderende informatiemaatschappij. In samenhang definiëren de volgende aspecten een competentie:

1. **Materiedeskundigheid:** kennis, begrippen, methoden en technieken, regels en procedures die specifiek zijn voor het beroep en waar een professional weet van moet hebben om tot goede resultaten te komen. Binnen een bedrijf zal de specifieke invulling hiervan nauw samen moeten hangen met de beoogde kerncompetenties van het bedrijf. Voor iedere afgestudeerde is dat anders. De afgestudeerde dient vaardig te zijn in het snel verzamelen, verwerken en bruikbaar maken van informatie en aanpakken die nodig zijn om een opdracht uit te voeren.
2. **Probleemaanpak vaardigheden:** het kunnen gebruiken van de verworven materiedeskundigheid om problemen uit het werkveld aan te kunnen pakken (kunnen analyseren, modellen en theorieën kunnen gebruiken bij het formuleren van ideeën, kennis uit verschillende kennisdomeinen kunnen integreren tot een consistente probleembeschrijving, etc.). De afgestudeerde dient vaardig te zijn in het doelgericht toewerken naar een gewenst resultaat vanuit een problematische situatie.
3. **Netwerkvaardigheden:** in staat zijn om binnen de context van het beroep functionele contacten aan te gaan met anderen (sociale en communicatieve vaardigheden, maar ook betrokkenheid bij

anderen). De afgestudeerde dient een goede basis te hebben voor verdere ontwikkeling van communicatieve vaardigheden, vaardigheid in (project)management en de vaardigheid te werken met een grote mate van zelfstandigheid in (multidisciplinaire) teams.

4. Zelfregulatie van competentiegroei en professionele ontwikkeling: een professional wordt geacht in staat te zijn om in relatie tot de kerncompetenties van het bedrijf nieuwe kennis op het spoor te komen, zich toe te eigenen en toe te passen. De afgestudeerde zal de eigen persoonlijke en professionele ontwikkeling ter hand moeten nemen. Reflectieve vaardigheden en meta-cognities dienen goed ontwikkeld te zijn.
5. Zelfregulatie van commitment: een professional wordt geacht in staat te zijn de eigen relationele verhouding tot het vak dat hij uitoefent, de organisatie waarin hij werkzaam is en de werkrelaties die hij in stand moet houden te verbeteren. De afgestudeerde zal in staat moeten zijn om constructieve bijdragen te leveren aan zichzelf en zijn omgeving. Bij het onderhouden van contacten zullen bijv. bestuurskundigen gebruik maken van materiedeskundigheid (informatie over het probleem, kennis van methoden), probleemaanpak-vaardigheden (zelfstandig opdrachten en taken uitvoeren) en werken met anderen (in verschillende rollen).

10.3.2 Kenmerken van opdrachten

Het doel van het opdrachtontwerpschema ligt besloten in de naam: het dient de groep docenten te ondersteunen bij de constructie van opdrachten. In het opdrachtontwerpschema worden kenmerken van het handelen opgesomd. Die kenmerken beschrijven de prestatie van de student, de omstandigheden waaronder de prestatie tot stand komt, de methoden, theorieën, tools en werkwijzen die nodig zijn om de prestatie te leveren, de rol die de student kan aannemen in de opdracht, de criteria die aan de producten en het gedrag zijn te stellen, enz. Als uitgangspunt voor voorbeeld 2 dienen de werkzaamheden van een bestuurskundige, bijv. in dienst van een provincie of een adviesbureau. Hoe ziet het kantoor eruit? Welke opdrachten worden gegeven? Welke problemen dienen zich aan? Welke prestaties worden geleverd? Tegen welke criteria worden deze beoordeeld? Welke bestuurskundige methoden staan de bestuurskundige ter beschikking? Welke rollen vervult de bestuurskundige?

Hoe komt een opdrachtontwerpschema tot stand? In ieder geval dient dat in een groep docenten te gebeuren, eventueel aangevuld met onderwijskundigen en andere externe partners. Het uitgangspunt is de handelende professional. In een open en creatieve analyse (eerst zoveel mogelijk opsommen, verwijderen kan altijd nog) worden de belangrijkste kenmerken van handelen in kaart gebracht in een begrijpelijke, contextgebonden terminologie. Het opdrachtontwerpschema van voorbeeld 2 kwam tot stand door het bestuderen van cursusmaterialen en bestuurskundige literatuur, gesprekken met docenten en voorbeelden van andere opdrachtontwerpschema's (bedrijfskunde, chemie, journalistiek, weg- en waterbouw).

In een analyse van een aantal opdrachtontwerpschema's onderscheidde Moerkerke de volgende algemene kenmerken van opdrachten: de student zal aan de hand van een probleem in een betekenisvolle context een serie handelingen gaan verrichten om tot een gewenst resultaat te komen. Bij het aanpakken van het probleem zal de student een aantal hulpmiddelen of tools ter beschikking staan, zal de student gebruik moeten maken van theorie en methodologie en zal de prestatie van de student aan de hand van -binnen de professie geaccepteerde- kwaliteitscriteria beoordeeld worden. Daarnaast dienen de opdrachtontwikkelaars rekening te houden met additionele zaken die alleen een rol spelen omdat er getoetst wordt. Deze toetsgebonden beslissingen bepalen onder meer de mate van zelfstandigheid van de student en de complexiteit van de opdracht.

Een 'vergeten' onderwerp in de opdrachtconstructie betreft de validiteit van het beoordelen. In het vaardigheidsonderwijs is men van meet af aan gewend geweest om de consistentie tussen beoordelaars

uit te leggen als overeenstemming tussen scores. De belangstelling voor de inhoudelijke motivering van de beoordeling is wat geringer. Moerkerke beschrijft een casus waarbij docenten Informatica voldoende numeriek consistent beoordeelden, maar waar de inhoudelijke analyse en aanwijzingen voor verbetering behoorlijk verschilden. De inhoudelijke waardering voor een computerprogramma was afhankelijk van de 'school' waartoe de docenten behoorden. De docenten hadden veelal een exacte achtergrond of een achtergrond in software engineering. Ten opzichte van elkaar verschilden de inzichten in wat een 'goed' programma zou moeten zijn. Stiggins is een van de weinigen die benadrukt dat in deze fase van opdrachtontwikkeling veel aandacht moet uitgaan naar de kwaliteitscriteria voor de prestaties. Door kwaliteitscriteria te benoemen en te definiëren wordt zowel aan de probleemoplosser (de student) als aan de beoordelaar (student, docent) duidelijk gemaakt wat 'goed' gedrag is in het vakgebied en wat een 'goed' product is. Ook voor het beoordelen geldt dat de uitgangspunten dienen aan te sluiten bij de inzichten in de werkvelden. Niet alleen de opdrachten maar ook de beoordelingsformulieren bepalen toetskwaliteiten als inhoudsvaliditeit, authenticiteit en representativiteit. In voorbeeld 2 (zie boven) worden (kwaliteits)criteria voor het handelen en voor de producten opgesomd.

XI-4. De opdrachten

Een opdracht nodigt een student uit om kennis en vaardigheden te tonen. Als eenmaal de prestatie geleverd is, wordt deze beoordeeld met behulp van een beoordelingsformulier. Veel beoordelingen in het onderwijs zijn gepland en voor de gelegenheid georganiseerd. In die gevallen worden de toetsmomenten in het rooster opgenomen. In het onderwijs kan echter ook gebruik gemaakt worden van beoordelingsformulieren in situaties die niet voor iedere individuele student geroosterd en gepland zijn. Het (project-)onderwijs gaat gepaard met allerlei 'natuurlijke' activiteiten als overleg, vergaderen, presenteren, e.d. In deze 'rijke' omgeving hoeven niet alle prestaties gepland te worden om toch zinvol beoordeeld te kunnen worden. Het verdient dan ook aanbeveling om niet alleen de precies omschreven opdrachten, maar ook de meer zich vanzelf aandienende probleemsituaties te beschouwen als zinvolle beoordelingsmomenten. Zowel geplande als meer natuurlijke probleemsituaties worden gespecificeerd met het opdrachtontwerpschema. Naast het specificeren en uitwerken van opgaven dient ook nog te worden bepaald hoeveel opdrachten voorgelegd dienen te worden om met enig vertrouwen te kunnen beoordelen of een bepaald onderdeel van een competentie voldoende beheerst wordt.

10.4.1 De constructie van opdrachten

Een opdracht wordt in twee stappen gemaakt. Bij het ontwerpen van de opdracht worden logische combinaties gezocht in het opdrachtontwerpschema. Per categorie worden een of meer elementen gekozen. Op deze manier ontstaat een verzameling van specificaties voor een opdracht. Dit wordt ook wel een profiel met opdrachtkenmerken genoemd. Een profiel definieert opdrachten van hetzelfde type. Uit onderzoek is gebleken dat deskundigen goed in staat zijn om aan de hand van profielen ontwerpen van opdrachten te bediscussiëren. Dit geldt zeker ook voor situaties waar verschillende disciplines nodig zijn om de toets te ontwerpen zoals simulaties van complexe processen of projectonderwijs.

De opdracht Ketenbeheer van voorbeeld 3 is te beschrijven aan de hand van een profiel gebaseerd op het opdrachtontwerpschema in voorbeeld 2. De handelingen die uitgelokt worden zijn informeren, argumenteren en overtuigen. De student dient een nota te schrijven. De context van de opdracht is een overheidsorganisatie, het type probleemsituatie is dat van ketenbeheer, de student neemt de rol aan van medewerker, dient ICT te gebruiken, de nota wordt beoordeeld aan de hand van criteria uit de argumentatieleer, criteria voor functionaliteit, leesbaarheid, volledigheid, enz. Bij het specificeren van de inhoud van de opdracht wordt in hoofdlijnen de gewenste vorm van een opdracht vastgelegd.

Voorbeeld 3: Casus Ketenbeheer (overgenomen van Arendsen, 1999)

Opdracht Ketenbeheer; contact onderhouden

In deze taak bent u medewerker van de algemene bestuursdienst van de provincie

Waddermeer. De provincie heeft vanwege haar ligging een sterke binding met de visserij. In de afweging tussen het economisch belang van de visserij en het milieubelang van de Waddenzee is uw bestuur voorstander van visserij. Maar tegelijk wil uw provincie rekening houden met integraal ketenbeheer. U bent mede verantwoordelijk voor het onderhouden van contacten met gemeenten, provincies en andere overheden. In het kader van integraal ketenbeheer is uw taak het aanleveren van argumenten om het beleid van uw provincie in dat verband met overtuiging over te dragen aan andere betrokken overheden. U dient er steeds rekening mee te houden dat uw provincie daarbij deel uit maakt van een netwerk met andere overheden en maatschappelijke groepen. In de casus waar u mee te maken heeft gaat het om de schelpdiervisserij in de Waddenzee. Het probleem waar u voor staat is het verzamelen van overtuigende argumenten voor het standpunt van uw provincie in later overleg met diverse betrokken andere overheden. In de casusbeschrijving vindt u nader omschreven waar het precies om gaat bij de schelpdiervisserij. Het product dat van u wordt verwacht is een beleidsnota waarin voor interne doeleinden argumenten op een rij gezet worden. Deze argumenten moeten gepresenteerd kunnen worden in overleg met andere overheden. Uitgangspunt is daarbij het standpunt van de eigen provincie, pro-visserij in het kader integraal ketenbeheer. Als bestuurskundige wordt van u verwacht dat u daarbij oog heeft voor het netwerk van betrokken actoren. De criteria waaraan het product moet voldoen vindt u aan het eind van de opdracht.

In de casusbronnen vindt u een belangrijke nota over de visserij van het Ministerie van Landbouw en visserij (LNV): Structuurnota zee en kustvisserij.

Over de visserij vindt u materiaal op de volgende plekken:

- overzicht visserij waddenzee etc.
- feiten en cijfers LNV over visserij
- algemeen LNV over visserijbeleid
- beheersplan Waddenzee
- nationaal Waddenzeebeleid

Over de Waddenzee vindt u materiaal in verband met de discussie over de visserij op:

- <http://www.waterland.net/interwad/discussie/discussie.html>
- <http://www.waterland.net/interwad/>
- discussieplatform VROM
- overzicht nota's LNV
- nota LNV schelpdiervisserij
- contact minister-kamer over schelpdiervisserij

Criteria

We verwachten van u een volledige nota, een leesbaar verhaal waarin u de vraag beantwoordt die u als opdracht heeft gekregen. Voor het beoordelen van het product hanteren wij de volgende criteria:

- heeft de nota een vorm die overeenkomt met het doel ervan (doelgroep, werkwijze, randvoorwaarden)
 - is de nota goed leesbaar
 - is het probleem gedefinieerd overeenkomstig de opdracht
 - worden argumenten in termen van integraal ketenbeheer overeenkomstig de belangen van de eigen organisatie gegeven
 - zijn de argumenten voldoende onderbouwd
 - in de zin van de argumentatieleer
-

-
- in de zin van door anderen ook te vinden documenten en literatuurverwijzingen
 - is het probleem geïnterpreteerd in termen van relevante beleidskundige theorieën
 - worden begrippen gebruikt in overeenstemming met de bekende beleidskundige literatuur
 - is op de juiste manier een analyse van het netwerk gemaakt
 - wordt de nota afgesloten met helder geformuleerde aanbevelingen
-

10.4.2 De vorm van opdrachten

Bij het ontwerpen van de opdrachten dient rekening te worden gehouden met eerder aangegeven randvoorwaarden zoals de tijd die student en beoordelaar kunnen besteden aan de opdracht. De inhoud en de randvoorwaarden bepalen uiteindelijk de vorm van de opdrachten. Wij gaan hier niet nader in op de wijze waarop de kosten van het beoordelen berekend worden. Deze zijn onderdeel van het interne bekostigingsstelsel van de onderwijsorganisatie. Wij concentreren ons hier op de opbrengst van de inspanning die toetsen heet: inzicht in de aard en het niveau van competentie van elke student. Bij de keuze van de opdracht dient het perspectief op de opbrengst te zijn: welke informatie is nodig en welk type opdracht verschaft die informatie?

Een werkmodel voor de koppeling van leerdoelen aan meest geëigende toetsvormen staat in voorbeeld 4. Het werkmodel is niet een voorschrift maar dient als oriëntatie op mogelijke toetsvormen. Voor een classificatie van leerdoelen wordt de indeling van Kessels gebruikt. In voorbeeld 4. worden als mogelijke opdrachten genoemd: gestandaardiseerde opgaven, authentieke taken voor individuele afname en teamwerk en standaardinstrumenten voor persoonlijke effectiviteit.

Voorbeeld 4. De meest geëigende toetsvormen en instrumenten voor het monitoren van groei uitgezet tegen verschillende aspecten van competentie

Materiedeskundigheid

toetsen met meerkeuze-opgaven of andere kleine opgaven

Probleemoplosvaardigheden

complexe vaardigheden: taken die leiden tot systematisch handelen dat leidt tot prestatie (incl. self assessment)

Netwerkvaardigheden

samenwerkend onderwijs: taken in groepen (inclusief self- en peerassessment)

Zelfregulatie professionele ontwikkeling

reflectie n.a.v. systematische zelfinschatting en inschatting door competentie manager

Zelfregulatie commitment

reflectie n.a.v. systematische zelfinschatting, inschatting door competentie manager,

Gestandaardiseerde opgaven

Gestandaardiseerde opgaven worden vooral gebruikt bij groepstentamens met een duur van 2-3 uur. Gestandaardiseerde opgaven zijn vooral geschikt voor het meten van kennis (denk- en doe-tijd van 2-3 minuten) en vaardigheden die worden aangewend bij problemen met een lage intensiteit (denk- en doe-tijd van 15 minuten). Alhoewel de actualiteit van basiskennis steeds sneller vermindert door snelle ontwikkelingen in kennis en technologie blijft op het moment van het functioneren zelf materiedeskundigheid van groot belang. Basiskennis is van belang voor het efficiënt en effectief aanpakken van problemen. Voor zover basiskennis wordt getoetst en getentamineerd is het aan te raden om vooral instrumenten te ontwikkelen die efficiënt zijn. Toetsen met meerkeuze-opgaven beperken de dure nakijktijd. Tentamens met open vragen kunnen informatiever zijn, maar leiden in veel gevallen tot een relatieve overbelasting van het personeel van wie de tijd beter ingezet kan worden voor het beoordelen van authentieke prestaties. Voor het vaststellen van deelvaardigheden kunnen evenwel goed schriftelijke toetsen met gestandaardiseerde opgaven gebruikt worden (ontwerpvaardigheden, programmeren, interpreteren van gegevens) mogelijk als voorbereiding op de uitvoering van meer reële complexe opdrachten ontleend aan de praktijk.

Authentieke taken

Authentieke taken zijn opdrachten waaruit een sterke betrokkenheid op het werkveld blijkt. De student wordt in de rol geplaatst van een actor in het werkveld die bepaalde taken krijgt. Dit kunnen grotere of kleinere taken zijn, waarbij een student kan tonen dat deze een product kan opleveren (een brief, een verslag, een nota, een ontwerp), of een complex proces kan sturen (een vergadering) of een complex proces kan managen (relatiebeheer, een productieproces). Om netwerkvaardigheden te toetsen worden opdrachten aan groepen studenten gegeven of aan studenten in een werkkring. Self-assessment en peer assessment aan de hand van beoordelingsformulieren vormen onderdeel van de instructie. De formele en informele beoordeling van producten, werkwijzen en informatie vindt plaats aan de hand van beoordelingsmethoden.

| Standaardinstrumenten voor zelfreflectie

De zelfregulerende vaardigheden worden in kaart gebracht door standaardinstrumenten. De afgestudeerde wordt geacht de eigen professionele groei en de eigen productiviteit te sturen.

10.4.3 De samenstelling van de toets

Een toets is samengesteld uit een aantal meetmomenten (de opdrachten) die elk tot doel hebben een (onderdeel van een) competentie te toetsen. Voor het samenstellen van de toets dient het benodigd aantal opdrachten te worden bepaald. Er moet een balans gezocht worden tussen de juiste intensiteit van de meting en het aantal keren dat de student de beoogde prestatie moet leveren. Met intensiteit bedoelen we de tijdsduur die de opdracht beslaat. Met sommige opdrachten is een student drie minuten bezig, met andere opdrachten enige uren of enige dagen. Opdrachten die leiden tot een realistisch complex ontwerp of onderzoek kunnen vele weken in beslag nemen. Het mag duidelijk zijn dat naarmate de intensiteit toeneemt, de wenselijkheid van het aantal herhalingen afneemt. Het is alleszins redelijk dat een toets met meerkeuze-opgaven uit een 30-tal opgaven bestaat. Opdrachten als de casus Ketenbeheer (voorbeeld 3) kunnen niet al te vaak herhaald worden vanwege de hoge intensiteit. De meetbetrouwbaarheid van de toets is mede afhankelijk van het aantal opdrachten waaruit de toets bestaat. De algemene regel luidt: hoe meer opdrachten een student aanpakt, des te betrouwbaarder het cijfer is.

XI-5. De beoordelingsformulieren

Het beoordelingsformulier is een hulpmiddel bij het scoren en beoordelen van de prestatie. Scoren is het toekennen van getallen aan waargenomen prestaties, in het algemeen in de vorm van scorepunten. Beoordelen is het toekennen van een waardering aan de waargenomen prestaties. Een waardering kan kwalitatief van aard zijn (goed/voldoende/onvoldoende), maar ook kwantitatief van aard (de bekende cijferschaal van 1 tot en met 10). Het beoordelingsformulier in voorbeeld 5 heeft dan ook een sectie ter ondersteuning van de scoring en een sectie ter ondersteuning van de beoordeling.

10.5.1 De beoordelaars en de beoordelingsformulieren

Beoordelingsformulieren hebben een tweeledig doel. Ten eerste worden deze gebruikt om de objectiviteit en betrouwbaarheid van de beoordelingen te vergroten. Objectiviteit van een toets is een kenmerk van die toets indien de persoon van de beoordelaar geen invloed uitoefent op de hoogte van de score en de uiteindelijke beoordeling. Betrouwbaarheid van een toets heeft betrekking op de mate waarin scores consistent of reproduceerbaar zijn. Uit onderzoek is meermalen gebleken dat het gebruik van beoordelingsformulieren leidt tot grotere objectivering en een hogere meetbetrouwbaarheid. In het onderwijs dient de beoordelingsmethode zoveel mogelijk analytisch van aard te zijn, omdat deze (a) leidt tot een betere en dus zorgvuldiger motivering van het oordeel en (b) informatie bevat waarmee het leerproces kan worden bijgestuurd. Voor de kwaliteit van de feedback is het van belang dat studenten de beschikking krijgen over de beoordelingsmethode, inclusief de modelantwoorden en de beoordelingscriteria. Een voorbeeld van een beoordelingsformulier is hieronder weergegeven:

Voorbeeld 5: Beoordelingsformulier behorend bij casus Ketenbeheer (overgenomen van Arendsen, 1999)

Instructie voor beoordelaars voor het beoordelen van de opdracht 'contact onderhouden'

De nota Ketenbeheer, het eindproduct van de opdracht contact onderhouden, beoordeelt u aan de hand van onderstaand beoordelingsschema.

Elk onderdeel scoort u met waardering: 5=goed/volledig; 4=voldoende; 3=twijfelachtig; 2=onvoldoende; 1=zeer onvoldoende

- a. Vorm en leesbaarheid van de nota
 - de nota heeft een vorm die overeenkomt met het doel (doelgroep, werkwijze, randvoorwaarden)
 - de opbouw van de nota in hoofdzaken is logisch en die opbouw is helder
 - het taalgebruik sluit aan bij die van de doelgroep
 - begrippen worden gedefinieerd en indien nodig uitgelegd
 - alle elementen van de tekst zijn relevant met het oog op de probleemstelling
 - de rode draad van het betoog wordt goed vastgehouden
 - b. Probleemdefiniëring
 - het probleem is gedefinieerd overeenkomstig de opdracht
 - de probleemstelling is helder geformuleerd en zo nodig in deelvragen uiteengelegd
 - c. Argumentatie
 - argumenten worden in termen van integraal ketenbeheer overeenkomstig de belangen van de eigen organisatie gegeven
 - bij een standpunt zijn steeds zowel argumenten pro als contra vermeld en daar is
-

-
- zo nodig een beredeneerde afweging tussen gemaakt
 - de herkomst van belangrijke uitspraken is duidelijk, zo nodig is de bron vermeld
 - argumenten zijn onderbouwd door verwijzingen naar gebruikte bronnen en literatuur
 - d. Probleemanalyse/koppeling theorie-praktijk
 - er is op de juiste manier een analyse van het netwerk gemaakt
 - het probleem is geïnterpreteerd in termen van relevante beleidskundige theorieën
 - begrippen worden gebruikt in overeenstemming met de bekende beleidskundige literatuur
 - e. Oplossing
 - de nota wordt afgesloten met helder geformuleerde aanbevelingen
 - de aanbevelingen sluiten aan bij de probleemstelling

Vaststellen resultaat opdracht Ketenbeheer

Bereken totaalscore door alle scores per onderdeel (17) bij elkaar op te tellen. Ligt de totaalscore tussen de 17 en 56 punten, dan is de opdracht onvoldoende en zal het product op de onvoldoende scorende elementen moeten worden bijgesteld. Bij een score van 57 punten of hoger is de opdracht voldoende.

De score-cijfer-transformatie luidt als volgt:

Score	Cijfer
17-33	1
34-39	2
40-45	3
46-51	4
52-56	5
57-62	6
63-68	7
69-73	8
74-79	9
80-85	10

In de toetspraktijk is van oudsher veel aandacht voor de oorzaken van geringe objectiviteit en betrouwbaarheid. Beoordelaarseffecten treden (doorgaans) onbewust en ongewild op, maar staan de objectiviteit en daarmee de betrouwbaarheid van het oordeel in de weg. In het algemeen worden de volgende bronnen onderscheiden:

- Het contaminatie-effect: de beoordelaar kent meerdere functies toe aan een beoordeling zoals het 'flatteren' van de beoordeling om de motivatie van een student op te vijzelen.
- Het halo-effect: de beoordelaar laat zich bij de beoordeling beïnvloeden door irrelevante aspecten van de geleverde prestatie zoals het hoger waarderen van een matige prestatie van een in het algemeen 'goede' student en het lager waarderen van een goede prestatie van een doorgaans 'zwakke' student.

- Normverschuiving: de neiging om zich tijdens het beoordelingsproces aan te passen aan het gemiddeld prestatieniveau van de studenten.
- Het sequentie-effect: de invloed op de beoordelaar die uitgaat van een reeks eerdere beoordelingen.
- Het signifisch effect: de invloed die optreedt indien verschillende beoordelaars hun taak verschillend opvatten, zoals bij het beoordelen van schriftelijke werkstukken: sommigen hechten meer waarde aan een correcte spelling, anderen aan stijl.

10.5.2 Het beoordelingsformulier als onderdeel van de beoordelingsmethode

Het beoordelingsformulier is de kern van de beoordelingsmethode. Een beoordelingsmethode dient uit drie onderdelen te bestaan:

- De beoordelingsinstructie: dit bestaat uit een opsomming van te hanteren criteria uit het opdrachtontwerpschema en een toelichting op de kwaliteitscriteria aan de hand van positieve en negatieve voorbeelden
- Een beoordelingsformulier: een formulier dat drie verschijningsvormen kan hebben, namelijk een checklist, een beoordelingsschaal of een productschaal. Met een checklist wordt gescoord of aspecten van het product of het gedrag aanwezig zijn of niet. De beoordelaar scoort per aspect met 'ja' of 'nee'. Met een beoordelingsschaal wordt de mate van waardering aangegeven (bijvoorbeeld een schaal van 1 tot en met 10 of een schaal met onvoldoende, voldoende en uitmuntend). Een bijzondere vorm van een beoordelingsschaal is een productschaal. In een productschaal wordt elk van de categorieën toegelicht aan de hand van beschrijvingen, foto's of voorbeeldproducten.
- Een training van docenten en studenten in het hanteren van het beoordelingsformulier

XI-6. Het registreren van toetsinformatie

De hier gepresenteerde werkwijze is bedoeld ter ondersteuning van projecten waarin opdrachten worden ontwikkeld. Door de introductie van competentiegericht onderwijs is de aandacht voor het registreren van toetsinformatie toegenomen. Immers, het opslaan van toetsinformatie stelt de opleiding in staat om de ontwikkeling van competenties te volgen. Een specifiek probleem voor competentiegericht onderwijs is het vaststellen van groei in de competentie naar het gewenste eindniveau van de opleiding. Hoe verhoudt het competentieniveau van de student zich ten opzichte van de eindtermen van de opleiding? Mogelijke instrumenten voor het 'monitoren' en plannen van groei in competentie zijn onder andere skills management systemen, voortgangstoetsen en portfolio's.

Een skills management systeem is een gestructureerde opsomming van vaardigheden. Voor elk van de vaardigheden wordt aangegeven wat het beheersingsniveau van de student is. Het zo ontstane profiel kan vergeleken worden met een gewenst profiel van bijvoorbeeld een bepaalde functie. Binnen het hoger onderwijs kunnen skills management systemen worden gebruikt om de positie te beschrijven die een student op een bepaald moment in de studie heeft ten opzichte van de eindtermen. De informatie over het beheersingsniveau van de student kan uit verschillende beoordelingsbronnen afkomstig zijn. Dit kan bijvoorbeeld een combinatie zijn van beoordelingen van medestudenten, de studenten zelf, docenten of, ingeval van stages, externe opdrachtgevers of de chef van de student. Student en competentie-manager/docent kunnen aan de hand van deze informatie de studievoortgang bespreken. Kennis van zwak ontwikkelde vaardigheden kan worden gebruikt om nieuwe leerervaringen te plannen. Kennis van sterk ontwikkelde vaardigheden kan dienen als uitgangspunt voor specialisatie. Een systeem van voortgangstoetsen met gesloten vragen is een bijzondere vorm van een skills management systeem waarmee men de ontwikkeling van de materiedeskundigheid in kaart brengt. In het PGO-onderwijs bestaat een lange traditie van voortgangstoetsen. Portfolio's zijn verzamelingen producten (bijvoorbeeld ontwerpen, papers, onderzoeksverslagen) waaruit blijkt dat een student bepaalde vaardigheden

beheerst (showcase-portfolio), eventueel aangevuld met beschrijvingen van de eigen ontwikkeling. De student heeft zelf de hand in het bepalen van de inhoud van het portfolio. De inhoud van het portfolio illustreert de competentiegroei van de student, geeft inzicht in de prestaties die geleverd zijn en dient tot reflectie op de kwalitatieve voortgang. Skills management systemen en portfolio assessment sluiten op elkaar aan. Skills management systemen beschrijven de positie die een student op een bepaald moment in de studie heeft ten opzichte van de eindtermen. Daarmee kan het leren binnen het curriculum 'gemonitord' en gestuurd worden. Met portfolio's kan het gesprek over de beheersing van vaardigheden worden verdiept.

In regelmatig te houden functioneringsgesprekken met een functionaris die men de competentie-manager zou kunnen noemen, wordt gereflecteerd over het functioneren in het verleden, de ambities voor de toekomst en worden eventueel afspraken gemaakt die betrekking hebben op versterking van de vaardigheden. Input voor dit gesprek is informatie over de leerresultaten. Deze informatie is beschikbaar in een skills management systeem en in portfolio's. De output zou een activiteitenplan kunnen zijn met individuele taken voor een volgende onderwijsperiode of aan het einde van een opleiding een persoonlijk ontwikkelplan waarin gereflecteerd wordt op toekomstige beroepsmatige activiteiten in relatie tot verdere persoonlijke ontwikkeling.

XII. Index

- analyseren, X-1, XI-1, XI-5
- antwoordmodel, I-2, I-4, I-5, II-2, IV-1, IV-4, V-1, V-2, V-3, V-5, V-6, V-7, V-12, V-13, V-14, V-15, VI-1, VI-2, VI-3, VI-4, VI-5, VI-6, VI-7, VI-8, VII-7, VII-9, VII-10, VII-11, VIII-1, VIII-5
- authenticiteit, XI-1, XI-7
- beoordelaarsbetrouwbaarheid, XI-1
- beoordelaarsinstructie, VII-7
- beoordelingscriteria, V-12, V-13, V-15, VI-8, IX-9, XI-11
- beoordelingsschaal, XI-13
- beoordelingsschema, IX-8, X-1, XI-11
- beroepspraktijk, XI-5
- betrouwbaarheidscoëfficiënt, VII-3
- bijzondere verplichting, I-1, I-5, III-1, III-2, **III-6**, IX-1, X-1, X-2
- certificeren, XI-2
- cesuur, VI-1, VI-4, VI-5, VII-1, VII-2, VII-11, VIII-5
- checklist, IV-7, IV-15, IV-16, V-8, V-14, VI-7, IX-8, XI-13
- cijfer, IX-7, IX-8, IX-9, XI-11, XI-12
- cijferschaal, VII-11, XI-11
- classificeren, XI-5
- competenties, XI-1, XI-5, XI-13
- complexe vaardigheden, XI-9, XI-14
- complexiteit, III-6, XI-6
- construct, VII-3
- correctievoorschrift, VI-1, VI-7, VI-8, VII-7
- correlatie, VII-2, VII-9, VII-10
- correlatiecoëfficiënt, VII-6, VII-9
- eindtermen, XI-1, XI-5, XI-13
- evalueren, II-1
- examencommissie, XI-2
- feedback, I-2, I-4, IV-17, IV-20, IV-21, IV-23, V-16, VIII-1, IX-7, IX-9, XI-1, XI-11
- formatieve toetsen, I-1
- gemiddelde score, VII-6, VII-7
- gesloten vragen, I-1, I-3, I-4, III-1, *III-6*, III-7, IV-1, IV-8, IV-9, IV-15, IV-16, IV-17, V-5, VI-1, VI-2, VI-3, VI-5, VI-6, VI-7, VII-1, VII-3, VII-6, VII-7, VII-8, VII-9, VIII-4, IX-1, IX-2, XI-14
- Gesloten vragen, I-3, I-4
- grensscore, VI-5
- herkansen, VIII-4
- inhoudsvaliditeit, I-4, I-5, XI-7
- interne consistentie, VII-2
- korte toetsen, I-1
- KR-20, VII-3
- meerkeuzevragen, I-3, III-6, IV-9, IV-11, IV-14, IV-15
- meetfout, VII-2
- modelantwoord, V-13, V-15
- mondeling tentamen, II-2, III-2, IX-1, IX-2, IX-3, IX-5, IX-7, IX-9, IX-10, IX-14, X-1
- normen, VII-6, VII-8, VII-10
- objectieve beoordeling, VI-5
- objectiviteit, I-2, IV-1, IV-4, IV-7, V-1, V-7, IX-3, IX-7, XI-11, XI-12
- onderwijs, I-1, XI-1, XI-2, XI-5, XI-7, XI-9, XI-11, XI-13
- open vragen, I-1, I-3, I-4, I-5, II-2, III-1, *III-6*, III-7, V-1, V-2, V-5, V-7, V-8, V-11, V-13, V-14, V-15, V-16, VI-2, VI-3, VI-5, VI-7, VII-1, VII-6, VII-7, VII-8, IX-4, IX-13, X-1, XI-10
- Open vragen*, I-3, I-4, V-6, VI-6
- prestatieniveau, XI-13
- Produceervragen*, *III-5*
- productschaal, XI-13

raadkans, III-6, IV-12, VI-4, VI-5
reproductie, I-4, III-4, III-6, III-7, V-1, V-2
score-cijfertransformatie, IX-8
scoreschaal, VI-3, VII-8
scoringsregel, VI-2, VI-5
scoringsvoorschrift, VI-4, VI-7, VII-7
signifisch effect, XI-13
simulaties, XI-7
spreiding, VIII-2, IX-7, IX-10
steekproef, I-3, VII-2, IX-3
stellingvragen, IV-9, IV-15, IV-23

studietoetsen, VII-3
summatief, I-1
Tentamenvorm, III-6
testconstructie, IV-6
toepassen, I-5, III-5, IX-1
toetsconstructie, XI-2
toetsmatrijs, III-1, III-3, III-4, III-5, III-6, III-7, VI-2, VI-7, VIII-1, VIII-2, VIII-4, VIII-5, IX-4
toetsvorm, III-1, IV-17, IX-1, XI-9
variantie, IV-9
werkprocessen, XI-5