



Toetsbeleid Pedagogische Wetenschappen en Onderwijswetenschappen

Inhoudsopgave

Voorwoord	3
Deel 1. Toetsbeleid	4
1.1 Doelen	4
1.2 Funderende principes	4
1.2.1 Uitgangspunten	4
1.2.2 Functies	6
1.2.3 Kwaliteitscriteria	6
1.3 Systeem van toetsing en beoordeling	7
1.3.1 Stelregels op curriculumniveau	7
1.3.2 Stelregels op cursusniveau	9
1.4 Verantwoordelijkheden	11
Deel 2. Implementatierichtlijnen	13
2.1 De toetscyclus	13
2.2 Richtlijnen voor toetsontwerp	13
2.3 Richtlijnen voor toetsconstructie	16
2.3.1 Tentamens met gesloten vragen	16
2.3.2 Tentamens met open vragen	17
2.3.2 Verslagen, werkstukken, praktische opdrachten en portfolio's	18
2.4 Richtlijnen voor toetsbeoordeling	19
2.4.1 Tentamens met gesloten vragen	19
2.4.1 Tentamens met open vragen	21
2.4.3 Verslagen, werkstukken, praktische opdrachten en portfolio's	22
2.5 Richtlijnen voor toetsevaluatie	22
Bronvermeldingen	24

Voorwoord

Dit rapport beschrijft het toetsbeleid van het cluster Pedagogische Wetenschappen en Onderwijswetenschappen (PWO) van het Onderwijsinstituut Faculteit Sociale Wetenschappen. Onder dit cluster vallen de volgende opleidingen:

- Bachelor Pedagogische Wetenschappen (inclusief het ALPO-traject)
- Bachelor Pedagogische Wetenschappen van Primair Onderwijs
- Master Pedagogische Wetenschappen
- Master Onderwijswetenschappen
- Master Curriculumontwikkeling voor Primair Onderwijs
- Premaster Pedagogische Wetenschappen
- Premaster Onderwijswetenschappen

Voor al deze opleidingen geeft het toetsbeleid een algemeen kader voor de systematiek van toetsing en beoordeling. Dit kader is richtinggevend voor de Onderwijs en Examen Regeling (OER) en de hieraan gekoppelde Regels & Richtlijnen (R&R) voor de examencommissie.

Deze versie van het toetsbeleid vervangt de versie uit 2012. Na 10 jaar was het hoog tijd om de tussentijdse verbeteringen en aanscherpingen die de opleidingen sinds 2012 hebben doorgevoerd in het toetsbeleid op te nemen. Het uiteindelijke resultaat is geïnspireerd op het toetsbeleid van de opleidingen Psychologie en Managementwetenschappen. Bij de totstandkoming van deze nieuwe versie van het toetsbeleid heeft de examencommissie een voortrekkersrol vervuld en tijdens het schrijven dankbaar gebruik gemaakt van de waardevolle suggesties van de opleidingscommissies en de PWO-directie.

Dit rapport kent een trechterstructuur. Na een beschrijving van de algemene doelen wordt ingegaan op de uitgangspunten van het toetsbeleid, de functies van toetsing in het hoger onderwijs en de kwaliteitscriteria waaraan goede toetsen moeten voldoen. Deze funderende principes worden verder geconcretiseerd in stelregels voor toetsing en beoordeling voor een gehele opleiding én de afzonderlijke cursussen binnen die opleiding. Bovendien wordt beschreven wie verantwoordelijk is voor de realisatie en borging van deze stelregels. In deel 2 worden de stelregels vertaald naar implementatierichtlijnen voor toetsing binnen cursussen.

Deel 1. Toetsbeleid

1.1 Doelen

Dit toetsbeleid vormt de ruggengraat van het systeem van toetsing en beoordeling van de PWO-opleidingen. Het doel van het toetsbeleid is vierledig en geeft:

1. de *opleidingen* een beschrijving van de gewenste systematiek van toetsing en beoordeling waarmee zij ervoor kunnen zorgen dat studenten de eindkwalificaties behalen;
2. *studenten* inzicht in de manier waarop hun leerprestaties worden getoetst zodat zij hun studeergedrag hierop kunnen afstemmen;
3. *examinatoren en docenten* een leidraad om weloverwogen beslissingen te nemen over de toetsing en beoordeling in hun cursussen; en
4. de *Examencommissie* een kader voor evaluatie en eventuele bijsturing van het toetsprogramma.

Dit toetsbeleid is niet het enige opleidingsdocument waarin aandacht wordt besteed aan toetsing en beoordeling. Het toetsbeleid moet dan ook in samenhang worden gezien met de opleidingsvisies van de PWO-opleidingen, het Handboek kwaliteitszorg onderwijs van de faculteit FSW, de onderwijsvisie van de Radboud Universiteit, de Onderwijs en Examen Regeling (OER) van de opleidingen en de hieraan gekoppelde Regels & Richtlijnen (R&R) van de Examencommissie PWO.

1.2 Funderende principes

In deze paragraaf worden de grondslagen van het PWO-toetsbeleid beschreven. Achtereenvolgens wordt ingegaan op de uitgangspunten van toetsing, de verschillende functies van toetsen in het wetenschappelijk onderwijs en de kwaliteitscriteria waaraan toetsen moeten voldoen.

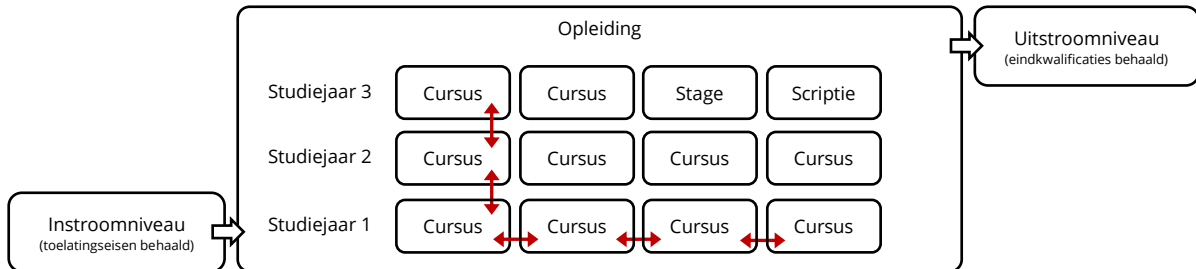
1.2.1 Uitgangspunten

De uitgangspunten van dit toetsbeleid zijn ontleend aan het model van *constructive alignment* (Biggs & Tang, 2011), dat veelvuldig wordt gebruikt in het hoger onderwijs. De centrale gedachte van dit model is dat het curriculum van een opleiding moet zijn afgestemd op de eindkwalificaties die studenten moeten bereiken. Dit betekent dat de leeractiviteiten, de leeruitkomsten en de manier waarop die worden getoetst, met elkaar in overeenstemming moeten zijn. Vanuit deze basisgedachte zijn vier uitgangspunten voor het PWO-toetsbeleid geformuleerd.

1. De componenten van het curriculum zijn op elkaar afgestemd

Een goed ontworpen curriculum wordt gekenmerkt door een logische samenhang tussen de onderwijsonderdelen. Dit betreft zowel de samenhang tussen onderdelen binnen een studiejaar (horizontale afstemming) als tussen verschillende studiejaar (verticale afstemming). Beide vormen worden geïllustreerd in Figuur 1.

Figuur 1. Horizontale (↔) en verticale (↑↓) afstemming van de toetsing binnen een opleiding



Een voorbeeld van horizontale afstemming is wanneer de inhoud van alle statistiekcursussen in een studiejaar op elkaar voortbouwen, zonder overlap en hiaten. Van verticale afstemming is bijvoorbeeld sprake bij een leerlijn beroepsvaardigheden: de vaardigheden die in het eerste studiejaar worden geoefend bereiden studenten voor op de beroepsvaardigheidencursus in het tweede jaar, die op zijn beurt voorbereidt op de cursus in het derde jaar. Bij verticale afstemming kan zowel sprake zijn van verbreding (elk jaar komen nieuwe, vaak complexere vaardigheden aan bod) als verdieping (in het tweede studiejaar worden de vaardigheden uit het eerste jaar opnieuw geoefend, maar in moeilijkere situaties). In dit laatste geval is sprake van een spiraalcurriculum.

In dit toetsbeleid heeft verticale afstemming alleen betrekking op de samenhang *binnen* een opleiding. Hoewel bacheloropleidingen voorbereiden op en toegang geven tot masteropleidingen, hebben zij hun eigen eindkwalificaties. Afstemming over opleidingen heen is weliswaar wenselijk, maar geen vereiste voor een goed toetsprogramma.

2. De componenten van een cursus zijn op elkaar afgestemd én op de eindkwalificaties

Wat geldt voor het gehele curriculum, geldt ook voor afzonderlijke cursussen: de leerdoelen, de leerstof, de leeractiviteiten en de toetsing moeten op een goed doordachte manier met elkaar samenhangen. Bovendien moeten de leerdoelen bijdragen aan het bereiken van de eindkwalificaties. Voor het ontwerpen van een cursus betekent dit dat de leerdoelen moeten worden afgeleid uit de eindkwalificaties van de opleiding. De leerdoelen moeten tevens op een heldere en meetbare manier worden geformuleerd; dit heeft voordelen voor studenten (zie uitgangspunt 3) en stelt docenten in staat om een passende toetsvorm te kiezen, uiteraard binnen de bestaande mogelijkheden (zie uitgangspunt 4).

3. Leerdoelen en toetsing beïnvloeden wat en hoe studenten leren

Iedere docent krijgt wel eens de vraag: "moeten we dit ook weten voor het tentamen?". Hoewel deze vraag suggereert dat zelfs universitaire studenten soms lijken te leren om studiepunten te halen in plaats van zichzelf te willen ontwikkelen en/of voor te bereiden op een toekomstige loopbaan, kan het zeker in de eerste studiejaar ook een signaal zijn dat studenten behoefte aan duidelijkheid hebben. Dit is een extra reden om de leerdoelen en toetsing goed op elkaar af te stemmen zodat studenten uit de leerdoelen kunnen afleiden wat ze op het tentamen voorgeschoteld krijgen. Dus als een cursus bijvoorbeeld is bedoeld om studenten inzicht bij te brengen, dan moet dit duidelijk tot uiting komen in de leerdoelen, de oefententamenvragen en de uiteindelijke toetsing.

4. Het toetsprogramma is tijdsefficiënt en kosteneffectief

Dit laatste uitgangspunt betreft de haalbaarheid of uitvoerbaarheid van de toetsing. In een ideale wereld waarin tijd en geld geen rol spelen, is het eenvoudig voor een opleiding om een volledige en elegante samenhang te creëren tussen de componenten uit het model van Biggs en Tang (2011). In de praktijk is dit echter niet mogelijk en daarom moet bij het ontwerpen van een toetsprogramma rekening worden gehouden met de randvoorwaarden waarbinnen een opleiding moet opereren. Dit betekent dat weloverwogen keuzes moeten worden gemaakt over de manier van toetsing, waarbij docenten soms genoeg moeten nemen met

een beperktere of minder passende toetsvorm en de opleiding soms extra middelen beschikbaar moet stellen om adequate toetsing van leerdoelen en eindkwalificaties mogelijk te maken.

1.2.2 Functies

In het onderwijs kunnen toetsen voor verschillende doeleinden worden gebruikt. Sanders (2010) onderscheidt vijf functies van toetsen, die worden beschreven in Tabel 1.

Tabel 1. Functies van toetsen in het hoger onderwijs

Functie	Omschrijving	Voorbeelden
Selecteren	Studenten tot een (deel van een) opleiding toelaten	Studenten moeten tot de beste 10% van hun lichter behoren om toegelaten te worden tot het honoursprogramma Studenten moeten alle B1 cursussen hebben behaald om met de afstudeerfase te mogen beginnen Studenten moeten de cursus Statistiek 1 hebben gehaald om te mogen deelnemen aan Statistiek 2
Plaatsen	Studenten tot een variant van een opleiding toelaten	Studenten die de Engelstalige mastervariant willen volgen, moeten een IELTS score van minimaal 6.0 hebben
Certificeren	Studenten een diploma geven	Studenten die de bacheloropleiding hebben afgerond, krijgen een BSc. diploma Studenten die de masteropleiding hebben afgerond, kunnen zich laten inschrijven in het beroepsregister
Monitoren/ diagnosticeren	De studievoortgang of ontwikkeling in kaart brengen	Studenten krijgen feedback op een conceptverslag Studenten krijgen verbeterpunten van hun stagebegeleider
Onderwijskwaliteit verbeteren	Een onderwijsonderdeel verbeteren	De opleiding of docent gebruikt de toetsresultaten om te bepalen welke cursusonderdelen moeten worden aangepast

Bij de eerste drie functies is sprake van summatieve toetsing: de resultaten worden gebruikt om het eindniveau te bepalen. De laatste twee functies zijn formatief van aard en bedoeld om tussentijds inzicht te krijgen in het niveau en verbetermogelijkheden te signaleren.

1.2.3 Kwaliteitscriteria

Om met zekerheid te kunnen zeggen dat een student die slaagt voor een toets ook daadwerkelijk voldoet aan de leerdoelen (en vice versa), moet de toetsing voldoen aan drie kwaliteitseisen.

1. Validiteit

Bij een goede toets weerspiegelen de resultaten de mate waarin een student de cursusdoelen heeft behaald (inhoudsvaliditeit). Dit betekent dat alle cursusdoelen naar rato moeten zijn vertegenwoordigd in de toetsing en de beoordeling alleen is gebaseerd op de vakspecifieke kennis en vaardigheden die tot de cursusinhoud behoren.

2. Betrouwbaarheid

In de psychometrische literatuur wordt het begrip betrouwbaarheid op verschillende manieren gedefinieerd: als de overeenstemming tussen toetsresultaten bij een herhaalde meting/beoordeling of als interne consistentie van de scores op vragen over hetzelfde onderwerp. Omdat een herhaalde afname van een toets in het onderwijs nauwelijks voorkomt (studenten die zijn geslaagd doen doorgaans geen hertentamen), richt het toetsbeleid zich op de betrouwbaarheid van de *beoordeling* van de toetsing. Bij tentamens met open vragen, verslagen of opdrachten die door verschillende docenten worden beoordeeld, kan de interbeoordelaars betrouwbaarheid worden gewaarborgd door een antwoordmodel of *rubric* te gebruiken. Bij tentamens met meerkeuzevragen heeft dit meestal weinig meerwaarde omdat van tevoren (en na collegiaal overleg) vaststaat wat het goede antwoord is.

Voor tentamens geldt bovendien het criterium van interne consistentie. Dit betekent dat een student alle tentamenvragen over hetzelfde onderwerp even goed (of even slecht) moet hebben gemaakt; het betreft dus een soort interne herhaalde meting. Hieruit valt af te leiden dat interne consistentie een irrelevant criterium is voor de psychometrische kwaliteit van toetsvragen over *verschillende* onderwerpen (Barchard, 2010; Peters, 2014).

3. Transparantie

De toetsing is transparant als studenten van tevoren weten hoe en waarop zij worden beoordeeld, welke beoordelingscriteria hiervoor worden gebruikt, en hoe de toetsresultaten inzicht geven in hun studievoortgang en eigen professionele ontwikkeling.

1.3 Systeem van toetsing en beoordeling

In deze paragraaf worden de funderende principes vertaald naar stelregels voor toetsing en beoordeling binnen de PWO-opleidingen. Deze stelregels zijn sturend maar geven docenten/ examinatoren wel voldoende ruimte om, indien de situatie hierom vraagt, afwijkende keuzes te maken—op voorwaarde dat wordt voldaan aan de funderende principes.

De stelregels zijn onderverdeeld naar curriculumniveau en cursusniveau. De stelregels op curriculumniveau zijn met name bedoeld voor opleidings- en leerlijncoördinatoren. De stelregels op cursusniveau bouwen hierop voort en zijn vooral bedoeld voor cursuscoördinatoren, docenten en examinatoren.

1.3.1 Stelregels op curriculumniveau

1. Het streefniveau van elk studiejaar is vastgesteld

Het curriculum heeft een duidelijke opbouw waarin wordt toegewerkt naar de eindkwalificaties. Bij meerjarige opleidingen kan hiertoe een grofmazige indeling worden gemaakt van het academisch niveau per studiejaar. De B1-fase is dan bijvoorbeeld inleidend, de B2 verbredend, en de B3 verdiepend. Op masterniveau is dan sprake van specialisatie. Een dergelijke indeling is ook mogelijk binnen eenjarige opleidingen. Zo kan het eerste semester van een masteropleiding theorievormend zijn en het tweede semester zich richten op de toepassing en van de geleerde theorieën in onderzoek en de beroepspraktijk.

Vanuit de gedachte van *constructive alignment* sluit het toetsprogramma van een opleiding aan bij deze niveauopbouw. Als het eerste studiejaar inleidend is met de nadruk op de ontwikkeling van basiskennis en -vaardigheden, dan moet de toetsing aantonen wat studenten kennen en kunnen. In de daaropvolgende studiejaar stijgt het academisch niveau waardoor de focus van de toetsing verschuift naar complexere verrichtingen zoals analyseren, evalueren en creëren (Anderson & Krathwohl, 2001; zie verder §2.2).

2. Elk studiejaar heeft een passende combinatie van toetsvormen

In het verlengde van de vorige stelregel geldt dat de toetsvormen moeten zijn afgestemd op het streefniveau van een studiejaar. Hierdoor kent de B1-fase bijvoorbeeld een andere combinatie van toetsvormen dan de B3-fase, met meer tentamens om te bepalen of studenten de basisbegrippen van het vakgebied kennen, en minder opdrachten en verslagen om de toepassing van die begrippen in authentieke situaties te beoordelen. Wanneer praktische randvoorwaarden hiertoe aanleiding geven (zie uitgangspunt 4), kan ook in de B3-fase met tentamens worden gewerkt, zij het met een ander karakter (bijvoorbeeld een open-boek tentamen), andersoortige vragen (bijvoorbeeld een casus beoordelen) of een andere weging van kennis- en toepassingsvragen.

Elk studiejaar dient daarnaast voldoende variatie in toetsvormen te bieden. Deze afwisseling is motiverend voor studenten en stelt hen in staat om op verschillende manieren te demonstreren dat zij over de gewenste kennis en vaardigheden beschikken. Dit geeft de opleiding een beter beeld van hun werkelijke capaciteiten.

3. De toetsing is evenwichtig verspreid over het studiejaar

Deze stelregel heeft betrekking op de studeerbaarheid van de opleiding. Allereerst moeten de cursussen goed zijn verdeeld over de onderwijsperiodes om piekbelasting te voorkomen. Op eenzelfde manier moet ook de toetsing van die cursussen zo goed mogelijk zijn verspreid over het studiejaar zodat studenten voldoende voorbereidingstijd hebben. In een onderwijssysteem met twee semesters van elk twee periodes wordt de toetsing sowieso verspreid over vier tentamenperiodes. Maar spreiding binnen die periodes kan soms lastig zijn, zeker voor verweven studenten, en de inlevermomenten voor verslagen kunnen de studiebelasting verder opvoeren. Daarom moet per opleiding en per studiejaar worden bekeken hoe de toetsing zo evenwichtig mogelijk kan worden geroosterd. Hierbij kan op het niveau van individuele cursussen worden besloten om met tussentijdse deeltolsten te werken of de opdrachten of verslagen vóór de tentamenperiodes in te leveren.

4. Het toetsprogramma van elk studiejaar is gebaseerd op een kosten-baten analyse

De ene toetsvorm is tijdsintensiever en dus duurder dan de andere. Bij het kiezen van het toetsmoment en de toetsvorm moeten deze kosten worden gezien in het licht van de baten: weegt de benodigde investering (tijd, geld) op tegen de informatie die met de toets wordt verkregen? Toetsing van welke kennis, vaardigheden en attitudes is ons op welk moment in de opleiding het meest waard? Voor het bewaken van een doelmatige inzet van middelen is het onder andere van belang om:

- Nauwkeurig te bepalen hoeveel tijd is gemoeid met een bepaalde toetsvorm. Een opdracht bedenken voor het schrijven van een essay kost veel minder tijd dan het samenstellen van een tentamen met meerkeuzevragen, maar bij de nakijktijd ligt dit precies andersom.
- Na te gaan of docenttijd op een slimme en creatieve manier kan worden verschoven van de ene naar de andere onderwijstaak. Als tijdens een cursus bijvoorbeeld wordt gewerkt met formatieve *peer feedback*, dan kan de docent meer tijd besteden aan de summatieve beoordeling van het eindverslag.
- De meeste tijd en het meeste geld te investeren in de toetsmomenten die voor het behalen en beoordelen van de eindkwalificaties het meest belangrijk zijn.
- De inzet van tijd anders te spreiden over de formatieve en summatieve toetsing, bijvoorbeeld door meer tijd te besteden aan feedback tijdens de cursus dan aan inzage na de cursus.

5. In elk studiejaar wordt effectief gebruik gemaakt van formatieve toetsing

In elk studiejaar moet ruimte zijn voor formatieve toetsing. Idealiter ontvangt elke student tijdens elke cursus minimaal één keer tussentijdse feedback om zijn/haar leerprestaties verder te verbeteren. Deze feedback is echter alleen effectief als de formatieve toetsing een goede afspiegeling vormt van de summatieve toetsing. Omdat feedback geven tijdsintensief is, is het wellicht niet mogelijk om in alle programmaonderdelen formatief te toetsen. In die situaties geldt dat de formatieve feedbackmomenten goed moeten worden gespreid over het studiejaar.

Ter aanvulling geldt dat de formatieve toetsing evenwichtig over de studie jaren moet zijn verdeeld. Wel kan de aard van de formatieve toetsing verschuiven. Bij dit laatste kan gedacht worden aan de geleidelijke invoering van peer feedback, waardoor de docent in de latere studie jaren een steeds minder actieve of leidende rol speelt en studenten steeds vaker elkaar van feedback voorzien.

1.3.2 Stelregels op cursusniveau

6. In elke cursus wordt op een valide manier getoetst

Een toets is valide als het een representatieve afspiegeling geeft van de leerdoelen en de leerstof. Daarnaast wordt de validiteit van een toets bepaald door de mate waarin wordt bijgedragen aan het realiseren van de eindkwalificaties van een opleiding. Wanneer geen of zeer weinig leerdoelen een directe relatie met de eindkwalificaties hebben, kan dit aanleiding zijn om de leerdoelen en cursusinhoud opnieuw te bezien. Bij het ontwerpen van een toets moet een docent/examinator ervoor zorgen dat alle leerdoelen zijn afgedekt in de mate waarin zij bijdragen aan de leeruitkomsten van de cursus. Hiervoor kan tijdens het ontwerpproces gebruik worden gemaakt van een toetsmatrix (of een soortgelijk hulpmiddel; zie §2.2).

7. Elke cursus heeft verschillende toetsmomenten en toetsvormen

Herhaaldelijk en op uiteenlopende manieren toetsen vergroot de betrouwbaarheid van de toetsing. Indien mogelijk heeft elke cursus daarom tenminste twee toetsmomenten en -vormen. Dit kan bijvoorbeeld gerealiseerd worden door halverwege de cursus een deeltentamen met meerkeuzevragen af te nemen en aan het einde van de cursus een deeltentamen met open vragen. Afhankelijk van de aard van de leerdoelen kan het tweede deeltentamen worden vervangen door bijvoorbeeld een opdracht.

In de cursushandleiding moet duidelijk zijn beschreven hoe zwaar elk toetsonderdeel meetelt voor het eindcijfer. Deze weging dient in overeenstemming te zijn met het belang van de leerdoelen die in het onderdeel worden getoetst en de tijd die studenten hieraan volgens de studieplanning zouden moeten besteden.

8. Toetsen worden jaarlijks veranderd om fraude te voorkomen

Tentamenvragen, casussen en opdrachten worden tegenwoordig razendsnel via sociale media gedeeld. Een tentamen, casus of opdracht in identieke vorm hergebruiken is daarom niet wenselijk. Bij tentamens dient jaarlijks een substantieel deel van de vragen te worden vervangen of aangepast; bij casussen en opdrachten gebeurt dit idealiter ook, hoewel het aantal mogelijk variaties hierbij vaak minder groot zal zijn. In plaats van steeds iets nieuws te bedenken, kunnen docenten eerder gebruikte vragen, casussen en opdrachten aanpassen, bijvoorbeeld door het databestand bij een statistiekopdracht te veranderen, een oude vraag over een nieuwe casus te stellen, of twee van de vier alternatieven bij een meerkeuzevraag te veranderen.

9. Studenten worden voorafgaand aan de cursus geïnformeerd over de toetsvorm en weging

Het eerder genoemde kwaliteitscriterium 'transparantie' stelt dat studenten van tevoren moeten weten waar ze aan toe zijn. In de online studiegids moet daarom vermeld staan welke toetsvorm(en) bij een cursus worden gebruikt, of er deelttoetsen zijn en zo ja, hoe de deelscores worden gewogen in het eindcijfer. In geval van werkstukken, stages en scripties dienen ook de beoordelingscriteria voor aanvang van de cursus bekend te worden gemaakt. Dit kan via de cursushandleiding die minimaal één week voor aanvang van de cursus via de online leeromgeving beschikbaar dient te zijn.

10. Een groepsbeoordeling weerspiegelt de bijdrage van alle individuele groepsleden

Samenwerken is belangrijk in het onderwijs en de latere beroepspraktijk. Door tijdens cursussen in kleine groepen samen aan een opdracht te werken, kunnen studenten hiermee oefenen. In veel gevallen wordt het groepsproduct beoordeeld om te bepalen of alle groepsleden de leerdoelen hebben bereikt en krijgen alle groepsleden hetzelfde eindcijfer. Dit is alleen te rechtvaardigen als de groepsleden in gelijke mate hebben bijgedragen aan het groepsproces en het eindproduct (wat overigens niet wil zeggen dat ze hetzelfde moeten

hebben gedaan of precies even veel tijd aan de opdracht moeten hebben besteed). Als een docent hierover twijfelt, dan kan via aanvullende toetsing worden overgestapt op een individuele beoordeling. Docenten kunnen bijvoorbeeld een reflectiegesprek met individuele groepsleden voeren, hen vragen om elkaars inbreng te beoordelen, of te laten aangeven wat hun eigen bijdrage is geweest aan de onderdelen van het eindproduct. Studenten moeten voor aanvang van de cursus worden geïnformeerd over deze mogelijke maatregelen, bij voorkeur via de cursushandleiding.

11. Bij open vragen en werkstukken wordt een transparant beoordelingsmodel gebruikt

Een beoordelingsmodel heeft diverse voordelen: het maakt inzichtelijk waarop en hoe studenten worden beoordeeld zodat zij hun studiegedrag hierop kunnen afstemmen, het vergroot de kans op interbeoordelaars overeenstemming en maakt inzichtelijk wat de kwaliteit van toetsvragen of toetsonderdelen is. Hoe het beoordelingsmodel er precies uit ziet hangt af van de toetsvorm. Bij meerkeuzetentamens is zo'n model eigenlijk niet meer dan een lijst van juiste antwoorden. Bij open vragen dient een antwoordsleutel met normeringsaanwijzingen te worden opgesteld en voor opdrachten en werkstukken een rubric (zie §2.3.2). In alle gevallen worden studenten geïnformeerd over de aard van het beoordelingsmodel en de criteria; correcte antwoorden en uitwerkingen worden alleen geïllustreerd in oefenvragen of voorbeeldverslagen.

12. Van tentamens met meerkeuzevragen wordt de psychometrische kwaliteit bepaald

Psychometrische analyse van meerkeuzetentamens is verplicht. De ondersteuningsdiensten leveren de examinerator hiertoe een zogeheten toetsrapport met eenvoudige toets- en itemkarakteristieken die zijn gebaseerd op de klassieke testtheorie. Op basis van deze indicatoren besluit de examinerator hoe om te gaan met meerkeuzevragen van onvoldoende kwaliteit. Aanwijzingen hiervoor staan in paragraaf 2.4.1.

13. Bij de beoordeling van tentamens met meerkeuzevragen wordt gecorrigeerd voor gokkans

Om de validiteit van toetsing te waarborgen, worden de scores van alle tentamens met meerkeuzevragen gecorrigeerd voor gokkans. Studenten worden hierover tijdig geïnformeerd, bij voorkeur via de cursushandleiding en tijdens het responsiecollege. Het verdient aanbeveling om studenten te wijzen op de consequenties van deze richtlijn: als zij een vraag niet weten kunnen ze beter gokken dan de vraag open laten omdat dit dubbel nadelig is.

14. Bij de beoordeling van tentamens wordt gebruik gemaakt van absolute normering

Absoluut normeren betekent dat de cijfers worden berekend op een van tevoren bedachte manier, ongeacht hoe goed of hoe slecht het tentamen is gemaakt. Deze werkwijze verdient de voorkeur boven relatief normeren waarbij van tevoren vaststaat dat een X-percentages studenten dient te slagen—dit is niet eerlijk als het tentamen bijvoorbeeld bijzonder goed is gemaakt, en leidt tot onterecht geslaagden wanneer het tentamen juist heel slecht is gemaakt. Absolute normering vereist dat de kwaliteit van de toets en moeilijkheidsgraad van de items van tevoren bekend is uit eerdere afnames bij een voldoende grote groep studenten. Omdat dit in de praktijk niet altijd mogelijk is, kunnen docenten besluiten de absolute normering enigszins bij te stellen. Suggesties hiervoor worden gegeven in paragraaf 2.4.

15. Toetsen en hun beoordelingsmodel worden onderworpen aan peer review

Tijdens het toetsontwerp wordt het vier-ogen principe toegepast. Dit vergroot de kwaliteit, validiteit en betrouwbaarheid van de toetsing. Omdat vrijwel alle cursussen bij PWO door minimaal twee docenten worden gegeven, kan meestal een direct betrokken collega worden ingeschakeld. Dit *peer review* systeem heeft ook betrekking op het beoordelingsmodel. Indien nodig kan ook de hulp van een toetsdeskundige van het Teaching and Learning Centre (TLC) worden ingeroepen.

16. Stages en praktijkwerkstukken worden beoordeeld door de begeleidend docent

Studieonderdelen die in of voor de praktijk worden uitgevoerd, worden beoordeeld door docenten die zijn verbonden aan de opleiding. Bij stages en praktijkwerkstukken is dit doorgaans de begeleidend docent. Indien

de cursusdoelen hiertoe aanleiding geven, kan de opdrachtgever of begeleider van de praktijkinstelling een adviserende stem hebben. Studenten worden voorafgaand aan hun stage of praktijkopdracht geïnformeerd over wie de beoordeling zal uitvoeren, volgens welke criteria, en hoe de verschillende onderdelen worden gewogen bij de bepaling van het eindcijfer. Deze informatie krijgen zij via de stage- of cursushandleiding, waarin ook het beoordelingsformulier moet zijn opgenomen. Bij stages hebben studenten bovendien recht op tenminste één formatieve beoordeling van het volledige stageverslag (de conceptversie); bij praktijkopdrachten verdient het aanbeveling om dit voorbeeld te volgen of in elk geval feedback te geven op tussenproducten en/of delen van het eindverslag.

17. Scripties worden beoordeeld door twee interne beoordelaars

Scriptieonderzoek kan zowel intern als extern worden uitgevoerd. Een externe scriptie mag worden uitgevoerd bij dezelfde instelling als waar de student stage heeft gelopen, op voorwaarde dat de scriptie een (grotendeels) op zichzelf staand onderzoek is. Stage en scriptie kunnen dus niet volledig worden gecombineerd. Tijdens het onderzoek geeft de scriptiebegeleider minimaal één formatieve beoordeling van de gehele scriptie. De eindbeoordeling wordt gedaan door de scriptiebegeleider en een docent van de opleiding die niet bij het onderzoek is betrokken (hierna: de tweede lezer). Bij een externe scriptie geldt, net als bij stages en praktijkopdrachten, dat de extern begeleider een adviserende stem in de eindbeoordeling kan hebben. Alle informatie over de beoordeling moet via de scriptiehandleiding bekend worden gemaakt.

Bij de eindbeoordeling gebruiken de scriptiebegeleider en de tweede lezer hetzelfde beoordelingsformulier. Zij bespreken hun oordelen om tot een eindcijfer te komen, waarbij de scriptiebegeleider een leidende rol heeft en de uiteindelijke beslissing neemt. Bij een verschil van 2 punten of meer wordt een onafhankelijke derde lezer ingeschakeld en wordt zijn/haar oordeel door de scriptiebegeleider meegenomen in de uiteindelijke beslissing over het cijfer. Als dit geen soelaas biedt, beslist de scriptiecoördinator of, indien hij/zij zelf bij de beoordeling is betrokken, de verantwoordelijk hoogleraar.

18. De toetsing wordt intern geëvalueerd

Na afloop van elk onderwijsonderdeel evalueert de examinator de kwaliteit van de toetsing. Indien deze evaluatie en/of de prestaties van studenten hiertoe aanleiding geven, worden voorstellen voor verbetering gedaan, die verder kunnen gaan dan de toetsing alleen (zie uitgangspunt 1). De resultaten van de evaluatie worden beschreven in een *teacher report*, waarin ook wordt ingaan op de resultaten van de studentevaluatie. Dit verslag wordt besproken in de Opleidingscommissie, die op haar beurt suggesties voor verdere verbetering kan geven.

1.4 Verantwoordelijkheden

Een goed functionerend systeem van toetsing en beoordeling vereist dat alle betrokken partijen hun verantwoordelijkheden kennen én nemen. Tabel 2 laat zien wie op welk niveau verantwoordelijk is voor de *constructive alignment* tussen toetsing en onderwijsinrichting, en daarmee voor de kwaliteit de toetsing en beoordeling.

Tabel 2. Verantwoordelijkheden voor de toetsing binnen de PWO opleidingen

Niveau	Uitvoering	Borging en controle	Advies
Opleiding	Opleidingsdirecteur	Examencommissie	Opleidingscommissie
Studiejaar/leerlijn	Jaar- of leerlijncoördinator	Examencommissie	Opleidingscommissie
Cursus	Cursuscoördinator	Examencommissie	Opleidingscommissie

De uitvoeringsverantwoordelijkheid op cursusniveau ligt bij de docent die de cursus coördineert. In de regel is dit ook de persoon die is aangewezen als examinator en zodoende verantwoordelijk is voor de validiteit,

betrouwbaarheid en transparantie van de toetsing en de beoordeling. In cursussen met twee of meer docenten heeft de cursuscoördinator een leidende rol en doorslaggevende stem bij de constructie van de toetsing en beoordeling. Ook zorgt de coördinator ervoor dat wordt voldaan aan de in paragraaf 1.3.2 genoemde stelregels op cursusniveau.

Een opleiding kan besluiten om een coördinator aan te wijzen die verantwoordelijk is voor de inhoud en organisatie van de onderwijsonderdelen in een bepaald studiejaar. Analoog hieraan kunnen ook coördinatoren worden aangesteld voor alle onderwijsonderdelen uit de opleiding over een bepaald onderwerp; denk bijvoorbeeld aan de coördinator van een doorlopende leerlijn beroepsvaardigheden. Deze coördinatoren moeten ervoor zorgen dat de onderwijsonderdelen in hun studiejaar/leerlijn voldoen aan de stelregels op curriculumniveau (zie §1.3.1). Voor een studiejaarcoördinator ligt het accent hierbij op de horizontale afstemming, voor een leerlijncoördinator op de verticale afstemming.

Op opleidingsniveau is de opleidingsdirecteur verantwoordelijk voor het beschrijven en jaarlijks actualiseren van de toetsprogramma's, in lijn met de stelregels op curriculumniveau. De opleidingsdirecteur kan zich desgewenst laten ondersteunen door de inhoudelijk verantwoordelijk hoogleraren, opleidingscoördinatoren en studieadviseurs. Bovendien zorgt de opleidingsdirecteur, in samenspraak met de Examencommissie, dat docenten over voldoende deskundigheid en middelen beschikken om de toetsing en beoordeling naar behoren uit te voeren, inclusief ondersteuning door een kwaliteitszorgteam of vergelijkbare ondersteuning.

De opleidingscommissie geeft gevraagd en ongevraagd advies over alle aspecten van onderwijskwaliteit. Hiertoe behoort ook de kwaliteit van toetsing en beoordeling. Jaarlijks bespreekt de opleidingscommissie in elk geval de teacher reports (zie stelregel 18) en het verslag van de screening van scripties.

De examencommissie controleert en bewaakt de kwaliteit van toetsing en beoordeling. De examencommissie doet dit door jaarlijks voor elk onderwijsonderdeel twee voldoende gekwalificeerde examinatoren aan te wijzen. Daarnaast stelt de examencommissie concrete richtlijnen op voor de uitvoering van het toetsbeleid (zie deel 2 van dit document). De examencommissie borgt de realisatie van het eindniveau van de opleiding door de toetsprogramma's te screenen aan de hand van de stelregels op curriculumniveau, en steekproefsgewijs toetsmateriaal en beoordelingen te screenen aan de hand van de stelregels op cursusniveau. Onder deze laatste taak valt ook de jaarlijkse screening van de eindwerkstukken van de opleiding. De examencommissie kan zich bij de uitvoering van deze taken laten ondersteunen door een toetscommissie.

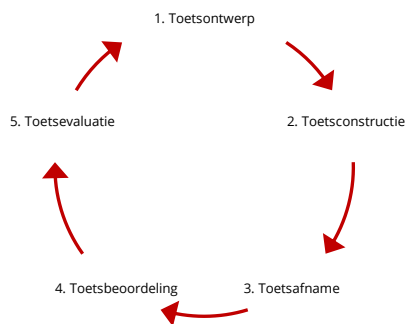
Deel 2. Implementatierichtlijnen

Deel 2 van het toetsbeleid geeft cursuscoördinatoren, examinatoren en docenten concrete handreikingen om de stelregels op cursusniveau toe te passen in hun eigen onderwijs. In paragraaf 2.1 wordt geïllustreerd hoe toetsing is verweven met (het ontwerpen van) een cursus. In de paragrafen daarna worden richtlijnen gegeven voor het ontwerpen, construeren, beoordelen en evalueren van diverse toetsvormen.

2.1 De toetscyclus

Toetsing loopt als rode draad door een cursus. Een bekwaam docent of examinator besteedt al bij de opzet van een cursus aandacht aan de toetsing, concretiseert de plannen uit het toetsontwerp tijdens de looptijd van de cursus en evalueert de kwaliteit van de toets nadat de beoordeling heeft plaatsgevonden. De evaluatieresultaten zijn het startpunt voor de toetsing tijdens het volgend studiejaar. Dit proces wordt geïllustreerd in Figuur 2.

Figuur 2. De toetscyclus



Het toetsontwerp gebeurt in de periode voor aanvang van de cursus. In deze eerste fase wordt besloten op welke manier de leerdoelen zullen worden getoetst. Dit vereist allereerst een analyse van de inhoud en het verrichtingsniveau van elk leerdoel (zie verder §2.2). Op basis hiervan wordt een passende toetsvorm gekozen en daarbij de vraagvorm(en) en/of het soort opdracht(en). In fase 2 wordt de toets geconstrueerd. Dit proces bestaat uit het formuleren van de toetsvragen of opdrachten en het opstellen van een antwoordsleutel en normeringsmodel. De toetsconstructie kan zowel voorafgaand als tijdens de cursus plaatsvinden. In deze fase moet de examinator bovendien de praktische voorbereidingen treffen zodat de toets in fase 3 op het geplande tijdstip kan worden afgenomen. In fase 4 wordt de psychometrische kwaliteit van de toetsvragen bepaald (voor opdrachten is dit vaak niet goed mogelijk) en besloten welke vragen buiten beschouwing moeten worden gelaten. Hierna worden de uitwerkingen van studenten beoordeeld volgens het antwoordmodel en wordt het behaalde aantal punten omgezet in een cijfer. Na afloop van de cursus wordt de toetsing geëvalueerd en worden voorstellen voor verbetering gedaan, die vanuit het idee van *constructive alignment* ook betrekking kunnen hebben op de leerdoelen en leerstof.

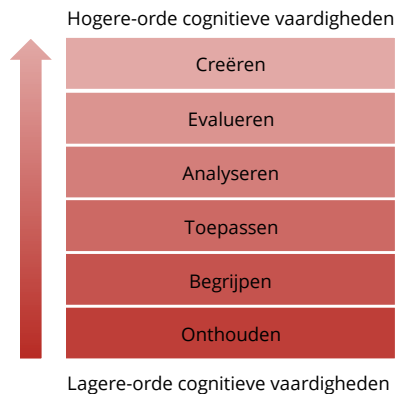
2.2 Richtlijnen voor toetsontwerp

Bij het ontwerpen van een toets maakt een docent/examinator verantwoorde keuzes over de inhoud en het verrichtingsniveau, de toetsvorm, en de vraagvorm of het soort opdracht. Voor deze onderdelen worden hieronder enkele richtlijnen gegeven.

1. Bepaal aan de hand van de leerdoelen de inhoud en het verrichtingsniveau van de toetsing

Leerdoelen beschrijven wat studenten aan het einde van een cursus moeten weten en kunnen, en vormen daarom een goed startpunt voor toetsontwerp. Als de leerdoelen op de juiste manier zijn geformuleerd, geven ze aan welke onderwerpen getoetst moeten worden en op welk verrichtingsniveau. Met dit laatste wordt bedoeld: op welke manier studenten moeten aantonen dat zij de leerstof/onderwerpen beheersen (zie §1.3.1, stelregel 1). Voor het bepalen van het verrichtingsniveau wordt vaak gebruik gemaakt van een taxonomie, zoals bijvoorbeeld die van Bloom (zie Figuur 3)

Figuur 3. Gereviseerde taxonomie van Bloom (Anderson & Krathwohl, 2001)



Neem bijvoorbeeld het leerdoel "Aan het einde van deze cursus ken je de vijf meest gebruikte nonparametrische testen". De *inhoud* van de toetsing is dan "de vijf meest gebruikte nonparametrische testen" en uit het cursusmateriaal kan eenvoudig worden afgeleid welke testen dit zijn. Het *verrichtingsniveau* is 'onthouden' omdat studenten deze testen moeten kennen; om dit aan te tonen moeten zij de testen kunnen aanduiden, opnoemen of opschrijven. Bij het leerdoel "Aan het einde van deze cursus kun je de vijf meest gebruikte nonparametrische toepassen" is de inhoud gelijk, maar het verrichtingsniveau hoger. Voor de toetsing betekent dit bijvoorbeeld dat studenten de formules van de vijf testen moeten kunnen gebruiken om een correcte berekening uit te voeren.

2. Bepaal de onderlinge verhouding van de leerdoelen in de toetsing

Niet elk leerdoel is even belangrijk en hiermee moet rekening worden gehouden in de toetsing. Om een evenredige verdeling te maken kan een toetsmatrix worden gebruikt. Een voorbeeld worden getoond in Figuur 3. Hieruit valt af te leiden een dergelijke matrix ook kan worden gebruikt voor werkstukken en opdrachten.

Op eenzelfde manier kan overigens worden aangetoond hoe de leerdoelen van de cursus aansluiten bij de eindkwalificaties van de opleiding. Omdat dit niet direct betrekking heeft op het toetsontwerp, wordt dit hier verder buiten beschouwing gelaten.

Figuur 3. Voorbeeld van een toetsmatrix

Leerdoel/ onderwerp	Verrichtingsniveau						Aantal vragen	Punten/ weging
	Onthouden	Begrijpen	Toepassen	Analyseren	Evalueren	Creëren		
1	1 vraag 5 punten	1 vraag 5 punten	2 vragen 10 punten				4	20
2		2 vragen 10 punten		1 vraag 10 punten			3	20
3	2 vragen 10 punten		2 vragen 15 punten				5	35
4	1 vraag 5 punten	1 vraag 10 punten		1 vraag 10 punten			3	25
Aantal vragen	4	5	4	2			15	100
Aantal punten	20	35	25	20				

3. Kies een toetsvorm die aansluit bij de inhoud en het verrichtingsniveau

In theorie kunnen vrijwel alle leerdoelen op verschillende manieren worden getoetst. In de praktijk zijn echter lang niet alle toetsvormen even geschikt voor elk leerdoel. In de voorbeelden bij richtlijn 1 lijkt het voor de hand te liggen om het eerste leerdoel met een tentamen te toetsen terwijl een opdracht meer geschikt is voor het tweede leerdoel. Om dit soort keuzes te kunnen maken, geeft Tabel 3 enkele concrete aanwijzingen.

In de literatuur worden naast tentamens en opdrachten nog tal van andere toetsvormen beschreven. Deze kunnen worden samengevat in vier hoofdcategorieën: tentamens, verslagen en werkstukken, praktische opdrachten/demonstraties, en portfolio's. Merk op dat in de OER uitsluitend wordt gesproken van tentamens en opdrachten; die laatste categorie omvat dan tevens de verslagen, werkstukken, portfolio's en alle andere toetsvormen waarin studenten hun kennis en vaardigheden op een toepassingsgerichte manier moeten laten zien.

Tabel 3. Handreiking voor het kiezen van een passende toets- en vraagvorm

Toets- en vraagvorm	Geschiktheid voor verrichtingsniveau	Praktische voordelen	Praktische nadelen
Tentamens met gesloten vragen ^a	Onthouden, begrijpen, toepassen	Korte nakijktijd Geschikt voor grote groepen studenten	Goede gesloten vragen formuleren is lastig
Tentamens met open vragen ^b	Onthouden, begrijpen, toepas- sen, analyseren, evalueren	Uitdagend voor studenten Antwoorden vereisen actieve reproductie	Nakijken kost veel tijd Minder geschikt voor grote groepen studenten
Verslagen en werkstukken	Toepassen, analyseren, evalueren, creëren	Uitdagend voor studenten Geschikt voor geïntegreerde toetsing van verschillende vaardigheden	Nakijken kost veel tijd Minder geschikt voor grote groepen studenten
Praktische opdrachten/ demonstraties	Toepassen, analyseren, evalueren, creëren	Authentieke manier van toetsing	Organisatie en afname zijn tijdrovend
Portfolio	Toepassen, analyseren, evalueren, creëren	Authentieke manier van toetsing	Beoordeling is niet per se betrouwbaar

^a Verzamelterm voor meerkeuzevragen, rangordevragen en selecteervragen.

^b Verzamelterm voor kort-antwoordvragen en essayvragen.

4. Kies een passende vraagvorm of opdracht

Als de toetsvorm bekend is, moet bij tentamens de vraagvorm worden gekozen; bij alle andere toetsvormen moet de aard van de opdracht worden bepaald. De uitwerking van deze richtlijn wordt in Tabel 4 geïllustreerd voor de vraagvorm bij tentamens. Tabel 3 geeft handreikingen voor het kiezen van een passende toets- en vraagvorm. Hierbij wordt niet alleen gekeken naar de aansluiting bij de aard van de leerdoelen, maar ook naar de kosten en baten (zie uitgangspunt 4).

Tabel 4. Enkele veelgebruikte vraagvormen voor tentamens

Vraagvorm	Omschrijving	Voorbeeld
Meerkeuzevragen	Studenten kiezen het juiste antwoord uit een beperkt aantal mogelijkheden	Een t-toets is een voorbeeld van een: A. Parametrische test B. Nonparametrische test C. Associatiemaat
Selecteervragen	Studenten kiezen een of meer juiste antwoorden uit een beperkt aantal mogelijkheden	Kruis aan voor welke doeleinden een correlatie kan worden gebruikt (meer dan 1 antwoord mogelijk). A. De samenhang tussen 2 variabelen beschrijven B. Het verschil tussen 2 groepen toetsen C. De effectgrootte bepalen
Rangordevragen	Studenten zetten de gegeven antwoordmogelijkheden in de juiste volgorde	Rangschik de soorten onderzoek van sterk naar zwak empirisch bewijs. A. Case studies B. Experimenten C. Survey-onderzoek
Kort-antwoordvragen	Studenten beantwoorden een open vraag in enkele woorden of één zin	Noem twee voorwaarden voor het uitvoeren van een ANOVA.
Essayvragen	Studenten beantwoorden een open vraag in enkele zinnen	Leg aan de hand van een voorbeeld uit wanneer een Bonferroni correctie nodig is.
Opdrachten	Studenten laten zien dat ze iets kunnen	Bereken met de gegevens uit de onderstaande tabel de samenhang tussen lengte en lichaamsgewicht.

2.3. Richtlijnen voor toetsconstructie

Omdat de richtlijnen voor toetsconstructie deels afhankelijk zijn van de gekozen toetsvorm, worden ze hieronder apart besproken.

2.3.1 Tentamens met gesloten vragen

5. Bedenk voldoende tentamenvragen

Als volgens richtlijn 2 een toetsmatrix is opgesteld, dan kan daarin het aantal vragen per leerdoel of onderwerp worden opgenomen. Wanneer in de matrix met een globalere aanduiding wordt gewerkt, moet als eerste stap bij de toetsconstructie worden bepaald uit hoeveel vragen het tentamen zal bestaan. Deze keuze wordt in de eerste plaats bepaald door de leerdoelen: over elk leerdoel dient ten minste één vraag te worden gesteld. Voor leerdoelen die belangrijker zijn voor de cursus kunnen twee of meer vragen worden opgesteld. Over het algemeen geldt hoe meer vragen hoe betrouwbaarder de toets—maar er is natuurlijk een bovengrens omdat studenten het tentamen in maximaal drie uur moeten kunnen maken. Daarom is het beter om de ondergrens vast te stellen: uit hoeveel vragen moet het tentamen minimaal bestaan? Van Berkel et al. (2014) noemen een minimum van 40, 60 en 80 vragen voor meerkeuzevragen met respectievelijk 4, 3 en 2 antwoordmogelijkheden.

6. Bepaal het optimale aantal antwoordmogelijkheden

Bij meerkeuzevragen moeten studenten doorgaans kiezen uit 2, 3, 4 of 5 alternatieven; bij rangorde en selecteervragen kan dit aantal hoger zijn. Het ideale aantal antwoordopties hangt af van hoeveel goede afleiders kunnen worden bedacht. Hoewel het in principe mogelijk is om vragen met een verschillend aantal antwoordmogelijkheden in één tentamen op te nemen, verdient dit niet de voorkeur. Kiest een docent hier toch voor, dan moet dit van tevoren duidelijk zijn voor studenten.

7. Formuleer duidelijke tentamenvragen en antwoordmogelijkheden

Een goede gesloten vraag stellen lijkt misschien eenvoudig, maar is het niet. Hieronder staan enkele suggesties die docenten kunnen helpen bij het formuleren van duidelijke vragen (Bergwerff et al., n.d.; Haladyna et al., 2002).

- Elke vraag gaat over één onderwerp
- Schrijf korte zinnen en gebruik een actieve schrijfstijl
- Vermijd ontkenningen, en als dit niet mogelijk is, arceer de ontkenning dan op een **duidelijke** en **eenduidige** manier
- Gebruik **nooit** dubbele ontkenningen
- Zorg dat het antwoord op de vraag niet (deels) kan worden afgeleid uit een andere vraag

Voor het formuleren van antwoordmogelijkheden geven Haladyna et al. (2002) soortgelijke aanbevelingen:

- Elke antwoordoptie bevat één antwoord/onderwerp
- Zorg dat één antwoordoptie verdedigbaar correct is en de andere opties verdedigbaar incorrect
- Zorg dat de antwoordopties onafhankelijk zijn; ze mogen elkaar niet deels overlappen
- Varieer de plaats van het correcte antwoord of zet de antwoordopties altijd in dezelfde logische volgorde (bijvoorbeeld alfabetisch of van weinig naar veel)
- Zorg dat de antwoordopties ongeveer even lang zijn
- Zorg dat de antwoordopties inhoudelijk en grammaticaal vergelijkbaar zijn
- Formuleer de antwoordopties positief, en als dit niet mogelijk is, arceer ontkenningen op een **duidelijke** en **eenduidige** manier
- Vermijd het gebruik van te stellige (nooit, geen enkele) of te open (eventueel, soms) begrippen en aanduidingen
- Gebruik typische denkfouten die studenten maken om afleiders te schrijven

2.3.2 Tentamens met open vragen

8. Bedenk voldoende tentamenvragen

Deze richtlijn komt grotendeels overeen met richtlijn 5 voor gesloten vragen. Het enige verschil is dat het aantal open vragen minder zal zijn omdat studenten het antwoord zelf moeten formuleren, wat doorgaans meer tijd kost dan het kiezen van het juiste antwoord. Tabel 5 geeft een voorzichtige indicatie van de tijd die studenten nodig hebben om een vraag van een bepaalde lengte te beantwoorden (Van Berkel et al., 2014).

Tabel 5. Indicatie beantwoordingstijden open vragen

Lengte van de vraag	Beantwoordingstijd
1 zin	1 minuut
¼ pagina A4	5 minuten
½ pagina A4	10 minuten
1 pagina A4	25 minuten
2 pagina A4	60 minuten

9. Formuleer duidelijke tentamenvragen

Veel van de aanbevelingen uit richtlijn 7 gelden ook voor open vragen. Ter aanvulling geven Bergwerff et al. (n.d.) de volgende suggesties:

- Maak duidelijk onderscheid tussen de inleiding op de vraag en de vraag/vragen zelf
- Splits de vraag op in deelvragen als je meer dan één onderwerp wilt toetsen
- Formuleer de vraag in de vorm van een opdracht en gebruik hierbij actieve woorden
- Geef zo specifiek mogelijk aan wat je in het antwoord verwacht (bijv. Noem **drie** redenen waarom...")
- Geef zo specifiek mogelijk aan hoe uitgebreid het antwoord moet zijn. Dit kan door het aantal zinnen te noemen of door antwoordkaders te geven
- Gebruik standaardformuleringen voor de verschillende verrichtingsniveaus om de vraagformulering consistenter te maken (bijv. "Noem...", "Leg uit...", "Bereken...")
- Vermeld bij elke (deel)vraag hoeveel punten studenten kunnen verdienen
- Zorg voor een duidelijke en consistente vormgeving

10. Formuleer een gedetailleerd antwoordmodel

Tijdens of direct na het formuleren van de vragen, moet ook een antwoordsleutel worden opgesteld. Hierin wordt per (deel)vraag beschreven welke antwoord(en) goed worden gerekend en hoeveel punten hiermee kunnen worden verdiend. Een goede antwoordsleutel geeft ook aan hoeveel punten voor een deels correct antwoord worden gegeven, dit vergroot namelijk de kans dat een docent alle studenten op dezelfde manier beoordeelt. Als het tentamen door twee of meer docenten wordt nagekeken, dan verhoogt een antwoordmodel bovendien de interbeoordelaars overeenstemming. Een voorbeeld van een antwoordmodel wordt getoond in Figuur 5.

Figuur 5. Voorbeeld van een antwoordmodel met normeringsaanwijzingen

Vraag 1

Johnson en Johnson (1984) noemen vijf voorwaarden voor succesvol samenwerkend leren.

1. Noem deze vijf voorwaarden (5 punten)
2. Leg in maximaal 6 zinnen uit of deze voorwaarden ook gelden voor cross-age peer tutoring. (7 punten)

Antwoordmodel

1. De vijf voorwaarden zijn: (1) positieve afhankelijkheid, (2) individuele verantwoordelijkheid, (3) goede inhoudelijke discussie, (4) goede sociale vaardigheden, (5) regulatie van het groepsproces. (1 punt per kenmerk, maximaal 5 punten).
2. Ja, al deze voorwaarden gelden ook voor cross-age peer tutoring (1 punt) omdat dit een vorm van samenwerkend leren is (1 punt) waarbij de jongere leerling afhankelijk is van de oudere leerling die de rol van tutor speelt (1 punt). Beide leerlingen moeten hun verantwoordelijkheid nemen en serieus met de leesopdracht aan de slag gaan, als een van beide verzaakt stagneert het leerproces (1 punt). [...]

2.3.2 Verslagen, werkstukken, praktische opdrachten en portfolio's

Deze deelparagraaf behandelt de richtlijnen voor alle andere toetsvormen, die omwille van de leesbaarheid worden aangeduid als opdracht.

11. Beschrijf duidelijk wat studenten moeten doen

Net als bij tentamenvragen moet de opdracht aan studenten helder zijn. Meestal is een opdrachtschrijving omvangrijker dan een tentamenvraag en wordt deze in de cursushandleiding of het cursusmateriaal gegeven. Voor de formulering van de opdracht gelden dezelfde aanbevelingen als voor tentamenvragen, plus de volgende extra aanwijzingen uit Bergwerff et al. (n.d.).

- Leg uit wat van studenten wordt verwacht. Beschrijf hierbij de verplichte onderdelen van de opdracht (bijv. Het verslag moet een theoretische inleiding bevatten...') en hoe deze moeten worden uitgewerkt

(... waarin je op basis van de cursusliteratuur tot ontwerprichtlijnen komt voor een klinische interventie)

- Geef een puntsgewijs overzicht van de eisen waaraan het eindproduct moet voldoen (bijvoorbeeld het aantal pagina's van het verslag of de duur van een mondelinge presentatie)
- Kondig aan hoe vaak en wanneer studenten tussentijdse feedback krijgen
- Geef specifiek aan wanneer en hoe studenten hun opdracht moeten inleveren, plus wat de gevolgen zijn als niet aan deze eisen wordt voldaan

12. Stel een rubric op voor de beoordeling

Een rubric beschrijft net als een antwoordmodel hoe de uitwerking van studenten moet worden beoordeeld. Het verschil is dat een rubric meestal wat algemener is omdat studenten meer vrijheid hebben bij het uitwerken van een opdracht. Verder blijft de essentie gelijk, zoals het voorbeeld in Figuur 6 illustreert. Merk op dat het aantal beoordelingscategorieën desgewenst kan worden uitgebreid.

Figuur 6. Voorbeeld van een rubric voor de beoordeling van opdrachten

Criterium	Beoordeling		
	Onvoldoende (0 punten)	Voldoende (1 punt)	Goed (2 punten)
Gebruik eigen literatuur	De student gebruikt geen bronnen of verwijst alleen naar de cursusliteratuur	De student gebruikt enkele eigen bronnen	De student gebruikt voornamelijk eigen bronnen
Gebruik voorbeelden	De student gebruikt geen voorbeelden	De student geeft 1 of 2 voorbeelden	De student geeft meer dan 2 voorbeelden
Opbouw van het verslag	Het verslag mist een duidelijke structuur	Het verslag volgt de APA structuur maar bevat enkele onlogische onderdelen	Het verslag is volgens de APA structuur opgebouwd

2.4 Richtlijnen voor toetsbeoordeling

Net als in de vorige paragraaf worden de richtlijnen per toetsvorm besproken.

2.4.1 Tentamens met gesloten vragen

13. Verwijder tentamenvragen met slechte psychometrische eigenschappen

De antwoorden op tentamens met gesloten vragen worden geanalyseerd door EvaluationService. Voor het beoordelen van de psychometrische kwaliteit van de vragen zijn drie itemkarakteristieken van belang:

- De f -waarde geeft aan hoe vaak elke antwoordoptie door studenten is gekozen (de f staat dus voor frequentie). Bij goede vragen is er een redelijk evenwichtige spreiding over de afleiders; dit kenmerk laat dus vooral zien hoe plausibel de afleiders waren.
- De p -waarde is een maat voor de proportie studenten die de vraag goed heeft beantwoord en geeft dus een indruk van de moeilijkheid van de vraag. Als vuistregel geldt dat de p -waarde ergens tussen de .20 en .80 moet liggen.
- De r_{it} en r_{ir} zijn correlatiematen die laten zien hoe sterk de score op een vraag samenhangt met de totale testscore (de item-test correlatie; r_{it}) of, iets preciezer berekend, met de scores op alle andere vragen (de item-rest correlatie; r_{ir}). Beide parameters geven een indruk van het onderscheidend vermogen van een vraag. Als vuistregel geldt dat de r_{it} en/of r_{ir} hoger dan .20 moeten zijn.

Docenten kunnen deze itemkarakteristieken gebruiken om slechte tentamenvragen op te sporen en eventueel buiten beschouwing te laten bij het berekenen van de cijfers. De aanwijzingen in Tabel 6 kunnen hierbij behulpzaam zijn.

Tabel 6. Interpretatie van 'bedenklijke' item-karakteristieken en mogelijke maatregelen

Waardes	Betekenis	Mogelijke maatregelen
$p < .20$ en $r_{ir} < .00$	De vraag is zeer vaak fout beantwoord; laag scorende studenten hadden de vraag vaker goed dan hoog scorende studenten	Verwijder de vraag of reken meerdere antwoordopties goed als de f -waarde daar aanleiding toe geeft
$p < .20$ en $r_{ir} < .20$	De vraag is zeer vaak fout beantwoord en maakt nauwelijks onderscheid tussen laag- en hoog scorende studenten	Verwijder de vraag of reken meerdere antwoordopties goed als de f -waarde daar aanleiding toe geeft
$p < .20$ en $r_{ir} > .20$	De vraag is zeer vaak fout beantwoord en maakt vooral onderscheid tussen de hoog scorende studenten	Handhaaf de vraag

14. Corrigeer voor gokkans

Nadat de vragen van slechte psychometrische kwaliteit uit het tentamen zijn verwijderd, kunnen de scores worden gecorrigeerd voor gokkans. Bij tentamens die met Cirrus worden afgenomen gebeurt dit automatisch, in andere gevallen moet een docent dit zelf doen met behulp van de volgende formule:

$$\text{Punten voor gokkans} = \text{Aantal vragen} \div \text{Aantal antwoordopties}$$

Bij een tentamen van 50 meerkeuzevragen met 4 antwoordopties is de correctie voor gokkans dus $50 \div 4 = 12.5$ punten. Als een tentamen bestaat uit vragen met een verschillend aantal antwoordopties, dan moet deze formule apart worden gebruikt voor elke subset van vragen met eenzelfde aantal antwoordopties. Hierna kan met de volgende formule de score van elke student worden gecorrigeerd voor gokkans:

$$\text{Gecorrigeerd puntenaantal} = \text{Aantal behaalde punten} - \text{Punten voor gokkans}$$

15. Bepaal de cesuur en de cijfers

Bij een systeem van absoluut normeren (zie §1.3.2, stelregel 14) wordt de grens tussen slagen en zakken (de cesuur) bepaald door het aantal punten dat nodig is voor een voldoende (een 5.5). De cesuur kan met de volgende formule worden berekend:

$$\text{Cesuur} = (5.5 \times (\text{Maximale score} - \text{Punten voor gokkans})) \div 10 + \text{Punten voor gokkans}$$

Dus bij een tentamen met 60 vierkeuzevragen is de cesuur $(5.5 \times (60 - 15)) \div 10 + 15 = 39.75$ punten en is een student dus geslaagd bij 40 punten of meer. Tot slot kan voor iedere student het cijfer worden berekend met de volgende formule:

$$\text{Cijfer} = (\text{Gecorrigeerd puntenaantal} \div (\text{Maximale score} - \text{Punten voor gokkans})) * 10$$

Als een tentamen zeer slecht is gemaakt, kunnen docenten besluiten de cesuur te verlagen, bijvoorbeeld door iedere student een half punt erbij te geven. Dit dient alleen in uitzonderlijke gevallen te gebeuren en eigenlijk alleen indien er concrete aanwijzingen zijn dat de cursus of tentamensessie anders is verlopen dan gebruikelijk of gewenst. In alle andere gevallen geeft het opplussen van de cijfers studenten een verkeerd beeld, namelijk dat hun huidige tentamenvoorbereiding voldoende is geweest, wat bij net-geslaagden in wezen dus niet zo is.

16. Handhaaf de normering bij hertentamens

Hertentamens worden op precies dezelfde manier nagekeken en beoordeeld als het eerste tentamen. De cijfers kunnen weliswaar lager uitvallen omdat de groep herkansers grotendeels zal bestaan uit studenten die eerder zijn gezakt, maar dit is dus geen reden om soepeler te normeren. Omgekeerd hoeft een versoepelde beoordeling van het eerste tentamen niet te worden doorgevoerd bij het hertentamen. Stel dat er wat problemen zijn geweest bij de afname van het eerste tentamen waardoor een docent besluit iedere student een half punt erbij te geven. Als de afname van het hertentamen probleemloos verloopt, is er dus geen enkele reden om de herkansers ook dat halve punt erbij te geven.

2.4.1 Tentamens met open vragen

17. Wees kritisch op tentamenvragen met slechte psychometrische eigenschappen

Net als bij tentamens met gesloten vragen, moet de psychometrische kwaliteit van open vragen worden bepaald voordat de cijfers kunnen worden berekend. Bij open vragen is het meestal niet mogelijk om naar de p -waardes te kijken en zijn f -waardes sowieso niet van toepassing. Maar docenten kunnen wel de gemiddelde score per vraag berekenen en de standaardafwijking; bovendien kan met behulp van SPSS de r_{it} of r_{ir} worden bepaald. De interpretatie van deze statistieken is minder vastomlijnd dan bij gesloten vragen (zie Tabel 6) en vereist enige ervaring met de tentamenvragen. Wordt een vraag van onvoldoende kwaliteit bevonden, dan vindt in de regel eerder een versoepeling van de beoordeling plaats dan dat de vraag wordt verwijderd, niet in de laatste plaats omdat tentamens met open vragen vaak een beperkt aantal vragen hebben.

18. Voorkom beoordelaarseffecten bij het nakijken

Het gebruik van een antwoordmodel draagt bij aan een consistente beoordeling van elke student, ook als een tentamen door twee of meer docenten wordt nagekeken. Een antwoordmodel kan echter niet alle mogelijke vormen van *bias* voorkomen. Tabel 7 beschrijft enkele beoordelaarseffecten en wat docenten hiertegen kunnen doen (Van Berkel et al., 2014).

Tabel 7. Beoordelaarseffecten en mogelijke maatregelen

Effect	Omschrijving	Mogelijke maatregelen
Normverschuiving	De docent wordt tijdens het nakijken steeds milder of strenger	Beoordeel het tentamen per vraag (i.p.v. per student) Varieer de nakijkvolgorde per vraag
Sequentie-effect	Na een aantal zeer slechte antwoorden beoordeelt de docent een redelijk goed antwoord te hoog (en vice versa)	Beoordeel het tentamen per vraag (i.p.v. per student) Varieer de nakijkvolgorde per vraag
Halo-effect	De docent laat zich bij de beoordeling leiden door de algemene indruk die hij/zij van een student heeft	Kijk geanonimiseerd na Laat elke docent de studenten beoordelen die hij/zij niet kent
Eerste indruk-effect	De docent laat zich bij de beoordeling leiden door een snelle eerste indruk van de kwaliteit van het antwoord	Onderdruk de neiging om snel te willen beoordelen
Contaminatie-effect	De docent beoordeelt stelselmatig te hoog of te laag omdat hij/zij (on)bewust een ander doel heeft met de beoordeling, bijvoorbeeld aantonen dat de cursus moeilijk is	Kalibreer het antwoordmodel door samen met een collega enkele tentamens te beoordelen

19. Bepaal de cesuur en de cijfers

Omdat bij tentamens met open vragen niet gecorrigeerd hoeft te worden voor gokkans, is de berekening van de cesuur en cijfers relatief eenvoudig:

$$\text{Cesuur} = (5.5 \times \text{Maximale score}) \div 10$$

Dus bij een tentamen waarvoor studenten maximaal 70 punten kunnen behalen, is de cesuur $(5.5 \times 70) \div 10 = 38.5$ punten.

De cijfers kunnen worden berekend met de volgende formule:

$$\text{Cijfer} = \text{Puntenaantal} \div \text{Aantal vragen}$$

20. Handhaaf de normering bij hertentamens

Deze richtlijn komt overeen met die voor tentamens met gesloten vragen; zie richtlijn 16.

2.4.3 Verslagen, werkstukken, praktische opdrachten en portfolio's

21. Bepaal de cesuur en de cijfers

Deze richtlijn komt goeddeels overeen met richtlijn 19. Voor verslagen en werkstukken waarbij een rubric wordt gebruikt voor de beoordeling, kan het handig zijn om niet elk criterium even zwaar te laten meewegen. Bij een verslag kan bijvoorbeeld de inhoudelijke kwaliteit voor 60% meetellen voor het cijfer, de reflectie voor 30% en de vormgeving van het verslag voor 10%. Ook kunnen minimumeisen worden gesteld, bijvoorbeeld dat studenten alleen een voldoende kunnen halen als ze voor alle drie de categorieën minimaal een 6 hebben gescoord.

22. Voorkom beoordelaarseffecten

Ondanks dat een rubric de beoordeling eenduidig en consistent kan maken, liggen ook hier beoordelaarseffecten op de loer. De meeste maatregelen uit Tabel 8 gelden daarom ook voor verslagen, werkstukken, praktische opdrachten en portfolio's. Als studenten ter plekke moeten demonstreren dat zij een praktische vaardigheid beheersen (bijvoorbeeld een slechtnieuwsgesprek voeren), kunnen beoordelaarseffecten worden tegengegaan door met twee beoordelaars te werken en/of de student op video op te nemen en later terug te kijken.

Het halo-effect verdient speciale aandacht omdat een docent niet alleen beoordeelt maar vaak ook begeleidt. Indrukken die zijn opgedaan tijdens de cursus kunnen de beoordeling op een positieve of negatieve manier kleuren en een valide oordeel in de weg staan. Om deze effecten tegen te gaan verdient het aanbeveling om opdrachten door een andere docent dan de begeleider te laten beoordelen. Bij scripties is een onafhankelijke beoordeling door een tweede beoordelaar verplicht (zie stelregel 17).

23. Geef goede, maar niet te specifieke suggesties voor verbetering

Studenten wiens opdracht als onvoldoende wordt beoordeeld, hebben recht op feedback waarmee zij hun werk kunnen verbeteren. Om te voorkomen dat de revisie een 'invuloefening' wordt, dient de feedback duidelijk en inhoudelijk te zijn *zonder* punt-voor-punt aan te geven hoe iets precies moet worden veranderd. Goede feedback geeft studenten duidelijke aanwijzingen om door eigen inspanning tot een voldoende eindproduct te komen, want alleen dan geeft het cijfer een valide indruk van de mate waarin de leerdoelen zijn bereikt.

2.5 Richtlijnen voor toetsevaluatie

Een belangrijk deel van de toetsevaluatie heeft al plaatsgevonden tijdens de beoordelingsfase; de richtlijnen die daar zijn beschreven geven docenten namelijk aanknopingspunten om hun vragen of opdrachten verder te verbeteren. Daarom worden in deze paragraaf twee afsluitende richtlijnen gegeven.

24. Verbeter of verwijder slechte tentamenvragen uit de itembank

Als bij de toetsbeoordeling blijkt dat tentamenvragen van onvoldoende kwaliteit zijn, pas deze dan direct aan of verwijder ze uit de itembank. Dit voorkomt dat in het volgend studiejaar met veel moeite moet worden teruggezocht welke vragen moesten worden verbeterd of verwijderd. Voor opdrachten geldt hetzelfde.

25. Reflecteer op de toetsing tijdens de cursusevaluatie

Na afloop van een cursus evalueren de docenten hoe het onderwijs is verlopen en welke verbeteringen kunnen worden doorgevoerd. De toetsing kan hiervoor belangrijke aanwijzingen geven. Vanuit het idee van constructieve alignment kunnen slecht gemaakte tentamenvragen bijvoorbeeld aanleiding zijn om na te gaan of de colleges over dat onderwerp goed waren, het studiemateriaal duidelijk genoeg was, studenten veel

vragen over de stof stelden, enzovoorts. Het omgekeerde geldt ook: als docenten vinden dat delen van de cursus niet optimaal zijn geweest, kan dit een verklaring zijn voor tegenvallende toetsresultaten.

De resultaten van de studentevaluatie kunnen het beeld dat docenten hebben bekrachtigen of juist tegenspreken. Besteed hier serieus aandacht aan, maar ga bij veranderingen niet over één nacht ijs. De respons op studentevaluaties is meestal laag en onvoldoende representatief voor de gehele groep studenten die de cursus heeft gevolgd; het zijn meestal de studenten die zeer tevreden of zeer ontevreden zijn die de evaluatie invullen. Als in opeenvolgende jaren dezelfde opmerking wordt gemaakt én de docent de kritieken terecht vindt, kan besloten worden iets aan de toetsing te veranderen. In alle andere gevallen volstaat een gemotiveerde reactie in het *teacher report* (zie §1.3.2, stelregel 18).

Bronvermeldingen

- Anderson, L. W., & Krathwohl, D. R. (Red.) (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Addison Wesley Longman.
- Barchard, K. (2010). Internal consistency reliability. In N. J. Salkind (Red.), *Encyclopedia of research design* (vol. 1) (pp. 616–619). Sage. <https://doi.org/10.4135/9781412961288.n191>
- Bergwerff, M., Dekker, M., Blankenstein, F. van, Romein, D., & Klaren, M. (n.d.). *Tips bij toetsen*. <https://www.universiteitleiden.nl/binaries/content/assets/science/onderwijs/sterrenkunde/tipsbijtoetsen.pdf>
- Berkel, H. van, Bakx, A., & Joosten-ten Brinke, D. (2014). *Toetsen in het hoger onderwijs*. Bohn, Stafleu, van Loghum.
- Biggs, J. & Tang, C. (2011). *Teaching for quality learning at university*. Open University Press.
- Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education*, 15(3), 309–333.
- Peters, G.-J. Y. (2014). The alpha and the omega of scale reliability and validity. *The European Health Psychologist*, 16, 56–69.
- Sanders, P. (Red.) (2010). *Toetsen op school: Hoger onderwijs*. Cito. <https://www.cito.nl/-/media/files/kennis-en-innovatie-onderzoek/toetsen-op-school/cito-toetsen-op-school-ho.pdf?la=nl-nl>