

# Wat doet Cito met de gegevens uit Wolf<sup>1</sup>?

## Inleiding

Het verzamelen van afnamegegevens in Wolf heeft primair tot doel de normeringsbeslissingen te ondersteunen. De uitvoering van de analyses op deze afnamegegevens staat tijdens de examencampagne altijd onder tijdsdruk. Per examen is vooraf een uiterste accordeerdatum bepaald. Meteen daarna start de analyse van dat examen. Toch kunnen ook nog na die uiterste accordeerdatum gegevens worden ingezonden. De vraag is 'Wat doet Cito met die gegevens?' Naast het verzamelen van afnamegegevens voor analysedoeleinden worden mits de Dienstverlenings- en verwerkersovereenkomst is getekend nog meer gegevens uitgewisseld. Sinds de invoering van de webbased versie is het ook mogelijk om de 2e correctie in Wolf uit te voeren. Wat gebeurt er met die gegevens? Worden die ook betrokken bij de normering? In deze notitie geven we openheid van zaken.

## Ondersteuning normeringsbeslissingen

Tijdens de examencampagne staat de verwerking van de gegevens en de analyse van de resultaten onder tijdsdruk. Voor de examinatoren (in de rol van 1e corrector) wordt in het correctievoorschrift een uiterste accordeerdatum vermeld. De gegevens die tot en met die datum worden geaccordeerd, worden gebruikt als basis voor de ondersteuning van de normeringsbeslissingen. Een overzicht van de gegevens per examen wordt ontsloten in het examenverslag, en zijn te vinden in de applicatie 'resultaten per examen'.

In het examenverslag publiceren we elk jaar een overzicht van de hoeveelheid gegevens die we van de examinatoren ontvangen hebben. Voor 2018 wordt de omvang van die gegevens weergegeven in onderstaande tabel.

schooltype	ingezonden	in steekproef	vervallen	boventallig	alleen voor groepsrapportage
vwo	283.099	233.733	131	199.198	49.196
havo	350.495	294.684	282	263.600	55.529
vmbo alg. vakken	354.391	315.103	151	274.514	39.135
vmbo beroeps	52.036	47.190	44	15.545	4.801
totaal (absoluut)	1.040.021	890.710	608	752.857	148.661
in procenten	100	85,6	0,1	74	14,3

Uit de tabel blijkt dat in 2018 bijna 86% van de gegevens die we van de examinatoren ontvingen, gebruikt wordt voor de ondersteuning van de normeringsbeslissingen. Dezelfde afnamegegevens worden ook gebruikt om de kwaliteit van examens en de afzonderlijke vragen in examens te evalueren.

De laatste kolom geeft een indruk van de hoeveelheid gegevens die geaccordeerd wordt na het verstrijken van de uiterste accordeerdatum.

## Genereren groepsrapportage

Voor alle examens met een substantieel aantal deelnemers wordt een groepsrapportage gegenereerd (voor de scholen waarvoor een Dienstverlenings- en verwerkersovereenkomst is getekend). In het verleden voelden docenten zich weleens onder druk gezet omdat men van alle groepen wel een groepsrapportage wilde ontvangen, maar het niet mogelijk was om het werk van alle kandidaten voor het verstrijken van de uiterste inzenddatum gereed te hebben.

<sup>1</sup> Wolf is de naam van het webbased computerprogramma voor scholen ten behoeve van de uitwisseling van afnamegegevens met Stichting Cito.

Dat is echter niet nodig. Het is namelijk ook mogelijk kandidaten te accorderen ná het verstrijken van de uiterste accordeerdatum. De gegevens van deze kandidaten worden niet meer betrokken bij de ondersteuning van de normering. De gegevens worden wanneer ze voor de datum van publicatie van de definitieve N-termen zijn ingezonden nog wel meegenomen bij het genereren van de groepsrapportage. Zo'n 14% van de gegevens werd in 2018 na de uiterste accordeerdatum ingezonden. Stichting Cito gebruikt deze gegevens alleen om ze te betrekken bij het genereren van de groepsrapportage. De standaard groepsrapportage wordt verzonden via de e-mail en is ook te downloaden vanuit het programma.

Sinds 2016 wordt er als experiment ook een groepsrapportage in Excel gegenereerd. Deze rapportage biedt de mogelijkheid om vragen uit het examen aan of uit te zetten en zo naar eigen inzicht te clusteren. De groepsrapportage in Excel biedt ook de mogelijkheid om kandidaten al dan niet buiten het overzicht te houden en het effect daarvan op de resultaten te bezien. De groepsrapportage in Excel wordt per groep op exact dezelfde gegevens gebaseerd.

## **2e correctie**

In de webbased versie van Wolf is het ook mogelijk om als gecommitteerde de 2e correctie af te werken (mits de Dienstverlenings- en verwerkersovereenkomst is getekend). Op basis van de poolinggegevens van DUO worden groepen doorgezet naar de school voor de 2e correctie. Wanneer op de gepoolde school een gecommitteerde aan de externe groep gekoppeld wordt, kan de 2e correctie tot en met de fiattering door de 1e corrector volledig worden afgewerkt. De Wolf-accountbeheerder kan de vastgestelde scores exporteren en zo uitwisselen met het leerling-administratiesysteem. Het primaat ligt hier bij een juiste afwerking van de correctieprocedure. Wolf is in dit opzicht een service waarvan de scholen gebruik kunnen maken. Scholen zijn niet verplicht om hier gebruik van te maken.

De gegevens van de 2e corrector worden niet gebruikt bij de ondersteuning van de normeringsbeslissing en evenmin bij de groepsrapportages. Ze zijn tot nog toe slechts incidenteel gebruikt om te analyseren welke vragen relatief veel verschillen tussen beoordelaars genereren. Daar gaan we hieronder iets verder op in.

Een van de conclusies uit een onderzoek naar de 2e correctie was dat verschillen in soepelheid tussen correctoren sterk samenhangen met kenmerken van het examen (Kuhlemeier & Kremers, 2012). De onderzochte antwoordmodellen boden de correctoren vaak onvoldoende steun en droegen daarmee bij aan verschillen tussen correctoren. Verschillen in soepelheid deden zich onder meer voor als het antwoordmodel de antwoorden van de leerlingen niet goed dekt en/of de deelscores niet en alleen de maximumscore in het antwoordmodel vermeld werden.

De resultaten van het beoordelaars-overeenstemmingsonderzoek hebben voornamelijk tot doel om zichtbaar te maken welk type vragen en bijbehorend antwoordmodel verantwoordelijk zijn voor verschillen tussen beoordelaars. Deze analyses worden pas na afronding van de normering uitgevoerd en kunnen een inspiratiebron zijn om de antwoordmodellen in toekomstige examens te verbeteren.

## **Evaluatie representativiteit gegevens**

Het is van belang dat de gegevens die gebruikt worden voor de ondersteuning van de normeringsgegevens een representatief beeld schetsen van de behaalde resultaten. Dat wordt op verschillende manieren en in twee fases gecontroleerd.

Vlak na de campagne wordt bekeken of er structurele verschillen bestaan in de gegevens die voor en na de uiterste accordeerdatum worden ingezonden. Ook wordt gecontroleerd of een gewogen steekproef tot andere resultaten leidt dan de huidige manier van steekproef trekken.

Na de campagne wordt de representativiteit gecheckt aan de hand van de DUO-data.

De conclusie is dat de Wolf-data representatief zijn en vrij nauwkeurig het effect van de normering op de feitelijk behaalde resultaten voorspelt.

## **Evaluatie kwaliteit examens**

Dezelfde afnamegegevens die gebruikt worden om de normeringsbeslissingen te ondersteunen worden ook gebruikt om de kwaliteit van de examens en de afzonderlijke vragen in de examens te evalueren.

Het leveren en aantonen van de kwaliteit van toetsen en examens is een belangrijk element van onze professionele verantwoordelijkheid. Betrouwbare informatie leveren is daarbij essentieel. De informatie die wij verzamelen, benutten we om de kwaliteit van de toetsen en examens te evalueren op hun meet-technische kwaliteiten. Die evaluatie bestaat uit een kwantitatief en kwalitatief onderdeel. Het kwantitatieve deel wordt voorzien van een professioneel oordeel en voor eventuele tekortkomingen worden oplossingen gezocht.

Daarbij maken we gebruik van signaalwaarden voor kwaliteiten als moeilijkheidsgraad, betrouwbaarheid, mogelijkheid van gender bias en analyse van de non-respons. Wanneer een examen een afwijkende signaalwaarde laat zien, is dat een bron voor verdere analyse. Zo kan er uitleg gegeven worden over de omstandigheden die hebben geleid tot een afwijkende score en zeker wordt ook gekeken welke verbeteracties kunnen worden uitgevoerd. Daarbij betrekken we ook het meer kwalitatieve oordeel van de docenten zoals dat uit de quickscan (beknopte Wolf-enquête) naar voren komt.

Daarmee ondersteunen de Wolf-gegevens in een soort van jaarlijkse cyclus het proces van continue verbetering van examens.