

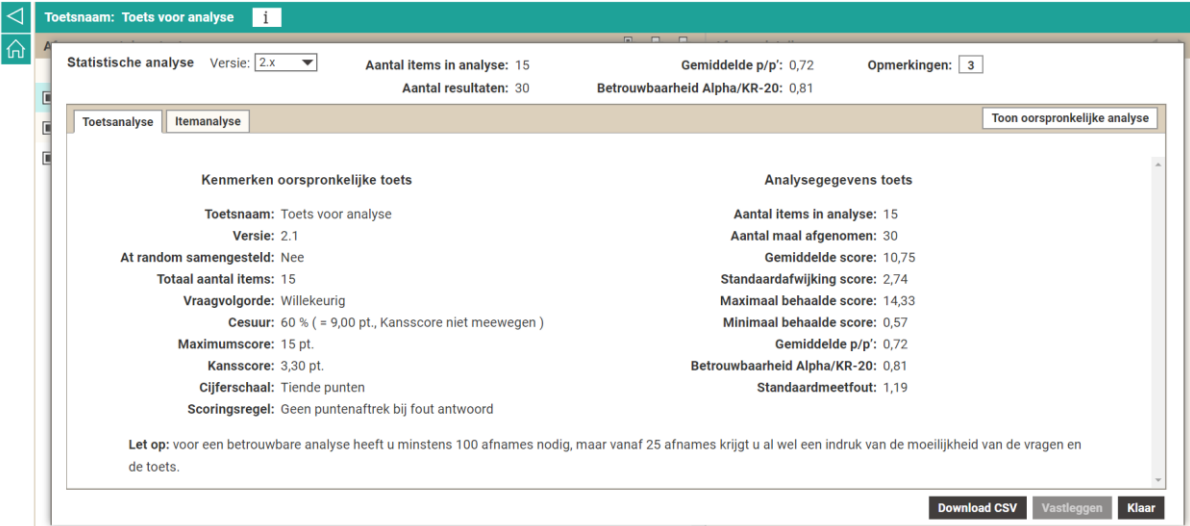
## Statistische analyse in TestVision

Dankzij digitaal toetsen is het eenvoudig mogelijk om analyses uit te voeren op de toetsresultaten. We beschrijven hier de belangrijkste mogelijkheden in TestVision.

In het domein *Demo* (<http://hu.testvision.nl/online/ontwikkelaar>) van TestVision kun je de analysegegevens van een voorbeeldtoets bekijken. Deze vind je in de map *Voorbeeldanalyse* in het onderdeel *Gepubliceerde toetsen*.

Toetsanalyse in TestVision .....	1
Itemanalyse in TestVision.....	2
Quick scan voor het beoordelen van de kwaliteit van de items .....	3

## Toetsanalyse in TestVision



Statistische analyse    Versie: 2.x    Aantal items in analyse: 15    Gemiddelde p/p: 0,72    Opmerkingen: 3  
Aantal resultaten: 30    Betrouwbaarheid Alpha/KR-20: 0,81

Kenmerken oorspronkelijke toets	Analysegegevens toets
Toetsnaam: Toets voor analyse	Aantal items in analyse: 15
Versie: 2.1	Aantal maal afgenomen: 30
At random samengesteld: Nee	Gemiddelde score: 10,75
Totaal aantal items: 15	Standaardafwijking score: 2,74
Vraagvolgorde: Willekeurig	Maximaal behaalde score: 14,33
Cesuur: 60 % (= 9,00 pt., Kansscore niet meewegen )	Minimaal behaalde score: 0,57
Maximumscore: 15 pt.	Gemiddelde p/p: 0,72
Kansscore: 3,30 pt.	Betrouwbaarheid Alpha/KR-20: 0,81
Cijferschaal: Tiende punten	Standaardmeetfout: 1,19
Scoringsregel: Geen puntenafrek bij fout antwoord	

Let op: voor een betrouwbare analyse heeft u minstens 100 afnames nodig, maar vanaf 25 afnames krijgt u al wel een indruk van de moeilijkheid van de vragen en de toets.

Download CSV    Vastleggen    Klaar

In TestVision kun je voor een afgenomen toets eenvoudig de toets- en itemanalyse bekijken. Op het tabblad *Toetsanalyse* zie je de toetsgegevens. In de linkerkolom zie je de gegevens die de toetsontwikkelaars van tevoren heeft ingesteld of van tevoren vastliggen, zoals de maximumscore, de kansscore en de cesuur. Rechts zie je de analysegegevens. Vooral interessant zijn daar natuurlijk de moeilijkheid, de betrouwbaarheid en de standaardmeetfout van de toets. Maar het is goed om ook te kijken wat de maximaal behaalde score is. Is er iemand die de maximale score heeft behaald? Dat geeft vertrouwen dat alle vragen voldoende duidelijk waren.

Wat de gewenste moeilijkheid is van je toets, hangt mede af van het doel ervan. Meestal is een moeilijkheid tussen 0,6 en 0,8 een prima resultaat. Voor de betrouwbaarheid hanteert men in de

literatuur vaak als norm dat deze minimaal 0,8 moet zijn. In de praktijk is dit vaak moeilijk haalbaar. Maar bij een betrouwbaarheid lager dan 0,6 is het toetsresultaat zo onbetrouwbaar dat je moet oppassen met de waarde die je eraan hecht. Een kandidaat die rond de cesuur scoort en is geslaagd, had dan net zo goed gezakt kunnen zijn (en andersom).

## Itemanalyse in TestVision

Vraagnaam	Vorm	MaxSc	Kans	D/P	# afd	GemSc	p/p'	p(corr)	r <sub>it</sub>	r <sub>ir</sub>	#	a-waarde	
1 EUM - Putdeksel - summatief	Eum	1	0,25	D	30	0,67	0,67	0,56	0,57	0,44	A	6	0,2
2 EUM - Voetbalstadion - formatief	Eum	1	0,17	D	30	0,93	0,93	0,92	0,80	0,76	B	20	0,667
3 EUM met video - Kensington - formatief	Eum	1	0,33	D	30	0,90	0,90	0,85	0,70	0,64	C	2	0,067
4 EUM met video - Promotie - summatief	Eum	1	0,25	D	30	0,67	0,67	0,56	0,49	0,34	D	0	0
5 Gelijke match - Organogram HU - formatief	MaG	1	0,33	P	30	0,90	0,90	0,85	0,63	0,57	-	2	0,067
6 Gelijke match - Straten/grachten - summatief	MaG	1	0,33	P	30	0,80	0,80	0,70	0,31	0,19			
7 Hotmatch - Provincies Nederland - formatief	HMa	1	0,00	P	30	0,84	0,84	0,84	0,67	0,60			
8 Invul - Fietsbellen - formatief	Inv	1	0,00	D	30	0,43	0,43	0,43	0,33	0,16			
9 Invulvraag - Nijntje - summatief	Inv	1	0,00	D	30	0,97	0,97	0,97	0,69	0,65			
10 MUM - Cabaretiers - formatief	Mum	1	0,30	P	30	0,66	0,66	0,52	0,67	0,60			
11 MUM - Woonplaatsen - summatief	Mum	1	0,34	P	30	0,53	0,53	0,29	0,72	0,65			
12 Numerieke invulvraag - Treden - formatief	InvN	1	0,00	D	30	0,23	0,23	0,23	0,33	0,18			
13 Ongelijke match - HU locaties - summatief	MaO	1	0,33	P	30	0,69	0,69	0,54	0,40	0,32			
14 Ongelijke match - Monopoly - formatief	MaO	1	0,33	P	30	0,83	0,83	0,75	0,71	0,67			
15 Rangschik - Hoogste gebouwen - formatief	Rs	1	0,33	P	30	0,68	0,68	0,52	0,54	0,42			

Op het tabblad Itemanalyse vind je de gegevens per item (toetsvraag). Naast de  $p$ -waarde (moeilijkheid van de vraag) vind je hier ook de  $r_{it}$ -waarde (het onderscheidend vermogen). TestVision geeft via een kleur aan hoe goed de vraag functioneert. Rode vragen functioneren niet goed, groene vragen zijn prima en oranje vragen zitten daartussenin. Een slecht functionerende vraag heeft bijvoorbeeld een erg lage of hoge  $p$ -waarde en is dan dus heel moeilijk of juist makkelijk. Ook een lage of negatieve  $r_{it}$ -waarde is een signaal dat er iets mis is met een vraag. Bij een negatieve  $r_{it}$ -waarde beantwoorden goede kandidaten de vraag fout en zwakkere kandidaten juist goed. Vaak blijkt zo'n vraag een strikvraag te zijn of is er een fout gemaakt bij het instellen van het correcte antwoord. Bij een-uit-meervragen zie je ook de  $a$ -waarden, de aantrekkelijkheid van de afleiders. De getallen achter het streepje -, betreft het aantal en de proportie studenten die de vragen hebben overgeslagen. Door de analysegegevens vast te leggen bij de vragen, kun je eenvoudig bij de toetsvragen zelf de vragen filteren en aanpassen. De goed functionerende vragen kun je natuurlijk mooi gebruiken als inspiratie voor nieuwe, vergelijkbare vragen!

De gecorrigeerde  $p$ -waarde ( $p(\text{corr})$ ) is interessant als je de moeilijkheid van verschillende vraagvormen met elkaar wilt vergelijken. In deze gecorrigeerde waarde is de kansscore verrekend, zodat de invloed daarvan is weggenomen. Stel dat twee vragen, driekeuze en vierkeuze, exact dezelfde  $p$ -waarde hebben, dan is een driekeuzevraag makkelijker dan de vierkeuzevraag. Dit zie je terug in de gecorrigeerde  $p$ -waarde.

In de voorbeeldtoets zijn geen open vragen opgenomen. Maar ook daarvoor berekent TestVision automatisch de itemgegevens op vergelijkbare wijze, zie het voorbeeld hieronder uit een andere toets in een andere database. Leuk detail is hier het grote verschil tussen  $r_{it}$ - en  $r_{ir}$ -waarde, dat wordt veroorzaakt door de geringe omvang van de toets (in totaal zijn er 11 punten te verdienen, waarvan 6 voor deze open vraag).

6	Beeldentuin 50 jaar	1	i	Open	6	0,00	P	3	1,00	0,17	0,17	0,89	0,05
---	---------------------	---	---	------	---	------	---	---	------	------	------	------	------

## Quick scan voor het beoordelen van de kwaliteit van de items

Itemgegevens	Oorzaak	Actie
<b><math>r_{it}</math>- en/of <math>r_{it}</math>-waarde &lt; 0?</b>	Sleutelfout	Sleutel aanpassen in TestVision
	Strikvraag of onduidelijke vraag	Vraag niet meer gebruiken en eventueel bij iedereen goed rekenen
<b><math>p</math>-waarde &lt; 0,4 en <math>r_{it}</math>-waarde &gt; 0,3?</b>	Moeilijke vraag die veel bijdraagt aan de betrouwbaarheid	Hergebruik van de vraag mogelijk  Parallele vragen maken voor nieuwe toetsen
<b><math>p</math>-waarde &lt; 0,4 en <math>r_{it}</math>-waarde &lt; 0,3?</b>	Moeilijke vraag die minder bijdraagt aan de betrouwbaarheid	Bij hergebruik eerst aanpassen: <ul style="list-style-type: none"> <li>• Vraagformulering duidelijk?</li> <li>• Voldoende relevant?</li> <li>• Afleiders echt fout?</li> </ul>
<b><math>p</math>-waarde 0,4 - 0,9 en <math>r_{it}</math>-waarde &gt; 0,3?</b>	Gemiddeld moeilijke vraag die veel bijdraagt aan de betrouwbaarheid	Hergebruik van de vraag mogelijk  Parallele vragen maken voor nieuwe toetsen
<b><math>p</math>-waarde 0,4 - 0,9 en <math>r_{it}</math>-waarde &lt; 0,3?</b>	Gemiddeld moeilijke vraag die minder bijdraagt aan de betrouwbaarheid	Bij hergebruik eerst aanpassen: Bij $p$ -waarde < 0,7 de vraagformulering en relevantie controleren. Bij $p$ -waarde > 0,7 niet-functionerende afleiders indien mogelijk vervangen
<b><math>p</math>-waarde &gt; 0,9 en <math>r_{it}</math>-waarde &gt; 0,2?</b>	Makkelijke vraag die nog wel wat bijdraagt aan de betrouwbaarheid	Kan hergebruikt worden voor het evenwicht makkelijke en moeilijke vragen in de toets
<b><math>p</math>-waarde &gt; 0,9 en <math>r_{it}</math>-waarde &lt; 0,2?</b>	Makkelijke vraag die weinig bijdraagt aan de betrouwbaarheid	Vraag niet meer gebruiken
<b><math>\alpha</math>-waarde &lt; 0,05?</b>	De afleider wordt vrijwel niet gekozen	Afleider vervangen of laten vervallen (zodat de kansscore van de vraag beter klopt)