



## De waarde van de rit-waarde

'De rit-waarde van een examenvraag geeft in een getal weer hoe de score op die vraag samenhangt met de totale score op het examen.' Deze ingewikkelde zin zullen we in dit artikel ontleden en toelichten. De rit-waarde is een statistische maat die iets zegt over de vraag. Je kunt de rit-waarden van de vragen in de centrale examens vinden in de toets- en itemanalyse<sup>1</sup>. Het doel van dit artikel is om duidelijk te maken wat de rit-waarde vertelt. Het gaat dus om het begrijpen van de zaken die een rol spelen bij het berekenen en interpreteren van deze waarde.

### Eendimensionaal

Een examen Engels moet vaststellen hoe goed iemand in Engels is. Dat klinkt als een open deur maar is wel een fundamenteel principe dat onder toetsen en examens ligt. Het suggereert namelijk dat elk mens beschikt over iets dat we 'kennis en vaardigheid in Engels' noemen en dat we die vaardigheid in één getal kunnen vangen. We noemen dat eendimensionaal. Dat vaardigheid in Engels niet eendimensionaal is, lijkt aannemelijk omdat de vaardigheid is op te splitsen in bijvoorbeeld leesvaardigheid, spreekvaardigheid, schrijfvaardigheid etc. Natuurlijk is er wel een verband tussen al deze vaardigheden maar iedereen kent wel voorbeelden in zijn of haar omgeving van iemand die heel makkelijk en goed spreekt en een ander die daar meer moeite mee heeft, terwijl ze op een toets leesvaardigheid misschien allebei wel even goed zijn. Met andere woorden: er is wel enige samenhang tussen spreekvaardigheid en leesvaardigheid maar het verband is niet keihard. Als je een groep mensen op volgorde van oplopende spreekvaardigheid zou zetten, dan is dat dus een andere volgorde dan wanneer je ze op basis van oplopende leesvaardigheid op een rijtje zet. Net als het vak Engels is bijna elk vak een verzameling van meerdere onderdelen. Een paar voorbeelden: het vmbo-examen wiskunde bevat meetkunde en algebra, het havo-examen aardrijkskunde bevat sociale en fysische geografie, het vwo-examen biologie bevat celbiologie en ecologie. De examens zijn over het algemeen gesproken dus niet volledig eendimensionaal.

### De eisen aan een centraal examen

Het centraal examen probeert zo goed mogelijk te bepalen hoe goed iemand werkelijk is. Het liefst zouden we duizenden vragen willen stellen maar dat gaat natuurlijk niet want dat kost veel te veel tijd. Het centrale examen blijft daarom beperkt tot enkele tientallen vragen. Deze set met vragen kan niet 'zo maar' bij elkaar worden gezocht. Om de set met vragen een examen te mogen noemen moet deze set nog wel aan een aantal vakinhoudelijke en toetstechnische eisen voldoen. Een paar voorbeelden van eisen:

- De vragen moeten van het goede niveau zijn en ze moeten in overeenstemming zijn met het examenprogramma en de uitwerking voor het centraal examen in de syllabus.
- De vragen moeten verdeeld zijn over de onderdelen van de centraal examenstof van het vak.
- De vragen moeten verdeeld zijn over vraagtypen en vraagvormen.
- De vragen moeten redelijkerwijs binnen de gestelde tijd gemaakt kunnen worden door een leerling die de stof voldoende beheerst.

Al deze eisen zorgen er dus voor dat het examen niet eenzijdig is maar een grote diversiteit laat zien aan vragen en onderwerpen.

### Samenhang en correlatie

Lange volwassenen zijn gemiddeld genomen zwaarder dan kleine ('korte') volwassenen. We spreken hier van samenhang tussen lengte en gewicht. De mate van samenhang kunnen we uitdrukken in een getal: de correlatiecoëfficiënt. We geven deze coëfficiënt weer met de letter  $r$ . Om iets meer gevoel te krijgen voor dit getal volgen hier een paar voorbeelden:

- Iemand heeft heel veel kubusvormige blokjes en bouwt daarmee torentjes van één blokje per laag. Het verband tussen de hoogte van de torentjes en het aantal blokjes in de toren is 'perfect'. In dit voorbeeld geldt:  $r = 1,0$ . Groter dan 1,0 kan de correlatiecoëfficiënt niet worden.
- Er bestaat geen enkele samenhang tussen het eindcijfer voor Engels en iemands lengte. Er is geen argument te verzinnen waarom lange leerlingen beter (of slechter) in Engels zouden zijn dan kleine leerlingen. In dit geval geldt:  $r = 0,0$ .

---

<sup>1</sup> Deze toets en itemanalyse is te vinden op [www.cito.nl](http://www.cito.nl) bij de examendocumenten (opgaven, correctievoorschrift etc.)

- Er zijn ook zaken die een omgekeerde samenhang vertonen. Als het ene element groter wordt, dan wordt het andere element kleiner. Bijvoorbeeld voor mensen tussen 15 en 99 jaar geldt: hoe hoger je leeftijd, hoe lager het sociaal-mediagebruik. In dat geval is de correlatiecoëfficiënt negatief.
- Het nu volgende voorbeeld is geïnspireerd op de samenhang tussen spreekvaardigheid en leesvaardigheid Engels, waar we het hierboven over hadden. We beschouwen de samenhang tussen het SE-cijfer voor Engels en het CE-cijfer voor Engels. Opgemerkt moet worden dat het SE-cijfer Engels een combinatie is van vele onderdelen zoals bijvoorbeeld spreekvaardigheid, kijk-  
luistervaardigheid, schrijfvaardigheid en kennis van literatuur. In het CE wordt vooral leesvaardigheid gemeten. We verwachten een positief verband tussen het SE-cijfer en het CE-cijfer. Dat zou je kunnen verklaren vanuit de gedachte dat woordenschat en taalgevoel bij beide een rol spelen. Het verband tussen het SE-cijfer en het CE-cijfer blijkt redelijk sterk te zijn maar zeker niet 'perfect' (zoals bij de blokjes in het eerste voorbeeld). De waarde van de correlatiecoëfficiënt ligt hierbij zo rond de 0,7.
- In de wereld van toetsen en examens is een hoge correlatie vrij zeldzaam. Bijvoorbeeld bij leerlingen op vmbo gl/tl geldt dat de samenhang tussen de eindcijfers voor rekenen en wiskunde iets sterker is dan de samenhang tussen Nederlands en Engels. Normale waarden voor deze correlaties zijn respectievelijk:  $r = 0,4$  en  $r = 0,3$ .

### De rit-waarde

De rit-waarde vertelt iets over de samenhang tussen de score op een vraag en de score van het hele examen. Of in vakjargon: De rit-waarde is de correlatie ( $r$ ) tussen de items score en de totaalscore. Vandaar *rit*. Het zegt dus iets over de mate waarin de vraag hetzelfde meet als alle andere vragen in het examen. Wat we graag willen, is dat de goede leerling de vraag goed beantwoordt en de minder goede leerling niet. Als de score op een vraag sterk samenhangt met de totaalscore dan is dat een teken dat de leerlingen die de vraag goed hadden beantwoord uiteindelijk op een hoge(re) eindscore uitkomen en dat de leerlingen die de vraag fout hadden beantwoord, waarschijnlijk op een lage(re) totaalscore uitkomen. We zeggen dan dat de vraag goed onderscheid maakt (in vakjargon: 'discrimineert') tussen leerlingen. De rit-waarden van vragen in een examen kunnen nogal variëren. Je kunt de rit-waarden van vragen in de centrale examens vinden in de toets- en itemanalyse (tia). Deze staan op [www.cito.nl](http://www.cito.nl). In de kolom 'Rit' staan getallen variërend van ongeveer 0 tot 50. Deze getallen moet je (net als percentages) eigenlijk door 100 delen om ze te kunnen vergelijken met de correlatiecoëfficiënten zoals hierboven uitgelegd. Hoe hoger de rit-waarde, hoe beter de vraag onderscheid maakt tussen leerlingen. Deze redenering kunnen we ook vanuit de leerling bekijken. Als een leerling een vraag met een lage rit-waarde goed heeft dan zegt dat nog niet zo veel over de totaalscore die hij op het examen mag verwachten. Als hij veel vragen met een hoge rit-waarde goed heeft dan is dat een signaal dat hij een hoge(re) totaalscore op het examen mag verwachten.

### Hoe groot moet de rit-waarde zijn?

De vraag die zich nu aandient is: wat zegt de rit-waarde over de kwaliteit van een vraag. Of met andere woorden: Welke rit-waarde is goed?

De rit-waarde komt uit de statistiek en is vooral bedoeld ter ondersteuning bij het maken van toetsen. Op de [website van Cito](http://website.vanCito.nl) staat een tabel waarin betekenis wordt gegeven aan rit-waarden. Daar staat dat een rit-waarde van onder de 20 onvoldoende is. Deze tabel is gemaakt voor situaties waarin een examenmaker uit een groot aantal vragen kan kiezen en waarbij de rit-waarden al tijdens de samenstelling van het examen bekend zijn. Dit is bijvoorbeeld het geval wanneer de vragen in een proeftoets hebben gezeten. In deze situatie lijkt het beter om de vragen met een rit-waarde van onder de 20 niet op te nemen in het examen. Bij het ontwikkelen van een centraal examen beschikken we echter zelden vooraf over de rit-waarden van vragen. Hierdoor is pas na afname bekend wat de rit-waarde van de examenvraag is. Achteraf kun je dan zien dat sommige vragen een lage rit-waarde hadden. De score op deze vragen hangt dus niet heel sterk samen met de totaalscore maar toch leveren deze vragen nog steeds wel een positieve bijdrage aan het onderscheid maken tussen goede en minder goede leerlingen.

### Het effect van verschillende vaardigheden

'Maar deugt een vraag met een rit-waarde van onder de 20 wel?' Natuurlijk moet kritisch gekeken worden naar de toegevoegde waarde van een dergelijke vraag en of de vraag wel gehandhaafd kan worden. Het kan zijn dat de vraag niet deugt maar het is ook heel goed mogelijk dat deze vraag een beroep doet op een andere vaardigheid of ander kennisgebied dan de rest van het centraal examen. Een centraal examen is immers zelden volledig eendimensionaal. En dat moeten we ook niet willen. Het is uitdrukkelijk niet de bedoeling dat aan de breedte van een vak wordt geschaafd door zoveel mogelijk gelijksoortige vragen te

stellen die (dus) allemaal een hoge rit-waarde hebben. Het moet geen streven zijn om zo eendimensionaal mogelijk te zijn. Het vak moet in de volle breedte geëxamineerd worden.

Bij het oefenen met examens is het dus niet verstandig om vragen met een lage rit-waarde over te slaan met de gedachte 'dat zal wel een slechte vraag zijn'. Natuurlijk, als je met beperkte tijd zo snel mogelijk een inschatting wilt hebben hoe je ervoor staat, is het effectief om alleen vragen met een hoge rit-waarde te oefenen. Maar het structureel overslaan van alle vragen met een lage rit-waarde kan tot gevolg hebben dat je bepaalde onderdelen van het vak niet oefent die wel bevroegd kunnen worden. Kortom, een vraag met een rit-waarde van minder dan 20 hoeft geen slechte vraag te zijn. We spreken hier bewust van "hoeft" omdat er ook andere redenen kunnen zijn voor een lage rit-waarden.

### **Het effect van het niveau van de vraag**

Er is nog een andere reden waarom een vraag een lage rit-waarde kan hebben. Dit is het geval als de vraag heel erg gemakkelijk is en meer dan 95% van de leerlingen de vraag goed heeft beantwoord. Omdat bijna iedereen de vraag goed had, discrimineert de vraag dus niet zo goed. Dit geldt ook voor extreem moeilijke vragen die door minder dan 5% van de leerlingen goed zijn beantwoord. Deze vragen kunnen inhoudelijk prima in het examen passen maar ze dragen weinig bij aan het scheiden van goede en minder goede leerlingen.

### **Kwaliteit van de vraag**

Een lage rit-waarde kan ook duiden op een slechte vraag. Een dubbelzinnigheid, een vage omschrijving of alternatieven bij een meerkeuzevraag die niet van elkaar te onderscheiden zijn. Daarnaast zijn er nog allerlei redenen waarom het antwoordgedrag van de leerlingen anders uitpakt dan de examenmakers vooraf hadden voorzien. Achteraf moet dan geconcludeerd worden dat de vraag niet deugt. Examenmakers gebruiken de rit-waarde dus om slechte vragen op het spoor te komen. Tijdens de normeringsvergadering worden deze vragen dan extra zorgvuldig onder de loep genomen om te onderzoeken of er gecompenseerd moet worden. In de praktijk blijkt echter dat zeker niet elke vraag met een lage rit-waarde een onvolkomen vraag is, waarvoor gecompenseerd moet worden.

### **Negatieve rit-waarde**

Ten slotte nog een opmerking over een speciale situatie: soms blijkt bij de analyse na afloop van de afname dat een vraag een negatieve rit-waarde heeft. Hier moet heel kritisch bezien worden of deze vraag wel gehandhaafd kan worden. Een dergelijke situatie kan bijvoorbeeld duiden op een 'strikvraag'. Goede leerlingen denken dan te ver door en gaan de mist in. Toch kan het voorkomen dat een dergelijke vraag uiteindelijk niet als ondeugdelijk wordt bestempeld omdat het bijvoorbeeld zo kan zijn dat deze vraag echt een ander onderwerp toetst dan alle andere vragen. Als een dergelijke vraag binnen het examenprogramma past dan kan om die reden de vraag toch gehandhaafd blijven.

### **Epiloog**

Het is niet zo moeilijk om een examen te maken waarin alle vragen een rit-waarde van 100 hebben: Gewoon tien keer dezelfde vraag stellen. Die 100 komt er trouwens alleen maar uit als de leerlingen ook steeds hetzelfde antwoord blijven geven. Toch zit er in dit onzinnige voorbeeld een belangrijke boodschap verstopt: in het algemeen kun je namelijk stellen dat hoe groter de diversiteit van de vragen in een examen is, hoe lager de rit-waarden zullen zijn.