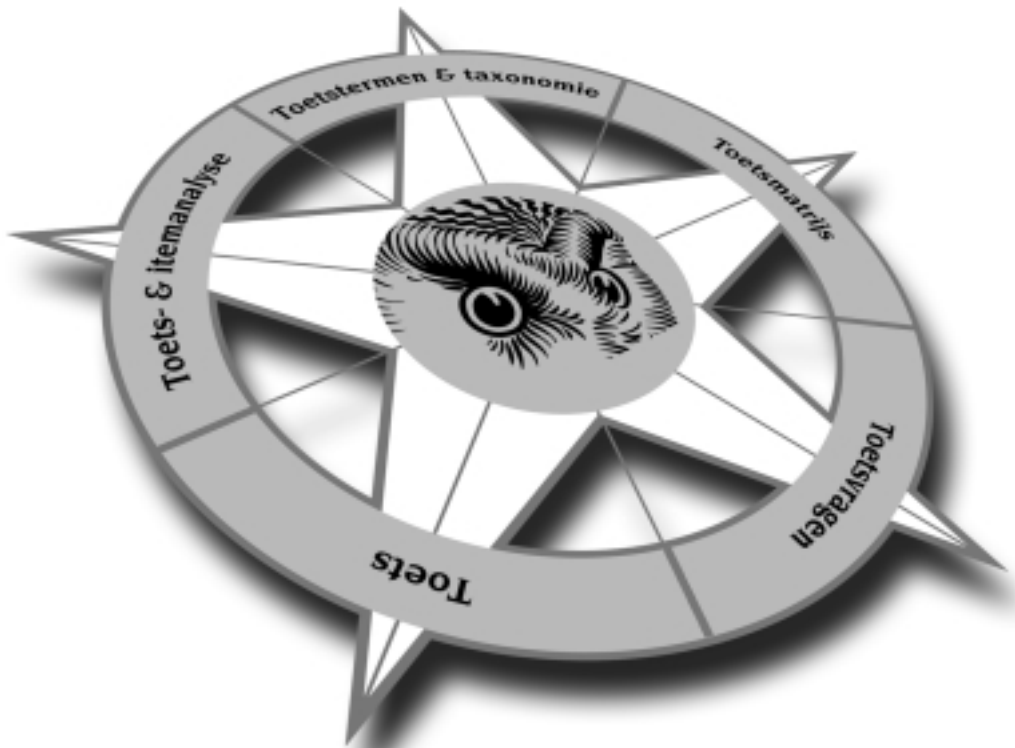


---

► **TOETSONTWIKKELING**  
*in de praktijk*

**HOE MAAK IK GOEDE VRAGEN EN TOETSEN?**



---

## Inhoudsopgave

1.	<b>Inleiding</b> . . . . .	7
1.1	Doel en doelgroep . . . . .	7
1.2	Toetspentagram . . . . .	8
1.3	Competenties toetsen . . . . .	8
2.	<b>Leerdoelen en toetstermen gebruiken</b> . . . . .	11
2.1	Wat is een leerdoel? . . . . .	11
2.2	Van leerdoel naar toetsterm . . . . .	12
2.3	Het belang van goede leerdoelen en toetstermen . . . . .	13
3.	<b>Een taxonomie gebruiken</b> . . . . .	15
3.1	Wat is een taxonomie? . . . . .	15
3.1.1	De taxonomie van Bloom . . . . .	15
3.1.2	De taxonomie van Romiszowski . . . . .	17
3.2	Een geschikte taxonomie kiezen . . . . .	18
3.3	De voordelen van een taxonomie . . . . .	19
4.	<b>Een toetsmatrijs gebruiken</b> . . . . .	21
4.1	Wat is een toetsmatrijs? . . . . .	21
4.2	Een toetsmatrijs invullen . . . . .	21
4.3	Het belang van toetstermen bepalen . . . . .	22
4.4	De voordelen van een toetsmatrijs . . . . .	23
5.	<b>Toetsvragen construeren</b> . . . . .	25
5.1	Inleiding . . . . .	25
5.2	Gesloten en open vragen . . . . .	25
5.2.1	Het verschil tussen gesloten en open vragen . . . . .	25
5.2.2	Een gesloten of een open vraag gebruiken . . . . .	25
5.2.3	Ordering en voorbeelden van vraagvormen . . . . .	26
5.2.4	Onderdelen van een vraag . . . . .	33
5.3	Gesloten vragen construeren . . . . .	34
5.3.1	Algemene constructietips . . . . .	34
5.3.2	Kwaliteitseisen en constructievoorschriften . . . . .	35
5.3.3	Afleidingsformuleringen . . . . .	64
5.3.4	De meest passende vraagvorm kiezen . . . . .	67
5.4	Open vragen construeren . . . . .	74
5.5	Feedback geven . . . . .	77
5.5.1	Het moment van feedback geven . . . . .	77
5.5.2	Het niveau van de feedback . . . . .	78
5.5.3	Tips voor het formuleren van feedback . . . . .	79
5.5.4	Leerstofverwijzingen opnemen . . . . .	80



<b>6.</b>	<b><i>Een toets samenstellen</i></b> . . . . .	<b>81</b>
6.1	Inleiding . . . . .	81
6.2	De functie van de toets . . . . .	81
6.2.1	De diagnostische functie . . . . .	81
6.2.2	De evaluatieve functie . . . . .	81
6.2.3	De selectieve functie . . . . .	82
6.3	Kwaliteitseisen voor de toets . . . . .	82
6.3.1	Validiteit . . . . .	82
6.3.2	Betrouwbaarheid . . . . .	83
6.3.3	Het belang van de eisen voor toetsen met een verschillende functie . . . . .	84
6.4	Toetsvragen selecteren . . . . .	85
6.5	Scoringsregels opstellen . . . . .	86
6.6	Een cesuur vaststellen . . . . .	87
6.6.1	De absolute en relatieve methode . . . . .	87
6.6.2	Een voorbeeld van beide methodes . . . . .	89
6.6.3	Het effect van de scoringsregels . . . . .	90
6.6.4	De kansscore in de cesuur verwerken . . . . .	91
6.6.5	Cijfers geven . . . . .	91
6.7	De toetsinstructie opstellen . . . . .	92
6.8	Een toetsplan opstellen en de toets vervaardigen . . . . .	93
<b>7.</b>	<b><i>Een toets- en itemanalyse interpreteren</i></b> . . . . .	<b>95</b>
7.1	Wat is een toets- en itemanalyse? . . . . .	95
7.2	Toetsindices interpreteren . . . . .	95
7.2.1	Frequentieverdeling . . . . .	96
7.2.2	Betrouwbaarheid . . . . .	97
7.2.3	Standaardmeetfout . . . . .	99
7.2.4	Gemiddelde moeilijkheid . . . . .	99
7.3	Itemindices interpreteren . . . . .	100
7.3.1	Moeilijkheid . . . . .	101
7.3.2	De kwaliteit van de afleiders . . . . .	103
7.3.3	Het onderscheidend vermogen van het item . . . . .	104
7.4	De toets- en itemanalyse gebruiken voor verbeteringen . . . . .	105
7.4.1	De toets zelf aanpassen . . . . .	106
7.4.2	De toetsvragen aanpassen . . . . .	106
7.4.3	Aanbevelingen opstellen voor toetsconstructie . . . . .	107
7.5	Validiteit beoordelen . . . . .	108
7.5.1	Inhoudsvaliditeit . . . . .	108
7.5.2	Begripsvaliditeit . . . . .	108
7.5.3	Criteriumvaliditeit . . . . .	108
	<i>Verder lezen</i> . . . . .	111
	<i>Index</i> . . . . .	113



De cursist:

- weet welke akten een rol spelen bij de hypotheek;
- weet wat registergoederen zijn;
- kan uitleggen wat de rangorde van hypotheek inhoudt;
- kan een notarisafrekening toelichten.

## 2.2 Van leerdoel naar toetsterm

In een leerdoel moeten minimaal de inhoud en het gedrag omschreven zijn. Wij spreken van een toetsterm als het leerdoel is verbijzonderd met de minimumprestatie en de voorwaarden. In feite geeft een toetsterm weer *op welke wijze* een leerdoel getoetst gaat worden.

Bij het formuleren van een toetsterm is het volgende van belang:

- er moet een *minimumprestatie* zijn bepaald in de vorm van bijvoorbeeld een tijdslimiet of een graad van nauwkeurigheid;
- er moeten *voorwaarden* zijn gesteld waaronder de kennis of vaardigheid aangetoond moet worden, zoals:
  - het gebruik van bijvoorbeeld een gegeven formule (informatie);
  - het gebruik van bijvoorbeeld een rekenmachine (hulpmiddelen).

Op basis van de leerdoelen in de vorige paragraaf zijn de volgende toetstermen geformuleerd.

De cursist:

- weet welke akten een rol spelen bij de hypotheek en kan deze rol in enkele regels beschrijven;
- weet wat registergoederen zijn, welke drie soorten registergoederen bestaan en kan van elke soort twee voorbeelden geven;
- kan aan de hand van de inschrijving van de hypotheek bij het Kadaster uitleggen wat de rangorde van hypotheek inhoudt;
- kan een notarisafrekening toelichten waarbij sprake is van een makelaarscourtage.

In het volgende voorbeeld is in de toetsterm niet alleen een minimumprestatie gegeven, maar ook een voorwaarde waaronder deze moet kunnen worden getoond.

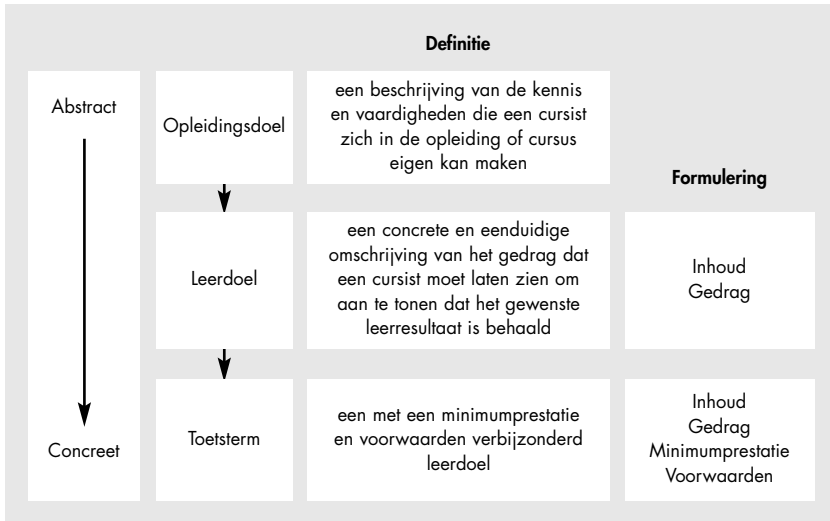
*Leerdoel*

De leerling kan rekenen met tafels.

*Toetsterm*

De leerling kan rekenen met de tafels van 1 tot en met 10 (minimumprestatie), zonder daarbij een rekenmachine te gebruiken (voorwaarde).

De volgende figuur toont via welke stappen u van een opleidingsdoel naar een toetsterm gaat.



Figuur 2.1: Van opleidingsdoel naar toetsterm

In de praktijk zult u het onderscheid tussen leerdoel en toetsterm wellicht niet altijd expliciet maken. Het is wel belangrijk dat u uw leerdoel of toetsterm zo duidelijk en volledig mogelijk formuleert. Zorg er tegelijkertijd voor dat u de toetsterm niet te concreet omschrijft! Het is namelijk belangrijk dat de toetstermen zodanig zijn geformuleerd dat het mogelijk is om er verschillende toetsvragen bij te maken. Een toetsterm mag daarom ook nooit geformuleerd zijn als toetsvraag.

*Te concreet geformuleerde toetsterm*

De leerling kan 3 maal 4 uit het hoofd uitrekenen.

### 2.3 Het belang van goede leerdoelen en toetstermen

Door leerstof en toetsvragen te ontwikkelen op basis van bij elkaar horende opleidingsdoelen, leerdoelen en toetstermen zal een sterke relatie tussen beide ontstaan. Zo voorkomt u dat de kandidaten na een toets zeggen dat ze bepaalde vragen niet konden beantwoorden omdat ze er geen leerstof over aangeboden hebben gekregen.

In de praktijk gaat dit echter niet altijd goed. Als de leerdoelen en toetstermen te abstract geformuleerd zijn, is de kans groot dat leerstof en toetsvragen niet op elkaar aansluiten. De leerstof en de toetsvragen zijn dan namelijk afhankelijk van de interpretatie die de ontwikkelaar aan de leerdoelen en toetstermen geeft. Bij concreet geformuleerde leerdoelen

Constructievoorschrift Vr-4: De vraag mag geen dubbele ontkenning bevatten.  
Kwaliteitseis: Efficiëntie

Dubbele ontkenningen zijn onnodig verwarrend. De kandidaat moet de vraag een aantal keer lezen om deze goed te kunnen begrijpen. De kans blijft bovendien groot dat de kandidaat de vraag verkeerd begrijpt.

Hoe het niet moet

Welke van de onderstaande beweringen is NIET waar?

- A. Pieter van Vollenhoven is lid van het koninklijk huis
- B. Prins Constantijn is geen lid van het koninklijk huis (\*)
- C. Prinses Máxima is lid van het koninklijk huis
- D. Prinses Irene is geen lid van het koninklijk huis

Deze vraag is vooral een lees oefening en meet dus niet op efficiënte wijze de aanwezige kennis. De vraag kan eenvoudig zonder dubbele ontkenning worden geformuleerd.

Hoe het wel moet

Wie van onderstaande personen is GEEN lid van het koninklijk huis?

- A. Pieter van Vollenhoven
- B. Prins Constantijn
- C. Prinses Máxima
- D. Prinses Irene (\*)

Constructievoorschrift Vr-5:	De vraag mag geen strikvraag zijn.
Constructievoorschrift St-6:	De stam moet kort en bondig geformuleerd zijn en mag geen overbodige informatie bevatten die de kandidaat kan afleiden.
Kwaliteitseis:	Objectiviteit Efficiëntie

Vanzelfsprekend is het niet verstandig een vraag door de formulering moeilijker te maken. Een vraag moet de kennis van de kandidaat meten, en niet zijn leesbaarheid. Formuleringen die de kandidaat bewust naar een verkeerd antwoord leiden, doen de kandidaat evenmin recht.

#### Hoe het niet moet

De visie van minister Heinsbroek op het Nederlandse televisieaanbod heeft veel aandacht gekregen.

Hij wil pulp en gewelddadige films van de Nederlandse tv-netten weren.

Van welk ministerie was Heinsbroek minister?

- A. het ministerie van Economische Zaken (\*)
- B. het ministerie van Justitie
- C. het ministerie van Onderwijs, Cultuur en Wetenschappen
- D. het ministerie van Sociale Zaken

De eerste zinnen van de stam zijn niet relevant voor het te geven antwoord en kunnen de kandidaat zelfs op het verkeerde been zetten, omdat de uitspraak meer voor de hand ligt bij de andere ministeries dan bij het ministerie van Economische Zaken. De eerste zinnen van de stam kosten de kandidaat bovendien onnodig extra leestijd.

#### Hoe het wel moet

Van welk ministerie was Heinsbroek in het kabinet Balkenende I minister?

- A. het ministerie van Economische Zaken (\*)
- B. het ministerie van Justitie
- C. het ministerie van Onderwijs, Cultuur en Wetenschappen
- D. het ministerie van Sociale Zaken

### 5.3.3 Afleiders formuleren

Bij het construeren van een gesloten vraag is het formuleren van het juiste antwoord vaak geen probleem. Des te lastiger is het om goede afleiders te bedenken. Goede afleiders zijn echter van groot belang: ze verkleinen de kans dat een kandidaat die de stof niet beheerst het juiste antwoord raadt. Een kandidaat die het juiste antwoord niet op basis van zijn kennis vindt, gaat foute antwoorden wegstrepen. Hoe duidelijker een alternatief herkenbaar is als afleider, hoe makkelijker zo'n kandidaat er een streep door kan zetten. Zoals wij al eerder opmerkten, zijn er diverse trainingen voor het maken van meerkeuzetoetsen die van deze techniek gebruikmaken.

In de vorige paragraaf hebben we diverse constructievoorschriften voor het formuleren van goede afleiders besproken. In deze paragraaf geven we nog wat extra tips voor het formuleren van goede afleiders. Wat betreft het aantal afleiders bij een toetsvraag is geen bindend voorschrift te geven. Een goede vuistregel is: formuleer bij voorkeur minimaal 2 en maximaal 5 afleiders. Kies liever voor een driekeuzevraag met 2 goede afleiders dan voor een vierkeuzevraag met een slechte derde afleider (zie ook constructievoorschrift Af-3).



**Tip: Gebruik verwant materiaal voor de afleiders**

Als u naar een omschrijving, definitie, kenmerk of formule van een bepaald begrip vraagt, kunt u als afleiders de omschrijvingen, definities, kenmerken of formules van verwante begrippen gebruiken. Let er daarbij op dat de omschrijvingen die u voor de afleiders gebruikt qua formulering vergelijkbaar zijn met de sleutel en met elkaar. Ook kunt u een omschrijving in de stam gebruiken en als afleiders de gebruikte termen voor verwante omschrijvingen.

Welke term gebruikt men voor het met een kaaskorst bedekken van een gerecht?

- A. flamberen
- B. gratineren (\*)
- C. marineren
- D. pocheren

Hetzelfde gaat op voor redenen, verklaringen, enzovoort. De beste afleiders vindt u vaak in verwante principes of redeneringen. Zorg er wel voor dat uw afleiders geen verwarring bij de kandidaat veroorzaken: ze moeten voldoende ver van het juiste antwoord afliggen!

Het ligt soms voor de hand om een synoniem van een alternatief te gebruiken als extra alternatief. Maar een handige kandidaat ziet dan vaak dat als één van beide alternatieven het juiste antwoord zou zijn, het andere alternatief ook juist moet zijn. Dat kan natuurlijk niet bij een één-uit-meervraag, dus kan de kandidaat beide afleiders wegstrepen. Oppassen dus met een dergelijke constructie! De kans is groot dat u daarmee constructievoorschrift Af-1 overtreedt.

Dus niet:

Welke term gebruikt men voor het met een kaaskorst bedekken van een gerecht?

- A. flamberen
- B. gratineren (\*)
- C. koken
- D. pocheren

Pocheren en koken hebben een vrijwel synonieme betekenis. Een kandidaat die dit weet, kan beide alternatieven wegstrepen.



**Tip: Gebruik logische fouten voor de afleiders**

Natuurlijk kunt u ook gebruikmaken van (denk)fouten die de kandidaten bij het betreffende onderwerp vaak maken. Bij rekensommen bijvoorbeeld is een afleider vaak eenvoudig te bedenken door gebruik te maken van veelvoorkomende rekenfouten. Laat u niet te veel meeslepen in het bedenken van (reken)fouten, en let erop dat de afleiders ook nog logisch blijven. U loopt anders het risico dat uw denkstappen ervoor zorgen dat uw afleider geen goed antwoord op de vraag meer kan zijn of inhoudelijk niet meer aansluit bij de vraagstelling.

Dus niet:

Het is 9:00 uur in de ochtend. Jan moet om 10:00 op plaats B zijn. Om er te komen, moet hij vanaf plaats A 10 minuten met de bus, 20 minuten met de trein en 5 minuten lopen.

Hoe laat moet Jan vertrekken van plaats A om op tijd op plaats B te arriveren?

- A. om 9:25 uur (\*)
- B. om 9:35 uur
- C. om 10:35 uur

De maker van deze vraag heeft gedacht met afleider C een logische rekenfout te gebruiken, door de tijden bij de gewenste aankomsttijd op te tellen in plaats van af te trekken. De kandidaat zal echter alternatief C zeer waarschijnlijk direct wegstrepen, nog voor hij met rekenen begint. De vertrektijd ligt namelijk later dan de gewenste tijd van aankomst. Als u na het bedenken van de afleiders nogmaals kritisch naar uw vraag kijkt, voorkomt u dit soort missers.

De gecorrigeerde  $p$ -waarde is altijd lager dan de gewone  $p$ -waarde (met uitzondering van de situatie waarin geldt:  $p=1$ ). De gecorrigeerde  $p$ -waarde kan zelfs negatief zijn: dan hebben minder kandidaten het antwoord goed beantwoord dan op basis van de raadkans verwacht kan worden.

De  $p$ -waarde van een driekeuze-opgave is 0,25. De gecorrigeerde  $p$ -waarde van dit item is  $-0,13$  ( $0,25 - (0,75/2)$ ).

Als u een toets hebt afgenomen onder de conditie 'correctie voor raden' (zie paragraaf 6.5), dan is het berekenen van de  $p(\text{corr})$ -waarde uiteraard niet relevant, omdat er geen sprake is van een raadkans die moet worden gecorrigeerd.

De moeilijkheid van polytome items wordt berekend door de gemiddelde score op de vraag te delen door de maximaal te behalen score. Het is dus *de proportie door de kandidaten behaalde punten*. De moeilijkheid van polytome vragen duiden we aan met  $p'$ -waarde. De  $p'$ -waarde kan negatief zijn en is maximaal 1.

De maximaal te behalen score op item 4 is 1 punt. Het is een matchvraag die gedeeltelijk wordt gescoord. De gemiddelde score op het item is 0,7. De  $p'$ -waarde van het item is 0,7 ( $0,7/1$ ). Item 5 is een open vraag waarvoor maximaal 5 punten te behalen zijn. De gemiddelde score van de kandidaten is 2 punten. De  $p'$ -waarde van het item is 0,4 ( $2/5$ ). Item 6 is een vraag waarbij een fout antwoord wordt bestraft met een punt aftrek. Voor een goed antwoord krijgt een kandidaat 2 punten. De kandidaten hebben een gemiddelde score van  $-0,24$  punt behaald. De  $p'$ -waarde van het item is  $-0,12$  ( $-0,24/2$ ).

Voor open vragen kan vanzelfsprekend geen  $p(\text{corr})$ -waarde berekend worden. Voor vragen met gedeeltelijke scoring is dit wel mogelijk. De berekening hiervan vindt via relatief ingewikkelde formules plaats, die wij hier verder niet behandelen. TestVision berekent deze waarden natuurlijk wel.

De  $p$ -waarde van een item ligt bij voorkeur ergens in het midden tussen 0 en 1. Als de  $p$ -waarde te dicht bij 0 of 1 ligt, geeft het item weinig informatie. De vraag is te makkelijk of te moeilijk, en kan geen goed onderscheid maken tussen kandidaten met een verschillende vaardigheid of kennisniveau. Een item heeft een maximale onderscheidende bijdrage bij een  $p$ -waarde van 0,5. Maar het is ook weer niet gewenst dat alle items in een toets een  $p$ -waarde van 0,5 hebben. Als al die items een perfect onderscheid tussen de goede en zwakke kandidaten zouden maken, dan krijgt de helft van de kandidaten een toetsscore 0 en de andere helft de maximale toetsscore en is er dus geen enkele rangorde zichtbaar onder de kandidaten. U kunt dan ook het best streven naar  $p(\text{corr})$ -waarden die rond de 0,5 liggen (bijvoorbeeld tussen 0,4 en 0,6). Daarnaast kunt u er voor kiezen om enkele makkelijke items op te nemen met een hoge  $p$ -waarde om de kandidaten te motiveren.

Bij de niet gecorrigeerde  $p$ -waarden moet rekening worden gehouden met de raadkans en ligt de ideale  $p$ -waarde hoger. Een item is veel te moeilijk als de  $p$ -waarde lager is dan de raadkans. De gecorrigeerde  $p$ -waarde is dan 0 of lager.

aantal alternatieven	ideale $p$ -waarde zonder correctie voor raden	ondergrens voor $p$ -waarde zonder correctie voor raden	ideale $p$ -waarde met correctie voor raden of ideale $p(\text{corr})$ -waarde
twee	rond 0,75	0,5	rond 0,5
drie	rond 0,67	0,33	rond 0,5
vier	rond 0,63	0,25	rond 0,5

Tabel 7.2: Ideale  $p$ -waardes voor één-uit-meervragen

Voor de open vragen ligt de ideale  $p$ -waarde altijd rond 0,50. Voor de polytome gesloten vraagvormen is de ideale  $p$ -waarde zonder correctie voor raden niet eenvoudig te berekenen, omdat ze afhankelijk is van de raadkans. Als het statistische programma dat u gebruikt de raadkans voor u berekent, dan kunt u de ideale  $p$ -waarde zelf berekenen met behulp van de volgende formule: ideale  $p$ -waarde =  $(1 + \text{de raadkans})/2$ . De ideale  $p$ -waarde ligt namelijk midden tussen de raadkans en de maximale  $p$ -waarde in.

Een matchvraag heeft een raadkans van 0,1. De ideale  $p$ -waarde van dit item is 0,55  $((1 + 0,10)/2)$ .

Bij een selectieve toets kunnen de normen voor de  $p$ -waarden anders liggen: als u de beste 5% kandidaten wilt selecteren, dan mag de  $p$ -waarde van de items relatief laag zijn, zodat u onderscheid kunt maken tussen 'goede' en 'heel goede' kandidaten. Als u echter de kandidaten wilt selecteren die een bepaalde bijscholing nodig hebben die alleen aan de 5% zwakste kandidaten wordt aangeboden, dan moet de toets juist items met relatief hoge  $p$ -waarden bevatten. Dergelijke toetsen geven vooral informatie over de kandidaten die rond de cesuur scoren. Over de andere kandidaten verkrijgt u met zo'n toets nauwelijks informatie. Bij een diagnostische toets wilt u vooral meten of de kandidaten de stof beheersen: in zo'n geval kunt u het best streven naar de hiervoor beschreven ideale  $p$ -waarden.

### 7.3.2 De kwaliteit van de afleiders

De  $a$ -waarde is de proportie kandidaten die een fout antwoord (een afleider) kiest. Elke afleider van een gesloten vraag heeft zijn eigen  $a$ -waarde. Bij een vraag met antwoordopties (zoals de matchvraag, zie paragraaf 5.2.4) heeft elke mogelijke antwoordoptie een  $a$ -waarde. Voor elk item moet de som van de  $p$ -waarde en de  $a$ -waarden gelijk zijn aan 1 (eventueel 1 minus de proportie kandidaten dat de opgave niet heeft beantwoord). Bij open vragen is er vanzelfsprekend geen sprake van  $a$ -waarden.