

Normering van het examen natuurkunde vwo 2016

Met een vooruitblik naar 2017

In 2016 is het examen natuurkunde vwo afgenomen volgens het nieuwe programma. In een vorig nummer van dit tijdschrift hebben we stilgestaan bij dit examen, de inhoudelijke opmerkingen en de aanvullingen op het CV. In dit artikel richten we ons op de statistische analyse en de totstandkoming van de normering. Ook brengen we onder de aandacht welke gegevens u over het afgelopen examen kunt downloaden op de website van Cito, hoe deze te interpreteren en te gebruiken bij uw eigen toetsen.

Het inschatten van het niveau van de leerlingen voor het examen natuurkunde vwo 2016 was lastiger dan voorheen, omdat het een nieuw programma betrof met een gewijzigde syllabus. Vooraf waren afspraken gemaakt over de normering, namelijk dat een examen dat niet zo goed scoorde, gecompenseerd zou worden in de N-term. Deze afspraak was ook nodig omdat de moeilijkheidsgraad niet, zoals gebruikelijk, vooraf via een pretest geschat kon worden.

2017

Bij het examen natuurkunde vwo 2017 is dezelfde manier van normering gevolgd als in 2016. Ook nu zijn voor diverse onderdelen géén pretesten gehouden. Bij de normering is gebruikgemaakt van de gegevens in Wolf, ingestuurd door de docenten. De resultaten van de statische berekeningen kunt u teruglezen op de website van Cito. Ga hiervoor naar http://www.cito.nl/onderwijs/voortgezet%20onderwijs/centrale_examens/schriftelijke_examens_havovwo/examens_havovwo_2016/vwo_ce_tv1 en klik dan op 'TIA' bij natuurkunde vwo.

Een TIA is een toets- en itemanalyse waarin

alle statische gegevens verzameld zijn van het betreffende examen.

p-waarde

Onder de p-waarde van een vraag, een opgave, een examen, verstaan we de gemiddelde percentuele score van alle leerlingen op het betreffende onderdeel. Als een vraag drie punten heeft en de gemiddelde score van alle leerlingen is twee, dan komt dat overeen met een p-waarde van 67. Bij een toets zoals dit examen met een maximumscore van 76 was de gemiddelde score van alle leerlingen 40,2. Dit leverde een p-waarde van 52,9. U herkent in de TIA de omschrijvingen zoals deze ook zijn opgenomen in de groepsrapportage van uw leerlingen die u van Cito krijgt. Daarop is te zien dat de verschillende opgaven zeer verschillend scoorden. De p-waarden geven een eerste indruk van de eigenschappen van de toets.

Rit-waarde

Bij de statische analyse wordt ook gekeken in welke mate een vraag discrimineert. Dat betekent in hoeverre de vraag het onderscheid blootlegt tussen meer en minder vaardige

leerlingen. Dit wordt aangegeven met de zogenaamde Rit-waarde. Rit-waarde staat voor *Response Item-Totaal correlatie*, een getal tussen 0 en 1. Hoe hoger dit getal, des te meer discrimineert de vraag. In de TIA vindt u van elke vraag de Rit-waarde. Is deze waarde erg laag, of zelfs negatief, dan is er met de vraag iets aan de hand. Een hoge Rit-waarde betekent dat vaardige leerlingen de vraag goed maken en minder vaardige leerlingen minder goed. Een negatieve Rit-waarde betekent dat de vaardigere leerlingen de vraag minder goed scoren dan de minder vaardigen. Dan is het van belang om uit te zoeken wat er met de vraag aan de hand is. Tijdens het proces van normering worden alle vragen met lage Rit-waarden nogmaals bekeken. De discriminerende werking van een vraag kan ook worden weergegeven middels nevenstaande figuur. Deze figuur gaat over de laagst-scorende vraag uit het examen; vraag 15, de 5-puntsvraag over de schaalfactoren. Zie figuur 1 voor de vraag en figuur 2 voor de statistische figuur.

Vraag 15

Deze vraag had een p-waarde van 20. In figuur 2 is de hele populatie van alle kandi-

Als je een grote en een kleine vogel vergelijkt, nemen we aan dat van de grote vogel alle afstanden (lengte, breedte, hoogte en dus ook slag grootte) k keer zo groot zijn als die van de kleine vogel. We spreken dan van schaalfactor k . De drie grootheden f , d en v in formule (1) hangen van k af. Deze afhankelijkheid (schaalwet) geven we aan met:

$$\dots \propto k^p$$

Hierin is:

- \propto evenredig met;
- k schaalfactor;
- p een getal, afhankelijk van de betreffende grootheid.

De vliegsnelheid hangt alleen af van de massa m en de

vleugeloppervlakte A van de vogel: $v \propto \sqrt{\frac{m}{A}}$.

De schaalwet voor de slagfrequentie luidt: $f \propto k^p$.

Sp 15 Voer de volgende opdrachten uit:

- Laat zien dat $v \propto k^{\frac{1}{2}}$.
- Bereken hiermee hoe groot getal p is in $f \propto k^p$.
- Vul de zin op de uitwerkbijlage aan.

Figuur 1: Vraag 15 van het examen n.a.v. 2016-1

daten verdeeld in vier groepen: de minste vaardige 25% (1), de vaardigste 25% (4) en de twee groepen daartussen. Dit is aangegeven met de cijfers 1 tot en met 4. Je zou deze gemiddeldes met staafdiagrammen kunnen weergeven en misschien wel moeten aangeven, maar dat zou ten koste gaan van de duidelijkheid. Van elk van de vier groepen (kwartielen) wordt de p-waarde berekend en weergegeven in de figuur. Als een vraag goed discrimineert, moet dit resulteren in een oplopende lijn. Dit is bij deze opgave het geval. Te zien is dat in de laagste groep minder dan 10% scoorde, maar in de hoogste groep bijna 50%. Dit is een voorbeeld van een vraag die ook onderscheid maakt tussen de vaardigere leerlingen: scoor je een 9 of een 10. Dat klopt ook want uit de TIA is af te lezen dat 7% van de leerlingen voor deze vraag de maximumscore van 5 punten behaalde, de bijbehorende Rit-waarde is extreem hoog.

Deze vraag is een voorbeeld waarmee we laten zien dat we de vernieuwing niet langzaam invoeren, maar meteen een stevige vraag over schaal en schaalafhankelijkheid stellen. In de Pilotexamens zijn overigens al een aantal vragen over dit onderwerp gesteld.

Trillingen binnen een molecuul

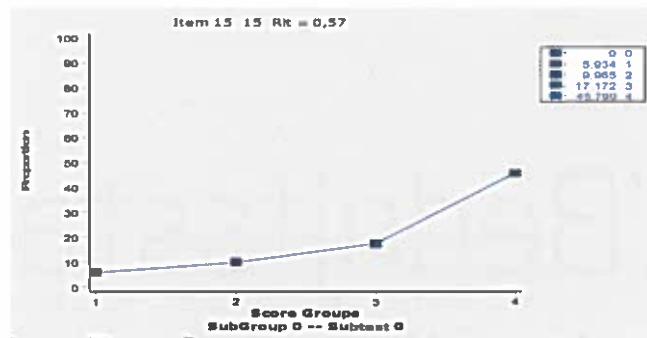
In de tabel rechtsboven staan de data van de vragen uit deze opgave. De laatste drie vragen behoren tot het domein Quantumwereld. Met name de score van vraag 19 viel veel lager uit dan vooraf ingeschat. Deze vraag hadden wij ingeschat als een repro-

ductievraag van basiskennis, maar is niet als zodanig ervaren. Als we deze vraag hadden kunnen pretesten, was dit vooraf mogelijk duidelijk geworden. Er zijn op een aantal pilotscholen wel een paar vragen bij een beperkt aantal leerlingen gepretest, maar daar zat deze vraag helaas niet bij.

Hoe hoger de **Rit-waarde** is, des te meer de vraag discrimineert

Normering

Bij de normering is uitgegaan van deze analyses. Deze analyses en de analyse van de gevolgen van de nieuwe slaag-zak regeling, resulteerden in de hoge N-term van 1,7. Daarnaast is, na de afname, het examen voorgelegd aan een panel van vijftien docenten. Aan hen werd gevraagd om per opgave het aantal scorepunten te schatten dat overeenkomt met een 5,5. Eerst maakten zij individueel een schatting, daarna konden ze



Figuur 2: Bijbehorende statische figuur

Vraag	max	p-waarde	Rit
16	3	64	38
17	3	34	45
18	4	42	48
19	3	28	42
20	2	67	39

Gegevens vraag 19

die, na discussie in de groep, bijstellen. Deze schatting werd later omgerekend naar een N-term. Deze kwam overeen met de uiteindelijke N-term. Het was natuurlijk niet zo dat de docenten in het panel konden zien welke N-term hun schattingen opleverde. Je zou dit als een soort validering door het onderwijsveld van de normeringsprocedure kunnen beschouwen. Deze methode van standard-setting is ook in 2017 toegepast. Er is een uitgebreide uitleg over het vaststellen van de N-term verschenen van het CvTE: de Normeringssystematiek van het Centrale examen vo. Daarin staan de besproken procedures van de normering en de aanpassing aan de nieuwe slaag- en zakregeling. Dit stuk is begin december aan alle examensecretarissen toegestuurd en is te vinden op www.examenblad.nl.

Ten slotte

Veel van onze examenvragen worden hergebruikt in PTA-toetsen en schoolexamens. Als u dat doet, kunt u gebruikmaken van de gegevens uit de TIA's. Na afname van de examencampagne zijn deze TIA's beschikbaar op de vermelde website van Cito. We vertrouwen erop met dit artikel en de publicatie van CvTE die eraan ten grondslag ligt, te verduidelijken hoe bij de normering gebruikgemaakt wordt van alle beschikbare gegevens. En dat daarbij een afweging gemaakt wordt volgens een procedure die vooraf vastligt, om daarmee te komen tot een eerlijke normering. ●