

TOELICHTING TOETSGBEUK

INLEIDING

De toetsen op deze site kunnen op verschillende manieren worden toegepast. Ze zijn vooral geschikt als indicatie voor de leerprestaties van leerlingen.

In het algemeen geldt dat we terughoudend moeten zijn wat betreft de psychometrische kwaliteit van de toetsen zoals ze hier worden gepresenteerd. Doorgaans zijn de instrumenten in andere, goed gereglementeerde omstandigheden afgenomen. Dat geeft geen garantie voor gebruik in andere omstandigheden, met andere populaties en met een andere, aangepaste samenstelling.

Met deze waarschuwing in het achterhoofd kunnen ze echter wel degelijk goede diensten bewijzen, als men wil weten hoe het staat met de kennis en kunde van leerlingen op de betreffende gebieden. Ze kunnen wat extra's bieden vergeleken met toetsen die door de leerkracht zelf moeten worden gemaakt. De opgaven zijn doorgaans al goed onderzocht op hun werkzaamheid. De somscore kan vaak vergeleken worden met een landelijk beeld.

De inzet van dit materiaal is vooral zinvol in samenhang met andere informatie die de leerkracht heeft over de leerlingen en het gegeven onderwijs in de betreffende onderwerpen.

HET MATERIAAL

De toetsen die uit de publicaties van PPOK zijn gedestilleerd bevatten opgaven die zijn geordend van makkelijk naar moeilijk. Het zijn opgaven die als voorbeeld hebben gediend bij de publicatie van de uitkomsten van peilingsonderzoek. Ze laten zien hoe het staat met het niveau van de leerlingen in Nederland in groep 8 van de basisschool. Het is daarom een representatieve verzameling van opgaven rond het landelijk gemiddelde. Dit gegeven maakt het mogelijk de uitkomsten op deze toetsen te interpreteren. Er kan worden vergeleken met dat landelijk gemiddelde.¹ Bovendien kunnen we de scores daarom voorzien van een niveau-indicatie.

Omdat de opgaven zijn gepubliceerd ter illustratie van het landelijke niveau zijn ze in volgorde van moeilijkheid geplaatst. Deze volgorde zorgt ervoor dat bij afname de kans op een goed antwoord van de leerling steeds kleiner wordt.² De gegeven volgorde kan worden gewijzigd, als men de toets als Word-document bewerkt. Er kan ook worden gekozen voor het pdf-format, maar dan is bewerking lastiger.

Bij de meeste studietoetsen wordt een volgorde gekozen waarin de moeilijkheidsgraad van de opgaven varieert. Soms is dat omdat er vooraf nog geen informatie is over de moeilijkheidsgraad van de opgaven. Vaak is dat echter ook omdat men de leerlingen een gevarieerde ervaring wil bieden met een afwisseling van succesansen. Er bestaat immers altijd de kans dat leerlingen, die de vragen steeds moeilijker vinden, voortijdig afhaken. Dat zou dan een onbetrouwbaar beeld kunnen geven. Als de opgaven worden aangeboden in de volgorde zoals ze nu staan dan moet bij de instructie dus duidelijk worden aangegeven dat alle opgaven moeten worden gemaakt, ook de laatste. Dat kan voor een aantal leerlingen lastig zijn.

¹ Bij PPOK betreft het meestal de prestaties van leerlingen aan het eind van de basisschool.

² Dit hangt uiteraard ook af van het onderwijsaanbod dat de betreffende leerlingen bij het betreffende onderwerp hebben genoten.

AFNAMECONDITIES

De toetsen kunnen worden afgedrukt en ingevuld zoals ze hier te vinden zijn. Het is ook mogelijk een antwoordblad te maken. Let dan wel op de opgaven met invulvelden of schrijfruimte.

Om de uitkomsten van de toets te kunnen vergelijken met de landelijke steekproef, moet de afname in omstandigheden plaatsvinden die vergelijkbaar zijn met die van het landelijke onderzoek. In dat geval geeft u de leerlingen de volgende aanwijzingen:

- Geef aan hoe ze de toets moeten maken:
 - invullen in het toetsboekje
 - Elke leerling werkt voor zich zelf
- Bespreek de afname: het aantal toetsen en de verwachte tijdsduur (deze is doorgaans ruim bemeten).
- Spreek af met de leerlingen wat ze moeten doen als ze klaar zijn. Zorg ervoor dat ze wat anders kunnen gaan doen, terwijl de andere leerlingen doorwerken, zodat ze elkaar niet storen.
- Eventuele tussentijdse vragen van leerlingen over opgaven beantwoordt u in de geest van:
 - Je moet invullen wat je denkt dat goed is;
 - Als je het echt niet weet, mag je de vraag overslaan.

SCORING EN INTERPRETATIE

Elke toets is voorzien van een document waarin de sleutels en de populatiewaarden worden gegeven (zie website Testadvisor). Ook is er een scoretabel waarmee de prestaties van de leerlingen kunnen worden vergeleken met de landelijke steekproef.

2

MAAK EEN REKENBLAD

Om inzicht te krijgen in de prestaties is het handig de resultaten bij elkaar in een overzicht te zetten. Dat kan het beste in een rekenblad in bijvoorbeeld Excel (horizontaal: de opgaven, verticaal: de namen van de leerlingen). In dat rekenblad geeft u per opgave per leerling aan of de vraag goed of fout is gemaakt. In de meeste gevallen geeft u een 1 of 0, voor respectievelijk een goed of fout antwoord. Als een leerling niets heeft ingevuld, noteert u ook een 0. We gaan er daarbij vanuit dat de leerling voldoende tijd had om elke opgave te maken. Vervolgens sommeert u over rijen en kolommen om totalen te krijgen en kunt u gemiddelden berekenen. Reserveer ook een paar rijen en kolommen voor het overnemen van de p-waarden uit het landelijke onderzoek en de eigen berekening (zie voorbeeld).

TOTAALSCORE PER LEERLING

Van elke leerling kunt u de totaalscore bepalen. Als u een overzicht gemaakt heeft in Excel staan deze totaalscores in de laatste kolom. U kunt de scores van uw leerlingen vergelijken met de landelijke uitkomst. Deze uitkomst is verdeeld in vier segmenten: Ver boven het gemiddelde- boven het gemiddelde- onder het gemiddelde – ver onder het gemiddelde. Deze indeling komt grotendeels overeen met de indeling in niveaus bij het LOVS van Cito in A, B, C en D/E. Aan de hand van deze tabel kunt u dus vaststellen in welk segment de leerling valt met zijn of haar score.

RESULTAAT VAN DE GROEP

Vervolgens telt u het aantal leerlingen dat in dat segment valt. Dat vergelijkt u met het landelijke beeld. Daar bevindt zich in elk segment 25%. Als bijvoorbeeld blijkt dat in uw groep meer dan de helft van de leerlingen in de segmenten 1 en 2 valt, steekt dat ongunstig af ten opzichte van het landelijke beeld: uw groep bezit minder kennis op het betreffende terrein dan landelijk is gemeten.

DE P-WAARDE

U kunt de gemiddelde score van de groep berekenen door de som van de totaalscores van de leerlingen te delen door het aantal leerlingen. Deze gemiddelde score zegt echter niet zo veel. Het is gebruikelijk om in plaats daarvan de zogenoemde p-waarde te bepalen, zowel van een opgave als van de hele toets.

De **p-waarde van een opgave** krijgt u door de scores van de leerlingen op te tellen en dat getal te delen het product van de maximale score van de opgave en het aantal leerlingen. Bij een opgave waarvoor de maximale score 1 is, is de p-waarde dus gelijk aan het aantal goede antwoorden gedeeld door het aantal leerlingen. In dat geval is het cijfer gelijk aan het percentage leerlingen dat de opgave correct heeft beantwoord.

Bij een opgave waarbij de maximale score bijvoorbeeld 3 is, is het iets ingewikkelder. Stel dat van de tien leerlingen er vier 3 punten, twee leerlingen 2 punten en twee leerlingen 1 punt halen (en de overige twee hebben 0 punten), dan is de p-waarde: $(4 \times 3 + 2 \times 2 + 2 \times 1) : (3 \times 10) = 18 : 30 = 0,60$. De p-waarde van een opgave is altijd een getal tussen 0,00 en 1,00. Een gemakkelijke opgave wordt door veel leerlingen goed gemaakt en krijgt dus een hoge p-waarde. Een moeilijke opgave heeft een lage p-waarde.

De **p-waarde van de toets** is gelijk aan het gemiddelde van de p-waarden van de opgaven. In het geval dat de maximale scores van de opgaven gelijk zijn is dat eenvoudig uit te rekenen. Bij toetsen waarin opgaven voorkomen met een maximumscore van meer dan 1 punt is de berekening wat complexer. Bij het bepalen van de p-waarde van de toets tellen deze opgaven zwaarder mee: een opgave met een maximumscore van 2 punten moet tweemaal worden meegeteld, een opgave met een maximumscore van 3 punten moet driemaal worden meegeteld. De p-waarde van de toets geeft aan hoe moeilijk de toets was voor de hele groep leerlingen. Als de p-waarde hoog is, is de toets gemakkelijk voor deze groep. Is die p-waarde lager dan bijvoorbeeld 0.50, dan hadden veel leerlingen daar moeite mee.

3

VERGELIJKING VAN UW GROEP MET DE LANDELIJKE P-WAARDEN

In de tabellen bij de toetsen zijn ook de p-waarde per opgave aangegeven, zoals gemeten in de landelijke PPON-peiling. Hiervoor kunt u ook een gemiddelde berekenen. Dat kunt u vergelijken met dat van uw eigen groep. De p-waarden van de opgaven en de gemiddelde p-waarde van elke toets geven aan hoe de toets gemaakt zou zijn door een representatieve groep leerlingen. Door die cijfers te vergelijken met uw eigen uitkomsten kunt u zowel per opgave eventuele hiaten in kennis opsporen, als ook per hele toets het kennisniveau van uw groep vergelijken met de landelijke situatie. Uiteraard kunt u ook per leerling nagaan of zich specifieke kennishiaten voordoen.

TOT SLOT

Deze toetsen dienen slechts als voorbeeld. De uitkomsten moeten met de nodige voorzichtigheid worden geïnterpreteerd en zijn vooral indicatief.

Het aantal opgaven per toets is vaak te klein om van een statistisch betrouwbare uitkomst te kunnen spreken³. Daarnaast is de grootte van de groep waar deze toetsen worden afgenomen van belang. Hoe meer leerlingen de toets hebben gemaakt, hoe beter de uitkomsten bruikbaar zijn. Bij een kleine groep kan men minder betekenis aan de uitkomst hechten dan bij een grote groep. In het peilingsonderzoek, dat als bron dient voor

³ De toets kan sterker gemaakt worden door het aantal opgaven te vergroten. Dat is bijvoorbeeld te realiseren door toetsen met aansluitende inhoud te combineren (bijv. Biologie of Natuurkunde en Techniek)).

de landelijke vergelijking, zijn de opgaven afgenomen bij grote groepen leerlingen (ca. 200 tot 600) en gaat het om langere toetsen.

Tot slot is het uiteraard de vraag in hoeverre de inhoud van de toets aansluit op het onderwijsaanbod voorafgaande aan de afname. Sluit het aangeboden onderwijs niet goed aan dan kan interpretatie van de uitkomsten moeilijk zijn. Als het onderwerpen betreft die op uw school (nog) nauwelijks aan bod zijn geweest, meet u vooral wat leerlingen langs andere weg al of niet hebben meegekregen. Daar kunnen leerlingen sterk in verschillen. U evalueert dan niet het effect van uw eigen onderwijs, maar meer wat leerlingen weten en kunnen, ongeacht waar dat vandaan komt.