

ToetsSpecial

Serie over Praktijktoetsing, deel 2

Praktijktoetsen en praktische opdrachten: hoe ontwikkel je ze?

Auteur: [Hans Kuhlemeier](#), onderwijskundige bij [Cito](#)
Datum: september 2017 (herziene versie); oorspronkelijke uitgave 2002
Internet: <http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>

ToetsWijzer

ToetsWijzer is de gids op internet voor iedereen in het onderwijs die betrokken is bij toetsing, examinering en evaluatie. ToetsWijzer bevat nieuws, internetlinks en achtergrondinformatie over toetsing in de breedste zin van het woord. Meer weten? Kijk dan op www.toetswijzer.nl.

Copyright © Stichting Cito

Het auteursrecht op het artikel 'Praktijktoetsen en praktische opdrachten: hoe ontwikkel je ze?', geschreven door Hans Kuhlemeier, berust bij Cito. Overname van het artikel, of onderdelen uit het artikel, is uitsluitend toegestaan na toestemming van de redactie van ToetsWijzer. Verzoeken tot overname van kopij dienen schriftelijk te worden gericht aan de [redactie](#) onder vermelding van het medium, de oplage en de doelgroep.

Inhoudsopgave

Inleiding	3
1 Specificeer de toets (stap 1)	4
Wat is een goede toets?.....	4
Hoe valide, betrouwbaar en bruikbaar?.....	6
Waarom wil ik toetsen?	7
Wat wil ik toetsen?.....	8
Wat zijn de randvoorwaarden?.....	9
2 Ontwikkel de opdrachten (stap 2)	10
Wat is het doel van de opdracht?	11
Is een taakanalyse noodzakelijk?	11
Hoe kies ik betekenisvolle contexten?	12
Hoe ontwerp ik zinvolle opdrachten?	12
Hoeveel opdrachten heb ik nodig?.....	13
Opdrachten al dan niet uitwisselbaar?	14
Hoe omvangrijk en hoe lang?	15
Welke vraagtypen?	15
Voordelen van gesloten vragen.....	16
Nadelen van gesloten vragen	16
Voordelen van open vragen	16
Nadelen van open vragen.....	16
Hoe formuleer ik de probleemstelling?	17
Welke verwerkings- of presentatievorm?.....	18
Samenwerking tussen leerlingen?	19
Hoeveel tijd en moeite?	20
3 Bepaal de beoordelingscriteria (stap 3)	21
Dekken de criteria de doelstelling?.....	21
Opdrachtspecifieke of opdracht overstijgende criteria?	22
Analytische of holistische beoordeling?	23
Criteria van tevoren of achteraf?	24
4 Ontwikkel een beoordelingsmiddel (stap 4)	26
Correctievoorschrift.....	26
Beoordelingsschema	27
Beoordelingsschaal.....	29
5 Probeer de toets uit en stel hem bij (stap 5)	32
Proeftoetsen	32
Validiteit	33
Profielanalyse	35
Betrouwbaarheid	36
Bruikbaarheid	37

Inleiding

In deze bijdrage geeft Hans Kuhlemeier achtergronden, vuistregels en tips voor de ontwikkeling van praktijktoetsen. Dit artikel is het tweede in een serie van drie over praktijktoetsen op Toetswijzer. In de eerste bijdrage is het begrip praktijktoets verhelderd. Het derde en laatste artikel over praktijktoetsen behandelt het ontwikkelen en gebruiken van beoordelingsschalen.

De drie artikelen over praktijktoetsing zijn te vinden op het volgende internetadres: <<http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>>.

Een praktijktoets is aanzienlijk moeilijker te ontwikkelen dan een schriftelijke toets. Praktijktoetsen trekken een zware wissel op de professionaliteit van de docent. Dit geldt zeker voor beginnende docenten die nog moeten leren om onderwijs en toetsing goed op elkaar af te stemmen. Over het maken van schriftelijke toetsen zijn talloze publicaties verschenen met vele voorbeelden van hoe het wel en niet moet. Zie bijvoorbeeld: <<http://www.toetswijzer.nl/berichten/toetsspecial-toetsconstructieproces-in-8-stappen>>. Aanwijzingen voor het maken van praktijktoetsen zijn dun gezaaid. Dit artikel bevat een stapsgewijze aanpak die de docent een eind op weg kan helpen.

Inhoud

De volgende stapsgewijze procedure vormt een nuttig hulpmiddel bij het maken van een praktijktoets:

- Specificeer de toets (stap 1)
- Ontwikkel de opdrachten (stap 2)
- Bepaal de beoordelingscriteria (stap 3)
- Ontwikkel een beoordelingsmiddel (stap 4)
- Probeer de toets uit en stel hem bij (stap 5)

In het vervolg van deze bijdrage worden de vijf stappen achtereenvolgens toegelicht. Lezers die meer geïnteresseerd zijn in praktische tips dan in achtergronden wordt aangeraden het eerste hoofdstuk slechts globaal door te nemen.

Reageren

Uw reacties, tips en vragen over praktijktoetsen zijn van harte welkom. U kunt hiervoor via e-mail contact opnemen met Hans Kuhlemeier van Cito: <hans.kuhlemeier@cito.nl>.

1 Specificieer de toets (stap 1)

Hoe maak ik een goede praktijktoets? De toetsconstructeur kan niet zomaar 'in het wilde weg' beginnen met het ontwerpen van de opdrachten van de praktijktoets. Eerst moet er over een aantal vragen goed worden nagedacht:

- Wat is de functie van de praktijktoets?
- Van welke doelstelling wil ik de beheersing vaststellen?
- Welke kennis, vaardigheden en/of attitudes wil ik toetsen?
- Welke belangrijke leerervaringen wil ik de leerlingen laten opdoen?
- Welke toets-, opdracht- en verwerkingsvormen wil ik gebruiken?
- Wanneer is er sprake van een voldoende prestatie?
- Hoe wil ik de prestaties van de leerlingen terugrapporteren?

Het proces waarin de contouren van de toets steeds duidelijker worden, noemen we de toetsspecificatie. De toetsspecificatie is de eerste stap in de constructie van een praktijktoets. Deze eerste fase resulteert in een toetsplan dat dient als richtsnoer voor de ontwikkeling.

Deze paragraaf over de toetsspecificatie behandelt vijf vragen waarover men moet nadenken alvorens te beginnen met het ontwikkelen van de opdrachten:

- Wat is een goede praktijktoets?
- Hoe valide, betrouwbaar en bruikbaar?
- Waarom wil ik toetsen?
- Wat wil ik toetsen?
- Wat zijn de randvoorwaarden?

Wat is een goede toets?

Een goede praktijktoets is onder meer valide, betrouwbaar en bruikbaar. Deze drie eisen zijn te beschouwen als algemene richtlijnen voor de ontwikkeling, evaluatie en verbetering van toetsmateriaal.

Validiteit

Een zelfgemaakte praktijktoets is valide als deze meet wat de docent bedoeld heeft ermee te meten. Toetsen kunnen verschillende doelen hebben, zoals voortgangscontrole, resultaatbepaling, determinatie en examinering. Bijgevolg zijn er ook uiteenlopende vormen van validiteit die een toets in verschillende mate kan bezitten. Een praktijktoets heeft niet alleen tot doel de prestaties van de leerlingen te beoordelen. Minstens even belangrijk is het leveren van een bijdrage aan het bereiken van onderwijsdoelen. Het is de bedoeling dat de leerlingen tijdens het maken van een praktijktoets waardevolle leerervaringen opdoen. Een belangrijk criterium voor de validiteit van een praktijktoets is de mate waarin deze leerervaringen zich inderdaad hebben voorgedaan.

Voor docenten zijn drie vormen van validiteit het meest van belang: validiteit van de consequenties, begripsvaliditeit en inhoudsvaliditeit.

Voor een praktijktoets is *validiteit van de consequenties* waarschijnlijk de belangrijkste vorm van validiteit. Een praktijktoets bezit consequentiële validiteit als het gebruik een positieve invloed heeft op het onderwijzen en leren en er geen negatieve (neven)effecten zijn. Een veel genoemd voorbeeld van een onbedoeld neveneffect stamt uit het wiskundeonderwijs.

Voorbeeld

In sommige landen ligt het accent bij de toetsing van wiskunde eenzijdig op theoretische rekenopgaven. Dit kan leerlingen ten onrechte op het idee brengen dat er bij wiskunde altijd iets uitgerekend moet worden en dat samenwerking en communicatie voor wiskundigen onbelangrijk zouden zijn. Volgens sommige toetsdeskundigen hebben praktijktoetsen in vergelijking met schriftelijke toetsen een positievere invloed op onderwijs en leren en zijn er minder negatieve (neven)effecten. Dit is echter zeker geen wet van Meden en Perzen. Als de uitvoering van een onderzoeksopdracht de leerlingen bijvoorbeeld zoveel tijd kost dat zij niet meer toekomen aan hun huiswerk, dan schort er iets aan de consequentiële validiteit.

Een toets is *begripsvalide* als deze het begrip-zoals-bedoeld daadwerkelijk meet. Als een afsluitende aardrijkskundetoets over verstedelijking in Nederland vooral leesvaardigheid zou meten is de toets niet begripsvalide. Een zelfgemaakte praktijktoets is begripsvalide als men aannemelijk kan maken dat de opdrachten de vaardigheid zoals bedoeld ook inderdaad meten en de beoogde leerervaringen zich inderdaad voordoen.

Een schriftelijke toets is *inhoudsvalide* als de opgaven representatief zijn voor de te toetsen kennis en vaardigheden. Zou een toets met meerkeuzevragen over verstedelijking in Nederland bijvoorbeeld alleen vragen over de topografie van Gelderland bevatten, dan is er iets mis met de inhoudsvaliditeit. Volgens een andere definitie is een toets inhoudsvalide als de inhoud ervan in het onderwijs voldoende behandeld is. Deze vorm van validiteit wordt ook wel curriculaire validiteit genoemd.

Van belang voor de inhoudsvaliditeit van een praktijktoets is de zogeheten *generaliseerbaarheid over de opdrachten*. De score op een praktijktoets is *generaliseerbaar* als men op grond van de gemaakte opdrachten een geldige conclusie kan trekken over de vaardigheid van de leerling. De opdrachten moeten met andere woorden representatief zijn voor de te meten vaardigheid. Zo zal één opstel niet genoeg zijn om een verantwoorde conclusie te kunnen trekken over de stelvaardigheid van de leerling. Daartoe moet men de leerling een aantal verschillende stelopdrachten voorleggen. Technisch gezien spreken we van generaliseerbaarheid als verschillende opdrachten over dezelfde vaardigheid tot een vergelijkbaar resultaat leiden.

Betrouwbaarheid

Bij de betrouwbaarheid gaat het erom hoe nauwkeurig de kennis en vaardigheden gemeten worden. Een toets is betrouwbaar als men staat kan maken op de meetresultaten, ongeacht het doel van de meting. Van een betrouwbare toets zijn de scores consistent, nauwkeurig en reproduceerbaar, kortom vrij van meetfouten. Voor de betrouwbaarheid van een praktijktoets zijn ten minste zes criteria van belang, namelijk objectiviteit, stabiliteit, specificiteit, moeilijkheid, differentiatie en toetslengte.

Een praktijktoets is *objectief* als de beoordeling op een objectieve wijze tot stand komt. Een beoordeling wordt bijvoorbeeld objectief genoemd als verschillende docenten onafhankelijk van elkaar aan een zelfde prestatie dezelfde score toekennen. In het derde artikel over praktijktoetsen behandelen we allerlei manieren om de objectiviteit van de beoordeling te verhogen, zie: <<http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>>.

Een score op een praktijktoets is *stabiel* als herhaalde afname onder vergelijkbare omstandigheden tot een zelfde resultaat leidt. Een duidelijke afname-instructie is hiervoor van groot belang.

Een praktijktoets is *specifiek* als alleen leerlingen met de vereiste kennis en vaardigheden de vragen en opdrachten tot een goed einde kunnen brengen.

Een praktijktoets is van de juiste *moeilijkheid* als deze voor de leerlingen noch te moeilijk noch te gemakkelijk is.

Een praktijktoets *differentieert* goed als deze een duidelijk onderscheid maakt tussen leerlingen met meer en minder kennis en vaardigheden.

Een praktijktoets is van de juiste *toetslengte* als deze voldoende opdrachten bevat voor een betrouwbare meting.

Bruikbaarheid

Een goede toets is bruikbaar voor de docent. Voor de bruikbaarheid zijn drie criteria van belang: efficiëntie, billijkheid en aanvaardbaarheid.

Een toets is *efficiënt* als de ermee verkregen informatie opweegt tegen de benodigde investering van tijd, moeite en kosten. Opstelvragen zijn doorgaans minder efficiënt dan meerkeuzevragen, aangezien de beoordeling de docent veel tijd kost.

Een toets is *billijk* als iedere leerling een even grote kans heeft om de toets met succes te kunnen maken (mits hij of zij zich voldoende heeft voorbereid). Als meisjes, gegeven een gelijke vaardigheid, de toets slechter maken dan jongens, is de toets niet billijk ofwel partijdig. Bügel & Sanders (1998) geven richtlijnen voor het ontwikkelen van onpartijdige toetsen:

<http://www.cito.nl/onderzoek%20en%20wetenschap/psychometrie/richtlijnen_onpartijdige_toetsen>.

Een toets is *aanvaardbaar* als leerlingen, docenten en ouders er blijk van geven de toets te accepteren als valide, betrouwbaar en bruikbaar.

Hoe valide, betrouwbaar en bruikbaar?

Een tweede belangrijk aandachtspunt bij de toetsspecificatie is het nadenken over de validiteit, betrouwbaarheid en bruikbaarheid. Deze drie kwaliteitseisen staan in een complexe driehoeksverhouding tot elkaar. Het een staat niet los van het ander.

Validiteit is het meest centrale begrip. Als een zelfgemaakte toets iets anders meet dan de docent eigenlijk had willen meten, dan zijn een hoge betrouwbaarheid en een hoge bruikbaarheid niet meer zo relevant. Het vervelende is dat betrouwbaarheid, validiteit en efficiëntie niet compenserend werken ten opzichte van elkaar. Het zijn dimensies van een verschillende orde. Een gebrek aan betrouwbaarheid kan men bij voorbeeld niet compenseren door een overschot aan validiteit. De toetsontwikkelaar zal altijd het een moeten afwegen tegen het ander.

Vaak is een hoge betrouwbaarheid en bruikbaarheid alleen maar haalbaar door welbewust concessies te doen aan de validiteit. Zo is enige tijd geleden voorgesteld de bekende stelopdracht in het examen te vervangen door de meerkeuze schrijfopdracht. De meerkeuzetoets is betrouwbaar, objectief, met de computer te scoren. Maar de validiteit als meting van de schrijfvaardigheid wordt over het algemeen als laag beschouwd. De leerling hoeft immers zelf niet meer te schrijven, maar alleen A, B, C of D aan te kruisen.

Ook kan een toets betrouwbaar en valide zijn, maar tegelijkertijd minder bruikbaar. Soms is een valide meting alleen maar mogelijk met een bewerkelijke afname- en beoordelingsprocedure. Zo kan men de discussievaardigheid alleen maar vaststellen door de leerlingen met elkaar te laten discussiëren. Ook zijn meerkeuzevragen niet geschikt om de vaardigheid in het uitvoeren van geografisch, historisch en economisch onderzoek in eigen omgeving vast te stellen. De afsluitingstoetsen basisvorming die de Citogroep sinds 1994 heeft ontwikkeld, bevatten een breed scala aan praktijktoetsen waarvan de betrouwbaarheid en validiteit zijn vastgesteld. Docenten bleken de praktijktoetsen echter weinig te gebruiken vanwege de ingewikkelde afname en de arbeidsintensieve beoordeling. Een vaksectie moet in zo'n geval bezien of er geen simpelere methode bestaat die wellicht wat minder valide lijkt, maar die wel efficiënter en bruikbaarder is.

Kortom: tijdens de toetsspecificatie zullen docenten de drie algemene kwaliteitseisen tegen elkaar moeten afwegen. Het moge duidelijk zijn dat dit geen volledig wetenschappelijke aangelegenheid is; het afwegingsproces bevat onvermijdelijk subjectieve momenten.

Waarom wil ik toetsen?

Het derde aandachtspunt tijdens de toetsspecificatie is het verkrijgen van meer duidelijkheid over de functie van de toets. Toetsen kunnen velerlei functies vervullen, zoals voortgangscontrole, resultaatbepaling, examinering, selectie en landelijke kwaliteitsbewaking. Voor praktijktoetsen in de onderbouw zijn instructie en oefening, voortgangscontrole en resultaatbepaling de drie belangrijkste functies.

In normale onderwijsleersituaties worden praktijktoetsen niet alleen gebruikt om de prestaties van de leerlingen te beoordelen, maar ook voor doeleinden van *instructie en oefening*. In de onderbouw vormt een praktijktoets vaak een integraal onderdeel van het onderwijsleerproces. Altijd is er de intentie om leerlingen al doende bepaalde onderwijsdoelstellingen te laten bereiken. De afname van een praktijktoets vergt vaak veel kostbare onderwijstijd. De toets is er niet alleen omwille van de toetsing, maar ook om de leerlingen waardevolle leerervaringen op te laten doen. Schriftelijke toetsen en proefwerken worden meestal afgenomen ter afsluiting van een serie lessen. In een praktijktoets vormen instructie, oefening en toetsing daarentegen één geheel. Een praktijktoets is meer dan alleen maar een middel om de prestaties van de leerlingen vast te stellen. Mits goed geconstrueerd kunnen ze een belangrijke rol spelen in het onderwijsleerproces. Tijdens de toetsing kunnen leerlingen een schat aan leerervaringen opdoen.

Toetsen voor *voortgangscontrole*, ook wel formatieve toetsen genoemd, geven zowel de docent als de leerlingen tussentijdse informatie over de kennis en vaardigheden. Het doel is de leerlingen al doende te laten zien waar zij goed en minder goed in zijn, welke onderdelen reeds beheerst worden en welke nog niet, zodat zij hun tekorten gericht kunnen wegwerken. De docent kan hiermee nagaan in hoeverre herhaling van de instructie of hernieuwde uitleg of oefening nodig zijn. De toetsuitslag heeft nooit zwaarwegende consequenties voor de leerling.

Toetsen voor *resultaatbepaling*, ook wel summatieve toetsen genoemd, geven inzicht in de kennis en vaardigheden van leerlingen aan het einde van een afgerond onderwijsleerproces. Het doel is niet zozeer het opsporen en verbeteren van lacunes of sturing van het onderwijsleerproces als wel het vaststellen van de kennis en vaardigheden van de leerlingen. De afname resulteert bijna altijd in een cijfer dat meetelt voor het rapport. De beslissingen kunnen zwaarwegende consequenties hebben voor de leerlingen, denk bijvoorbeeld aan overgangsbeslissingen of aan plaatsing in leerwegen en opleidingstypen.

Is er eenmaal voor een bepaalde toetsfunctie gekozen, dan is het zaak een passende toetsvorm te kiezen. Wie de leerlingen aan het einde van een serie lessen op hun prestaties wil afrekenen, kan vaak het beste voor een schriftelijke toets kiezen. Die is meestal sneller, efficiënter en betrouwbaarder dan een praktijktoets.

Wil men de leerlingen laten zien waar zij goed en minder goed in zijn en hen van hun fouten laten leren, dan ligt een praktijktoets het meest voor de hand. Uiteraard kan een praktijktoets ook voor doeleinden van resultaatbepaling worden gebruikt. Maar dan moeten de opdrachten en de beoordeling wel aan zeer hoge eisen voldoen.

Voorbeeld

Een schriftelijke toets met meerkeuzevragen kan een prima middel zijn om de kennis van bepaalde natuurkundige principes te toetsen. Als middel om vast te stellen of de leerlingen deze principes in nieuwe situaties daadwerkelijk kunnen toepassen, is een meerkeuzetoets niet zo geschikt. Daartoe kan men de leerlingen bijvoorbeeld beter een natuurkundig onderzoek laten uitvoeren.

Wat wil ik toetsen?

De vierde stap in de ontwikkeling van een praktijktoets is de doelspecificatie. Men kan niets onderwijzen en toetsen als er niet eerst goed is nagedacht over wat men precies met het onderwijs wil bereiken. Eerst moet duidelijk zijn wat de leerlingen precies moeten kennen en kunnen. Een goede doelstelling is niet alleen helder en goed geformuleerd, maar moet ook belangrijk zijn voor de leerlingen. Over welk type doelstellingen hebben we het dan? Praktijktoetsen zijn een geëigend middel om de beheersing van algemene en vakoverstijgende onderwijsdoelen vast te stellen. In de onderbouw van het voortgezet onderwijs wordt een onderscheid gemaakt in leren uitvoeren, leren leren, leren communiceren, leren reflecteren op het leerproces en leren reflecteren op de toekomst. Algemene onderwijsdoelen zijn vanuit hun aard betrekkelijk abstracte uitspraken over de eigenschappen en veranderingen die men bij de leerlingen wil bewerkstelligen. We illustreren dit aan de hand van twee voorbeelden uit de kerndoelen basisvorming 1998-2003.

Voorbeeld 1 Elementaire sociale conventies in acht nemen

Deze algemene doelstelling van de basisvorming is voor velerlei uitleg vatbaar en derhalve op uiteenlopende wijze te onderwijzen en te toetsen. Wij hebben de auteurs van de kerndoelen gevraagd om verduidelijking. Bij elementaire sociale conventies moeten we vooral denken aan zaken als 'dankjewel zeggen', 'niet door elkaar praten', 'beleefd zijn tegen volwassenen' en 'niet meteen erop slaan als je in een discussie niet direct gelijk krijgt'. Dit type doelstelling is tot op zekere hoogte toetsbaar. Zo kan bij de afname van een discussieopdracht worden geregistreerd of een leerling zijn of haar gesprekspartners laat uitspreken en tekenen van betrokkenheid vertoont. Elementaire sociale conventies kunnen ook aan bod komen bij de schrijftoetsen. Zo kan bij de briefconventies beoordeeld worden of een formele brief niet te informeel is en of aanhef en afsluiting passen bij het doel van de tekst en de status van de geadresseerde.

Voorbeeld 2 Omgaan met formele en informele afspraken, regels en procedures

Ook deze algemene doelstelling van de basisvorming kan men op uiteenlopende wijze concretiseren. Volgens de auteurs van de kerndoelen valt onder meer te denken aan allerlei aspecten van werkprocedures en klassenmanagement. Als voorbeelden worden genoemd: 'Niet met scharen door de klas lopen', 'niet zo'n rotzooi in de klas maken' en 'volgens afspraak echt na tien minuten stoppen'. Het toetsen van dit soort zaken is moeilijk, maar niet onmogelijk. Zo kan men de leerling een plan laten maken voor veldonderzoek in de eigen omgeving en later beoordelen in hoeverre dat plan gerealiseerd is. Ook ligt het voor de hand de afspraak-, regel- en proceduregevoeligheid mee te nemen als een apart aandachtspunt bij studie-, onderzoeks- en samenwerkingsopdrachten.

Deze twee voorbeelden laten zien dat men algemene doelstellingen eerst concreet moet maken alvorens te beginnen aan het ontwikkelen van les- en toetsmateriaal. Het is zaak de doelstelling zo te formuleren dat duidelijk is wat de leerlingen precies moeten kennen en kunnen.

Tip 1 Leerdoelen formuleren, hoe doe je dat?

Onderstaande richtlijnen voor het concretiseren van algemene doelen in zogeheten leerdoelen zijn ontleend aan het bekende boekje van Mager (1974) getiteld 'Leerdoelen formuleren, hoe doe je dat?'.

- Geef zo duidelijk mogelijk aan wat een leerling die de doelstelling bereikt heeft, moet kennen en kunnen (zo mogelijk in termen van direct waarneembaar gedrag).
- Geef zo concreet mogelijk aan op welke leerinhouden of leerstof de leerling de beschreven vaardigheden moet kunnen toepassen.
- Geef aan onder welke omstandigheden of voorwaarden een leerling die de doelstelling bereikt heeft de vaardigheden moet kunnen toepassen.
- Geef een zo duidelijk mogelijke omschrijving van de prestatie die minimaal nodig is wil een leerling er blijk van geven de doelstelling te beheersen.

Tip 2 Ontwikkel indicatoren

Voor het ontwikkelen van een praktijktoets zijn helder geformuleerde doelstellingen niet altijd voldoende. Vaak zijn er ook *indicatoren* nodig. De doelomschrijving bij een praktijktoets heeft meestal betrekking op een langetermijndoelstelling. Vrijwel altijd gaat het om een nog tamelijk algemene beschrijving van wat de leerlingen over een langere periode moeten kennen en kunnen. Indicatoren laten daarentegen zien wat een leerling bij de opdrachten moet kennen en kunnen. Indicatoren zijn hulpmiddelen bij het specificeren van de te toetsen kennis en vaardigheden *voordat* men met de ontwikkeling van de opdrachten begint. Doordat indicatoren veel specifiekere zijn dan doelstellingen, vormen ze een goed uitgangspunt voor het ontwikkelen van de opdrachten. Indicatoren kan men dus opvatten als de link tussen doelstelling en opdrachten.

Een voorbeeld moge dit verduidelijken. De algemene onderwijsdoelstelling 'informatievaardigheid' zou men bijvoorbeeld kunnen verduidelijken aan de hand van de volgende indicatoren:

- De leerlingen moeten kunnen aangeven hoe zij het internet kunnen gebruiken voor het verzamelen van informatie.
- De leerlingen moeten met een zoekmachine op het internet informatie kunnen verzamelen.
- De leerlingen moeten de betrouwbaarheid en actualiteit van informatie op het internet kunnen beoordelen.

De hierboven als tweede genoemde indicator zou bijvoorbeeld aanleiding kunnen zijn tot het ontwikkelen van een praktische opdracht waarbij de leerlingen via het internet een aantal oorzaken voor het uitbreken van de Golfoorlog moeten opsporen.

Tip 3 Wanneer een praktijktoets inzetten?

De voorbereiding, uitvoering en beoordeling van een praktijktoets kosten over het algemeen veel tijd en energie. Dit geldt niet alleen voor de docent, maar ook voor de leerlingen. Gebruik een praktijktoets daarom vooral voor breed toepasbare onderwijsdoelen die van belang zijn voor het leven van alle dag en/of de latere beroepspraktijk.

Wat zijn de randvoorwaarden?

Het nadenken over de randvoorwaarden is het vijfde en laatste aandachtspunt tijdens de specificatie van een praktijktoets. Hieronder staan enkele vragen die hiervoor van belang zijn:

- Aan welk type leerlingen wordt de praktijktoets voorgelegd? Aan leerlingen in de theoretische leerweg kunnen bijvoorbeeld hogere eisen worden gesteld dan aan hun leeftijdsgenoten in de beroepsgerichte leerwegen.
- Is er sprake van een homogeen of heterogeen samengestelde klas? In een heterogene klas zal men wat meer variatie moeten aanbrengen in de opdrachten en de criteria, zodat alle leerlingen, ongeacht hun niveau, leerstijl en belangstelling, aan hun trekken komen.
- Maken de leerlingen de opdrachten alleen en/of in groepjes? Indien in groepjes, worden de leerlingen dan individueel beoordeeld en/of is er sprake van een groepsbeoordeling?
- Zijn er voldoende hulpmiddelen, materiaal en gereedschap voorhanden (bijv. computer met internet) en is dat voor de leerlingen makkelijk verkrijgbaar of toegankelijk? Zijn de activiteiten en het gereedschap voldoende veilig?
- Hoeveel hulp en begeleiding kunnen de leerlingen krijgen en op welke tijdstippen kunnen zij daarvan gebruik maken? Wellicht van nog groter belang is dat de docent de leerlingen van tevoren duidelijk maakt wanneer hij *niet* beschikbaar is voor hulp en welke begeleiding hij per se *niet* wil geven. Ook is het verstandig afspraken te maken over de eventuele gevolgen van hulp en begeleiding voor de beoordeling.
- Hoeveel tijd krijgen de leerlingen?
- Waar voeren de leerlingen de opdrachten uit? Op school of ook daarbuiten?

2 Ontwikkel de opdrachten (stap 2)

Na de specificatie van de praktijktoets volgt het ontwikkelen van de opdrachten. Dit is de tweede stap in het constructieproces.

De opdrachten van een praktijktoets bepalen welke activiteiten de leerlingen uitvoeren en welke leerervaringen zij daarbij opdoen. De beheersing van een vaardigheid kan men niet beoordelen als de leerlingen deze vaardigheid niet daadwerkelijk toepassen. Soms roept een opdracht iets heel anders op bij de leerlingen dan de docent oorspronkelijk bedoelde. De opdracht bleek niet duidelijk genoeg, de leerlingen kwamen tijd tekort, ze hadden onvoldoende voorkennis of ze vonden de opdracht niet uitdagend of interessant. Een ander voorbeeld. Naar het idee van de docent waren de leerlingen hard aan het werk met een onderzoeksopdracht in de eigen omgeving. Maar hun onderzoeksverslagen gaven weinig uitsluitel over de vaardigheden die zij hadden moeten toepassen. Roepen de opdrachten inderdaad de gewenste activiteiten en vaardigheden op? Dit is het belangrijkste criterium voor de keuze en ontwikkeling van de opdrachten van een praktijktoets. Vermijd opdrachten die alleen maar leuk zijn om te doen, maar die geen uitsluitel geven over de beheersing van de vaardigheid.

Een praktijktoets kan zeer diverse opdrachten bevatten. Deze verscheidenheid kan tot uiting komen in onder meer de openheid van de vraag- of probleemstelling (gesloten versus open), de opdrachtvorm (schriftelijk versus praktisch), de authenticiteit (al dan niet levensecht) en de afname (individueel versus groepsgewijs). Vrijwel alle typen praktische opdrachten verdienen een plaats in het onderwijs. Het ene type is niet beter dan het andere. Wel is lang niet iedere opdracht even geschikt voor elk type doelstelling, leerstof of leerlingen. De keuze is aan de docent. Hij of zij staat voor de moeilijke taak precies die opdrachten te ontwikkelen die gegeven de situatie het beste passen. Hoe kiezen we nu de juiste opdrachten? In dit hoofdstuk geven we achtergrondinformatie, vuistregels en tips voor het ontwikkelen van de opdrachten van een praktijktoets. We doen dat aan de hand van een aantal vragen waarvoor de docent bij de ontwikkeling van de praktijktoets komt te staan:

- Wat is het doel van de opdracht?
- Is een taakanalyse noodzakelijk?
- Hoe kies ik betekenisvolle contexten?
- Hoe ontwerp ik zinvolle opdrachten?
- Hoeveel opdrachten heb ik nodig?
- Opdrachten al dan niet uitwisselbaar?
- Hoe omvangrijk en hoe lang?
- Welke vraagtypen?
- Hoe formuleer ik de probleemstelling?
- Welke verwerkings- of presentatievorm?
- Samenwerking tussen leerlingen?
- Hoeveel tijd en moeite?

Wat is het doel van de opdracht?

Men kan niet zomaar beginnen met het maken van een opdracht van een praktijktoets. Eerst moet duidelijk zijn wat het doel van de opdracht precies is. Hieronder staan zeven vragen over wat de leerlingen na het voltooien van de opdracht moeten kennen en kunnen. Als de docent deze vragen duidelijk en gedetailleerd ten overstaan van een collega kan beantwoorden, kan er worden begonnen met de verdere uitwerking van de opdracht.

- Welke algemene vaardigheid wordt er geoefend en getoetst? Denk bijvoorbeeld aan lange-termijn doelstellingen op het gebied van de informatievaardigheden, probleemoplossingsvaardigheden, onderzoeks- en ontwerpvaardigheden of presentatievaardigheden.
- Gaat het inderdaad om een vaardigheid waarvan de beheersing van belang is voor het dagelijks leven en/of de latere beroepspraktijk? Is de desbetreffende vaardigheid toepasbaar in een breed scala aan situaties en toepasbaar op een brede range aan problemen?
- Welke kennis (bijv. begrippen, principes, relaties) moeten de leerlingen hebben en eventueel leren toepassen?
- Is duidelijk wat de leerlingen moeten kunnen? Welke problemen moeten zij leren oplossen? Welke werkwijzen moeten zij leren toepassen? Welke producten moeten zij opleveren?
- Is duidelijk aan welke eisen de werkwijze en/of het product moet voldoen?
- Welke sociaal-emotionele vaardigheden worden er geoefend en getoetst (indien van toepassing)? Bijvoorbeeld: zelfstandig werken, samenwerken, nauwkeurig werken of kritisch leren denken?
- Is de beoogde vaardigheid onderwijsbaar en toetsbaar gegeven de groep leerlingen en de beschikbare tijd en middelen? Is de vaardigheid misschien te complex voor de minder vaardige leerlingen? Beschikken zij wel over voldoende voorkennis? Biedt de vaardigheid de betere leerlingen wel voldoende uitdaging?

Is een taakanalyse noodzakelijk?

Onderwijsdoelen zijn vrijwel altijd geformuleerd in algemene termen. Ze geven zelden volledige duidelijkheid over de veranderingen en eigenschappen die men bij de leerlingen beoogt. Leerlingen leren echter stap voor stap. Ingewikkelde vaardigheden als musiceren, een experiment uitvoeren, een onderzoek doen of een discussie voeren worden niet in één keer aangeleerd. Een vaardigheid bestaat uit een aantal deelvaardigheden of stappen. De docent moet dan een keuze maken: toetsing van de vaardigheid in zijn geheel dan wel toetsing van één of meer onderdelen ervan? In sommige gevallen is de vaardigheid zo complex dat een taakanalyse noodzakelijk is. Daarbij wordt de vaardigheid uiteengelegd in een aantal activiteiten, stappen of deelvaardigheden. De te beoordelen aspecten vloeien vaak rechtstreeks uit de taakanalyse voort.

Voorbeeld van een taakanalyse

Bij complexe vaardigheden kan de uitkomst van een taakanalyse bestaan uit een omschrijving van de relevante situaties of contexten waarin de vaardigheid gedemonstreerd moet worden, de deeltaakdelen of deelhandelingen die de leerling moet uitvoeren (in chronologische of logische volgorde) en de criteria waaraan de uitvoering moet voldoen. Onderstaand voorbeeld is afkomstig uit het handboek voor de beoordeling van leerresultaten in het gezondheidszorgonderwijs (Schotten & Robroek, 1997). Het betreft een beschrijving van de deelhandelingen van de vaardigheid 'het wassen van een patiënt' in chronologische volgorde.

Leerinhoud: het wassen van een patiënt
• verpleegkundig dossier raadplegen
• patiënt voorbereiden
• juiste werkomgeving creëren
• handen wassen
• juiste benodigdheden klaarzetten
• patiënt in juiste houding helpen
• privacy van de patiënt bewaken
• patiënt uitkleden en sieraden af (laten) doen
• gezicht wassen
• handen en armen wassen

Een waarschuwing is hier echter op zijn plaats. Als stelregel verdient het aanbeveling niet alles ineens te willen toetsen. Probeer het simpel te houden. Toets complexe vaardigheden niet alleen in onderlinge samenhang. Maak ook opdrachten bij duidelijk afgrensbare activiteiten, stappen of deeltaakdelen die mogelijk meer gericht te onderwijzen zijn.

Hoe kies ik betekenisvolle contexten?

Om de leerlingen te motiveren is het verstandig de opdrachten aan te bieden in voor hen relevante contexten. Hoe kiest men dergelijke contexten? Het is in dit verband van belang dat de opdrachten aansluiten bij de beleevingswereld van de leerlingen. Hoe bereikt men dat? Men kan proberen zich in hun situatie te verplaatsen. Wat interesseert de leerlingen? Wat vinden zij belangrijk? Welke hobby's en baantjes hebben zij? Tegelijkertijd kan men zich de vaardigheid voor ogen nemen die de leerlingen nog verder moeten ontwikkelen. Hoe wil ik dat de leerlingen de opdrachten precies uitvoeren? Wat weten zij al van het onderwerp? Wat kunnen zij al wel en wat moeten ze nog leren? Het is de kunst deze twee invalshoeken – interesse en vaardigheid - bij het maken van een praktijktoets met elkaar in verband te brengen. Stel dat het ontwikkelen van internetvaardigheden op het lesprogramma staat en The Lord of the Rings toevallig net de allerlaatste hype is. Het ligt dan voor de hand de leerlingen een zoekopdracht over dit onderwerp te laten uitvoeren. Kortom: een relevante context slaat een brug tussen enerzijds de interesses van de leerlingen en anderzijds de vaardigheden die zij nog moeten ontwikkelen.

Hoe ontwerp ik zinvolle opdrachten?

Zodra men een betekenisvolle context of thema voor een opdracht gevonden heeft, kan er worden begonnen met de verdere uitwerking van de activiteiten van de leerlingen. In een praktijktoets hebben deze activiteiten een tweeledig doel. Ze dienen enerzijds om van te leren en anderzijds om de beheersing van het geleerde vast te stellen. Hieronder staan enkele suggesties die wellicht steun kunnen geven bij het ontwerpen van zinvolle

opdrachten:

- Zorg ervoor dat de activiteiten passen bij de onderwijsdoelstelling die men wil ontwikkelen en waarvan men de beheersing wil beoordelen.
- Zorg ervoor dat de opdrachten belangrijk, motiverend en uitdagend zijn voor de leerlingen en goed aansluiten bij wat zij al weten en kunnen.
- Zorg ervoor dat de opdrachten uitvoerbaar en haalbaar zijn. Maak duidelijk hoeveel tijd de leerlingen aan de opdracht mogen besteden. En zorg ervoor dat de toegewezen tijd voldoende is om de opdracht met succes te voltooien. Maak de opdrachten noch te complex noch te gemakkelijk. Het is verstandig de opdrachten eerst zelf bij een aantal leerlingen uit te proberen en op grond van de resultaten te verbeteren.
- Omschrijf de kwaliteitseisen waaraan de werkwijze en/of het product moet voldoen.
- Geef duidelijk aan welke hulpmiddelen (materialen, computer, gereedschap) de leerlingen mogen/moeten gebruiken. En zorg ervoor dat alle leerlingen gelijke toegang tot deze hulpmiddelen hebben (met name computergebruik).
- Maak de opdrachten zo zelfsturend mogelijk, zodat de docent zo weinig mogelijk hoeft te interveniëren. Verstrek de leerlingen daartoe duidelijke, schriftelijke en stapsgewijze instructies. Verduidelijk de instructie zo nodig aan de hand van beeldmateriaal. De docent hoeft dan niet iedere stap in het uitvoeringsproces uitvoerig te introduceren en te begeleiden.
- Leg de verantwoordelijkheid voor de verslaglegging zoveel mogelijk bij de leerlingen. Geef duidelijk aan wanneer en hoe de leerlingen moeten rapporteren.
- Geef duidelijk aan hoeveel hulp en begeleiding de leerlingen mogen vragen. En op welke momenten dat geoorloofd is. Maak ook duidelijk welke gevolgen eventuele hulp en begeleiding heeft voor de beoordeling. Over het algemeen geldt: hoe ouder de leerlingen, hoe zelfstandiger zij de opdrachten van de praktijktoets kunnen uitvoeren en hoe minder hulp en begeleiding zij nodig hebben.
- Bij langere praktijktoetsen is het raadzaam om op gezette tijden klassikale instructie en uitleg in te lassen. Bepaal van tevoren welke momenten of fasen in het uitvoeringsproces daartoe het meest geschikt zijn. Het kan zinvol zijn de leerlingen enkele voorbeelden van goede en minder goede prestaties te demonstreren, zodat duidelijk wordt wat van hen verwacht wordt.
- Geef duidelijk aan waar de leerlingen de opdrachten uitvoeren. In de les, elders in de school en/of ook buiten het schoolgebouw?
- Het ontwikkelen van bruikbare opdrachten kost veel tijd en moeite. Ontwikkel de opdrachten daarom nooit alleen. Doe het altijd met één of meer collega's. Als er ook docenten van andere vakken meewerken is de kans groter dat er echt vakoverstijgende en breed toepasbare vaardigheden worden aangeleerd.

Tip

Probeer tijdens het ontwerpen van een opdracht een zo concreet mogelijk beeld te krijgen van de activiteiten van de leerlingen. Ga voor uzelf na welke activiteiten u wilt zien en welke u liever niet wilt zien. Probeer de kenmerken van de uit te voeren werkwijze en/of het te vervaardigen product zo helder mogelijk op het netvlies te krijgen. Anticipeer niet alleen op wenselijke kenmerken en handelingen, maar ook op voor de hand liggende fouten en verkeerde benaderingen.

Hoeveel opdrachten heb ik nodig?

Is er eenmaal voor een praktijktoets gekozen, dan moet men beslissen hoeveel opdrachten er nodig zijn. Hoe complexer de vaardigheid, hoe meer opdrachten er nodig zijn voor een adequate dekking. Voor het vaststellen van de beheersing van een eenvoudige vaardigheid die de leerling slechts in één situatie hoeft toe te passen zijn minder opdrachten nodig dan voor een complexe, breed toepasbare vaardigheid. Hieronder lichten we aan de hand van drie voorbeelden toe waarom wij vinden dat een praktijktoets meestal uit meer opdrachten moet bestaan.

Voorbeeld 1 Het leerlingenperspectief (billijkheid)

Stel een docent wil de vaardigheid in het oplossen van natuurkundige problemen toetsen bij tien leerlingen. Hij maakt tien korte opdrachten die hij op basis van toeval aan de leerlingen toewijst. Elke leerling maakt dus slechts één van de tien opdrachten. Kan de docent de uitslag van de toetsing dan gebruiken om conclusies te trekken over de probleemoplossingsvaardigheid van een individuele leerling? Wij hopen dat u zegt: Natuurlijk niet! Een leerling kan bijvoorbeeld een goed ontwerp maken voor het ene experiment, maar bij een andere type experiment hopeloos falen. Een goed of slecht resultaat op één opdracht kan het gevolg zijn van stom geluk of pure pech. Een leerling kan die ene opdracht toevallig al een eens eerder tegengekomen zijn. Een leerling kan toevallig geen enkele affiniteit hebben met de context waarin de opdracht gegoten is. Als een leerling meer opdrachten maakt, kan deze een tegenvaller bij de ene opdracht compenseren met een gelukje bij een andere opdracht. Vandaar dat men een leerling vrijwel altijd meer opdrachten zal voorleggen als men een gedegen uitspraak wil doen over zijn of haar vaardigheid.

Voorbeeld 2 Het vakinhoudelijk perspectief (dekking of generaliseerbaarheid)

Stel de vaksectie Nederlands wil de stelvaardigheid van de leerlingen vaststellen. De docenten ontwikkelen daartoe een flink aantal schrijfp opdrachten: een excuusbrief, een sollicitatiebrief, een memo, een verslag van een vergadering, een brief met een verzoek aan een officiële instantie, een beschrijving van een object, een traditioneel opstel et cetera. Elke leerling krijgt een andere stelopdracht voorgelegd. Kan de docent nu op basis van die ene schrijfprestatie een conclusie trekken over de stelvaardigheid van de leerling? Anders gezegd: kan hij op grond van dat ene schrijfproduct generaliseren naar het volledige domein van schrijftaken die een leerling zou moeten beheersen? Ook nu is het antwoord natuurlijk negatief. De betere vragen zouden zijn:

- Hoeveel verschillende opdrachten moet ik een leerling laten maken om een valide conclusie te kunnen trekken over zijn of haar vaardigheid?
- Als ik een leerling even zoveel andere opdrachten had gegeven, zou de conclusie over de vaardigheid van die leerling dan heel anders zijn geweest?

Voorbeeld 3 Het praktisch perspectief (geheimhouding)

Meer opdrachten verdienen ook de voorkeur als men de praktijktoetsen wil gebruiken voor resultaatbepaling in verschillende klassen. Een vaksectie wil bijvoorbeeld een gemeenschappelijk proefwerk geven in alle tweede klassen. Vanwege de geheimhouding is het vaak niet mogelijk om alle leerlingen precies dezelfde opdracht te geven. Omdat de leerling iets zichtbaar moet demonstreren is de kans op het bekend worden van die ene opdracht erg groot. In deze situatie verdient het aanbeveling om elke leerling een afzonderlijke steekproef van opdrachten uit een grotere verzameling opdrachten voor te leggen. Dit stelt wel hoge eisen aan de kwaliteit van de set opdrachten. Vanuit overwegingen van billijkheid moeten die opdrachten qua moeilijkheid, inhoud, afname-omstandigheden en beoordeling zo gelijkwaardig mogelijk zijn. In toetstechnische termen: elke set opdrachten moet men kunnen opvatten als een a-selecte steekproef uit een zelfde vaardigheids- of taakdomein.

De drie voorbeelden laten zien dat één opdracht meestal te weinig is om de beheersing van een complexe vaardigheid vast te stellen. Hierin verschilt een praktische opdracht in een praktijktoets overigens niet principieel van een meerkeuze-item in een schriftelijke toets. Vanuit deze optiek is een praktische opdracht niet meer of minder dan een 'één-itemtoets' met een navenant lage betrouwbaarheid (ook al wordt die ene opdracht op meer aspecten beoordeeld). Wie een hoge betrouwbaarheid en validiteit nastreeft, zal de leerling zoveel mogelijk opdrachten willen voorleggen.

Opdrachten al dan niet uitwisselbaar?

Een goede praktijktoets bestaat uit meer opdrachten die alle een beroep doen op dezelfde vaardigheid. De opdrachten kunnen onder meer verschillen in de openheid van de vraagstelling, de moeilijkheid en de thematiek. De uitwisselbaarheid van de opdrachten heeft als voordeel dat men de leerlingen keuzemogelijkheden kan bieden en dat men rekening kan houden met hun specifieke belangstelling, vaardigheidsniveau en leerstijl.

Het ontwikkelen van een procedurele vaardigheid bestaat veelal uit het aanleren van een aantal stappen, meestal in een vaste volgorde. Bij het doen van onderzoek wordt bijvoorbeeld een onderscheid gemaakt in de fasen van oriëntatie, planning, uitvoering, commentaar, presentatie en reflectie. In de onderbouw is het doorgaans niet verstandig om de volledige onderzoekscyclus in zijn totaliteit te toetsen. In het begin is het vaak beter om voor elke onderzoeksfase een afzonderlijke praktische opdracht te ontwikkelen. De praktijktoets bestaat dan uit een aantal opdrachten die ieder op een andere fase in het onderzoeksproces betrekking hebben. Voor elke fase is er een opdracht met afzonderlijke instructies, criteria en (tussen)producten. In dit geval zijn de opdrachten van de praktijktoets geenszins uitwisselbaar. Het bieden van keuzemogelijkheden aan de leerlingen is dan niet zinvol.

Hoe omvangrijk en hoe lang?

Hoe lang moet of mag een opdracht duren en hoeveel tijd mogen de leerlingen eraan besteden? Voor praktijktoetsen bestaat geen minimum of maximum tijdslimiet. Wat reëel is, zal ook afhangen van het belang van de te ontwikkelen vaardigheid in het leerplan. Over het algemeen geldt: hoe jonger de leerlingen hoe korter de opdrachten. Gaat het om een niet al te complexe vaardigheid dan kan men de leerlingen soms meer opdrachten in één lesuur laten maken. Doet de praktijktoets echter een beroep op vakoverstijgende vaardigheden en moet de leerling leerinhouden van verschillende vakken integreren, dan dient men meer tijd in te plannen. Een docent die nog maar net met praktijktoetsen bezig is, doet er goed aan klein te beginnen. Het ontwikkelen en het opdoen van ervaring kosten veel tijd en energie. Begin daarom met een korte praktijktoets waarvan de leerlingen de opdrachten in hooguit enkele lessen kunnen uitvoeren.

Welke vraagtypen?

In de praktische en eventueel ook schriftelijke opdrachten van een praktijktoets is plaats voor verschillende vraagtypen. We onderscheiden hier alleen gesloten vragen (o.a. meerkeuzevragen) en open vragen (kort-antwoordvragen en lang-antwoordvragen). De openheid van een opdracht hangt ten dele af van de gehanteerde vraagtypen. Het spreekt vanzelf dat een gesloten opdracht vooral gesloten vraagtypen zal bevatten, terwijl een open opdracht vooral uit open vragen zal bestaan. Voor het construeren van gesloten en open vragen zijn twee handzame brochures met vele tips beschikbaar via de databank op Toetswijzer <www.toetswijzer.nl>:

- Toetsen met gesloten vragen (door Moelands, Noijons en Rem)
- Toetsen met open vragen (door Erkens en Moelands)

Hier volstaan we met het geven van een schematisch overzicht van de voor- en nadelen van open en gesloten vragen.

Voordelen van gesloten vragen

- De correctie kost weinig tijd. De antwoorden op geprecodeerde vragen kan de docent zelfs nakijken met een mal.
- De beoordeling is volledig objectief. Afgezien van slordigheidsfouten, komt elke docent tot dezelfde score.
- De formuleervaardigheid van de leerling speelt geen rol.
- Er is veel kennis en inzicht afvraagbaar in korte tijd.

Nadelen van gesloten vragen

- Niet alle soorten doelen en leerstof kunnen met gesloten vragen worden getoetst.
- Gesloten vragen doen vaak een sterker beroep op leesvaardigheid en kennis van de wereld dan open vragen; deze 'taalbarrière' kan bijvoorbeeld leerlingen met een anderstalige achtergrond benadelen.
- Leerlingen die de stof onvolledig beheersen kunnen 'gokgedrag' vertonen. De raadkans wordt groter naarmate de vraag minder alternatieven kent. Er moeten dan veel vragen gemaakt worden en dat is kostbaar en tijdrovend.
- De constructie is wat moeilijker en tijdrovender, zeker als men niet alleen feitenkennis wil toetsen, maar ook inzicht en vaardigheden.

Voordelen van open vragen

- Open vragen zijn meer geschikt als er veel goede antwoorden mogelijk zijn of als een probleem op verschillende manieren aangepakt en opgelost kan worden.
- De leerling heeft een grotere vrijheid in het beantwoorden van de vraag.
- De antwoorden op open vragen geven de docent meer informatie over de mate waarin de leerlingen de stof beheersen en waar zich de problemen precies voordoen.
- De kans dat de leerling het goede antwoord raadt, is (bijkans) nul.

Nadelen van open vragen

- Het beantwoorden van een open vraag kost relatief veel afnametijd. Daardoor kan men in dezelfde tijd minder vragen stellen. Dat kan ten koste gaan van de betrouwbaarheid.
- Open vragen zijn minder objectief. De docent moet ze zelf beoordelen. Hierbij kunnen allerlei storende beoordelaarseffecten optreden. Zo kunnen leerlingen met een duidelijk handschrift en een goede schriftelijke uitdrukkingsvaardigheid een voordeel hebben. Bij een kort-antwoordvraag geldt het laatste bezwaar niet, maar het handschrift blijft een punt in het voordeel van de duidelijke schrijver. Dit probleem is deels te verhelpen met correctievoorschriften, beoordelingsschema's en beoordelingsschalen. Ook beeldschermtoetsen kunnen in het geval van kort-antwoordvragen soelaas bieden.
- Het is niet gemakkelijk om de vraag of opdracht zo te formuleren dat het voor alle leerlingen duidelijk is wat voor antwoord of gedrag er van hen verlangd wordt.
- Het beoordelen kost de docent doorgaans meer tijd en moeite. Dit geldt uiteraard meer voor lang-antwoordvragen dan voor kort-antwoordvragen.

Hoe formuleer ik de probleemstelling?

Het is van groot belang dat de opdrachten van een praktijktoets duidelijk zijn. Een goede instructie voorkomt dat de leerlingen te vaak voor uitleg terug moeten naar de docent. Hoe kunnen we de probleemstelling het beste formuleren? En uit welke onderdelen moet de instructie bestaan? Hieronder geven we enkele suggesties voor het formuleren van een probleemstelling, het instrueren van de leerlingen en het correct gebruik van taal.

De probleemstelling

Een goede probleemstelling voor een opdracht voldoet onder meer aan de volgende eisen:

- Het probleem moet vakinhoudelijk correct zijn.
- Er moeten verschillende goede oplossingen mogelijk zijn.
- Een mogelijke oplossing mag niet alleen op grond van algemene ontwikkeling of gezond verstand te vinden zijn.
- De probleemstelling moet de leerling voldoende houvast geven om het probleem op te kunnen lossen; van de andere kant mag de oplossing niet reeds in de probleemstelling vervat zijn.

De instructie van de leerlingen

Het Algemeen Pedagogisch Studiecentrum (APS) heeft richtlijnen ontwikkeld voor de formulering van de instructies bij omvangrijke opdrachten in het examendossier, zie: <http://www.aps.nl/vaardigheden/toetsen_van_praktische_opdrachte.htm>. Deze richtlijnen kunnen ook bruikbaar zijn voor de kortere opdrachten in de basisvorming.

Richtlijnen voor de instructie van de leerlingen

- *Waarom?* Geef aan waarom de leerling de opdracht moet doen, wat de opbrengst is en waarom dit onderdeel van belang is.
- *Wat?* Geef aan wat de leerling precies moet gaan doen. Puntsgewijs en overzichtelijk geformuleerd. De opdracht kan zeer gesloten zijn, en een andere keer kan men de leerling meer keuzevrijheid geven.
- *Hoe?* Geef aan hoe de leerling de opdracht moet uitvoeren. Een docent kan een procedure willen meegeven of hij kan alleen vermelden dat een bepaalde procedure gevolgd moet worden. Ook hier kan men variëren in de mate van openheid. Wordt de hele procedure gegeven, moet de leerling een reeds aangeleerde procedure bewandelen of mag hij een eigen procedure doorlopen?
- *Hulp(middelen).* Geef aan welke hulpmiddelen de leerling mag/moet gebruiken en wanneer hij hulp van de docent of anderen mag vragen. Ook hier geldt weer dat de hulpmiddelen voorgeschreven kunnen worden of dat de leerling zelf de hulpmiddelen kan kiezen.
- *Samenwerking (wie).* Geef aan of de leerling alleen werkt of mag/moet samenwerken met andere leerlingen. Vermeld zo nodig de minimale en/of maximale groepsgrootte.
- *Verdeling van de taken.* Als er sprake is van groepswerk, geef dan aan hoe de verdeling van de verschillende taken over de verschillende leerlingen is of zou moeten zijn.
- *Tijd.* Geef aan hoeveel tijd de leerling heeft. Niet alleen de totale tijd, maar zo nodig ook de verdeling tussen de tijd in en buiten de les.
- *Afnamecondities.* Geef aan wat er onder toezicht gedaan wordt en wat er door de leerling buiten de les gedaan wordt.
- *Beoordelingscriteria en de weging.* Vermeld de beoordelingscriteria en eventueel de weging van de criteria ten opzichte van elkaar.
- *Producten (resultaat).* Geef aan welke producten de praktische opdracht als uitkomst moet hebben en wat de vorm van de presentatie is.
- *Verwerking (presentatie).* Geef zo nodig aan op welke wijze de leerling de resultaten moet presenteren.

Het taalgebruik

De probleemstelling bij de opdrachten moet kort, eenvoudig en duidelijk zijn. Bovendien moet de probleemstelling taalkundig correct zijn. Hieronder staan enkele aanwijzingen voor helder en bondig taalgebruik. Eerst wordt een van voorbeelden voorzien probleem benoemd en vervolgens wordt aangegeven hoe de docent dit probleem kan verhelpen.

Tip Helder en bondig taalgebruik

1. Lange zinnen met in het begin veel onbekende informatie
Voorbeeld: De bikkelharde intimidatie waarmee de Praagse regering werd overstelpt door Hitlers diplomaten bleef niet zonder gevolgen: de splitsing van Tsjecho-Slowakije werd door de regering in Praag geaccepteerd.
Remedie: anders ordenen en splitsen; eerst bekende informatie, dan nieuwe gegevens.
Voorbeeld: Hitlers diplomaten overstelpten de regering in Praag met hun keiharde intimidatie. De Duitse politiek slaagde. De regering in Praag accepteerde de splitsing van Tsjechië en Slowakije.
2. Passieve zinnen
Voorbeeld: Er werd gebouwd...
Remedie: actief maken (voorbeeld: Ze bouwden...).
3. Dubbele ontkenningen
Voorbeeld: Het is beslist niet onmogelijk dat... (zie ook het voorbeeld bij punt 6).
Remedie: vervangen door bevestigende zinnen (voorbeeld: Het is waarschijnlijk dat...).
4. Tussenvoegsels en nuanceringen
Voorbeelden: maar liefst, vooral, meestal, vaak, dan ook en dus.
Remedie: schrappen; als ze écht nodig zijn, dan extra benadrukken door een andere vraagformulering.
5. Vriendschappelijkheidsformuleringen
Bijvoorbeeld: Probeer uit te leggen... of: Kun je aangeven wie...
Remedie: schrappen (voorbeeld: Leg uit... of: Geef aan...).
6. Verwijswoorden
Voorbeeld: deze, hierom, daardoor, dit... (De intimidatie was niets ontziend. Dit...).
Remedie: schrappen; herhaal het woord waarnaar verwezen wordt, ook al lijkt het dubbelop.
Voorbeeld: De intimidatie was niets ontziend. De intimidatie kreeg succes.
7. Complexe voorzetsels
Voorbeeld: met betrekking tot, door middel van, ten behoeve van.
Remedie: vervangen door eenvoudige alternatieven (bijvoorbeeld: over, door, voor).

Welke verwerkings- of presentatievorm?

De leeractiviteiten van de leerlingen moeten ergens hun neerslag in vinden. De verwerkings- of presentatievorm kan een schriftelijk verslag of een werkstuk zijn, maar ook een mondelinge presentatie, een uitvoering of een product. Het is van belang dat de verwerkingsvorm past bij de vraag- of probleemstelling en de leeractiviteiten van de leerlingen.

Tip Varieer de verwerkingsvormen

Praktijktoetsen doen vaak ongewild een sterk beroep op de leesvaardigheid van de leerlingen. Daarnaast brengt het uitvoeren van de opdrachten dikwijls moeizaam schrijfwerk met zich mee. Als dit lees- of schrijfwerk geen onlosmakelijk onderdeel van de te toetsen vaardigheid is, loopt de validiteit gevaar. In de Tweede Fase blijken docenten een vrij eenzijdig accent te leggen op schriftelijke verwerkingsvormen. Het schriftelijk verslag en het werkstuk worden soms te pas en te onpas gebruikt. Het maken ervan kost de leerlingen veel tijd, terwijl docenten veel energie moeten steken in de beoordeling ervan. Leerlingen klagen dan ook nogal eens over 'veel van hetzelfde'. Om de overladenheid van het programma en de werklast voor docenten en leerlingen terug te brengen is het verstandig de verwerkingsvormen te variëren en met name wat vaker voor minder arbeidsintensieve verwerkingsvormen te kiezen. Wissel bijvoorbeeld een geschreven verslag (essay, artikel, onderzoeksverslag, verhalend verslag, recensie, enquêteverslag, interviewverslag) af met een mondelinge voordracht (uiteenzetting, beschouwing of betoog, forumdiscussie, debat, voordracht, posterpresentatie), een ontworpen product met documentatie, een (multi)mediapresentatie of een presentatie van toneel, dans of muziek. Een actief schoolbeleid kan ertoe bijdragen dat niet alle docenten dezelfde verwerkingsvormen hanteren. Onderling overleg en afstemming zijn daartoe noodzakelijk. De school kan de gemaakte afspraken vastleggen in een plan van toetsing en afsluiting.

Samenwerking tussen leerlingen?

Bij de ontwikkeling van een praktijktoets staat de docent voor de keuze de opdrachten individueel dan wel in groepjes te laten maken. Voor samenwerking tussen leerlingen pleiten verschillende argumenten.

In vergelijking met klassikaal-frontaal onderwijs vergt samenwerkend leren een grotere mate van activiteit en betrokkenheid van de leerling. Mits adequaat toegepast, hebben activerende werkvormen als leerpsychologisch voordeel dat de leerstof beter en dieper verwerkt wordt en mede daardoor beter beklijft en breder toepasbaar is. Daarnaast is het oefenen van samenwerkend leren van belang met het oog op het vervolgonderwijs. Zo wordt in de Tweede Fase van havo/vwo sterker dan voorheen een beroep gedaan op zelfstandig leren in groepsverband. Samenwerkingsvaardigheden kunnen dan goed van pas komen. Ten slotte heeft samenwerking een maatschappelijk belang. Vrijwel alle beroepen vereisen tegenwoordig sociale en communicatieve vaardigheden die nodig zijn om effectief te kunnen samenwerken.

Wanneer is samenwerking tussen leerlingen nu zinvol in de praktijk? Samenwerking kan onder meer zinvol zijn in de volgende drie gevallen:

- Als de school het belangrijk vindt dat leerlingen leren om met elkaar samen te werken.
- Als er grote verschillen in voorkennis zijn en de docent deze verschillen wil nivelleren voordat de leerlingen aan de praktijktoets beginnen. We illustreren dit aan de hand van een voorbeeld. Stel een docent wil de leerlingen een presentatie over een technisch onderwerp laten verzorgen. Sommige leerlingen weten over dat onderwerp echter veel meer dan anderen. Groepswerk is een middel om leerlingen elkaars voorkennis te laten delen.
- Als verdeling van rollen en taken functioneel is en voordeel oplevert. Als een groepje leerlingen bijvoorbeeld informatie moet opzoeken over een veelzijdig onderwerp, is het verstandig dat niet alle leerlingen tegelijkertijd naar dezelfde informatie op zoek gaan. Ook is het van belang dat niet altijd dezelfde leerlingen dezelfde rollen vervullen of dezelfde taken uitvoeren.

Het afnemen van een afzonderlijke toets voor samenwerkingsvaardigheden achten wij overigens vrij zinloos. De kwaliteit van de samenwerking kan men het beste over een langere periode beoordelen, aan de hand van een aantal praktijktoetsen. Samenwerking is dan één van de beoordelingsaspecten waar de docent telkens op let.

Hoeveel tijd en moeite?

Schriftelijke opdrachten zijn doorgaans efficiënter dan praktische opdrachten. De afname gaat snel, de beoordeling is vrijwel nooit moeilijk en papier is makkelijk te bewaren. Deze voordelen zijn groter naarmate de vraagstelling meer gesloten is en het probleem meer gestructureerd. *Mondelinge antwoorden* zijn vanwege het vluchtige karakter moeilijker te beoordelen. *Producten* zijn vaak de enige mogelijkheid om vast te stellen of de leerlingen daadwerkelijk iets kunnen (in plaats van erover te praten of te schrijven). De kwaliteit van een product kan de docent vaak vrij eenvoudig vaststellen. De afname kost evenwel meer tijd en de opslag neemt doorgaans meer ruimte in beslag. Heeft de docent erg weinig tijd, dan zijn de leerlingen mogelijk beter af met schriftelijke opdrachten.

3 Bepaal de beoordelingscriteria (stap 3)

De derde stap in de ontwikkeling van een praktijktoets is het uitwerken van de criteria, dat wil zeggen de eisen waaraan de prestaties van de leerlingen moeten voldoen. Criteria zijn er niet alleen om het werk van de leerlingen beter te kunnen beoordelen. Ze kunnen ook een belangrijke positieve bijdrage leveren aan het leren. Goede criteria geven de leerlingen aanwijzingen over wat zij moeten doen om kwalitatief hoogwaardig werk af te leveren. Ze bieden leerlingen met een nog niet voldoende prestatie gerichte informatie over wat zij nog moeten verbeteren.

Toelichting Wat zijn criteria?

Docenten die pas beginnen met praktijktoetsing zijn geneigd zich vooral te richten op de opdrachten. Een praktijktoets is echter meer dan een verzameling leuke en uitdagende opdrachten. Het is niet voldoende dat leerlingen bezig zijn met betekenisvolle activiteiten en levenssechte problemen. De leerlingen moeten niet alleen weten wat er getoetst wordt en waarom, maar ook aan welke criteria hun prestaties moeten voldoen wil er sprake zijn van een onvoldoende, net voldoende, goede of excellente prestatie. Beoordelingscriteria maken duidelijk wat er van de leerlingen wordt verwacht. Ze geven een antwoord op de vraag: 'Hoe goed is goed genoeg?' Ze definiëren beheersingsniveaus en maken duidelijk wat we verstaan onder een zwakke, middelmatige of goede prestatie.

Criteria zijn de basis voor een beoordelingsmiddel, zoals een checklist, observatieschema, beoordelingsschema of beoordelingsschaal. Om een beoordelingsmiddel te kunnen ontwerpen, moet er eerst worden nagedacht over de criteria.

Goede criteria liggen niet zomaar voor het oprapen. Docenten moeten ze zelf ontwikkelen. Er zijn verschillende soorten criteria met ieder hun eigen voor- en nadelen. De keuze is aan de docent. Wat een goede keuze is, hangt onder meer af van a) de visie op onderwijs en leren en b) het type doelstelling dat men wil onderwijzen en waarvan men de beheersing wil vaststellen. Hieronder bespreken we enkele vragen die van belang zijn voor het ontwikkelen van goede criteria:

- Dekken de criteria de doelstelling?
- Opdrachtspecifieke of opdracht overstijgende criteria?
- Analytische of holistische beoordeling?
- Criteria van tevoren of achteraf?

Dekken de criteria de doelstelling?

De beoordelingscriteria moeten de onderwijsdoelstelling zo goed mogelijk dekken. Het is belangrijk dat alle belangrijke kennis, vaardigheden en attituden in het criterium vertegenwoordigd zijn. De criteria moeten betrekking hebben op wat gegeven de doelstelling belangrijk is en niets anders. Vaak zien we dat de beoordeling wordt beperkt tot kwaliteitsaspecten die makkelijk te scoren en te beoordelen zijn. Als een belangrijk aspect nooit in de beoordeling terugkomt, zullen leerlingen daar na verloop van tijd vrijwel automatisch minder aandacht aan schenken. Kortom, de criteria moeten dekken wat gegeven de doelstelling belangrijk is en het onbelangrijke vermijden.

Voorbeeld

De vaksectie Frans heeft de eerste helft van het schooljaar systematisch aandacht besteed aan de uitbreiding van de actieve woordenschat van de leerlingen. De nieuwe woorden zijn aangebracht in diverse taalgebruikssituaties. Halverwege het schooljaar wil de vaksectie Frans weten hoe het met de woordenschat van de leerlingen gesteld is. De docenten ontwerpen daartoe een aantal mondelinge vocabulaireopdrachten bij verschillende taalgebruikssituaties. Een van de docenten vindt deze mondelinge toetsvorm echter te omslachtig. In plaats daarvan kiest hij voor schriftelijke toetsing in de vorm van een tweetalige woordenlijst Nederlands-Frans waarbij hij spelfouten zwaar aanreken. Dit geeft de leerlingen het verkeerde signaal over wat het belangrijkste is bij de uitbreiding van hun actieve mondelinge woordenschat.

Opdrachtspecifieke of opdracht overstijgende criteria?

Een tweede aandachtspunt bij het ontwikkelen van criteria is de keuze tussen opdrachtspecifieke of opdracht overstijgende criteria.

We spreken van *opdrachtspecifieke criteria* als de kwaliteitseisen alleen toepasbaar zijn op één specifieke opdracht.

Opdracht overstijgende criteria zijn daarentegen niet gebonden aan één specifieke opdracht en aan één vast afnamemoment, maar zijn van toepassing op meer opdrachten over dezelfde algemene vaardigheid.

Voorbeeld 1 Een opdrachtspecifiek criterium

In de wereld van arbeid en beroep is het kunnen uitleggen van een proces aan anderen een belangrijke vaardigheid. Een vmbo-team heeft een praktijktoets ontwikkeld die een aantal praktische opdrachten over verschillende productieprocessen bevat. Een van de meer gesloten opdrachten gaat over het produceren van bakstenen. De leerlingen geven hun uitleg aan de hand van een serie plaatjes die zij eerst in de goede volgorde moeten leggen. De uitleg van de leerling wordt op video opgenomen. Naderhand scoort de docent de aan- of afwezigheid van de volgende inhoudselementen:

- 1 introductie
- 2 afgraven van klei of leem
- 3 blijft liggen
- 4 neemt water op
- 5 naar mengmachine voor vermenging met water
- 6 roeren tot een glad beslag
- 7 in bakken persen (vormbakpers)
- 8 drogen in open schuren
- 9 bakken op hoge temperatuur
- 10 temperatuur en duur van het bakken afhankelijk van de gewenste hardheid

Voorbeeld 2 Een opdracht overstijgend criterium

Al vanaf het eerste leerjaar probeert de sectie Nederlands de stelvaardigheid van de leerlingen systematisch te ontwikkelen. Daarbij confronteren de docenten de leerlingen met een groot aantal functionele schrijfopdrachten. Niet alleen informatieve teksten, maar ook overtuigende en verhalende teksten. De kwaliteit van de schrijfproducten wordt telkenmale beoordeeld op dezelfde vijf algemene kwaliteitscriteria ofwel beoordelingsaspecten:

- doel- en publiekgerichtheid
- inhoud
- stijl
- organisatie
- technische conventies (bijv. spelling en interpunctie)

Opdrachtspecifieke en opdracht overstijgende criteria hebben hun eigen voor- en nadelen. Opdracht overstijgende criteria hebben in beginsel een hogere onderwijskundige waarde. Doordat de leerlingen dezelfde criteria op verschillende opdrachten toepassen, raken de kwaliteitseisen op den duur geïnternaliseerd. De leerlingen leren zo hun eigen werk te beoordelen, ook al zijn de opdrachten anders en ook al hebben ze de beoordelingsvoorschriften niet op papier voor zich. Worden er slechts opdrachtspecifieke kenmerken beoordeeld, dan krijgen de leerlingen alleen maar een idee van wat goed is bij deze specifieke opdracht in deze specifieke situatie. Kortom, opdracht overstijgende criteria zijn meer geschikt om leerlingen te helpen bij de transfer van kennis en vaardigheden in nieuwe situaties met nieuwe problemen.

Hiermee zij niet gezegd dat opdrachtspecifieke criteria te allen tijde verwerpelijk zijn. Het gebruik ervan hangt af van het doel van het onderwijs. Opdrachtspecifieke criteria zijn bijvoorbeeld bij uitstek geschikt als het gaat om duidelijk afgebakende kenniselementen of om procedures die in een vaste volgorde moeten worden doorlopen. Het gebruik ervan is ook zinvol als een hoge mate van betrouwbaarheid noodzakelijk is, bijvoorbeeld als er belangrijke beslissingen in het geding zijn. Het verschil tussen opdrachtspecifieke en

-overstijgende criteria is met andere woorden niet zo zwart-wit als hierboven geschetst. In de praktijk doen zich allerlei mengvormen voor.

Het verschil tussen opdrachtspecifieke en -overstijgende criteria loopt grotendeels parallel aan dat tussen indirecte beoordeling en directe beoordeling.

Indirecte beoordeling aan de hand van opdrachtspecifieke criteria lijkt op het scoren van een meerkeuzetoets. Voor elk goed antwoord of voor elk correct kenmerk krijgt de leerling een of meer punten. De leerling produceert bijvoorbeeld een grafiek, een berekening of een verklaring die aan verschillende, nauwkeurig omschreven kenmerken moet voldoen. De vaardigheid van de leerling wordt hierbij niet direct beoordeeld, maar afgeleid uit het aantal behaalde punten.

Bij *directe beoordeling* geeft de docent daarentegen geen punten voor afzonderlijke kwaliteitskenmerken. De kwaliteit wordt direct beoordeeld aan de hand van een opdracht overstijgend criterium, ook wel een algemeen of gegeneraliseerd criterium genoemd. De docent beoordeelt bijvoorbeeld niet of een grafiek aan bepaalde, van tevoren omschreven kenmerken voldoet, maar wel óf er een grafiek gebruikt is en of dat gebruik passend en correct is. Bij samenwerkingsvaardigheid geeft de docent bijvoorbeeld een oordeel over de kwaliteit van de samenwerking of van het opgeleverde product. Dit soort directe oordelen zijn opdracht overstijgend en van toepassing op allerlei opdrachten, problemen of situaties.

Tip Streef naar evenwicht

Kies een belangrijke vakoverstijgende doelstelling. Ontwikkel een aantal gelijksoortige opdrachten voor afname gespreid over een langere periode. Begin met een eenvoudige opdracht overstijgende beoordelingsschaal. Vul deze langzaam maar zeker aan met regels voor de scoring van opdrachtspecifieke kwaliteitskenmerken. Zo kan men de inhoud van schrijfproducten de eerste tijd direct beoordelen met een beoordelingsschaal (bijvoorbeeld een vijf-puntsschaal). Als de leerlingen dit kwaliteitsaspect na verloop van tijd enigszins onder de knie hebben, kan men beginnen met het indirect scoren van de aan- of afwezigheid van bepaalde opdrachtspecifieke kenmerken (bijvoorbeeld het tellen van relevante inhoudselementen in een sollicitatiebrief of een brief aan een officiële instantie). Hiermee wordt een evenwicht bereikt tussen enerzijds directe, opdracht overstijgende beoordeling op algemene kwaliteitsdimensies en anderzijds indirecte, gedetailleerde scoring op opdrachtspecifieke kenmerken.

Analytische of holistische beoordeling?

We laten de opdrachtspecifieke criteria nu even voor wat ze zijn en richten ons geheel en al op de opdracht overstijgende criteria. Binnen deze categorie zijn er verschillende ontwerpopties. Een daarvan is de keuze tussen analytische en holistische beoordeling.

Bij *holistische beoordeling* geeft de docent één algemeen oordeel over de kwaliteit van een werkwijze of een product.

Bij *analytische beoordeling* worden daarentegen verschillende oordelen op verschillende kwaliteitsdimensies gegeven. Elke dimensie krijgt een afzonderlijke score, waardoor een profiel ontstaat van de sterke en zwakke punten van de leerling. Het geven van één allesomvattend cijfer voor de kwaliteit van een gedocumenteerde schrijfofdracht is een voorbeeld van holistische beoordeling. Geeft de docent daarentegen (ook) cijfers voor afzonderlijke kwaliteitsdimensies – zoals doel- en publiekgerichtheid, inhoud, stijl, organisatie en technische conventies – dan is er sprake van analytische beoordeling.

De voordelen van holistische beoordeling zijn evident: het is eenvoudig, kost weinig tijd en geeft een duidelijk overzicht van de prestaties van de leerlingen. Holistische beoordeling is met name bruikbaar als men zich op snelle wijze een beeld wil vormen van de leerlingprestaties. Wel moet het mogelijk zijn om de kwaliteit van de werkwijze of het product op één algemene dimensie uit te drukken. Mits het aantal beoordelingsaspecten niet te groot is, heeft analytische beoordeling echter een hogere onderwijskundige waarde:

- Het past bij uitstek bij het beoordelen van complexe vaardigheden.
- Het is vooral geschikt als men de leerlingen van tevoren een idee wil geven van de eisen waaraan hun prestatie moet voldoen.
- Het is vooral geschikt als men de leerlingen wil leren hoe zij de kwaliteit van hun werk zelf kunnen beoordelen.
- Het geeft de docent en de leerlingen gedetailleerde feedback over hetgeen al wel beheerst wordt en hetgeen nog nadere uitleg en oefening behoeft.
- Analytische criteria passen bij een onderwijsaanpak waarin complexe vaardigheden stapsgewijs worden aangeleerd en geëvalueerd. Leerlingen kunnen een nieuwe vaardigheid vaak niet in een keer onder de knie krijgen. Een stapsgewijze aanpak verdient dan de voorkeur.

Tip

De volgende suggestie kan nuttig zijn bij het ontwikkelen van criteria bij een bepaalde opdracht. Stel dat u een leerling die de opdracht slecht gemaakt heeft moet uitleggen wat hij of zij niet goed gedaan heeft, waarom dat zo is en wat deze leerling zou moeten doen om een hogere score te behalen. Als de leerling begrijpt wat hij of zij de volgende keer beter moet doen, dan zijn de criteria waarschijnlijk voldoende gespecificeerd.

Criteria van tevoren of achteraf?

Bij omvangrijkere praktijktoetsen die meer tijd vergen, is het raadzaam de beoordelingscriteria van tevoren aan de leerlingen bekend te maken. Leerlingen moeten niet alleen weten wat er getoetst wordt en waarom, maar ook aan welke eisen hun prestaties moeten voldoen wil er sprake zijn van een voldoende of uitmuntende prestatie. Door de criteria van tevoren bekend te maken kan de autonomie van de docent enigszins worden doorbroken. Leerlingen worden dan beter in staat gesteld de opdrachten zelfstandig uit te voeren.

Een beoordelingsmiddel kan meer aspecten of dimensies omvatten die de docent ieder afzonderlijk moet beoordelen. Als de docent een allesomvattend cijfer wil geven, zal hij of zij de verschillende scores of oordelen moeten wegen tot één totaalcijfer. Het is uiteraard het gemakkelijkst om alle oordelen even zwaar te laten meetellen. Er kunnen echter redenen zijn om het ene aspect zwaarder te laten wegen dan het andere. De docent moet de leerlingen dan wel van tevoren op de hoogte stellen van de precieze weging, zodat zij zich optimaal kunnen voorbereiden.

Het ontwikkelen van goede criteria vereist een schoolbrede aanpak. Er is afstemming en samenwerking binnen en tussen vaksecties nodig. Het is niet alleen inefficiënt als iedere docent zijn eigen procedures en criteria ontwikkelt voor bijvoorbeeld het schrijven van een werkstuk of scriptie, het doen van onderzoek of het presenteren van eigen werk. Ook leerlingen worden erdoor benadeeld, omdat zij bij elk vak weer met een andere aanpak geconfronteerd worden.

Tip 1 Bespreek de criteria van tevoren

Alvorens de leerlingen met de opdrachten aan de slag gaan, is het verstandig de criteria eerst expliciet te bespreken. Dit kan vaak het beste in de vorm van een kort onderwijsleergesprek:

- Vertel de leerlingen dat hun werk op een aantal aspecten beoordeeld zal worden (de kwaliteitseisen).
- Licht elk beoordelingsaspect toe.
- Geef voor elk beoordelingsaspect een voorbeeld van een slechte en een goede uitwerking.
- Vraag de leerlingen nog meer voorbeelden van slechte en goede uitwerkingen aan te dragen.
- Geef de leerlingen een voorbeeld van een slechte uitwerking en vraag hoe ze deze kunnen verbeteren.
- Zorg ervoor dat de leerlingen de criteria tijdens het uitvoeren van de opdrachten regelmatig kunnen raadplegen (bijvoorbeeld door hen de beoordelingsschaal van tevoren ter hand te stellen).

Tip 2 Schoolbrede aanpak

Het systematisch aanleren van vaardigheden in de onderbouw vereist een schoolbrede aanpak. Verdeel de 'grote' vaardigheden in eerste instantie over de vakken(clusters). Wijs de onderzoeksvaardigheden bijvoorbeeld toe aan de bètavakken en de presentatievaardigheden aan de talen. Ontwikkel voor elke belangrijke vaardigheid een set beoordelingscriteria en zorg ervoor dat die set bij elk vak op (vrijwel) dezelfde wijze wordt toegepast. Hanteer voor vaardigheden die bij meer vakken terugkomen zo mogelijk dezelfde beoordelingsmiddelen. De publicatie van Henk Lindeman (APS, 2002) over het toetsen van vaardigheden en het examendossier bevat nadere informatie over een schoolbrede aanpak. De uitgave is beschikbaar via internet: <http://www.aps.nl/vaardigheden/toetsen_van_praktische_opdrachte.htm>. Zie voor meer informatie over (het beoordelen van) vaardigheden bij APS ook: <<http://www.aps.nl/vaardigheden>>.

4 Ontwikkel een beoordelingsmiddel (stap 4)

Is er duidelijkheid over de leerlingactiviteiten en de beoordelingscriteria, dan is het zaak een passend beoordelingsmiddel te ontwikkelen. Dit is de vierde stap in het constructieproces van een praktijktoets.

Praktijktoetsen kunnen uit gesloten en open opdrachten bestaan. Een meerkeuzetoets is een voorbeeld van een gesloten opdracht die volledig objectief is na te kijken. Alle goed te rekenen antwoorden zijn immers van tevoren vastgelegd. Slordigheidsfoutjes daargelaten, zou elk antwoord van elke docent dezelfde score moeten krijgen. Opdrachten met meerkeuzevragen worden gescoord aan de hand van een correctiesleutel: een lijst van de goed te rekenen antwoorden.

Een praktijktoets zal doorgaans één of meer open opdrachten bevatten. Bij het nakijken van open vragen en opdrachten speelt de subjectiviteit van de beoordelaar altijd een rol. Er zijn tenminste drie beoordelingsmiddelen die de beoordeling objectiever kunnen maken:

- Correctievoorschrift
- Beoordelingsschema
- Beoordelingsschaal

Correctievoorschrift

Als het gaat om afgebakende en direct afvraagbare kennis, dan ligt een correctievoorschrift het meest voor de hand. Het is geschikt voor het nakijken van antwoorden op vragen van het type waarop slechts één goed antwoord of een beperkte set van goede antwoorden mogelijk is. De criteria zijn vrijwel altijd opdrachtspecifiek.

Een correctievoorschrift bij een opdracht met open vragen is een lijst met richtlijnen voor de docent. Het bestaat doorgaans uit een antwoordmodel, een scoringsvoorschrift en een beoordelaarsinstructie.

Een *antwoordmodel* bevat voor elke open vraag een opsomming van de goede en foute antwoorden of inhoudselementen; soms worden ook minder goede antwoorden vermeld.

Een *scoringsvoorschrift* vermeldt de maximaal haalbare toetscore, het aantal scorepunten per vraag en de scoring van foute, gedeeltelijk goede en geheel goede antwoorden. Indien nodig wordt omschreven wanneer bonuspunten worden toegekend c.q. aftrekpunten in mindering worden gebracht.

Voorbeeld Antwoordmodel met bijbehorend scoringsvoorschrift

MAXIMALE SCORE	3 punten
Zeven antwoorden:	
- koolhydraten/zetmeel	
- water	
- eiwit	
- vet	
- vitamines	
- mineralen	
- voedingsvezels	
7 antwoorden correct	3 punten
4 of 5 antwoorden correct	2 punten
3 antwoorden correct	1 punt
minder dan 3 antwoorden correct	0 punten

Ten slotte kan het correctievoorschrift een aanvullende *beoordelaarsinstructie* bevatten. Die bestaat uit algemene richtlijnen ten aanzien van de wijze waarop de leerlingenprestaties gescoord moeten worden.

Voorbeeld Beoordelaarsinstructie

- Bij de beoordeling van een vraag worden slechts hele punten toegekend.
- Ontbreekt voor een vraag elke prestatie of is er sprake van een volledig foutief antwoord dan dient u voor deze vraag 0 punten toe te kennen.
- Indien een gegeven antwoord niet in het antwoordmodel voorkomt en dit antwoord op grond van aantoonbare vakinhoudelijke argumenten als 'juist' of 'gedeeltelijk juist' gekwalificeerd kan worden, moet het aantal beschikbare punten geheel of gedeeltelijk aan het gegeven antwoord worden toegekend naar analogie van het antwoordmodel.
- Indien 1 punt beschikbaar is en een gegeven antwoord niet volledig is of alleen gedeeltelijk juist is, dan dient u voor deze vraag 0 punten toe te kennen.
- Indien een leerling meer antwoorden (in de vorm van voorbeelden, redenen, argumenten en dergelijke) geeft dan er expliciet gevraagd worden, dan komen alleen de eerstgenoemde antwoorden voor beoordeling in aanmerking.
- Indien er slechts één antwoord expliciet gevraagd wordt, wordt dus alleen het eerst gegeven antwoord in de beoordeling betrokken.

Beoordelingsschema

Sommige open vragen ontlocken langere antwoorden aan de leerlingen. Het lukt dan vaak niet om alle relevante elementen in het antwoordmodel onder te brengen. Voorbeelden hiervan zijn vragen naar een argumentatie, een samenvatting, een tekening, een presentatie, een brief of een opstel. Een *beoordelingsschema of -lijst* biedt dan uitkomst. Het bevat een omschrijving van de criteria aan de hand waarvan de docent de antwoorden beoordeelt.

Een bekend voorbeeld van een beoordelingsschema is een zogeheten checklist. Met een checklist kan de docent afvinken of de kenmerken, eigenschappen of elementen van een werkwijze of een product al dan niet aanwezig zijn. De docent scoort 'ja' of 'nee' of 'aanwezig' of 'afwezig'.

Een fraai voorbeeld van hoe docenten een beoordelingsschema onderwijskundig en toetstechnisch verantwoord kunnen inzetten in het onderwijs is afkomstig van het Pieter Nieuwland College. Zie: <http://www.pieternieuwland.nl/Menu_Items/Vakken/doponieuw/index.html>.

Op deze school wordt het leerdoel 'Het houden van een mondelinge presentatie' onderwezen aan de hand van een gerichte instructie (Hoe bereik je je spreekdoel?), een beoordelingsschema en een zelfevaluatieformulier.

De leerlingen krijgen het beoordelingsschema van tevoren aangereikt, zodat de presentator ook zelf weet waarop hij of zij beoordeeld wordt.

Beoordelingsschema mondelinge presentatie				
Naam spreker:				
Naam beoordelaar:				
Taal:				
Datum:				
Onderwerp:				
	Zeer goed	Goed	Zwak	Onvoldoende
Houding				
Stemgebruik				
Taalgebruik				
Inhoud				
Mediagebruik				
Nabespreking				

Met het zelfevaluatieformulier kan de leerling gaandeweg zicht krijgen op de vorderingen en de verbeteringen die nodig zijn.

Zelfevaluatieformulier mondelinge presentatie		
Naam:		
Taal:		
Datum:		
Onderwerp:		
	Ja	Nee
Houding		
Ik sta of zit rustig voor het publiek		
Ik richt mij tot iedereen		
Stemgebruik		
Ik spreek duidelijk verstaanbaar		
Ik spreek niet te snel		
Taalgebruik		
Ik spreek de betreffende taal gemakkelijk		
Ik hoef niet vaak naar woorden te zoeken		
Inhoud		
Mijn onderwerp is interessant voor dit publiek		
Ik weet genoeg van het onderwerp af		
Mediagebruik		
Ik heb de werking van tevoren uitprobeerde		
Dit middel past bij mijn presentatie		
Nabespreking		
Ik reageer vriendelijk op vragenstellers		
Ik kan adequaat antwoorden		

Beoordelingsschaal

Bij praktische opdrachten met een open karakter zijn de reacties van de leerlingen vaak zeer gevarieerd. De beoordeling is dan uiterst lastig. Dan kan een beoordelingsschaal uitkomst bieden. Een beoordelingsschaal beschrijft een werkwijze of een product van een leerling op verschillende kwaliteitsniveaus. Met een beoordelingsschaal kan de docent op een glijdende schaal aangeven in welke mate kennis, vaardigheden of houdingen aanwezig zijn. De glijdende schaal bestaat uit meerdere schaalpunten die een bepaalde positie of rangorde aangeven. Een voorbeeld van een eenvoudige beoordelingsschaal voor de beheersing van een beroepspraktijkvaardigheid is: 1 = beginner, 2 = gevorderd en 3 = expert.

De omschrijving van de schaalpunten van een beoordelingsschaal kan meer of minder gedetailleerd zijn. Over het algemeen geldt: hoe gedetailleerder, hoe meer steun de docent heeft bij het beoordelen en hoe minder de oordelen van verschillende docenten van elkaar zullen verschillen. Een objectieve beoordeling is bijvoorbeeld belangrijk als verschillende docenten van een vaksectie een gemeenschappelijke praktijktoets moeten beoordelen. Te veel detail maakt de beoordelaarstaak echter onnodig complex. Nog belangrijker is echter dat de mate van detaillering is afgestemd op wat de leerlingen aankunnen.

Een beoordelingsschaal is meer dan alleen een middel om de prestatie op een praktijktoets te waarderen. Een goede schaal schept heldere verwachtingen en een gemeenschappelijk kader voor kwaliteitsbeoordeling en -verbetering.

Er zijn zeer veel soorten beoordelingsschalen. We bespreken hier alleen descriptieve beoordelingsschalen en productschalen. Voor meer informatie over het ontwikkelen van beoordelingsschalen en het gebruik ervan verwijzen we naar het derde artikel over praktijktoetsen op Toetswijzer: <<http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>>.

Descriptieve beoordelingsschalen

In een *descriptieve beoordelingsschaal* wordt de kwaliteit van de werkwijze of het product verduidelijkt door de schaalpunten van vakinhoudelijke omschrijvingen te voorzien. Deze omschrijvingen maken duidelijk hoe prestaties van verschillende kwaliteit zich van elkaar onderscheiden. De beoordeling vindt plaats door het werk van een leerling te vergelijken met de omschreven kwaliteitsniveaus.

Voorbeeld Descriptieve beoordelingsschaal

De schaal 'Wanneer is vis vers?' is een fraai voorbeeld van een descriptieve beoordelingsschaal. De schaal is bruikbaar als onderdeel van een praktijktoets over het thema 'Het beoordelen van de versheid van voedingsmiddelen'. Toepassingsmogelijkheden zijn het biologie-onderwijs, het horeca-onderwijs of de lessen praktische sectororiëntatie. De versheid van vis wordt op drie aspecten beoordeeld, te weten uiterlijk, consistentie en geur. Per beoordelingsaspect moeten de leerlingen letten op verschillende kenmerken, zoals de toestand van de huid, ogen en kieuwen. De schaalpunten zijn: 3 = zeer vers, 2 = vers, 1 = matig vers en 0 = niet geschikt voor consumptie. De kenmerken zijn direct waarneembaar en nauwkeurig omschreven. De schaal is ontleend aan de website van Schmidt Zeevis Rotterdam, zie: <http://www.schmidtzeervis.nl/html/visrecepten.html>.

EG-beoordelingsschaal versheid vis

Verse vis is veerkrachtig. Druk met de vinger op de vis; de afdruk van uw vinger mag niet in de vis achter blijven. De ogen moeten nog helder zijn en bol staan. Zolang de vis vers is zijn de kieuwen rood. Vis moet neutraal ruiken en niet naar vis. Hoe sterker de vislucht, hoe ouder de vis.

Criteria waarderingscijfers				
Te beoordelen elementen	3 zeer vers	2 vers	1 matig vers	0 niet geschikt voor consumptie
Uiterlijk				
Huid	heldere, metaalachtige glanzende kleur; geen verkleuring; glashelder doorzichtig slijm	heldere kleur, maar zonder glans; ietwat troebel slijm	verbleekte, matte kleur; melkachtig slijm	doffe kleur; ondoorzichtig slijm
Ogen	convex (bol); doorzichtig hoornvlies; zwarte glanzende pupil	bol, maar iets ingevallen; hoornvlies iets glanzend; zwarte, doffe pupil	vlak, glanzend hoornvlies; ondoorzichtige pupil	in het midden hol; melkachtig hoornvlies; grijze pupil
Kieuwen	helle kleur; geen slijm	blekere kleur; lichte sporen; helder slijm	verbleekt; ondoorzichtig slijm	geelachtig; melkachtig slijm
Spierweefsel	blauwachtig doorschijnend glad, glanzend; zonder verandering van de oorspronkelijke kleur	fluweelachtig, wasachtig stroef; lichte verkleuring	licht ondoorzichtig	ondoorzichtig
Kleur langs wervelkolom	geen verkleuring	licht rose	rose	rood
Organen	nieren en resten van andere organen, alsmede bloed in de aorta, helder rood	nieren en resten van andere organen dof rood; bloed verbleekt	nieren, resten van andere organen en bloed bleek rood	nieren, resten van andere organen en bloed bruinachtig
Consistentie				
Spierweefsel	stevig elastisch; glad van oppervlak	verminderde elasticiteit	iets week (slap); verminderde elasticiteit; oppervlak wasachtig (fluweelachtig) dof	week (slap) schubben laten gemakkelijk los de huid korrelig oppervlak
Wervelkolom	breekt, in plaats van los te laten	blijft hechten	laat enigszins los	laat gemakkelijk los
Buikvlies	sluit volledig aan	sluit aan	laat enigszins los	laat gemakkelijk los
Geur				
Huid, kieuwen, buik	als van zeewier	niet als van zeewier maar toch niet slecht	lichtelijk zuur	zuur

Productschalen

Soms zijn de reacties van de leerlingen zo divers of ongrijpbaar dat zelfs een descriptieve beoordelingsschaal nog te weinig sturing biedt, hoe goed de schaalpunten ook worden toegelicht. Het is dan raadzaam de schaalpunten te voorzien van zogeheten ankers. Een anker is een voorbeeld van een concreet leerlingenproduct dat het 'niveau' van het schaalpunt illustreert. Zo'n van ankers voorziene schaal wordt wel een *productschaal* genoemd. Een anker kan de vorm hebben van een tastbaar product, zoals een brief, opstel, tekening, foto of bouwconstructie. Bij vluchtige prestaties, zoals een mondelinge presentatie, een dans of een muziekstuk, kan een anker uit een audio- of video-opname bestaan.

Het beoordelen met een productschaal kan men het beste zien als een sorteertaak. De docent beoordeelt elk product door dat telkens te vergelijken met de voorbeeldproducten van de beoordelingsschaal. De docent kent een score toe door het te beoordelen product als het ware op de schaal te plaatsen.

Een productschaal biedt de beoordelaar een concreet referentiekader waarin de kwaliteit van het werk van de leerlingen zichtbaar wordt. Productschalen zijn bij uitstek geschikt om de beoordeling te objectiveren en de leerlingen vertrouwd te maken met de beoordelingscriteria. Docenten en leerlingen ontwikkelen hiermee een gemeenschappelijke definitie van kwaliteit. Mits op de juiste wijze gebruikt, maakt een goede schaal duidelijk wat de docent van de leerlingen verwacht (criteria), hoe goed zij de opdracht hebben uitgevoerd (feedback) en wat zij moeten doen om hun prestatie te verbeteren (remediëring). Een productschaal helpt leerlingen bij het beoordelen en verbeteren van hun werkwijzen en producten.

Het nadeel is echter dat het maken van een goede productschaal geen sinecure is. Het zinvol gebruik van een beoordelingsschaal stelt hoge eisen aan de professionaliteit van de docent. Docenten die zelf een beoordelingsschaal willen maken en geïnteresseerd zijn in gebruik ervan verwijzen wij naar het derde artikel over praktijktoetsen op Toetswijzer: <http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>.

5 Probeer de toets uit en stel hem bij (stap 5)

Als de praktijktoets met de instructies voor de leerlingen in conceptvorm gereed is, is het verstandig deze op hun deugdelijkheid te beproeven. Het vaststellen en verbeteren van de kwaliteit van een praktijktoets is de vijfde stap in het constructieproces.

Een praktijktoets moet aan vergelijkbare kwaliteitseisen voldoen als een schriftelijke toets. Ook praktijktoetsen moeten valide, betrouwbaar en bruikbaar zijn. Bij een praktijktoets worden deze eisen echter anders ingevuld en zijn er enkele extra problemen te overwinnen. De kwaliteit van een praktijktoets is veel moeilijker vast te stellen dan die van een toets met meerkeuzevragen. Dit onderdeel bevat achtergronden en richtlijnen voor het vaststellen en verbeteren van de kwaliteit van een zelfgemaakte praktijktoets.

Deze hebben betrekking op:

- Proeftoetsen
- Validiteit
- Betrouwbaarheid
- Bruikbaarheid

Proeftoetsen

Het uittesten van de praktijktoets en het begeleidend materiaal gebeurt in de zogeheten pretestfase. Deze kan bestaan uit een try-out en/of een pretest. Try-outs en pretests worden uitgevoerd om één of meer van de volgende redenen:

- het verkrijgen van gegevens over onder meer de moeilijkheid van de opdrachten;
- het verkrijgen van gegevens over de vakinhoudelijke, opgaventechnische en taalkundige kwaliteit van de opdrachten;
- het samenstellen van het beoordelingsschema en/of de beoordelingsschaal;
- het vaststellen van de afnametijd;
- het uitproberen van de procedures voor afname, scoring, beoordeling, verwerking, analyse en rapportage.

Pretest

Een belangrijk doel van een pretest is slechte vragen en opdrachten te signaleren. Dit kan door de vragen en opdrachten ter commentaariëring voor te leggen aan interne en externe deskundigen op vakinhoudelijk en toetstechnisch gebied. In een pretest worden de opgaven door een groot aantal leerlingen gemaakt (normaliter honderd of meer). Daarbij bepaalt de toetsontwikkelaar zowel de vakinhoudelijke als de psychometrische kwaliteit van de opgaven. Dit laatste houdt in dat de antwoorden van de leerlingen met de computer geanalyseerd worden. Voor docenten is dit vrijwel altijd te arbeidsintensief. Een try-out behoort vaak wel tot de mogelijkheden.

Try-out

Ook in een try-out onderwerpt men de praktijktoets aan een kritisch onderzoek. Er gelden echter minder strenge onderzoeksvoorwaarden dan bij een pretest. Vaak wordt een try-out gehouden met een vrij klein aantal leerlingen (steekproef) en worden aan de representativiteit van de steekproef geen hoge eisen gesteld. De bij een try-out verkregen gegevens bevatten alleen aanwijzingen voor het verbeteren van de opgaven en opdrachten en kunnen niet worden gebruikt voor het doen van uitspraken over de groep personen waarvoor de opgaven zijn bedoeld. Het doel is elementaire tekortkomingen op te sporen van vakinhoudelijke en opgaventechnische aard. Analyse van antwoorden op open vragen kan informatie opleveren over de inhoudelijke kwaliteit van de opgave. Ook kan men leerlingen vragen de termen die zij niet kennen te onderstrepen of de zinnen waarmee zij moeite hebben aan te geven. Ook kunnen leerlingen gevraagd worden naar de moeilijkheidsgraad van de opgaven en opdrachten, de aantrekkelijkheid (motivatie) en de tijdsduur.

Verder kan men kijken naar de uitvoerbaarheid van de procedures voor afname, correctie en/of beoordeling.

Validiteit

Docenten in de onderbouw gebruiken praktijktoetsen niet alleen voor doeleinden van resultaatbepaling en voortgangscntrole, maar ook als instructie- en oefenmateriaal. De toetsing moet daarom ook voldoen aan eisen die normaliter aan goed onderwijs worden gesteld. Deze eisen vallen buiten het bestek van deze publicatie en blijven daarom verder onbesproken.

Hoe bepaalt men nu of een praktijktoets valide is? Het vaststellen ervan is veel moeilijker dan het zo op het eerste gezicht lijkt. Als docenten meer uitsluitel willen, zullen zij de inhoud van de praktijktoets grondig moeten analyseren en de leerlingen moeten observeren en interviewen tijdens de uitvoering van de opdrachten. Hieronder geven wij enkele suggesties voor het vaststellen van de validiteit van praktijktoetsen. We maken daarbij een onderscheid in:

- validiteit van de consequenties;
- begripsvaliditeit;
- inhoudsvaliditeit.

Validiteit van de consequenties

Bij dit type validiteit gaat het om het gebruik van de toets. Wat getoetst wordt, heeft invloed op wat onderwezen wordt. Hoe getoetst wordt, beïnvloedt hoe er onderwezen wordt. Elke vorm van toetsing heeft neveneffecten. Deze kunnen positief zijn. Een serie interessante, uitdagende praktijktoetsen kan er bij voorbeeld toe leiden dat leerlingen een positievere houding aannemen ten opzichte van het vak en de docent die dat vak geeft. Een docent kan onder de druk van de veranderde exameneisen ertoe overgaan om bepaalde wenselijk geachte vaardigheden te onderwijzen. De negatieve neveneffecten treden echter meer op de voorgrond. Het pad van de toetsgeschiedenis ligt bezaaid met goed bedoelde pogingen die verkeerd uitpakten. De toetspraktijk kent vele voorbeelden van onbedoelde en meestal ongewenste neveneffecten. Zo kan een eenzijdig accent op meerkeuzevragen leiden tot minder variatie in leerdoelen en leeractiviteiten. Over het algemeen geldt: hoe groter het belang van de toetsing voor de school en de leerlingen, hoe groter de onbedoelde neveneffecten.

Voorbeelden van ongewenste neveneffecten

- Toetsen kost tijd. Hoe meer toetsen, proefwerken of overhoringen docenten geven, hoe minder tijd er overblijft voor onderwijs, instructie en oefening. De voorbereiding, afname en correctie van evaluatiemiddelen vergen tijd die wellicht beter anders besteed had kunnen worden.
- Toetsing kan leiden tot een versmalling van het onderwijsaanbod. Een eenzijdig accent op meerkeuzetoetsen in het examen kan ertoe leiden dat docenten vooral die leerstof behandelen die gemakkelijk met deze toetsvorm getoetst kan worden (teaching-to-the-test). Daardoor krijgen moeilijker toetsbare kennis en vaardigheden onvermijdelijk minder aandacht. Omgekeerd kan een te sterk accent op praktijktoetsing bijdragen aan een verwaarlozing van de direct afvraagbare kennis van feiten en principes. Over het algemeen geldt: hoe groter het belang van de toetsing voor de school en de leerlingen, hoe groter dit soort onbedoelde neveneffecten.
- Teaching-to-the-test kan landelijke referentiegegevens bij toetsen onbruikbaar maken. Een fraai voorbeeld hiervan is het zogeheten Lake Wobegon-effect. Begin jaren tachtig ontdekte een Amerikaanse onderzoeker de statistische anomalie dat het toetsgemiddelde in alle staten die een bepaalde toets gebruikten boven het landelijke gemiddelde lag. Deze kunstmatige inflatie van testcores werd het Lake Wobegon-effect genoemd naar het in de VS bekende verhaal over een fictieve samenleving waarin alle vrouwen sterk zijn, alle mannen er goed uitzien en alle kinderen zich boven het gemiddelde bevinden.
- Een eenzijdig gebruik van meerkeuzetoetsen kan ertoe bijdragen dat leerlingen zich alleen nog maar inspinnen voor leerstof die met deze toetsvorm beoordeeld kan worden (what you assess is what you get). Een eenzijdig accent op toetsing voor resultaatbepaling kan ertoe bijdragen

dat leerlingen alleen nog maar hun best doen voor activiteiten die met een cijfer beloond kunnen worden.

- Toetsing kan leerlingen opzadelen met een eenzijdig beeld van het vakgebied. Een sectie scheikunde geeft bijvoorbeeld alleen maar proefwerken met meerkeuzevragen over geïsoleerde feiten en principes. Bij leerlingen kan dan de indruk ontstaan dat het bij scheikunde altijd gaat om problemen waarover iedereen het eens is en waarvoor slechts één oplossing bestaat. Hadden de leerlingen regelmatig laboratoriumproeven gedaan, waarbij zij eigen hypothesen toetsen en eigen conclusies trekken, dan hadden zij een geheel ander beeld van het vak gekregen. Een ander voorbeeld is het verkeerde idee dat er bij wiskunde altijd iets uitgerekend moet worden en dat samenwerking en communicatie voor wiskundigen onbelangrijk zouden zijn. Een eenzijdig accent op praktijktoetsing zou ook verkeerd kunnen uitpakken. Het kan er wellicht toe leiden dat leerlingen feiten, begrippen en principes als minder belangrijk gaan beschouwen.
- Het veelvuldig werken met aftrekpunten kan ertoe leiden dat leerlingen bang worden om fouten te maken en risico's uit de weg gaan. Denk bijvoorbeeld aan een leerling die uit angst om fouten te maken moeilijke woorden vermijdt. Als de toetsing het accent legt op wat de leerlingen niet kunnen in plaats van wat zij wel kunnen, kan dat negatieve gevolgen hebben voor het zelfbeeld en de motivatie.
- Te moeilijke toetsen kunnen het zelfvertrouwen van leerlingen ondermijnen en hun motivatie negatief beïnvloeden ('dit is zo moeilijk, dit is voor mij niet weggelegd'). Omgekeerd kunnen te makkelijke toetsen leiden tot onverschilligheid en demotivatie ('dit is zo makkelijk, daar hoef ik me niet voor in te spannen').
- Toetsen kunnen misbruikt worden om leerlingen te disciplineren en te straffen. De toetsing geeft de leerlingen dan ongewenste signalen over hoe macht, rollen en status in het klaslokaal verdeeld zijn.

Voor het vaststellen van de consequentiële validiteit van een praktijktoets zijn onder meer de volgende vragen van belang:

- Heeft de praktijktoets een positieve invloed op onderwijzen en leren?
- Heeft het gebruik van de toets onbedoelde (neven)effecten? Een goede praktijktoets maximaliseert de positieve (neven)effecten en minimaliseert de negatieve.
- Geeft de toets de leerlingen de juiste boodschap over wat belangrijk is om te leren?
- Is de verzamelde informatie relevant voor de te nemen beslissing? Is de 'bewijsvoering' voor de te nemen beslissing voldoende? Als de leerling bijvoorbeeld een lage beoordeling krijgt en het desbetreffende onderdeel moet overdoen, is die beslissing dan gerechtvaardigd gegeven de inhoud en de lengte van de toets?
- Stimuleert de praktijktoets de nieuwsgierigheid van de leerlingen? Worden de leerlingen gemotiveerd om te leren?

Begripsvaliditeit

Een zelfgemaakte praktijktoets is begripsvalide als de docent aannemelijk kan maken dat de opdrachten de vaardigheid zoals bedoeld ook inderdaad meten en de beoogde leerervaringen zich inderdaad voordoen. Voor het vaststellen van de begripsvaliditeit zijn onder meer de volgende vragen relevant:

- Is er een duidelijke omschrijving van de vaardigheid die de toets beoogt te meten?
- Zijn de beoordelingscriteria duidelijk over wat de leerlingen moeten kennen en kunnen en passen ze bij de vaardigheid die onderwezen en getoetst wordt?
- Roepen de opdrachten de beoogde vaardigheden op? Treden de bedoelde waardevolle leerervaringen inderdaad op? Maken de leerlingen zoals bedoeld gebruik van 'hogere-orde denkprocessen' en passen zij inderdaad complexe vaardigheden toe? Als men bijvoorbeeld samenwerkingsvaardigheid met een praktijktoets wil vaststellen, passen de leerlingen deze vaardigheid dan ook inderdaad toe? Of komen de leerlingen nauwelijks toe aan overleg, afstemming en taakverdeling?
- Zijn de problemen die de leerlingen moeten oplossen echt nieuw voor hen? Of kunnen zij ook een goede beoordeling krijgen door uit het hoofd geleerde kennis blind toe te passen?
- Doen de opdrachten niet onbedoeld een te sterk beroep op de taalvaardigheid van de leerlingen (in plaats van op de vaardigheid die men eigenlijk had willen oefenen en evalueren)?

Meer uitsluitel over de begripsvaliditeit kan men slechts tot op zekere hoogte van achter het bureau verkrijgen. Bestudering van de praktijktoets met bijbehorende documentatie is meestal niet voldoende. Vaak zal men de leerlingen ook tijdens het uitvoeren van de opdrachten moeten observeren. Ook de hardop denkmethode kan meer uitsluitel geven. Vraag de leerlingen tijdens de uitvoering hardop te vertellen waar zij precies mee bezig zijn, hoe zij de vraagstelling interpreteren, hoe zij de opdrachten uitvoeren, hoe zij gegevens interpreteren of conclusies trekken et cetera.

Tip

Hoe weet ik nu of mijn praktijktoets bij de leerlingen de juiste leeractiviteiten heeft opgeroepen en een beroep heeft gedaan op de beoogde kennis en vaardigheden? Een handig hulpmiddel is een gezamenlijke analyse en bespreking van het werk van de leerlingen. Bestudeer eerst samen met uw collega's de opdrachten met de bijbehorende beoordelingsmiddelen. Formuleer vervolgens enkele verwachtingen ten aanzien van hetgeen de leerlingen van deze opdrachten zouden moeten leren. Leg uw collega's vervolgens een steekproef uit het werk van de leerlingen voor. Probeer in gezamenlijk overleg een conclusie te trekken over hetgeen de leerlingen daadwerkelijk van de opdrachten geleerd hebben. Contrasteer het geleerde vervolgens met hetgeen de leerlingen hadden moeten leren. Als er sprake is van een discrepantie, kijk dan nog eens goed naar de opdrachten en ga na of deze wellicht verbeterd moeten worden.

Inhoudsvaliditeit

Een praktijktoets is inhoudsvalide als de opdrachten representatief zijn voor de te meten vaardigheid. Ook moeten de leerlingen voldoende in de gelegenheid zijn geweest om zich de te meten kennis en vaardigheden eigen te maken. Voor het vaststellen van de inhoudsvaliditeit zijn onder meer de volgende vragen relevant:

- Vormen de opdrachten een representatieve steekproef uit de verzameling van alle opdrachten die men bij de te toetsen vaardigheid zou kunnen ontwikkelen?
- Worden er geen belangrijke kennis en vaardigheden over het hoofd gezien? Worden er onbedoeld triviale aspecten in de beoordeling betrokken?
- Sluit de inhoud van de toets goed aan bij wat de leerlingen al wisten en konden?
- Zijn de leerlingen tijdens de toetsing voldoende in de gelegenheid geweest om zich de gemeten kennis en vaardigheden eigen te maken?

Om een algemene vaardigheid goed te dekken is vaak een flink aantal opdrachten nodig. Door een gebrek aan tijd en middelen is het meestal niet mogelijk om al die opdrachten te ontwikkelen en af te nemen. Als het aantal opdrachten klein is in verhouding tot het aantal opdrachten dat nodig is voor een goede dekking, is er geen sprake meer van een representatieve steekproef uit het 'toetsdomein'. De uitslag van de praktijktoets wordt dan te zeer afhankelijk van de opdrachten die toevallig in de toets zijn opgenomen. Anders gezegd: generalisatie van de toetsscore naar het toetsdomein is dan niet meer geoorloofd. Het vaststellen van de generaliseerbaarheid is verre van eenvoudig. Wie hierover meer te weten wil komen, kan zich de volgende vraag stellen: 'Zou de leerling een sterk afwijkende beoordeling hebben gekregen als deze een andere opdracht over dezelfde vaardigheid gemaakt zou hebben?' Als het antwoord op deze vraag negatief is, schort er mogelijk iets aan de inhoudsvaliditeit.

Profielanalyse

Technisch gezien is er sprake van inhoudsvaliditeit (generaliseerbaarheid) als verschillende opdrachten over dezelfde vaardigheid tot een vergelijkbaar resultaat leiden. Een profielanalyse kan hierover wellicht meer uitsluitel geven. Met een profielanalyse kan de docent vaststellen of leerlingen die het goed doen op de ene opdracht ook hoog scoren op andere, overeenkomstige opdrachten. Het volgende voorbeeld laat zien hoe een docent zelf een profielanalyse kan uitvoeren.

Voorbeeld van een profielanalyse

Ter ontwikkeling en beoordeling van de vaardigheid in het doen van onderzoek heeft een sectie natuur- en scheikunde vijf opdrachten ontwikkeld. De leerlingen hebben deze onderzoeksopdrachten in de loop van het schooljaar gemaakt. De docent heeft hen voor elke opdracht een cijfer gegeven. Hebben leerlingen met hoge cijfers op de ene opdracht ook hoge cijfers op de andere opdracht? Het scoreprofiel laat zien dat de leerlingen behoorlijk consistent zijn in hun scores op de vijf opdrachten. De docent kan dan ook met een gerust hart een afsluitend cijfer geven voor de vaardigheid in het ontwerpen, uitvoeren en presenteren van onderzoek.

Scoreprofiel van de praktijktoets 'Het ontwerpen, uitvoeren en presenteren van natuurkunde-experimenten'

Klas: 3A

Leerling	Cijfer				
	Opdracht 1	Opdracht 2	Opdracht 3	Opdracht 4	Opdracht 5
Ben Schreuder	6	7	8	7	8
Esther van Ginkel	3	4	5	3	4
George Hoeven	2	3	2	3	4
Barbara Meurders	5	6	7	7	6
Bas Offenbergh	7	5	6	6	7
Annelies Kaak	4	3	2	3	4
Martin van Roekel	8	9	8	9	9

In ons voorbeeld is de samenhang tussen de opdrachten gelukkig hoog. Dit hoeft echter niet altijd het geval te zijn. Een lage samenhang betekent dat een laag cijfer op de ene opdracht vaak samengaat met een hoog cijfer op een andere opdracht. Een mogelijke verklaring voor dit uiteenlopen van de scores is dat de opdrachten een beroep doen op geheel verschillende vaardigheden. Een inhoudelijke analyse van de opdrachten kan hierover meer uitsluitsel geven.

Betrouwbaarheid

Een belangrijk aspect van betrouwbaarheid is objectiviteit. De score op een praktijktoets is *objectief* als de beoordeling op een objectieve wijze tot stand komt. Voor het bepalen van de objectiviteit van de beoordeling zijn onder meer de volgende drie vragen van belang:

- Geeft de docent verschillende leerlingen voor een zelfde prestatie een zelfde beoordeling?
- Geeft de docent voor een zelfde prestatie een zelfde beoordeling als zijn of haar collega's?
- Is de hoogte van de beoordeling afhankelijk van het moment waarop de docent de prestatie beoordeelt? Als de docent dezelfde prestatie enige tijd later nogmaals beoordeelt, valt het oordeel dan hoger of lager uit dan de eerste beoordeling?

Tip

Verzamel een aantal producten van zwakke, middelmatige en goede leerlingen. Vraag uw collega's deze producten aan de hand van de door u gemaakte beoordelingsschaal te beoordelen. Vergelijk uw oordelen met die van uw collega's en bespreek eventuele discrepanties. Pas de beoordelingsschaal en de beoordelingsinstructies desgewenst aan.

Bruikbaarheid

Een derde eis waaraan praktijktoetsen moeten voldoen is bruikbaarheid. Voor de bruikbaarheid van een praktijktoets is vooral de efficiëntie van belang. Een praktijktoets is *efficiënt* als de ermee verkregen informatie opweegt tegen de benodigde investering van tijd, moeite en kosten. Dit geldt niet alleen voor de docent, maar ook voor de leerlingen. Praktijktoetsen zijn over het algemeen minder efficiënt dan schriftelijke toetsen. Enkele adviezen zijn dan ook op hun plaats:

- Wie een praktijktoets overweegt, dient zich altijd af te vragen of de leerlingen zich diezelfde kennis en vaardigheden niet ook op een andere, minder omslachtige wijze eigen kunnen maken. Is een klassikale les of een onderwijsleergesprek niet een veel efficiënter middel? Zijn de leerlingen met een schriftelijke toets niet veel beter af?
- Gebruik praktijktoetsen en beoordelingsschalen vooral voor breed toepasbare vaardigheden die van belang zijn voor het leven van alledag en de latere beroepspraktijk.
- Begin klein. Probeer niet al te veel algemene vaardigheden tegelijkertijd te onderwijzen en te beoordelen. Beoordeel niet al te veel aspecten per vaardigheid.
- Het ontwikkelen van de opdrachten en de beoordelingsschalen is arbeidsintensief. Ga ervan uit dat het de eerste keer zelfs erg veel tijd kost. Doe het daarom nooit alleen, maar altijd samen met één of meer collega's. Volg zo mogelijk het voorbeeld van een collega die al meer ervaring met praktijktoetsen heeft opgedaan.
- Wees bereid om risico's te nemen en te leren van fouten. De beste praktijktoetsen hebben een lange ontwikkelhistorie achter de rug en functioneren beter naarmate docenten ze vaker gebruiken.
- Het is moeilijk van tevoren in te schatten hoeveel tijd de voorbereiding, uitvoering en presentatie van een nieuwe praktijktoets de leerlingen zullen kosten. Probeer te voorkomen dat leerlingen er onevenredig veel tijd aan besteden, ook al zijn zij nog zo gemotiveerd om de opdrachten tot een goed einde te brengen. Laat de leerlingen bijvoorbeeld de tijdsbesteding in een logboek bijhouden. Grijp in zodra de uitvoering te veel tijd en moeite kost.

Derde artikel over praktijktoetsen

Het derde artikel over praktijktoetsen op Toetswijzer bevat praktische informatie over het *ontwikkelen* van beoordelingsschalen en het *gebruik* ervan. Tevens worden achtergrondinformatie, vuistregels en tips gegeven ter verhoging van de *objectiviteit* van de beoordeling. Het artikel bevat verder een *checklist* waarmee docenten de kwaliteit van een zelfgemaakte praktijktoets kunnen beoordelen. Tot slot is een overzicht van de geraadpleegde literatuur opgenomen. U vindt het artikel op het volgende internetadres: <<http://www.toetswijzer.nl/berichten/toetsspecial-praktijktoetsen-en-praktische-opdrachten>>