



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

1.

# Projectplan

## Prestatiemonitor 2020

Opdrachtgever:  
Stuurgroep:  
Projectleider:

Directie Kennis  
Platform Toezicht (Plato)

**5.1.2.e**

Versie 1.1 maart 2020

## **1. Inleiding**

### **1.1 Achtergrond project**

Voor de uitvoering van onze waarborgfunctie monitoren we de prestaties van besturen en scholen. Hiervoor voeren we jaarlijks een prestatieanalyse uit. Dit doen we om tijdig mogelijke risico's ten aanzien van de onderwijskwaliteit en het financieel beheer te kunnen detecteren.

Als er op basis van de prestatieanalyse geen vermoeden van risico's is dan wordt de analyse afgesloten. Besturen worden dan niet actief over de uitkomst van de prestatieanalyse op de hoogte gesteld. Als er een vermoeden van risico's is dan voeren we een expertanalyse uit.

De prestatiemonitor is een informatieproduct dat wordt gebruikt ter ondersteuning van de prestatieanalyse. In 2017-2018 is begonnen met het ontwikkelen van de prestatiemonitor. Dit projectplan beschrijft de doelen en resultaten van de prestatiemonitor 2020. Mogelijkerwijs wordt de opdracht van dit projectplan in 2020 aangepast aan de ontwikkelingen rondom de maatschappelijke opgaven van het onderwijs.

### **1.2 Opdracht**

Zorg voor verdere ontwikkeling, productie en evaluatie van de prestatiemonitor.

#### Ontwikkeling

- Ontwikkel nieuwe prestatiemonitors voor besturen, speciaal onderwijs, middelbaar beroepsonderwijs en hoger onderwijs.
- Ontwikkel nieuwe prestatiemonitors, gemeenten kinderopvang/vve en nieuwkomers voortgezet onderwijs (onder voorbehoud)
- Zorg voor verdere ontwikkeling van bestaande prestatiemonitors.

#### Productie(wijze)

- Verbeter de visuele presentatie van de prestatiemonitor in Power BI, zodat de prestatiemonitor vanuit verschillende invalshoeken bekeken kan worden.

#### Evaluatie

- Evalueer en valideer eerder ontwikkelde prestatiemonitors.

## **2. Opzet prestatiemonitor 2020**

### **2.1 Doel**

Net als in 2018 en 2019 is het overkoepelende doel van het project om te komen tot inrichting van een prestatiemonitor die

- aansluit bij het vernieuwde toezicht
- onnodige sectorale verschillen vermijdt
- gebruiksvriendelijk is en
- optimaal gebruik maakt van beschikbare innovatieve mogelijkheden

Daarbij streven we ernaar om in 2020 een prestatiemonitor op bestuursniveau beschikbaar te maken.

## 2.2 Beoogde resultaten

Om het overkoepelende doel van dit project te bereiken zijn de volgende resultaten nodig:

### *Inhoud*

- Ontwikkelen nieuwe prestatie-monitors. Voor 2020 zijn dit
  - de prestatie-monitor besturen (bst), inclusief financiële continuïteit en doelmatigheid,
  - speciaal onderwijs (so),
  - middelbaar beroepsonderwijs (mbo)
  - en hoger onderwijs (ho).
- Ontwikkel nieuwe prestatie-monitors. Voor 2020 zijn dit, onder voorbehoud, een prestatie-monitor voor
  - gemeenten, kinderopvang/vve (ko/vve),
  - nieuwkomers voortgezet onderwijs.
- Verder ontwikkelen van prestatie-monitors die in 2019 ontwikkeld zijn. Dit betreft de prestatie-monitor
  - basisscholen,
  - speciaal basisonderwijs,
  - nieuwkomers primair onderwijs,
  - samenwerkingsverbanden po,
  - samenwerkingsverbanden vo,
  - voortgezet speciaal onderwijs,
  - voortgezet onderwijs.
- Aan de hand van de ontwikkeling van een prestatie-monitor wordt bepaald of én hoe sectorspecifieke producten blijven bestaan, die momenteel gebruikt worden voor de prestatie-analyse.
- Verkenning integratie indicatoren prestatie-monitor en stelsel-monitor. Hoe kunnen we de prestatie-monitor en stelsel-monitor meer met elkaar verbinden? En hoe kunnen we de maatschappelijke opgaven voor het onderwijs en de ijkpunten hierbij betrekken.
- Zorgen voor interne communicatie en scholingsactiviteiten binnen de toezichtdirecties.

### *Productie(wijze)*

- Visualisatie van prestatie-monitor in Power BI uitbreiden, zodat de prestatie-monitor vanuit verschillende invalshoeken bekeken kan worden.
- Verkenning van mogelijkheden om het rekenmodel uit te breiden zodat meer complexe criteria kunnen worden gebruikt.

### *Evaluatie*

- Valideren verschillende prestatie-monitors door uitkomsten expertanalyse te analyseren.

## 2.3 Afbakening

Het project betreft "alleen" de prestatie-monitor: instellingsoverzichten ter ondersteuning van de prestatie-analyse. Het project gaat niet over instellingsoverzichten in andere categorieën, zoals de verdiepingsinformatie en de beoordelingsmodellen. Het project gaat ook niet over informatieproducten op hogere aggregatieniveaus, zoals de stelsel-monitor.

## **2.4 Relatie met andere projecten**

Dit project heeft een relatie met verschillende andere projecten.

- **Onderwijsresultaten**

Onderwijsresultaten zijn een belangrijk onderdeel van de prestatiemonitor. De verdere ontwikkeling van deze modellen heeft dan ook direct consequenties voor de inhoud van de prestatiemonitor.

- **Verdiepingsinformatie**

De prestatiemonitor hangt samen met het project verdiepingsinformatie. Beiden maken namelijk deels gebruik van dezelfde databronnen. Afstemming tussen beide projecten vindt binnen de directie Kennis plaats. Daarnaast fungeert de klankbordgroep prestatiemonitor ook als klankbord voor de verdiepingsinformatie.

- **Stelselmonitor**

De prestatiemonitor hangt samen met de stelselmonitor. Beiden maken namelijk deels gebruik van dezelfde databronnen en indicatoren. Afstemming tussen beide projecten vindt binnen de directie Kennis plaats.

- **Maatschappelijke opgaven**

De ontwikkelingen rondom de maatschappelijke opgaven van het onderwijs worden meegenomen in dit project. Indien nodig wordt de opdracht van dit project aangepast.

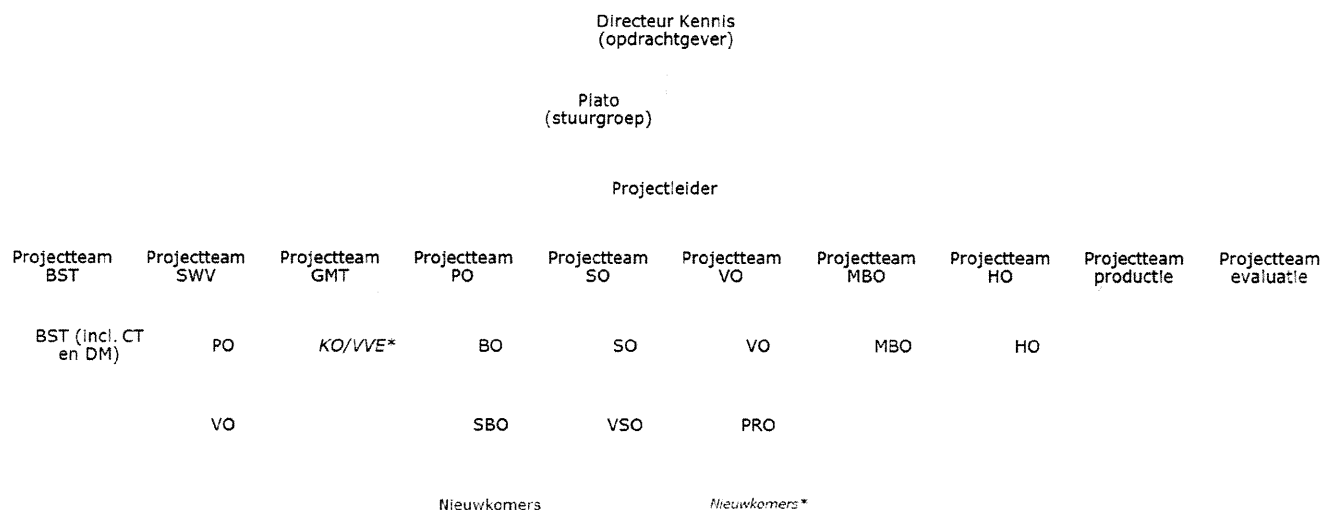


### 3. Organisatie

#### 3.1 Projectorganisatie

Het platform toezicht (Plato) treedt op als stuurgroep. De stuurgroep ondersteunt de projectleiding bij het toewijzen van middelen en menskracht. De stuurgroep monitort voortgang en tussenresultaten. Afwijkingen ten aanzien van het projectplan worden aan Plato voorgelegd. Waar nodig stuurt de stuurgroep bij. Uiteindelijk verleent de stuurgroep decharge voor de opdracht.

Voor het project als geheel is de projectleider verantwoordelijk. De projectleider organiseert de afstemming tussen projectteams en bewaakt de uitgangspunten. Binnen het project zijn er verschillende projectteams Inhoud, een projectteam



Productie en een projectteam Evaluatie. Zie figuur 1.

Voor het onderdeel Inhoud is per prestatie-monitor een projectteam actief. Elk van deze projectteams wordt geleid door een vertegenwoordiger van Kennis en/of een vertegenwoordiger van de toezichtdirectie. Deze staan in nauw contact met de projectleider voor de onderlinge afstemming.

Figuur 1. geeft een schematische weergave van de projectorganisatie in 2020.  
\*onder voorbehoud

### 3.2 Interne en externe communicatie

#### *Interne communicatie*

Intern zijn er twee groepen ingesteld waar de voortgang van het project wordt besproken en waarin ontwerpkeuzes worden bediscussieerd. Ten eerst is er een kennisoverleg prestatie-monitor. Dit overleg vindt elke maand plaats. Als tweede vindt er afstemming plaats met een sectoroverstijgende klankbordgroep. Deze groep komt ongeveer eens in de twee maanden bijeen en fungeert ook als klankbordgroep voor het project Verdiepingsinformatie.

Om de vier maanden rapporteert het project schriftelijk aan de opdrachtgever en aan de hiervoor aangewezen portefeuillehouder binnen het Plato-overleg. Deze rapportage wordt (kort) besproken in Plato. Eind 2020 wordt een eindverslag geschreven.

#### *Externe communicatie*

Afspraken over externe communicatie worden gemaakt in de sectoroverstijgende werkgroep.

### 4. Projectplanning

#### 4.1. Benodigde capaciteit

Het project zal ook dit jaar een aanzienlijke capaciteit vergen, zowel vanuit de directie Kennis als vanuit de betrokken sectordirecties en Rekenschap/Juridische Zaken.

Rol	Organisatie-onderdeel	Medewerker	Benodigde dagen
Projectleider	Kennis	5.1.2e	20
Projectteam ontwikkeling BST (incl. continuïteit en doelmatigheid)	Rekenschap – BST		5
	Toezicht SWV		3
	Toezichtdirectie PO- BST		3
	Toezichtdirectie VO - BST		3
	Toezichtdirectie SO - BST		3
	Toezichtdirectie MBO- BST		3
	Toezichtdirectie HO- BST		3
	Kennis - BST		30
Projectteam doorontwikkeling SWV voor PO en VO	Toezicht SWV – PO VO		10
	Kennis		15

Rol	Organisatie- onderdeel	Medewerker	Benodigde dagen
Projectteam GMT voor KO/VVE	Toezichtdirectie PO – KO/VVE	5.1.2.e	-
	Kennis – KO/VVE		-
Projectteam doorontwikkeling PO voor BO, SBO en NWK	Toezichtdirectie PO – BO		10
	Kennis – BO		15
	Toezichtdirectie PO – SBO		10
	Kennis - SBO		15
	Toezichtdirectie PO – NWK		5
	Kennis - NWK		15
Projectteam (door)ontwikkeling SO voor SO en VSO	Toezichtdirectie SO		10
	Kennis - SO		15
	Toezichtdirectie VSO		10
	Kennis – VSO		15
Projectteam (door)ontwikkeling VO, PRO en NWK	Toezichtdirectie VO		10
	Kennis		15
	Toezichtdirectie VO		-
	Kennis		-
	Toezichtdirectie VO - NWK	-	
	Kennis - NWK	-	
Projectteam ontwikkeling MBO	Toezichtdirectie MBO	10	
	Kennis	30	
Projectteam ontwikkeling HO	Toezichtdirectie HO	10	
	Kennis	30	
Projectteam productie	Kennis	25	
Projectteam evaluatie	Kennis	15	
Totaal			363

Hieronder de totale capaciteit per organisatieonderdeel.

Organisatieonderdeel	Benodigde dagen
Kennis	255
Rekenschap – BST	5
Toezicht SWV – PO VO	13
Toezichtdirectie PO	28
Toezichtdirectie SO	23
Toezichtdirectie VO	13
Toezichtdirectie MBO	13
Toezichtdirectie HO	13

Organisatieonderdeel	Benodigde dagen
Totaal	363

#### 4.2. Fasering

Onderstaand overzicht toont de planning van het project.

2020	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Aug	Sep	Okt	Nov	Dec	Streefdatum
<b>Prestatiemonitors</b>													
<b>Besturen</b>													
Besturen (incl. financieel beheer)												•	31-12-2020
<b>PAO</b>													
Primair onderwijs							•						1-7-2020
Voortgezet onderwijs							•						1-7-2020
<b>Gemeenten</b>													
KO/VVE													onbekend
<b>PO</b>													
Basisscholen										•			1-10-2020
Speciaal basisonderwijs							•						1-7-2020
Nieuwkomers							•						1-7-2020
<b>SO</b>													
Speciaal onderwijs								•					1-8-2020
Voortgezet speciaal onderwijs								•					1-8-2020
<b>VO</b>													
Voortgezet onderwijs			•							•			30-9-2020
Praktijkonderwijs												•	31-12-2020
Nieuwkomers													onbekend
<b>MBO</b>													
Bekostigd							•						1-7-2020
<b>HO</b>													
Bekostigd												•	31-12-2020
<b>Productie(wijze)</b>													
Ontwikkeling visualisatie PowerBI				•				•				•	31-12-2020
<b>Evaluatie</b>													
Validatie prestatiemonitor												•	31-12-2020



# Projectplan

## Prestatiemonitor 2021

Opdrachtgever:

Stuurgroep:

Programmamanager AMMO:

Projectleider prestatie-monitor:

Directie Kennis

Platform Toezicht (Plato)

5.1.2e

## 1. Inleiding

### 1.1 Achtergrond project

Dit projectplan Prestatiemonitor 2021 is onderdeel van het programma AMMO. AMMO staat voor:

**A**lignment  
**M**onitoring  
**M**aatschappelijke  
**O**pgaven.

Het AMMO programma heeft als doel richting en sturing te geven aan de (door)ontwikkeling van de informatievoorziening ten behoeve van het vernieuwde toezicht. In onderstaand schema zijn de belangrijkste informatieproducten voor het vernieuwde toezicht 2021 ingedeeld naar functie. Dit projectplan gaat over de prestatie-monitor.

	Proportionaliteit	
	Risico	Analyse
Waarborgen	Prestatiemonitor Signalenmonitor	Verdiepingsinformatie Oordelenmonitor Opbrengstenmodellen
Stimuleren	Stelselmonitor (ingang stelsel)	Stelselmonitor (ingang bestuur)

In 2017-2018 is begonnen met het ontwikkelen van de prestatie-monitor. De prestatie-monitor is een informatieproduct dat wordt gebruikt ter ondersteuning van de jaarlijkse prestatieanalyse, zoals omschreven in het onderzoekskader 2017.

Voor de uitvoering van onze waarborgfunctie monitoren we de prestaties van besturen en scholen. Dit doen we om tijdig mogelijke risico's ten aanzien van de onderwijskwaliteit en het financieel beheer te kunnen detecteren. We noemen dit de jaarlijks prestatieanalyse.

Als er op basis van de prestatieanalyse geen vermoeden van risico's is dan wordt de analyse afgesloten. Besturen worden dan niet actief over de uitkomst van de prestatieanalyse op de hoogte gesteld. Als er een vermoeden van risico's is dan voeren we een expertanalyse uit.

Inmiddels zijn de onderzoekskaders na evaluatie van het kader 2017 bijgesteld en per 1 augustus 2021 worden de nieuwe onderzoekskaders 2021 gebruikt in het toezicht. In het vernieuwde toezicht willen we in de uitvoering o.a. meer gebruik maken van proportionaliteit (zie schema). Dit betekent dat we de intensiteit van het toezicht op het bestuur en de scholen afstemmen op de kwaliteit van het bestuur. De gegevens uit de prestatie-monitor dragen bij aan de inschatting van de kwaliteit van het bestuur door inzicht te geven in het al dan niet aanwezig zijn van (mogelijke) risico's voor de onderwijskwaliteit en/of risico's voor de kwaliteit van het bestuur.

Dit projectplan beschrijft de doelen en resultaten van de prestatie-monitor 2021.

## **1.2 Opdracht**

Zorg voor verdere ontwikkeling, productie en evaluatie van de prestatie-monitor.

### Ontwikkeling

- Zorg voor aansluiting bij AMMO, het vernieuwde toezicht en het onderzoekskader 2021.
- Zorg voor onderhoud en verdere ontwikkeling van de bestaande prestatie-monitors.
- Ontwikkel een nieuwe prestatie-monitor voor gemeenten kinderopvang/vve.

### Productie(wijze)

- Ontwikkel indicatoren zoveel mogelijk sectoroverstijgend.
- Verbeter de visuele presentatie van de prestatie-monitor in Power BI.

### Evaluatie

- Evaluer en valideer eerder ontwikkelde prestatie-monitors.
- Opzetten van werkwijze voor jaarlijkse evaluatie.



## 2. Opzet prestatiemonitor 2021

### 2.1 Doel

Het overkoepelende doel van het project is om te komen tot inrichting van een prestatiemonitor die

- aansluit bij ontwikkelingen binnen AMMO, het vernieuwde toezicht en het bijgestelde onderzoekskader 2021
- onnodige sectorale verschillen vermijdt
- gebruiksvriendelijk is en
- optimaal gebruik maakt van beschikbare innovatieve mogelijkheden, o.a. door gebruik te maken van het data science programma.

Daarbij streven we ernaar om in 2021 alle prestatiemonitors op te nemen in het reguliere werkproces.

### 2.2 Beoogde resultaten

Om het overkoepelende doel van dit project te bereiken zijn de volgende resultaten nodig:

#### *Inhoud*

- Aansluiten bij ontwikkelingen binnen AMMO. Denk hierbij aan inhoudelijke, functionele en technische uitlijning.
- Zorgen voor aansluiting bij het vernieuwde toezicht en het bijgestelde onderzoekskader 2021. Indien de vernieuwing van het toezicht consequenties heeft voor de inrichting en/of het gebruik van de prestatiemonitor zal de prestatiemonitor hierop aangepast worden.
- Onderhoud en verder ontwikkelen van prestatiemonitors die eerder ontwikkeld zijn. Dit betreft de prestatiemonitors
  - besturen geïntegreerd, (we onderzoeken of het mogelijk is om niet-bekostigde instellingen in deze monitor op te nemen. Zo niet, dan ontwikkelen we een afzonderlijke prestatiemonitor besturen niet-bekostigd).
  - besturen samenwerkingsverbanden po,
  - besturen samenwerkingsverbanden vo,
  - basisscholen,
  - speciaal basisonderwijs,
  - nieuwkomers primair onderwijs,
  - speciaal onderwijs,
  - voortgezet speciaal onderwijs,
  - voortgezet onderwijs,
  - praktijkonderwijs,
  - middelbaar beroepsonderwijs
  - hoger onderwijs (mogelijk geheel onderbrengen bij besturen).
- Ontwikkel nieuwe prestatiemonitors. Voor 2021 is dit een prestatiemonitor
  - gemeenten, kinderopvang/vve (ko/vve)
- Afbakenen inhoud prestatiemonitor besturen en prestatiemonitors op niveau object van toezicht.
- Zorgen voor interne communicatie en (gezamenlijke) scholingsactiviteiten binnen de toezichtdirecties. Het ontwikkelen van een informatieproduct is nl. breder dan het product zelf. In eerdere jaren is dit als belangrijk verbeterpunt naar voren gekomen. En zeker nu, in het kader van het vernieuwde toezicht, is het van belang dat we voldoende tijd steken in communicatie en scholing over het gebruik van de prestatiemonitor. We gaan zorgen voor bekendheid van de prestatiemonitor binnen de gehele organisatie.

#### *Productie(wijze)*

- We ontwikkelen indicatoren zo veel mogelijk per thema/databron in plaats van afzonderlijk per sector. Uitgangspunt hierbij is sectoroverstijgend waar het kan en sectorspecifiek waar het moet.
- Visualisatie van de prestatie-monitor in Power BI verbeteren, zodat deze voor alle sectoren voldoet.

#### *Evaluatie*

- Valideren verschillende prestatie-monitors door de relatie tussen risicoscores en de uitkomsten expertanalyses, en de relatie tussen risicoscores en toezichtresultaten te analyseren.
- Opzetten van een werkwijze ten aanzien van jaarlijkse evaluatie. Het gaat hierbij om een jaarlijkse planning, een format voor verzamelen van gegevens, een evaluatiemethodiek, een gestandaardiseerde rapportagevorm en afspraken ten aanzien van communicatie met sectoren.
- Verbeteren van de registratie(mogelijkheid) en het gebruik van de uitkomsten van expertanalyses.
- Opzetten van een methodiek om de bijdrage van individuele indicatoren aan de voorspelkracht van prestatie-monitors inzichtelijk te maken. Dit doen we in samenwerking met het data science programma.
- Verkenning van mogelijkheden om het rekenmodel uit te breiden zodat meer complexe criteria kunnen worden gebruikt. We passen het rekenmodel alleen aan wanneer uit evaluatie blijkt dat dit de voorspelkracht ten goede komt. Dit doen we in samenwerking met het data science programma.

### **2.3 Afbakening**

Het project betreft “alleen” de prestatie-monitor: instellingsoverzichten ter ondersteuning van de prestatie-analyse. Het project gaat niet over instellingsoverzichten in andere categorieën, zoals de verdiepingsinformatie en de beoordelingsmodellen. Het project gaat ook niet over informatieproducten op hogere aggregatieniveaus, zoals de stelselmonitor. Daarnaast gaat het project ook niet over de verbinding tussen alle informatieproducten. Hiervoor is AMMO ingericht. Wel is het zo dat besluiten van AMMO ten aanzien van inhoudelijke, functionele en technische uitlijning van informatieproductie van invloed zijn op de verdere ontwikkeling van de prestatie-monitor.

### **2.4 Relatie met andere projecten**

Dit project heeft een relatie met verschillende andere projecten.

- AMMO

Dit project valt onder het programma Alignment Monitoren Maatschappelijke Opgaven (AMMO). De prestatie-monitor is één van de monitoren onder dit programma.

- Verdiepingsinformatie

De prestatie-monitor hangt samen met het project verdiepingsinformatie. Beiden maken namelijk deels gebruik van dezelfde databronnen. Afstemming tussen beide projecten vindt binnen de directie Kennis, binnen AMMO, plaats. Daarnaast fungeert de klankbordgroep prestatie-monitor ook als klankbord voor de verdiepingsinformatie.

- Stelselmonitor

De prestatie-monitor hangt samen met de stelselmonitor. Beiden maken namelijk deels gebruik van dezelfde databronnen en indicatoren. Afstemming tussen beide projecten vindt binnen de directie Kennis, binnen AMMO, plaats.

- Onderwijsresultaten

Onderwijsresultaten zijn een belangrijk onderdeel van de prestatie-monitor. De verdere ontwikkeling van deze modellen heeft dan ook direct consequenties voor de inhoud van de prestatie-monitor.

- Evaluatie en bijstelling onderzoekskaders

De prestatie-monitor is een informatieproduct ter ondersteuning van het toezicht dat gestructureerd is aan de hand van de kwaliteitsgebieden en standaarden uit het onderzoekskader 2017. Evaluatie van het toezichtskader en bijstelling van het onderzoekskader in 2021 hebben daardoor direct een relatie met dit project.

- Procesnetwerk

Het procesnetwerk werkt de komende tijd, samen met het procesverbeteroverleg instellingstoezicht, aan het uitwerken en aanpassen van de procesbeschrijving 4JO. In augustus 2021 zijn deze vernieuwde processen klaar om het herziene onderzoekskader uit te kunnen voeren. Daarbij ook van belang dat de prestatie-monitors, als ondersteunende instrumenten, aansluiten bij het vernieuwde toezicht.

- Data science programma

Voor het project is het van belang om op de hoogte te zijn van de ontwikkelingen binnen het data science programma. Een van de doelen van de prestatie-monitor is immers om optimaal gebruik te maken van innovatieve mogelijkheden.

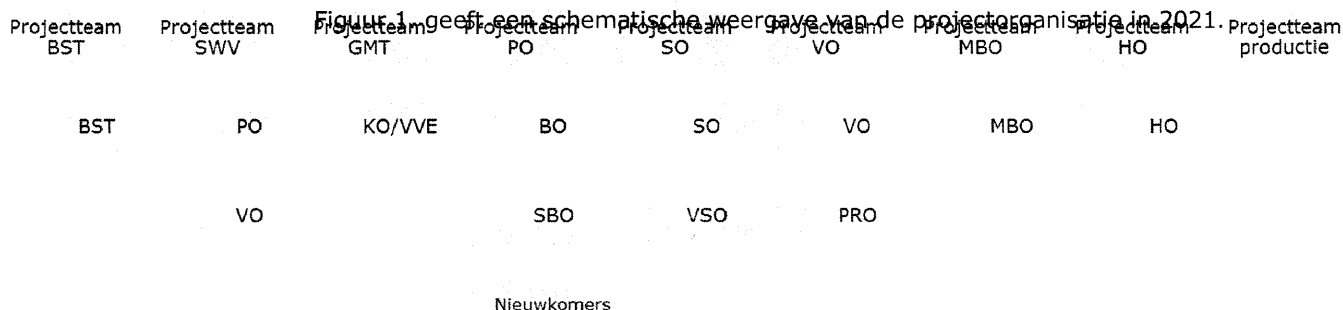
### 3. Organisatie

#### 3.1 Projectorganisatie

Het platform toezicht (Plato) treedt op als stuurgroep. De stuurgroep ondersteunt de projectleiding bij het toewijzen van middelen en menskracht. De stuurgroep monitort voortgang en tussenresultaten. Afwijkingen ten aanzien van het projectplan worden aan Plato voorgelegd. Waar nodig stuurt de stuurgroep bij. Uiteindelijk verleent de stuurgroep decharge voor de opdracht.

Voor het project als geheel is de projectleider verantwoordelijk. De projectleider organiseert de afstemming tussen projectteams en bewaakt de uitgangspunten. Binnen het project zijn er verschillende projectteams Inhoud, een projectteam Productie en een projectteam Evaluatie. Zie figuur 1.

Voor het onderdeel Inhoud is per prestatieindicator een projectteam actief. Elk van deze projectteams wordt vertegenwoordigd door een kennisambtenaar en/of een vertegenwoordiger van de toezichtdirectie. De wisselwerking tussen collega's van Kennis en de toezichtsectoren in een projectteam zorgen voor een succesvolle inhoudelijke ontwikkeling van een prestatieindicator. De projectteams staan in nauw contact met de projectleider voor de onderlinge afstemming.



### **3.2 Interne en externe communicatie**

#### *Interne communicatie*

Intern zijn er drie groepen waar de voortgang van het project wordt besproken en waarin ontwerpkeuzes worden bediscussieerd. Deze interne communicatie zorgt voor continuïteit van het project.

Ten eerst is er een kennisoverleg prestatie-monitor. Dit overleg vindt elke maand plaats. We bespreken hier de voortgang en stemmen de prestatie-monitors waar mogelijk op elkaar af.

Als tweede vindt er afstemming plaats met een sectoroverstijgende klankbordgroep. Deze groep komt ongeveer eens in de twee maanden bijeen en fungeert ook als klankbordgroep voor het project Verdiepingsinformatie.

Als derde worden de ontwikkelingen van het project één keer per maand in de programmagroep AMMO besproken. Voor 2021 verwachten we dat besluiten van AMMO ten aanzien van inhoudelijke, functionele en technische uitlijning van invloed zullen zijn op de verdere ontwikkeling van de prestatie-monitor.

Om de vier maanden rapporteert het project schriftelijk aan de opdrachtgever en aan de hiervoor aangewezen portefeuillehouder binnen het Plato-overleg. Deze rapportage wordt (kort) besproken in Plato. Eind 2021 wordt een eindverslag geschreven.

#### *Externe communicatie*

Afspraken over externe communicatie worden gemaakt in de sectoroverstijgende klankbordgroep.

## 4. Projectplanning

### 4.1. Benodigde capaciteit

Het project zal ook in 2021 een aanzienlijke capaciteit vergen, zowel vanuit de directie Kennis als vanuit de betrokken sectordirecties en Rekenschap/Juridische Zaken.

Rol	Organisatie-onderdeel	Medewerker	Benodigde dagen
Projectleider	Kennis	5.1.2.e	20
Projectteam ontwikkeling BST	Rekenschap – BST		5
	Toezicht SWV		5
	Toezichtdirectie PO-BST		5
	Toezichtdirectie VO - BST		5
	Toezichtdirectie SO - BST		5
	Toezichtdirectie MBO- BST		5
	Toezichtdirectie HO-BST		5
	Kennis - BST		30
Projectteam doorontwikkeling SWV voor PO en VO	Toezicht SWV – PO VO		10
	Kennis		15
Projectteam GMT voor KO/VVE	Toezichtdirectie PO – KO/VVE		10
	Kennis – KO/VVE		15
Projectteam doorontwikkeling PO voor BO, SBO en NWK	Toezichtdirectie PO – BO		75
	Kennis – BO		20
	Toezichtdirectie PO – SBO		10
	Kennis - SBO		15
	Toezichtdirectie PO – NWK		5
	Kennis - NWK		15
Projectteam (door)ontwikkeling SO voor SO en VSO	Toezichtdirectie SO		10
	Kennis - SO		20
	Toezichtdirectie VSO		10
	Kennis – VSO		20
Projectteam (door)ontwikkeling VO en PRO	Toezichtdirectie VO		20
	Kennis		15
	Toezichtdirectie VO - PRO		10

Rol	Organisatie-onderdeel	Medewerker	Benodigde dagen
	Kennis	<b>5.1.2.e</b>	15
Projectteam ontwikkeling MBO	Toezichtdirectie MBO		10
	Kennis		15
Projectteam ontwikkeling HO	Toezichtdirectie HO		10
	Kennis		15
Projectteam productie	Kennis		25
Projectteam evaluatie	Kennis		15
<b>Totaal</b>			<b>485</b>

Hieronder de totale capaciteit per organisatieonderdeel.

Organisatieonderdeel	Benodigde dagen
Kennis	270
Rekenschap – BST	5
Toezicht SWV – PO VO	15
Toezicht GMT – KO VVE	10
Toezichtdirectie PO	95
Toezichtdirectie SO	25
Toezichtdirectie VO	35
Toezichtdirectie MBO	15
Toezichtdirectie HO	15
<b>Totaal</b>	<b>485</b>

#### 4.2. Fasering

Deze paragraaf beschrijft de planning van het project, gevolgd door een schematisch overzicht van de planning.

Deze planning gaat uit van maximaal twee leveringen per **prestatiemonitor** per jaar: in het voorjaar (maart-mei) en rond de start van het nieuwe schooljaar (september).

Belangrijke overwegingen bij deze planning zijn:

1. beschikbaarheid van gegevens

Met name maart en september zijn periodes waarin veel gegevens beschikbaar komen. Het gaat in maart om de verwerking van inschrijvingen/examens o.b.v. 1cijferbestanden en personeelsdata en in september om de definitieve financiële gegevens en ziekteverzuim.

2. capaciteit die nodig is voor verversing van de prestatiemonitor

Het evalueren en opleveren van een nieuwe versie van een prestatiemonitor kost tijd. Daarom is het nodig om een doorlooptijd aan te houden tussen het beschikbaar komen van de data en het opleveren van een prestatiemonitor. Daarnaast is het wenselijk om het aantal leveringen enigszins beperkt te houden, en in ieder geval niet vaker te leveren dan strikt noodzakelijk.

3. aansluiting bij het werkproces

Momenteel wordt in veel sectoren eenmaal per jaar een planning gemaakt op basis van een analyse van de meest risicovolle ovt's (hetzij opleidingen/scholen, hetzij besturen). Voor veel sectoren gaat het om de start van het schooljaar

(dus rond september). Het lijkt onwenselijk om de rangorde van besturen regelmatig aan te passen, omdat hiermee het overzicht op risicosorteringen minder kan worden. Wel is het natuurlijk zo dat er aan de hand van binnenkomende ernstige signalen te allen tijde onderzoeksactiviteiten opgestart kunnen worden (d.w.z. zonder prestatie-monitor).

Wat betreft **productie** zijn er drie oplevermomenten gepland van een mogelijke update/verbetering van de rapportage in Power BI.

Voor de **evaluatie** is het streven om voor iedere prestatie-monitor een paar maanden voor oplevering van een nieuwe versie de evaluatie gereed te hebben, bijv. januari vo en in mei po. Voor de evaluatie ontwikkelen we dit jaar een jaarlijkse werkwijze.

2021	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Aug	Sep	Okt	Nov	Dec	Streefdatum
<b>Prestatiemonitors</b>													
<b>Besturen</b>													
Besturen <sup>1</sup>				•					•				15-9-2021
<b>PAO</b>													
Primair onderwijs		•							•				1-3-2021
Voortgezet onderwijs		•							•				1-3-2021
<b>Gemeenten</b>													
KO/VVE									•				1-9-2021
<b>PO</b>													
Basisscholen									•				6-9-2021
Speciaal basisonderwijs									•				6-9-2021
Nieuwkomers						•							1-7-2021
<b>SO</b>													
Speciaal onderwijs									•				15-9-2021
Voortgezet speciaal onderwijs									•				15-9-2021
<b>VO</b>													
Voortgezet onderwijs			•						•				7-3-2021
Praktijkonderwijs			•										14-3-2021
<b>MBO</b>													
Subgroep x niveau					•								15-5-2021
<b>HO</b>													
Bekostigd <sup>2</sup>									•				15-9-2021
<b>Productie(wijze)</b>													
Ontwikkeling visualisatie PowerBI				•				•				•	31-12-2021
<b>Evaluatie</b>													
Validatie prestatie-monitors	•				•								1-5-2021

<sup>1</sup> Mogelijk niet-bekostigde instellingen toevoegen aan prestatie-monitor besturen geïntegreerd dan wel afzonderlijke prestatie-monitor ontwikkelen voor besturen niet-bekostigd.

<sup>2</sup> Mogelijk geheel onderbrengen bij de prestatie-monitor besturen geïntegreerd.





Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

## **EVALUATIE VOORSPELKRACHT PRESTATIEMONITOR PO NK 2020**

**5.1.2.e**

februari 2021

## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 4</b>
<b>2</b>	<b>Registratie van de expertanalyse t.b.v. de evaluatie 5</b>
<b>3</b>	<b>Kwaliteit van de risicoschatting 6</b>
3.1	Indeling uitkomsten expertanalyse 6
3.2	Verdeling risico-scholen op basis van rang in de prestatie-monitor 7
3.3	Foutanalyse 8
3.4	Foutanalyse alleen aan de hand van oordelen 9
<b>4</b>	<b>Conclusie 12</b>
<b>5</b>	<b>Aanbevelingen 13</b>
5.1	Aanbevelingen voor toekomstige ontwikkelingen van de PM PO NK 13
5.2	Aanbevelingen voor de registratie van de expertanalyse 13
<b>Bijlage I: Risico-indicatoren per kwaliteitsgebied 14</b>	
<b>Bijlage II: Indicatoranalyse 16</b>	
Conclusie 17	

## Samenvatting

In dit document wordt een beschrijving gegeven van de evaluatie van de voorspelkracht van de Prestatiemonitor PO nieuwkomers (PO-NK) aan de hand van de registratie van de expertanalyse door de sector. Deze evaluatie laat een redelijk positief beeld zien van de voorspelkracht van de Prestatiemonitor PO-NK. 92% van de scholen die door analisten of inspecteurs als risicovol zijn aangeduid kwamen ook als risicovol uit de prestatie-monitor. Daarnaast kwamen alle scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de hoogste risicogroepen. Daarmee kan de Prestatiemonitor PO-NK een belangrijke bijdrage leveren aan risicogericht toezicht binnen de sector.

Desalniettemin mistte de prestatie-monitor nog 1 risicovolle school volgens de inschatting van de analist. Er bestaat dus ook enige ruimte voor verbetering van risicoschatting door de prestatie-monitor. Bovendien zijn er ook enkele onzekerheden bij het interpreteren van de uitkomsten van de expertanalyse die de evaluatie bemoeilijken. Om de evaluatie te vergemakkelijken en risicoschatting te verbeteren doen wij aan het eind van dit document enkele aanbevelingen. Deze zullen in de toekomst de evaluatie van de prestatie-monitor verbeteren.

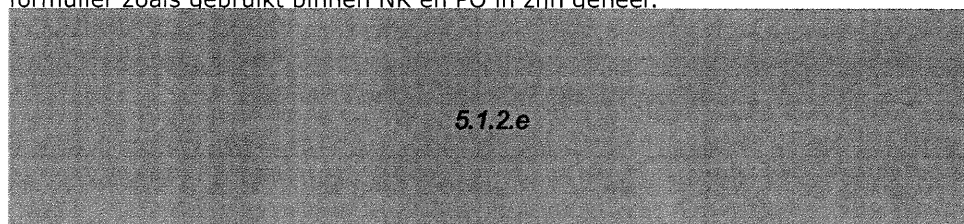
Ten slotte is het goed te benadrukken dat risicoschatting nooit alleen op basis van de prestatie-monitor moet plaatsvinden, maar altijd in combinatie met andere bronnen (zoals kwalitatieve kennis van inspecteurs en analisten over scholen).

## **1 Inleiding**

De directie Kennis ontwikkelt jaarlijks de zogenaamde prestatie-monitors voor de verschillende sectoren. De prestatie-monitors beogen inspecteurs en analisten te ondersteunen bij het maken van risicoprioritering, gebaseerd op data. Sinds 2018 bestaat er ook een Prestatiemonitor PO-Nieuwkomers (de PM PO-NK). In bijlage I is te zien welke indicatoren er per kwaliteitsgebied zijn opgenomen in de PM PO-NK. Om de werking van de prestatie-monitors te borgen en te verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de prestatie-monitor. Voor deze evaluatie worden de risicoscores van de prestatie-monitor afgezet tegen enerzijds de beoordeling van risico's op scholen aan de hand van de expertanalyse (deskresearch); en anderzijds tegen de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een inspectie. In dit document wordt verslag gedaan van de evaluatie zoals uitgevoerd in december 2020.

## 2 Registratie van de expertanalyse t.b.v. de evaluatie

Om de risicoscores van de prestatiemonitor te vergelijken met de uitkomsten van de expertanalyse is het nodig om de uitkomsten van de expertanalyse op gestructureerde wijze vast te leggen. Dit jaar is er voor de tweede keer met een dergelijke registratie begonnen. Dit is gebeurd aan de hand van een Excelformulier dat binnen de sector is bijgehouden. Hieronder een indruk van de structuur van dat formulier zoals gebruikt binnen NK en PO in zijn geheel:



Dit formulier vormde de basis voor de evaluatie die hieronder wordt beschreven, en is daarmee erg waardevol. Dat uiteindelijk de registratie nagenoeg compleet is gemaakt is een compliment waard. Wij erkennen als Kennis hoeveel werk daarin is gegaan. Zonder deze investering was deze evaluatie niet mogelijk geweest. Voor een volgende versie zien wij een volgend verbeterpunt. Waar we regelmatig tegen aanliepen bij de evaluatie is dat het niet altijd duidelijk is of een school met een hoge risicoscore volgens de analist eigenlijk *niet* risicovol is (een fout van de Prestatiemonitor), of dat een school mogelijk wel risicovol is maar dat er andere redenen waren om bij de school geen kwaliteitsonderzoek bij risico's (KO risico) uit te voeren, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. Vaak stonden er echter wel beschrijvingen in het document die (deels handmatig) zijn gebruikt om toch duiding te geven aan de uitkomsten van de expertanalyse. Meer daarover in de volgende secties. Voor de toekomst raden wij aan om te zorgen voor een duidelijke scheiding in de registratie tussen de kwaliteit van de risicoschatting en de genomen vervolgstappen m.b.t. toezicht.

### 3 Kwaliteit van de risicoschatting

Om de kwaliteit van de risicoschatting van de Prestatiemonitor PO-NK inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken hoe de uitkomst van de expertanalyse door de analisten en het huidige oordeel zich verhouden tot de rang die scholen hebben in de prestatie-monitor, en of de school volgens de prestatie-monitor in de risicogroep valt. De risicogrens vanaf wanneer een ovt werd meegenomen in de expertanalyse (EA) is in 2020 gelegd bij risicoscore  $\geq 0$ . 10. Hiermee vallen 33 van de 83 scholen in de risicocategorie.

#### 3.1 Indeling uitkomsten expertanalyse

Voor de foutanalyse hebben we de uitkomst van de expertanalyse ingedeeld naar 3 mogelijke uitkomsten: *EA: KO Risico*; *EA: Wel risico, geen KO*; *EA: Geen risico (en dus geen KO)*. In sommige gevallen is er echter besloten om bijvoorbeeld geen KO-risico uit te voeren omdat een school 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende school die nog even de tijd krijgt om herstelplan uit te voeren). Om deze scholen ook in beeld te brengen zijn voor de scholen die geen KO-risico krijgen maar die wel een huidig oordeel hebben ook de huidige oordelen meegenomen: *Zeër zwak*; *Onvoldoende*; *Voldoende*; en *Goed*. Oordelen worden meegenomen als ze gegeven zijn sinds augustus 2017.

Daarmee komen we tot de volgende categorieën:

1. **EA: KO risico**  
Als de analist een voorstel tot KO risico doet.
2. **Oordeel Onvoldoende of Zeër zwak**  
Er is geen expertanalyse uitgevoerd, maar deze school heeft het oordeel *Onvoldoende* of *Zeër zwak*. Deze scholen tellen mee als risico in de foutanalyse.
3. **EA: Wel risico, geen KO**  
Als de analist geen voorstel doet voor een KO risico, maar wel risico's op de school constateert. Dit is bijvoorbeeld als vertrouwen is in het bestuur er een herstelonderzoek nog gaande is of net is afgerond of een verificatie plaats gaat vinden. Deze scholen tellen mee als risico in de foutanalyse.
4. **EA: Geen KO risico**  
De analist heeft de expertanalyse uitgevoerd, maar geen risico's geconstateerd die voldoende zijn voor een KO-risico.
5. **Oordeel: Voldoende**  
Er is geen expertanalyse uitgevoerd, en deze school heeft het oordeel *Voldoende*. Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed**  
Er is geen expertanalyse uitgevoerd, maar deze school heeft het oordeel *Goed*. Deze scholen tellen mee als geen risico in de foutanalyse.

Zoals te zien in Tabel 1 heeft een analist voor 10 scholen een voorstel gedaan voor een KO risico, 9 van deze scholen vallen in de risicogroep van de prestatie-monitor (top 33), 1 valt daarbuiten. 2 scholen hebben het oordeel *Onvoldoende* of *Zeër zwak*, beide vallen in de risicogroep van de prestatie-monitor. 5 scholen hebben het oordeel *Goed*, 2 hiervan vallen in de risicogroep van de prestatie-monitor, 3 vallen erbuiten.

**Tabel 1: Aantal scholen per uitkomst van de expertanalyse en risico op de prestatie-monitor.**

	Risico op de prestatie-monitor	
	Risico (1 – 33)	Geen risico (34 – 83)

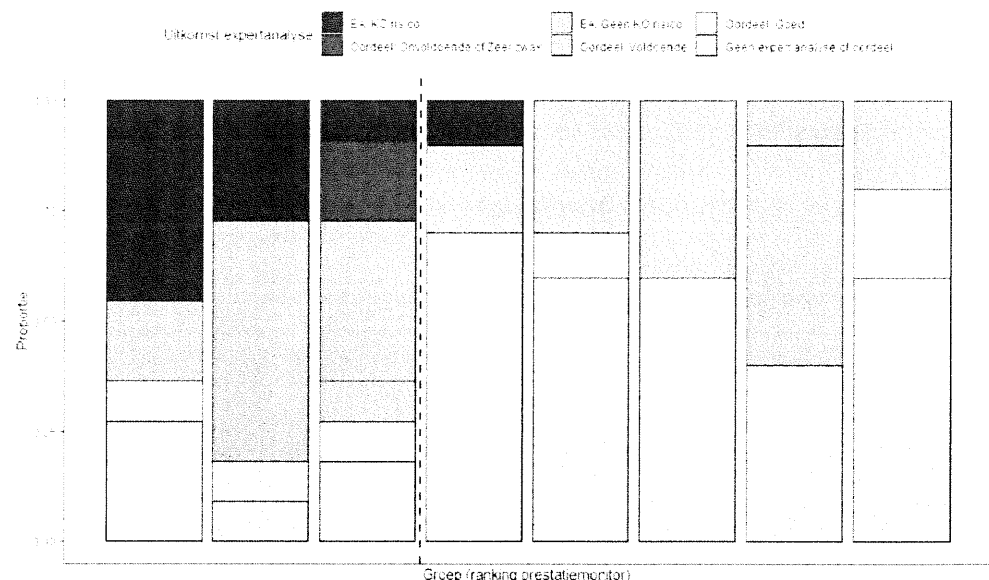
Uitkomst expertanalyse	EA: KO risico	9	1
	Oordeel: <i>Onvoldoende of Zeer zwak</i>	2	0
	EA: Wel risico, geen KO	0	0
	EA: Geen KO risico	12	1
	Oordeel: <i>Voldoende</i>	1	16
	Oordeel: <i>Goed</i>	2	3
	EA: Geen expertanalyse of oordeel	4	31

### 3.2

#### Verdeling risico-scholen op basis van rang in de prestatie-monitor

De uitkomsten van de expertanalyse kunnen verdeeld worden over de rangorde in de prestatie-monitor. Hiervoor zijn de scholen onderverdeeld in groepen op basis van hun rang in de prestatie-monitor. De 33 scholen die risico hebben volgens de prestatie-monitor zijn opgesplitst in drie groepen van elk 11 scholen. De overige 50 scholen zijn verdeeld in 5 groepen van elk 10 scholen (Figuur 1).

In Figuur 1 valt te zien dat alle scholen die risico hebben (EA: KO risico, Oordeel: *Onvoldoende of Zeer zwak* en EA: Wel risico, geen KO) in de top van de prestatie-monitor zitten. Naast scholen die in één van de risicogroepen vallen zitten er ook nog 1 risicoschool net onder de risicogrens in groep 4. Wel is het zo dat voor bijna één derde van de scholen afgelopen jaar geen expert analyse uitgevoerd en dat er ook geen recent oordeel is. Hierdoor hebben we geen zicht op de onderwijskwaliteit van deze scholen en of onze risicoschatting terecht is.



**Figuur 1: De relatie tussen scores op de prestatie-monitor en waargenomen risico's.** Scholen zijn gesorteerd op basis van de PM-risicoscore, en samengevoegd in 8 groepen. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de expertanalyse (EA) of Oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste twee groepen bevatten de 33 scholen die risico hebben op de prestatie-monitor, de 5 groepen daarna bestaan elk uit 10 scholen.

### 3.3

#### Foutanalyse

Foutanalyse is een methode om de kwaliteit van risicoschatting te kwantificeren. Dit is zinvol omdat daarmee op termijn de kwaliteit van risicoschattingen over de tijd en over verschillende sectoren/kwaliteitsgebieden vergeleken kan worden. Foutanalyse berust op een onderverdeling van uitkomsten in vier groepen (voortkomend uit Tabel 1):

- **Echt positieven (terecht alarm):** Zowel de prestatie-monitor als de expertanalyse detecteren risico's voor een school
- **Fout negatieven (gemist geval):** De prestatie-monitor detecteert geen risico's, maar de expertanalyse constateert dat er wél risico's zijn op een school.
- **Fout positieven (loos alarm):** De prestatie-monitor detecteert risico, maar de expertanalyse constateert dat er géén risico's zijn op een school.
- **Echt negatieven (terecht genegeerd):** Zowel de prestatie-monitor als de expertanalyse detecteren geen risico's voor een school.

Zo krijgen we Tabel 2.

**Tabel 2: Foutanalyse op de uitkomst van de prestatie-monitor en de expertanalyse**

		Risico op de prestatie-monitor	
		Risico	Geen risico
Risico volgens de analist	Risico	11 (Echt positieven)	1 (Fout negatieven)
	Geen risico	20 (Fout positieven)	51 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Op zowel sensitiviteit als op specificiteit scoort de prestatie-monitor heel aardig. 85% van de risicovolle scholen is boven de gestelde risicogrens gekomen. Dit betekent dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die mogelijk risico's hebben (en daarmee ook aan het onderscheiden van scholen waar waarschijnlijk niks aan de hand is). Zeker als we dit vergelijken met het alternatief wat gewoon het trekken van willekeurige steekproeven is. Ook op specificiteit scoort de prestatie-monitor aardig 71% van de niet risicovolle scholen worden correct afgewezen door de prestatie-monitor.



Op precisie scoort de prestatie-monitor minder (35% van de onderzochte scholen is vervolgens gewaardeerd als risicovol). Dit betekent dat er relatief veel scholen moeten worden bekeken in de expertanalyse omdat ze hoog scoren op de prestatie-monitor, maar waar onvoldoende aanleiding is voor een KO risico. Een verklaring hiervoor is dat het optimaliseren van de sensitiviteit en specificiteit vaak ten koste gaat van de precisie. Een lage precisie betekent dat er meer capaciteit nodig is van de analisten voor de expertanalyse. Het is dus van belang een balans te vinden tussen een hoge sensitiviteit en specificiteit nastreven en te zorgen dat de precisie niet te laag wordt, omdat dit zal leiden tot overvragen van de analisten. Een andere mogelijke verklaring hiervoor is dat er geen KO risico wordt gepland omdat er vertrouwen in het bestuur is en/of omdat de risico's bekend zijn, maar dat dit vanuit de huidige manier van documenteren niet duidelijk wordt en gebruikt kan worden in de evaluatie van de prestatie-monitor. Dit zou verbeterd kunnen worden door de manier bij het documenteren van de expertanalyse een onderscheid te maken tussen of de analist risico's constateert op een school en het doen van een voorstel tot een toezichtsactiviteit. In Bijlage II is te zien welke indicatoren er per kwaliteitsgebied ontwikkeld zijn.

### 3.4

#### **Foutanalyse alleen aan de hand van oordelen**

Naast de kwaliteit van de risicoschatting bepalen aan de hand van de uitkomst van expertanalyses, kunnen we ook kijken naar de uitkomst van recente oordelen. Deze manier van analyseren heeft als nadeel dat er minder data beschikbaar is (er worden namelijk meer expertanalyses uitgevoerd dan oordelen gegeven). Het grote voordeel van het doen van een foutanalyse aan de hand van oordelen is dat de data betrouwbaarder is, omdat de school ook daadwerkelijk bezocht is.

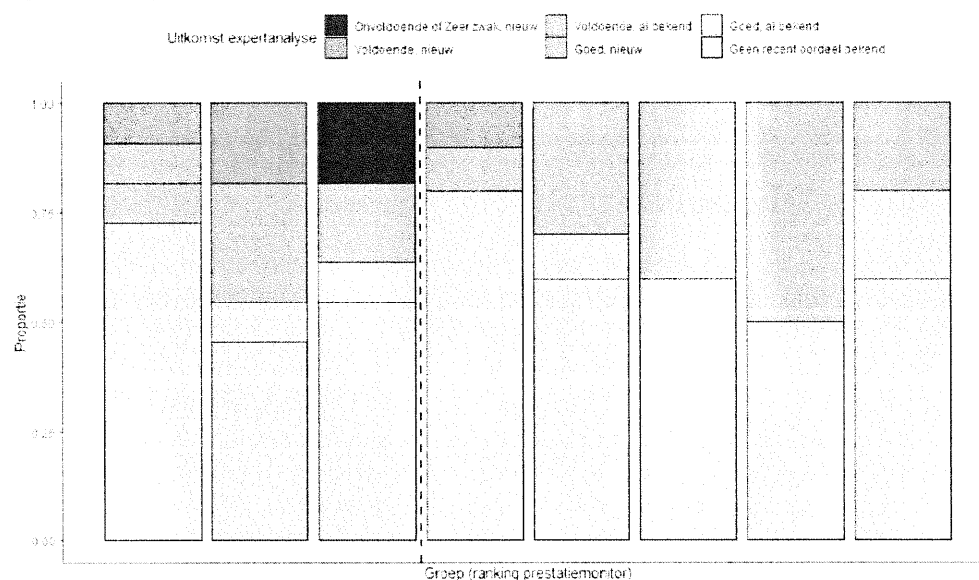
Voor de foutanalyse delen we oordelen in zeven categorieën:

1. **Oordeel: Onvoldoende of Zeer zwak, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
2. **Oordeel: Onvoldoende of Zeer zwak, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
3. **Oordeel: Voldoende, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
4. **Oordeel: Voldoende, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
5. **Oordeel: Goed, al nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.

Op dezelfde manier als Figuur 1 kunnen we de scholen over 8 groepen verdelen. Hier hebben de eerste de eerste 3 staven elk 11 scholen die risico hebben volgens de prestatie-monitor en de 5 groepen daarna bestaan steeds uit 10 scholen.

Voor ongeveer twee derde van de top 33 van de prestatie-monitor is er een recent oordeel bekend. Bij de 2 scholen is het oordeel *Onvoldoende* of *Zeer zwak* gegeven.

Er zijn geen scholen met het oordeel *Onvoldoende* of *Zeer zwak* buiten de risicogroep van de prestatiemonitor



**Figuur 2: De relatie tussen scores op de prestatiemonitor en recente oordelen, sinds 1 augustus 2017. Scholen zijn gesorteerd op basis van de PM-risicoscore, en samengevoegd in 8 groepen. Iedere staaf geeft de proportie weer van hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatiemonitor. De eerste groep bevatten elk 11 scholen die risico hebben op de prestatiemonitor, de 5 groepen daarna bestaan steeds uit 10 scholen.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 2, krijgen we Tabel 3.

**Tabel 3: Foutanalyse op de uitkomst van de prestatiemonitor en recente oordelen**

		Risico op de prestatiemonitor	
		Risico	Geen risico
Risico volgens recent oordeel	Risico	2 (Echt positieven)	0 (Fout negatieven)
	Geen risico	29 (Fout positieven)	52 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

De bovenstaande analyse aan de hand van de oordelen gegeven sinds augustus 2017 laat dit zien dat ook aan de hand van recente oordelen de sensitiviteit van de prestatiemonitor hoog is. De prestatiemonitor PO-NK detecteert 100% van de scholen die het oordeel Onvoldoende of Zeer zwak hebben, hiermee mist de prestatiemonitor geen enkele risicovolle school. Daarom valt te concluderen dat de prestatiemonitor een goede bijdrage kan leveren aan het detecteren van scholen die in aanmerking komen voor een KO risico onderzoek en het oordeel *Onvoldoende* of *Zeer zwak* krijgen. Wel is het zo dat er weinig recente oordelen bekend zijn en dat het dus onzeker is hoeveel scholen dus daadwerkelijk onvoldoende zijn. Daarnaast scoort de prestatiemonitor op precisie erg laag, slechts 6% van de scholen die onderzocht zijn kregen uiteindelijk het oordeel *Onvoldoende* na een KO risico. Dit betekent dat er relatief veel scholen moeten worden bekeken in de expertanalyse omdat ze hoog scoren op de prestatiemonitor, maar waar de risico's uiteindelijk niet voldoende zijn voor het oordeel *Onvoldoende* of *Zeer zwak*. Een verklaring hiervoor is dat het optimaliseren van de sensitiviteit en specificiteit vaak ten koste gaat van de precisie. Een lage precisie betekent dat er meer capaciteit nodig is van de analisten voor de expertanalyse. Dus de hoge score op sensitiviteit voor de prestatiemonitor leidt hier ook tot de lage score op precisie.

## 4 Conclusie

De foutanalyse schetst een overwegend positief beeld van de kwaliteit van risicoschatting door de Prestatiemonitor PO-NK: 85% van de scholen die als risicovol uit de expertanalyse komen -of die al een onvoldoende oordeel hebben- kwamen terecht in de hoogste risicogroepen. Daarnaast kwamen alle scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de hoogste risicogroepen. Dit suggereert dat er toegevoegde waarde is voor het gebruik van de PM PO-NK bij het prioriteren van scholen in het toezicht.

Tegelijkertijd is er ook ruimte voor verbetering. Immers, niet álle risicovolle scholen kwamen in de twee hoogste risicogroepen terecht. Verder bleken relatief veel scholen die boven de risicogrens uitkwamen toch voldoende. Dat roept de vraag op of het niet mogelijk zou zijn om de risicogrens te verhogen. Dit zou echter samen moeten gaan met verbeterde risicoschatting (nét boven de grens vonden we immers al vrij veel risicoscholen). Een van de belangrijkste bijdragen van de methode van evaluatie zoals hier beschreven is dat de ontwikkeling van de voorspelkracht in de komende jaren ook gemonitord kan worden. Daarmee kan er gericht en meetbaar gewerkt worden aan verbetering van risicoschatting.

Daarnaast is een perfecte risicoschatting onmogelijk. Risico's veranderen over de tijd, en er zijn ook risico's waar we op dit moment simpelweg geen data over hebben (zoals bijvoorbeeld de kwaliteit/ervaring van bestuurders, hoewel hier wel aan gewerkt wordt vanuit het datascience programma). Het detecteren van risicoscholen zal daarom altijd moeten bestaan uit een *combinatie* van zowel datagedreven risicoschatting (zoals in de prestatie-monitor) als het gebruik van de impliciete kennis van domeinexperts zoals inspecteurs of analisten, en bijvoorbeeld signalen.

## 5 Aanbevelingen

### 5.1 Aanbevelingen voor toekomstige ontwikkelingen van de PM PO NK

- 1) Zoals beschreven in dit document bestaat er onzekerheid in de schatting van kwaliteitsmaten van risicoschatting. Dit komt omdat we simpelweg niet goed weten in hoeverre er risico's bestaan op scholen die een lage risicoscore hebben en het volgens de prestatiemonitor dus juist goed doen (en die ook niet in beeld zijn bij de COI). Hoewel we niet alle scholen kunnen en willen onderzoeken, zou het wel mogelijk zijn om onderzoeken uit te voeren onder scholen met een lage risicoscore aan de hand van steekproeven. De prestatiemonitor zou ook gebruikt kunnen worden om op basis van de rangorde een steekproef te trekken en deze scholen te bekijken, als vorm van proportionaliteit in het toezicht. Ook is het een optie, gezien het geringe aantal nieuwkomersscholen, om voor één jaar de gehele populatie mee te nemen in de expertanalyse. Om zo meer kennis op te doen over alle scholen en daarna over te gaan op het trekken van steekproeven aan de hand de rangorde in de prestatiemonitor. Dit kan helpen om de *prevalentie* van onvoldoende scholen inzichtelijk te maken. Dit zou sterk kunnen bijdragen aan de verdere verbetering van (datagedreven) risicogericht toezicht.
- 2) Als laatste kunnen de uitkomsten van de evaluatie ook gebruikt worden om verbeteringen in indicatorselectie en weging toe te passen op de prestatiemonitor. In Bijlage II (zie hieronder) hebben we hiertoe een aanvullende analyse uitgevoerd.

### 5.2 Aanbevelingen voor de registratie van de expertanalyse

- 1) Voor deze analyse is gebruik gemaakt van de registratie van de expertanalyse in een Exceltabel. Voor toekomstige evaluatie en verbetering van de prestatiemonitor zou het echter helpen om 'tot een nog duidelijkere scheiding te komen tussen het bestaan van risico's (ofwel, vindt de analist de risicoscore terecht) en de beslissing voor vervolgtoezicht (wordt er een KO-risico uitgevoerd). Met name het eerste is namelijk belangrijk voor een goede evaluatie van de prestatiemonitor. Hiertoe stellen wij voor om de kolom '*KO-risico? Voorstel analist*' op te splitsen in twee kolommen. Bijvoorbeeld één kolom '*Zijn er risico's op de school?*' en één kolom '*KO-risico voorstel*'.
- 2) Daarnaast waren er dit jaar problemen met dataregistratie over de expertanalyse. Dit resulteerde in fouten die handmatig omgezet moesten worden bij de evaluatie. De oorzaak hiervan ligt bij Proza dat niet geschikt is voor een dergelijke registratie. Voor komend jaar is het goed om te kijken of de procedure van het registreren van de expertanalyse effectiever kan.
- 3) Scholen die volgens de prestatiemonitor niet in de risicogroep vallen, maar waar wel een KO risico voorstel gedaan wordt door de analist zijn zeer interessant voor de evaluatie. Daarom is het handig als uit de documentatie duidelijk wordt waarom er een KO risico voorstel en hoe deze school in beeld is gekomen (bijvoorbeeld via de contactinspecteur of via onderwijsresultaten).

## Bijlage I: Risico-indicatoren per kwaliteitsgebied

De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

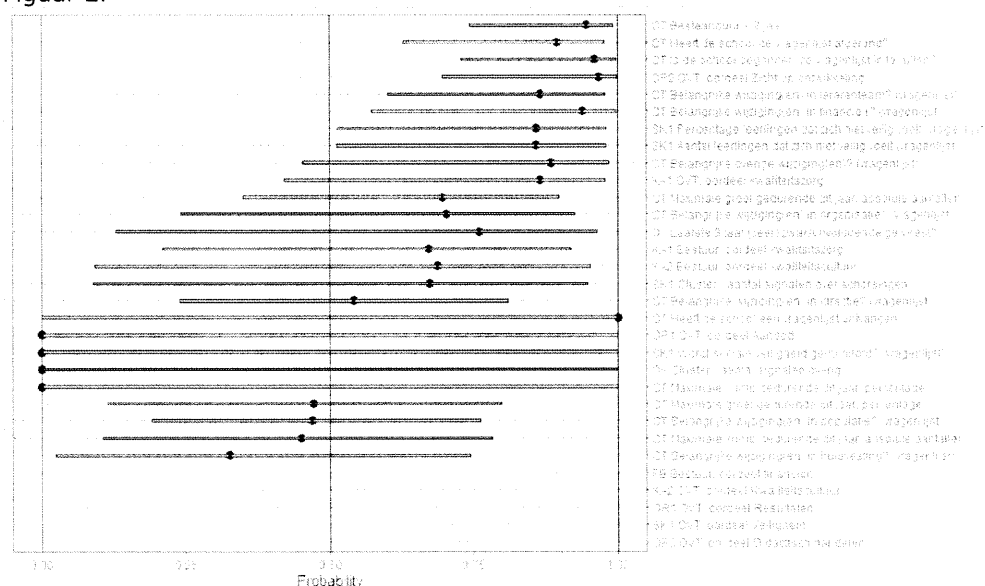
Onderwijsresultaten	Onderwijsproces	Schoolklimaat	Kwaliteitszorg en ambitie	Financieel beheer	Context	Overig
OVT: oordeel resultaten (OR1) <i>Weging = 1</i>	OVT: oordeel aanbod (OP1) <i>Weging = 1</i>	OVT: oordeel veiligheid (SK1) <i>Weging = 1</i>	OVT: oordeel kwaliteitszorg (KA1) <i>Weging = 1</i>	Bestuur oordeel financiën (FB) <i>Weging = 1</i>	Bestaansduur < 2 jaar(CT) <i>Weging = 1</i>	Laatste 3 jaar (zeer) zwak/onvoldoende geweest (OV) <i>Weging = 1</i>
	OVT: oordeel zicht op ontwikkeling (OP2) <i>Weging = 1</i>	Cluster – aantal signalen over schorsingen (SK1) <i>Weging = 1</i>	Bestuur: oordeel kwaliteitszorg (KA1) <i>Weging = 1</i>		Heeft de school een vragenlijst ontvangen (CT) <i>Weging = 2</i>	Cluster: aantal signalen overig (OV) <i>Weging = 1</i>
	OVT: oordeel didactisch handelen (OP3) <i>Weging = 1</i>	Wordt sociale veiligheid gemonitord? (SK1) <i>Weging = 2</i>	OVT: oordeel kwaliteitscultuur (KA2) <i>Weging = 1</i>		Is de school begonnen de vragenlijst in te vullen (CT) <i>Weging = 2</i>	
		Aantal leerlingen dat zich niet veilig voelt (SK1) <i>Weging = 2</i>	Bestuur: oordeel kwaliteitscultuur (KA2) <i>Weging = 1</i>		Heeft school de vragenlijst afgerond (CT) <i>Weging = 15</i>	
		Percentage leerlingen dat zich niet veilig voelt (SK1) <i>Weging = 1</i>			Maximale groei gedurende dit jaar, absolute aantallen (CT) <i>Weging = 0,5</i>	
					Maximale krimp gedurende dit jaar, absolute aantallen (CT) <i>Weging = 0,5</i>	
					Maximale groei	

gedurende dit jaar, percentage (CT) <i>Weging = 0,5</i>
Maximale krimp gedurende dit jaar, percentage (CT) <i>Weging = 0,5</i>
Belangrijke wijzigingen in organisatie? (CT) <i>Weging = 1</i>
Belangrijke wijzigingen in populatie? (CT) <i>Weging = 1</i>
Belangrijke wijzigingen in directie? (CT) <i>Weging = 1</i>
Belangrijke wijzigingen in lerarenteam? (CT) <i>Weging = 1</i>
Belangrijke wijzigingen in huisvestiging? (CT) <i>Weging = 1</i>
Belangrijke wijzigingen in financiën? (CT) <i>Weging = 1</i>
Belangrijke overige wijzigingen? (CT) <i>Weging = 1</i>

## Bijlage II: Indicatoranalyse

De uitkomst van de expertanalyse kan gebruikt worden om de betrouwbaarheid en voorspellende waarde van verschillende indicatoren te bepalen. In de onderstaande analyse is hiertoe per school bekeken in welke mate er statistische samenhang bestaat tussen de score op een individuele indicator en de uitkomsten van de expertanalyse/oordelen. Hiervoor is een GLM (Generalized Linear Model) gebruikt met de indicator als onafhankelijke variabele en de uitkomst van de expertanalyse als uitkomst (afhankelijke) variabele. Hiervoor zijn de volgende categorieën gecodeerd als 'risico': EA: KO *risico*, Oordeel: *Onvoldoende of Zeer zwak en EA: Wel risico, geen KO*.

De uitkomsten van deze analyse zijn kansen (probabilities) tussen 0 en de 1 die aangeven hoe sterk een indicator is geassocieerd met de uitkomst van de expertanalyse. Een probability van 1 geeft hierbij aan dat risico op de indicator altijd overeenkomt met risico volgens de expertanalyse. Een probability van 0 geeft hierbij aan dat risico op de indicator altijd overeenkomt met géén risico volgens de expertanalyse (dus de indicator voorspelt de andere kant op dan we verwachten). Een probability van 0.5 geeft aan dat risico op de indicator met beide uitkomsten van de expertanalyse even sterk is geassocieerd, dit is vergelijkbaar met het opgooien van een munt. De resultaten van deze analyse worden weergegeven in Figuur 2.



**Figuur 2: Probability waarmee indicatoren uit de prestatie-monitor de uitkomst van de expertanalyse voorspellen. De horizontale balken geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van meest naar minst betrouwbaar.**

Er zijn 10 indicatoren die een hoge probability hebben voor het voorspellen van risico als uitkomst van de expertanalyse:

- CT Bestaansduur < 2 jaar
- CT Heeft de school de vragenlijst afgerond?
- CT Is de school begonnen de vragenlijst in te vullen?
- OP2 OVT: oordeel Zicht op ontwikkeling



- *CT Belangrijke wijziging(en) in lerarenteam? (vragenlijst)*
- *CT Belangrijke wijziging(en) in financiën? (vragenlijst)*
- *SK1 Percentage leerlingen dat zich niet veilig voelt (vragenlijst)*
- *SK1 Aantal leerlingen dat zich niet veilig voelt (vragenlijst)*
- *CT Belangrijke overige wijziging(en)? (vragenlijst)*
- *KA1 OVT: oordeel kwaliteitszorg*

Deze indicatoren zijn betrouwbare indicatoren. Voor indicatoren uit deze groep met een laag gewicht (bijvoorbeeld: *CT Bestaansduur < 2 jaar* of *OP2 OVT: oordeel Zicht op ontwikkeling*) is het overwegen waard om dit gewicht te verhogen.

Er zijn ook indicatoren met een probability van onder de 0,50:

- *OP1 OVT: oordeel Aanbod*
- *SK1 Wordt sociale veiligheid gemonitord? (vragenlijst)*
- *OV Cluster - aantal signalen overig*
- *CT Maximale krimp gedurende dit jaar, percentage*
- *CT Maximale groei gedurende dit jaar, percentage*
- *CT Belangrijke wijziging(en) in populatie? (vragenlijst)*
- *CT Maximale krimp gedurende dit jaar, aantallen*
- *CT Belangrijke wijziging(en) in huisvesting? (vragenlijst)*

Deze indicatoren voorspellen het tegenovergestelde van wat we zouden verwachten, namelijk dat risico op één van deze indicatoren juist geen risico voorspelt als uitkomst van de expertanalyse. De eerste vier hebben een hele lage probability van net boven de 0. Dit komt doordat er op beide indicatoren geen enkele school is die risico heeft op de indicator en waar de analist vervolgens een KO risico voorstel doet. Daarnaast zijn alle drie de indicatoren slecht gevuld, met elk slecht 1 of 2 scholen die risico hebben op de indicator.

Voor *CT Maximale groei gedurende dit jaar, percentage*, *CT Belangrijke wijziging(en) in populatie? (vragenlijst)*, *CT Maximale krimp gedurende dit jaar, aantallen* en *CT Belangrijke wijziging(en) in huisvesting? (vragenlijst)* zijn deze effecten minder groot, maar ook deze indicator voorspelt het tegenovergesteld van wat we zouden verwachten. Al deze 8 indicatoren werken dus eigenlijk een getrouwe risicoschatting door de prestatie-monitor tegen.

Tot slot zijn er ook nog indicatoren die gebaseerd zijn op een oordeel vanuit de inspectie, maar die niet gevuld zijn. Daarom valt de betrouwbaarheid van deze indicatoren ook niet te beoordelen.

### **Conclusie**

Er zitten 10 indicatoren in de Prestatiemonitor PO-NK die (zeer) betrouwbare voorspellers van de uitkomst van de expertanalyse zijn. Daarnaast zijn er ook 8 indicatoren waarvan deze analyse laat zien dat ze (op het moment) geen betrouwbare voorspellers van de uitkomst van de expertanalyse zijn. Voor deze indicatoren is het goed als de werkgroep bespreekt of ze deze willen behouden.



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

## **EVALUATIE VOORSPELKRACHT PRESTATIEMONITOR PO BO 2020**

5.1.2.e

februari 2021

## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 4</b>
<b>2</b>	<b>Registratie van de expertanalyse t.b.v. de evaluatie 5</b>
<b>3</b>	<b>Kwaliteit van de risicoschatting 6</b>
3.1	Indeling uitkomsten expertanalyse 6
3.2	Verdeling risico-scholen op basis van rang in de prestatiemonitor 7
3.3	Foutanalyse 8
3.4	Foutanalyse alleen aan de hand van oordelen 10
<b>4</b>	<b>Conclusie 13</b>
<b>5</b>	<b>Aanbevelingen 14</b>
5.1	Aanbevelingen voor toekomstige ontwikkelingen van de PM PO BO 14
5.2	Aanbevelingen voor de registratie van de expertanalyse 14
<b>Bijlage I: Foutanalyse per kwaliteitsgebied 15</b>	
<b>Bijlage II: Risico-indicatoren 17</b>	
<b>Bijlage III: Indicatoranalyse 18</b>	
Conclusie 19	

## Samenvatting

In dit document wordt een beschrijving gegeven van de evaluatie van de voorspelkracht van de Prestatiemonitor PO-BO aan de hand van de registratie van de expertanalyse door de sector. Deze evaluatie laat een overwegend positief beeld zien van de voorspelkracht van de Prestatiemonitor PO-BO. 88% van de scholen die door analisten of inspecteurs als risicovol zijn aangeduid kwamen ook als risicovol uit de prestatie-monitor. Aan dit percentage kleven echter belangrijke onzekerheden, met name omdat er slechts weinig scholen die een lage risicoscore hebben aan een expertanalyse worden onderworpen. Daarnaast kwam 75% van de scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de risicogroep van de prestatie-monitor. Hoewel wij denken dat de Prestatiemonitor PO-BO een belangrijke bijdrage kan leveren aan risicogericht toezicht binnen de sector verdient de mogelijkheid tot evaluatie de aandacht voor komend jaar.

Hoewel een groot deel van de scholen die als risicovol uit de prestatie-monitor kwamen ook volgens de expertanalyse risico's hebben mist de prestatie-monitor daarmee nog steeds een deel van de scholen die volgens de expertanalyse risico's bevatten. In totaal mist de prestatie-monitor 49 die naar inschatting van de analist wel risicovol zijn. Ook mist de prestatie-monitor 18 scholen die het oordeel *Onvoldoende* of *Zeer zwak* hebben gekregen. Er bestaat dus ook duidelijk ruimte voor verbetering van risicoschatting door de prestatie-monitor.

Daarnaast zijn er zoals genoemd ook enkele onzekerheden bij het interpreteren van de uitkomsten van de expertanalyse die de evaluatie bemoeilijken. Om de evaluatie te vergemakkelijken en risicoschatting te verbeteren doen wij aan het eind van dit document enkele aanbevelingen. Deze zullen in de toekomst de werking en evaluatie van de prestatie-monitor verbeteren.

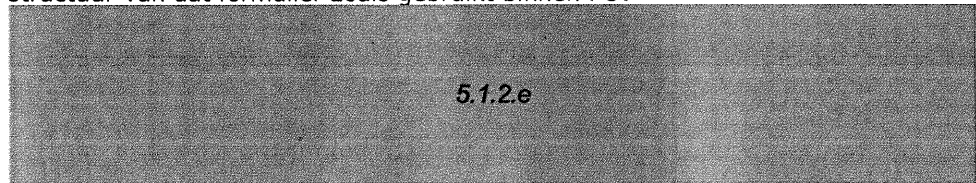
Ten slotte is het goed te benadrukken dat risicoschatting nooit alleen op basis van de prestatie-monitor moet plaatsvinden, maar altijd in combinatie met andere bronnen (zoals kwalitatieve kennis van inspecteurs en analisten over ovt's).

## **1 Inleiding**

De directie kennis ontwikkeld jaarlijks de zogenaamde Prestatiemonitors voor de verschillende sectoren. De prestatimonitors beogen inspecteurs en analisten te ondersteunen bij het maken van risicoprioritering, gebaseerd op data. Sinds 2018 bestaat er ook een Prestatiemonitor PO-Basisonderwijs (de PM PO-BO). Om de werking van de prestatimonitors te borgen en verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de Prestatiemonitor. Voor deze evaluatie worden de risicoscores van de prestatimonitor afgezet tegen enerzijds de beoordeling van risico's op scholen aan de hand van de expertanalyse (desk-research); en anderzijds tegen de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een schoolbezoek. In dit document wordt verslag gedaan van de evaluatie zoals uitgevoerd in december 2020.

## 2 Registratie van de expertanalyse t.b.v. de evaluatie

Om de risicoscores van de prestatie-monitor te vergelijken met de uitkomsten van de expertanalyse is het nodig om de uitkomsten van de expertanalyse op gestructureerde wijze vast te leggen. Dit jaar is er voor de tweede keer een dergelijke registratie bijgehouden. Dit is gebeurd aan de hand van een Excelformulier dat binnen de sector is bijgehouden. Hieronder een indruk van de structuur van dat formulier zoals gebruikt binnen PO:



Dit formulier vormde de basis voor de evaluatie die hieronder wordt beschreven, en is daarmee erg waardevol. Dat uiteindelijk de registratie nagenoeg compleet is gemaakt is een compliment waard. Wij erkennen als Kennis hoeveel werk daarin is gegaan. Zonder deze investering was deze evaluatie niet mogelijk geweest. Voor een volgende versie zien wij een volgend verbeterpunt. Waar we regelmatig tegen aanliepen bij de evaluatie is dat het niet altijd duidelijk is of een school met een hoge risicoscore volgens de analist eigenlijk *niet* risicovol is (een fout van de Prestatiemonitor), of dat een school mogelijk wel risicovol is maar dat er andere redenen waren om bij de school geen KO risico uit te voeren, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. Vaak stonden er echter wel beschrijvingen in het document die (deels handmatig) zijn gebruikt om toch duiding te geven aan de uitkomsten van de expertanalyse. Meer daarover in de volgende secties. Voor de toekomst raden wij aan om te zorgen voor een duidelijke scheiding in de registratie tussen de kwaliteit van de risicoschatting en de genomen vervolgstappen m.b.t. toezicht.

### 3 Kwaliteit van de risicoschatting

Om de kwaliteit van de risicoschatting van de Prestatiemonitor PO-BO inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken naar hoe de uitkomst van de expertanalyse door de analisten en het huidige oordeel zich verhouden tot de rang die scholen hebben in de prestatie-monitor, en of de school volgens de prestatie-monitor in de risico groep valt. e risicogrens vanaf wanneer een ovt werd meegenomen in de expertanalyse (EA) is in 2020 gelegd bij risicoscore  $\geq 0.29$ . Hiermee vallen 953 van de 6492 scholen in de risicocategorie.

#### 3.1 Indeling uitkomsten expertanalyse

Voor de foutanalyse hebben we de uitkomst van de expertanalyse ingedeeld naar 3 mogelijke uitkomsten: *EA: KO Risico*; *EA: Wel risico, geen KO*; *EA: Geen risico (en dus geen KO)*. In sommige gevallen is er echter besloten om bijvoorbeeld geen KO-risico uit te voeren omdat een school 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende school die nog even de tijd krijgt om herstelplan uit te voeren). Om deze scholen ook in beeld te brengen zijn voor die scholen die geen KO-risico krijgen maar die wel een huidig oordeel hebben ook de huidige oordelen meegenomen: *Zeer zwak*; *Onvoldoende*; *Voldoende*; en *Goed*. Oordelen worden meegenomen als ze gegeven zijn sinds augustus 2017.

Daarmee komen we tot de volgende categorieën:

1. **EA: KO risico**  
Als de analist een voorstel tot KO risico doet
2. **Oordeel Onvoldoende of Zeer zwak**  
Er is geen expertanalyse uitgevoerd, maar deze school heeft het oordeel *Onvoldoende* of *Zeer zwak*. Deze scholen tellen mee als risico in de foutanalyse.
3. **EA: Wel risico, geen KO**  
Als de analist geen voorstel doet voor een KO risico, maar wel risico's op de school constateert. Dit is bijvoorbeeld als vertrouwen is in het bestuur er een herstelonderzoek nog gaande is of net is afgerond of een verificatie plaats gaat vinden. Deze scholen tellen mee als risico in de foutanalyse.
4. **EA: Geen KO risico**  
De analist heeft de expertanalyse uitgevoerd, maar geen risico's geconstateerd die aanleiding zijn voor een KO-risico.
5. **Oordeel: Voldoende**  
Er is geen expertanalyse uitgevoerd en deze school heeft het oordeel *Voldoende*. Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed**  
Er is geen expertanalyse uitgevoerd en deze school heeft het oordeel *Goed*. Deze scholen tellen mee als geen risico in de foutanalyse.

Zoals te zien in Tabel 1 heeft een analist voor 65 scholen een voorstel gedaan voor een KO risico. 60 hiervan waren ook gedetecteerd door de prestatie-monitor (top 953). Op 204 scholen uit de top van de prestatie-monitor detecteerde de analist wel risico's maar deed deze geen voorstel voor een KO risico, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. Daarnaast hebben er 80 scholen uit de top 953 een oordeel *Onvoldoende* of *Zeer zwak*, 31 scholen buiten de top 953 hebben dit oordeel. Ook zijn er 17 scholen die het oordeel *Goed* hebben, maar ook in de top 953 van de prestatie-monitor staan, er zijn 156 scholen die het oordeel *Goed* hebben en buiten de top 953 staan.

In totaal zijn er 47 scholen waar de prestatie-monitor geen risico detecteerde, maar de analist wel risico's zag of waar de school een oordeel *Onvoldoende* of *Zeer zwak* heeft. Bij 5 van deze 47 stelt de analist ook een KO risico voor (Tabel 1). Bij de overige 42 scholen zijn het risico's die al bekend zijn, is er vertrouwen in het bestuur of het betreft een oordeel *Onvoldoende* of *Zeer zwak*.

**Tabel 1: Aantal scholen per uitkomst van de expertanalyse en risico op de prestatie-monitor.**

		Risico op de prestatie-monitor	
		Risico (1 – 953)	Geen risico (954 – 6492)
Uitkomst expertanalyse	EA: KO risico	60	5
	Oordeel: <i>Onvoldoende</i> of <i>Zeer zwak</i>	80	31
	EA: Wel risico, geen KO	204	11
	EA: Geen KO risico	575	24
	Oordeel: <i>Voldoende</i>	4	275
	Oordeel: <i>Goed</i>	17	156
	EA: Geen expert analyse of oordeel	13	5037

### 3.2

#### Verdeling risico-scholen op basis van rang in de prestatie-monitor

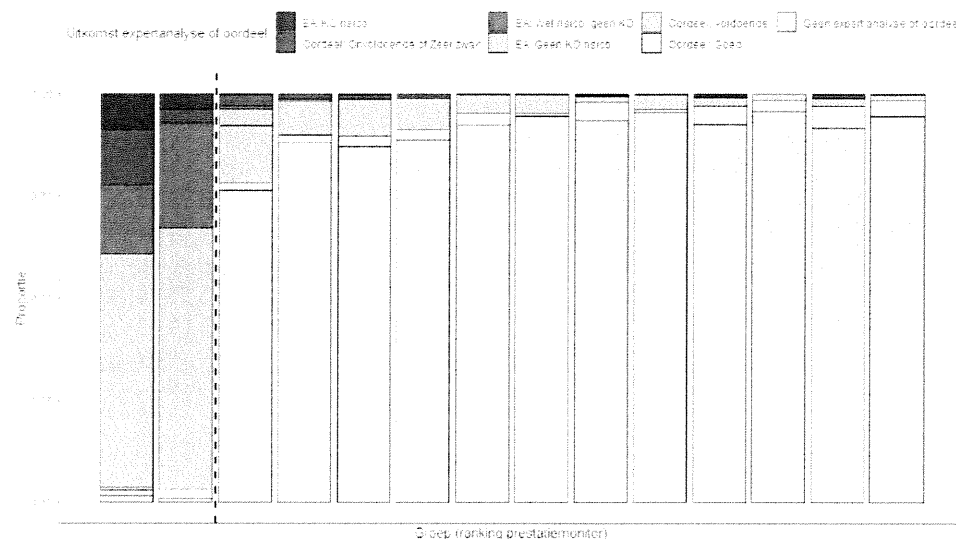
De uitkomsten van de expertanalyse kunnen verdeeld worden over de rangorde in de prestatie-monitor. Hiervoor zijn de scholen onderverdeeld in groepen op basis van hun rang in de prestatie-monitor. De 953 scholen die risico hebben volgens de prestatie-monitor zijn opgesplitst in twee groepen van resp. 476 en 477 scholen. De overige 5539 scholen zijn verdeeld in 12 groepen van 461 of 462 scholen (Figuur 1).

In Figuur 1 valt te zien dat het gros van de scholen die een KO-risico kregen -of al een onvoldoende oordeel hadden- ook in de hoogste risico categorieën terecht zijn gekomen. Daarnaast bevatten deze groepen ook een aantal scholen waarbij wel risico's zijn waargenomen, maar waarvoor niet tot een KO risico is besloten.

Deze resultaten laten zien dat er een behoorlijke voorspellende waarde lijkt uit te gaan van de prestatie-monitor. Desalniettemin worden niet alle risicovolle of onvoldoende scholen in de hoogste risicocategorieën geplaatst. In de derde groep, bijvoorbeeld, constateert de expertanalyse risico op 5 scholen. Dit is iets hoger dan in andere groepen buiten de top van de prestatie-monitor (groep 4 en hoger), waar het aantal scholen met risico volgens de expertanalyse varieert tussen de 1 en de 4. Ook zit er nog een relatief groot deel met scholen die het oordeel *Onvoldoende* of *Zeet zwak* hebben, net onder de grens van risico volgens de prestatie-monitor: In groep 3 zijn dit 12 scholen, in groep 4 zijn dit er 6 en in groep 5 en 6 zijn dit er beide 5. Vanaf groep 6 en hoger (de laagste helft) komen dergelijke risicovolle scholen echter nog maar sporadisch voor.

Een belangrijk aspect aan deze analyse is echter dat voor bijna alle scholen in groepen 3 tot en met 14 er afgelopen jaar geen expert analyse uitgevoerd en is er ook geen recent oordeel is gegevens (de grijze delen in Figuur 1). In feite hebben we geen goede basis om te bepalen hoeveel van deze scholen voldoende of onvoldoende zijn. Dit vormt een belangrijk probleem bij het uitvoeren en evalueren van risicogericht toezicht.





**Figuur 1: De relatie tussen scores op de prestatiemonitor en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 14 groepen. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de expertanalyse (EA) of Oordeel voorkomt per groep op basis van de rangorde in de prestatiemonitor. De eerste twee groepen zijn de 953 scholen die risico hebben op de prestatiemonitor, de 12 groepen daarna bestaan steeds uit 462/461 scholen.**

### 3.3

#### Foutanalyse

Foutanalyse is een methode om de kwaliteit van risicoschatting te kwantificeren. Dit is zinvol omdat daarmee op termijn de kwaliteit van risicoschattingen over de tijd en over verschillende sectoren/kwaliteitsgebieden vergeleken kan worden. Foutanalyse berust op een onderverdeling van uitkomsten in vier groepen (voortkomend uit Tabel 1):

- **Echt positieven (terecht alarm):** Zowel de prestatiemonitor als de expertanalyse detecteren risico's voor een school
- **Fout negatieven (gemist geval):** De prestatiemonitor detecteert geen risico's, maar de expertanalyse constateert dat er wél risico's zijn op een school.
- **Fout positieven (loos alarm):** De prestatiemonitor detecteert risico, maar de expertanalyse constateert dat er géén risico's zijn op een school.
- **Echt negatieven (terecht genegeerd):** Zowel de prestatiemonitor als de expertanalyse detecteren geen risico's voor een school.

Zo krijgen we Tabel 2.

**Tabel 2: Foutanalyse op de uitkomst van de prestatie-monitor en de expertanalyse**

		Risico op de prestatie-monitor	
		Risico	Geen risico
Risico volgens de analist	Risico	344 (Echt positieven)	47 (Fout negatieven)
	Geen risico	609 (Fout positieven)	5492 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Hoewel er belangrijke onzekerheden zijn over de niet-bezochte scholen kunnen we wel kijken naar de scholen die wel aan een expertanalyse zijn onderworpen (of een recent oordeel hebben gehad). Vanuit die blik zien we dat zowel op sensitiviteit als op specificiteit de prestatie-monitor heel behoorlijk lijkt te presteren. Zo is 88% van de bekende risicovolle scholen boven de gestelde risicogrens gekomen. Dit betekent dat de prestatie-monitor een bijdrage lijkt te kunnen leveren aan het detecteren van scholen die mogelijk risico's hebben. Op precisie scoort de prestatie-monitor minder (36% van de onderzochte scholen is vervolgens gewaardeerd als risicovol). Dit betekent dat er relatief veel scholen moeten worden bekeken in de expertanalyse omdat ze hoog scoren op de prestatie-monitor, maar waar onvoldoende aanleiding is voor een KO risico. Een verklaring hiervoor is dat het optimaliseren van de sensitiviteit en specificiteit vaak ten koste gaat van de precisie (precisie wordt bijvoorbeeld hoger wanneer we slechts de 100 scholen met de allerhoogste risicoscores onderzoeken, maar dan wordt de sensitiviteit echter veel lager). Een lage precisie betekent dus dat er relatief veel capaciteit nodig is geweest van de analisten voor de expertanalyse. Het is dus van belang een balans te vinden tussen een hoge sensitiviteit en specificiteit nastreven en te zorgen dat de precisie niet te laag wordt, omdat dit zal leiden tot overvragen van de analisten. Een andere mogelijke verklaring hiervoor is dat er geen KO risico wordt gepland omdat er vertrouwen in het bestuur is en/of omdat de risico's bekend zijn, maar dat dit vanuit de huidige manier van documenteren niet duidelijk wordt en gebruikt kan worden in de evaluatie van de prestatie-monitor. Dit zou verbeterd kunnen worden door de manier bij het documenteren van de expertanalyse een onderscheid te maken tussen of de analist risico's constateert op een school en het doen van een voorstel tot een toezichtsactiviteit.

In Bijlage I hebben we op dezelfde manier een foutanalyse per kwaliteitsgebied gedaan en in Bijlage II is te zien welke indicatoren er per kwaliteitsgebied ontwikkeld zijn..

### 3.4 Foutanalyse alleen aan de hand van oordelen

Naast de kwaliteit van de risicoschatting bepalen aan de hand van de uitkomst van expertanalyses en recente oordelen, kunnen we ook kijken naar de uitkomst van recente oordelen. Deze manier van analyseren heeft als nadeel dat er minder data beschikbaar is (er worden namelijk meer expertanalyses uitgevoerd dan oordelen gegeven). Het grote voordeel van het doen van een foutanalyse aan de hand van oordelen is dat de data betrouwbaarder is omdat de school ook daadwerkelijk bezocht is.

Voor de foutanalyse delen we oordelen in zeven categorieën:

1. **Oordeel: Onvoldoende of Zeer zwak, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
2. **Oordeel: Onvoldoende of Zeer zwak, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
3. **Oordeel: Voldoende, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
4. **Oordeel: Voldoende, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
5. **Oordeel: Goed, al nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 juli 2020 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
7. **Geen recent oordeel bekend**  
Deze scholen tellen mee als geen risico in de foutanalyse

Op dezelfde manier als Figuur 1 kunnen we de scholen over 12 groepen verdelen. Hier bestaan de eerste 2 staven uit de 43 scholen die risico hebben volgens de prestatie-monitor en de overige 10 groepen daarna bestaan steeds uit 23/22 scholen. Voor ongeveer twee derde van de top 43 van de prestatie-monitor is er een recent oordeel bekend. Bij de 4 scholen uit de risicogroepen is het oordeel *Onvoldoende* of *Zeer zwak* gegeven. Daarnaast zijn er 2 scholen met het oordeel *Onvoldoende* of *Zeer zwak* die buiten de risicogroep van de prestatie-monitor vallen.



**Figuur 2: De relatie tussen scores op de prestatie monitor en recente oordelen, sinds 1 augustus 2017. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 14 groepen. Iedere staaf geeft de proportie weer van de hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatie monitor. De eerste twee groepen zijn de 953 scholen die risico hebben op de prestatie monitor, de 12 groepen daarna bestaan steeds uit 462/461 scholen.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 2, krijgen we Tabel 3.

**Tabel 3: Foutanalyse op de uitkomst van de prestatie monitor en recente oordelen**

		Risico op de prestatie monitor	
		Risico	Geen risico
Risico volgens recent oordeel	Risico	55 (Echt positieven)	18 (Fout negatieven)
	Geen risico	898 (Fout positieven)	5521 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.

- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Deze analyse aan de hand van de oordelen gegeven sinds augustus 2017 laat zien dat ook aan de hand van recente oordelen de sensitiviteit van de prestatie-monitor hoog is. De prestatie-monitor PO-BO detecteert 75% van de scholen die het oordeel *Onvoldoende* of *Zeet zwak* hebben, hiermee mist de prestatie-monitor 18 risicovolle scholen. Hiermee valt te concluderen dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die in aanmerking komen voor een KO risico onderzoek en het oordeel *Onvoldoende* of *Zeet zwak* krijgen. Wel is het zo dat er weinig recente oordelen bekend zijn en dat het dus onzeker is hoeveel scholen dus daadwerkelijk onvoldoende zijn. Dit is met name een probleem voor de scholen die een lage risicoscore krijgen op de prestatie-monitor. Omdat er voor scholen met een lage risicoscore minder recente oordelen bekend zijn.

## 4 Conclusie

Binnen de mogelijkheden voor evaluatie schets de hier uitgevoerde analyse een overwegend positief beeld over de kwaliteit van risicoschatting door de Prestatiemonitor PO-BO: 88% van de scholen die als risicovol uit de expertanalyse komen -of die al een onvoldoende oordeel hebben- kwamen terecht in de hoogste risicogroepen. Daarnaast kwam 75% van de scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de risicogroep van de prestatie-monitor. Dit suggereert dat er toegevoegde waarde is voor het gebruik van de PM PO-BO bij het prioriteren van scholen in het toezicht.

Tegelijkertijd is er ook ruimte voor verbetering. Immers, niet álle risicovolle scholen kwamen in de twee hoogste risicogroepen terecht. Het zal dus een belangrijk doel zijn om de kwaliteit van risicoschatting verder te verbeteren. Een van de belangrijkste bijdragen van de methode van evaluatie zoals hier beschreven is dat de ontwikkeling van de voorspelkracht in de komende jaren ook gemonitord kan worden. Daarmee kan er gericht en meetbaar gewerkt worden aan verbetering van risicoschatting.

Daarnaast is een belangrijk aandachtspunt dat er veel scholen zijn waar we weinig over weten (de scholen die onder de risicogrens vallen). Dit maakt dat de evaluatie van de kwaliteit van de risicoschatting ook in belangrijke mate beperkt is. Dit is urgent omdat de kwaliteit van de risicoschatting een van de belangrijkste factoren is die de invulling van risicogericht toezicht bepaalt.

Als laatste is het goed om op te merken dat een perfecte risicoschatting onmogelijk is. Risico's veranderen immers over de tijd, en er zijn ook risico's waar we op dit moment simpelweg geen data over hebben (zoals bijvoorbeeld de kwaliteit/ervaring van bestuurders, hoewel hier wel aan gewerkt wordt vanuit het datascience programma), of de relatie tussen bestuurders, schoolleiders en leraren. Het detecteren van risicoscholen zal daarom altijd moeten bestaan uit een *combinatie* van zowel datagedreven risicoschatting (zoals in de prestatie-monitor) als het gebruik van de impliciete ('zachte') kennis van domeinexperts zoals inspecteurs of analisten, en bijvoorbeeld signalen.

## 5 Aanbevelingen

### 5.1 Aanbevelingen voor toekomstige ontwikkelingen van de PM PO BO

- 1) Zoals beschreven in dit document bestaat er onzekerheid in de schatting van kwaliteitsmaten van risicoschatting. Dit komt omdat we simpelweg niet goed weten in hoeverre er risico's bestaan op scholen die een lage risicoscore hebben en het volgens de prestatiemonitor dus juist goed doen (en die ook niet in beeld zijn bij de COI). Hoewel we niet alle scholen kunnen en willen onderzoeken, zou het wel mogelijk zijn om onderzoeken uit te voeren onder scholen met een lage risicoscore aan de hand van steekproeven. De prestatiemonitor zou ook gebruikt kunnen worden om op basis van de rangorde een steekproef te trekken en deze scholen te bekijken, als vorm van proportionaliteit in het toezicht. Dit kan helpen om de *prevalentie* van onvoldoende scholen inzichtelijk te maken. Dit zou sterk kunnen bijdragen aan de verdere verbetering van (datagedreven) risicogericht toezicht.
- 2) Als laatste kunnen de uitkomsten van de evaluatie ook gebruikt worden om verbeteringen in indicatorselectie en weging toe te passen op de prestatiemonitor. In Bijlage III (zie hieronder) hebben we hiertoe een aanvullende analyse uitgevoerd.

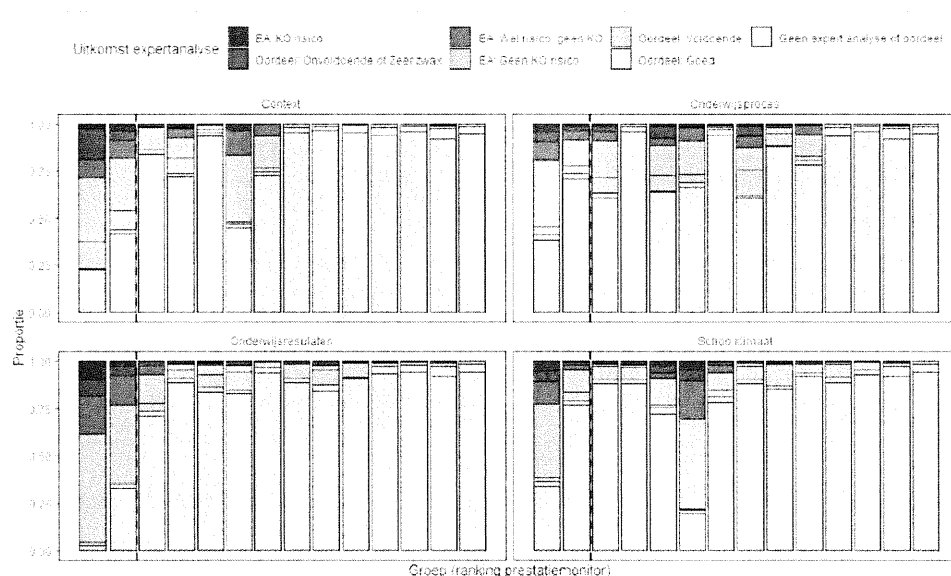
### 5.2 Aanbevelingen voor de registratie van de expertanalyse

- 1) Voor deze analyse is gebruik gemaakt van de registratie van de expertanalyse in een Exceltabel. Voor toekomstige evaluatie en verbetering van de prestatiemonitor zou het echter helpen om 'tot een nog duidelijkere scheiding te komen tussen het bestaan van risico's (ofwel, vindt de analist de risicoscore terecht) en de beslissing voor vervolgtoezicht (wordt er een KO-risico uitgevoerd). Met name het eerste is belangrijk voor een goede evaluatie van de prestatiemonitor.  
Hiertoe stellen wij voor om de kolom '*KO-risico? Voorstel analist*' op te splitsen in twee kolommen. Bijvoorbeeld één kolom '*Zijn er risico's op de school?*' en één kolom '*KO-risico voorstel*'.
- 2) Daarnaast waren er dit jaar problemen met data-registratie over de expertanalyse. Dit resulteerde in fouten die handmatig omgezet moesten worden bij de evaluatie. De oorzaak hiervan ligt bij Proza dat niet geschikt is voor een dergelijke registratie. Voor komend jaar is het goed om te kijken of de procedure van het registreren van de expertanalyse effectiever kan.
- 3) Scholen die volgens de prestatiemonitor niet in de risicogroep vallen, maar waar wel een KO risico voorstel gedaan wordt door de analist zijn zeer interessant voor de evaluatie. Daarom is het handig als uit de documentatie duidelijk wordt waarom er een KO risico voorstel en hoe deze school in beeld is gekomen (bijvoorbeeld via de contactinspecteur of via onderwijsresultaten).

## Bijlage I: Foutanalyse per kwaliteitsgebied

De risicoscores in de prestatie-monitor zijn opgebouwd uit risicoscores op verschillende kwaliteitsgebieden. Bij de Prestatiemonitor PO-BO zijn dit *Onderwijsresultaten*; *Onderwijsproces*; *Schoolklimaat*; *Context*; *Kwaliteitszorg en Ambitie*; en *Financieel Beheer*. In Bijlage II is een overzicht opgenomen van welke indicatoren onder welk kwaliteitsgebied vallen.

Een zelfde soort figuur als Figuur 1 kan gemaakt worden per kwaliteitsgebied. Hierbij is de uitkomst van de expertanalyse door de analist nog steeds genomen als uitkomst-maat. Echter zijn de groepen op de x-as nu verdeeld op basis van de rangorde op één kwaliteitsgebied. Dit is mogelijk voor vier van de zes kwaliteitsgebieden, omdat de kwaliteitsgebieden *Kwaliteitszorg en ambitie* en *Financieel beheer* elk maar gevuld worden door één indicator en er dus geen rangorde gemaakt kan worden op deze kwaliteitsgebieden.



**Figuur 3: De relatie tussen scores op de prestatie-monitor per kwaliteitsgebied en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore per kwaliteitsgebied, en samengevoegd in 14 groepen. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de expertanalyse/oordeel voorkomt per groep op basis van de rangorde. De eerste twee groepen zijn de 953 scholen die risico hebben op de prestatie-monitor, de 12 groepen daarna bestaan steeds uit 462/461 scholen.**

Uit Figuur 3 kan de sensitiviteit, specificiteit en precisie per kwaliteitsgebied te berekenen. Zo krijgen we Tabel 3.

**Tabel 3: Waardes voor de foutanalyse per kwaliteitsgebied en de gehele prestatie-monitor.**

	Sensitiviteit	Specificiteit	Precisie
Context	0,58	0,88 - 0,93	0,24 - 0,35
Onderwijsproces	0,35	0,87 - 0,95	0,14 - 0,31
Onderwijsresultaten	0,76	0,89 - 0,92	0,31 - 0,38



Schoolklimaat	0,34	0,87 – 0,95	0,14 – 0,31
<b>Gehele prestatiemonitor</b>	<b>0,88</b>	<b>0,90</b>	<b>0,36</b>

In Tabel 3 is te zien dat de verdeling van risico-scholen op het kwaliteitsgebied *Onderwijsresultaten* het meest overeenkomt met de uitkomsten van de expertanalyse/oordelen, en daarmee het meest lijkt op de relatie zoals waargenomen voor de totaal-risicoscore van de prestatiemonitor. Dit is ook logisch omdat de twee indicatoren met de zwaarste weging *Percentage 1S/2F behaald* en *Gemiddelde Eindtoetsscore* beide onder het kwaliteitsgebied *Onderwijsresultaten* valt. Voor de overige drie kwaliteitsgebieden komt verdeling van risico-scholen minder goed overeen met de algehele rangorde in de prestatiemonitor. Wel is voor alle drie deze kwaliteitsgebieden zo dat de groepen met de laagste rang in de prestatiemonitor niet gevuld zijn met risico scholen. Dus alle kwaliteitsgebieden detecteren de risico scholen in elk geval tot op een zekere hoogte. Daarnaast is het zo dat de prestatiemonitor als geheel een betere voorspelkracht heeft dan elk kwaliteitsgebied apart.

## Bijlage II: Risico-indicatoren

De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

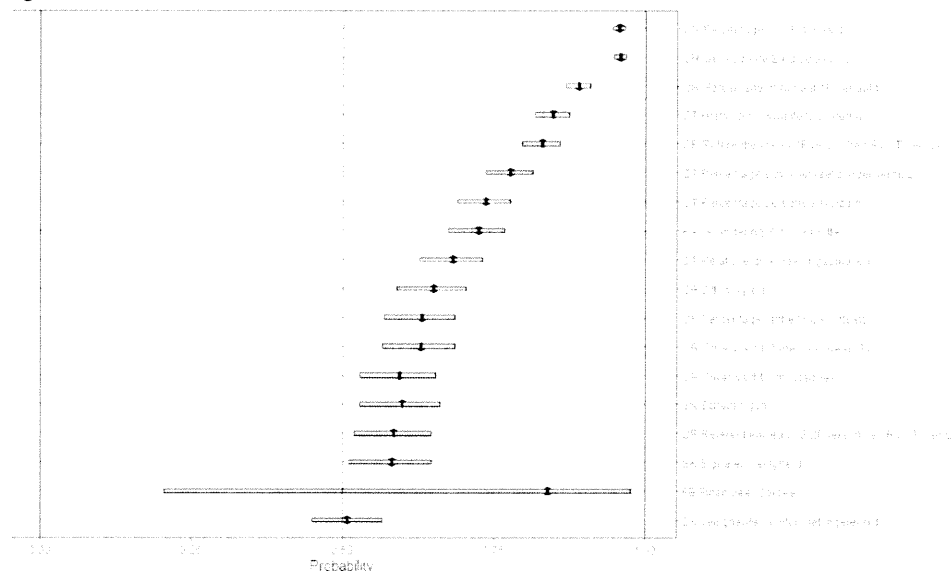
Onderwijs- resultaten	Onderwijs- proces	School- klimaat	Kwaliteitszo- rg en ambitie	Financieel beheer	Context
Gemiddelde Eindtoetsscor- e (OR1) <i>Weging = 4</i>	Onverwach- t Schooladvi- es (OP2) <i>Weging = 2</i>	Signalen Veiligheid (SK1) <i>Weging = 1</i>	Verandering Schoolgrootte (KA2) <i>Weging = 2</i>	Financieel Oordeel (FB) <i>Weging = 1</i>	Historische Kwaliteitsoordeel en (CT) <i>Weging = 2</i>
Percentage minimaal 1F behaald (OR1) <i>Weging = 1</i>	Zittenblijv- ers (OP2) <i>Weging = 2</i>	Schorsingen (SK1) <i>Weging = 1</i>			Percentage personeelsverzui- m (CT) <i>Weging = 2</i>
Percentage 1S/2F behaald (OR1) <i>Weging = 4</i>	Referentie- niveau 1F verschillen (OP3) <i>Weging = 1</i>	Veiligheidsm- onitor niet ingeleverd (SK1) <i>Weging = 1</i>			Percentage schoolwissel zonder verzuim (CT) <i>Weging = 1</i>
Onverwacht Onderwijs- niveau 3VO (OR3) <i>Weging = 1</i>	Referentie- niveau 2F verschillen (OP3) <i>Weging = 1</i>				Relatieve groei leerlingaantallen (CT) <i>Weging = 1</i>
	Percentage ontheffing eindtoets (OP8) <i>Weging = 1</i>				

Bron: Van Gulik-Gulikers, I. (2021) *De prestatiemonitor bekeken vanuit de werkgroep prestatiemonitor basisonderwijs*

## Bijlage III: Indicatoranalyse

De uitkomst van de expertanalyse kan gebruikt worden om de betrouwbaarheid en voorspellende waarde van verschillende indicatoren te bepalen. In onderstaande analyse is hiertoe per school bekeken in welke mate er statistische samenhang bestaat tussen de score op een individuele indicator en de uitkomsten van de expertanalyse/oordelen. Hiervoor is een GLM (Generalized Linear Model) gebruikt met de indicator als onafhankelijke variabele en de uitkomst van de expertanalyse als uitkomst (afhankelijke) variabele. Hiervoor zijn de volgende categorieën gecodeerd als 'risico': EA: *KO risico*, Oordeel: *Onvoldoende of Zeer zwak en EA: Wel risico, geen KO*.

De uitkomsten van deze analyse zijn kansen (probabilities) tussen 0 en de 1 die aangeven hoe sterk een indicator geassocieerd is met de uitkomst van de expertanalyse. Een probability van 1 geeft hierbij aan dat risico op de indicator altijd overeenkomt met risico volgens de expertanalyse. Een probability van 0 geeft hierbij aan dat risico op de indicator altijd overeenkomt met géén risico volgens de expertanalyse (dus de indicator voorspelt de andere kant op dan we verwachten). Een probability van 0.5 geeft aan dat risico op de indicator met beide uitkomsten van de expertanalyse even sterk is geassocieerd, dit is vergelijkbaar met het opgooien van een munt. De resultaten van deze analyse worden weergegeven in Figuur 3.



**Figuur 4: Probability waarmee indicatoren uit de prestatiemonitor de uitkomst van de expertanalyse voorspellen. De horizontale balken geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van meest naar minst betrouwbaar.**

OR percentage 1S/2F behaald en OR gemiddelde Eindtoetsscore -met beide een probability van 0,96- de beste voorspellers van risico als de uitkomst van de expertanalyse (Figuur 4). Dit komt overeen met de verwachtingen omdat deze indicatoren beide ook het grootste gewicht in de prestatiemonitor hebben (een gewicht van 4). Indicatoren die laag scoren en waarbij de ondergrens van het 95% betrouwbaarheidsinterval onder 0,50 ligt, zijn:

*FB Financieel Oordeel*  
*SK Veiligheidsmonitor niet ingeleverd*

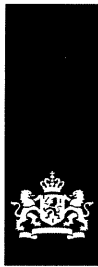
Voor deze indicatoren is het nog niet uitgesloten dat het mogelijk betrouwbare indicatoren zijn. Dit zal verder gemonitord moeten worden. Voor *FB Financieel Oordeel* is de vulling van de indicator de oorzaak dat de indicator onbetrouwbaar is, omdat deze voor slechts 4 scholen gevuld is. Aangezien dit inspectie oordelen betreft, valt dit niet direct te verbeteren. Dit kan echter nog steeds zeer veel waarde bevatten omdat de puntschatting wel héél hoog ligt. Sociale veiligheid monitor is in het schooljaar 2019-2020 compleet niet gevuld en in de jaren hiervoor slechts gebrekkig. Met verbetering van de vulling van de data kan deze indicator mogelijk meer voorspellend werken.

Voor de indicatoren *OR Percentage minimaal 1F behaald* is het te overwegen deze een hoger gewicht toe te kennen. Deze indicator heeft een hoge betrouwbaarheid, maar heeft wel slechts een gewicht van 1 in de prestatie-monitor. Mogelijk zou de verzwaring van deze indicator wel resulteren in een sterkere focus op de eindtoets, waardoor de vraag is of dit een wenselijk effect is. Waardoor vooral scholen met een hoge schoolweging risicovol zullen scoren.

### **Conclusie**

De twee indicatoren met het hoogste gewicht in de prestatie-monitor, *OR percentage 1S/2F behaald* en *OR gemiddelde Eindtoetsscore*, zijn ook de meest betrouwbare voorspellers van de uitkomst van de expertanalyse. De indicator, *OR Percentage minimaal 1F behaald* heeft nog een laag gewicht in de prestatie-monitor. Deze zou volgend jaar misschien een hoger gewicht kunnen krijgen.

Er zijn geen indicatoren met een negatief effect op de voorspelling. Wel zijn er twee indicatoren waar de ondergrens van het 95% betrouwbaarheidsinterval onder de 0,5 is. Voor alle drie de indicatoren zullen we dit de komende periode monitoren in hoeverre het zinvolle toevoegingen blijven aan de prestatie-monitor.



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

6.

## **EVALUATIE VOORSPELKRACHT PRESTATIEMONITOR SBO 2020**

**5.1.2.e**

februari 2021

## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 4</b>
<b>2</b>	<b>Registratie van de expertanalyse t.b.v. de evaluatie 5</b>
<b>3</b>	<b>Kwaliteit van de risicoschatting 6</b>
3.1	Indeling uitkomsten expertanalyse 6
3.2	Verdeling risico-scholen op basis van rang in de prestatie-monitor 7
3.3	Foutanalyse 8
3.4	Foutanalyse alleen aan de hand van oordelen 10
<b>4</b>	<b>Conclusie 13</b>
<b>5</b>	<b>Aanbevelingen 14</b>
5.1	Aanbevelingen voor toekomstige ontwikkelingen van de PM SBO 14
5.2	Aanbevelingen voor de registratie van de expertanalyse 14
<b>Bijlage I: Foutanalyse per kwaliteitsgebied 15</b>	
<b>Bijlage II: Risico-indicatoren 17</b>	
<b>Bijlage III: Indicatoranalyse 18</b>	
Conclusie 19	

## Samenvatting

In dit document wordt een beschrijving gegeven van de evaluatie van de voorspelkracht van de Prestatiemonitor SBO aan de hand van de registratie van de expertanalyse door de sector. Deze evaluatie laat een redelijk positief beeld zien van de voorspelkracht van de Prestatiemonitor SBO. 76% van de scholen die door analisten of inspecteurs als risicovol zijn aangeduid kwamen ook als risicovol uit de prestatie-monitor. Daarnaast kwam 67% van de scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de risicogroep van de prestatie-monitor. Daarmee kan de Prestatiemonitor SBO een belangrijke bijdrage leveren aan risicogericht toezicht binnen de sector.

Desalniettemin mist de prestatie-monitor daarmee nog steeds een deel van de risicovolle scholen. In totaal mist de prestatie-monitor 10 risicovolle scholen. Er bestaat dus ook duidelijk ruimte voor verbetering van risicoschatting door de prestatie-monitor. Bovendien zijn er ook enkele onzekerheden bij het interpreteren van de uitkomsten van de expertanalyse die de evaluatie bemoeilijken. Om de evaluatie te vergemakkelijken en risicoschatting te verbeteren doen wij aan het eind van dit document enkele aanbevelingen. Deze zullen in de toekomst de evaluatie van de prestatie-monitor verbeteren.

Ten slotte is het goed te benadrukken dat risicoschatting nooit alleen op basis van de prestatie-monitor moet plaatsvinden, maar altijd in combinatie met andere bronnen (zoals kwalitatieve kennis van inspecteurs en analisten over scholen).

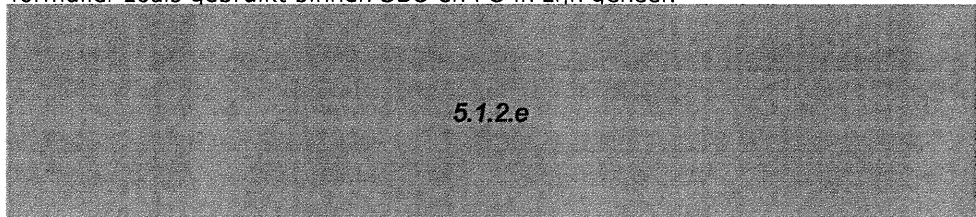
## **1 Inleiding**

De directie Kennis ontwikkelt jaarlijks de zogenaamde prestatie-monitors voor de verschillende sectoren. De prestatie-monitors beogen inspecteurs en analisten te ondersteunen bij het maken van risicoprioritering, gebaseerd op data. Sinds 2019 bestaat er ook een Prestatiemonitor Speciaal Basisonderwijs (de PM SBO). Om de werking van de prestatie-monitors te borgen en te verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de prestatie-monitor. Voor deze evaluatie worden de risicoscores van de prestatie-monitor afgezet tegen enerzijds de beoordeling van risico's op scholen aan de hand van de expertanalyse (deskresearch); en anderzijds tegen de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een inspectie. In dit document wordt verslag gedaan van de evaluatie zoals uitgevoerd in december 2020.



## 2 Registratie van de expertanalyse t.b.v. de evaluatie

Om de risicoscores van de prestatie-monitor te vergelijken met de uitkomsten van de expertanalyse is het nodig om de uitkomsten van de expertanalyse op gestructureerde wijze vast te leggen. Dit jaar is er voor tweede keer met een dergelijke registratie begonnen. Dit is gebeurd aan de hand van een Excelformulier dat binnen de sector is bijgehouden. Hieronder een indruk van de structuur van dat formulier zoals gebruikt binnen SBO en PO in zijn geheel:



5.1.2.e

Dit formulier vormde de basis voor de evaluatie die hieronder wordt beschreven, en is daarmee erg waardevol. Dat uiteindelijk de registratie nagenoeg compleet is gemaakt is een compliment waard. Wij erkennen als Kennis hoeveel werk daarin is gegaan. Zonder deze investering was deze evaluatie niet mogelijk geweest. Voor een volgende versie zien wij een volgend verbeterpunt. Waar we tegen aanliepen bij de evaluatie is dat het niet altijd duidelijk is of een school met een hoge risicoscore volgens de analist eigenlijk *niet* risicovol is (een fout van de Prestatiemonitor), of dat een school mogelijk wel risicovol is maar dat er andere redenen waren om bij de school geen kwaliteitsonderzoek bij risico (KO risico) uit te voeren, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. Vaak stonden er echter wel beschrijvingen in het document die (deels handmatig) zijn gebruikt om toch duiding te geven aan de uitkomsten van de expertanalyse. Meer daarover in de volgende secties. Voor de toekomst raden wij aan om te zorgen voor een duidelijke scheiding in de registratie tussen de kwaliteit van de risicoschatting en de genomen vervolgstappen m.b.t. toezicht.

### 3 Kwaliteit van de risicoschatting

Om de kwaliteit van de risicoschatting van de Prestatiemonitor SBO inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken hoe de uitkomst van de expertanalyse door de analisten en het huidige oordeel zich verhouden tot de rang die scholen hebben in de prestatie-monitor, en of de school volgens de prestatie-monitor in de risicogroep valt. De risicogrens vanaf wanneer een ovt werd meegenomen in de expertanalyse (EA) is in 2020 gelegd bij risicoscore  $\geq 0.23$ . Hiermee vallen 43 van de 269 scholen in de risicocategorie. Hiernaast zaten er ook nog drie SBO-scholen in de dataset die niet voorkwamen in de prestatie-monitor, deze scholen zijn niet meegenomen in deze analyse.

#### 3.1 Indeling uitkomsten expertanalyse

Voor de foutanalyse hebben we de uitkomst van de expertanalyse ingedeeld naar 3 mogelijke uitkomsten: *EA: KO Risico*; *EA: Wel risico, geen KO*; *EA: Geen risico (en dus geen KO)*. In sommige gevallen is er echter besloten om bijvoorbeeld geen KO-risico uit te voeren omdat een school 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende school die nog even de tijd krijgt om herstelplan uit te voeren). Om deze scholen ook in beeld te brengen zijn voor de scholen die geen KO-risico krijgen maar die wel een huidig oordeel hebben ook de huidige oordelen meegenomen: *Zeer zwak*; *Onvoldoende*; *Voldoende*; en *Goed*. Oordelen worden meegenomen als ze gegeven zijn sinds augustus 2017. Daarmee komen we tot de volgende categorieën:

1. **EA: KO risico**  
Als de analist een voorstel tot KO risico doet.
2. **Oordeel Onvoldoende of Zeer zwak**  
Er is geen expertanalyse uitgevoerd, maar deze school heeft het oordeel *Onvoldoende* of *Zeer zwak*. Deze scholen tellen mee als risico in de foutanalyse.
3. **EA: Wel risico, geen KO**  
Als de analist geen voorstel doet voor een KO risico, maar wel risico's op de school constateert. Dit is bijvoorbeeld als vertrouwen is in het bestuur er een herstelonderzoek nog gaande is of net is afgerond of een verificatie plaats gaat vinden. Deze scholen tellen mee als risico in de foutanalyse.
4. **EA: Geen KO risico**  
De analist heeft de expertanalyse uitgevoerd, maar geen risico's geconstateerd die aanleiding zijn voor een KO-risico.
5. **Oordeel: Voldoende**  
Er is geen expertanalyse uitgevoerd en deze school heeft het oordeel *Voldoende*. Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed**  
Er is geen expertanalyse uitgevoerd en deze school heeft het oordeel *Goed*. Deze scholen tellen mee als geen risico in de foutanalyse.

Zoals te zien in Tabel 1 heeft een analist voor 7 scholen een voorstel gedaan voor een KO risico, deze vallen allemaal in de risicogroep van de prestatie-monitor (top 43). Op 8 scholen uit de top van de prestatie-monitor detecteerde de analist wel risico's maar deed deze geen voorstel voor een KO risico, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. 6 SBO scholen hebben het oordeel *Onvoldoende* of *Zeer zwak*, 4 van deze scholen vallen in de risicogroep van de prestatie-monitor en 2 vallen daarbuiten. 8 scholen hebben het oordeel *Goed*, 1 hiervan valt in de risicogroep van de prestatie-monitor, 7 vallen erbuiten.

In totaal zijn er 6 scholen waar de prestatie-monitor geen risico detecteerde, maar de analist wel risico's zag of waar de school een oordeel *Onvoldoende* of *Zeer zwak* heeft. Bij 4 van deze scholen constateert de analist wel risico's, maar deed de analist geen voorstel voor een KO risico. 2 scholen hebben het oordeel *Onvoldoende* of *Zeer zwak*. Er is geen enkele school buiten de risicogroep van de prestatie-monitor waar een analist een voorstel heeft gedaan voor een KO risico.

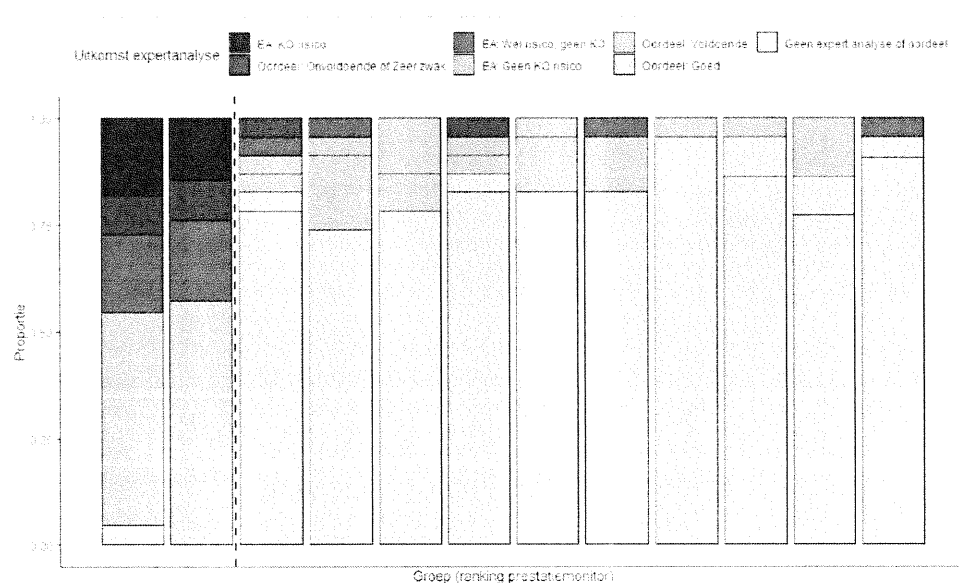
**Tabel 1: Aantal scholen per uitkomst van de expertanalyse en risico op de prestatie-monitor.**

		Risico op de prestatie-monitor	
		Risico (1 – 43)	Geen risico (44 – 269)
Uitkomst expertanalyse	EA: KO risico	7	0
	Oordeel: <i>Onvoldoende</i> of <i>Zeer zwak</i>	4	2
	EA: Wel risico, geen KO	8	4
	EA: Geen KO risico	23	9
	Oordeel: <i>Voldoende</i>	0	17
	Oordeel: <i>Goed</i>	1	7
	EA: Geen expertanalyse of oordeel	0	187

### 3.2

#### Verdeling risico-scholen op basis van rang in de prestatie-monitor

De uitkomsten van de expertanalyse kunnen verdeeld worden over de rangorde in de prestatie-monitor. Hiervoor zijn de scholen onderverdeeld in groepen op basis van hun rang in de prestatie-monitor. De 43 scholen die risico hebben volgens de prestatie-monitor zijn opgesplitst in twee groepen van resp. 22 en 21 scholen. De overige 226 scholen zijn verdeeld in 10 groepen van 23 of 22 scholen (Figuur 1). In Figuur 1 valt te zien dat alle scholen waar een KO risico voorstel is gedaan in de bovenste twee groepen van prestatie-monitor zitten. Voor de scholen met het oordeel *Onvoldoende* is dit niet het geval, 4 van de 6 scholen die het oordeel *Onvoldoende* hebben zitten in één van de twee risicogroepen van de prestatie-monitor. De overige 2 *Onvoldoende* scholen zitten in de groepen 3 en 6. Deze scholen scoren dus allemaal nog wel redelijk hoog op de prestatie-monitor. Daarnaast zijn er nog 4 scholen waarbij wel risico's zijn waargenomen, maar waarvoor niet tot een KO risico is besloten, die niet in de risicogroepen van de prestatie-monitor vallen. Hiervan is er zelfs één school die in de laatste risicogroep valt. Voor bijna de helft van de scholen in groepen 3 tot en met 12 is er afgelopen jaar geen expert analyse uitgevoerd en is er ook geen recent oordeel. Hierdoor weten we niet hoeveel van deze scholen voldoende of onvoldoende zijn.



**Figuur 1: De relatie tussen scores op de prestatie-monitor en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een uitkomst van de expertanalyse (EA) of Oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste twee groepen bevatten de 43 scholen die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 23/22 scholen.**

### 3.3

#### Foutanalyse

Foutanalyse is een methode om de kwaliteit van risicoschatting te kwantificeren. Dit is zinvol omdat daarmee op termijn de kwaliteit van risicoschattingen over de tijd en over verschillende sectoren/kwaliteitsgebieden vergeleken kan worden. Foutanalyse berust op een onderverdeling van uitkomsten in vier groepen (voortkomend uit Tabel 1):

- **Echt positieven (terecht alarm):** Zowel de prestatie-monitor als de expertanalyse detecteren risico's voor een school
- **Fout negatieven (gemist geval):** De prestatie-monitor detecteert geen risico's, maar de expertanalyse constateert dat er wél risico's zijn op een school.
- **Fout positieven (loos alarm):** De prestatie-monitor detecteert risico, maar de expertanalyse constateert dat er géén risico's zijn op een school.
- **Echt negatieven (terecht genegeerd):** Zowel de prestatie-monitor als de expertanalyse detecteren geen risico's voor een school.

Zo krijgen we Tabel 2.

**Tabel 2: Foutanalyse op de uitkomst van de prestatie-monitor en de expertanalyse**

		Risico op de prestatie-monitor	
		Risico	Geen risico
Risico volgens de analist	Risico	19 (Echt positieven)	6 (Fout negatieven)
	Geen risico	24 (Fout positieven)	220 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Zowel op sensitiviteit als op specificiteit scoort de prestatie-monitor heel aardig. 76% van de risicovolle scholen is boven de gestelde risicogrens gekomen. Dit betekent dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die mogelijk risico's hebben (en daarmee ook aan het onderscheiden van scholen waar waarschijnlijk niks aan de hand is). Zeker als we dit vergelijken met het alternatief wat gewoon het trekken van willekeurige steekproeven is. Ook is er geen enkel KO risico voorstel gedaan voor een school die onder de risicogrens zat. Alle risicovolle scholen die de prestatie-monitor "miste" betrof scholen waar wel risico's zijn geconstateerd, maar geen KO risico voorstel is gedaan of scholen die het oordeel *Onvoldoende* hebben.

Op precisie scoort de prestatie-monitor minder (44% van de onderzochte scholen is vervolgens gewaardeerd als risicovol). Dit betekent dat er relatief veel scholen moeten worden bekeken in de expertanalyse omdat ze hoog scoren op de prestatie-monitor, maar waar onvoldoende aanleiding is voor een KO risico. Een verklaring hiervoor is dat het optimaliseren van de sensitiviteit en specificiteit vaak ten koste gaat van de precisie. Een lage precisie betekent dat er meer capaciteit nodig is van de analisten voor de expertanalyse. Het is dus van belang een balans te vinden tussen een hoge sensitiviteit en specificiteit nastreven en te zorgen dat de precisie niet te laag wordt, omdat dit zal leiden tot overvragen van de analisten. Een andere mogelijke verklaring hiervoor is dat er geen KO risico wordt gepland omdat er vertrouwen in het bestuur is en/of omdat de risico's bekend zijn, maar dat dit vanuit de huidige manier van documenteren niet duidelijk wordt en gebruikt kan worden in de evaluatie van de prestatie-monitor. Dit zou verbeterd kunnen worden door de manier bij het documenteren van de expertanalyse een onderscheid te maken tussen of de analist risico's constateert op een school en het doen van een voorstel tot een toezichtsactiviteit.

In Bijlage I hebben we op dezelfde manier een foutanalyse per kwaliteitsgebied gedaan en in Bijlage II is te zien welke indicatoren er per kwaliteitsgebied ontwikkeld zijn.

### 3.4

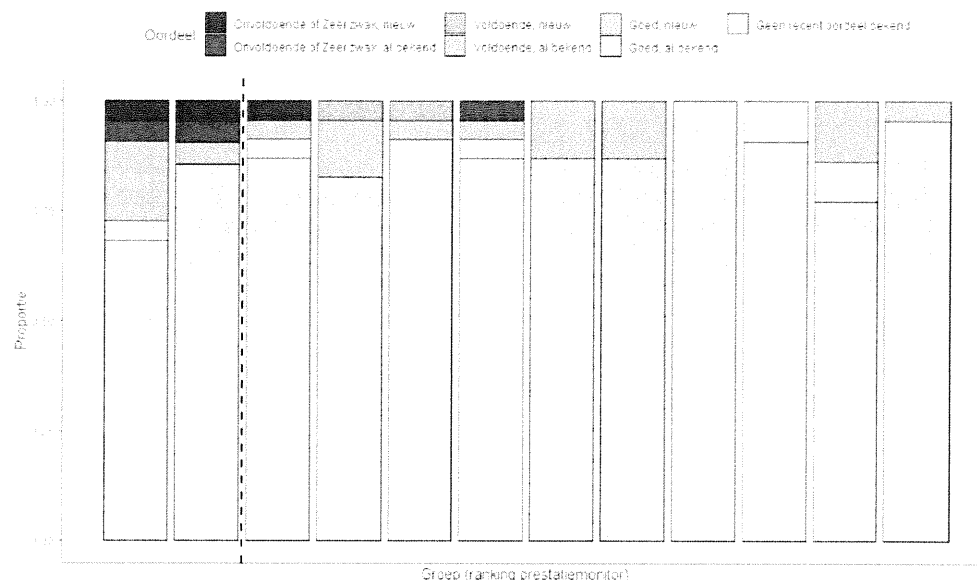
#### **Foutanalyse alleen aan de hand van oordelen**

Naast de kwaliteit van de risicoschatting bepalen aan de hand van de uitkomst van expertanalyses en recente oordelen, kunnen we ook kijken naar de uitkomst van recente oordelen. Deze manier van analyseren heeft als nadeel dat er minder data beschikbaar is (er worden namelijk meer expertanalyses uitgevoerd dan oordelen gegeven). Het grote voordeel van het doen van een foutanalyse aan de hand van oordelen is dat de data betrouwbaarder is omdat de school ook daadwerkelijk bezocht is.

Voor de foutanalyse delen we oordelen in zeven categorieën:

1. **Oordeel: Onvoldoende of Zeer zwak, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
2. **Oordeel: Onvoldoende of Zeer zwak, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
3. **Oordeel: Voldoende, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
4. **Oordeel: Voldoende, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
5. **Oordeel: Goed, al nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
7. **Geen recent oordeel bekend**  
Voor deze scholen is geen recent oordeel bekend.  
Deze scholen tellen mee als geen risico in de foutanalyse.

Op dezelfde manier als Figuur 1 kunnen we de scholen over 12 groepen verdelen. Hier bestaan de eerste 2 staven uit de 43 scholen die risico hebben volgens de prestatiemonitor en de overige 10 groepen daarna bestaan steeds uit 23/22 scholen. Voor ongeveer twee derde van de top 43 van de prestatiemonitor is er een recent oordeel bekend. Bij de 4 scholen uit de risicogroepen is het oordeel *Onvoldoende* of *Zeer zwak* gegeven. Daarnaast zijn er 2 scholen met het oordeel *Onvoldoende* of *Zeer zwak* die buiten de risicogroep van de prestatiemonitor vallen.



**Figuur 2: De relatie tussen scores op de prestatie-monitor en recente oordelen, sinds 1 augustus 2017. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 8 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste groep bevatten elk 11 scholen die risico hebben op de prestatie-monitor, de 5 groepen daarna bestaan steeds uit 10 scholen.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 2, krijgen we Tabel 3.

**Tabel 3: Foutanalyse op de uitkomst van de prestatie-monitor en recente oordelen**

		Risico op de prestatie-monitor	
		Risico	Geen risico
Risico volgens recent oordeel	Risico	4 (Echt positieven)	2 (Fout negatieven)
	Geen risico	39 (Fout positieven)	224 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.

- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

De bovenstaande analyse aan de hand van de oordelen gegeven sinds augustus 2017 laat dit zien dat ook aan de hand van recente oordelen de sensitiviteit van de prestatie-monitor hoog is. De prestatie-monitor SBO detecteert 67% van de scholen die het oordeel Onvoldoende of Zeer zwak hebben, hiermee mist de prestatie-monitor 2 risicovolle scholen. Daarom valt te concluderen dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die in aanmerking komen voor een KO risico onderzoek en het oordeel Onvoldoende of Zeer zwak krijgen. Wel is het zo dat er erg weinig recente oordelen bekend zijn en dat het dus onzeker is hoeveel scholen dus daadwerkelijk onvoldoende zijn. Dit is met name een probleem voor de scholen die een lage risicoscore krijgen op de prestatie-monitor. Omdat er voor scholen met een lage risicoscore minder recente oordelen bekend zijn.



## 4 Conclusie

De hier uitgevoerde analyse schetst een overwegend positief beeld over de kwaliteit van risicoschatting door de Prestatiemonitor SBO: 76% van de scholen die als risicovol uit de expertanalyse komen -of die al een onvoldoende oordeel hebben- kwamen terecht in de hoogste risicogroepen. Daarnaast kwam 67% van de scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de risicogroep van de prestatie-monitor. Dit suggereert dat er toegevoegde waarde is voor het gebruik van de PM SBO bij het prioriteren van scholen in het toezicht.

Tegelijkertijd is er ook ruimte voor verbetering. Immers, niet alle risicovolle scholen kwamen in de twee hoogste risicogroepen terecht. Het zal dus een belangrijk doel zijn om de kwaliteit van risicoschatting verder te verbeteren. Een van de belangrijkste bijdragen van de methode van evaluatie zoals hier beschreven is dat de ontwikkeling van de voorspelkracht in de komende jaren ook gemonitord kan worden. Daarmee kan er gericht en meetbaar gewerkt worden aan verbetering van risicoschatting.

Daarnaast is een perfecte risicoschatting onmogelijk. Risico's veranderen over de tijd, en er zijn ook risico's waar we op dit moment simpelweg geen data over hebben (zoals bijvoorbeeld de kwaliteit/ervaring van bestuurders, hoewel hier wel aan gewerkt wordt vanuit het datascience programma). Het detecteren van risicoscholen zal daarom altijd moeten bestaan uit een *combinatie* van zowel datagedreven risicoschatting (zoals in de prestatie-monitor) als het gebruik van de impliciete kennis van domeinexperts zoals inspecteurs of analisten, en bijvoorbeeld signalen.

## 5 Aanbevelingen

### 5.1 Aanbevelingen voor toekomstige ontwikkelingen van de PM SBO

- 1) Zoals beschreven in dit document bestaat er onzekerheid in de schatting van kwaliteitsmaten van risicoschatting. Dit komt omdat we simpelweg niet goed weten in hoeverre er risico's bestaan op scholen die een lage risicoscore hebben en het volgens de prestatiemonitor dus juist goed doen (en die ook niet in beeld zijn bij de COI). Hoewel we niet alle scholen kunnen en willen onderzoeken, zou het wel mogelijk zijn om onderzoeken uit te voeren onder scholen met een lage risicoscore aan de hand van steekproeven. De prestatiemonitor zou ook gebruikt kunnen worden om op basis van de rangorde een steekproef te trekken en deze scholen te bekijken, als vorm van proportionaliteit in het toezicht. Dit kan helpen om de *prevalentie* van onvoldoende scholen inzichtelijk te maken. Dit zou sterk kunnen bijdragen aan de verdere verbetering van (datagedreven) risicogericht toezicht.
- 2) Als laatste kunnen de uitkomsten van de evaluatie ook gebruikt worden om verbeteringen in indicatorselectie en weging toe te passen op de prestatiemonitor. In Bijlage III (zie hieronder) hebben we hiertoe een aanvullende analyse uitgevoerd.

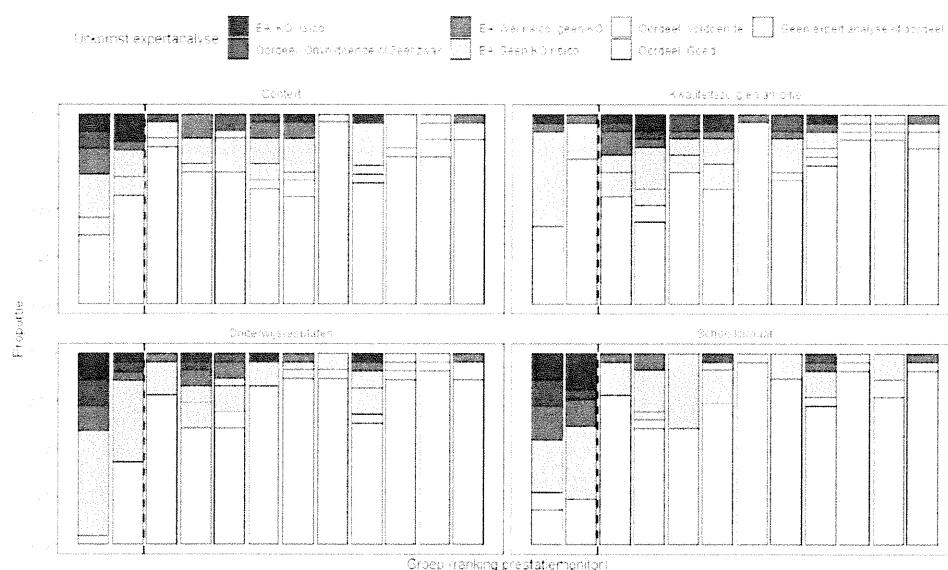
### 5.2 Aanbevelingen voor de registratie van de expertanalyse

- 1) Voor deze analyse is gebruik gemaakt van de registratie van de expertanalyse in een Exceltabel. Voor toekomstige evaluatie en verbetering van de prestatiemonitor zou het echter helpen om 'tot een nog duidelijkere scheiding te komen tussen het bestaan van risico's (ofwel, vindt de analist de risicoscore terecht) en de beslissing voor vervolgtoezicht (wordt er een KO-risico uitgevoerd). Met name het eerste is belangrijk voor een goede evaluatie van de prestatiemonitor.  
Hiertoe stellen wij voor om de kolom '*KO-risico? Voorstel analist*' op te splitsen in twee kolommen. Bijvoorbeeld één kolom '*Zijn er risico's op de school?*' en één kolom '*KO-risico voorstel*'.
- 2) Daarnaast waren er dit jaar problemen met dataregistratie over de expertanalyse. Dit resulteerde in fouten die handmatig omgezet moesten worden bij de evaluatie. De oorzaak hiervan ligt bij Proza dat niet geschikt is voor een dergelijke registratie. Voor komend jaar is het goed om te kijken of de procedure van het registreren van de expertanalyse effectiever kan.
- 3) Scholen die volgens de prestatiemonitor niet in de risicogroep vallen, maar waar wel een KO risico voorstel gedaan wordt door de analist zijn zeer interessant voor de evaluatie. Daarom is het handig als uit de documentatie duidelijk wordt waarom er een KO risico voorstel en hoe deze school in beeld is gekomen (bijvoorbeeld via de contactinspecteur of via onderwijsresultaten).

## Bijlage I: Foutanalyse per kwaliteitsgebied

De risicoscores in de prestatie-monitor zijn opgebouwd uit risicoscores op verschillende kwaliteitsgebieden. Bij de Prestatiemonitor SBO zijn dit *Onderwijsresultaten; Schoolklimaat; Context; Kwaliteitszorg en Ambitie; en Financieel Beheer*. In Bijlage II is een overzicht opgenomen van welke indicatoren onder welk kwaliteitsgebied vallen.

Een zelfde soort figuur als Figuur 1 kan gemaakt worden per kwaliteitsgebied. Hierbij is de uitkomst van de expertanalyse door de analist nog steeds genomen als uitkomstmaat. Echter zijn de groepen op de x-as nu verdeeld op basis van de rangorde op één kwaliteitsgebied. Dit is mogelijk voor vier van de vijf kwaliteitsgebieden, omdat de kwaliteitsgebieden *Financieel beheer* maar gevuld wordt door één indicator en er dus geen rangorde gemaakt kan worden op dit kwaliteitsgebied.



**Figuur 2 De relatie tussen scores op de prestatie-monitor per kwaliteitsgebied en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore per kwaliteitsgebied, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de expertanalyse/oordeel voorkomt per groep op basis van de rangorde. De eerste twee groepen zijn de 43 scholen die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 23/22 scholen.**

Uit figuur 2 kan de sensitiviteit, specificiteit en precisie per kwaliteitsgebied te berekenen. Zo krijgen we Tabel 3.

**Tabel 3: Waardes voor de foutanalyse per kwaliteitsgebied en de gehele prestatie-monitor.**

	Sensitiviteit	Specificiteit	Precisie
Context	0,44	0,87 - 0,95	0,26 - 0,48
Kwaliteitszorg en ambitie	0,12	0,84 - 0,93	0,07 - 0,17
Onderwijsresultaten	0,48	0,87 - 0,91	0,28 - 0,36

Schoolklimaat	0,72	0,90 – 0,93	0,42 - 0,53
<b>Gehele prestatiemonitor</b>	<b>0,76</b>	<b>0,90</b>	<b>0,44</b>

In Tabel 3 is te zien dat de verdeling van risicoscholen op het kwaliteitsgebied *Schoolklimaat* het meest overeenkomt met de uitkomsten van de expertanalyse/oordelen, en daarmee het meest lijkt op de relatie zoals waargenomen voor de totaal-risicoscore van de prestatiemonitor. Ook voor de kwaliteitsgebieden *Onderwijsresultaten* en *Context* komt de verdeling van risicoscholen enigszins overeen met de algehele rangorde in de prestatiemonitor. Voor het kwaliteitsgebied *Kwaliteitszorg en Ambitie* is dit niet het geval en dit kwaliteitsgebied scoort dan ook erg laag op sensitiviteit en precisie. Tot slot is te zien dat de prestatiemonitor als geheel een betere voorspelkracht heeft dan elk kwaliteitsgebied apart.

## Bijlage II: Risico-indicatoren

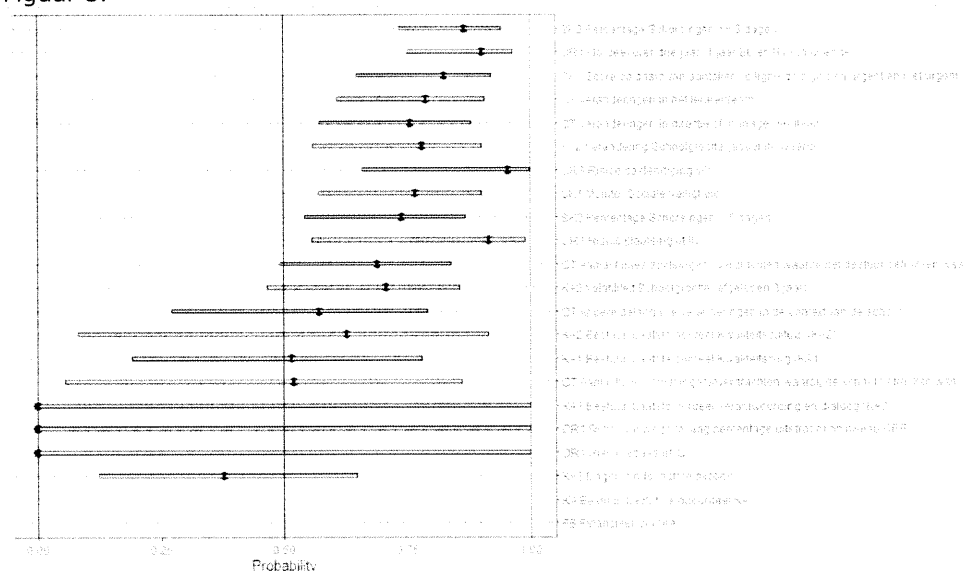
De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

Onderwijsresultaten	Schoolklimaat	Kwaliteitszorg en ambitie	Financieel beheer	Context
Oordeel over drie jaar (OR1) <i>Weging = 2</i>	Aantallen veiligheidssignalen (SK1) <i>Weging = 2</i>	Dagen sinds laatste bezoek (KA3) <i>Weging = 2</i>	Financieel Oordeel (FB) <i>Weging = 1</i>	School: betrokken bij fusies / splitsingen / overdrachten (CT) <i>Weging = 2</i>
Risico advies irt IQ (OR1) <i>Weging = 2</i>	Monitor Sociale Veiligheid niet ingeleverd (SK1) <i>Weging = 2</i>	Laatste eindoordeel KA (KA) <i>Weging = 1</i>		Bestuur: betrokken bij fusies / splitsingen / overdrachten (CT) <i>Weging = 1</i>
Risico plaatsing irt IQ (OR1) <i>Weging = 2</i>	Percentage Schorsingen < 5 dagen (SK2) <i>Weging = 2</i>	Laatste oordeel Kwaliteitszorg (KA1) <i>Weging = 1</i>		Veranderingen in de directie of managementteam (CT) <i>Weging = 0,33</i>
Risico uitstroom op OPP (OR1) <i>Weging = 2</i>	Percentage Schorsingen >= 5 dagen (SK2) <i>Weging = 2</i>	Laatste oordeel Kwaliteitscultuur (KA2) <i>Weging = 1</i>		Veranderingen in het lerarenteam (CT) <i>Weging = 0,33</i>
Risico bestendiging VO (OR3) <i>Weging = 2</i>		Verandering Schoolgrootte (KA2) <i>Weging = 1</i>		Belangrijke veranderingen school (CT) <i>Weging = 0,33</i>
		Volatiliteit Schoolgrootte (KA2) <i>Weging = 1</i>		
		Laatste oordeel Verantwoording en dialoog (KA3) <i>Weging = 1</i>		

## Bijlage III: Indicatoranalyse

De uitkomst van de expertanalyse kan gebruikt worden om de betrouwbaarheid en voorspellende waarde van verschillende indicatoren te bepalen. In onderstaande analyse is hiertoe per school bekeken in welke mate er statistische samenhang bestaat tussen de score op een individuele indicator en de uitkomsten van de expertanalyse/oordelen. Hiervoor is een GLM (Generalized Linear Model) gebruikt met de indicator als onafhankelijke variabele en de uitkomst van de expertanalyse als uitkomst (afhankelijke) variabele. Hiervoor zijn de volgende categorieën gecodeerd als 'risico': EA: *KO risico*, Oordeel: *Onvoldoende of Zeer zwak* en EA: *Wel risico*, geen KO.

De uitkomsten van deze analyse zijn kansen (probabilities) tussen 0 en de 1 die aangeven hoe sterk een indicator geassocieerd is met de uitkomst van de expertanalyse. Een probability van 1 geeft hierbij aan dat risico op de indicator altijd overeenkomt met risico volgens de expertanalyse. Een probability van 0 geeft hierbij aan dat risico op de indicator altijd overeenkomt met géén risico volgens de expertanalyse (dus de indicator voorspelt de andere kant op dan we verwachten). Een probability van 0.5 geeft aan dat risico op de indicator met beide uitkomsten van de expertanalyse even sterk is geassocieerd, dit is vergelijkbaar met het opgooien van een munt. De resultaten van deze analyse worden weergegeven in Figuur 3.



**Figuur 3: Probability waarmee indicatoren uit de prestatiemonitor de uitkomst van de expertanalyse voorspellen. De horizontale balken geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van meest naar minst betrouwbaar.**

SK2 Percentage Schorsingen >= 5 dagen en OR1 Oordeel over drie jaar (1 jaar BL en RW voldoende) zijn de beste voorspellers van risico als de uitkomst van de expertanalyse. Omdat beide indicatoren goed gevuld zijn en een hoge probability hebben. Deze indicatoren hebben beide ook een hoog gewicht in de prestatiemonitor, dit komt dus overeen met de verwachtingen.

De indicatoren *CT Veranderingen in directie of managementteam* en *CT Veranderingen in het lerarenteam* hebben beide een goede voorspellende waarde van de uitkomst van de expertanalyse, maar slechts een gewicht in prestatie-monitor van 0,33. Voor aankomend jaar is het overwegen waard om deze indicatoren een hoger gewicht in de prestatie-monitor mee te geven.

Vier indicatoren hebben een probability van onder de 0,50.

- *KA3 Bestuur: Laatste oordeel Verantwoording en dialoog (KA3)*
- *OR1 Risico vanwege te laag percentage uitstroom op niveau OPP*
- *OR1 Risico advies irt IQ*
- *KA3 Dagen sinds laatste bezoek*

Deze indicatoren voorspellen het tegenovergestelde dan we zouden verwachten, namelijk dat risico op één van deze indicatoren juist *geen risico* voorspelt als uitkomst van de expert analyse.

*KA3 Bestuur: Laatste oordeel Verantwoording en dialoog (KA3)*, *OR1 Risico vanwege te laag percentage uitstroom op niveau OPP* en *OR1 Risico advies irt IQ* hebben een hele lage probability van net boven de 0. Dit komt doordat er op deze indicatoren geen enkele school is die risico heeft op de indicator en waar de analist vervolgens een KO risico voorstel doet. Daarnaast zijn deze indicatoren slecht gevuld, met alle drie slechts tussen de 4 en 6 scholen die risico hebben op de indicator. Dit verklaart ook de grote betrouwbaarheidsintervallen. Voor *KA3 Dagen sinds laatste bezoek* zijn deze effecten minder groot, maar ook deze indicator voorspelt het tegenovergesteld van wat we zouden verwachten. Alle vier deze indicatoren werken dus eigenlijk een getrouwe risicoschatting door de prestatie-monitor tegen.

### **Conclusie**

De twee indicatoren, *OR1 Oordeel over drie jaar (1 jaar BL en RW voldoende)* en *SK2 Percentage Schorsingen  $\geq 5$  dagen*, zijn de beste voorspellers van de uitkomst van de expert analyse. Deze indicatoren hebben ook al het gewicht 2 gekregen wat nu het hoogst is in de prestatie-monitor. Daarnaast hebben de indicatoren *CT Veranderingen in directie of managementteam* en *CT Veranderingen in het lerarenteam* momenteel een laag gewicht in de prestatie-monitor, maar vertonen ze wel goede voorspelkracht. Voor al deze vier indicatoren is het overwegen waard het gewicht te verhogen.

De indicatoren *KA3 Bestuur: Laatste oordeel Verantwoording en dialoog (KA3)*, *OR1 Risico vanwege te laag percentage uitstroom op niveau OPP*, *OR1 Risico advies irt IQ* en *KA3 Dagen sinds laatste bezoek* scoren onder de 0,50 hiermee zijn het onbetrouwbare indicatoren en werken ze een goede risico-indicatie door de prestatie-monitor zelfs tegen. Het is aan de werkgroep om te overwegen of deze indicatoren een toegevoegde waarde hebben en in de prestatie-monitor moeten blijven.



## **EVALUATIE VOORSPELKRACHT PRESTATIEMONITOR VO 2020**

5.1.2e

februari 2021



## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 4</b>
<b>2</b>	<b>Registratie van de expertanalyse t.b.v. de evaluatie 5</b>
<b>3</b>	<b>Kwaliteit van de risicoschatting 6</b>
3.1	Uitvoering expertanalyse 6
3.2	Indeling uitkomsten expertanalyse 6
3.3	Verdeling risico-ovt's op basis van rang in de prestatiemonitor 7
3.4	Foutanalyse 8
3.5	Foutanalyse alleen aan de hand van oordelen 10
<b>4</b>	<b>Conclusie 13</b>
<b>5</b>	<b>Aanbevelingen 14</b>
5.1	Aanbevelingen voor toekomstige ontwikkelingen van de PM VO 14
5.2	Aanbevelingen voor de registratie van de expertanalyse 14
<b>6</b>	<b>Bijlage I: Foutanalyse per kwaliteitsgebeid 15</b>
<b>7</b>	<b>Bijlage II: Analyse top 600 17</b>
<b>8</b>	<b>Bijlage III: Risico-indicatoren 18</b>
<b>9</b>	<b>Bijlage IV: Indicatoranalyse 20</b>
9.1	Afstemming grenswaarden individuele indicatoren (Alleen voor Kennis) 21
9.2	Conclusie 22

## Samenvatting

In dit document wordt een beschrijving gegeven van de evaluatie van de voorspelkracht van de Prestatiemonitor VO aan de hand van de registratie van de expertanalyse door de sector. Deze evaluatie laat een overwegend positief beeld zien van de voorspelkracht van de Prestatiemonitor VO. 77% van de objecten van toezicht (ovt's; hier opleidingsoorten) die door analisten of inspecteurs als risicovol zijn aangeduid kwamen ook als risicovol uit de prestatie-monitor. Daarnaast kwam 80% van de ovt's die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de hoogste risicogroepen. Daarmee kan de Prestatiemonitor VO een belangrijke bijdrage leveren aan risicogericht toezicht binnen de sector.

Desalniettemin mist de prestatie-monitor daarmee nog steeds een deel van de risicovolle ovt's. In totaal mist de prestatie-monitor ten minste 69 risicovolle ovt's. Er bestaat dus ook duidelijk ruimte voor verbetering van risicoschatting door de prestatie-monitor. Bovendien zijn er ook enkele onzekerheden bij het interpreteren van de uitkomsten van de expertanalyse die de evaluatie bemoeilijken. Om de evaluatie te vergemakkelijken en risicoschatting te verbeteren doen wij aan het eind van dit document enkele aanbevelingen. Deze zullen in de toekomst de evaluatie van de prestatie-monitor verbeteren.

Ten slotte is het goed te benadrukken dat risicoschatting nooit alleen op basis van de prestatie-monitor moet plaatsvinden, maar altijd in combinatie met andere bronnen (zoals kwalitatieve kennis van inspecteurs en analisten over ovt's).

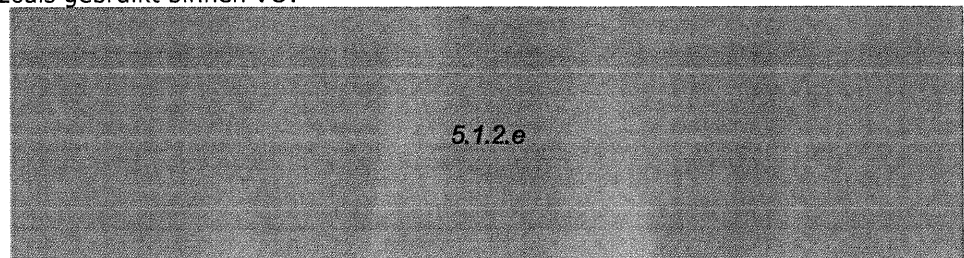
## **1 Inleiding**

De directie Kennis ontwikkelt jaarlijks de zogenaamde prestatie-monitors voor de verschillende sectoren. De prestatie-monitors beogen inspecteurs en analisten te ondersteunen bij het maken van risicoprioritering, gebaseerd op data. Sinds 2020 bestaat er ook een Prestatiemonitor Voortgezet Onderwijs (de PM VO).

Om de werking van de prestatie-monitors te borgen en te verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de Prestatiemonitor. Voor deze evaluatie worden de risicoscores van de prestatie-monitor afgezet tegen enerzijds de beoordeling van risico's op ovt's aan de hand van de expertanalyse (deskresearch) en anderzijds tegen de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een inspectie. In dit document wordt verslag gedaan van de evaluatie zoals uitgevoerd in december 2020.

## 2 Registratie van de expertanalyse t.b.v. de evaluatie

Om de risicoscores van de prestatie-monitor te vergelijken met de uitkomsten van de expertanalyse is het nodig om de uitkomsten van de expertanalyse op gestructureerde wijze vast te leggen. Dit jaar is er voor het eerst met een dergelijke registratie begonnen. Dit is gebeurd aan de hand van een Excel-formulier dat binnen de sector is bijgehouden. Hieronder een indruk van de structuur van dat formulier zoals gebruikt binnen VO:



Dit formulier vormde de basis voor de evaluatie die hieronder wordt beschreven, en is daarmee erg waardevol. Dat uiteindelijk de registratie nagenoeg compleet is gemaakt is een compliment waard. Wij erkennen als Kennis hoeveel werk dat zal hebben gekost. Zonder deze investering was deze evaluatie niet mogelijk geweest. Wat vooral veel toegevoegde waarde heeft is dat er ook veel ovt's zijn bekeken die geen risico hadden volgens de prestatie-monitor en dat er gedocumenteerd is hoe deze ovt's wel in beeld zijn gekomen. Daarnaast is het van grote toegevoegde waarde dat niet alleen de KO-risico voorstellen zijn opgenomen als mogelijke toezichtsactiviteit, maar ook bestuursgesprek/bestuurscontact, het geven van een opdracht, het meenemen in een 4JOB of een herstelonderzoek.

Desalniettemin is er nog winst te behalen bij het documenteren van de expertanalyse. Waar we regelmatig tegen aanliepen bij de evaluatie is dat het niet altijd duidelijk is of een ovt met een hoge risicoscore in de prestatie-monitor volgens de analist eigenlijk *niet* risicovol is (een fout van de Prestatie-monitor), of dat een ovt mogelijk wel risicovol is maar dat er andere redenen waren om bij de ovt geen KO risico uit te voeren, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat het ovt net bezocht is; of omdat er een herstelonderzoek loopt. Vaak stonden er echter wel beschrijvingen in het document die (deels handmatig) zijn gebruikt om toch duiding te geven aan de uitkomsten van de expertanalyse. Meer daarover in de volgende secties. Voor de toekomst raden wij aan om te zorgen voor een duidelijke scheiding in de registratie tussen de kwaliteit van de risicoschatting en de registratie van de genomen vervolgstappen m.b.t. toezicht. Daarnaast lijkt het erop dat het doel en gebruik van het document niet voor elke analist duidelijk was. Soms valt een al gepland herstelonderzoek onder een voorstel tot een toezichtsactiviteit een andere keer niet. Daarom raden wij aan dat er een duidelijkere structuur en gebruik van het document wordt afgesproken.

### 3 Kwaliteit van de risicoschatting

Om de kwaliteit van de risicoschatting van de Prestatiemonitor VO inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken hoe de uitkomst van de expertanalyse door de analisten en het huidige oordeel zich verhouden tot de rang die ovt's hebben in de prestatie-monitor, en of het ovt volgens de prestatie-monitor in de risicogroep valt.

#### 3.1 Uitvoering expertanalyse

De grens voor risico in de prestatie-monitor is in 2020 gelegd bij de 300 hoogst scorende ovt's van de in totaal 3397 ovt's. Naast de top van de prestatie-monitor zijn er ook andere redenen voor het doen van een expertanalyse. Een ovt wordt bekeken als:

- Het in de top 300 van de prestatie-monitor staat.
- De onderwijsresultaten onvoldoende of onvoldoende tenzij zijn.
- Er een te groot verschil is tussen het SE-CE.
- Het wordt aangedragen door de contactinspecteur.
- Eén van de andere ovt's van hetzelfde bestuur risico heeft op minimaal één van de bovenstaande indicatoren.

In totaal zijn er 1858 expertanalyses uitgevoerd.

#### 3.2 Indeling uitkomsten expertanalyse

Voor de foutanalyse hebben we de uitkomst van deze expertanalyses (EA) teruggecodeerd naar 3 mogelijke uitkomsten: *EA: wel toezichtsactiviteit*; *EA: Wel risico, geen toezichtsactiviteit*; *EA: Geen risico (en dus geen toezichtsactiviteit)*. In sommige gevallen is er echter besloten om bijvoorbeeld geen KO-risico uit te voeren omdat een ovt 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende ovt die nog even de tijd krijgt om herstelplan uit te voeren). Om deze ovt's ook in beeld te brengen zijn voor die ovt's waar geen voorstel tot een toezichtsactiviteit is ook de huidige oordelen meegenomen: *Zeer zwak*; *Onvoldoende*; *Voldoende*; en *Goed*. Alleen de oordelen die gegeven zijn sinds 1 augustus 2017 zijn meegenomen.

Daarmee komen we tot de volgende categorieën:

1. **EA: wel toezichtsactiviteit**

Als de analist een voorstel tot een toezichtsactiviteit doet.

2. **Oordeel Onvoldoende of Zeer zwak**

Er is geen expertanalyse uitgevoerd, maar deze ovt heeft het oordeel *Onvoldoende* of *Zeer zwak*. Deze ovt's tellen mee als risico in de foutanalyse.

3. **EA: Wel risico, geen toezichtsactiviteit**

Als de analist geen voorstel doet voor een toezichtsactiviteit, maar wel risico's op de ovt constateert. Dit is bijvoorbeeld als vertrouwen is in het bestuur, er een herstelonderzoek nog gaande is of net is afgerond, of een verificatie plaats gaat vinden. Deze ovt's tellen in de foutanalyse mee als risico.

4. **EA: Geen toezichtsactiviteit**

De analist heeft de expertanalyse uitgevoerd, maar geen risico's geconstateerd die voldoende zijn voor een toezichtsactiviteit.

5. **Oordeel: Voldoende**

Er is geen expertanalyse uitgevoerd, en deze ovt heeft het oordeel *Voldoende*. Deze ovt's tellen mee als geen risico in de foutanalyse.

#### 6. Oordeel: Goed

Er is geen expertanalyse uitgevoerd, maar deze ovt heeft het oordeel *Goed*. Deze ovt's tellen mee als geen risico in de foutanalyse.

Zoals te zien in Tabel 1 heeft een analist voor 183 ovt's een voorstel gedaan voor een toezichtsactiviteit. 148 hiervan waren ook gedetecteerd door de prestatie-monitor (top 300). Op 26 ovt's uit de top van de prestatie-monitor detecteerde de analist wel risico's maar deed deze geen voorstel voor een toezichtsactiviteit, bijvoorbeeld omdat er vertrouwen is in het bestuur of omdat het ovt net bezocht is. Ook hebben er 52 ovt's uit de top 300 een oordeel *Onvoldoende* of *Zeër zwak*, 14 ovt's buiten de top 300 hebben dit oordeel. Alle 125 ovt's met het oordeel *Goed* vallen buiten de top 300 van de prestatie-monitor. In totaal zijn er 69 ovt's waar de prestatie-monitor geen risico detecteerde, maar de analist wel risico's zag of waar het ovt een oordeel *Onvoldoende* of *Zeër zwak* heeft. Bij 35 van deze 72 stelt de analist ook een toezichtsactiviteit voor (Tabel 1). Bij de overige 34 zijn het risico's die al bekend zijn, is er vertrouwen in het bestuur of het betreft een oordeel *Onvoldoende* of *Zeër zwak*.

**Tabel 1: Aantal ovt's per uitkomst van de expertanalyse en risico op de prestatie-monitor.**

			Risico op de prestatie-monitor	
			Risico (1 – 300)	Geen risico (301 – 3397)
Uitkomst expertanalyse	Risico	EA: wel toezichtsactiviteit	148	35
		Oordeel: <i>Onvoldoende</i> of <i>Zeër zwak</i>	52	14
		EA: Wel risico, geen toezichtsactiviteit	26	20
	Geen risico	EA: Geen toezichtsactiviteit	55	749
		Oordeel: <i>Voldoende</i>	6	139
		Oordeel: <i>Goed</i>	0	125
		EA: Geen expert analyse of oordeel	13	2015

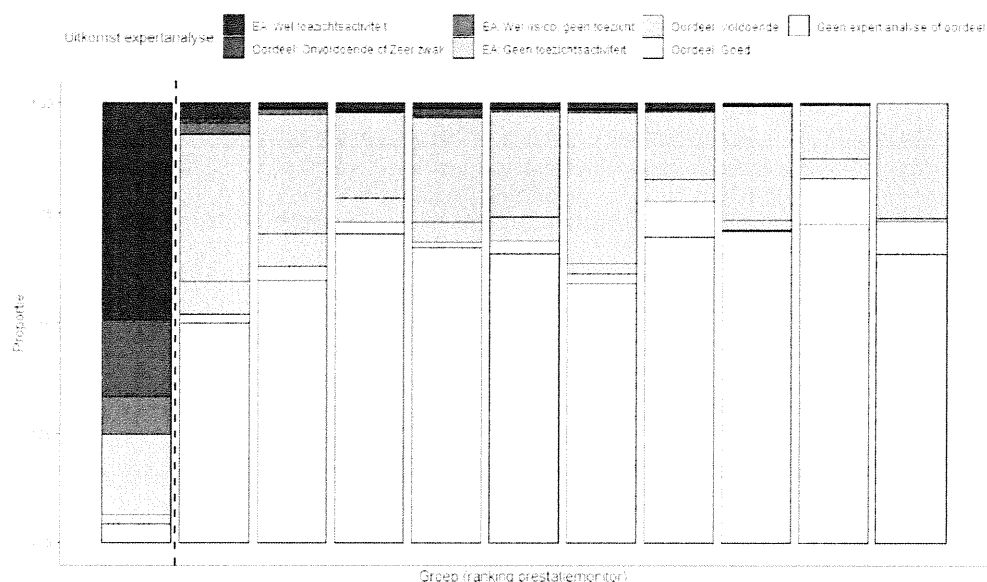
### 3.3

#### Verdeling risico-ovt's op basis van rang in de prestatie-monitor

De uitkomsten van de expertanalyse kunnen verdeeld worden over de rangorde in de prestatie-monitor. Hiervoor zijn de ovt's onderverdeeld in groepen op basis van hun rang in de prestatie-monitor. De eerste groep bevat de 300 ovt's die risico hebben volgens de prestatie-monitor. De overige 3097 ovt's zijn verdeeld in 10 groepen van 310 of 309 ovt's (Figuur 1).

In Figuur 1 valt te zien dat het grootste deel van de ovt's die risico hebben volgens de expertanalyse of het oordeel *Onvoldoende* hebben in de eerste groep zitten. Dit laat zien dat er aardig wat voorspelkracht van de prestatie-monitor uitgaat. Echter niet alle risico-ovt's worden in de hoogste groep geplaatst. In de tweede groep, bijvoorbeeld, constateert de expertanalyse risico op 11 ovt's. Dit is iets hoger dan in andere groepen buiten de top van de prestatie-monitor (groep 3 en hoger), waar het aantal ovt's met risico volgens de expertanalyse varieert tussen de 0 en de 5. Ook zit er nog een relatief groot deel met ovt's die het oordeel *Onvoldoende* of *Zeër zwak* hebben, net onder de grens van risico volgens de prestatie-monitor: In groep 2 zijn dit 7 ovt's en in groep 5 zijn dit er 6. In de overige groepen zijn dit er allemaal 0, 1 of 2. Voor meer dan de helft van de groepen 2 tot en met 11 is er afgelopen jaar geen expertanalyse uitgevoerd en is er ook geen

recent oordeel bekend (sinds augustus 2017). Hierdoor weten we niet hoeveel van deze ovt's voldoende of onvoldoende zijn.



**Figuur 1: De relatie tussen scores op de prestatie monitor en waargenomen risico's. Ovt's zijn gesorteerd op basis van de PM-risicoscore, en samengevoegd in 11 groepen. Iedere staaf geeft de proportie weer van hoe vaak een uitkomst van de expertanalyse (EA) of Oordeel voorkomt per groep op basis van de rangorde in de prestatie monitor. De eerste groep behelst de 300 ovt's die risico hebben op de prestatie monitor, de 10 groepen daarna bestaan steeds uit 310/309 ovt's.**

### 3.4

#### Foutanalyse

Foutanalyse is een methode om de kwaliteit van risicoschatting te kwantificeren. Dit is zinvol omdat daarmee op termijn de kwaliteit van risicoschattingen over de tijd en over verschillende sectoren/kwaliteitsgebieden vergeleken kan worden. Foutanalyse berust op een onderverdeling van uitkomsten in vier groepen (voortkomend uit Tabel 1):

- **Echt positieven (terecht alarm):** Zowel de prestatie monitor als de expertanalyse detecteren risico's voor een ovt.
- **Fout negatieven (gemist geval):** De prestatie monitor detecteert geen risico's, maar de expertanalyse constateert dat er wél risico's zijn op een ovt.
- **Fout positieven (loos alarm):** De prestatie monitor detecteert risico, maar de expertanalyse constateert dat er géén risico's zijn op een ovt.
- **Echt negatieven (terecht genegeerd):** Zowel de prestatie monitor als de expertanalyse detecteren geen risico's voor een ovt.

Zo krijgen we Tabel 2.

**Tabel 2: Foutanalyse op de uitkomst van de prestatie-monitor en de expertanalyse**

		Risico op de prestatie-monitor	
		Risico	Geen risico
Risico volgens de analist	Risico	226 (Echt positieven)	69 (Fout negatieven)
	Geen risico	74 (Fout positieven)	3028 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle ovt's dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct ovt's waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal ovt's met daadwerkelijk risico, binnen het aantal geselecteerde ovt's door de test.

Zowel op sensitiviteit als op specificiteit scoort de prestatie-monitor heel behoorlijk. Zo is 75% van de risicovolle ovt's in de top 300 van de prestatie-monitor gekomen. Dit betekent dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van ovt's die mogelijk risico's hebben (en daarmee ook aan het onderscheiden van ovt's waar waarschijnlijk niks aan de hand is). Zeker als we dit vergelijken met het alternatief: het trekken van willekeurige steekproeven van 300 ovt's. Bij het trekken van een dergelijke steekproef zou gemiddeld slechts 8,8% van de ovt's in de selectie daadwerkelijk risico's hebben. Er vanuit gaande dat we het grootste deel van de risico's in beeld zouden hebben. Overigens is het zeer waarschijnlijk dat er onvoldoende ovt's in het deel van de populatie bestaan die niet zijn onderzocht. In totaal mist de prestatie-monitor nu 69 risico-ovt's op een totaal van 3397. Op 35 van deze risico-ovt's is een voorstel tot een toezichtsactiviteit gedaan. Van deze risico-ovt's zijn er 4 in beeld gekomen via het onderwijsresultatenmodel. 4 andere ovt's zaten al in een traject, maar is er alsnog gekozen voor een aanvullende toezichtsactiviteit. De overige 27 ovt's zijn alleen in beeld gekomen doordat er één of meerdere andere ovt's van hetzelfde bestuur risico hadden en toen alle ovt's van dit bestuur bekeken zijn.

Ook qua precisie scoort de Prestatie-monitor VO heel behoorlijk: 75% van de bekeken risico-ovt's is vervolgens ook door de analist beoordeeld als risicovol. De score op precisie hangt vanzelfsprekend samen met de score op sensitiviteit en specificiteit. We zouden de sensitiviteit kunnen verhogen door de risicogroep te verhogen (dus bijvoorbeeld de bovenste 600 i.p.v. 300 aan een expertanalyse te



onderwerpen). Maar dit zou betekenen dat hiervoor meer ovt's bekeken moeten worden en dus leiden tot een lagere precisie. Het is dus van belang een balans te vinden tussen een hoge sensitiviteit en specificiteit en een niet te lage precisie. Het lijkt erop dat deze balans bij de Prestatiemonitor VO momenteel heel aardig aanwezig is.

In Bijlage I hebben we op dezelfde manier een foutanalyse per kwaliteitsgebied gedaan. In Bijlage II is een analyse gedaan op alleen de top 600 van de prestatie-monitor om verder te analyseren hoe de risicoscore in de prestatie-monitor de expertanalyse beïnvloed.

### 3.5 Foutanalyse alleen aan de hand van oordelen

Naast de kwaliteit van de risicoschatting bepalen aan de hand van de uitkomst van expertanalyses en recente oordelen, kunnen we ook alleen kijken naar de uitkomst van recente oordelen. Deze manier van analyseren heeft als nadeel dat er minder data over beschikbaar is (er worden meer expertanalyses uitgevoerd dan oordelen gegeven). Het grote voordeel van het doen van een foutanalyse aan de hand van oordelen is dat de data betrouwbaarder is omdat, het ovt ook daadwerkelijk bezocht is.

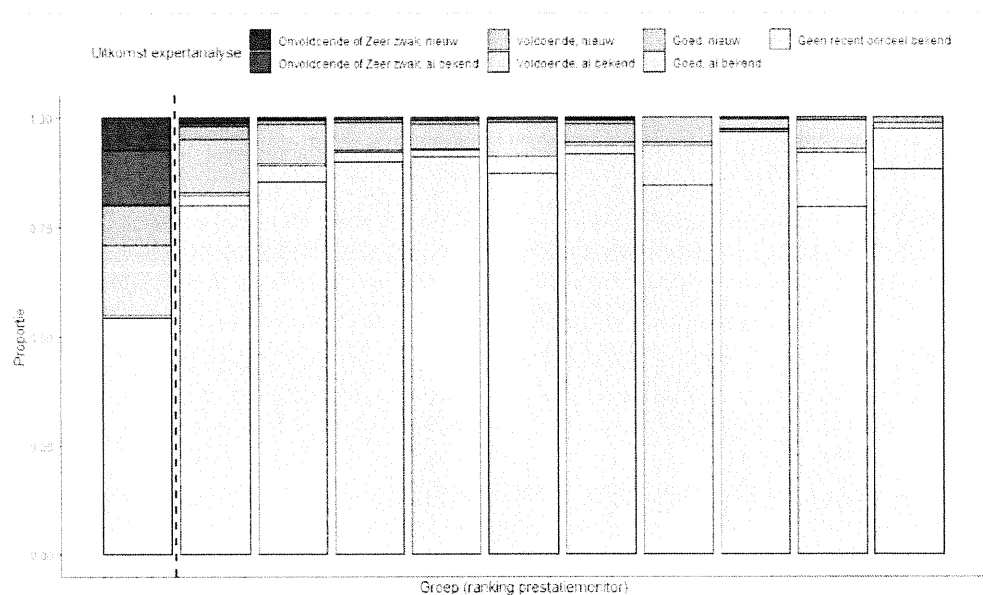
Voor de foutanalyse delen we oordelen in zeven categorieën:

1. **Oordeel: Onvoldoende of Zeer zwak, nieuw**  
Deze ovt's hebben tussen 1 juni 2020 en 19 februari 2021 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze ovt's tellen mee als risico in de foutanalyse.
2. **Oordeel: Onvoldoende of Zeer zwak, al bekend**  
Deze ovt's hebben tussen 1 augustus 2017 en 30 mei 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze ovt's tellen mee als risico in de foutanalyse.
3. **Oordeel: Voldoende, nieuw**  
Deze ovt's hebben tussen 1 juni 2020 en 19 februari 2021 het oordeel *Voldoende* gekregen.  
Deze ovt's tellen mee als geen risico in de foutanalyse.
4. **Oordeel: Voldoende, al bekend**  
Deze ovt's hebben tussen 1 augustus 2017 en 30 mei 2020 het oordeel *Voldoende* gekregen.  
Deze ovt's tellen mee als geen risico in de foutanalyse.
5. **Oordeel: Goed, al nieuw**  
Deze ovt's hebben tussen 1 juni 2020 en 19 februari 2021 het oordeel *Goed* gekregen.  
Deze ovt's tellen mee als geen risico in de foutanalyse.
6. **Oordeel: Goed, al bekend**  
Deze ovt's hebben tussen 1 augustus 2017 en 30 mei 2020 het oordeel *Goed* gekregen.  
Deze ovt's tellen mee als geen risico in de foutanalyse.

Op dezelfde manier als Figuur 1 kunnen we de ovt's over 11 groepen verdelen. Hier hebben de eerste 300 ovt's risico volgens de prestatie-monitor en de 10 groepen daarna bestaan steeds uit 310 of 309 ovt's.

Voor iets minder dan de helft van de top 300 van de prestatie-monitor is er een recent oordeel bekend. Bij de helft van deze recente oordelen voor de top van de prestatie-monitor is het oordeel *Onvoldoende* of *Zeer zwak* gegeven. Daarnaast zijn er ook ovt's met het oordeel *Onvoldoende* of *Zeer zwak* die niet in de top van de prestatie-monitor terecht komen. In totaal 'mist' de prestatie-monitor nu 15 ovt's die wel een oordeel *Onvoldoende* of *Zeer zwak* hebben. Van deze 15 ovt's hebben er 6 sinds 1 juni 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen en zijn dus gemist door de prestatie-monitor, maar wel via een andere weg in beeld gekomen.

Van deze 6 zijn ovt's er 5 in beeld gekomen voor een expertanalyse doordat een ander ovt van hetzelfde bestuur risico had, 1 ovt is helemaal niet in beeld gekomen en is dus ook geen expertanalyse op uitgevoerd.



**Figuur 2: De relatie tussen scores op de prestatie monitor en recente oordelen, sinds 1 augustus 2017. Ovt's zijn gesorteerd op basis van de PM-risicoscore, en samengevoegd in 11 groepen. Iedere staaf geeft de proportie weer van hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatie monitor. De eerste groep behelst de 300 ovt's die risico hebben op de prestatie monitor, de 10 groepen daarna bestaan steeds uit 310/309 ovt's.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 2, krijgen we Tabel 3.

**Tabel 3: Foutanalyse op de uitkomst van de prestatie monitor en recente oordelen**

		Risico op de prestatie monitor	
		Risico	Geen risico
Risico volgens recent oordeel	Risico	60 (Echt positieven)	15 (Fout negatieven)
	Geen risico	240 (Fout positieven)	3082 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle ovt's dat in beeld is).

- **Specificiteit** -of selectiviteit- is de mate waarin een test correct ovt's waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal ovt's met daadwerkelijk risico, binnen het aantal geselecteerde ovt's door de test.

## 4 Conclusie

Deze analyse schetst een overwegend positief beeld van de kwaliteit van risicoschatting door de Prestatiemonitor VO: 77% van de ovt's die als risicovol uit de expertanalyse komen -of die al een onvoldoende oordeel hebben- kwamen terecht in de hoogste risicogroepen. Daarnaast kwam 80% van de ovt's die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de hoogste risicogroepen. Dit suggereert dat er toegevoegde waarde is voor het gebruik van de PM VO bij het prioriteren van ovt's in het toezicht. Zeker omdat er afgelopen jaar ook via andere wegen risico-ovt's zijn gedetecteerd. Hieruit zijn 69 risico-ovt's naar voren gekomen, waarvan er op 35 ook daadwerkelijk een voorstel tot een toezichtsactiviteit is gedaan.

Dit betekent tegelijkertijd dat er ruimte is voor verbetering. Want niet alle risico-ovt's kwamen in de hoogste risicogroep van de prestatie-monitor terecht. Het zal dus een belangrijk doel zijn om de kwaliteit van risicoschatting verder te verbeteren. Een van de belangrijkste bijdragen van de methode van evaluatie zoals hier beschreven is dat de ontwikkeling van de voorspelkracht in de komende jaren gemonitord kan worden. Daarmee kan er gericht en meetbaar gewerkt worden aan verbetering van risicoschatting.

Daarnaast is een perfecte risicoschatting onmogelijk. Risico's veranderen over de tijd, en er zijn risico's waar we op dit moment simpelweg geen data over hebben (bijvoorbeeld de kwaliteit/ervaring van bestuurders, hoewel hier wel aan gewerkt wordt vanuit het datascience programma). Het detecteren van risico-ovt's zal daarom altijd moeten bestaan uit een *combinatie* van zowel datagedreven risicoschatting (zoals in de prestatie-monitor) als het gebruik van de impliciete kennis van domeinexperts zoals inspecteurs of analisten, en bijvoorbeeld signalen.

## 5 Aanbevelingen

### 5.1 Aanbevelingen voor toekomstige ontwikkelingen van de PM VO

- 1) Zoals beschreven in dit document bestaat er onzekerheid in de schatting van kwaliteitsmaten van risicoschatting. Dit komt omdat we simpelweg niet goed weten in hoeverre er risico's bestaan op ovt's die laag scoren op de prestatieindicator (en die ook niet in beeld zijn bij de COI of onder een bestuur vallen met andere risico-ovt's). Hoewel we niet alle ovt's kunnen en willen onderzoeken, zou het wel mogelijk zijn om onderzoeken uit te voeren onder ovt's met een lage risicoscore aan de hand van steekproeven. De prestatieindicator zou ook gebruikt kunnen worden om op basis van de rangorde een steekproef te trekken en deze ovt's te bekijken, als vorm van proportionaliteit in het toezicht. Dit kan helpen om de *prevalentie* van onvoldoende ovt's inzichtelijk te maken. Dit zou sterk kunnen bijdragen aan de verdere verbetering van (datagedreven) risicogericht toezicht, maar ook aan het inzicht dat de IvhO op stelselniveau heeft over het bestaan van onvoldoende ovt's.
- 2) De uitkomsten van de evaluatie kunnen ook gebruikt worden om verbeteringen in indicatorselectie en weging toe te passen op de prestatieindicator. In Bijlage IV (zie hieronder) hebben we hiertoe een aanvullende analyse uitgevoerd.

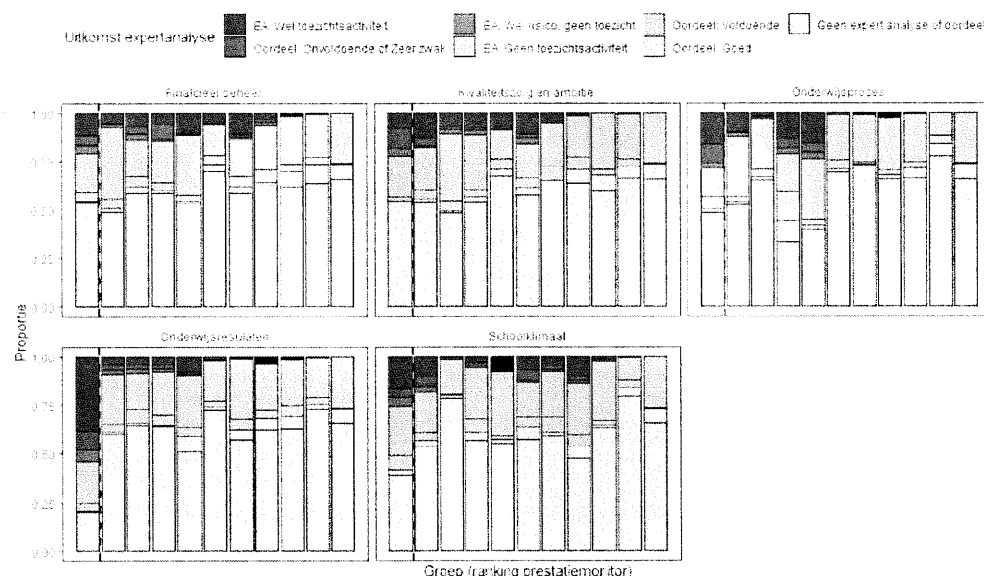
### 5.2 Aanbevelingen voor de registratie van de expertanalyse

- 1) Voor deze analyse is gebruik gemaakt van de registratie van de expertanalyse in een Exceltabel. Voor toekomstige evaluatie en verbetering van de prestatieindicator zou het echter helpen om tot een nog duidelijkere scheiding te komen tussen het bestaan van risico's (ofwel, vindt de analist de risicoscore terecht) en de beslissing voor vervolgtoezicht (wordt er een voorstel gedaan voor een toezichtsactiviteit). Voornamelijk het eerste is belangrijk voor een goede evaluatie van de prestatieindicator. Hiertoe stellen wij voor om de kolom '*Voorstel: uitvoeren toezichtactiviteit? ja/nee*' op te splitsen in twee kolommen. Bijvoorbeeld één kolom '*Zijn er risico's op het ovt?*' en één kolom '*Voorstel toezichtactiviteit? ja/nee*'.
- 2) Daarnaast waren er dit jaar problemen met dataregistratie over de expertanalyse. Dit resulteerde in fouten die handmatig omgezet moesten worden bij de evaluatie. Mogelijkerwijs valt er winst te behalen door duidelijk te communiceren wat het doel is van het bijhouden van de resultaten van de expertanalyse.

## 6 Bijlage I: Foutanalyse per kwaliteitsgebied

De risicoscores in de prestatie-monitor zijn opgebouwd uit risicoscores op verschillende kwaliteitsgebieden. Bij de Prestatiemonitor VO zijn dit *Onderwijsresultaten; Onderwijsproces; Schoolklimaat; Context; Kwaliteitszorg en Ambitie; en Financieel Beheer*. In Bijlage III is een overzicht opgenomen van welke indicatoren onder welk kwaliteitsgebied vallen.

Een zelfde soort figuur als Figuur 2, kan daarom ook gemaakt worden per kwaliteitsgebied. Hierbij is de uitkomst van de expertanalyse door de analist nog steeds genomen als uitkomst maat. Echter zijn de groepen op de x-as nu verdeeld op basis van de rangorde op één kwaliteitsgebied. Dit is mogelijk voor vijf van de zes kwaliteitsgebieden, omdat het kwaliteitsgebied *Context* slechts gevuld wordt door één indicator en er dus geen rangorde gemaakt kan worden op dit kwaliteitsgebied.



**Figuur 3: De relatie tussen scores op de prestatie-monitor per kwaliteitsgebied en waargenomen risico's. ovt's zijn gesorteerd op basis van de PM-risicoscore per kwaliteitsgebied, en samengevoegd in 11 groepen. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de expertanalyse/oordeel voorkomt per groep op basis van de rangorde. De eerste groep is de top 300 ovt's die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 310/309 ovt's.**

Op basis van Figuur 2 kunnen we de sensitiviteit, specificiteit en precisie per kwaliteitsgebied berekenen. Zo krijgen we Tabel 3.

**Tabel 4: Waardes voor de foutanalyse per kwaliteitsgebied en de gehele prestatie-monitor.**

	Sensitiviteit	Specificiteit	Precisie
Financieel beheer	0,21	0,92	0,21
Kwaliteitszorg en ambitie	0,23	0,92	0,22
Onderwijsproces	0,29	0,93	0,29
Onderwijsresultaten	0,55	0,96	0,54

Schoolklimaat	0,26	0,93	0,26
<b>Gehele prestatieindicator</b>	<b>0,77</b>	<b>0,98</b>	<b>0,75</b>

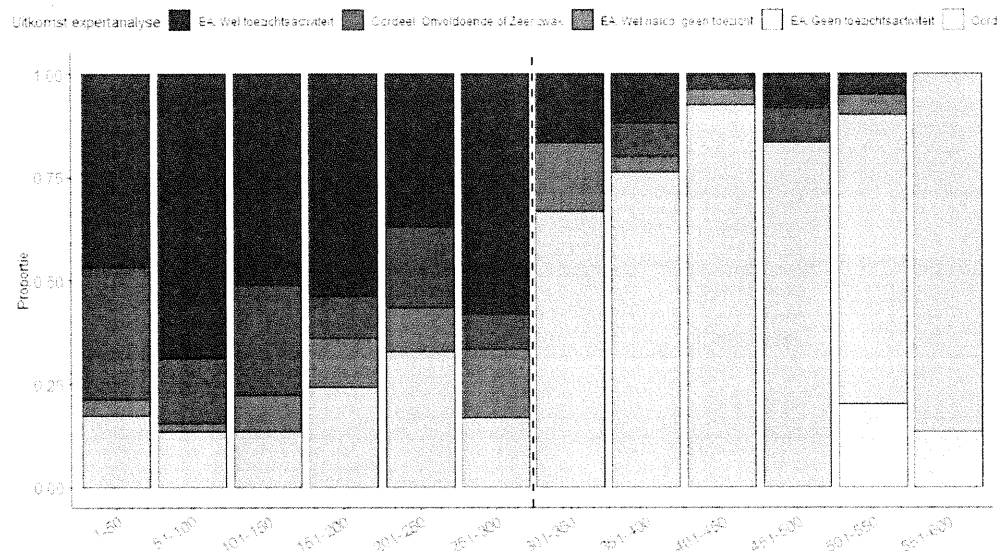
In Tabel 3 is te zien dat de verdeling van risico-overtredingen op het kwaliteitsgebied *Onderwijsresultaten* het meest overeenkomt met de uitkomsten van de expertanalyse/oordelen, en daarmee het meest lijkt op de relatie zoals waargenomen voor de totale risicoscore van de prestatieindicator. Dit valt ook te verwachten omdat de indicator *Berekend eindoordeel ORM onvoldoende* verreweg de grootste weging in de prestatieindicator heeft. Voor de overige kwaliteitsgebieden komt verdeling van risico-overtredingen minder goed overeen met de complete rangorde in de prestatieindicator. Wel is het voor al deze kwaliteitsgebieden zo dat de groepen met de laagste rang in de prestatieindicator niet gevuld zijn met risico-overtredingen. Dus alle kwaliteitsgebieden detecteren risico-overtredingen in elk geval tot op een zekere hoogte. Daarnaast is het zo dat de gehele prestatieindicator een betere voorspelkracht heeft dan elk kwaliteitsgebied los.

## 7 Bijlage II: Analyse top 600

Omdat de prestatie-monitor een risicomodel is, verwachten we dat ovt's die hoger scoren in de prestatie-monitor eerder een voorstel tot toezichtsactiviteit krijgen. Dit zou onafhankelijk moeten zijn van waar de risicogrens getrokken wordt. Als we de ovt's opdelen in groepen op basis van hun rang, verwachten we dus dat er in de eerste groep de meeste risico's worden geconstateerd, dat in de laatste groep de minste risico worden geconstateerd en dat de proportie aan ovt's met risico in de groepen hier tussenin geleidelijk afneemt. Er zou dus geen harde grens moeten zijn tussen groepen die nog onder de top van de prestatie-monitor vallen (top 300) en groepen die daar niet onder vallen.

In Figuur 3 zijn de 600 hoogst scorende ovt's uit de prestatie-monitor verdeeld in groepen van 50 ovt's om te zien of we in de Prestatie-monitor VO een geleidelijk afname in het aantal risico-ovt's zien. Omdat er in de top 300 meer expertanalyses zijn uitgevoerd dan bij lager scorende ovt's, zijn de ovt's waar geen expertanalyse is uitgevoerd of die het oordeel *Voldoende* hebben niet meegenomen in deze analyse.

Figuur 3 laat zien dat er wel een harde grens tussen de top 300 en de groepen daarna is i.p.v. een geleidelijke afname. Daarom is het goed om het gebruik van de prestatie-monitor in de expertanalyse te bekijken en te heroverwegen. Bijvoorbeeld of er niet te veel wordt afgegaan of een ovt boven de risicogrens zit i.p.v. te kijken naar de losse indicatoren en de onderliggende data.



**Figuur 4: De relatie tussen scores op de prestatie-monitor en waargenomen risico's voor de 600 hoogst scorende ovt's uit de prestatie-monitor. De ovt's zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van elk 50 ovt's. De ovt's waar geen expert analyse op is uitgevoerd, of die het oordeel voldoende hadden zijn vervolgens uit deze groepen gefilterd en niet meegenomen in deze analyse. Iedere staaf geeft de proportie weer van hoe vaak een uitkomst van de expertanalyse (EA) of oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste 6 groepen zijn de top 300 uit de prestatie-monitor, dit zijn dus de ovt's die risico hebben volgens de prestatie-monitor.**



## 8 Bijlage III: Risico-indicatoren

De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

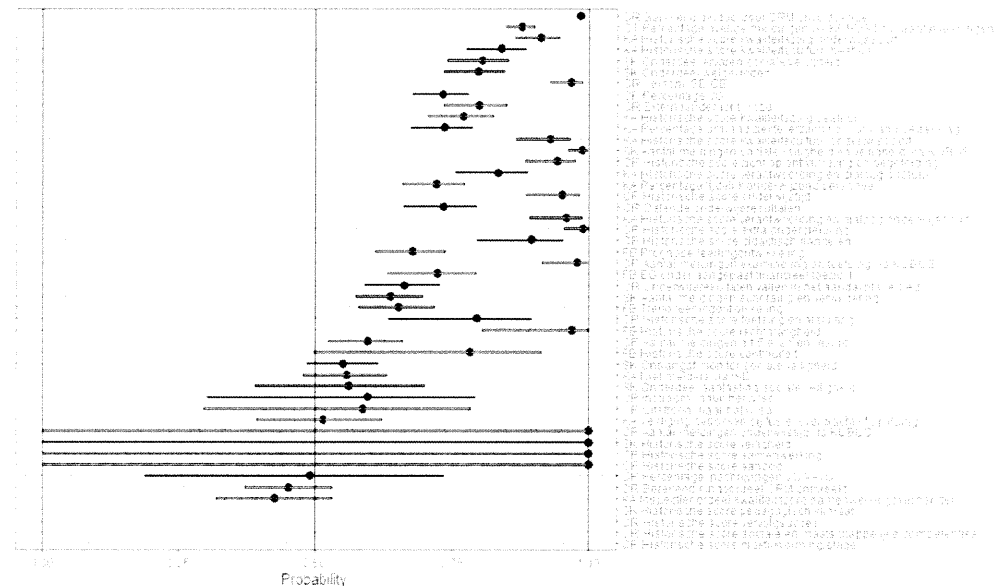
Onderwijsresultaten	Onderwijsproces	Schoolklimaat	Kwaliteitszorg en ambitie	Financieel beheer	Context
Berekend eindoordeel ORM onvoldoende (OR1) <i>Weging = 10</i>	Historische score aanbod (OP1) <i>Weging = 2</i>	Ontvangst monitor sociale veiligheid (SK) <i>Weging = 1</i>	Inspectieoordeel kwaliteitszorg SWV (KA1) <i>Weging = 1</i>	BG onder aangepast financieel toezicht (FB) <i>Weging = 1</i>	Percentage over melding t.o.v. aantal leerlingen (CT) <i>Weging = 5</i>
Onderwijsresultaten vallen in het aandachtsgebied (OR1) <i>Weging = 1</i>	Percentage inschrijvingen VO VAVO (OP2) <i>Weging = 1</i>	Onderdeel ervaren sociale veiligheid (SK1) <i>Weging = 1</i>	Inspectieoordeel kwaliteitszorg onderwijssoort (KA1) <i>Weging = 1</i>	Trend leerontwikkeling (FB1) <i>Weging = 1</i>	
ORM eindoordeel ontbreekt (OR1) <i>Weging = 1</i>	Percentage VSV (OP2) <i>Weging = 1</i>	Onderdeel aantasting sociale veiligheid (SK1) <i>Weging = 1</i>	Inspectieoordeel kwaliteitszorg bestuur (KA1) <i>Weging = 1</i>	Prognose leerontwikkeling (FB1) <i>Weging = 1</i>	
Dalende onderwijsresultaten (OR1) <i>Weging = 1</i>	Historische score zicht op ontwikkeling en begeleiding (OP3) <i>Weging = 1</i>	Aantal meldingen schorsing en verwijdering (SK1) <i>Weging = 1</i>	Vestiging betrokken bij fusie, overdracht of splitsing (KA2) <i>Weging = 1</i>	Historische score continuïteit (FB1) <i>Weging = 1</i>	
Verskil SE-CE (OR1) <i>Weging = 1</i>	Historische score didactisch handelen (OP3) <i>Weging = 1</i>	Aantal meldingen (sociale) veiligheid (SK1) <i>Weging = 5</i>	Percentage omvang ziekteverzuim (KA2) <i>Weging = 1</i>	Historische score rechtmatigheid (FB2) <i>Weging = 1</i>	
Historische score sociale en maatschappelijke competenties (OR2) <i>Weging = 1</i>	Instroom vanuit het (V)SO (OP4) <i>Weging = 1</i>	Historische score veiligheid (SK1) <i>Weging = 1</i>	Percentage tijdelijk onderwijzend personeel (KA2) <i>Weging = 1</i>		
Extern rendement VMBO (OR3) <i>Weging = 1</i>	Uitstroom naar het (V)SO (OP4) <i>Weging = 1</i>	Onderdeel welbevinden (SK2) <i>Weging = 1</i>	Historische score kwaliteitscultuur onderwijssoort (KA2) <i>Weging = 1</i>		

Historische score vervolgsucces (OR3) <i>Weging = 1</i>	Historische score extra ondersteuning (OP4) <i>Weging = 1</i>	Historische score pedagogische klimaat (SK2) <i>Weging = 1</i>	Historische score kwaliteitscultuur bestuur (KA2) <i>Weging = 1</i>
	Aantal meldingen onderwijstijd (OP5) <i>Weging = 5</i>		Niet melders via ISD (KA3) <i>Weging = 1</i>
	Historisch score onderwijstijd (OP5) <i>Weging = 1</i>		Historische score verantwoording en dialoog onderwijssoort (KA3) <i>Weging = 1</i>
	Historisch score samenwerking (OP6) <i>Weging = 1</i>		Historische score verantwoording en dialoog bestuur (KA3) <i>Weging = 1</i>
	Historisch score praktijkvorming/stage (OP7) <i>Weging = 1</i>		
	Aantal meldingen art 5 examenbesluit (OP8) <i>Weging = 1</i>		
	Aantal meldingen examinering en toetsing (OP8) <i>Weging = 5</i>		
	Historische score toetsing en afsluiting (OP8) <i>Weging = 1</i>		

## 9 Bijlage IV: Indicatoranalyse

De uitkomst van de expertanalyse kan gebruikt worden om de betrouwbaarheid en voorspellende waarde van verschillende indicatoren te bepalen. In de onderstaande analyse is hiertoe per ovt bekeken in welke mate er statistische samenhang bestaat tussen de score op een individuele indicator en de uitkomsten van de expertanalyse/oordelen. Hiervoor is een GLM (Generalized Linear Model) gebruikt met de indicator als onafhankelijke variabele en de uitkomst van de expertanalyse als uitkomst (afhankelijke) variabele. Hiervoor zijn de uitkomsten van de volgende categorieën gecodeerd als 'risico': EA: *wel toezichtsactiviteit*, *Oordeel: Onvoldoende of Zeer zwak en EA: Wel risico, geen toezichtsactiviteit*.

De uitkomsten van deze analyse zijn kansen (probabilities) tussen 0 en de 1 die aangeven hoe sterk een indicator is geassocieerd met de uitkomst van de expertanalyse. Een probability van 1 geeft hierbij aan dat risico op de indicator altijd overeenkomt met risico volgens de expertanalyse. Een probability van 0 geeft hierbij aan dat risico op de indicator altijd overeenkomt met géén risico volgens de expertanalyse (dus de indicator voorspelt de andere kant op dan we verwachten). Een probability van 0.5 geeft aan dat risico op de indicator met beide uitkomsten van de expertanalyse even sterk is geassocieerd, dit is vergelijkbaar met het opgooien van een munt. De resultaten van deze analyse worden weergegeven in Figuur 4.



**Figuur 5: Probability waarmee indicatoren uit de prestatiemonitor de uitkomst van de expertanalyse voorspellen. De horizontale balken geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van minst naar meest betrouwbaar.**

Zoals Figuur 4 laat zien is *OR Berekend eindoordeel ORM onvoldoende* de meeste betrouwbare voorspeller van risico als de uitkomst van de expertanalyse met een probability van 0,98. Dit komt overeen met de verwachting omdat deze indicator met een gewicht van 10, verreweg de zwaarstwegende indicator is. Verder zijn er 4 indicatoren die helemaal niet gevuld zijn op dit moment:

- *OP Historische score praktijkvorming/stage*

- *OR Historische score sociale en maatschappelijke competenties*
- *OR Historische score vervolgsucces*
- *SK Historische score pedagogisch klimaat*

Dit zijn allemaal indicatoren die gebaseerd zijn op historische oordelen vanuit de inspectie. Dus de vulling van de indicator valt niet direct te verbeteren. Daarnaast is er nu ook niet over te oordelen of het een betrouwbare indicator is. Ditzelfde geldt ook voor de andere indicatoren die gebaseerd zijn op historische inspectie oordelen, maar een lage betrouwbaarheid.

Er zijn drie indicatoren waarbij de probability lager is dan 0,50:

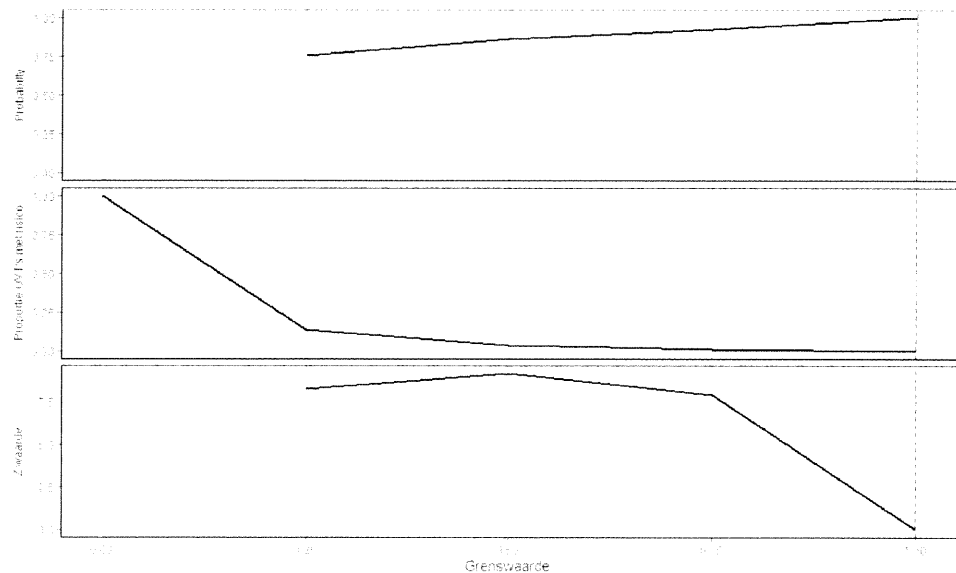
- *OR Berekend eindoordeel ORM ontbreekt*
- *KA Inspectieoordeel kwaliteitszorg samenwerkingsverbanden*
- *OP Percentage inschrijvingen VO VAVO*

Risico op deze indicator voorspelt dus juist *geen risico* als uitkomst van de expertanalyse. Dit is tegenovergesteld van wat we verwachten en werkt een betrouwbare risico-indicatie van de prestatie-monitor dus tegen. Daarom is het goed om deze indicatoren te heroverwegen en binnen de werkgroep te bespreken of deze indicator van toegevoegde kan waarde zijn.

## 9.1

### **Afstemming grenswaarden individuele indicatoren (Alleen voor Kennis)**

In een verdere analyse is gekeken of er winst te behalen valt door de grenswaarden van individuele indicatoren aan te passen. De indicator '*Indicator OP Aantal meldingen onderwijstijd via KUBUS*' kwam hiervoor in aanmerking. Deze indicator heeft op het moment een lage betrouwbaarheid omdat de indicator slecht gevuld is. Dit is echter aan te passen door de grenswaarde voor deze indicator anders te leggen. Momenteel is de grenswaarde zodanig gedefinieerd dat risicopunten worden toebedeeld als gemiddeld over drie jaar één of meer meldingen over onderwijstijd zijn geregistreerd. Figuur 5 vergelijkt voor verschillende grenswaarden de probability, proportie van ovt's met risico en Z-waarde voor de indicator *OP Aantal meldingen onderwijstijd via KUBUS*. De Z-waarde geeft hier de betrouwbaarheid van de indicator weer, een hogere Z-waarde betekent dat de indicator een meer betrouwbare voorspeller van de uitkomst van de expertanalyse is. Zoals in Figuur 5 te zien is geeft een grenswaarde van 0,75 een veel grotere Z-waarde, dan de huidige grenswaarde van 1,0. Daarom is het overwegen waard om een lagere grenswaarde te kiezen voor de indicator *OP Aantal meldingen onderwijstijd via KUBUS*.



**Figuur 6: De probability, proportie ovt's met risico en Z-waarde voor de indicator *OP Aantal meldingen onderwijstijd via KUBUS*. De grenswaarde is de minimale score die een ovt moet hebben om risico te krijgen op deze indicator, de gestippelde lijn geeft de huidige grenswaarde in de prestatiemonitor weer. De Z-waarde is een waarde die aangeeft hoe betrouwbaar een indicator is.**

## 9.2

### Conclusie

De indicator met het hoogste gewicht in de prestatiemonitor, *OR Berekend eindoordeel ORM onvoldoende*, is ook de meest betrouwbare voorspeller van de uitkomst van de expertanalyse. Enkele indicatoren die gebaseerd zijn op historische oordelen vanuit de inspectie scoren laag op betrouwbaarheid, maar zijn (waarschijnlijk) nog steeds van toegevoegde waarde voor de prestatiemonitor. De indicatoren *OR Berekend eindoordeel ORM ontbreekt*, *KA Inspectieoordeel kwaliteitszorg samenwerkingsverbanden* en *OP Percentage inschrijvingen VO VAVO* scoren onder de 0,50 en zijn hiermee onbetrouwbare indicatoren die een goede risico-indicatie door de prestatiemonitor tegenwerken. Het is aan de werkgroep om te overwegen of deze indicatoren een toegevoegde waarde hebben en in de prestatiemonitor moeten blijven. Voor de indicator *OP Aantal meldingen onderwijstijd via KUBUS* kan een lagere grenswaarde overwogen worden om zo een betrouwbaardere indicator te verkrijgen.



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

8.

## **EVALUATIE VOORSPELKRACHT PRESTATIEMONITOR (V)SO 2020**

5.1.2e

April 2021

## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 4</b>
<b>2</b>	<b>Uitvoering van risicoschatting en registratie van voorstel KO-risico t.b.v. de evaluatie 5</b>
<b>3</b>	<b>Kwaliteit van de risicoschatting 6</b>
3.1	Indeling uitkomsten 6
3.2	Kwaliteit van de risicoschatting prestatiemonitor 6
3.2.1	Verdeling risico-scholen op basis van rang in de prestatiemonitor 8
3.2.2	Foutanalyse 10
3.2.3	Foutanalyse alleen aan de hand van oordelen 12
<b>4</b>	<b>Conclusie 17</b>
<b>5</b>	<b>Aanbevelingen voor toekomstige ontwikkeling van PM VSO en SO 18</b>
<b>Bijlage I: VSO - Foutanalyse per kwaliteitsgebied 19</b>	
<b>Bijlage II: SO – Foutanalyse per kwaliteitsgebied 21</b>	
<b>Bijlage III: VSO – Risico-indicatoren 23</b>	
<b>Bijlage IV: SO – Risico-indicatoren 24</b>	
<b>Bijlage V: Foutanalyse op de 'oude' prestatieanalyse 25</b>	
VSO: Foutanalyse op de 'oude' prestatieanalyse 25	
SO: Foutanalyse op de 'oude' prestatieanalyse 26	
Vergelijking tussen de prestatiemonitor en 'oude' prestatieanalyse 26	
<b>Bijlage VI: Indicatoranalyse 28</b>	
VSO – Indicatoranalyse 28	
SO – Indicatoranalyse 29	
Lege indicatoren 30	

## Samenvatting

In dit document beschrijven we de evaluatie van de voorspelkracht van de Prestatiemonitor VSO en de Prestatiemonitor SO aan de hand van de risicoschatting van de inspecteur en recente oordelen. Deze evaluatie laat een redelijk positief beeld zien van de voorspelkracht van beide prestatie-monitors. De Prestatiemonitor VSO detecteert 67% van de risicovolle scholen en de Prestatiemonitor SO detecteert 61% van de risicovolle scholen. Daarmee kunnen de prestatie-monitors een belangrijke bijdrage leveren aan risicogericht toezicht binnen de sector. Desalniettemin missen beide prestatie-monitors nog steeds een deel van de risicovolle scholen. De Prestatiemonitor VSO mist in totaal 22 risicovolle scholen en de Prestatiemonitor SO mist 17 risicovolle scholen. Er bestaat dus duidelijk ruimte voor verbetering van risicoschatting door de prestatie-monitor.

De Prestatiemonitors VSO en SO zijn afgelopen jaar op een andere manier ingezet binnen de sector SO dan in de andere sectoren, dit maakt vergelijkingen van de evaluatie van de Prestatiemonitors VSO en SO met andere Prestatiemonitors moeilijk. Binnen de sector SO zijn er afgelopen jaar door inspecteurs inschattingen gemaakt over de mate van risico's op OKE's. Hierdoor zijn veel OKE's beoordeeld, wat gunstig is, maar tegelijkertijd is de risicoschatting minder sterk gebaseerd op data, en is er per ovt minder tijd besteedt aan de expertanalyse, dan bij andere sectoren. De sector SO geeft aan dat dit samenhangt met het feit dat over leerlingen in het SO maar beperktere hoeveelheid kwantitatieve gegevens beschikbaar is (bijvoorbeeld in veel mindere mate onderwijsresultaten). Om de mogelijkheden tot evaluatie van de voorspelkracht te voor komend jaar te verbeteren en vergelijking met andere Prestatiemonitors beter mogelijk te maken, doen wij aan het eind van dit document enkele aanbevelingen.



## **1 Inleiding**

De directie Kennis ontwikkelt jaarlijks de zogenaamde prestatie-monitors voor de verschillende sectoren. De prestatie-monitors beogen inspecteurs en analisten te ondersteunen bij het maken van risicoprioritering, gebaseerd op data. Sinds 2018 bestaat de Prestatiemonitor Voortgezet Speciaal Onderwijs (de PM VSO) en sinds 2019 bestaat ook de Prestatiemonitor Speciaal Onderwijs (de PM SO).

Om de werking van de prestatie-monitors te borgen en te verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de prestatie-monitor. Voor deze evaluatie worden de risicoscores van de prestatie-monitor afgezet tegen enerzijds de inspecteursinschatting van risico op de school; en anderzijds tegen de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een inspectie. In dit document wordt verslag gedaan van de evaluatie zoals uitgevoerd in maart en april 2021.

## 2 Uitvoering van risicoschatting en registratie van voorstel KO-risico t.b.v. de evaluatie

Het bepalen van risicovolle OKE's is afgelopen jaar binnen de sector SO in verschillende stappen uitgevoerd. Eerst zijn in juni quickscans uitgevoerd op de besturen die nog moesten worden bezocht met een 4JO om in te schatten of deze bezocht gingen worden met een compact of regulier 4JO. Hierin zijn de onderliggende scholen meegenomen op basis van hun kwaliteit, mate van ontwikkeling en het contact tussen de inspecteur en de school. Vervolgens heeft de contactinspecteur in september na het opleveren van de Prestatiemonitor inschattingen gedaan per OKE of deze risicovol is aan de hand van de eigen actuele kennis en de gegevens in de Prestatiemonitor. Deze inschattingen hebben vervolgens geleid tot KO-risico voorstellen per OKE. Tot slot zijn gedurende ABCD-bijeenkomst (gezamenlijke bijeenkomst tussen SO, rekenschap en Kennis) alle besturen bekeken en vanuit de verscheidene invalhoeken beredeneerd of hier (aanvullende) toezichtsactiviteiten dienden plaats te vinden. De KO-risico voorstellen en risicoschattingen per OKE door de inspecteur zijn in een Excelformulier bijgehouden. Hieronder een indruk van de structuur van dat formulier zoals gebruikt binnen (V)SO:

Algemene gegevens en verdelen		In te vullen september 2020					KO-risico nog te plannen	van een andere dachthier
1	2	3	4	5	6	7		
5.1.2.e								

Dit formulier vormde de basis voor de evaluatie die hieronder wordt beschreven. Dat de informatie over de inschattingen van risico en KO-risico zo compleet en overzichtelijk is vastgelegd verdient een groot compliment. Wij erkennen als Kennis hoeveel werk daarin is gegaan. Zonder deze investering was deze evaluatie niet mogelijk geweest.

In vergelijking met andere sectoren bestaan er verschillen in de manier waarop tot voorstel van toezichtsactiviteiten is gekomen. In tegenstelling tot andere sectoren is binnen VSO en SO de risicoschatting niet uitgevoerd door analisten, maar door de contactinspecteur. Hierbij wordt sterker geleund op de kennis over scholen en besturen van de contactinspecteur en minder op gestructureerde data over b.v. onderwijsresultaten. Niet in de laatste plaats omdat deze gestructureerde data voor de sector SO voor een deel ontbreekt en minder bestaat uit kwantitatieve objectieve data in vergelijking met andere sectoren. Dit komt onder andere doordat een zeer beperkt deel van leerlingen in SO en VSO deelneemt aan Eindtoetsen of Eindexamens. Wel moet opgemerkt worden dat vergelijkbare uitdagingen ook spelen in het Speciaal Basisonderwijs en het onderwijs aan Nieuwkomers. Hoewel ook hier vaak data over onderwijsresultaten ontbreekt werken deze sectoren ook met uitgebreide expertanalyses op risicovolle ovt's, waarin de gegevens uit de Prestatiemonitor een belangrijke rol spelen.

De contactinspecteurs binnen SO hadden daarnaast aanzienlijk minder tijd per OVT beschikbaar ten opzichte van de gemiddelde expertanalyse (risicoschatting) door een analist van een andere sector. Deze factoren beperken de vergelijking tussen deze evaluatie en de evaluaties uitgevoerd op andere Prestatiemonitors.

### 3 Kwaliteit van de risicoschatting

Om de kwaliteit van de risicoschatting van de Prestatiemonitor VSO en SO inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken hoe de uitkomst van de inschatting van risico's door de inspecteur en het huidige oordeel zich verhouden tot de rang die scholen hebben in de prestatie-monitor, en of de school volgens de prestatie-monitor in de risicogroep valt. Vooraf is vastgesteld dat ongeveer 25% van de ovt's binnen het risicogebied zouden vallen. Hiermee vallen voor VSO 85 van de 341 scholen in de risicocategorie en voor SO 79 van de 308 scholen binnen de risicocategorie.

#### 3.1 Indeling uitkomsten

Voor de foutanalyse hebben we de uitkomst van de risicoschattingen ingedeeld naar 3 mogelijke uitkomsten: *Inschatting inspecteur: KO Risico; Inschatting inspecteur: Wel risico, geen KO; Inschatting inspecteur: Geen risico (en dus geen KO)*. In sommige gevallen is er besloten om geen KO-risico uit te voeren omdat een school 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende school die nog even de tijd krijgt om herstelplan uit te voeren). Om deze scholen ook in beeld te brengen zijn voor de scholen die geen KO-risico krijgen maar die wel een huidig oordeel hebben ook de huidige oordelen meegenomen: *Zeër zwak; Onvoldoende; en Voldoende; en waardering Goed*. Oordelen worden meegenomen als ze gegeven zijn sinds augustus 2017.

Daarmee komen we tot de volgende categorieën:

1. **Inschatting inspecteur: KO-risico**  
Als de inspecteur een voorstel tot KO-risico doet.
2. **Oordeel Onvoldoende of Zeër zwak**  
Deze school heeft het oordeel *Onvoldoende* of *Zeër zwak*. Deze scholen tellen mee als risico in de foutanalyse.
3. **Inschatting inspecteur: Wel risico, geen KO**  
Als er geen voorstel voor een KO-risico wordt gedaan, maar wel risico's op de school worden geconstateerd. Dit is bijvoorbeeld wanneer er vertrouwen is in het bestuur of als er een herstelonderzoek nog gaande is of net is afgerond. Deze scholen tellen mee als risico in de foutanalyse.
4. **Inschatting inspecteur: Geen KO-risico**  
Er is een risicoschatting gemaakt, maar er zijn geen risico's geconstateerd die aanleiding zijn voor een KO-risico.
5. **Oordeel: Voldoende**  
Er is geen risicoschatting uitgevoerd en deze school heeft het oordeel *Voldoende*. Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Waardering: Goed**  
Deze school heeft het oordeel *Goed*. Deze scholen tellen mee als geen risico in de foutanalyse.

#### 3.2 Kwaliteit van de risicoschatting prestatie-monitor

Zoals te zien in Tabel 1 is er voor 13 VSO-scholen een voorstel gedaan voor een KO-risico, 8 hiervan vallen in de risicogroep van de prestatie-monitor (top 25%). Op 33 scholen uit de top van de prestatie-monitor worden wel risico's gedetecteerd maar geen voorstel gedaan voor een KO-risico, bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. 6 scholen hebben het oordeel *Onvoldoende* of *Zeër zwak*, 3 van deze scholen vallen in de risicogroep van de prestatie-monitor en 3 vallen daarbuiten. 13 scholen hebben de waardering *Goed*, al deze scholen vallen buiten de risicogroep van de

prestatie-monitor.

In totaal zijn er 22 scholen waar de prestatie-monitor geen risico detecteerde, maar waar de inspecteur inschatte dat er wel risico's zijn of waar de school een oordeel *Onvoldoende* of *Zeer zwak* heeft. Bij 5 van deze scholen is een KO-risicovoorstel gedaan door de inspecteur. Daarnaast zijn er 14 waar de inspecteur wel risico's constateerde, maar deze risico's niet genoeg aanleiding waren voor een KO-risico onderzoek. Tot slot vallen er 3 scholen met het oordeel *Onvoldoende* of *Zeer zwak* buiten de risicogroep van de prestatie-monitor.

**Tabel 1: Aantal scholen per uitkomst van de risicoschatting en risico op de prestatie-monitor VSO.**

		Risico op de prestatie-monitor	
		Risico (rang 1 – 85)	Geen risico (rang 86 – 341)
Inschatting inspecteur en recent oordeel	Inschatting inspecteur: KO-risico	8	5
	Oordeel: <i>Onvoldoende</i> of <i>Zeer zwak</i>	3	3
	Inschatting inspecteur: Wel risico, geen KO	33	14
	Inschatting inspecteur: Geen KO-risico	29	195
	Oordeel: <i>Voldoende</i>	7	22
	Waardering: <i>Goed</i>	0	13
	Geen inschatting inspecteur of oordeel bekend	5	4

In Tabel 2 is te zien dat 3 van de 5 SO-scholen waar een KO-risicovoorstel is gedaan in de risicogroep van de prestatie-monitor vallen. Daarnaast zijn er voor 40 scholen wel risico's gedetecteerd maar is er voor deze scholen geen voorstel voor een KO-risico onderzoek gedaan. 23 scholen hebben de waardering goed, 21 hiervan vallen buiten de risicogroep van de prestatie-monitor, 2 SO-scholen met de waardering goed vallen wel binnen de risicogroep van de prestatie-monitor.

Er zijn 17 scholen waar de prestatie-monitor geen risico's detecteerde maar waar de inspecteur inschatte dat er wel risico's waren. Bij 2 van deze scholen is ook een voorstel tot KO-risico gedaan. Op de overige 15 SO-scholen waren de gedetecteerde risico's niet genoeg om over te gaan tot een KO-risico.

**Tabel 2: Aantal scholen per uitkomst van de risicoschatting en risico op de prestatie-monitor SO.**

		Risico op de prestatie-monitor	
		Risico (rang 1 – 79)	Geen risico (rang 80 – 308)
Inschatting inspecteur en recent oordeel	Inschatting inspecteur: KO-risico	3	2
	Oordeel: <i>Onvoldoende</i> of <i>Zeer zwak</i>	0	0
	Inschatting inspecteur: Wel risico, geen KO	25	15
	Inschatting inspecteur: Geen KO-risico	41	174

Oordeel: <i>Voldoende</i>	1	15
Waardering: <i>Goed</i>	2	21
Geen inschatting inspecteur of oordeel bekend	7	2

### 3.2.1 Verdeling risico-scholen op basis van rang in de prestatie-monitor

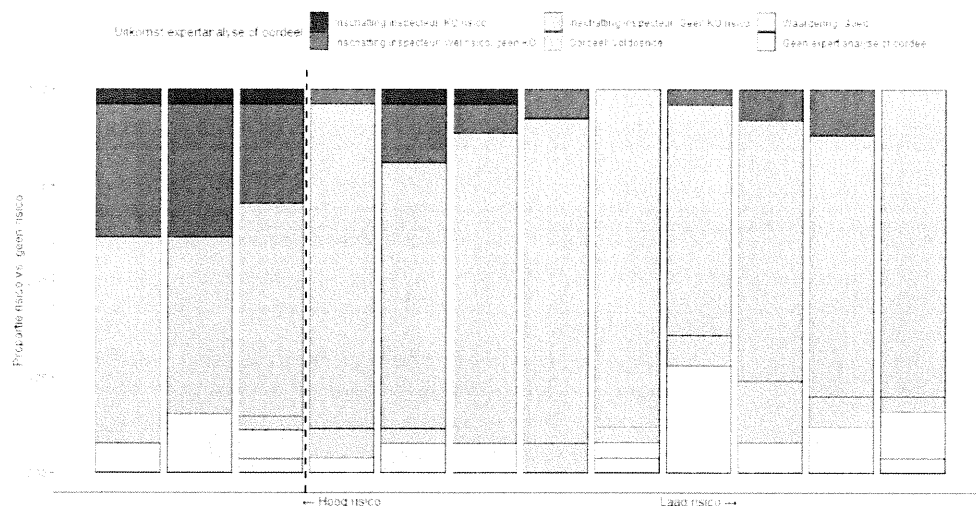
De uitkomsten van de risicoschatting kunnen verdeeld worden over de rangorde in de prestatie-monitor. Hiervoor zijn de scholen onderverdeeld in groepen op basis van hun rang in de prestatie-monitor.

#### 3.2.1.1 VSO: Verdeling risico-scholen op basis van rang in de prestatie-monitor

De 85 VSO-scholen die risico hebben volgens de prestatie-monitor zijn opgesplitst in drie groepen van resp. 28, 28 en 29 scholen. De overige 256 scholen zijn verdeeld in 9 groepen van 28 of 29 scholen (Figuur 1).

In Figuur 1 valt te zien dat het gros van de scholen die een KO-risico kregen -of al een onvoldoende oordeel hadden- ook in de hoogste risicocategorieën terecht zijn gekomen. Daarnaast bevatten deze groepen ook een aantal scholen waarbij wel risico's zijn waargenomen, maar waarvoor niet tot een KO-risico is besloten. Deze resultaten laten zien dat er een behoorlijke voorspellende waarde lijkt uit te gaan van de prestatie-monitor. Desalniettemin worden niet alle risicovolle of onvoldoende scholen in de hoogste risicocategorieën geplaatst.

Met name in de categorie *Inschatting inspecteur: Wel risico, geen KO* worden nog veel scholen gemist, 14 in totaal, een deel hiervan scoort ook erg laag in de prestatie-monitor en lijkt dus helemaal niet gedetecteerd te zijn. Er zijn 5 scholen waar een KO-risicovoorstel is gedaan maar die niet in de risicogroep van de prestatie-monitor zitten, 4 hiervan vallen net onder de risicogrens. De 3 scholen met het oordeel *Onvoldoende* of *Zeer zwak* die niet in de risicogroep van de prestatie-monitor vallen zitten alle 3 ook relatief dicht tegen de risicogrens aan.



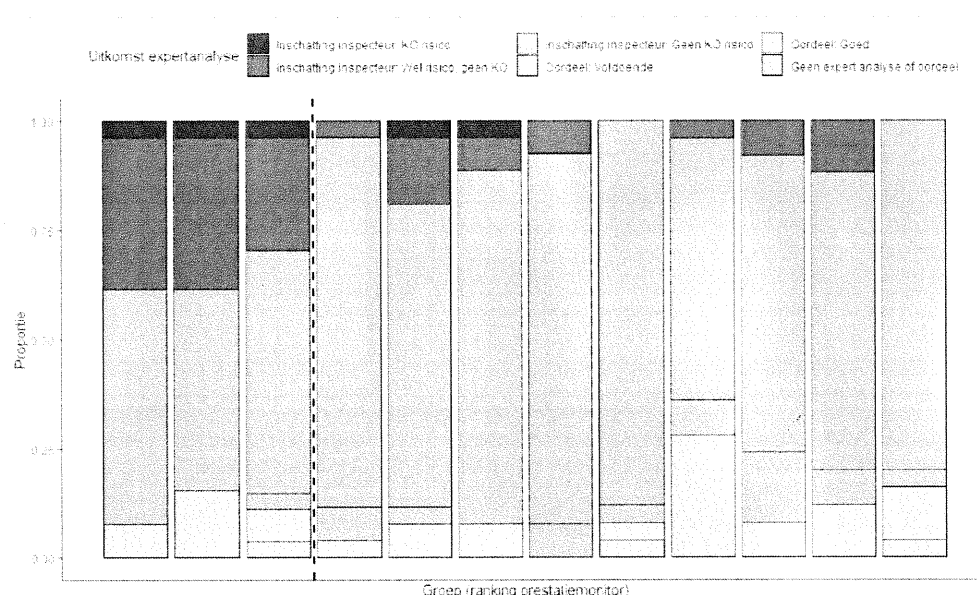
**Figuur 1: De relatie tussen scores op de prestatie-monitor VSO en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een uitkomst van de inschatting van de inspecteur of Oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste drie groepen bevatten de 85 scholen die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 28/29 scholen.**

### 3.2.1.2

#### SO: Verdeling risico-scholen op basis van rang in de prestatie-monitor

De 79 VSO-scholen die risico hebben volgens de prestatie-monitor zijn opgesplitst in drie groepen van resp. 26, 26 en 27 scholen. De overige 229 scholen zijn verdeeld in 9 groepen van 26 of 25 scholen. Figuur 2.

In Figuur 2 is te zien dat het grootste deel van de SO-scholen die risico hebben volgens de risicoschatting in de hoogste risicocategorieën van de prestatie-monitor terecht zijn gekomen. Wel is het zo dat twee scholen waar een voorstel voor een KO-risico is gedaan onder de risicogrens in de prestatie-monitor vallen. Deze zitten in groep 5 en groep 6. Daarnaast zijn er nog 6 scholen waar de inspecteur wel risico's inschat maar geen voorstel voor een KO-risico doet, die erg laag scoren in de prestatie-monitor.



**Figuur 2: De relatie tussen scores op de prestatie-monitor SO en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een uitkomst van de inschatting van de inspecteur of Oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste drie groepen bevatten de 79 scholen die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 25/26 scholen.**

### 3.2.2

#### Foutanalyse

Foutanalyse is een methode om de kwaliteit van risicoschatting te kwantificeren. Dit is zinvol omdat daarmee op termijn de kwaliteit van risicoschattingen over de tijd en over verschillende sectoren/kwaliteitsgebieden vergeleken kan worden. Foutanalyse berust op een onderverdeling van uitkomsten in vier groepen (voortkomend uit Tabel 1 en Tabel 2):

- **Echt positieven (terecht alarm):** Er zijn risico's op een school zowel volgens de prestatie-monitor als volgens de inschatting van de inspecteur of het recente oordeel.
- **Fout negatieven (gemist geval):** De prestatie-monitor detecteert geen risico's, maar volgens de inschatting van de inspecteur of het recente oordeel zijn er wel risico op de school.

- **Fout positieven (loos alarm):** De prestatie-monitor detecteert risico, maar volgens de inschatting van de inspecteur of het recente oordeel zijn er geen risico's
- **Echt negatieven (terecht genegeerd):** Zowel volgens de prestatie-monitor als de inschatting van de inspecteur en het recente oordeel zijn er geen risico's op deze school.

### 3.2.2.1

VSO: Foutanalyse

Vanuit de data van Tabel 1 kunnen we de foutanalyse voor de prestatie-monitor bepalen, zo krijgen we Tabel 3.

**Tabel 3: Foutanalyse op de uitkomst van de prestatie-monitor VSO en de risicoschatting van de inspecteur**

		Risiko op de prestatie-monitor	
		Risiko	Geen risiko
Risiko volgens de inschatting inspecteur en oordeel	Risiko	44 (Echt positieven)	22 (Fout negatieven)
	Geen risiko	41 (Fout positieven)	234 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Zowel op sensitiviteit als op specificiteit scoort de prestatie-monitor VSO heel aardig. 67% van de risicovolle scholen is boven de gestelde risicogrens gekomen. Dit betekent dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die mogelijk risico's hebben (en daarmee ook aan het onderscheiden van scholen waar waarschijnlijk niks aan de hand is). Zeker als we dit vergelijken met het alternatief: het trekken van een willekeurige steekproef. Op precisie scoort de prestatie-monitor minder (52% van de onderzochte scholen is vervolgens gewaardeerd als risicovol). Dit betekent dat er relatief veel scholen moeten worden bekeken in de expertanalyse (omdat ze hoog scoren op de prestatie-monitor, maar waar onvoldoende aanleiding is voor een KO-risico). Een verklaring hiervoor is dat het optimaliseren van de sensitiviteit en specificiteit vaak ten koste gaat van de precisie. Een lage precisie betekent dat er meer capaciteit nodig is van de analisten voor de expertanalyse (wanneer deze uitgevoerd worden).

In Bijlage I hebben we op dezelfde manier een foutanalyse per kwaliteitsgebied gedaan en in Bijlage III is te zien welke indicatoren er per kwaliteitsgebied ontwikkeld zijn. In Bijlage V hebben we een foutanalyse op de 'oude' prestatieanalyse gedaan.

### 3.2.2.2

SO: Foutanalyse

Vanuit de data van Tabel 2 kunnen we de foutanalyse voor de prestatie-monitor bepalen, zo krijgen we Tabel 4.

**Tabel 4: Foutanalyse op de uitkomst van de prestatie-monitor SO en de risicoschatting van de inspecteur**

		Risiko op de prestatie-monitor	
		Risiko	Geen risiko
Risiko volgens de inschatting inspecteur en oordeel	Risiko	28 (Echt positieven)	17 (Fout negatieven)
	Geen risiko	51 (Fout positieven)	212 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Op sensitiviteit scoort de prestatie-monitor SO aardig. Iets minder dan twee derde van de risicoscholen worden gedetecteerd door de prestatie-monitor SO. Ook op specificiteit scoort de prestatie-monitor SO aardig. Wel valt op dat op precisie de prestatie-monitor SO minder scoort.

In Bijlage II hebben we op dezelfde manier een foutanalyse per kwaliteitsgebied gedaan en in Bijlage IV is te zien welke indicatoren er per kwaliteitsgebied ontwikkeld zijn. In Bijlage V hebben we een foutanalyse op de 'oude' prestatieanalyse gedaan.

### 3.2.3

*Foutanalyse alleen aan de hand van oordelen*

Dezelfde foutanalyse kan ook uitgevoerd worden op alleen de recente oordelen in plaats van op zowel de recente oordelen als de inschatting van de inspecteur. Deze manier van analyseren heeft als nadeel dat er minder data beschikbaar is (voor meer scholen heeft een inspecteur afgelopen jaar namelijk een inschatting van risico gedaan, dan dat er oordelen worden gegeven). Het grote voordeel van het doen



van een foutanalyse aan de hand van oordelen is dat de data betrouwbaarder is omdat de school ook daadwerkelijk bezocht is.

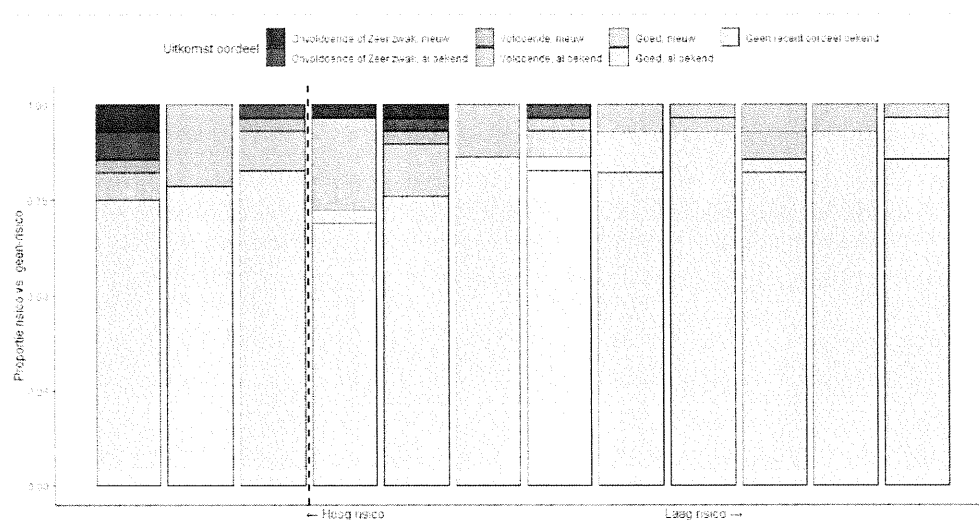
Voor de foutanalyse delen we oordelen in zeven categorieën:

1. **Oordeel: Onvoldoende of Zeer zwak, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
2. **Oordeel: Onvoldoende of Zeer zwak, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Onvoldoende* of *Zeer zwak* gekregen.  
Deze scholen tellen mee als risico in de foutanalyse.
3. **Oordeel: Voldoende, nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
4. **Oordeel: Voldoende, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Voldoende* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
5. **Waardering: Goed, al nieuw**  
Deze scholen hebben tussen 1 september 2020 en 19 februari 2021 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
6. **Waardering: Goed, al bekend**  
Deze scholen hebben tussen 1 augustus 2017 en 31 augustus 2020 het oordeel *Goed* gekregen.  
Deze scholen tellen mee als geen risico in de foutanalyse.
7. **Geen recent oordeel bekend**  
Voor deze scholen is geen recent oordeel bekend.  
Deze scholen tellen mee als geen risico in de foutanalyse.

#### 3.2.3.1

VSO: Foutanalyse aan de hand van oordelen

Op dezelfde manier als Figuur 1 kunnen we de scholen over 12 groepen verdelen. Voor ongeveer 20% van de top 85 van de prestatie-monitor is er een recent oordeel bekend. Bij de 5 scholen uit de risicogroepen is het oordeel *Onvoldoende* of *Zeer zwak* gegeven. Daarnaast zijn er 4 scholen met het oordeel *Onvoldoende* of *Zeer zwak* die buiten de risicogroep van de prestatie-monitor vallen.



**Figuur 3: De relatie tussen scores op de prestatiemonitor en recente oordelen, sinds 1 augustus 2017. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatiemonitor. De eerste drie groepen bevatten de 85 scholen die risico hebben op de prestatiemonitor, de 10 groepen daarna bestaan steeds uit 28/29 scholen.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 1 krijgen we Tabel 5.

**Tabel 5: Foutanalyse op de uitkomst van de prestatiemonitor en recente oordelen**

		Risico op de prestatiemonitor	
		Risico	Geen risico
Risico volgens recent oordeel	Risico	5 (Echt positieven)	4 (Fout negatieven)
	Geen risico	80 (Fout positieven)	252 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.

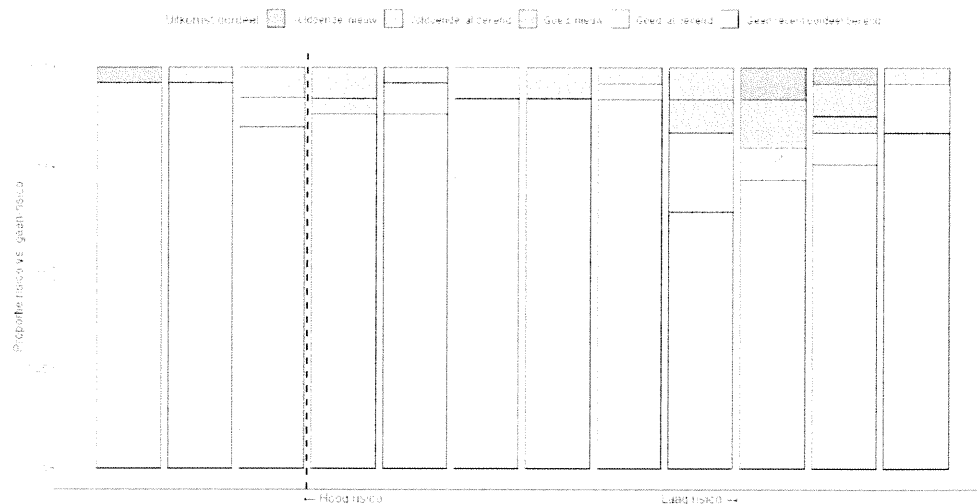
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

De bovenstaande analyse aan de hand van de oordelen gegeven sinds augustus 2017 laat dit zien dat ook aan de hand van recente oordelen de sensitiviteit van de prestatie-monitor redelijk is. De prestatie-monitor VSO detecteert 56% van de scholen die het oordeel Onvoldoende of Zeer zwak hebben, hiermee mist de prestatie-monitor 4 risicovolle scholen. Daarom valt te concluderen dat de prestatie-monitor een goede bijdrage lijkt te kunnen leveren aan het detecteren van scholen die in aanmerking komen voor een KO-risico onderzoek en het oordeel Onvoldoende of Zeer zwak krijgen. Wel is het zo dat er erg weinig recente oordelen bekend zijn en dat het dus onzeker is hoeveel scholen dus daadwerkelijk onvoldoende zijn. Dit is met name een probleem voor de scholen die een lage risicoscore krijgen op de prestatie-monitor, omdat voor deze scholen minder recente oordelen bekend zijn.

### 3.2.3.2

SO: Foutanalyse aan de hand van oordelen

Op dezelfde manier als Figuur 2 kunnen we de scholen over 12 groepen verdelen. Voor ongeveer 8% van de top 79 van de prestatie-monitor is er een recent oordeel bekend. Bij de geen van deze scholen uit de risicogroepen is het oordeel *Onvoldoende* of *Zeer zwak* gegeven.



**Figuur 4: De relatie tussen scores op de prestatie-monitor en recente oordelen, sinds 1 augustus 2017. Scholen zijn gesorteerd op basis van de PM-risicoscore, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van hoe vaak een oordeel voorkomt per groep op basis van de rangorde in de prestatie-monitor. De eerste drie groepen bevatten de 79 scholen die risico hebben op de prestatie-monitor, de 10 groepen daarna bestaan steeds uit 25/26 scholen.**

Als we de recente oordelen (sinds augustus 2017) indelen in dezelfde groepen als Tabel 2 krijgen we Tabel 6.

**Tabel 6: Foutanalyse op de uitkomst van de prestatie-monitor en recente oordelen**

	Risico op de prestatie-monitor
--	--------------------------------

		Risico	Geen risico
Risico volgens recent oordeel	Risico	0 (Echt positieven)	0 (Fout negatieven)
	Geen risico	79 (Fout positieven)	229 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

Omdat recentelijk niet het oordeel *Onvoldoende* of *Zeer zwak* is gegeven, valt de sensitiviteit ook niet te berekenen. Op sensitiviteit scoort de prestatiemonitor SO redelijk 74% van de 'voldoende' scholen werd terecht niet gedetecteerd door de prestatiemonitor.

## 4 Conclusie

De hier uitgevoerde analyse schetst een redelijk positief beeld over de kwaliteit van risicoschatting door de Prestatiemonitor VSO en de Prestatiemonitor SO: respectievelijk 67% en 62% van de scholen die als risicovol zijn volgens de inschatting van de inspecteur of het recente oordeel kwamen terecht in de hoogste risicogroepen. Daarnaast kwam voor de Prestatiemonitor VSO 56% van de scholen die uiteindelijk een onvoldoende oordeel kregen (of behielden) terecht in de risicogroep van de prestatie-monitor. Dit suggereert dat er toegevoegde waarde is voor het gebruik van de prestatie-monitors VSO en SO bij het prioriteren van scholen in het toezicht.

Tegelijkertijd is er ook ruimte voor verbetering. Immers, niet alle risicovolle scholen kwamen in de twee hoogste risicogroepen terecht. Het zal dus een belangrijk doel zijn om de kwaliteit van risicoschatting verder te verbeteren. Wat verder opviel binnen de sector (V)SO was dat relatief veel ovt's zijn onderzocht door inspecteurs, maar dat die expertanalyse relatief kort was, en dat daarin -vergeleken met andere sectoren- de rol van de Prestatiemonitors relatief klein was. Dit komt voor een deel omdat er in het (V)SO relatief minder kwantitatieve gegevens beschikbaar zijn dan in bijvoorbeeld het PO, waar meer gestandaardiseerde (toets)gegevens beschikbaar zijn over alle leerlingen. De sector (V)SO heeft ook nadrukkelijk de wens uitgesproken dat het belangrijk is dat er meer kwantitatieve data beschikbaar komt. Desalniettemin zagen we in deze evaluatie dat ondanks de beperktere hoeveelheid kwantitatieve gegevens in het (V)SO de datagedreven risicoschatting toch een vrij goede voorspelkracht heeft. Dit laat zien dat er wel degelijk informatiewaarde in de beschikbare gegevens zit en dat het dus zond zou zijn om deze gegevens onvoldoende te gebruiken.

Een van de belangrijkste bijdragen van de methode van evaluatie zoals hier beschreven is dat de ontwikkeling van de voorspelkracht in de komende jaren ook gemonitord kan worden. Daarmee kan er gericht en meetbaar gewerkt worden aan verbetering van risicoschatting door de Prestatiemonitor.

Ten slotte is het goed om te benoemen dat een perfecte risicoschatting onmogelijk is. Risico's veranderen over de tijd, en er zijn ook risico's waar we op dit moment simpelweg geen data over hebben (zoals bijvoorbeeld de kwaliteit/ervaring van bestuurders, hoewel hier wel aan gewerkt wordt vanuit het datascience programma). Het detecteren van risicoscholen zal daarom altijd moeten bestaan uit een *combinatie* van zowel datagedreven risicoschatting (zoals in de prestatie-monitor) als het gebruik van de kennis van domeinexperts zoals inspecteurs of analisten, en bijvoorbeeld signalen.

## 5            **Aanbevelingen voor toekomstige ontwikkeling van PM VSO en SO**

1. Binnen de huidige werkwijze bij SO is gekozen om veel (97%) scholen, beknopt te bekijken. De brede dekking is gedeeltelijk, ook vergeleken met andere sectoren, een positief gegeven. De keerzijde is echter dat er per ovt minder tijd besteedt is, en dat de gegevens uit de Prestatiemonitor daarin een beperktere rol gespeeld hebben. Daardoor bestaat het risico dat er nu nog onvoldoende gebruik wordt gemaakt van beschikbare gegevens m.b.t. risico's. Wij stellen daarom voor om meer proportionaliteit in te bouwen in de jaarlijkse analyse. Een uitgebreidere expertanalyse voor de meest risicovolle ovt's is waardevoller voor het evalueren en verbeteren van de prestatie-monitor en deze werkwijze biedt minder kans op het missen van risicovolle ovt's. Deze selectie zou gebaseerd moeten worden op basis van zowel beschikbare data als op basis van actuele kwalitatieve kennis van inspecteurs.

Deze werkwijze maakt het mogelijk om zowel de prestatie-monitor (scholen die boven de gestelde risicogrenzen vallen) als de actuele kennis van de contactinspecteur en analist te gebruiken om scholen te selecteren die een uitgebreidere analyse of onderzoek behoeven. Hierdoor combineren we twee waardevolle informatiebronnen om te komen tot een KO-risico onderzoek: expertkennis én datagedreven risicoschatting.

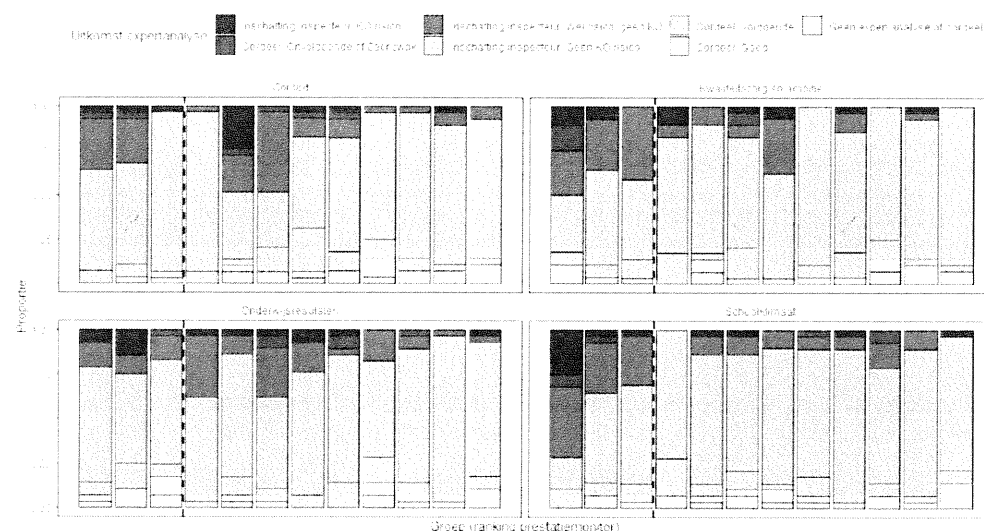
Een mooie bijvangst is dat deze werkwijze tot meer uniformiteit in werkwijze zal leiden tussen sectoren, en daarmee ook de evaluatie van datagedreven risicoschatting inzichtelijker maken.

2. Wanneer we de inschatting van de contactinspecteur combineren met de meest recente oordelen, zijn 19% van de VSO-scholen en 15% van SO-scholen risicovol. De huidige grenswaarde voor de definiëren van wel/geen risico in de prestatie-monitor ligt bij 25%, waardoor er weinig bewegingsruimte is. Wij adviseren het risicogebied te verhogen naar 33% om zekerder te worden dat er minder scholen worden gemist.
3. Momenteel wordt er alleen gedocumenteerd wanneer een KO-risicovoorstel wordt gedaan. Het kan ook interessant zijn om andere toezichtsactiviteiten te gaan vastleggen, zoals bestuurscontact; formeel bestuursgesprek of een opdracht tot eigen onderzoek/analyse. Wij stellen voor dat deze opties worden toegevoegd aan het Excelformulier waarin de uitkomsten worden vastgelegd.
4. Als laatste kunnen de uitkomsten van de evaluatie ook gebruikt worden om verbeteringen in indicatorselectie en weging toe te passen op de prestatie-monitor. In Bijlage VI hebben we hiertoe een aanvullende analyse uitgevoerd. Wij adviseren hierin een aantal indicatoren het komende jaar kritisch te bekijken en eventueel na een volgende foutenanalyse te verwijderen uit de prestatie-monitor. De geldt met name voor de indicatoren: einduitstroom, '3 keer rood', de tussentijdse uitstroom en fusies/splitsingen op bestuursniveau.

## Bijlage I: VSO - Foutanalyse per kwaliteitsgebied

De risicoscores in de prestatie-monitor zijn opgebouwd uit risicoscores op verschillende kwaliteitsgebieden. Bij de Prestatiemonitor VSO zijn dit *Onderwijsproces; Onderwijsresultaten; Schoolklimaat; Context; Kwaliteitszorg en Ambitie; en Financieel Beheer*. In Bijlage III is een overzicht opgenomen van welke indicatoren onder welk kwaliteitsgebied vallen.

Eenzelfde soort figuur als Figuur 1 kan gemaakt worden per kwaliteitsgebied. Hierbij wordt nog steeds de inschatting van de inspecteur en recente oordelen gebruikt als uitkomstmaat. Echter zijn de groepen op de x-as nu verdeeld op basis van de rangorde op één kwaliteitsgebied. Dit is mogelijk voor vier van de zes kwaliteitsgebieden, omdat het kwaliteitsgebied *Financieel beheer* maar gevuld wordt door één indicator (omdat de indicator FB Aangepast financieel toezicht in afgelopen jaar leeg is) en het kwaliteitsgebied *Onderwijsproces* door een programmeerfout in de indicator *OP5 Aantal aanvragen onderwijstijdvermindering*, ook slechts gevuld is door één indicator. Voor deze twee kwaliteitsgebieden valt dus geen rangorde te maken.



**Figuur 5: De relatie tussen scores op de prestatie-monitor per kwaliteitsgebied en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore per kwaliteitsgebied, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de risicoschatting/oordeel voorkomt per groep op basis van de rangorde. De eerste drie groepen zijn de 85 scholen die risico hebben op de prestatie-monitor, de 9 groepen daarna bestaan steeds uit 28/29 scholen.**

Uit Figuur 5 kan de sensitiviteit, specificiteit en precisie per kwaliteitsgebied te berekenen. Zo krijgen we Tabel 9.

**Tabel 9: Waardes voor de foutanalyse per kwaliteitsgebied en de gehele prestatie-monitor.**

	Sensitiviteit	Specificiteit	Precisie
Context	0,30	0,76 - 0,77	0,24 - 0,24

Kwaliteitszorg en ambitie	0,55	0,82 - 0,84	0,42 - 0,45
Onderwijsresultaten	0,27	0,76 - 0,77	0,21 - 0,23
Schoolklimaat	0,58	0,83 - 0,84	0,45 - 0,46
<b>Gehele prestatiemonitor</b>	<b>0,67</b>	<b>0,85</b>	<b>0,52</b>

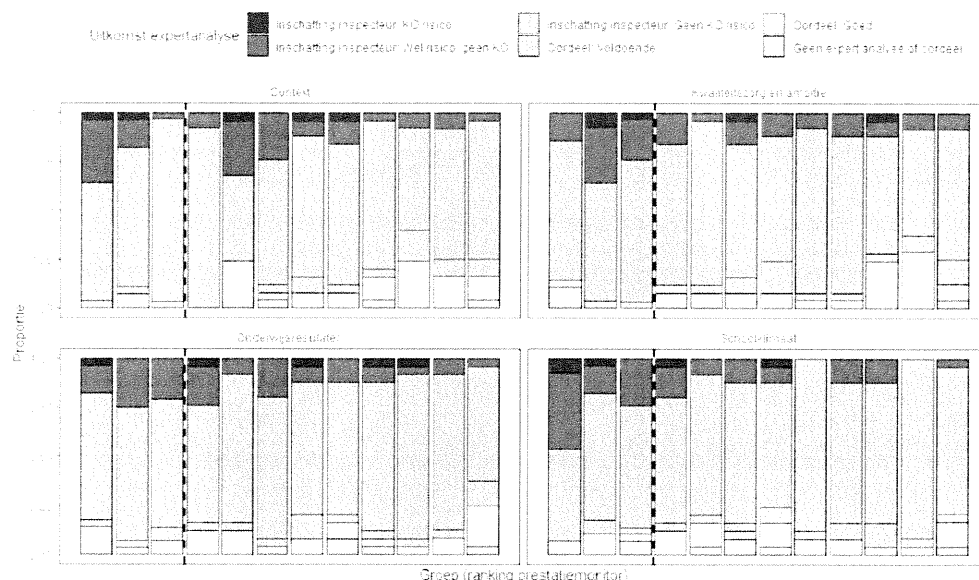
In Tabel 9 is te zien dat de verdeling van risicoscholen op de kwaliteitsgebieden *Kwaliteitszorg en ambitie* en *Schoolklimaat* het meest overeenkomt met de uitkomsten van de risicoschatting/oordelen, en daarmee het meest lijken op de relatie zoals waargenomen voor de totaal-risicoscore van de prestatiemonitor. Voor de kwaliteitsgebieden *Context* en *Onderwijsresultaten* is dit niet het geval en deze kwaliteitsgebieden scoren dan ook erg laag op sensitiviteit en precisie. Omdat het kwaliteitsbied *Onderwijsresultaten* formeel niet wordt beoordeeld, is de lage sensitiviteit op dit kwaliteitsbied deels te verwachten. Tot slot is te zien dat de prestatiemonitor als geheel een betere voorspelkracht heeft dan elk kwaliteitsgebied apart.



## Bijlage II: SO – Foutanalyse per kwaliteitsgebied

De risicoscores in de prestatie-monitor zijn opgebouwd uit risicoscores op verschillende kwaliteitsgebieden. Bij de Prestatiemonitor SO zijn dit *Onderwijsproces; Onderwijsresultaten; Schoolklimaat; Context; Kwaliteitszorg en Ambitie; en Financieel Beheer*. In Bijlage IV is een overzicht opgenomen van welke indicatoren onder welk kwaliteitsgebied vallen.

Eenzelfde soort figuur als Figuur 2 kan gemaakt worden per kwaliteitsgebied. Hierbij is de uitkomst van de risicoschatting door de inspecteur de uitkomstmaat. Echter zijn de groepen op de x-as nu verdeeld op basis van de rangorde op één kwaliteitsgebied. Dit is mogelijk voor vier van de zes kwaliteitsgebieden, omdat het kwaliteitsgebied *Financieel beheer* maar gevuld wordt door één indicator (omdat de indicator *FB Aangepast financieel toezicht in afgelopen jaar leeg* is) en het kwaliteitsgebied *Onderwijsproces* door een programmeerfout in de indicator *OP5 Aantal aanvragen onderwijstijdvermindering*, ook slechts gevuld is door één indicator. Voor deze twee kwaliteitsgebieden valt dus geen rangorde te maken.



**Figuur 6: De relatie tussen scores op de prestatie-monitor per kwaliteitsgebied en waargenomen risico's. Scholen zijn gesorteerd op basis van de PM-risicoscore per kwaliteitsgebied, en verdeeld over 12 groepen van gelijke omvang. Iedere staaf geeft de proportie weer van de hoe vaak een uitkomst van de risicoschatting/oordeel voorkomt per groep op basis van de rangorde. De eerste drie groepen zijn de 79 scholen die risico hebben op de prestatie-monitor, de 9 groepen daarna bestaan steeds uit 25/26 scholen.**

Uit Figuur 6 kan de sensitiviteit, specificiteit en precisie per kwaliteitsgebied te berekenen. Zo krijgen we Tabel 10.

**Tabel 10: Waardes voor de foutanalyse per kwaliteitsgebied en de gehele prestatie-monitor.**

	Sensitiviteit	Specificiteit	Precisie
Context	0,36	0,74	0,19

Kwaliteitszorg en ambitie	0,47	0,76 - 0,77	0,25 - 0,26
Onderwijsresultaten	0,40	0,75 - 0,76	0,21 - 0,22
Schoolklimaat	0,56	0,77 - 0,78	0,29 - 0,31
<b>Gehele prestatiemonitor</b>	<b>0,62</b>	<b>0,81</b>	<b>0,35</b>

In Tabel 10 is te zien dat de verdeling van risicoscholen op het kwaliteitsgebied *Schoolklimaat* het meest overeenkomt met de uitkomsten van de risicoschatting/oordelen, en daarmee het meest lijkt op de relatie zoals waargenomen voor de totaal-risicoscore van de prestatiemonitor. Voor de kwaliteitsgebieden *Context* en *Onderwijsresultaten* is dit niet het geval en deze kwaliteitsgebieden scoren dan ook erg laag op sensitiviteit en precisie. Tot slot is te zien dat de prestatiemonitor als geheel een betere voorspelkracht heeft dan elk kwaliteitsgebied apart.

## Bijlage III: VSO – Risico-indicatoren

De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

Onderwijspro- ces	Onderwijsresul- taten	Schoolklimaat	Kwaliteitszorg en ambitie	Financieel beheer	Context
Bestendiging Ontwikkelingsper- spectief (OP2) <i>Weging = 2</i>	Einduitstroom (OR1) <i>Weging = 2</i>	Percentage signalen Veiligheid Urgent (SK1) <i>Weging = 2</i>	Bestuur: Laatste Oordeel Kwaliteitszorg (KA1) <i>Weging = 1</i>	Aangepast financieel toezicht (FB) <i>Weging = 1</i>	Aantal fusies/splitsingen /overdrachten waarbij het bestuur betrokken was (CT) <i>Weging = 1</i>
Aantal aanvragen onderwijstijdver- mindering (OP5) <i>Weging = 2</i>	IQ irt Uitstroom (OR1) <i>Weging = 2</i>	Percentage signalen Veiligheid Niet- Urgent (SK1) <i>Weging = 2</i>	Bestuur: Laatste Oordeel Kwaliteitscultuur (KA2) <i>Weging = 1</i>	Aangepast financieel toezicht in het afgelopen jaar(FB) <i>Weging = 1</i>	Aantal fusies/splitsingen /overdrachten waarbij het bestuur betrokken was (CT) <i>Weging = 2</i>
	Tussentijdse uitstroom (OR1) <i>Weging = 2</i>	Monitor Sociale Veiligheid (SK1) <i>Weging = 2</i>	Verandering Schoolgrootte (absoluut) (KA2) <i>Weging = 2</i>		
	Verschil SE en CE cijfer (OR1) <i>Weging = 2</i>	Percentage VI Signalen (SK2) <i>Weging = 2</i>	Volatiliteit Schoolgrootte (afgelopen 3 jaar) (KA2) <i>Weging = 1</i>		
	OKE 3 jaar risicovol oude berekening (OR3) <i>Weging = 2</i>	Percentage korte Schorsingen (< 5 dagen) (SK2) <i>Weging = 2</i>	Dagen sinds laatste bezoek (KA3) <i>Weging = 2</i>		
	Plaatsbestendigin- g (OR3) <i>Weging = 2</i>	Percentage lange Schorsingen (>= 5 dagen) (SK2) <i>Weging = 2</i>	Bestuur: Laatste oordeel Verantwoording en dialoog (KA3) <i>Weging = 1</i>		

## Bijlage IV: SO – Risico-indicatoren

De verschillende risico-indicatoren zijn gekoppeld aan de kwaliteitsgebieden uit het onderzoekskader.

Onderwijsproces	Onderwijsresultaten	Schoolklimaat	Kwaliteitszorg en ambitie	Financieel beheer	Context
Bestending Ontwikkelingsperspectief (OP2) <i>Weging = 2</i>	Einduitstroom (OR1) <i>Weging = 2</i>	Percentage signalen Veiligheid Urgent (SK1) <i>Weging = 2</i>	Bestuur: Laatste Oordeel Kwaliteitszorg (KA1) <i>Weging = 1</i>	Aangepast financieel toezicht (FB) <i>Weging = 1</i>	Aantal fusies/splitsingen /overdrachten waarbij het bestuur betrokken was (CT) <i>Weging = 1</i>
Aantal aanvragen onderwijstijdvermindering (OP5) <i>Weging = 2</i>	IQ irt Uitstroom (OR1) <i>Weging = 2</i>	Percentage signalen Veiligheid Niet-Urgent (SK1) <i>Weging = 2</i>	Bestuur: Laatste Oordeel Kwaliteitscultuur (KA2) <i>Weging = 1</i>	Aangepast financieel toezicht in het afgelopen jaar (FB) <i>Weging = 1</i>	Aantal fusies/splitsingen /overdrachten waarbij het bestuur betrokken was (CT) <i>Weging = 2</i>
	Tussentijdse uitstroom (OR1) <i>Weging = 2</i>	Monitor Sociale Veiligheid (SK1) <i>Weging = 2</i>	Procentuele verandering schoolgrootte (KA2) <i>Weging = 2</i>		
	OKE 3 jaar risicovolle berekening (OR3) <i>Weging = 2</i>	Percentage VI Signalen (SK2) <i>Weging = 2</i>	Volatiliteit Schoolgrootte (afgelopen 3 jaar) (KA2) <i>Weging = 1</i>		
	Plaatsbestending (OR3) <i>Weging = 2</i>	Percentage korte Schorsingen (< 5 dagen) (SK2) <i>Weging = 2</i>	Dagen sinds laatste bezoek (KA3) <i>Weging = 2</i>		
		Percentage lange Schorsingen (>= 5 dagen) (SK2) <i>Weging = 2</i>	Bestuur: Laatste oordeel Verantwoording en dialoog (KA3) <i>Weging = 1</i>		

## Bijlage V: Foutanalyse op de 'oude' prestatieanalyse

Vóór het ontwikkelen van de prestatie-monitors VSO en SO maakt (V)SO gebruik van de Kennisanalyse. Deze analyse gaf jaarlijks een risicoschatting per ovt. Wanneer de uitkomst 3 jaar op rij risicovol was, het befaamde '3 keer rood', werd er overgegaan op de toezichtsactiviteiten. Deze 'oude' prestatieanalyse is als losse indicator opgenomen in de beide prestatie-monitors. Hierdoor kunnen we de sensitiviteit, specificiteit en precisie van de 'oude' prestatieanalyse ook uitrekenen, om deze vervolgens te vergelijken met de kengetallen van de prestatie-monitor. Omdat deze 'oude' prestatieanalyse geen gebruik maakt van een rangorde is er echter geen figuur van te maken, zoals wel kan bij de prestatie-monitor (Figuur 1 en Figuur 2). Daarnaast is het belangrijk om te benadrukken dat het vergelijken van beide analysemethoden in termen van beter of slechter in het voorspellen van de risico's voorbarig is, hier gaan we in de conclusie verder op in.

### VSO: Foutanalyse op de 'oude' prestatieanalyse

Er zijn 15 VSO-scholen die volgens de 'oude' prestatieanalyse risico hebben (Tabel 7). Volgens de inschatting van de inspecteurs en de recente oordelen hebben 5 van deze 15 scholen ook daadwerkelijk risico's. Daarnaast zijn er 61 risicovolle VSO-scholen die de 'oude' prestatieanalyse niet detecteert.

**Tabel 7: Foutanalyse op de uitkomst van de 'oude' prestatieanalyse en risicoschattingen en recente oordelen**

		Risico op de 'oude' prestatieanalyse	
		Risico	Geen risico
Risico volgens de inschatting inspecteur en oordeel	Risico	5 (Echt positieven)	61 (Fout negatieven)
	Geen risico	10 (Fout positieven)	265 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

### SO: Foutanalyse op de 'oude' prestatieanalyse

Volgens de 'oude' prestatieanalyse zijn er 19 risicovolle scholen (Tabel 8). Volgens de inschatting van de inspecteurs en de recente oordelen heeft slechts 1 van deze 19 scholen ook daadwerkelijk risico's. Daarnaast zijn er 44 risicovolle SO-scholen die de 'oude' prestatieanalyse niet detecteert.

**Tabel 8: Foutanalyse op de uitkomst van de 'oude' prestatieanalyse en risicoschattingen en recente oordelen**

		Risico op de 'oude' prestatieanalyse	
		Risico	Geen risico
Risico volgens de inschatting inspecteur en oordeel	Risico	1 (Echt positieven)	44 (Fout negatieven)
	Geen risico	18 (Fout positieven)	245 (Echt negatieven)

Met foutanalyse kunnen we de drie maten uitrekenen:

- **Sensitiviteit** -of gevoeligheid- is de mate waarmee een test correct risico's kan detecteren (het % risicovolle scholen dat in beeld is).
- **Specificiteit** -of selectiviteit- is de mate waarin een test correct scholen waar geen risico's aanwezig zijn ook afwijst.
- **Precisie** is de verhouding van het aantal scholen met daadwerkelijk risico, binnen het aantal geselecteerde scholen door de test.

### Vergelijking tussen de prestatiemonitor en 'oude' prestatieanalyse

Zowel voor VSO als SO detecteert de prestatiemonitor op het eerste gezicht meer risicovolle scholen dan de 'oude' prestatieanalyse. De prestatiemonitor VSO detecteert 67% van de risicovolle scholen en de 'oude' prestatieanalyse slechts 8% van de risicovolle scholen volgens de inschatting van de inspecteur en de recente oordelen. De prestatiemonitor SO detecteert 62% en de 'oude' prestatieanalyse 2% van de risicovolle scholen.

Dit grote verschil in het percentage gedetecteerde risicovolle scholen wordt grotendeels verklaard doordat de 'oude' prestatieanalyse hier in het nadeel is. Ten eerste omdat de sector SO het afgelopen jaar de prestatiemonitor heeft gebruikt als risicodetectiemethode en de confirmation bias in het voordeel van de prestatiemonitor is. Ten tweede omdat dat de 'oude' prestatieanalyse maar tussen de 15 en 20 risicovolle scholen detecteert en er afgelopen jaar voor VSO en SO respectievelijk 66 en 45 risicovolle scholen waren volgens de inschatting van inspecteurs en oordelen. In feite zullen we hierdoor de sensitiviteit, specificiteit en precisie van de 'oude' prestatieanalyse onderschatten.

De voornaamste conclusie die we kunnen trekken hiermee is dat de prestatiemonitor ten opzichte van de oude prestatieanalyse wezenlijk andere scholen detecteert. Of

er echt sprake is van een 'betere' detectie kunnen we nooit meten op deze manier, maar de verschillen zijn wel van een dusdanige orde, dat we het veilig achtten aan te nemen dat er vooruitgang is gemaakt.

De prestatiemonitor scoort met name beter op sensitiviteit, de 'oude' prestatieanalyse scoort beter op specificiteit. In andere woorden de prestatiemonitor detecteert sneller een niet-risicovolle school (loos alarm) en de 'oude' prestatieanalyse mist sneller een risicovolle school. Gezien de waarborgfunctie van de inspectie en de plek van de prestatiemonitor in het toezicht, hechten we meer belang aan het niet missen van risicovolle scholen. De prestatiemonitor sluit dus beter aan bij de functie die het moet vervullen.

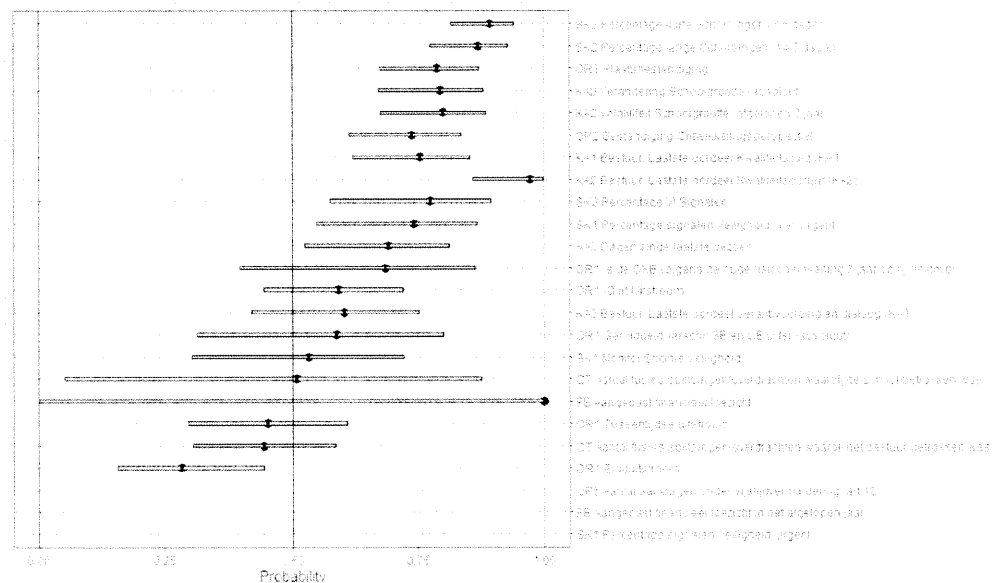
## Bijlage VI: Indicatoranalyse

De uitkomst van de risicoschattingen en recente oordelen kunnen gebruikt worden om de betrouwbaarheid en voorspellende waarde van verschillende indicatoren te bepalen. In onderstaande analyse is hiertoe per school bekeken in welke mate er statistische samenhang bestaat tussen de score op een individuele indicator en de uitkomsten van de risicoschatting/oordelen. Hiervoor is een GLM (Generalized Linear Model) gebruikt met de indicator als onafhankelijke variabele en de uitkomst van de risicoschatting/oordelen als uitkomst (afhankelijke) variabele. Hiervoor zijn de volgende categorieën gecodeerd als 'risico': Inschatting inspecteur: *KO-risico*, Oordeel: *Onvoldoende of Zeer zwak en Inschatting inspecteur: Wel risico, geen KO*.

De uitkomsten van deze analyse zijn kansen (probabiliteiten) tussen 0 en de 1 die aangeven hoe sterk een indicator geassocieerd is met de uitkomst van de risicoschatting/oordelen. Een probability van 1 geeft hierbij aan dat risico op de indicator altijd overeenkomt met risico volgens de risicoschatting/oordelen. Een probability van 0 geeft hierbij aan dat risico op de indicator altijd overeenkomt met géén risico volgens de risicoschatting/oordelen (dus de indicator voorspelt de andere kant op dan we verwachten). Een probability van 0.5 geeft aan dat risico op de indicator met beide uitkomsten van de risicoschatting/oordelen even sterk is geassocieerd, dit is vergelijkbaar met het opgooien van een munt.

### VSO – Indicatoranalyse

De resultaten van de analyse op indicatorniveau voor de Prestatiemonitor VSO worden weergegeven in Figuur 7.



**Figuur 7: Probability waarmee indicatoren uit de Prestatiemonitor VSO de uitkomst van de risicoschatting/oordelen voorspellen. De horizontale balken geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van meest naar minst betrouwbaar.**

SK2 Percentage korte Schorsingen (< 5 dagen) en SK2 Percentage lange Schorsingen (>= 5 dagen) zijn de beste voorspellers van risico als de uitkomst van



de risicoschatting/oordelen. Omdat beide indicatoren goed gevuld zijn en een hoge probability hebben. Deze indicatoren hebben beide ook een hoog gewicht in de prestatie-monitor, dit komt dus overeen met de verwachtingen.

Drie indicatoren hebben een probability van onder de 0,50.

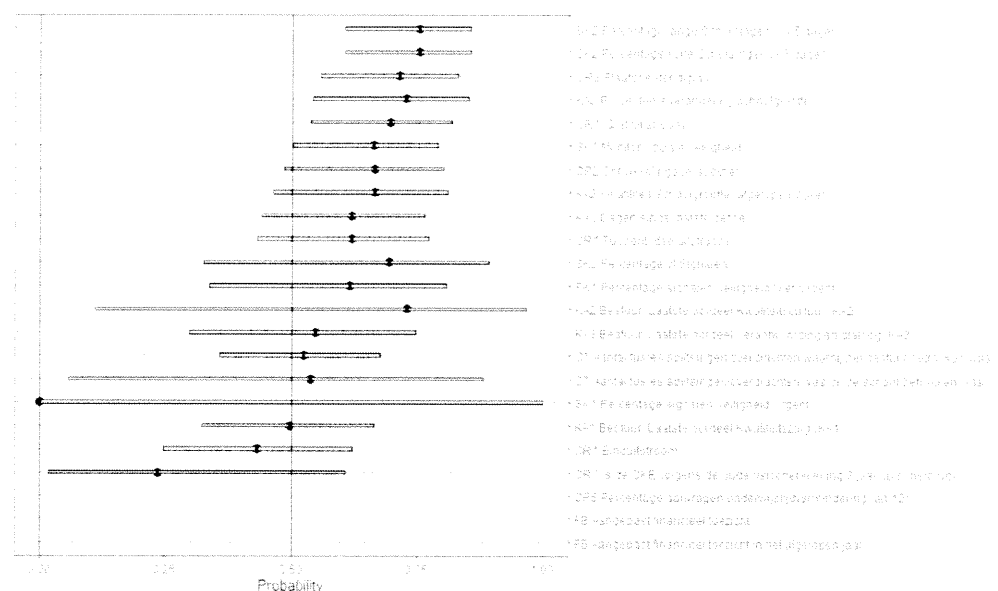
- *OR1 Tussentijdse uitstroom*
- *CT Aantal fusies/splitsingen/overdrachten waarbij het bestuur betrokken was*
- *OR1 Einduitstroom*

Deze indicatoren voorspellen op het moment niet wat we zouden verwachten, namelijk dat risico op één van deze indicatoren juist *geen risico* voorspelt als uitkomst van de risicoschatting door de inspecteur. Dit kan verschillende oorzaken hebben. Voor de indicator *OR1 Tussentijdse uitstroom* geldt dat tussentijdse uitstromen erg afneemt binnen het VSO, hierom voorspelt de indicator wellicht niet zoals we zouden verwachten. Voor de indicator *CT Aantal fusies/splitsingen/overdrachten waarbij het bestuur betrokken was* geldt dat al bekend is dat deze indicator misschien niet weergeeft wat de werkgroep graag zou willen meten, hierom stond een herijking al op de planning. Voor de indicator *OR1 Einduitstroom* is de uitkomst opvallend en deze moet verder uitgezocht worden. Wellicht is er een samenhang met een groei in einduitstroom onbekend. Voor al deze drie indicatoren geldt dat het op dit moment onbetrouwbare indicatoren lijken en dat ze een effectieve risico-indicatie door de prestatie-monitor tegen werken. Daarom is het zinvol om uit te zoeken of deze drie indicatoren goed werken of dat er aanpassingen nodig zijn.

Tot slot valt op dat er in vergelijking tot andere prestatie-monitors weinig verschil is in gewicht tussen verschillende indicatoren. Het is het overwegen waard of sommige betrouwbare indicatoren (zowel vanuit deze analyse als vanuit de kennis van de werkgroep) een hoger gewicht kunnen krijgen.

## SO – Indicatoranalyse

De resultaten van de analyse op indicatorniveau voor de Prestatiemonitor VSO worden weergegeven in Figuur 8.



**Figuur 8: Probability waarmee indicatoren uit de Prestatiemonitor SO de uitkomst van de risicoschatting/oordelen voorspellen. De horizontale balken**

**geven de betrouwbaarheidsintervallen weer en de indicatoren zijn geordend van meest naar minst betrouwbaar.**

*SK2 Percentage korte Schorsingen (< 5 dagen)* en *SK2 Percentage lange Schorsingen (>= 5 dagen)* zijn de beste voorspellers van risico als de uitkomst van de risicoschatting/oordelen. Omdat beide indicatoren goed gevuld zijn en een hoge probability hebben. Deze indicatoren hebben beide ook een hoog gewicht in de prestatie-monitor, dit komt dus overeen met de verwachtingen.

Vier indicatoren hebben een probability van onder de 0,50.

- *SK1 Percentage signalen Veiligheid Urgent*
- *KA1 Bestuur: Laatste oordeel Kwaliteitszorg (KA1)*
- *OR1 Einduitstroom*
- *OR1 Is de OKE volgens de oude risicoberekening 3 jaar op rij risicovol*

Over de indicator *SK1 Percentage signalen Veiligheid Urgent* valt niet veel te zeggen omdat de indicator slecht gevuld is. Slechts één school had in de afgelopen jaren een urgent veiligheidssignaal. Alle andere scholen hadden 0% op deze indicator. *OR1 Is de OKE volgens de oude risicoberekening 3 jaar op rij risicovol* is de indicator die weergeeft hoe de risicodetectie vóór de prestatie-monitor werd gedaan. Daarom is het interessant dat deze indicator op het moment niet voorspelt wat we zouden verwachten.

Al deze indicatoren voorspellen op het moment dus niet wat we zouden verwachten, namelijk dat risico op één van deze indicatoren juist *geen risico* voorspelt als uitkomst van de risicoschatting door de inspecteur. Dit kan verschillende oorzaken hebben. Voor al deze drie indicatoren geldt dat het op dit moment onbetrouwbare indicatoren lijken en dat ze een effectieve risico-indicatie door de prestatie-monitor tegen werken. Daarom is het zinvol om uit te zoeken of deze drie indicatoren goed werken of dat er aanpassingen nodig zijn.

Tot slot valt op dat ook voor de Prestatie-monitor SO er in vergelijking tot andere prestatie-monitors weinig verschil is in gewicht tussen verschillende indicatoren. Het is het overwegen waard of sommige betrouwbare indicatoren (zowel vanuit deze analyse als vanuit de kennis van de werkgroep) een hoger gewicht kunnen krijgen.

**Lege indicatoren**

Voor zowel SO als VSO kan er voor een aantal indicatoren geen uitspraak gedaan worden over de voorspelkracht, omdat deze indicatoren leeg zijn.

Dit zijn betreft de volgende indicatoren:

- *OP5 Percentage aanvragen onderwijstijdvermindering*
- *FB Aangepast financieel toezicht*
- *FB Aangepast financieel toezicht in afgelopen jaar*

Dat de indicator *OP5 Percentage aanvragen onderwijstijdvermindering* leeg is ligt aan een codeerfout. Deze fout is ondertussen al hersteld en vanaf komend jaar kunnen we wel de werking gaan testen. Ten aanzien van financieel toezicht, kunnen we pas oordelen over de voorspelkracht wanneer er daadwerkelijk een SO-bestuur onder toezicht geplaatst gaat worden. Daarnaast kan overwogen worden deze indicator te verplaatsen naar de prestatie-monitor besturen.



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

9.

## **RISICOSCHATting MET DE PRESTATIEMONITORS: EVALUATIE EN AANBEVELINGEN**

5.1.2.e

Juni 2021

## INHOUD

### **Samenvatting 3**

<b>1</b>	<b>Inleiding 5</b>
<b>2</b>	<b>Registratie van expertanalyses t.b.v. de evaluatie 6</b>
<b>3</b>	<b>Kwaliteit van risicoschatting door de Prestatiemonitors in 2020 7</b>
3.1	Indeling uitkomsten expertanalyses en oordelen in categorieën 7
3.2	Kwaliteit van de risicoschatting Prestatiemonitor 8
3.2.1	Vergelijking van de Prestatiemonitors 8
3.2.2	Prestatiemonitors vergeleken met gelijkwaardige risicogrens 10
<b>4</b>	<b>Conclusies 13</b>
<b>5</b>	<b>Aanbevelingen 14</b>
5.1	Rol van de expertanalyses in verbetering van de Prestatiemonitors. 14
5.2	Steekproeftrekking ter voorkoming van tunnelvisie 14
5.3	Gebruik van de Prestatiemonitors in risicogericht toezicht. 15
<b>Bijlage I: Overzicht sectorspecifieke evaluaties van de Prestatiemonitor 16</b>	
<b>Bijlage II: Gedetecteerde risico-ovt's per Prestatiemonitor 17</b>	

## Samenvatting

De Inspectie van het Onderwijs (IvHO) zet sinds enkele jaren steeds sterker in op risicogericht toezicht. Hiermee wordt de toezichtcapaciteit zo efficiënt mogelijk ingezet en is er meer ruimte voor bestuurs- en stelseltoezicht. In de risicogerichte werkwijze spelen (datagedreven) informatieproducten voor risicoschatting een belangrijke rol. Binnen de IvHO wordt die rol voor een belangrijk deel vervuld door de Prestatiemonitors. De Prestatiemonitors hebben als doel de analisten en inspecteurs te ondersteunen bij het prioriteren van ovt's voor nader onderzoek in een expertanalyse<sup>1</sup>. De uitkomst van de expertanalyse kan een voorstel zijn tot verdere toezichtsactiviteiten, zoals een inspectiebezoek. Een duurzame, betrouwbare inzet van Prestatiemonitors kent een aantal belangrijke voorwaarden, zoals een kwalitatief goede risicoschatting en een werkwijze die ruimte biedt voor evaluatie en verbetering van die risicoschatting. Om deze reden is er dit jaar voor het eerst een uitgebreide evaluatie uitgevoerd van de voorspelkracht van zes Prestatiemonitors<sup>2</sup> (BO, VO, VSO, SO, SBO en NK) over 2020. Deze evaluaties hebben we uitgevoerd aan de hand van de registratie van expertanalyses door de sectoren en recente oordelen. De sectorspecifieke evaluaties zijn gedocumenteerd in vijf rapporten (Bijlage I voor een overzicht). Het doel van het voorliggende rapport is om een overkoepelende evaluatie uit te voeren van de voorspelkracht van deze zes Prestatiemonitors, en te reflecteren op en adviseren over de werkwijzen waarbinnen de Prestatiemonitors worden toegepast.

De evaluatie van de huidige risicoschatting laat een overwegend positief beeld zien van de voorspelkracht van de Prestatiemonitors. Alle Prestatiemonitors detecteren een fors deel van de risicovolle ovt's. Dat betekent dat het gebruik van de Prestatiemonitors een duidelijke toegevoegde waarde heeft bij het prioriteren van scholen/opleidingen in het toezicht.

Er bestaat echter ook ruimte voor verbetering. Zo worden niet alle ovt's die volgens analisten en inspecteurs risicovol zijn voldoende opgepikt door de Prestatiemonitors, wat verdere verbetering van de Prestatiemonitors noodzakelijk maakt. Het huidige proces van de expertanalyse is op dit moment echter onvoldoende ingericht op het structureel werken aan verbetering van risicoschatting en risicogericht toezicht. Om dit te realiseren doen wij enkele aanbevelingen voor aanpassingen in werkwijzen. Deze aanbevelingen zijn erop gericht om te komen tot een sectoroverstijgende en cyclische werkwijze voor datagedreven risicoschatting die gericht is op doorlopende evaluatie en verbetering.

1. De registratie van de expertanalyse verschilt tussen de sectoren en daarnaast is het vaak onvoldoende duidelijk of de risicoschatting van de Prestatiemonitors niet goed aansluit bij de inschatting van analisten of inspecteurs, of dat er om andere redenen van de risicoschatting is afgeweken. Hiertoe stellen wij voor om tot een duidelijke sectoroverstijgende registratie van expertanalyses te komen.
2. De sectoren verschillen in het deel van de onder toezicht staande ovt's waar zeer weinig over bekend is vanuit expertanalyses en recente oordelen. Dit gemis aan kennis over groepen ovt's kan leiden tot een vorm van tunnelvisie (een focus op al bekende risico's) en vormt daarmee een risico op zich. Om dit te ondervangen stellen wij voor om jaarlijks op basis van steekproeven aanvullende onderzoeken uit te voeren. De benodigde ruimte zou moeten komen uit het zo proportioneel als mogelijk invullen van onder andere onze 4JO's. Naast het

<sup>1</sup> Met expertanalyse doelen we hier op het proces van desk-research door de analist (of inspecteur) als deel van het jaarlijks monitoren van scholen.

<sup>2</sup> Voor andere Prestatiemonitors waaronder de Prestatiemonitor besturen en de Prestatiemonitor MBO is deze evaluatie over 2020 nog niet mogelijk.

verschaffen van informatie over risico's t.b.v. de verbetering van risicoschatting voorzien we hiermee ook in een representatiever beeld van de kwaliteit van onderwijs op stelselniveau, iets dat nu ontbreekt in de Staat van het Onderwijs.

3. Ten slotte zien we dat er ook verschillen tussen sectoren bestaan in de rol van zowel de Prestatiemonitor als de domeinexperts m.b.t. de schatting van risico's. Dit kan ertoe leiden dat de uiteindelijke risicoschatting te weinig, of juist te veel, leunt op de gegevens in de Prestatiemonitor. Het is onwenselijk als gegevens over risico's niet gebruikt worden en daarmee risicovolle ovt's over het hoofd worden gezien. Tegelijkertijd zijn niet alle risico's in gegevens te vangen, en kan expertkennis sneller aansluiten bij nieuwe risico's die mogelijk nog niet goed in data zichtbaar zijn. Datagedreven risicoschatting en kennis o.b.v. kwalitatieve, of 'zachte', informatie moet daarom in samenhang gebruikt worden. Het is daarom belangrijk om in de werkprocessen rond risicogericht toezicht de rol van zowel datagedreven risicoschatting als de kennis van domeinexperts (analisten en inspecteurs) over risico's beter te verankeren.

## **1 Inleiding**

De Directie Kennis ontwikkelt in samenwerking met de toezichtsectoren jaarlijks de zogenaamde Prestatiemonitors voor de verschillende sectoren. De Prestatiemonitors beogen inspecteurs en analisten te ondersteunen bij het prioriteren en onderzoeken van ovt's, gebaseerd op data over mogelijke risicofactoren. Daarmee geven de Prestatiemonitors invulling aan ten minste twee van de uitgangspunten uit het waarderingskader. De Prestatiemonitors dragen bij aan Waarborgfunctie door domeinexperts te helpen bij het prioriteren van scholen die risico's vertonen m.b.t. de basiskwaliteit. Daarnaast dragen de Prestatiemonitors bij aan de invulling van Proportionaliteit (toezicht op maat) omdat ze niet alleen risicovolle ovt's in beeld helpen brengen, maar in veel gevallen ook handvatten bieden over waar de mogelijke risico's mee samen hangen. Deze kennis kan gebruikt worden om toezichtsactiviteiten gericht in te zetten.

Gezien de rol die de Prestatiemonitors vervullen binnen de IvHO is het belangrijk dat de kwaliteit van de Prestatiemonitors goed geborgd is. Bijvoorbeeld op het vlak van datakwaliteit, maar ook of de risicoschatting voldoende aansluit bij de risico's zoals domeinexperts deze kennen. Dit laatste sluit aan op een bredere maatschappelijke discussie over het gebruik van algoritmes, en de zorg dat algoritmen (zoals de Prestatiemonitors) onwenselijke impact hebben op de prioritering van onder toezicht staanden door overheden. Het is daarom belangrijk dat risicoschatting continue wordt geëvalueerd en verbeterd om het aan te laten sluiten bij inzichten over risico's door domeinexperts.

Om de kwaliteit van de Prestatiemonitors te borgen en continue te verbeteren wordt er vanaf dit jaar door de directie Kennis een jaarlijkse evaluatie uitgevoerd op de kwaliteit van risicoschatting door de verschillende Prestatiemonitors. De afgelopen periode is deze evaluatie uitgevoerd op de Prestatiemonitors BO, VO, VSO, SO, SBO en NK over 2020 (voor onder andere de Prestatiemonitor besturen en MBO is deze evaluatie niet mogelijk over 2020). Het doel van deze evaluaties is om de voorspelkracht van de Prestatiemonitors te monitoren en te verbeteren, zowel door verbetering van indicatoren en grenswaardes als het doen van aanbevelingen om de plek van de Prestatiemonitor in werkprocessen zoals de expertanalyses te verbeteren. Voor de evaluatie worden de risicoscores van de Prestatiemonitor afgezet tegen risicoschattingen door domeinexperts. Dit gebeurt op basis van de beoordeling die voortkomt uit de expertanalyse; en de daadwerkelijke oordelen die door inspecteurs zijn gegeven tijdens een inspectie. In deze overkoepelende evaluatie vergelijken we de werkwijze en voorspelkracht van verschillende Prestatiemonitors. Daarnaast doen we aanbevelingen om de werkwijze en Prestatiemonitors verder te verbeteren en meer op één lijn te krijgen tussen sectoren en werkgroepen.

## **2 Registratie van expertanalyses t.b.v. de evaluatie**

Om de risicoscores van de Prestatiemonitor te vergelijken met de uitkomsten van de expertanalyse is het nodig om de uitkomsten van de expertanalyses op gestructureerde wijze vast te leggen. Afgelopen jaar is dit door alle sectoren vastgelegd aan de hand van een Excelformulier dat binnen iedere sector is bijgehouden. Deze formulieren zijn van groot belang voor de evaluatie die in hoofdstuk 3 wordt beschreven. Alle sectoren verdienen complimenten dat dit zo gestructureerd is vastgelegd afgelopen jaar. Er waren echter wel onderling grote verschillen in het gebruik en hoe deze formulieren gevuld zijn.

Zowel de sector PO (Prestatiemonitor BO, SBO en PO Nieuwkomers) als SO (Prestatiemonitor VSO en SO) hebben beide zeer overzichtelijke formulieren gebruikt voor de registratie van de risicoschatting. De manier van werken binnen de sector SO verschilde echter van andere sectoren doordat de expertanalyses zijn uitgevoerd door inspecteurs aan de hand van met name kwalitatieve kennis (de kennis in hun hoofden). In vergelijking met de andere sectoren zijn de expertanalyses in de sector SO ook minder gebaseerd op data en is er minder tijd besteed per expertanalyse. Hierdoor bestaat de kans dat er op SO-scholen risico's spelen waar de IvHO al wel data over heeft, maar die onvoldoende mee worden gewogen in de risico-inschatting door inspecteurs.

Binnen de sector PO zaten er soms datafouten in het formulier, waarschijnlijk vanwege het grote aantal mensen wat met het hetzelfde formulier werkte. Binnen de sector VO viel juist de uitgebreide documentatie op. Zo zijn niet alleen voorstellen tot risico-onderzoeken door analisten gedocumenteerd, maar ook andere toezichtsactiviteiten, zoals bestuurscontact; formeel bestuursgesprek; een opdracht tot eigen onderzoek/analyse of het meenemen in een 4JOB of een herstelonderzoek, wat een gedetailleerder beeld oplevert.

Tot slot zijn binnen alle sectoren voornamelijk de voorstellen tot risico-onderzoeken of andere toezichtsactiviteiten vastgelegd. Er zijn echter soms redenen om niet over te gaan tot een toezichtsactiviteit, terwijl de risico's die Prestatiemonitor detecteert wel terecht zijn. Dit gebeurt bijvoorbeeld omdat er vertrouwen in het bestuur is; omdat de school net bezocht is; of omdat er een herstelonderzoek loopt. Hierdoor was het soms onduidelijk of een school met een hoge risicoscore volgens de analist of inspecteur eigenlijk niet risicovol is (een fout van de Prestatiemonitor), of dat een school wel risicovol is maar dat er andere redenen zijn om niet over te gaan tot een toezichtsactiviteit. Hiervoor is het van belang dat duidelijk en structureel gedocumenteerd wanneer er niet wordt overgaan tot een toezichtsactiviteit, maar de risico's die geconstateerd worden door de Prestatiemonitor wel worden onderschreven door analist of inspecteur. In de aanbevelingen komen de hierboven genoemde verbeterpunten terug.



### 3 Kwaliteit van risicoschatting door de Prestatiemonitors in 2020

Om de kwaliteit van de risicoschatting van de Prestatiemonitors inzichtelijk te maken hebben we een zogenaamde foutanalyse uitgevoerd. Hiervoor hebben we gekeken hoe de uitkomst van risicoschatting door analisten of inspecteur en het huidige oordeel zich verhouden tot de positie die scholen hebben in de Prestatiemonitor (m.a.w.; hoe 'hoog' ze staan).

De grens voor het risicogebied in de Prestatiemonitor is per sector verschillend, en is bepaald door een sectorspecifieke werkgroep. Het percentage ovt's in de risicogroep verschilt dus tussen sectoren en is mede afhankelijk van de beschikbare (capaciteit aan) analisten en inspecteurs voor expertanalyses, en hoe uitgebreid de expertanalyses zijn uitgevoerd. Zowel de relatieve capaciteit en omvang van de expertanalyses lijkt te verschillen tussen de sectoren.

#### 3.1 Indeling uitkomsten expertanalyses en oordelen in categorieën

Voor de foutanalyse hebben we de uitkomst van de expertanalyse door de analist of inschatting door de inspecteur ingedeeld naar 3 mogelijke uitkomsten: 1) *Expertanalyse: Risico-onderzoek of andere toezichtsactiviteit*; 2) *Expertanalyse: Wel risico, geen toezichtsactiviteit*; 3) *Expertanalyse: Geen risico (en dus geen KO)*. In sommige gevallen is er bijvoorbeeld besloten om geen risico-onderzoek uit te voeren omdat een school 'al in beeld' is vanwege het huidige oordeel (bijvoorbeeld een onvoldoende school die nog even de tijd krijgt om herstelplan uit te voeren). Om deze scholen ook in beeld te brengen zijn voor de scholen die geen risico-onderzoek krijgen maar die wel een huidig oordeel hebben de huidige oordelen meegenomen<sup>3</sup>: *Zeer zwak*; *Onvoldoende*; *Voldoende*; en *Goed*. Oordelen worden meegenomen als ze gegeven zijn sinds augustus 2017.

Daarmee komen we tot de volgende categorieën:

##### Ovt's met risico's:

1. **Expertanalyse: Risico-onderzoek of andere toezichtsactiviteit**  
Als de analist of inspecteur een voorstel tot risico-onderzoek of een andere toezichtsactiviteit doet.
2. **Oordeel Onvoldoende of Zeer zwak**  
Deze school heeft het oordeel *Onvoldoende* of *Zeer zwak*. Deze scholen tellen mee als risico in de foutanalyse.
3. **Expertanalyse: Wel risico, geen toezichtsactiviteit**  
Als er geen voorstel voor een risico-onderzoek of een andere toezichtsactiviteit wordt gedaan, maar wel risico's op de school worden geconstateerd. Dit is bijvoorbeeld wanneer er vertrouwen is in het bestuur of als er een herstelonderzoek nog gaande is of net is afgerond. Deze scholen tellen mee als risico in de foutanalyse.

##### Ovt's zonder risico's:

1. **Expertanalyse: Geen toezichtsactiviteit**  
De analist of inspecteur constateert geen risico's die voldoende aanleiding geven voor een toezichtsactiviteit.
2. **Oordeel: Voldoende**  
Er is geen expertanalyse uitgevoerd afgelopen jaar maar deze school heeft het oordeel *Voldoende*. Deze scholen tellen mee als geen risico in de foutanalyse.

---

<sup>3</sup> Waar we spreken over een oordeel kan het ook gaan om de waardering Goed.

### 3. Waardering: Goed

Er is geen expertanalyse uitgevoerd afgelopen jaar maar deze school heeft het oordeel *Goed*. Deze scholen tellen mee als geen risico in de foutanalyse.

#### Ovt's zonder recente risicoschatting door analist of inspecteur:

##### 1. Geen expertanalyse uitgevoerd of recent oordeel gegeven

Deze ovt's tellen mee als geen risico in de foutanalyse.

## 3.2

### Kwaliteit van de risicoschatting Prestatiemonitor

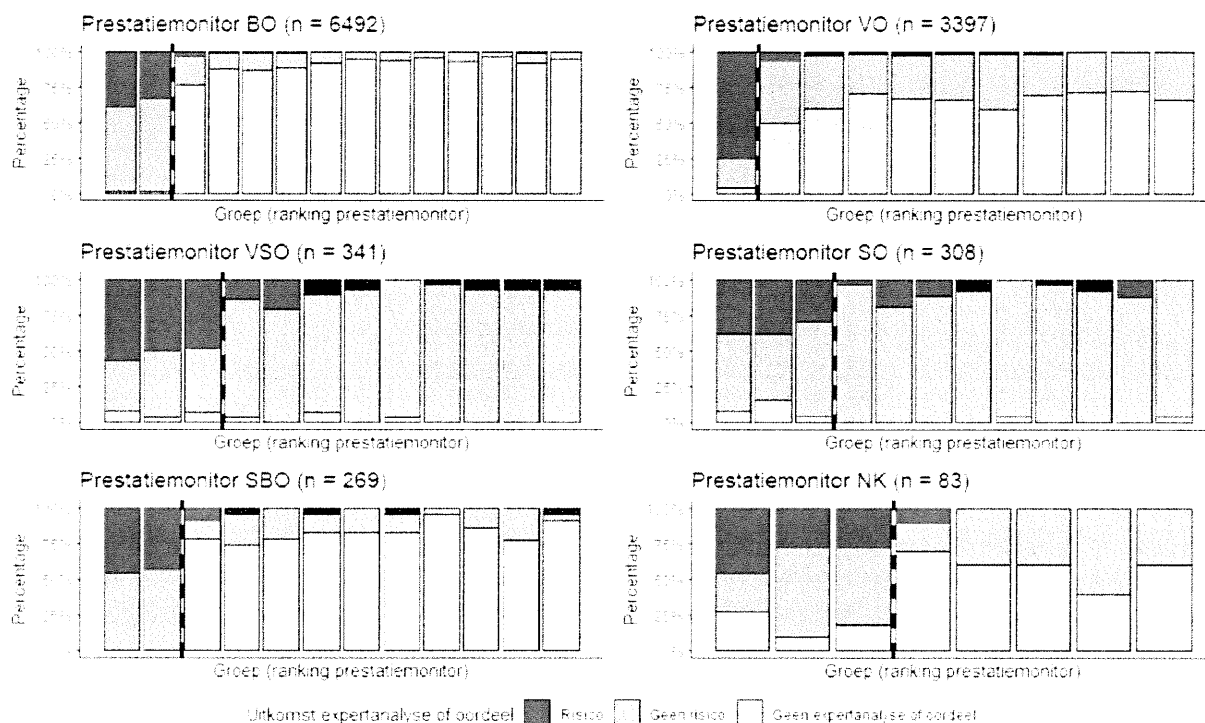
De verschillende evaluaties per Prestatiemonitor hebben een aantal technische verbeterpunten naar voren gebracht, zoals foutieve indicatoren, adviezen voor aanpassingen van grenswaardes en wegingen of het weglaten van indicatoren. Deze verbeterpunten zijn verder toegelicht in de desbetreffende evaluaties (zie Tabel 2 in Bijlage II voor verwijzingen) en gedeeld met de ontwikkelaars van de desbetreffende Prestatiemonitors. In deze overkoepelend evaluatie richten we ons met name op de kwaliteit van risicoschatting van de Prestatiemonitors en de vergelijking van de Prestatiemonitoren onderling.

#### 3.2.1

##### *Vergelijking van de Prestatiemonitors*

De uitkomsten van de expertanalyse en recente oordelen kunnen vergeleken worden met de rangorde in de Prestatiemonitor. Hiervoor zijn de scholen onderverdeeld in groepen op basis van hun rang in de Prestatiemonitor. Figuur 1 biedt een eerste inzicht in de kwaliteit van risicoschatting van de Prestatiemonitors voor de verschillende sectoren. In het figuur zijn de verschillende sectoren opgenomen als aparte panelen. Ter verduidelijking: het paneel linksboven geeft weer in welke mate de risicoschatting door de Prestatiemonitor BO aansluit bij de risicoschatting door de analisten/inspecteurs. In het figuur staan alle scholen in het PBO, gesorteerd op basis van hun risicoscore, waarbij risicovolle scholen aan de linkerkant staan, en scholen met een lage risicoscore rechts. Scholen zijn voor dit figuur samengevoegd in 14 groepen (van ongeveer 475 scholen elk) die zijn weergegeven als balken. De meest linker balk omvat de 475 scholen met de hoogste risico's. Verticaal binnen een balk staat het percentage dat weergeeft hoe vaak de verschillende uitkomsten van de risicoschattingen en oordelen voorkomen. Rode kleuren geven scholen weer die volgens de expertanalyse of huidige oordelen risicovol zijn, groene kleuren zijn scholen waar volgens de expertanalyse of oordelen juist geen risico's zijn. Grijs vlakken geven scholen weer waar geen recent oordeel of expertanalyse over heeft plaatsgevonden.

De gestippelde lijn geeft de risicogrens weer zoals deze bepaald is door de werkgroep, de groepen links van deze grens vallen dus in de risicogroep. Als de Prestatiemonitor volledig perfect zou zijn, zouden dus alle risico-ovt's (de rode kleuren) helemaal links in het figuur moeten staan (en liefst links van de gestippelde lijn) en alle niet-risicovolle ovt's juist aan de rechterkant. Voor de Prestatiemonitor BO zien we dat het merendeel van de risicovolle ovt's links van de gestippelde lijn valt. Dit suggereert dat de risicoschatting van de Prestatiemonitor relatief vaak goed aansluit bij de uiteindelijke uitkomst van de expertanalyse. Op eenzelfde wijze kan ook de kwaliteit van risicoschatting in de andere sectoren bekeken worden.



**Figuur 1: De relatie tussen scores op de Prestatiemonitor en waargenomen risico's tijdens de expertanalyse. Ovt's zijn gesorteerd op basis van de PM-risicoscore, en samengevoegd in groepen die gekozen zijn om tot afgebakende groepen boven en onder de risicogrens te komen. Iedere staaf geeft het percentage weer van risicovolle ovt's volgens de expertanalyse of oordeel. De groep(en) links van de gestippelde lijn vallen in de risicogroep van de betreffende Prestatiemonitor, volgens de risicogrens zoals deze bepaald is door de werkgroep. *N.B. in het (V)SO behelst de expertanalyse een korter proces, waardoor er relatief meer ovt's überhaupt beoordeeld zijn. Deze expertanalyses behelzen daarmee een groter deel van de populatie maar zijn minder uitgebreid, waardoor de kans bestaat dat er risico's over het hoofd worden gezien.***

Het totaalbeeld dat uit Figuur 1 naar voren komt is overwegend positief. Voor elke Prestatiemonitor is het percentage risicovolle ovt's immers groter binnen de risicogroepen (links van de stippellijnen), dan daarbuiten. Wel missen alle Prestatiemonitors nog een deel van de risicovolle ovt's in hun sector.

De Prestatiemonitor NK is volgens deze analyse de best presterende Prestatiemonitor. Bijna alle risicovolle ovt's vallen in één van de risicogroepen en de enige risicovolle NK-voorziening die gemist wordt valt net onder de risicogrens. Ook de Prestatiemonitors BO, VO en SBO presteren alle drie vrij goed, maar missen wel nog een deel van de risicovolle ovt's. Met name bij de Prestatiemonitor BO en de Prestatiemonitor SBO valt echter op dat over veel ovt's geen kennis is vanuit een expertanalyse of recent oordeel (de grijze vlakken rechts van de stippellijn zijn in deze sectoren sterk oververtegenwoordigd). Dit hangt waarschijnlijk samen met het grote aantal ovt's dat binnen deze sector valt. Desalniettemin vergroot dit aspect de kans op het onzichtbaar blijven van risico's, en bemoeilijkt daarmee ook de mogelijkheid om te testen of de Prestatiemonitors wel de juiste risico's aan het licht brengen. Dit betekent dus dat de kwaliteit van de risicoschatting voor deze Prestatiemonitors een stuk *onzekerder* is dan voor de andere Prestatiemonitors. Bij de Prestatiemonitor VO zien we daarentegen dat er relatief veel ovt's bekeken die géén risico hadden volgens de Prestatiemonitor. Dit zorgt voor een duidelijker beeld

op de 'laag-risico' populatie. Dit draagt bij aan de kwaliteit van de evaluatie van de Prestatiemonitor VO.

De Prestatiemonitors VSO en SO missen juist relatief meer risicovolle ovt's dan de andere Prestatiemonitors, maar hier bestaat wel een duidelijker beeld van de scholen/opleidingen die onder de risicogrens vallen. En ook zien we dat deze Prestatiemonitors nog een redelijk goed risicodetecterend vermogen hebben (het gros van de risicovolle ovt's valt immers links van de stippellijn). Zoals eerder benoemd is een belangrijk aandachtspunt hierbij echter dat voor de risicoschatting binnen (V)SO in mindere mate gebruik is gemaakt van gegevens in de Prestatiemonitor. Daarmee wijkt de werkwijze dus af van de andere sectoren. Dit beperkt de vergelijking met de andere sectoren.

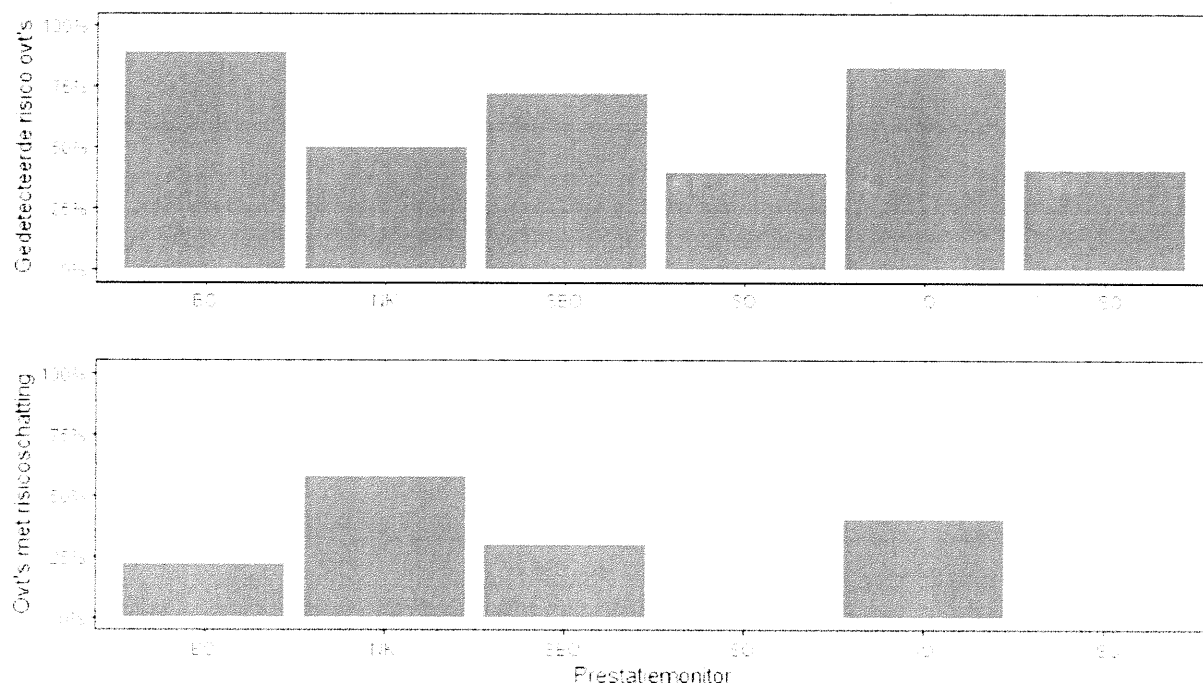
### 3.2.2 *Prestatiemonitors vergeleken met gelijkwaardige risicogrens*

De waarde vanaf wanneer een ovt in de risicogroep valt wordt bepaald door een sectorspecifieke werkgroep prestatie-monitor. Zie Tabel 2 (in Bijlage II) voor de percentielgrenzen per sector. Deze risicogrenzen bepalen het percentage aan ovt's dat überhaupt in de risicogroep valt. Dit verschilt soms behoorlijk tussen sectoren. Waar bij de Prestatiemonitor NK 37% van de ovt's in de risicogroep vallen is dit bij de Prestatiemonitor VO slechts 9%. Dit valt enerzijds te verklaren doordat er veel minder nieuwkomers-voorzieningen dan ovt's binnen VO zijn (en het dus makkelijker is om over een groot deel van de ovt's expertanalyses uit te voeren). Anderzijds kiest VO ervoor ook veel expertanalyses uit te voeren op risico-ovt's die via een andere weg gedetecteerd zijn (bijvoorbeeld via onderwijsresultaten of kennis van de contactinspecteur). Willen we echter de verschillende Prestatiemonitors als risicomodel met elkaar vergelijken, dan is het goed om ze te vergelijken aan de hand van een gelijk percentage van ovt's dat binnen de risicogroep valt. Voor de volgende analyse hebben we dit percentage voor elke Prestatiemonitor op 15% gezet. We kiezen voor 15% omdat dit het percentage is dat BO -de grootste sector- hanteert en omdat dit ook het percentage is dat vaak in de Prestatiemonitor wordt gebruikt voor losse indicatoren.

Als belangrijkste uitkomstmaat nemen we hier het percentage risicovolle ovt's dat is gedetecteerd door de Prestatiemonitor. Dat wil zeggen; welk aandeel van de ovt's die uit de expertanalyse (of o.b.v. recent oordeel) als risicovol geclassificeerd zijn, komen in de bovenste 15% van de Prestatiemonitors terecht<sup>4</sup>. Een ander belangrijk punt dat we aanhaalden is de mate van vertrouwen die we in de evaluatie van de risicoschatting kunnen hebben, omdat voor sommige sectoren veel minder recente onderzoeken of expertanalyses zijn uitgevoerd bij laag-risico ovt's.

---

<sup>4</sup> Zoals zichtbaar in de bijlage staat de risicogrens voor VO op 9%, waardoor er bij VO ook scholen binnen het 15<sup>e</sup> percentiel niet zijn onderzocht. Omdat VO echter ook relatief veel laag-scorende ovt's heeft onderzocht passen we de risicogrens van 15% toe.



**Figuur 2: Per Prestatiemonitor het percentage aan risicovolle ovt's dat is gedetecteerd (bovenste paneel), en het percentage ovt's waarvan een risicoschatting bekend is (onderste paneel). Voor het bovenste paneel is de risicogrens van de Prestatiemonitor gezet op het 15<sup>e</sup> percentiel. Voor het onderste paneel laat zien welk aandeel van alle ovt's een recent oordeel heeft of is onderzocht in een expertanalyse. Dit paneel geeft dus een indicatie van de mate van betrouwbaarheid van de evaluatie van de risicoschatting. Op basis van de gegevens lijkt de Prestatiemonitor BO de best presterende Prestatiemonitor, omdat maar liefst 89% van de ovt's die op basis van de expertanalyse of een recent oordeel als risicovol worden aangeduid ook in de bovenste 15% van de Prestatiemonitor terechtkwamen. Twee andere goed presterende Prestatiemonitors zijn VO en SBO, met respectievelijk 83% en 72% van de risicovolle ovt's dat gedetecteerd wordt.**

In het bovenste paneel van Figuur 2 wordt, per Prestatiemonitor, weergegeven welk percentage van de ovt's die volgens de expertanalyse risicovol waren ook in de bovenste 15% van de respectievelijke Prestatiemonitor terechtkwamen. Het onderste paneel laat zien welk aandeel van alle ovt's een recent oordeel heeft of is onderzocht in een expertanalyse. Dit paneel geeft dus een indicatie van de mate van betrouwbaarheid van de evaluatie van de risicoschatting.

Op basis van de gegevens lijkt de Prestatiemonitor BO de best presterende Prestatiemonitor, omdat maar liefst 89% van de ovt's die op basis van de expertanalyse of een recent oordeel als risicovol worden aangeduid ook in de bovenste 15% van de Prestatiemonitor terechtkwamen. Twee andere goed presterende Prestatiemonitors zijn VO en SBO, met respectievelijk 83% en 72% van de risicovolle ovt's dat gedetecteerd wordt.

Echter is bij de Prestatiemonitors BO en SBO relatief weinig bekend over de ovt's die een lage risicoscore kregen (paneel 2). Voor SBO is slechts 30% recent onderzocht, en voor BO slecht 22%. Het is erg waarschijnlijk dat ook in deze sectoren er in feite risicovolle ovt's bestaan onder de groep die een lage score krijgen uit de Prestatiemonitor, maar die vanwege de sterk risicogerichte werkwijze buiten beeld blijven. Het is daardoor niet uit te sluiten, en wellicht zelfs aannemelijk, dat er voor deze Prestatiemonitors een te rooskleurig beeld uit de analyse naar voren komt.

Hoewel dit probleem inherent is aan een risicogerichte werkwijze, en dus bij alle sectoren in enige mate speelt, is het probleem bij bijvoorbeeld de Prestatiemonitor VO minder aanwezig. Bij VO is over 40% van de ovt's informatie bekend vanuit de expertanalyses en recente oordelen. Daarnaast kwamen bij VO ook ovt's in aanmerking voor een expertanalyse via andere risicodetectiemethoden (bijvoorbeeld kennis van de COI of het onderwijsresultatenmodel). Zoals eerder ook

uit Figuur 1 naar voren kwam hebben we ook over ovt's met een lage risicoscore dus nog een relatief goed beeld (ongeveer 25% van de VO-ovt's die onder de risicogrens vallen is alsnog onderzocht). Wel is het ook hier belangrijk om te noemen dat de huidige werkwijze binnen VO het probleem niet volledig oplost: de 'laag-risico' ovt's die wel onderzocht zijn doorgaans niet willekeurig geselecteerd. Daarmee kunnen deze resultaten alsnog een vertekend beeld geven.

De Prestatiemonitors NK, VSO en SO lijken het hier minder goed te doen. Desalniettemin is ook voor deze prestatie-monitoren de risicoschatting boven kans en heeft daarmee dus toegevoegde waarde. Zoals eerder genoemd speelt bij de Prestatiemonitors SO en VSO wel de relatief andere werkwijze binnen de expertanalyse, wat de resultaten wellicht minder makkelijk te vergelijken maakt. Mogelijk heeft de Prestatiemonitor bijvoorbeeld risico's opgepikt die onvoldoende onder de aandacht zijn gekomen van de inspecteur.

De Prestatiemonitor NK was met de eigen gekozen risicogrens van 37% nog de best presterende Prestatiemonitor, waar slechts 1 risicovolle ovt werd gemist. Met een uniforme risicogrens van 15% presteert deze Prestatiemonitor echter een stuk minder goed. Dit laat zien dat de goede score in de eerste analyse voor een deel kwam omdat er binnen de werkgroep PO-NK een vrij ruime risicogrens toegepast is. De specifieke risicodetectie van de hoog-risicovolle ovt's voor de PM-NK kan dus nog verbeterd worden, want deze eindigen niet bovenin in de Prestatiemonitor.

Ten slotte is het bij de hier uitgevoerde analyses ook belangrijk om ingedachte te houden dat het vergelijken van de uitkomst van de expertanalyse met de positie in de rangorde van scholen in de Prestatiemonitor enige vorm zelfbevestigingsbias kan hebben. Analisten en inspecteurs wordt immers gevraagd om de Prestatiemonitor ook te gebruiken bij het vormen van hun oordeel tijdens de expertanalyse. Aanvullende verkenningen (hier niet uitgebreid beschreven) hebben laten zien dat dit probleem in enige mate speelt: ovt's die nét boven de risicogrens vielen leken een grotere kans te hebben om als 'risicovol' te worden aangeduid door analisten/inspecteurs dan ovt's die daar net onder vielen (terwijl die ovt's in feite dus weinig van elkaar verschillen in datagedreven risicoschatting). Deze observatie laat zien dat het belangrijk is om in de communicatie te blijven benadrukken dat analisten en inspecteurs een eigen inschatting moeten maken van risico's, en zich vrij moeten voelen om af te wijken van de risicoschatting van de Prestatiemonitor.

## 4 Conclusies

De hier uitgevoerde analyse schetst een overwegend positief beeld over de kwaliteit van risicoschatting door de Prestatiemonitors. Alle Prestatiemonitors detecteren een groot deel van de risicovolle-ovt's. Dit laat zien dat er een duidelijke toegevoegde waarde is voor het gebruik van de Prestatiemonitors bij het prioriteren van scholen in het toezicht.

Tegelijkertijd is er ruimte voor verbetering. Immers, niet álle risicovolle ovt's kwamen in de hoogste risicogroepen van de Prestatiemonitors terecht. Dit maakt het belangrijk om ook op een sectoroverstijgende manier te werken aan een duurzame inzet op verbetering van de risicoschatting. Uit de evaluaties zijn hiervoor een aantal belangrijke aandachtspunten naar voren. Ten eerste viel op dat de registratie van de uitkomsten van de expertanalyse op verschillende wijze plaatsvindt in verschillende sectoren. Dit bemoeilijkt de evaluatie van risicoschatting. Ten tweede viel met name bij SBO en BO op dat er relatief veel ovt's zijn waar weinig over bekend is. Hoewel dit inherent is aan een risicogerichte werkwijze, ontbreekt bij er in deze sectoren een robuust beeld van de kwaliteit van onderwijs bij scholen die lage risicoscores krijgen van de Prestatiemonitor. Dit is belangrijk omdat dit er voor zorgt dat de IvhO in feite geen representatief stelselbeeld heeft van het aandeel scholen dat de basiskwaliteit niet haalt. Dit gebrek in een beeld op stelselniveau vertaalt zich ook door in een onzekerheid over de uitkomsten van evaluatie voor deze sectoren. We weten immers niet hoeveel extra onvoldoende scholen zich in de 'grijze vlakken' bevinden. Hoewel in mindere mate, speelt dit probleem ook bij de andere sectoren omdat de laag-risico ovt's die wel onderzocht zijn doorgaans niet willekeurig zijn gekozen. Ten slotte zagen we dat de rol van de Prestatiemonitor in de expertanalyse verschilt tussen de sectoren (een bijvoorbeeld grote rol in PO, en een juist een relatief kleine rol in (V)SO). Daarmee bestaat het risico dat er te veel, of juist te weinig gebruik wordt gemaakt van de gegevens die beschikbaar zijn gemaakt in de Prestatiemonitor. Het is daarom belangrijk om toe te werken naar een duidelijkere en sectoroverstijgende positionering van de rol van expertkennis en datagedreven inzichten in risicogericht toezicht. In de volgende sectie doen we naar aanleiding van deze conclusies drie aanbevelingen.

## 5 Aanbevelingen

De rol van datagedreven risicoschatting binnen de IvhO wordt steeds groter. Dit is belangrijk omdat het helpt om risicogericht toezicht vorm te geven, en daarmee zowel de toezichtlast voor scholen en besturen te beperken en om toezichtcapaciteit proportioneel en efficiënt in te zetten. Dit biedt daarnaast ruimte voor een grotere rol voor bestuurs- en stelselgericht toezicht. Een duurzaam gebruik van risicoschatting vraagt echter om een meer uniforme en cyclische werkwijze met een belangrijke plek voor evaluatie en verbetering. Hierbij wordt inspectie capaciteit zo efficiënt mogelijk ingezet om enerzijds de ovt's met de (bekende typen) risico's te identificeren en te laten beoordelen door domeinexpert, en anderzijds om de risicomodellen te verifiëren opdat de toekomstige kwaliteit ervan ook gewaarborgd blijft. Om dit te realiseren doen wij enkele aanbevelingen voor aanpassingen in werkwijzen.

### 5.1 Rol van de expertanalyses in verbetering van de Prestatiemonitors.

In de evaluatie merkten we dat de registratie van de expertanalyses verschilde tussen de sectoren. Gestructureerde registratie van de expertanalyses is belangrijk zodat deze gebruikt kunnen worden in een cyclisch proces van het evalueren en verbeteren van Prestatiemonitors. Hiervoor is het belangrijk dat van alle ovt's die bekeken/beoordeeld worden binnen een sector er op een uniforme wijze geregistreerd wordt of er volgens de analist of inspecteur risico's zijn, ook als de betreffende ovt's niet via de Prestatiemonitor in beeld zijn gekomen.

Daarnaast is het belangrijk om niet alleen de risico-onderzoeken te registreren. Zo wordt er in sommige gevallen voor gekozen om geen risico-onderzoek uit te voeren, terwijl de inspecteur of analist wel risico's constateert op het ovt. Dit kan zijn omdat er vertrouwen is in het bestuur; er al een herstelonderzoek gaande is; of omdat het ovt al in een traject zit. Voor toekomstige evaluatie en verbetering van de Prestatiemonitor zou het helpen om tot een nog duidelijkere scheiding te komen tussen het bestaan van risico's (ofwel, vindt de analist de risicoscore terecht) en de beslissing voor vervolgtoezicht (wordt er een risico-onderzoek uitgevoerd). Met name het eerste is belangrijk voor een goede evaluatie van de Prestatiemonitor.

Hiertoe stellen wij voor om de registratie van het oordeel door de analist op te splitsten in 'Zijn er risico's op de school?' en 'Voorstel voor toezichtactiviteit'.

*Eerste stappen die al gezet zijn: In het nieuwe Format 4JO wordt op ons verzoek aan analisten gevraagd om de risicoschatting van de Prestatiemonitors en die van de expertanalyse als geheel apart te registreren (en eventuele verschillen te duiden). Voor de registratie van de Jaarlijkse Prestatie Analyse (JPA) wordt gewerkt aan een uniforme wijze van vastlegging van de kwaliteit van risicoschatting.*

### 5.2 Steekproeftrekking ter voorkoming van tunnelvisie

Voor veel sectoren ontbreekt momenteel het beeld van risico's op ovt's die laag scoren in de Prestatiemonitor. Ovt's die laag scoren in de Prestatiemonitor worden minder vaak onderworpen aan een expertanalyse en daarnaast zijn er over deze ovt's ook minder recente oordelen bekend. Dit zorgt ervoor dat we relatief weinig kennis hebben over ovt's die niet boven de risicogrens van de Prestatiemonitor komen. Dit kan ertoe leiden dat er structureel bepaalde risico's gemist worden door de Prestatiemonitor, maar, belangrijker nog; ook in het toezicht als geheel. Dergelijke onzekerheden kunnen zelfs leiden tot een vorm van tunnelvisie in ons risicogericht toezicht (omdat we alleen ovt's onderzoeken op basis van al bekende



risicoprofielen). Een belangrijk gevolg hiervan is dat de IvhO momenteel geen representatief stelselbeeld heeft van het aandeel scholen/opleidingen dat er niet in slaagt om de basiskwaliteit te realiseren. Het aandeel bekende onvoldoende ovt's waarover de IvhO rapporteert vormen in feite slechts een de *ondergrens* van het werkelijke aantal. Dit belemmert ons zicht op de basiskwaliteit op stelselniveau. Om dit te ondervangen stellen wij voor om jaarlijks op basis van steekproeven onderzoeken uit te voeren. Dit tevens in navolging van de aanbevelingen o.b.v. recent onderzoek van de Vrije Universiteit dat in samenwerking met de IvhO is uitgevoerd<sup>5</sup>. Dergelijke steekproeven zouden idealiter op twee niveaus plaatsvinden:

1) Het doel van de Prestatiemonitors is om op basis van gegevens de inzet van expertanalyse zo efficiënt mogelijk in te zetten. Omdat zowel de expertanalyses als de Prestatiemonitor met name gebaseerd zijn op beschikbare gegevens (in tegenstelling tot de meer kwalitatieve informatie die wordt opgehaald tijdens een schoolbezoek) is het belangrijk om de Prestatiemonitors te ijken aan de hand van uitkomsten van expertanalyses. *Om de Prestatiemonitors goed te ijken zijn daarom expertanalyses nodig die zijn uitgevoerd op deels willekeurig geselecteerde ovt's (steekproef niveau 1; expertanalyse).*

2) Een belangrijker probleem vormt echter het eerder genoemde gebrek aan een representatief beeld van de onderwijskwaliteit op stelselniveau. Dergelijke stelselbeelden zijn belangrijk om te kunnen monitoren en rapporteren over de kwaliteit van het onderwijsbestel als geheel, met name waar het gaat om het realiseren van de basiskwaliteit (bijvoorbeeld in de Staat van het Onderwijs). *Hiervoor is het noodzakelijk om ook steekproefsgewijs ovt's te bezoeken en te beoordelen (steekproef niveau 2; onderzoeken op ovt's).* Dergelijke inzichten op stelselniveau zullen vervolgens ook van grote waarde zijn bij het evalueren en ijken van de Prestatiemonitors als instrument van risicoschatting.

Een mogelijk bezwaar tegen de voorgestelde steekproeven is dat deze op gespannen voet lijken te staan met het uitgangspunt van de proportionaliteit. Bij het prioriteren van ovt's voor risico-onderzoeken speelt proportionaliteit een grote rol: ovt's met lage risico's behoeven immer weinig, of zeer gerichte, interventies; en bij ovt's met grotere, maar specifieke risico's kan capaciteit ook gericht worden ingezet. Bij het uitvoeren van steekproefsgewijze onderzoeken daarentegen is het belangrijk om onderzoeken wel breder in te zetten. D.w.z.; bij een schoolbezoek in het kader van een steekproef is het belangrijk om een breed palet aan kwaliteitsgebieden te onderzoeken, juist omdat deze onderzoeken verkennend van aard zijn. Hoewel dit op het eerste gezicht niet lijkt aan te sluiten op een proportionele werkwijze is dit nodig om ook op langere termijn capaciteit proportioneel in te kunnen blijven zetten. De capaciteit die nodig is voor het uitvoeren van verkenningen zal vrij moeten komen uit de capaciteit die beschikbaar komt door de meer proportionele werkwijze in zowel de risicogerichte onderzoekslijn, als die van de 4-jaarlijkse bestuursonderzoeken. Zo is het is zinvoller om risico-onderzoeken bij goed functionerende besturen over te laten aan het bestuur zelf en onze capaciteit juist te richten op het goed onderzoeken van steekproeven. Hiermee dienen we namelijk zowel het instellingtoezicht als het stelseltoezicht.

*Eerste stappen die al gezet zijn: Dit jaar zullen binnen de sector VO 30 semi-willekeurig geselecteerde ovt's worden onderworpen aan een expertanalyse, waarbij de analist niet actief wordt verteld dat het om een steekproef gaat. In andere sectoren is dit nog niet opgezet. Er is nog in geen van de sectoren een steekproef*

---

<sup>5</sup> Savi, A. O., Cornelisz, I., Sjerps, M. J., Greup, S. L., Bres, C. M., & van Klaveren, C. (2020). Balancing Trade-offs in the Detection of Schools at Risk. <https://edarxiv.org/q6z4r/>

*opgezet voor het uitvoeren van daadwerkelijke risico-onderzoeken bij scholen/opleidingen.*

### **5.3**

#### **Gebruik van de Prestatiemonitors in risicogericht toezicht.**

Ten slotte viel bij de evaluatie op dat de rol van de Prestatiemonitor en de rol van de domeinexpert verschilt tussen de sectoren. Het is belangrijk om hier een sectoroverstijgende werkwijze in vast te stellen. Bij het maken van de selectie van ovt's die bekeken worden in een expertanalyse, en bij het nemen van beslissingen over verdere toezichtsactiviteiten, is het belangrijk om de plek van zowel de Prestatiemonitors als die van de domeinexpert (bijvoorbeeld de contactinspecteur of analist) goed te verankeren. Dit betekent enerzijds dat de experts bij de prioritering van ovt's, maar ook in het uiteindelijke expertanalyse, kennisnemen van de gegevens in de Prestatiemonitor. Deze bevat immers een veelheid aan gegevens waar inspecteurs en analisten waarschijnlijk niet volledig zicht op hebben voor alle besturen/scholen waar ze toezicht op houden. Anderzijds is het belangrijk dat analisten en inspecteurs zich te allen tijde vrij moeten voelen om af te wijken van de risicoschatting van de prestatie-monitor, bijvoorbeeld op basis van aanvullende, meer kwalitatieve informatie. Dat wil zeggen; datagedreven risicoschatting dient altijd gewogen en beoordeeld worden door analisten en inspecteurs. De uiteindelijke keuze voor het overgaan tot een toezichtactiviteit ligt daarmee bij de analist en/of inspecteur. Wij raden aan om zowel de rol van de Prestatiemonitors als de rol van de domeinexpert een duidelijke plaats te geven in de procesbeschrijvingen/werkinstructies.

## **Bijlage I: Overzicht sectorspecifieke evaluaties van de Prestatiemonitor**

**Tabel 1: Prozalocaties van de sectorspecifieke evaluaties van de voorspelkracht van de Prestatiemonitor**

Rapport	Prozalink
Evaluatie voorspelkracht Prestatiemonitor PO BO 2020	<a href="https://proza.ocw.local/otcs/cs.exe/app/nodes/27218299">https://proza.ocw.local/otcs/cs.exe/app/nodes/27218299</a>
Evaluatie voorspelkracht Prestatiemonitor SBO 2020	<a href="https://proza.ocw.local/otcs/cs.exe/app/nodes/27201584">https://proza.ocw.local/otcs/cs.exe/app/nodes/27201584</a>
Evaluatie voorspelkracht Prestatiemonitor PO NK 2020	<a href="https://proza.ocw.local/otcs/cs.exe/app/nodes/27191786">https://proza.ocw.local/otcs/cs.exe/app/nodes/27191786</a>
Evaluatie voorspelkracht Prestatiemonitor VO 2020	<a href="https://proza.ocw.local/otcs/cs.exe/app/nodes/26865644">https://proza.ocw.local/otcs/cs.exe/app/nodes/26865644</a>
Evaluatie voorspelkracht Prestatiemonitor (V)SO 2020	<a href="https://proza.ocw.local/otcs/cs.exe/app/nodes/27383900">https://proza.ocw.local/otcs/cs.exe/app/nodes/27383900</a>

## Bijlage II: Gedetecteerde risico-ovt's per Prestatiemonitor

**Tabel 2: Per Prestatiemonitor het aantal ovt's wat onder een Prestatiemonitor valt, hoeveel procent hiervan in het risicogebied van de Prestatiemonitor vallen, over hoeveel ovt's er in totaal informatie bekend is en het percentage gedetecteerde risico-ovt's.**

Prestatiemonitor	Totaal aantal ovt's	Percentage ovt's in het risicogebied	Percentage ovt's met risicoschatting of oordeel	Percentage correct gedetecteerd risico-ovt's	Percentage correct gedetecteerde risico-ovt's met 15%-risicogebied
Prestatiemonitor NK	83	37%	58%	92%	50%
Prestatiemonitor BO	6492	15%	22%	88%	89%
Prestatiemonitor VO	3397	9%	40%	77%	83%
Prestatiemonitor SBO	269	16%	30%	76%	72%
Prestatiemonitor VSO	341	25%	97%	67%	41%
Prestatiemonitor SO	308	26%	97%	62%	40%

# **Algemene handleiding prestatiemonitors**

Toelichting voor gebruikers

## *Inhoudsopgave*

Inhoudsopgave.....	2
Inleiding.....	2
Plaats in het toezicht.....	3
Huidige Prestatiemonitors.....	3
Waar is de prestatie monitor te vinden?.....	4
<i>Toegang tot het DWH-portaal.....</i>	<i>4</i>
Via het startmenu.....	4
Via het Rijksportaal.....	4
<i>De rapportage van de prestatie monitor in Power BI.....</i>	<i>5</i>
<i>Printen Power BI rapport.....</i>	<i>10</i>
<b>Hoe is de prestatie monitor opgebouwd?.....</b>	<b>11</b>
<b>Uitleg berekening risicoscore prestatie monitor.....</b>	<b>11</b>
<b>Verantwoord gebruik van risicoscores.....</b>	<b>13</b>
<b>Actualiteit van de gegevens in de prestatie monitor.....</b>	<b>13</b>
<b>Beheer: eigenaarschap en ontwikkeling.....</b>	<b>13</b>
<b>Feedback: vragen en opmerkingen over de prestatie monitor.....</b>	<b>13</b>
<b>Bijlage I Kwaliteitsgebieden en standaarden uit het onderzoekskader 2017, per prestatie monitor.....</b>	<b>14</b>
<b>Bijlage II Kwaliteitsgebieden en standaarden uit het onderzoekskader 2021, per prestatie monitor.....</b>	<b>15</b>

## *Inleiding*

In deze handleiding wordt voor gebruikers een nadere toelichting gegeven op:

- De plaats van de prestatie monitor in het toezicht.
- De vindplaats en het gebruik van de prestatie monitor in het datawarehouse (dwh).
- De opbouw van de prestatie monitor.
- Uitleg berekening risicoscore.
- Gebruik risicoscore.
- De actualiteit van de gegevens en beheer van de prestatie monitor.
- Bij wie u terecht kunt voor feedback en vragen over de prestatie monitor.

### *Plaats in het toezicht*

Voor de uitvoering van onze waarborgfunctie monitoren we de ontwikkeling en prestaties van besturen, scholen en opleidingen. Hiervoor voeren we jaarlijks een prestatie- en risicoanalyse uit (de JPRA). Dit doen we om tijdig eventuele risico's te detecteren en inzicht te krijgen in het functioneren van onze objecten van toezicht.

Als er op basis van de jaarlijkse prestatie- en risicoanalyse geen vermoedens van risico's zijn dan wordt de analyse afgesloten. De objecten van toezicht worden dan niet actief over de uitkomst van de analyse geïnformeerd. Als er een vermoeden van risico's is dan volgen er verdere stappen in het toezicht, zoals in eerste instantie een expertanalyse (desk-research).

Daarnaast maken we in de uitvoering van ons toezicht gebruik van proportionaliteit en maatwerk. Dit betekent dat we de intensiteit van het toezicht op het bestuur en de scholen afstemmen op de kwaliteit van objecten van toezicht. De gegevens uit de prestatiemonitor dragen bij aan de inschatting van de kwaliteit daarvan door inzicht te geven in het al dan niet aanwezig zijn van (mogelijke) risico's voor de onderwijskwaliteit en/of risico's voor de kwaliteit van het bestuur.

De prestatiemonitors zijn interne informatieproducten die wordt gebruikt ter ondersteuning van de JPRA. In 2017-2018 is begonnen met het ontwikkelen van de eerste prestatiemonitors.

Redenen voor het ontwikkelen van de prestatiemonitor zijn:

- aansluiten bij het vernieuwde toezicht,
- vermijden onnodige sectorale verschillen, o.a. in verschijningsvorm, infrastructuur en productiewijze,
- een gebruiksvriendelijk informatieproduct en
- optimaal gebruik maken van de beschikbare innovatieve mogelijkheden.

De prestatiemonitor dient niet als basisinformatie voor de expertanalyse, daar is de verdiepinginformatie voor ontwikkeld.

### *Huidige Prestatiemonitors*

In onderstaande tabel 1 staat weergegeven welke prestatiemonitors momenteel beschikbaar zijn.

Tabel 1. Overzicht van de verschillende prestatiemonitors augustus 2023

Sector	Prestatiemonitor	Analyse_code
BST – Bestuurlijk toezicht	besturen geïntegreerd	PM_BST_BST
BST – Bestuurlijk toezicht	Besturen financieel	PM_BST_FIN
PO – Primair onderwijs	basisonderwijs	PM_PO_BO
PO – Primair onderwijs	speciaal basisonderwijs	PM_PO_SBO
PO – Primair onderwijs	nieuwkomers	PM_PO_NK
SO – Speciaal onderwijs	speciaal onderwijs	PM_EC_SO
SO – Speciaal onderwijs	voortgezet speciaal onderwijs	PM_EC_VSO
VO – Voortgezet onderwijs	voortgezet onderwijs	PM_VO_VO
VO – Voortgezet onderwijs	praktijkonderwijs	PM_VO_PRO
VO – Voortgezet onderwijs	Eerste opvang anderstaligen	PM_VO_EOA

De meest recente editie en versie van de Prestatiemonitor wordt automatisch weergegeven in Power-BI. De editie wordt aangeduid door de samenvoeging van het jaar en de maand waarin de editie is verschenen (JJJJMM), de versie bestaat uit jaar, maand en dag (JJJJMMDD). Voor meer informatie verwijs ik u naar de paragraaf Actualiteit van de gegevens in de prestatiemonitor.

## Waar is de prestatie-monitor te vinden?

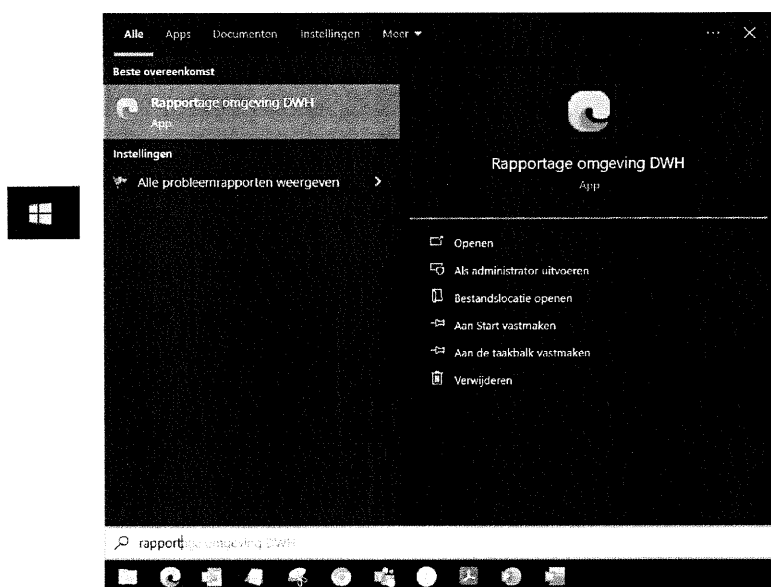
Sinds 2019 is de prestatie-monitor beschikbaar in Power BI, te benaderen vanaf de startpagina van de rapportage omgeving DWH. Power BI is een selfservice BI tool. Met deze tool kunt u zelf gegevens inzien, filteren en sorteren.

## Toegang tot het DWH-portaal

De Power BI weergave van de prestatie-monitor is voor alle IvHO-medewerkers te benaderen via de Rapportage omgeving van het DWH.

### Via het startmenu

Door in het startmenu een deel van 'Rapportage omgeving DWH' in te typen komt de App Rapportage omgeving DWH naar voren. Door op deze App te klikken gaat u naar de startpagina van de veelgebruikte informatieproducten (zie Figuur 3. Overzicht DWH-rapportages ).

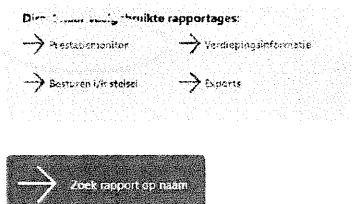


Figuur 1. Verwijzing naar DWH-portaal via startmenu

Hierna verschijnt het onderstaande scherm, zie Figuur 3. Overzicht DWH-rapportages . Voor de weergave in Power BI klikt u linksboven, in het gele vak, op Prestatiemonitor. Het is ook mogelijk om te zoeken op het rapport of rechts op het lichtblauwe vakje te klikken (en daarna op Verder).

## Rapportage omgeving DWH

12/07/2021 10:02:15

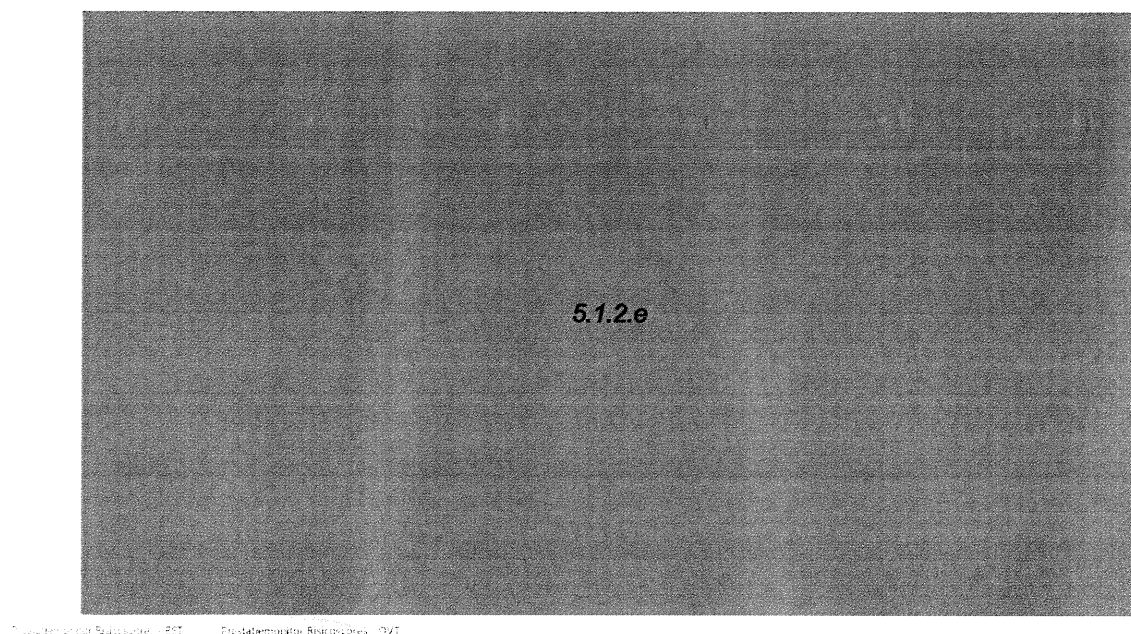


Figuur 3. Overzicht DWH-rapportages



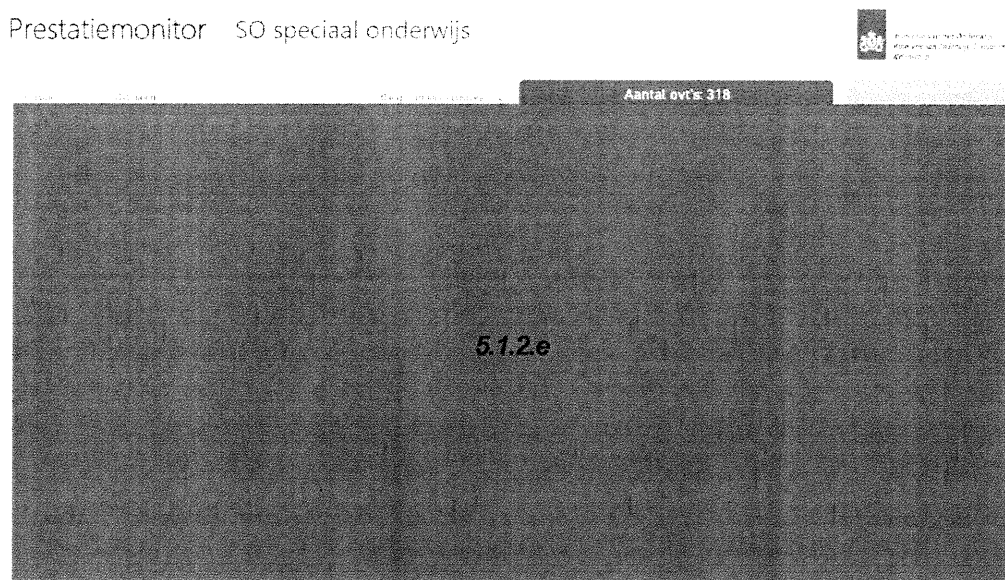
## De rapportage van de Prestatiemonitor in Power BI

Voor de startpagina van de prestatie-monitor. Selecteer op deze pagina onderaan eerst het tabblad besturen (BST) of objecten van toezicht (OVT). Selecteer daarna de prestatie-monitor die u wilt bekijken. Zie ook de instructie in het scherm.



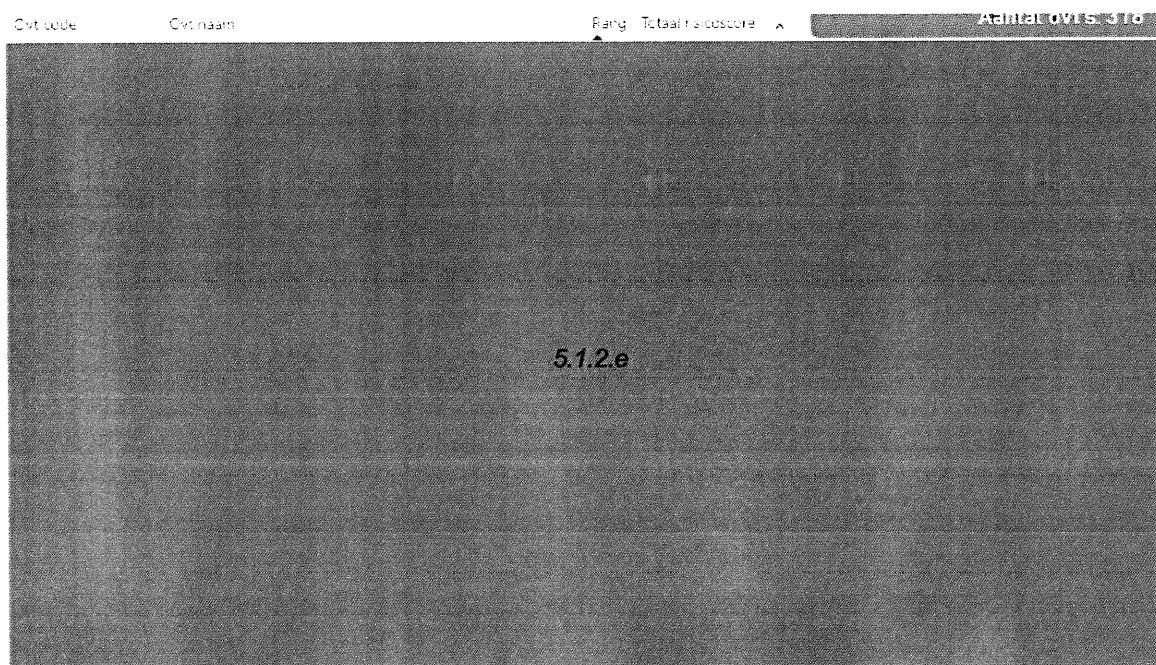
Figuur 4. Startpagina prestatie-monitor in Power BI.

Na de selectie krijgt u een totaal overzicht te zien met de risicoscores van alle besturen/objecten van toezicht, zie Figuur 5. Daarnaast is te zien op hoeveel objecten van toezicht de prestatie-monitor betrekking heeft en welke oordelen er zijn gegeven. Ook wordt de kaart van Nederland getoond met de objecten van toezicht, waarbij instellingen met hogere risicoscores rood gekleurd zijn.



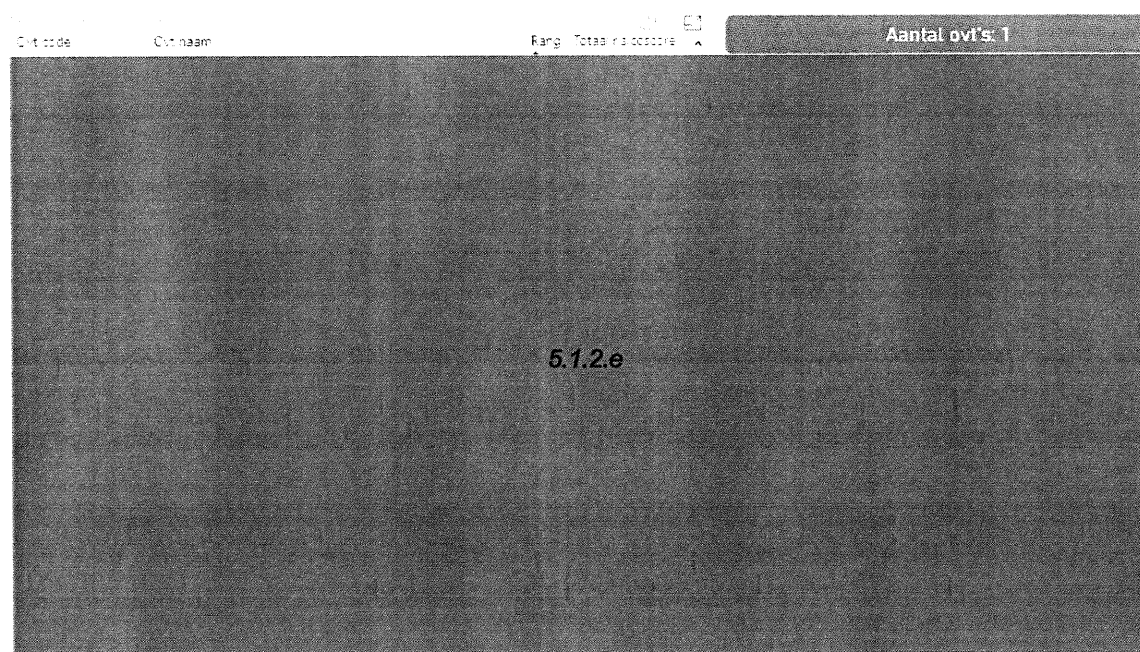
Figuur 5. Overzichtspagina Power BI met risicoscores van alle objecten van toezicht

Door in de linker tabel met de muis over een object van toezicht te gaan verschijnt er aanvullende informatie (het ovt en de deelscores op de kwaliteitsgebieden), zie Figuur 6. Wanneer u met de linker muisknop een object van toezicht aanklikt wordt op de kaart ingezoomd, en worden rechts onderaan de vigerende oordelen op bestuursniveau getoond.



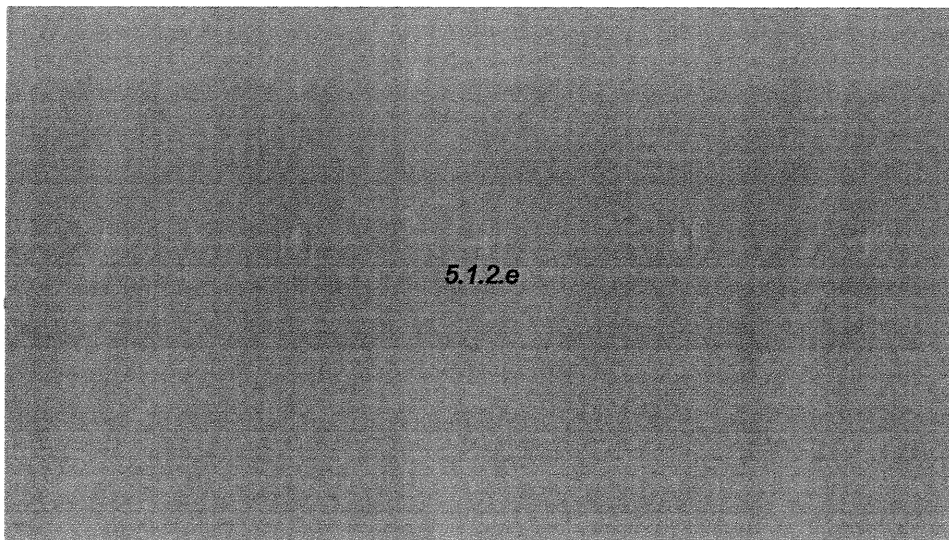
Figuur 6. Overzichtspagina Power BI met aanvullende informatie van een object van toezicht

Met de rechtermuisknop opent u de extra opties, zie Figuur 7. Specifiek de optie 'Onderzoeken' is interessant. Hiermee gaat u een niveau dieper in de data. Een veel gebruikte optie is het 'Indicatoren' overzicht. Het 'Kwaliteitsgebieden' overzicht biedt een overzicht op de vulling en risico's per kwaliteitsgebied als geheel (dit zal hier niet dieper worden besproken, maar neem vooral eens een kijkje).



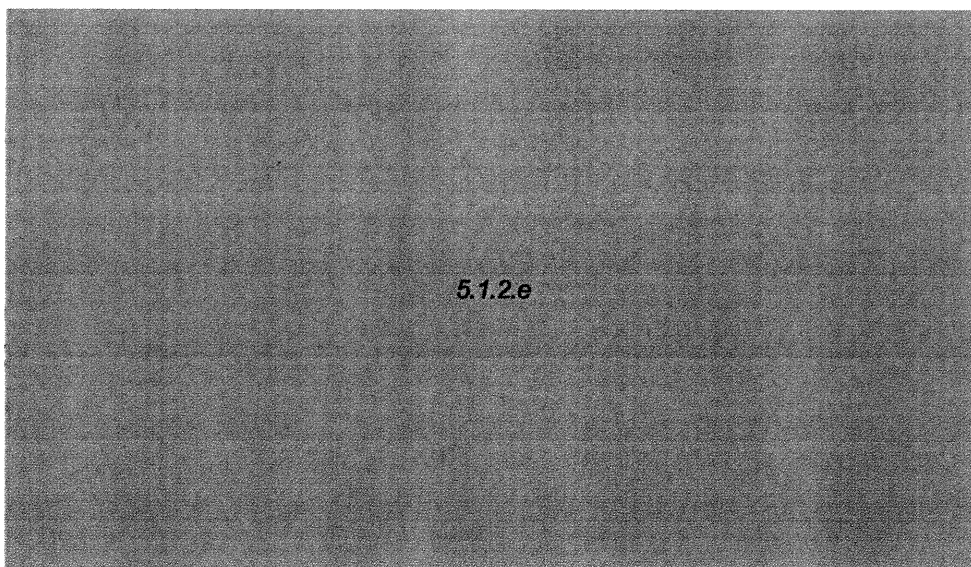
Figuur 7. Overzichtspagina Power BI met mogelijkheid om naar tabel- of grafiekoverzicht te gaan.

Het indicatorenoverzicht omvat het totaaloverzicht van een object van toezicht: alle scores op de individuele indicatoren zijn hier te zien.



Figuur 8. Indicatoren overzicht prestatie monitor in Power BI.

Door een indicator te selecteren wordt inzicht gegeven in de ontwikkeling van de waarde op de indicator. Op deze manier kan voor alle indicatoren bekeken worden hoe de waarden zich ontwikkelen.



Figuur 9. Ontwikkeling van indicatorscores zichtbaar maken

Daarnaast wordt beschikbare aanvullende informatie getoond (onderin). In de Prestatiemonitor wordt alleen extra informatie getoond die gekoppeld is aan een indicator. Er is geen extra informatie toegevoegd zoals bijv. de adresgegevens, toezichthistorie of regio-informatie. Voor dit soort informatie worden aparte informatieproducten ontwikkeld.

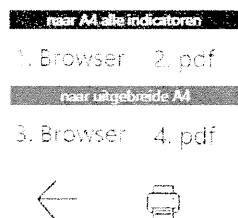
In het overzicht is rechtsonderaan de metadata te bekijken. De metadata geeft achtergrondinformatie over een indicator. Het gaat bijvoorbeeld om de berekeningswijze, weegfactor, bronnen etc. Zie Figuur 11. Toelichting bij indicator.

Indicatornummer	PM_6C_SQ_1028	
Korte naam	Aanvragen onderwijsbijdravenverhoging (art.12)	
Beschrijving/definitie	Het aantal aangevraagde onderwijsbijdravenverhogingen uitgedrukt als percentage van het aantal leerlingen.	v
Berekeningswijze	Per OKE wordt het aantal aanvragen onderwijsbijdravenverhogingen gedeeld door het aantal leerlingen (op 1 oktober) op het OKE en weergegeven als percentage.	v
Selecties	Enkel aanvragen t.a.v. WBO art. 12 worden meegenomen in de berekening (dus niet op basis van art. 36). Alleen toegewezen of afgewezen aanvragen tellen mee.	v
Signaleringswaarde	Boven grens: 10. Onder grens: -geen-	
Toelichting signaleringswaarde	De 15% scholen met de meeste aanvragen (rekeninghoudend met de schoolgrootte) lopen risico. Voor de drie jaren wordt het 88e percentiel berekend. Dit wordt gemiddeld om tot de grenswaarde te komen.	
Weegfactor indicator	2 (6,15% van het totaal)	v
Weging per jaar	2021-2022: 3x, 2020-2021: 1x, 2019-2020: 1x	

**Figuur 11. Toelichting bij indicator**

## Printen Power BI rapport

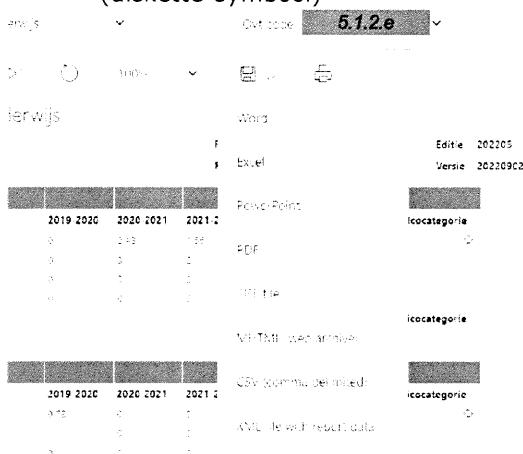
Vanuit power BI kunnen ook exports van data per instelling worden gehaald.



Figuur 12. Knoppen rechtsonder

Dit gaat als volgt:

1. Open het rapport in de webbrowser van Google Chrome. Indien het rapport al geopend is in Edge kunt u de URL (de link) kopiëren vanuit de Edge naar de webbrowser van Google Chrome.
2. Kies een van de 4 opties: Kies daarvoor of een rapport in de browser of als een pdf-bestand gewenst is. Kies ook of alleen de indicatoren ('naar A4 alle indicatoren'), of ook de aanvullende informatie gewenst is ('uitgebreide A4').  
De pdf variant biedt direct een handzame rapportage. Via de browserversie zijn er echter meer mogelijkheden, zoals het exporteren als een Excel bestand (zie stap 3)
3. (optioneel) Om te Exporteren naar een ander format: Kies bijvoorbeeld '3.Browser' (uitgebreide A4). Let op: het laden kan lang duren. Kies dan voor de exportfunctie (diskette-symbool)



Figuur 13. Exporteren vanuit browser rapport

## Hoe is de prestatie-monitor opgebouwd?

De prestatie-monitor bevat een beknopte selectie van indicatoren die gezamenlijk inzicht geven in eventuele risico's bij scholen en besturen ten aanzien van de (onderwijs)kwaliteit.

De prestatie-monitor kent verschillende varianten: afhankelijk van de onderwijssector en het type object is de inhoud verschillend. Tegelijkertijd wordt de prestatie-monitor gekenmerkt door een grote mate van sectoroverstijgende uniformiteit in de verschijningsvorm, infrastructuur en productiewijze.

Binnen de Prestatie-monitor zijn indicatoren veelal gekoppeld aan de kwaliteitsgebieden en standaarden van het onderzoekskader 2021. Deze koppeling heeft geen invloed op de totaal-risicoscore, maar helpt bij de duiding van risico's in kwaliteitsgebieden. In sommige gevallen had een indicator ook aan twee (of meer) standaarden gekoppeld kunnen worden, maar daarin is dan altijd de meest relevante gekozen. In theorie kan de prestatie-monitor risico-indicatoren bevatten voor elk kwaliteitsgebied en elke standaard, maar in de praktijk is dit nog niet mogelijk, omdat



geschikte databronnen (nog) ontbreken, met name op het gebied van bijvoorbeeld Onderwijsproces.

Naast indicatoren die gerelateerd zijn aan deugdelijkheidseisen in het onderzoekskader, kan de prestatieindicator indicatoren bevatten die betrekking hebben op de zogenaamde "overige wettelijke vereisten". Ten slotte is het denkbaar dat de prestatieindicator ook indicatoren bevat die niet direct te relateren zijn aan een bepaalde standaard of aan overige wettelijke vereisten. We noemen dit contextindicatoren.

### *Uitleg berekening risicoscore Prestatiemonitor*

Om de totale risicoscore van een object van toezicht te berekenen is een algemeen rekenmodel ontwikkeld (toegepast op alle Prestatiemonitors). Het rekenmodel is opgebouwd uit een aantal stappen.

#### *Berekenen trendwaarde*

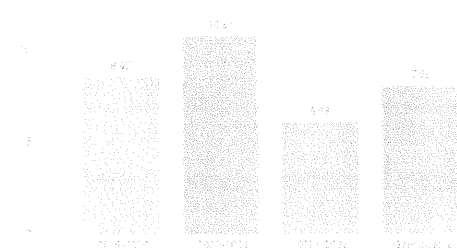
Allereerst wordt per indicator een trendwaarde berekend. Deze trendwaarde is het (gewogen) gemiddelde over verschillende jaren. Een indicator kan een waarde op één of meerdere jaren hebben. Het rekenmodel biedt de mogelijkheid om de score van bepaalde jaren zwaarder mee te laten wegen. In Tabel 3 en Figuur 15 hieronder is te zien hoe een weging over jaren werkt.

Tabel 3. Voorbeeld weging tussen jaren

Jaar	Aantal jaar geleden	Weging jaren	Proportionele weging
2022-2023	1	3	$3/6 = 0,5$
2021-2022	2	2	$2/6 = 0,33$
2020-2021	3	1	$1/6 = 0,17$

In Figuur 15 is gebruik gemaakt van de weging tussen jaren zoals te zien in Tabel 3. Aan de wegingen is te zien dat de meer recente gegevens zwaarder meetellen. De trendwaarde in het voorbeeld is als volgt te berekenen:

#### **Indicator**



Figuur 15. Voorbeeld rekenmodel prestatieindicator

Als alle jaren even belangrijk zijn is het gewicht van alle jaren gelijk. Zonder weging zou de trendwaarde in het voorbeeld uit Figuur 16 op een andere waarde uitkomen, namelijk

#### *Trendwaarde ten opzichte van norm/grenswaarde*

De trendwaarde wordt vervolgens vergeleken met een norm/grenswaarde. De onder en/of bovengrens wordt gezamenlijk bepaald door Kennis en de toezichtsector: dit kan bijvoorbeeld een vaste waarde zijn of een relatieve waarde zoals een percentielscore. De vergelijking van de trendwaarde gebeurt op basis van 'gelijk aan' of 'groter dan' de grenswaarde/norm. Als de trendwaarde voldoet aan de norm krijgt deze de waarde 0 (= groen), voldoet de trendwaarde niet dan krijgt deze de waarde 1 (= rood).

In het voorbeeld van Figuur 15 (uit de Prestatiemonitor PO-BO) is de grenswaarde vastgesteld op een gemiddeld verzuim van 9%. De gemiddelde trendwaarde blijft daar onder. De school krijgt daarom op deze indicator géén risico's toegekend.



### Berekenen totale risicoscore

Vervolgens worden de scores op alle indicatoren, met behulp van een weegfactor per indicator, omgezet naar een totale risicoscore. Zie Tabel 4 voor een voorbeeld. Door een gewicht toe te kennen aan een indicator is het mogelijk om bepaalde indicatoren zwaarder mee te laten wegen. Als alle indicatoren even belangrijk zijn is het gewicht van alle indicatoren gelijk. De totale risicoscore is een score tussen 0 en 1. Dit is een relatieve score die bedoeld is om een ordening van objecten van toezicht te kunnen maken. Deze score heeft geen verdere interpretatie.

### Berekenen subtotaal kwaliteitsgebied

Het subtotaal per kwaliteitsgebied wordt op een vergelijkbare manier berekend als de totale risicoscore, zie Tabel 4. Het enige verschil is dat voor deze berekening de indicatoren binnen één kwaliteitsgebied worden meegenomen.

Tabel 4. Voorbeeld berekening subtotaal kwaliteitsgebied en totale risicoscore

Indicatoren bij kwaliteitsgebied	Score	Weging	Proportionele weging	Berekening score kwaliteitsgebied	(sub)totalen
Indicator1	0	5	$5/10 = 0,5$		
Indicator2	1	3	$3/10 = 0,3$		
Indicator3	0	2	$2/10 = 0,2$		
Kwaliteitsgebied A	1	10	$10/10 = 1$		
Indicator 4	1	2	$2/2 = 1$		
Kwaliteitsgebied B	1	2	$1/1 = 1$		
Totale risicoscore	2	12	$12/12 = 1$		

### Verantwoord gebruik van risicoscores

De prestatie-monitor bevat een selectie van indicatoren die analisten en/of inspecteurs kunnen ondersteunen bij hun beeldvorming over mogelijke risico's bij besturen en scholen. Het is echter belangrijk om daarbij te benoemen dat deze gegevens niet *alle* risico's in kaart kunnen brengen: omdat er simpelweg niet over alle risico's data beschikbaar is en omdat data onbetrouwbaar kan zijn.

Wij bevelen inspecteurs en analisten van harte aan om de informatie die beschikbaar is gesteld in de prestatie-monitor te gebruiken bij de beeldvorming omtrent risico's. Tegelijkertijd roepen we ook op om ten alle tijden kritisch naar de gegevens te blijven kijken en om ook andere bronnen (met meer kwalitatieve, of 'zachte' informatie) te raadplegen om tot een meer compleet beeld van een object van toezicht te komen.

### Actualiteit van de gegevens in de prestatie-monitor

Vooralsnog wordt er voor verschillende sectoren één keer per jaar een nieuwe prestatie-monitor opgeleverd. Voor besturen wordt twee keer per jaar een prestatie-monitor opgeleverd. Voor besturen staat oplevering gepland in april en september. Bij oplevering van een prestatie-monitor staan de gegevens op dat moment vast, de gegevens worden in de tussentijd niet (automatisch) verversd. Het is belangrijk te beseffen dat de Prestatie-monitor daarin verschilt van het informatieproduct 'de Verdiepingsinformatie'.

We spreken van een nieuwe editie als er een nieuwe prestatiemonitor wordt opgeleverd. We spreken van een nieuwe versie als er binnen een editie verbeteringen of aanpassingen worden gedaan.

### ***Beheer: eigenaarschap en ontwikkeling***

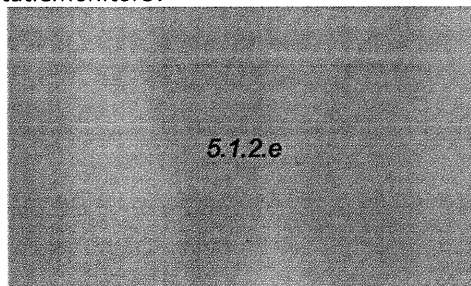
Elke prestatiemonitor wordt gezamenlijk ontwikkeld door een toezichtsector en Kennis. Kennis houdt regie over het onderhoud en de ontwikkeling van alle prestatiemonitors. De samenhang tussen de Prestatiemonitors en andere informatieproducten wordt bepaald vanuit het programma Toezicht Informatie Producten (TIP).

### ***Feedback: vragen en opmerkingen over de prestatiemonitor***

Voor vragen en opmerkingen kunt u terecht bij **5.1.2.e** (@owinsp.nl) of bij de sectorvertegenwoordiger in de inspectiebrede klankbordgroep prestatiemonitor.

#### ***Sectorleden werkgroepen Prestatiemonitors:***

RS  
Samenwerkingsverbanden  
PO BO  
PO SBO  
PO nieuwkomers  
SO en VSO  
VO  
MBO  
HO  
KO/VVE





*Bijlage I Kwaliteitsgebieden en standaarden uit het onderzoekskader 2021, per prestatiemonitor*

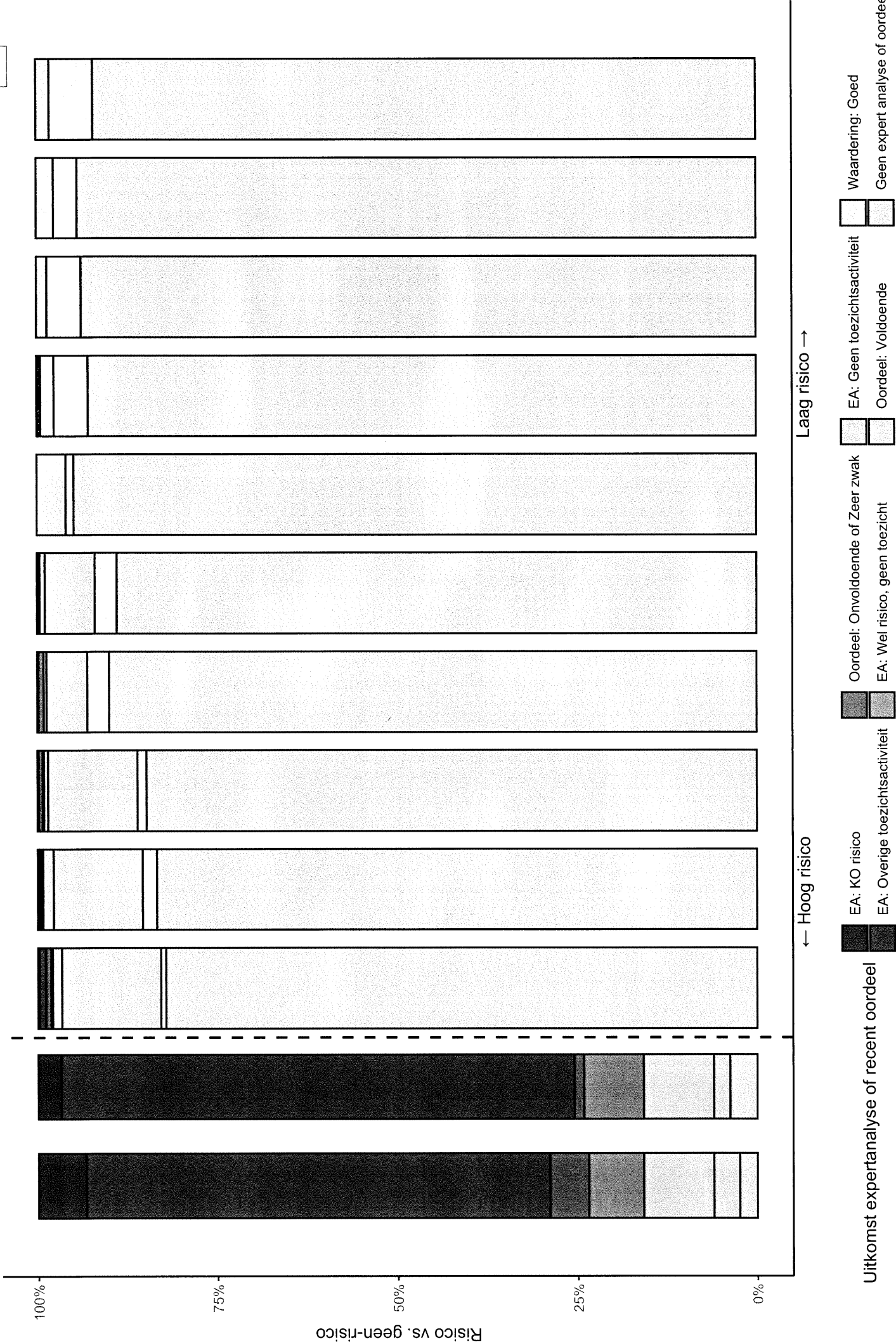
**Vernieuwd kader Besturen scholen en samenwerkingsverbanden**

KWALITEITSGEBIEDEN EN STANDAARDEN		BST SWV	BST SCHOLEN
<b>RPO REALISATIE PASSEND ONDERWIJS</b>			
RPO1	Dekkend netwerk van voorzieningen	•	
RPO2	Regionale samenwerking	•	
RPO3	Advies en toewijzing extra ondersteuning	•	
<b>BAK BESTURING, KWALITEITSZORG EN AMBITIE</b>			
BAK1	Visie, ambities en doelen	•	•
BAK2	Uitvoering en kwaliteitscultuur	•	•
BAK3	Evaluatie, verantwoording en dialoog	•	•

**Vernieuwd kader Scholen/objekten van toezicht**

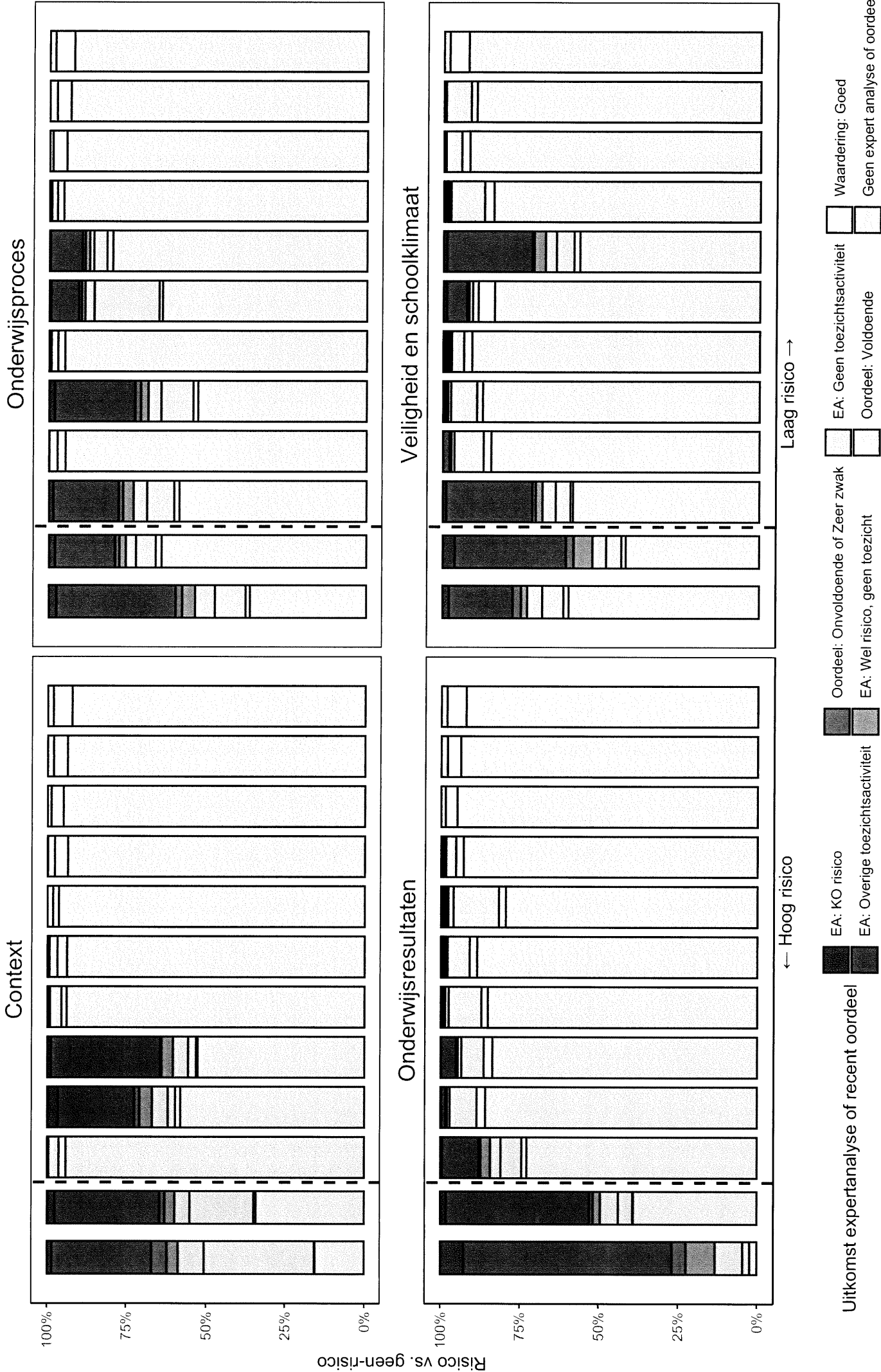
KWALITEITSGEBIEDEN EN STANDAARDEN		PO			(V)SO	VO	MBO	
		BO	SBO	NWK	(V)SO	BEK	PRO	BEK
<b>OP ONDERWIJSPROCES</b>								
OP1	Aanbod	•	•	•	•	•	•	•
OP2	Zicht op ontwikkeling en begeleiding, <i>Ontwikkeling en begeleiding(MBO)</i>	•	•	•	•	•	•	•
OP3	Pedagogisch-didactisch handelen	•	•	•	•	•	•	•
OP4	Onderwijstijd	•	•	•	•	•	•	
OP5	Praktijkvorming/stage <i>Beroepspraktijkvorming(MBO)</i>				•		•	•
OP6	Afsluiting	•	•	•	•	•		•
<b>VS VEILIGHEID EN SCHOOLKLIMAAT</b>								
VS1	Veiligheid	•	•	•	•	•	•	•
VS2	Schoolklimaat	•	•	•	•	•	•	•
<b>BD DIPLOMERING</b>								
BD1	Borging diplomering							•
<b>OR ONDERWIJSRESULTATEN</b>								
OR1	Resultaten <i>Studiesucces(MBO)</i>	•	•	•	•	•	•	•
OR2	Sociale en maatschappelijke competenties	•	•	•	•	•	•	
OR3	Vervolgsucces							•
<b>SKA STUREN, KWALITEITSZORG EN AMBITIE</b>								
SKA1	Visie, doelen en ambities	•	•	•	•	•	•	•
SKA2	Sturing op uitvoering en kwaliteitscultuur	•	•	•	•	•	•	•
SKA3	Evaluatie, verantwoording en dialoog	•	•	•	•	•	•	•

PM BO 2021 - PM rang en uitkomst EA of oordeel

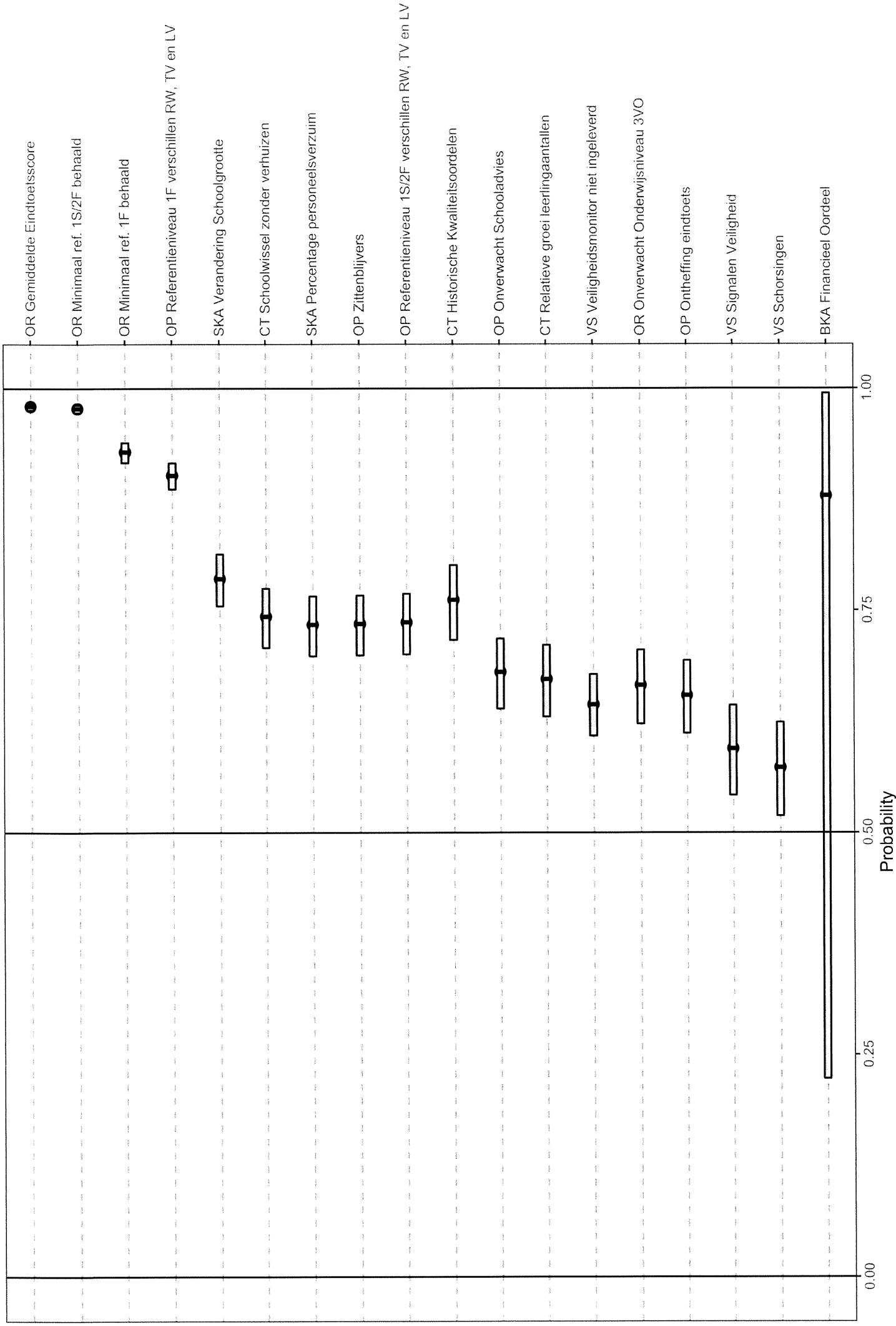


# PM BO 2021 - PM rang en uitkomst EA of oordeel

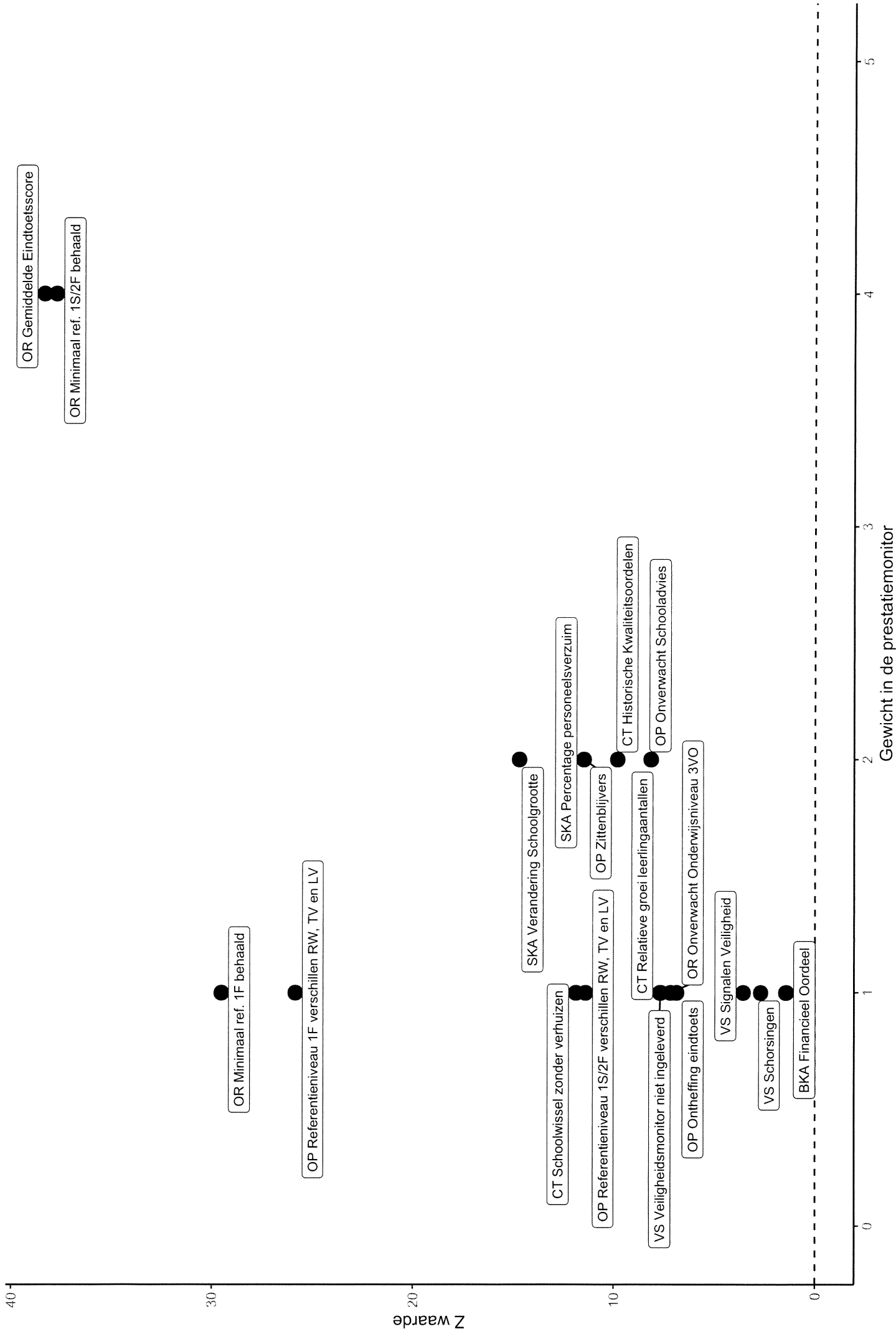
Per kwaliteitsgebied



# PM BO 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



# PM BO 2021 - Relatie voorspelkracht en gewicht

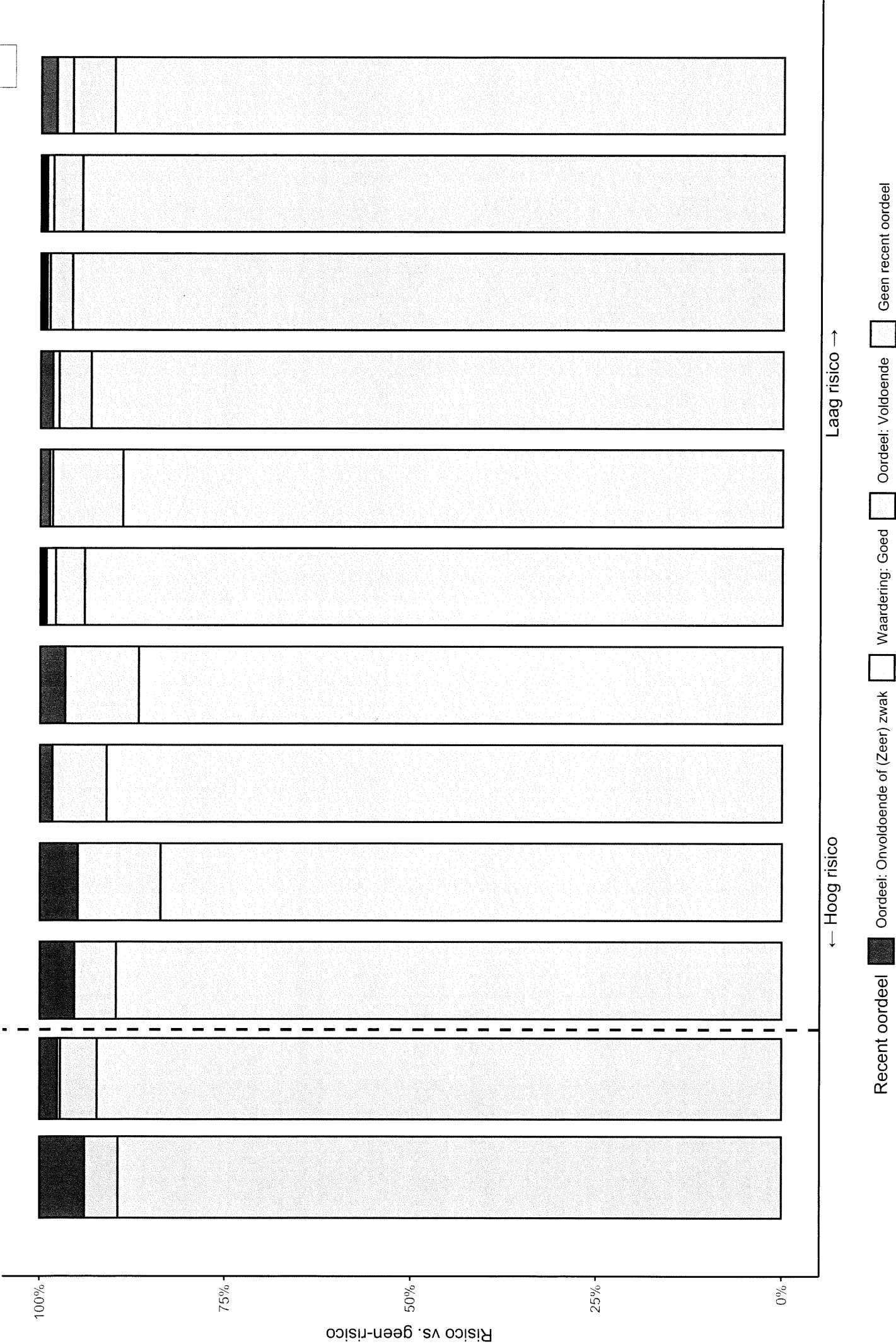


Prestatiemonitor BO 2021 - Gevoeligheid

Gevoeligheid	Context	Onderwijsproces	Onderwijsresultaten	Veiligheid en schoolklimaat	Totaal
Sensitiviteit	0.47	0.40	0.78	0.42	0.96
Specifiteit	0.91	0.90	0.95	0.90	0.98
Precisie	0.41	0.35	0.69	0.37	0.84

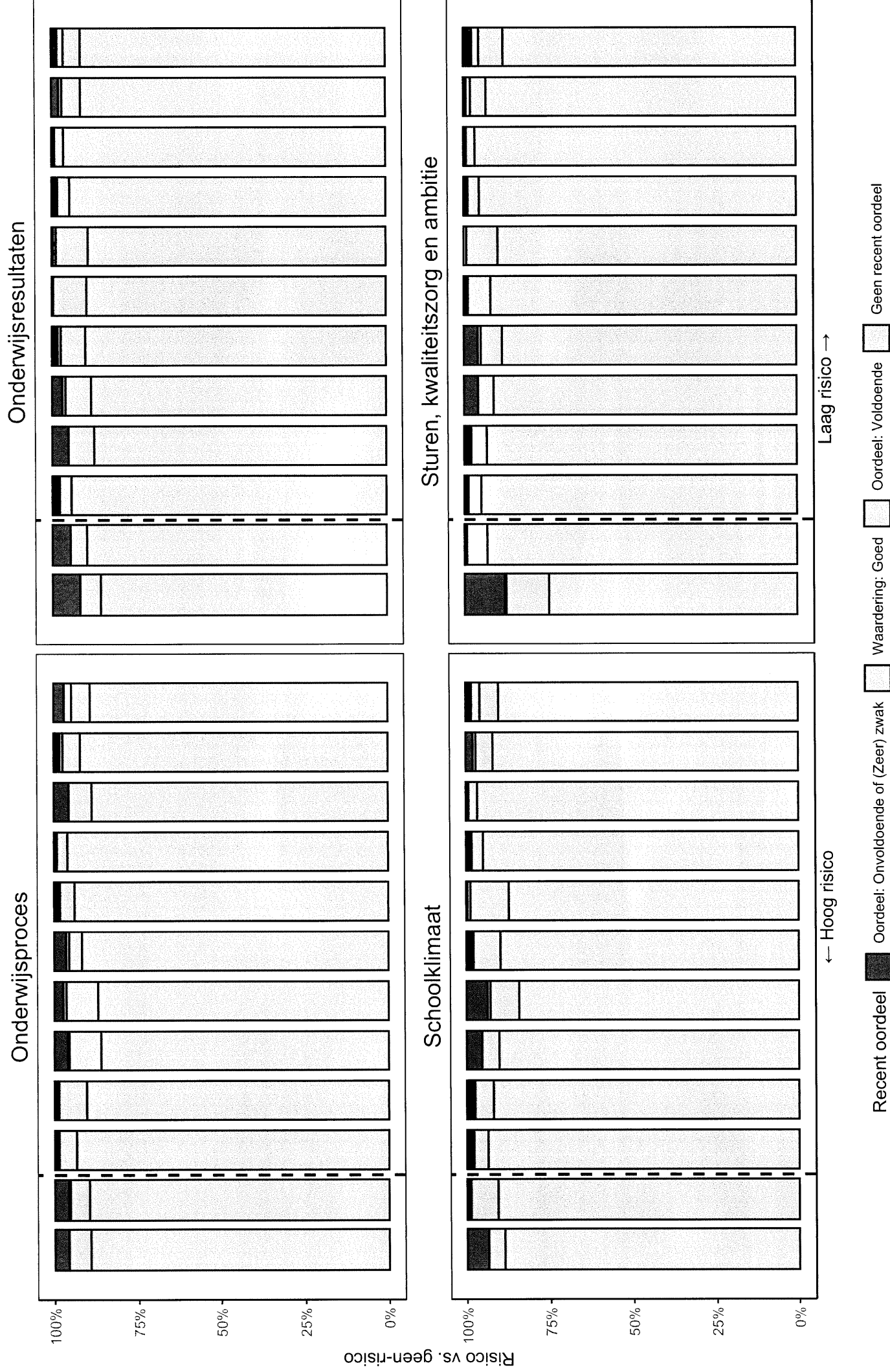
PM MBO 2021 - PM rang en oordeel

12.



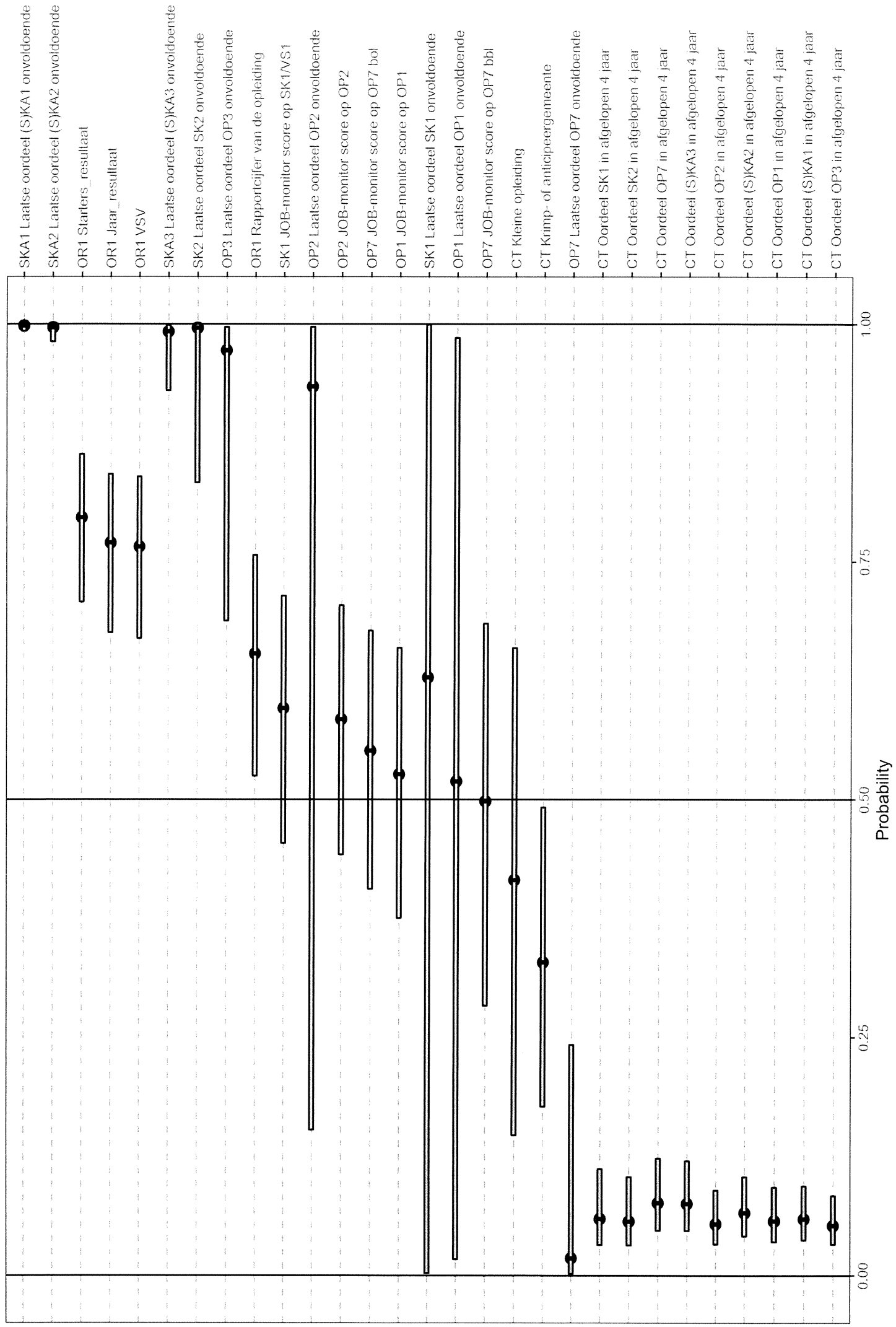
# PM MBO 2021 - PM rang en oordeel

Per kwaliteitsgebied

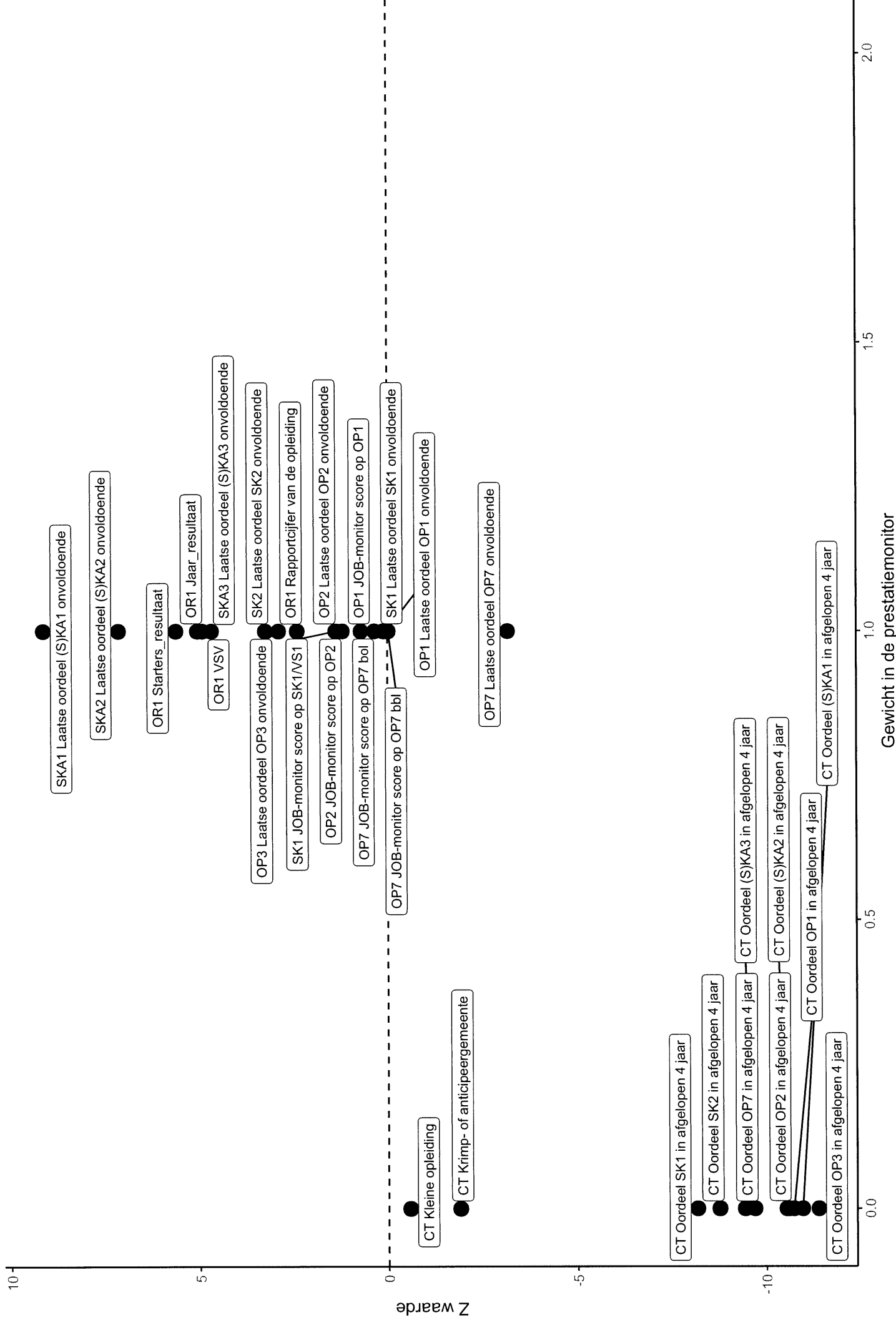




# PM MBO 2021 - Voorspelkracht indicatoren oordeel



# PM MBO 2021 - Relatie voorspelkracht en gewicht

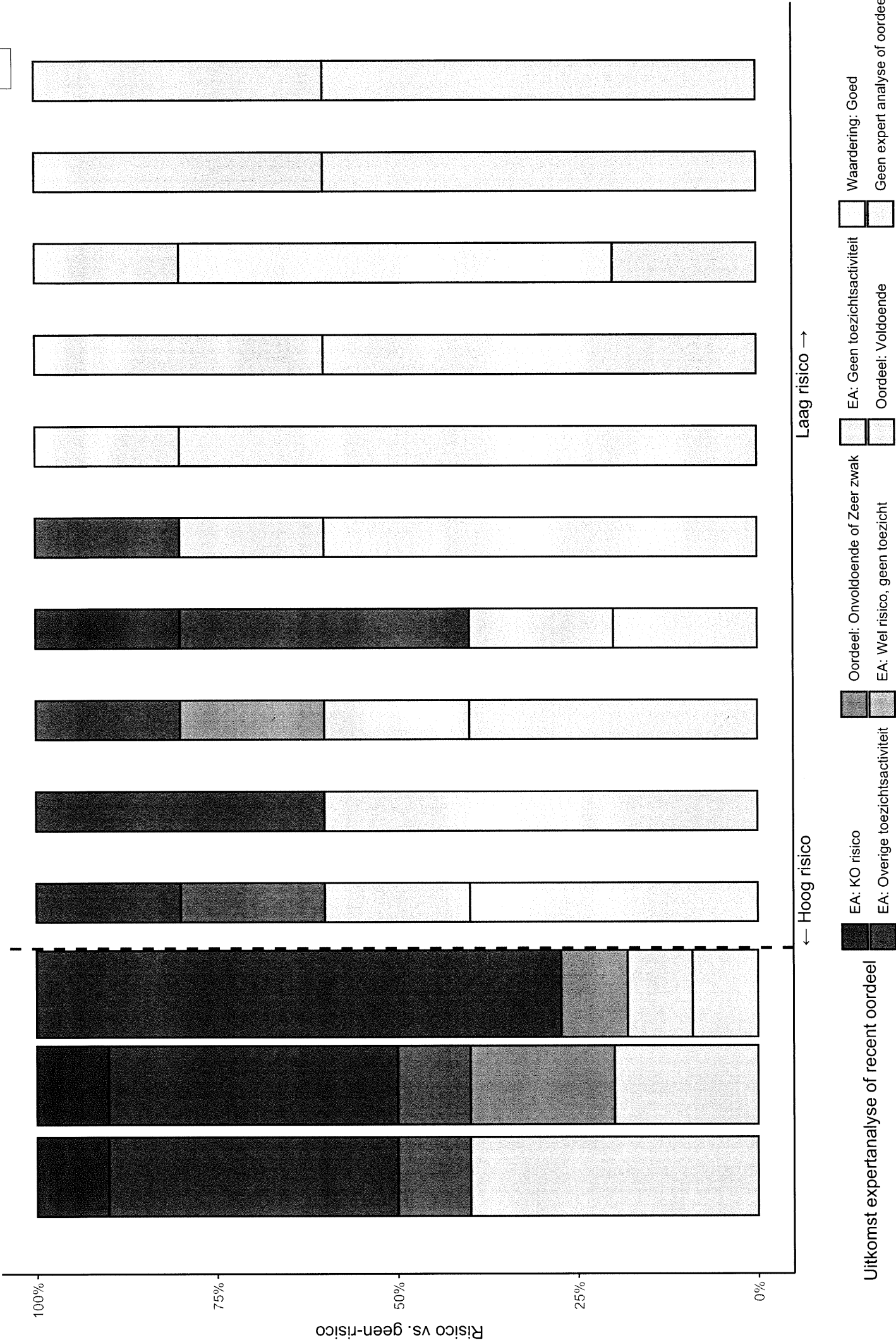


Prestatiemonitor MBO 2021 - Gevoeligheid

<b>Gevoeligheid</b>	<b>Sturen Kwaliteitszorg en ambitie</b>	<b>Onderwijsproces</b>	<b>Onderwijsresultaten</b>	<b>Schoolklimaat</b>	<b>Totaal</b>
Sensitiviteit	0.43	0.27	0.45	0.24	0.28
Specificiteit	0.83	0.83	0.83	0.83	0.83
Precisie	0.07	0.04	0.07	0.04	0.04

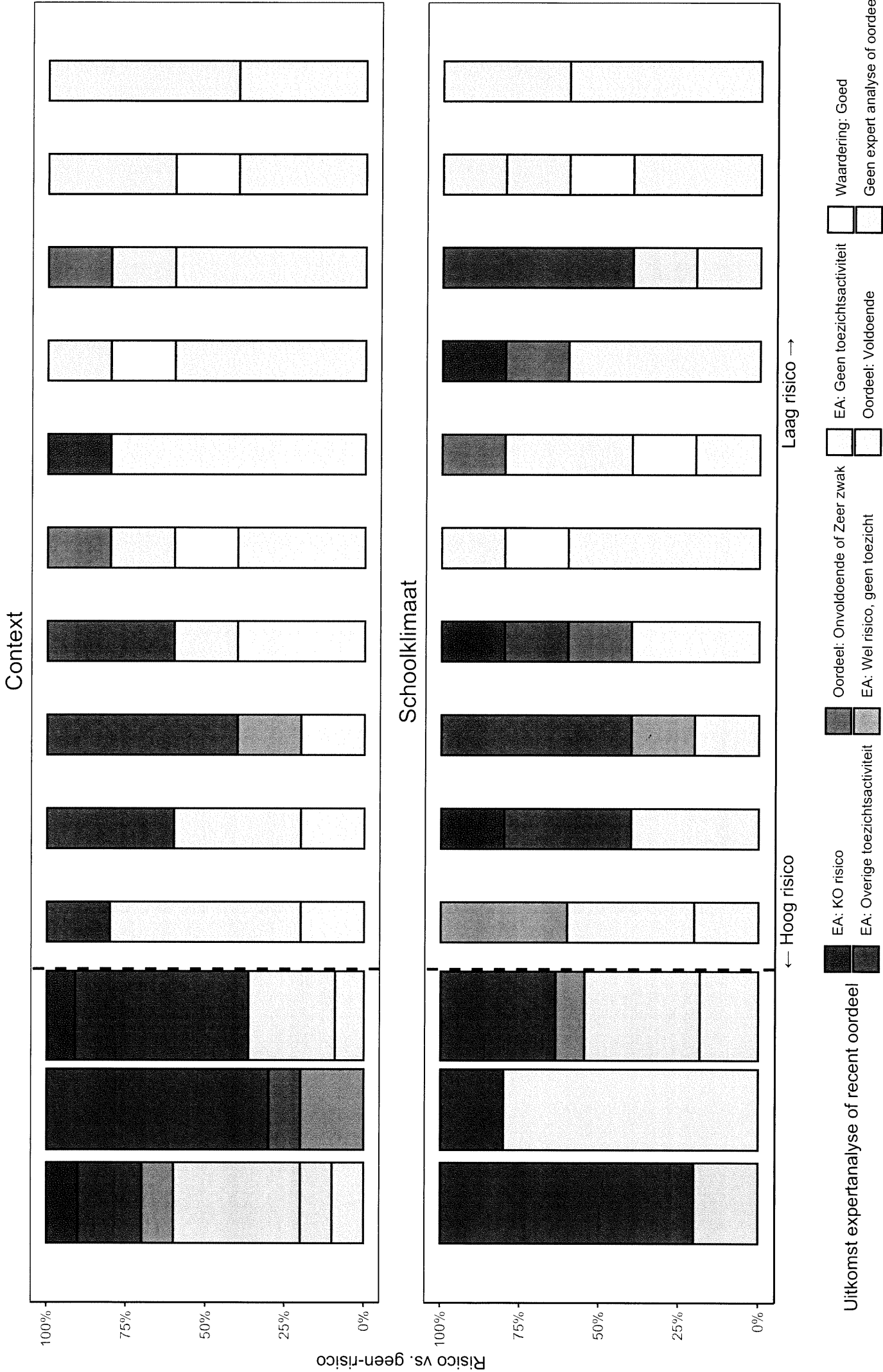
PM NK 2021 - PM rang en uitkomst EA of oordeel

13.

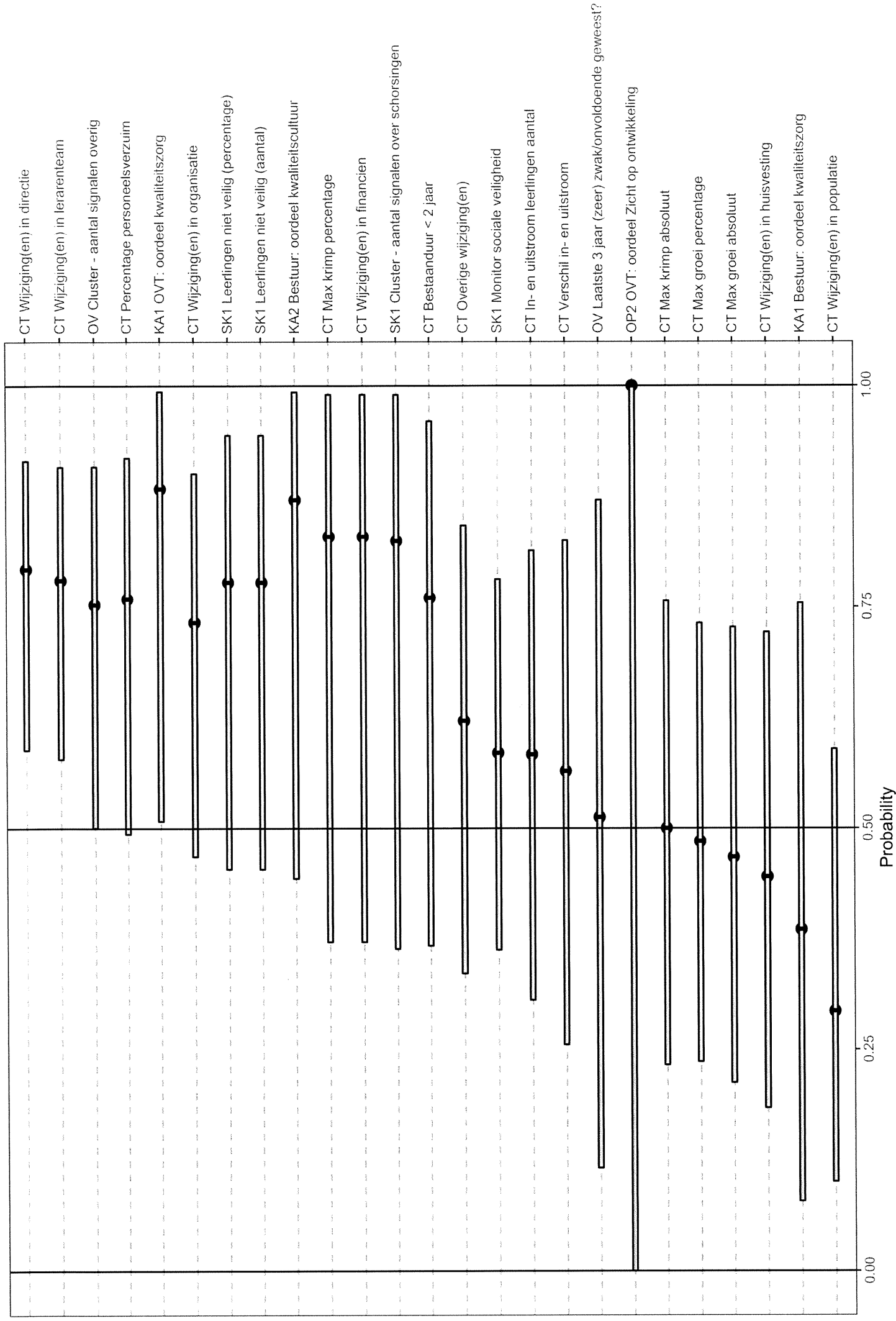


# PM NK 2021 - PM rang en uitkomst EA of oordeel

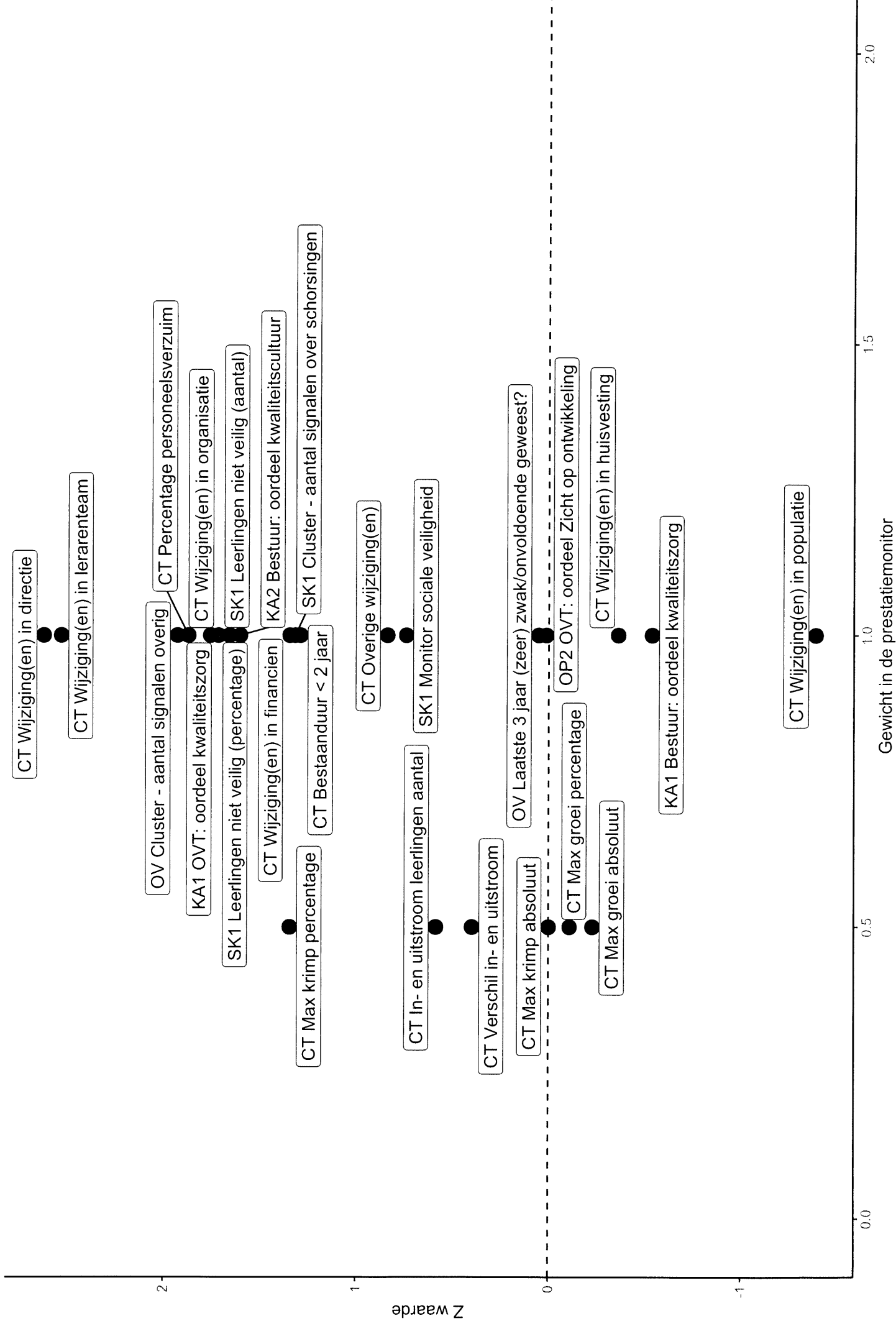
Per kwaliteitsgebied



# PM NK 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



# PM NK 2021 - Relatie voorspelkracht en gewicht

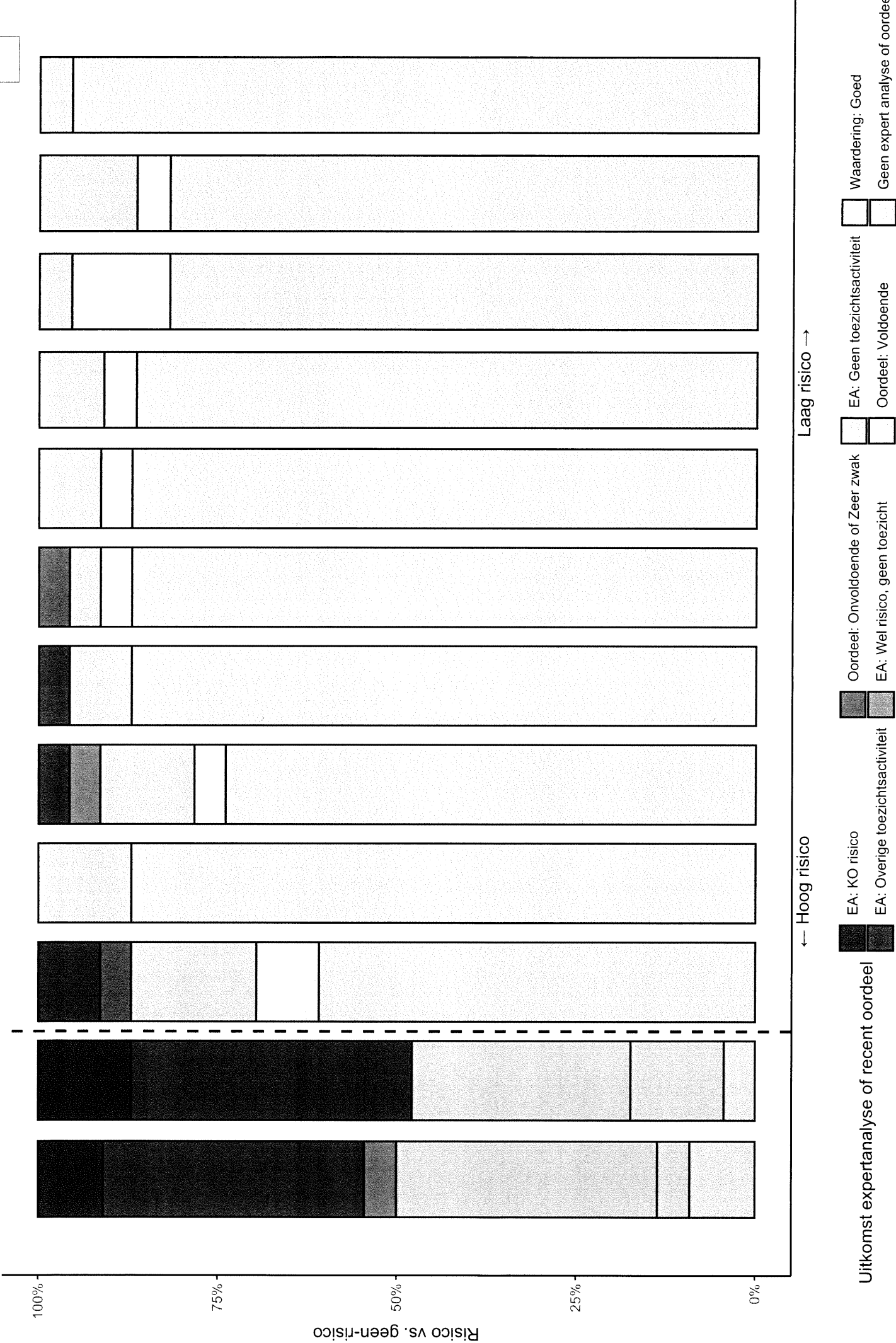


Prestatiemonitor NK 2021 - Gevoeligheid

Gevoeligheid	Context	Schoolklimaat	Totaal
Sensitiviteit	0.64	0.45	0.70
Specifiteit	0.79	0.67	0.83
Precisie	0.68	0.48	0.74



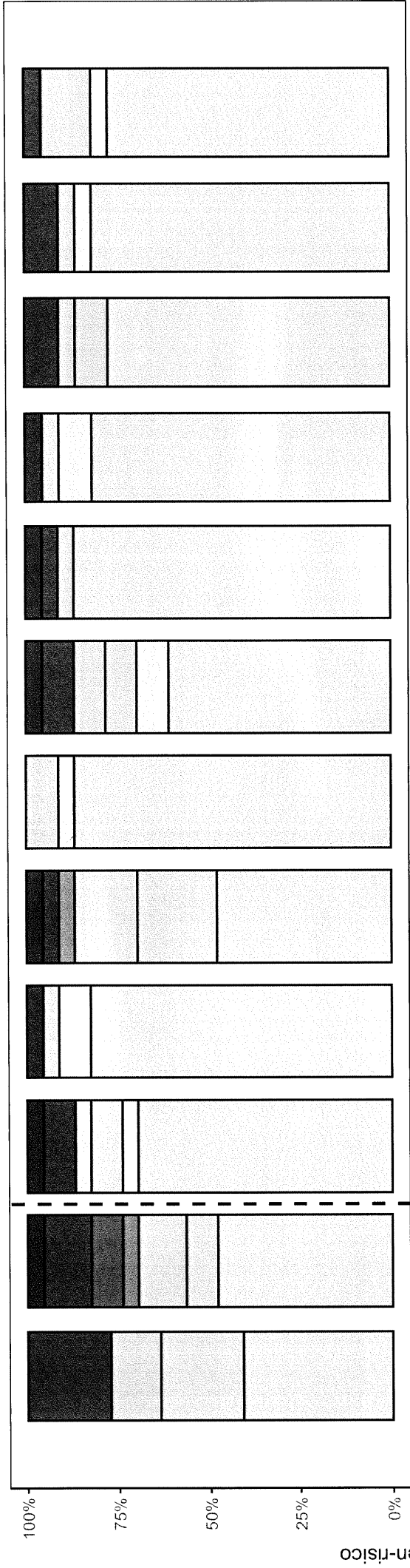
PM SBO 2021 - PM rang en uitkomst EA of oordeel



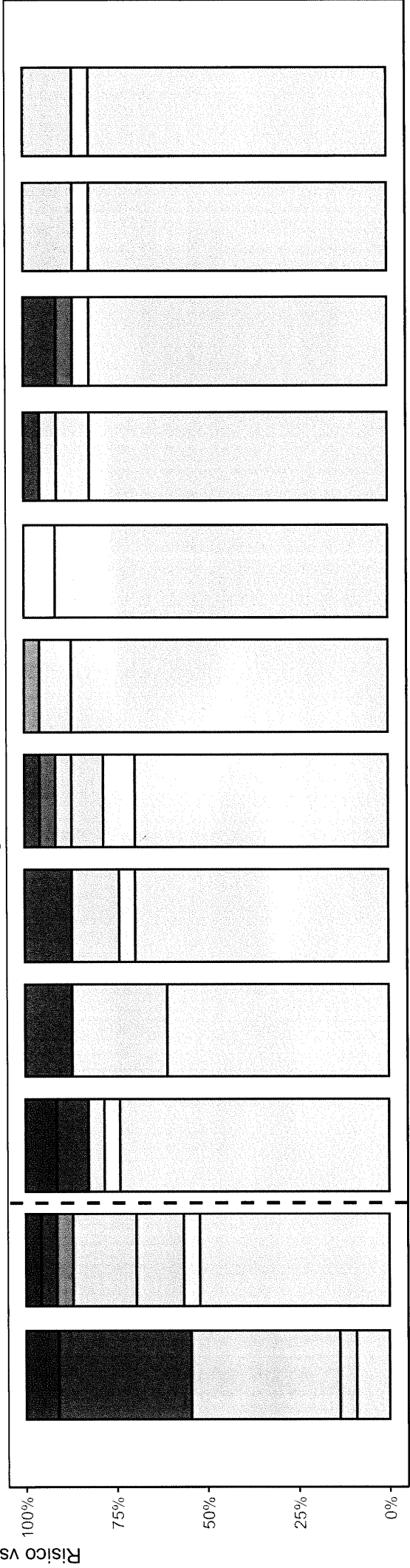
# PM SBO 2021 - PM rang en uitkomst EA of oordeel

Per kwaliteitsgebied

## Context



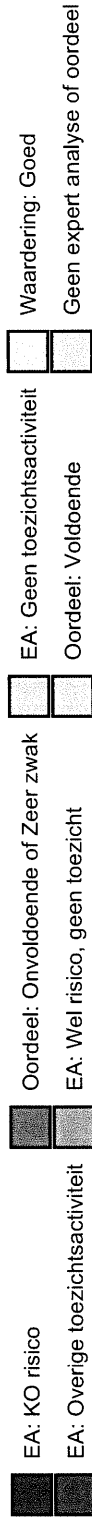
## Veiligheid en schoolklimaat



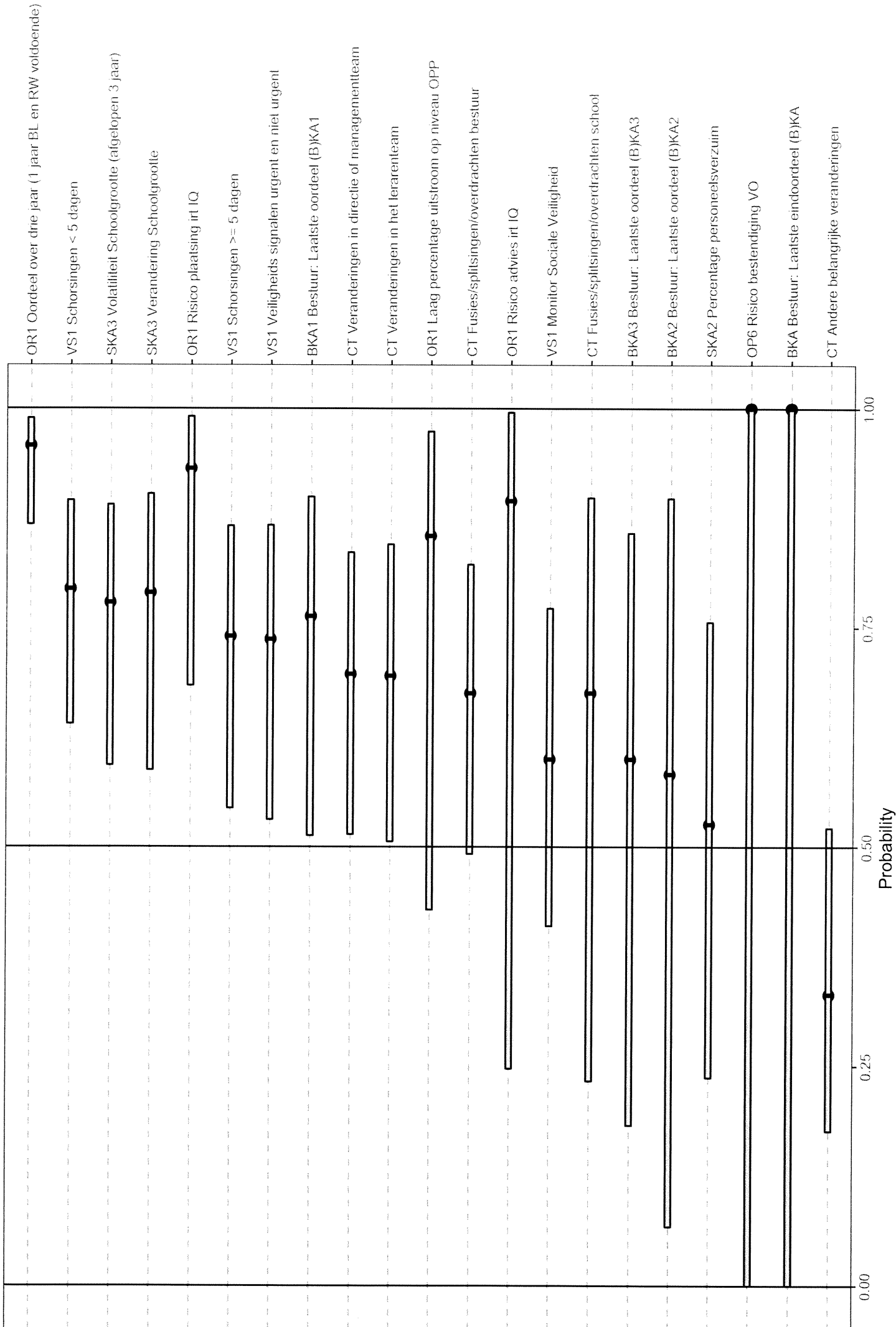
← Hoog risico

Laag risico →

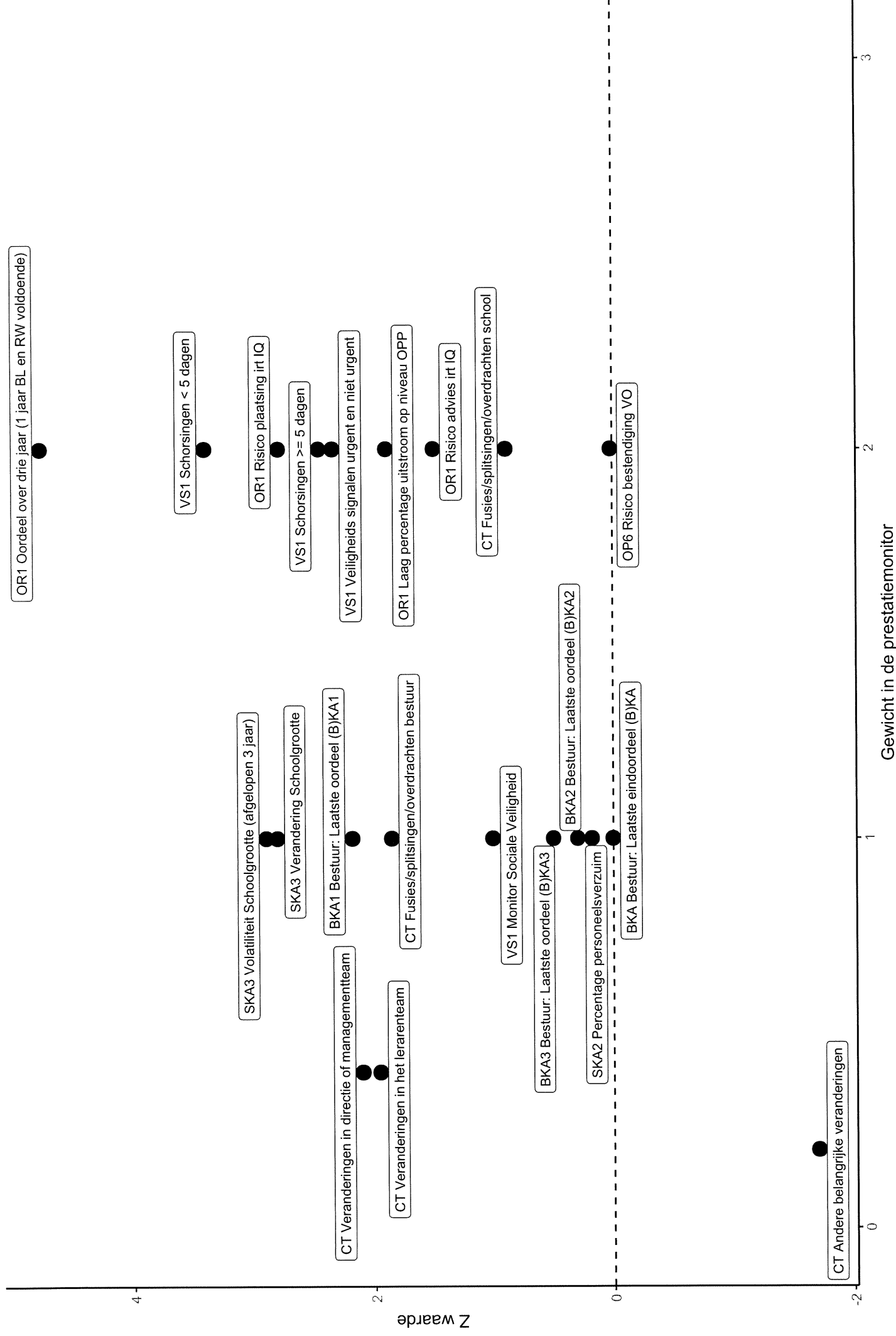
Uitkomst expertanalyse of recent oordeel



# PM SBO 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



# PM SBO 2021 - Relatie voorspelkracht en gewicht

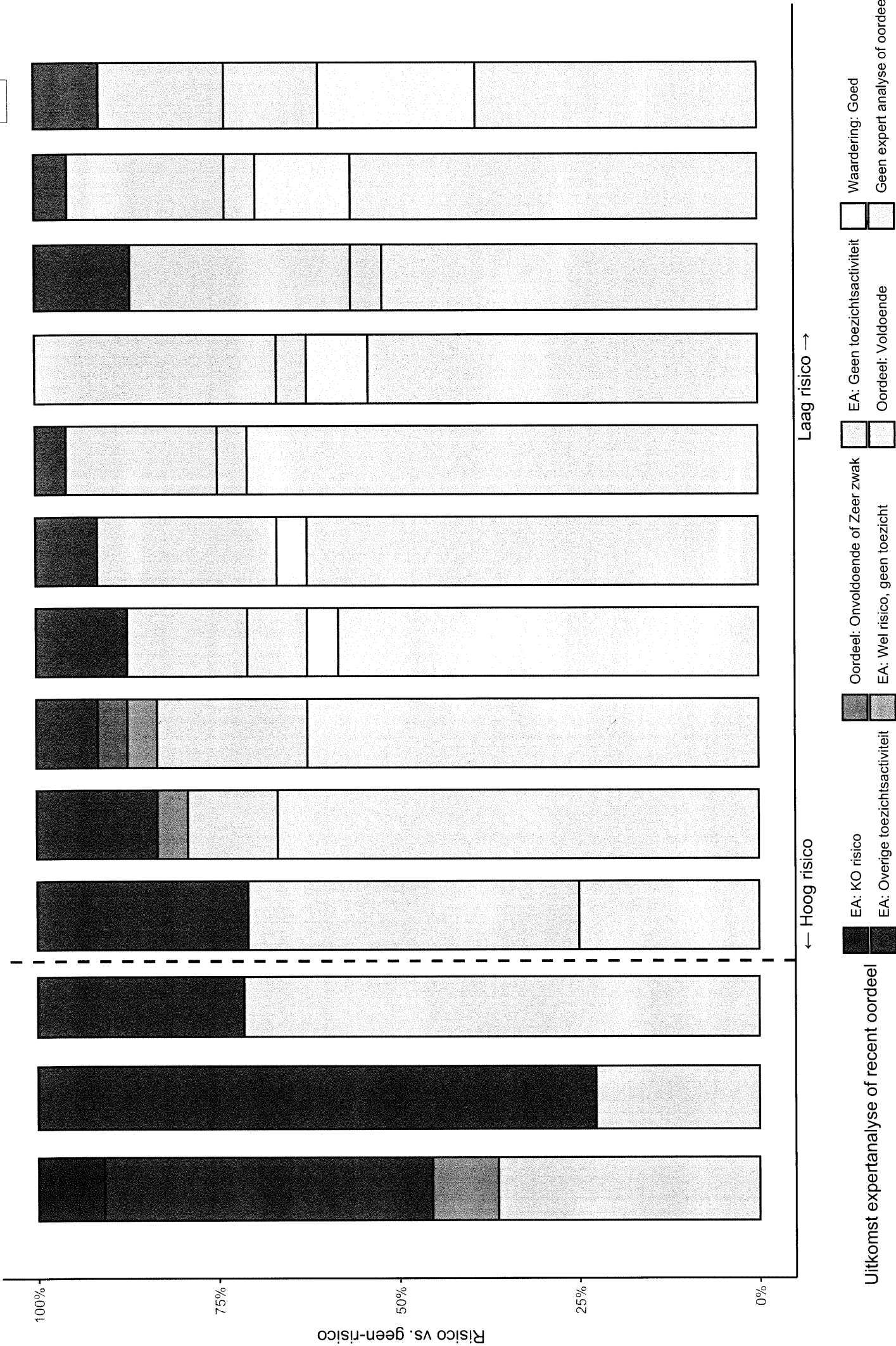


Prestatiemonitor SBO 2021 - Gevoeligheid

Gevoeligheid	Context	Veiligheid en schoolklimaat	Totaal
Sensitiviteit	0.40	0.43	0.77
Specifiteit	0.86	0.87	0.91
Precisie	0.27	0.29	0.51

PM SO 2021 - PM rang en uitkomst EA of oordeel

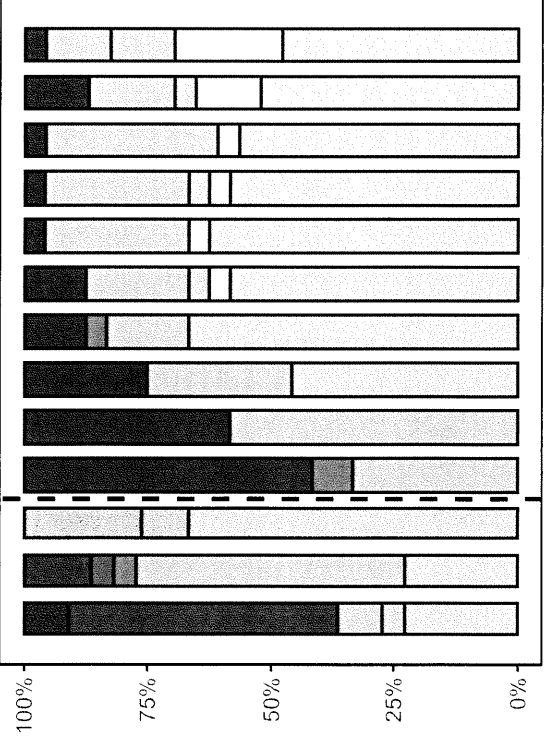
15.



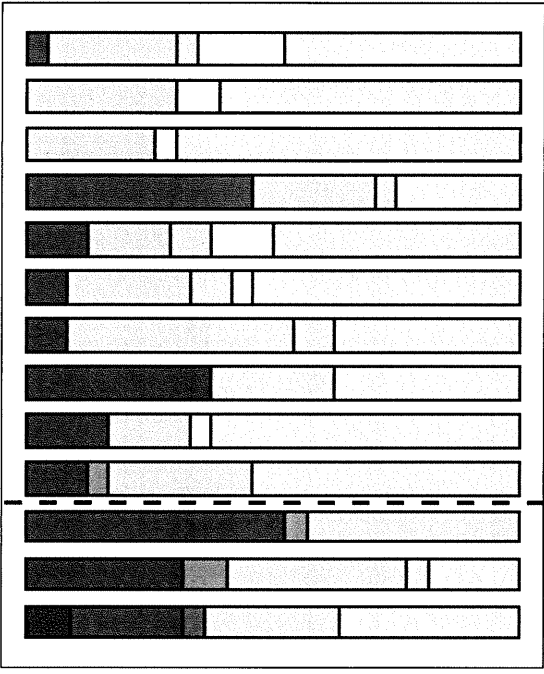
# PM SO 2021 - PM rang en uitkomst EA of oordeel

Per kwaliteitsgebied

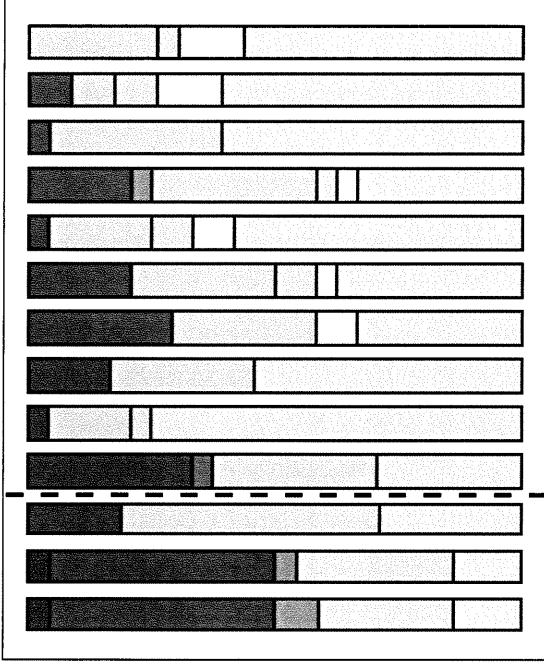
## Besturing, kwaliteitszorg en ambitie



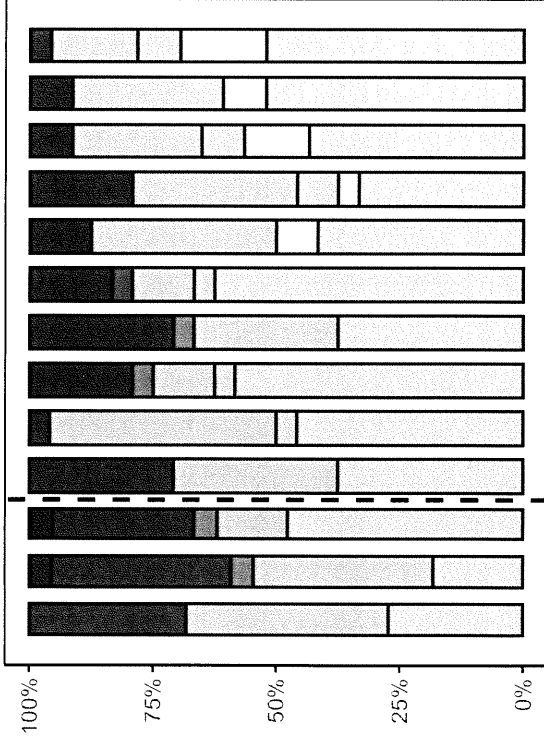
## Context



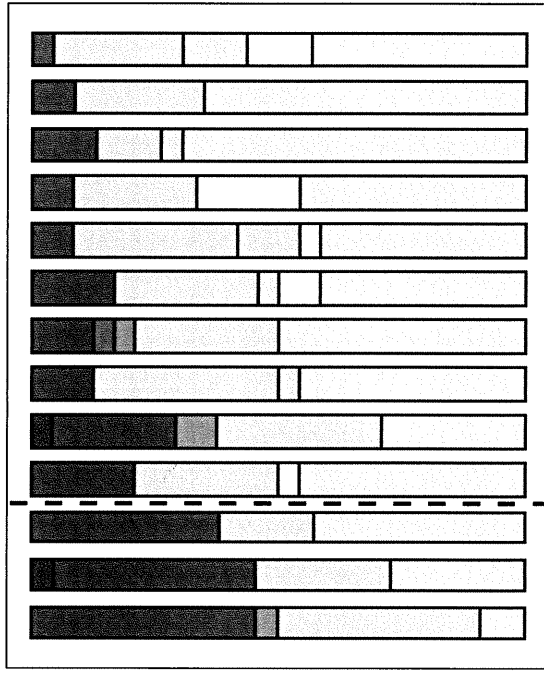
## Onderwijsresultaten



## Sturen, kwaliteitszorg en ambitie



## Veiligheid en schoolklimaat



← Hoog risico

→ Laag risico

Uitkomst expertanalyse of recent oordeel

EA: KO risico  
EA: Overige toezichtsactiviteit

Oordeel: Onvoldoende of Zeer zwak  
Oordeel: Wel risico, geen toezicht

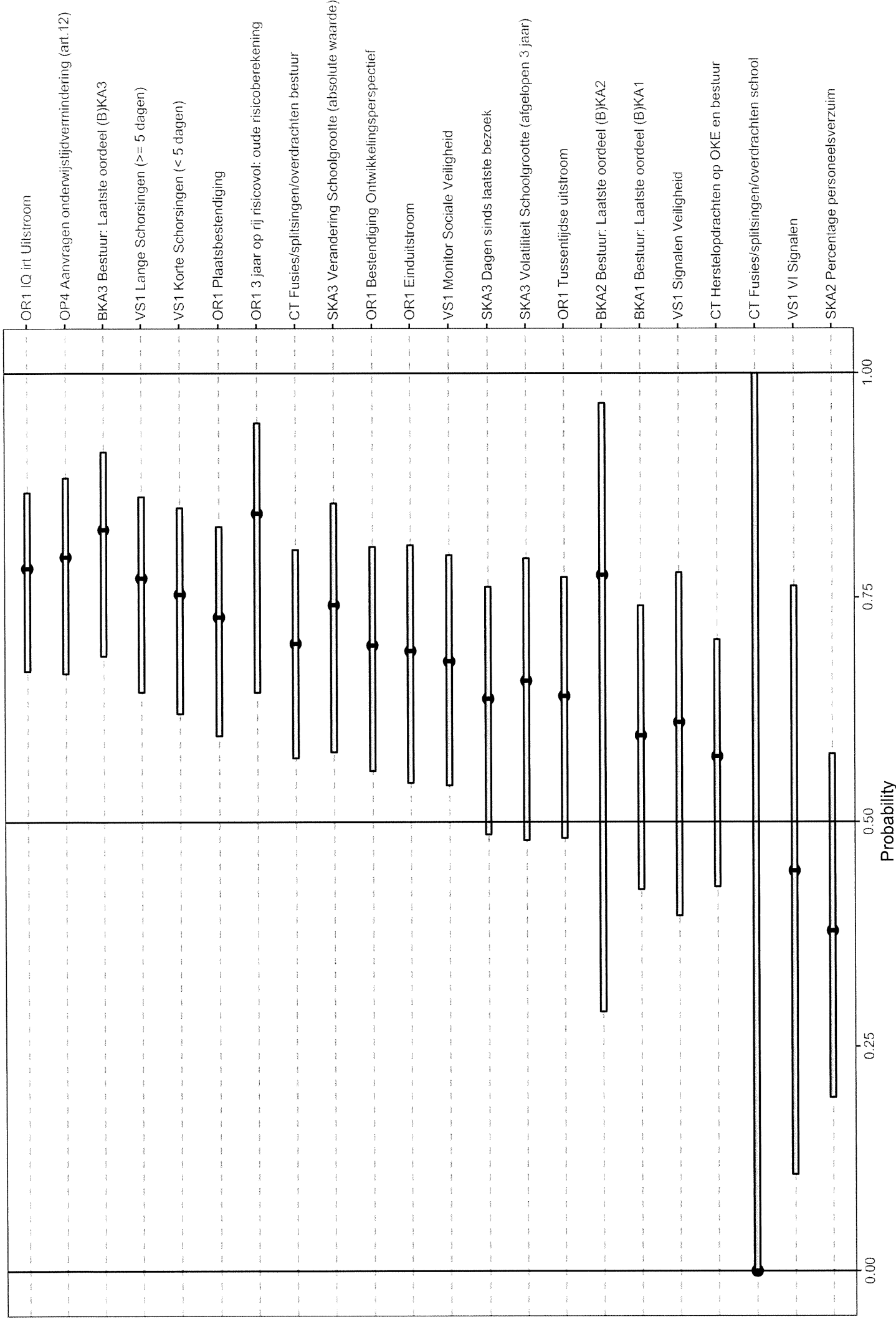
Oordeel: Voldoende  
Waardering: Goed

EA: Geen toezichtsactiviteit  
Oordeel: Voldoende

EA: Geen toezichtsactiviteit  
Oordeel: Voldoende

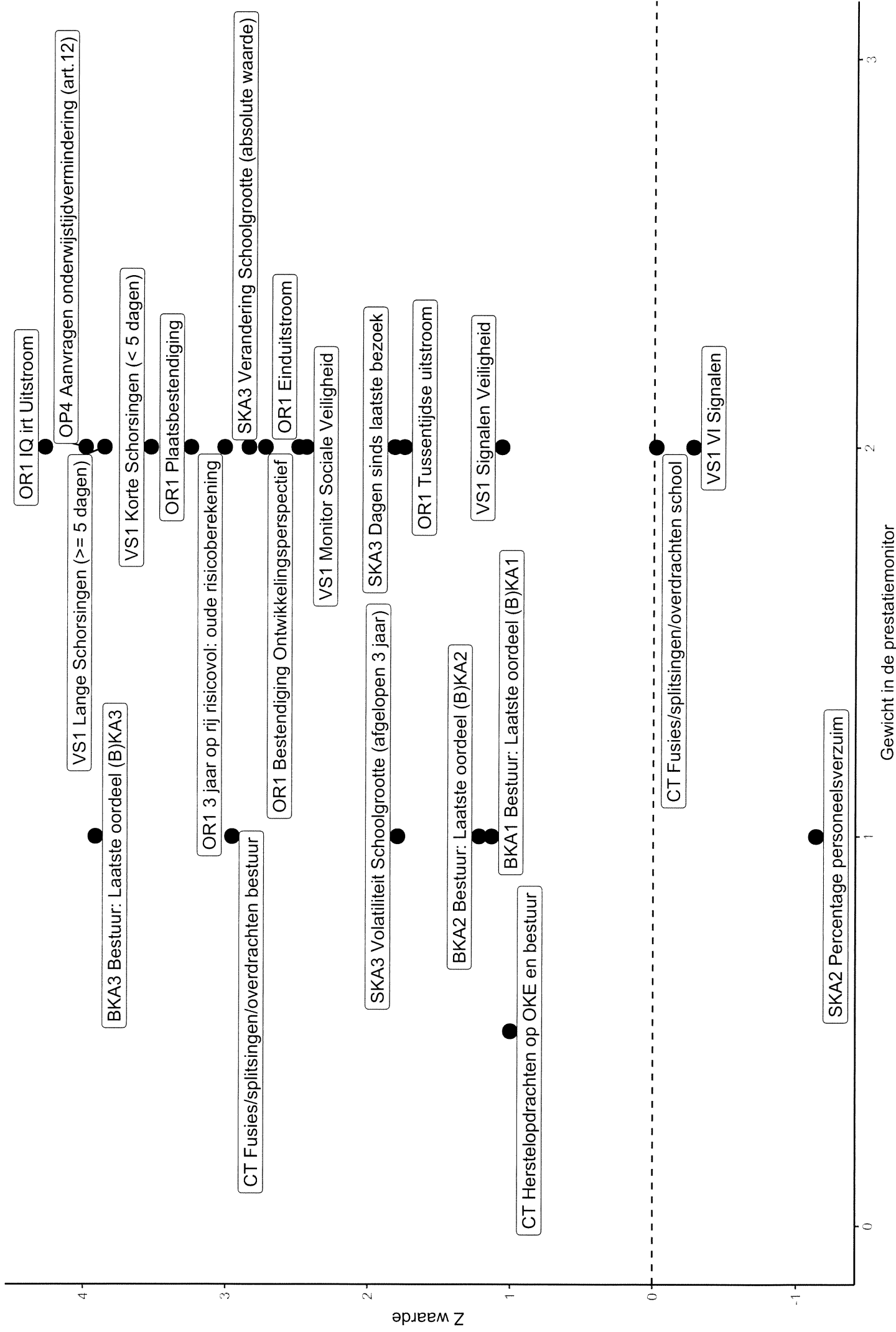
Waardering: Goed  
Geen expert analyse of oordeel

# PM SO 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel





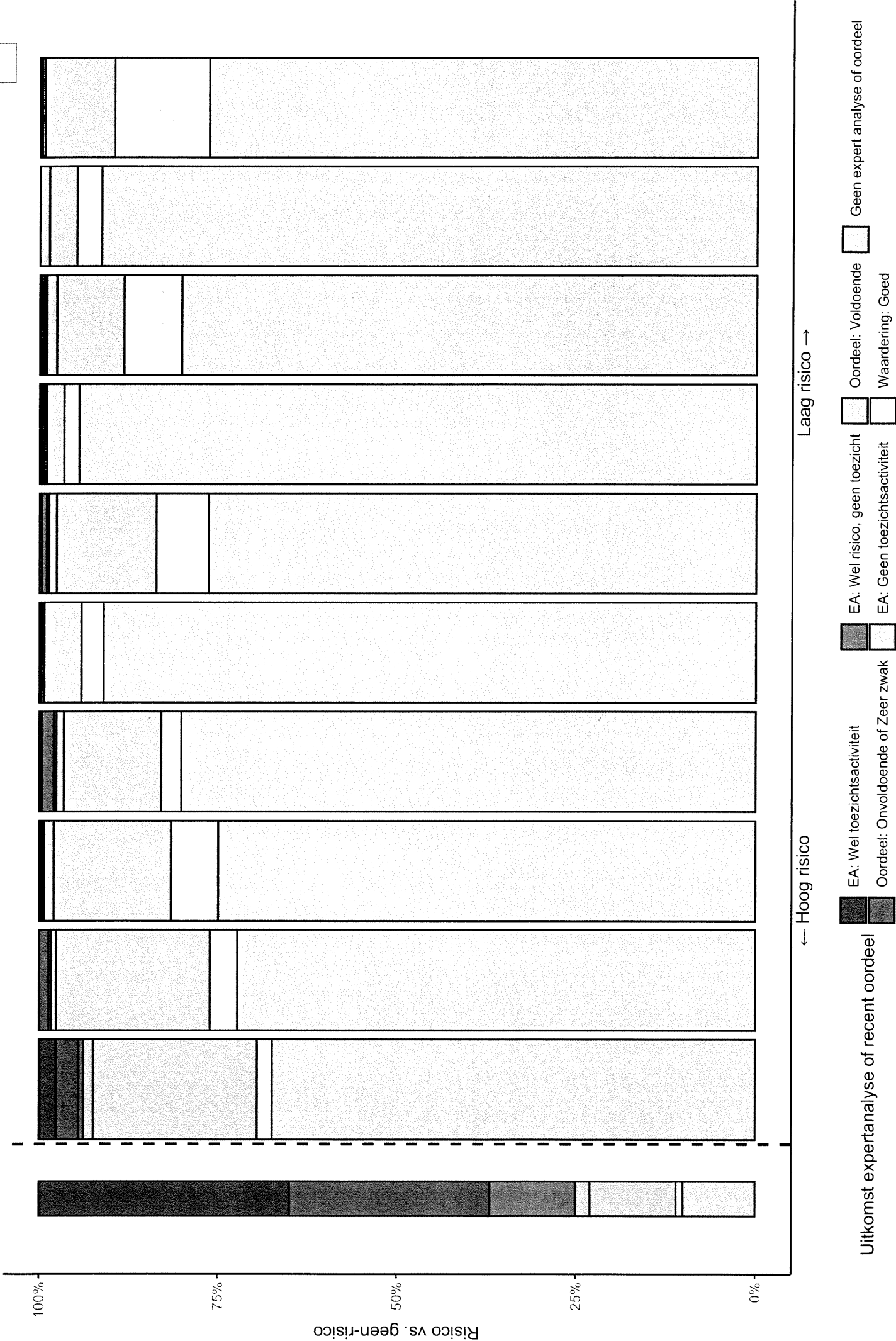
# PM SO 2021 - Relatie voorspelkracht en gewicht



Prestatiemonitor SO 2021 - Gevoeligheid

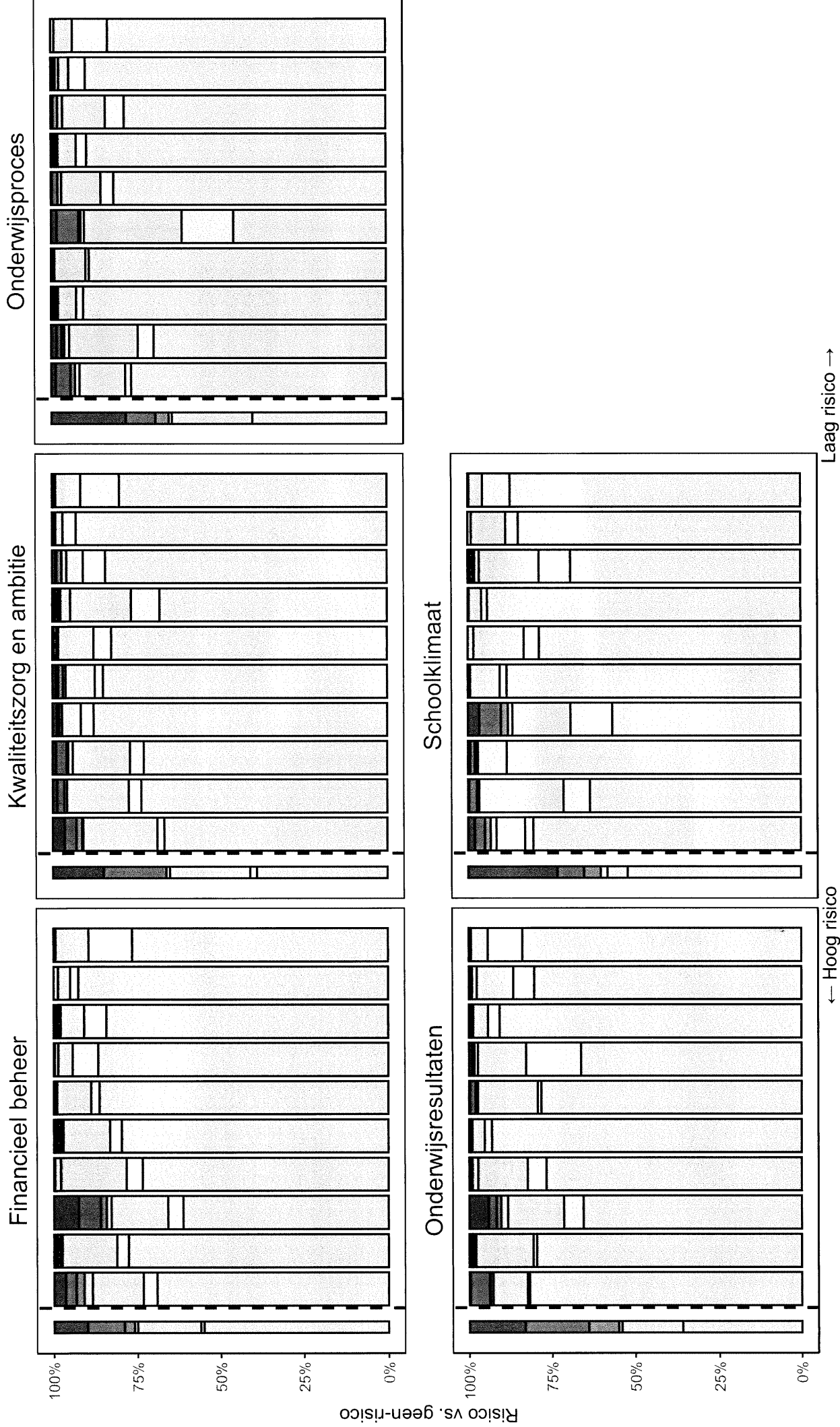
Gevoeligheid	Besturing, kwaliteitszorg en ambitie	Context	Onderwijsresultaten	Sturen, kwaliteitszorg en ambitie	Veiligheid en Schoolklimaat	Totaal
Sensitiviteit	0.29	0.45	0.45	0.38	0.45	0.68
Specifiteit	0.81	0.85	0.85	0.83	0.85	0.84
Precisie	0.29	0.45	0.45	0.38	0.45	0.54

PM VO 2021 - PM rang en uitkomst experanalyse of oordeel

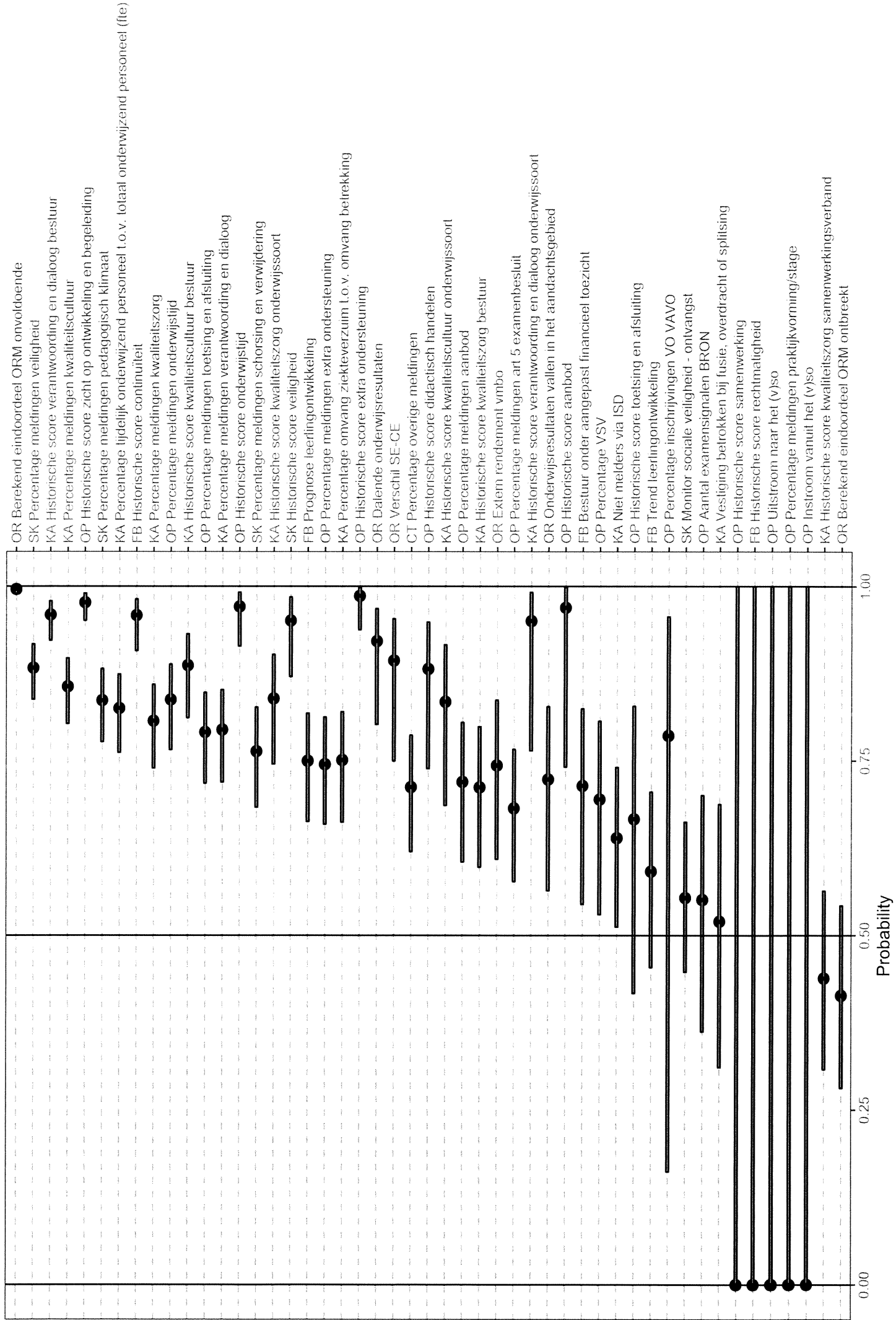


# PM VO 2021 - PM rang en uitkomst experanalyse of oordeel

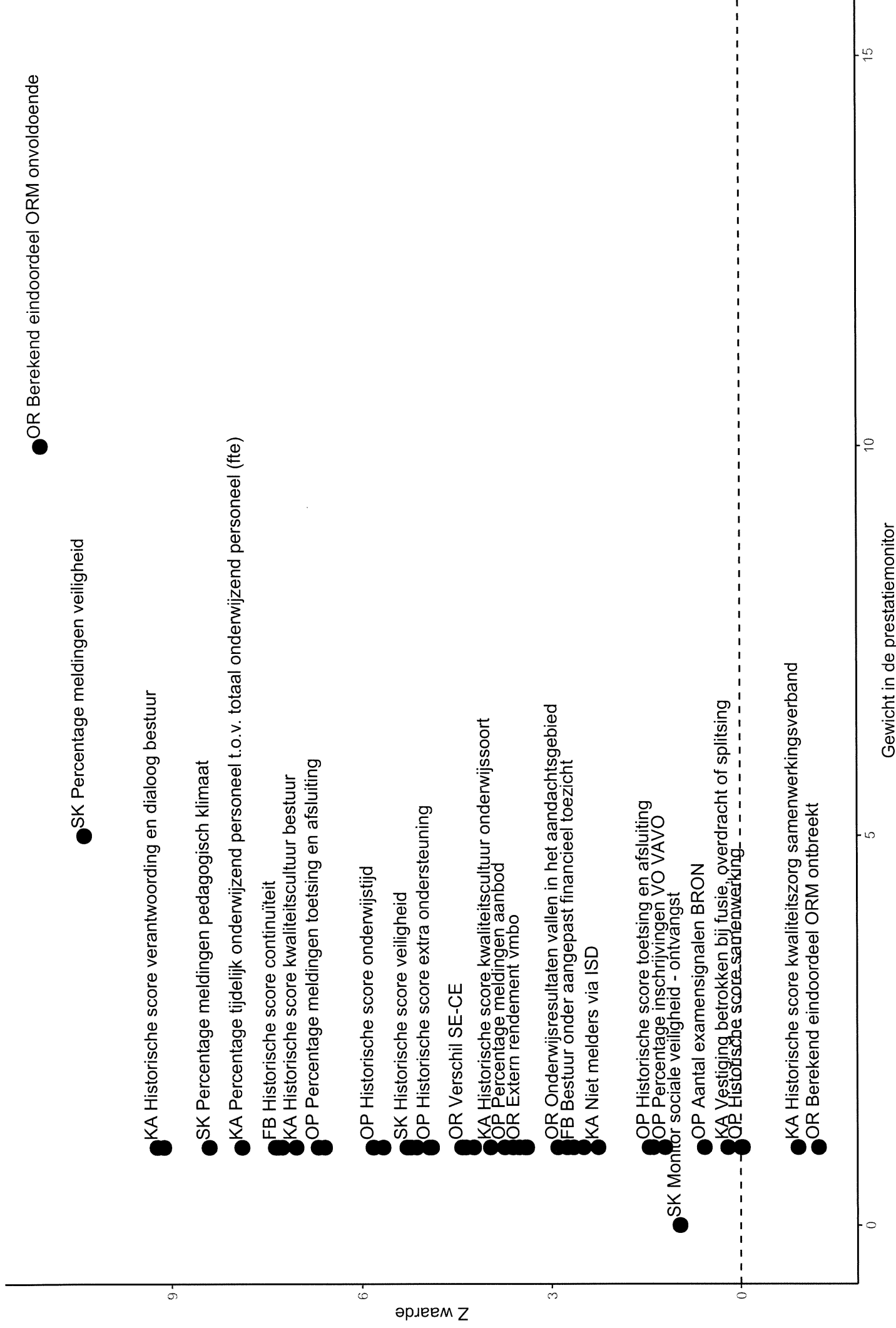
Per kwaliteitsgebied



# PM VO 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



# PM VO 2021 - Relatie voorspelkracht en gewicht

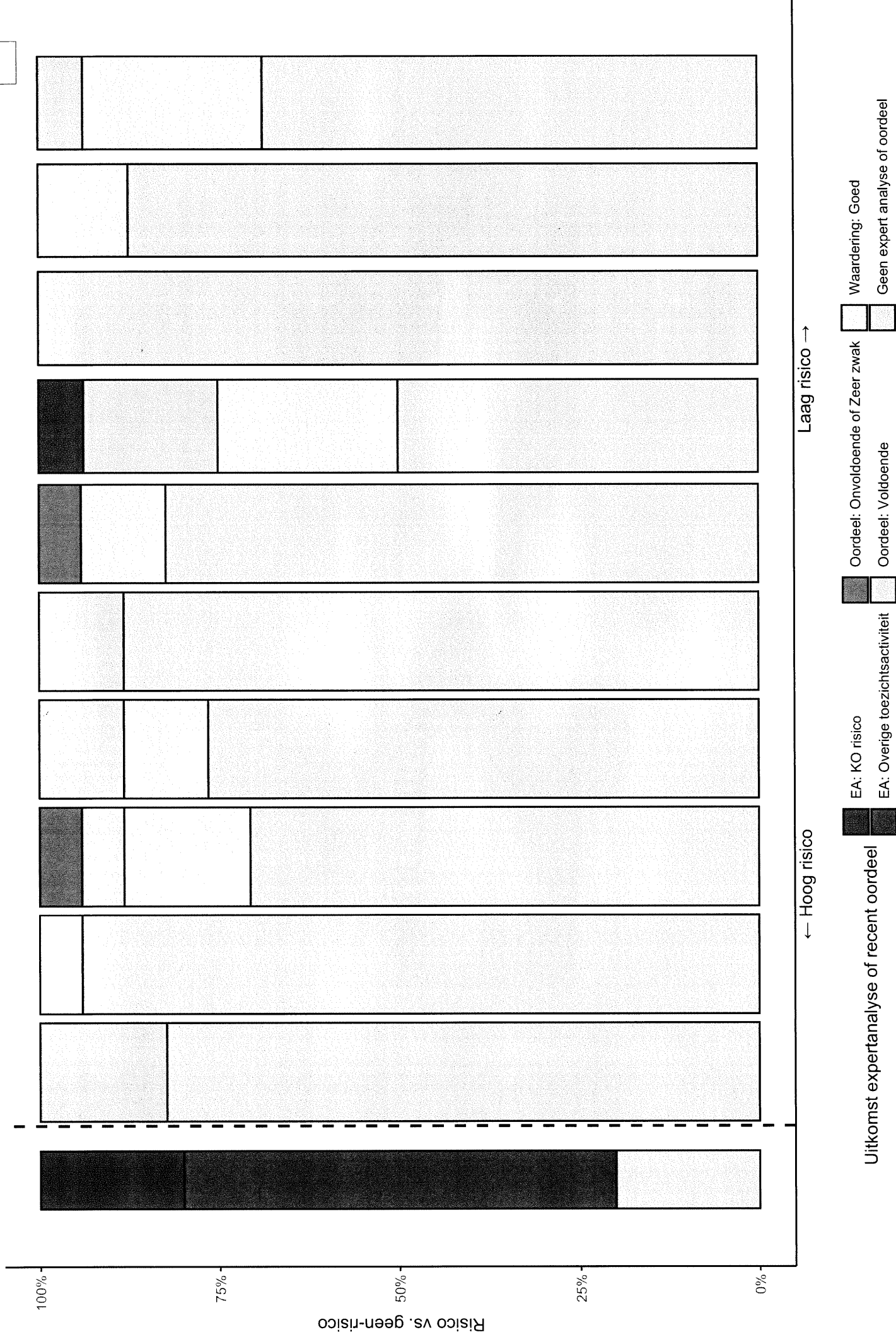


# Prestatiemonitor VO 2021 - Gevoeligheid

Gevoeligheid	Financieel beheer	Kwaliteitszorg en ambitie	Onderwijsresultaten	Onderwijsproces	Schoolklimaat	Totaal
Sensitiviteit	0.20	0.29	0.38	0.30	0.34	0.64
Specificiteit	0.97	0.98	0.98	0.98	0.98	0.99
Precisie	0.24	0.34	0.45	0.35	0.40	0.82

# PM VO PRO 2022 - PM rang en uitkomst EA of oordeel

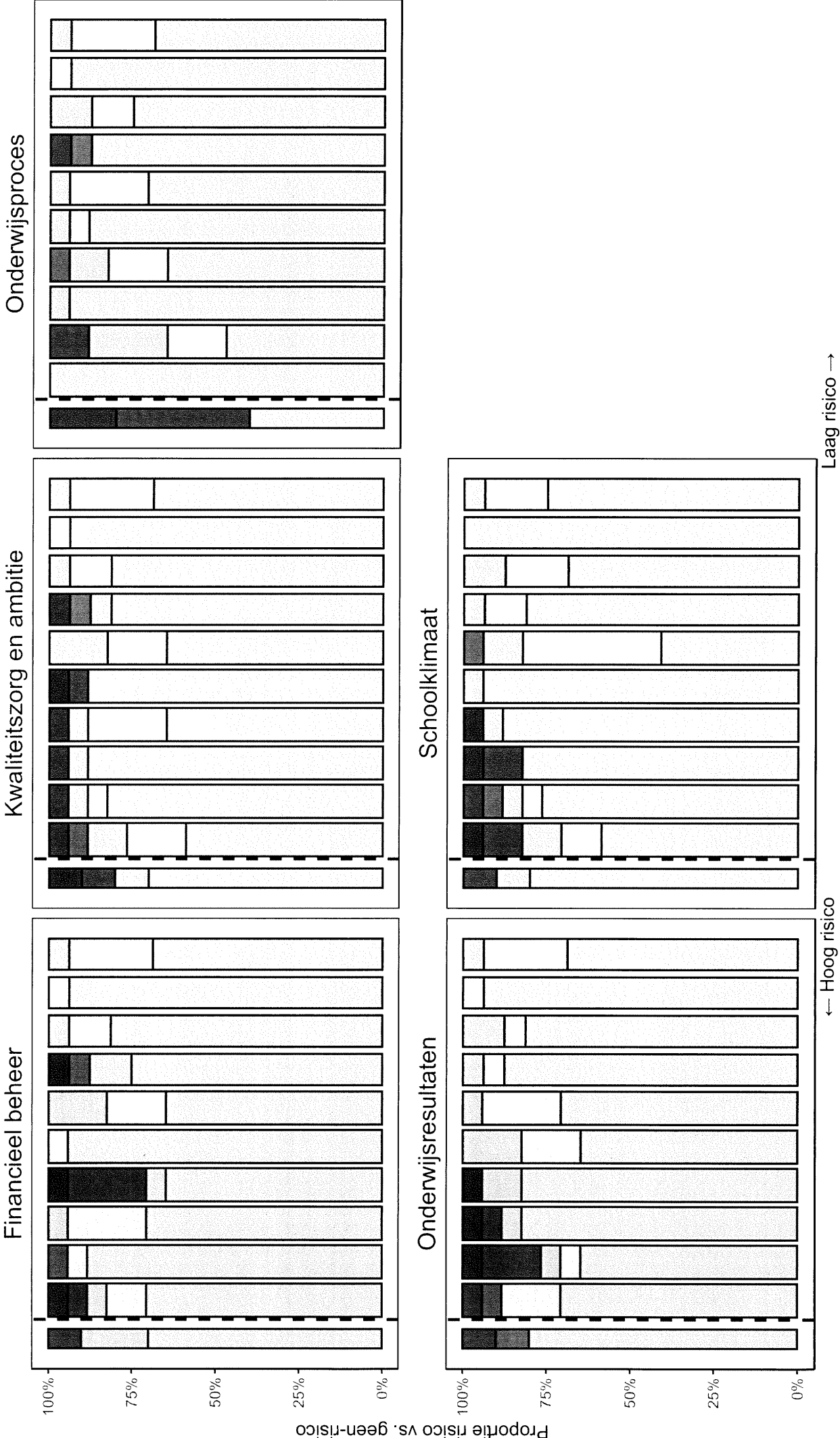
17.





# PM VO PRO 2022 - PM rang en uitkomst EA of oordeel

Per kwaliteitsgebied



Uitkomst expertanalyse of recent oordeel

EA: KO risico

EA: Overige toezichtsactiviteit

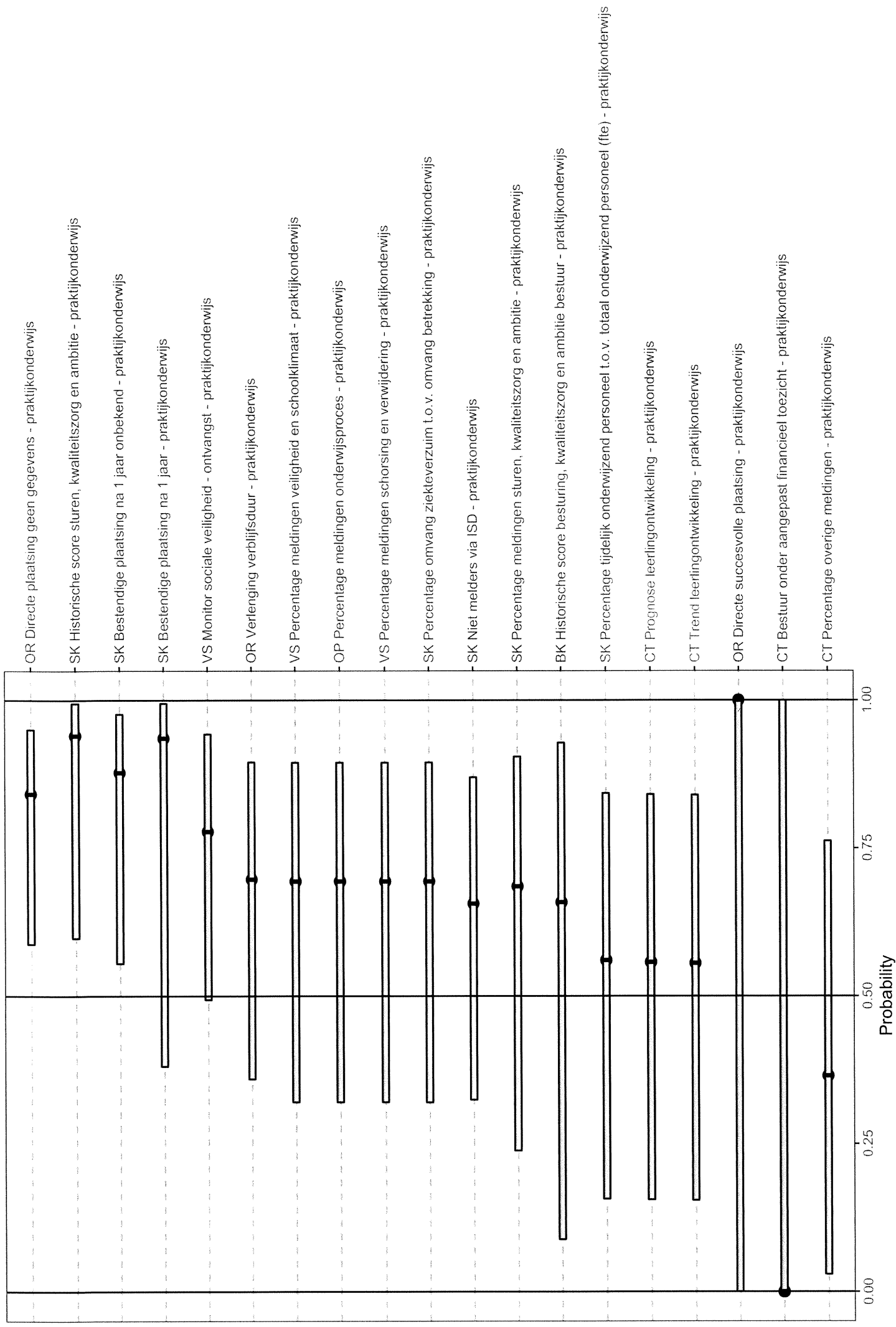
Oordeel: Onvoldoende of Zeer zwak

Oordeel: Voldoende

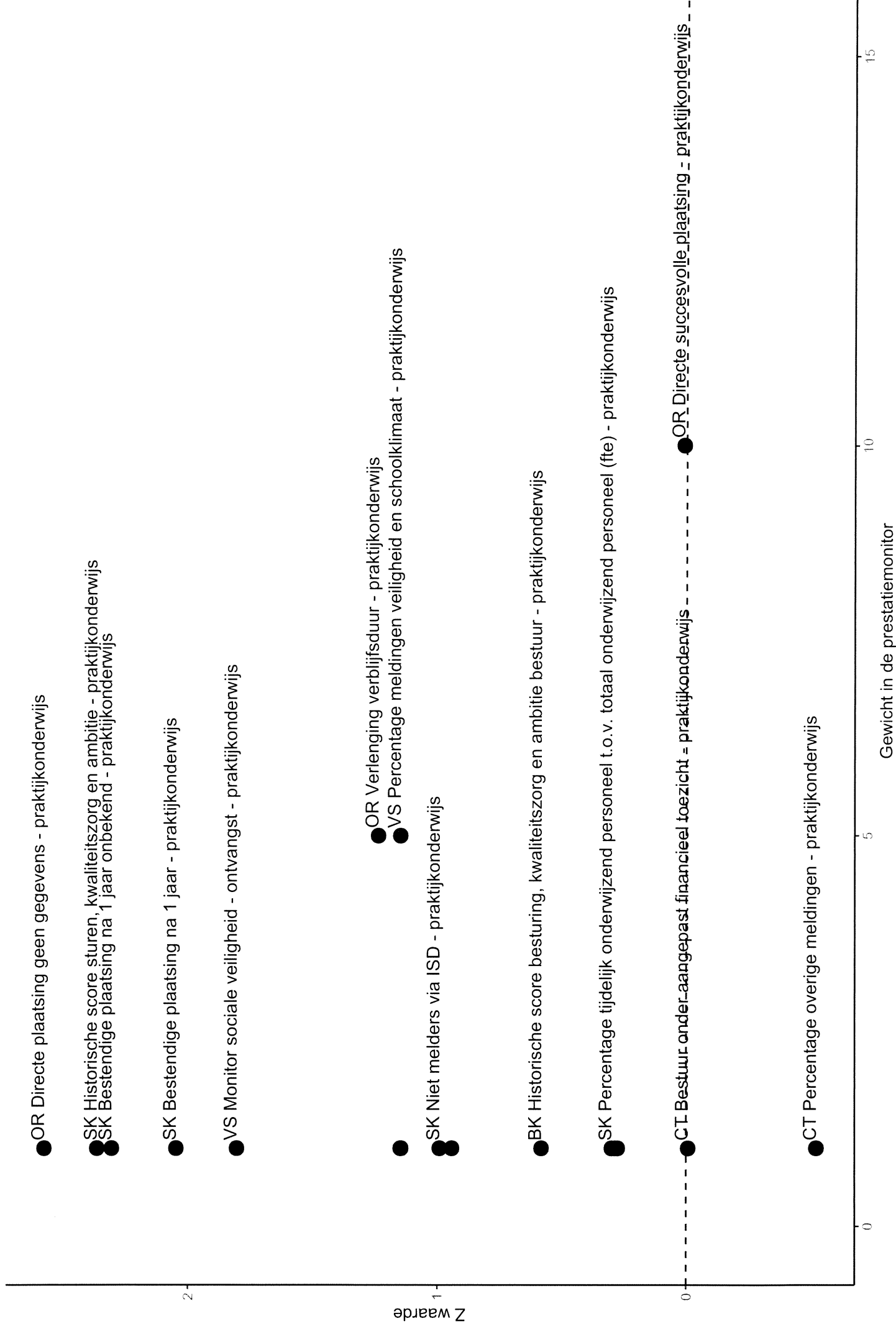
Waardering: Goed

Geen expert analyse of oordeel

# PM VO PRO 2022 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



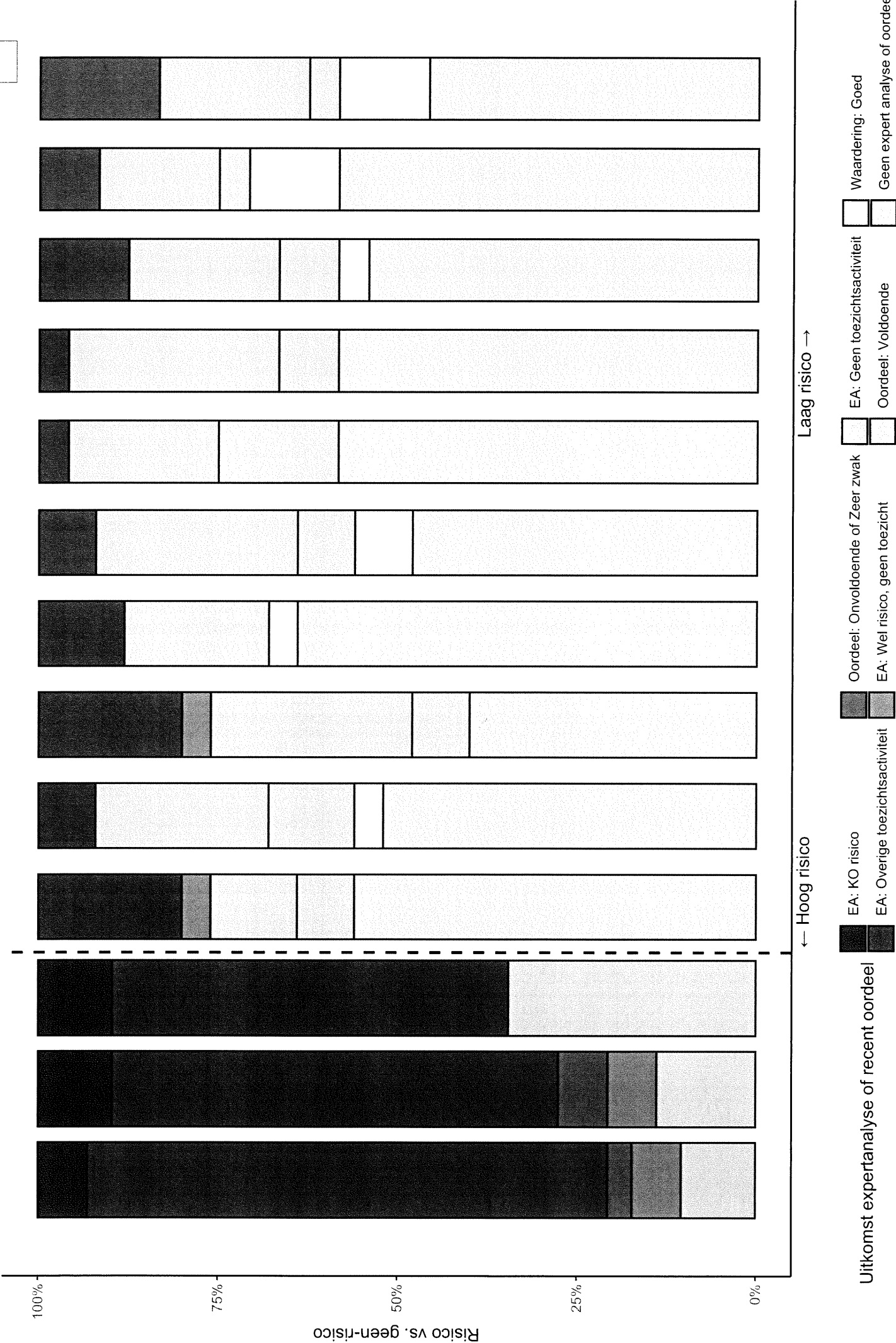
# PM VO PRO 2022 - Relatie voorspelkracht en gewicht



Prestatiemonitor VO PRO 2022 - Gevoeligheid

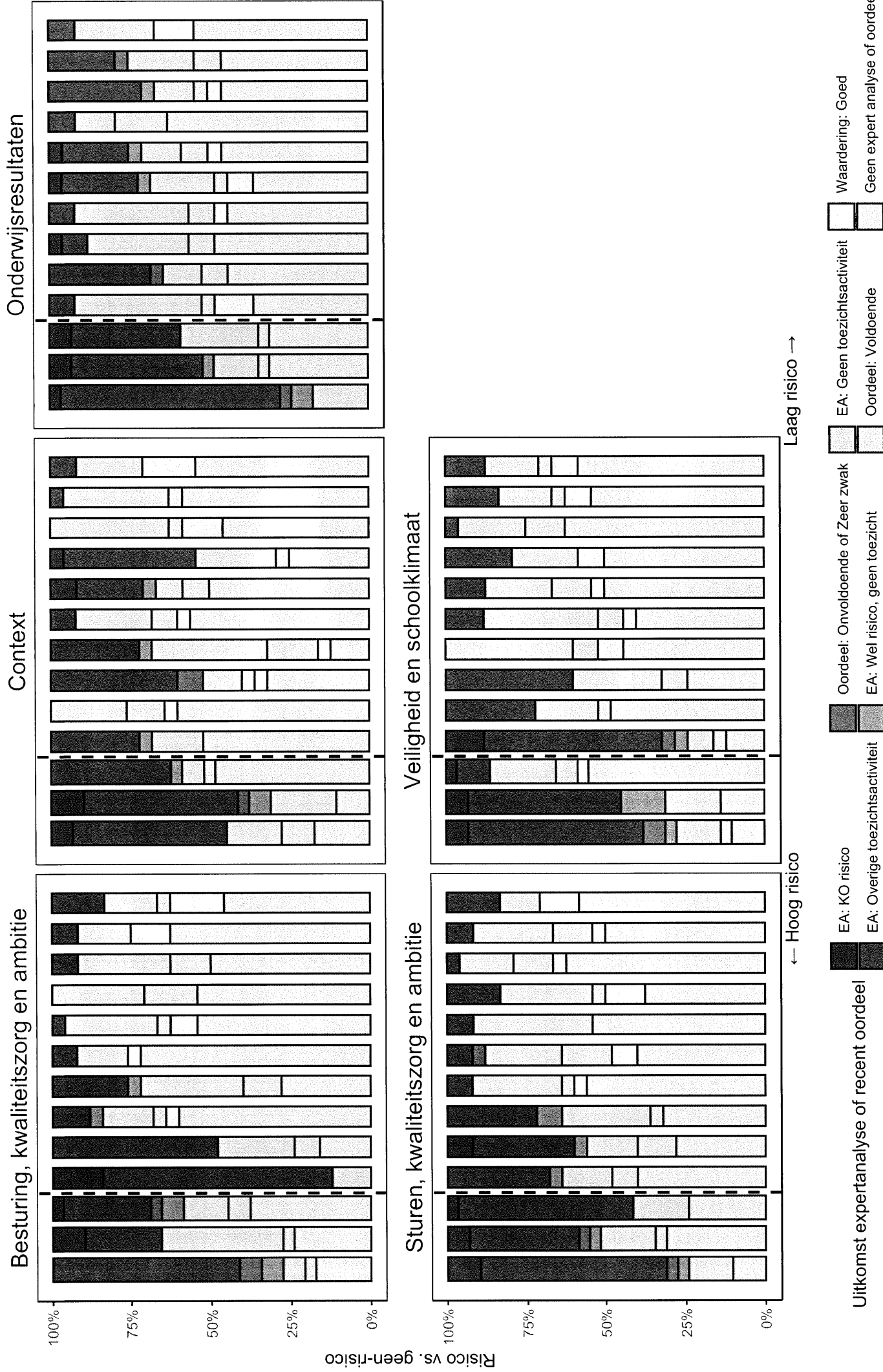
Gevoeligheid	Financieel beheer	Kwaliteitszorg en ambitie	Onderwijsresultaten	Onderwijsproces	Schoolklimaat	Totaal
Sensitiviteit	0.09	0.18	0.18	0.55	0.09	0.73
Specifiteit	0.95	0.95	0.95	0.98	0.95	0.99
Precisie	0.10	0.20	0.20	0.60	0.10	0.89

PM VSO 2021 - PM rang en uitkomst EA of oordeel

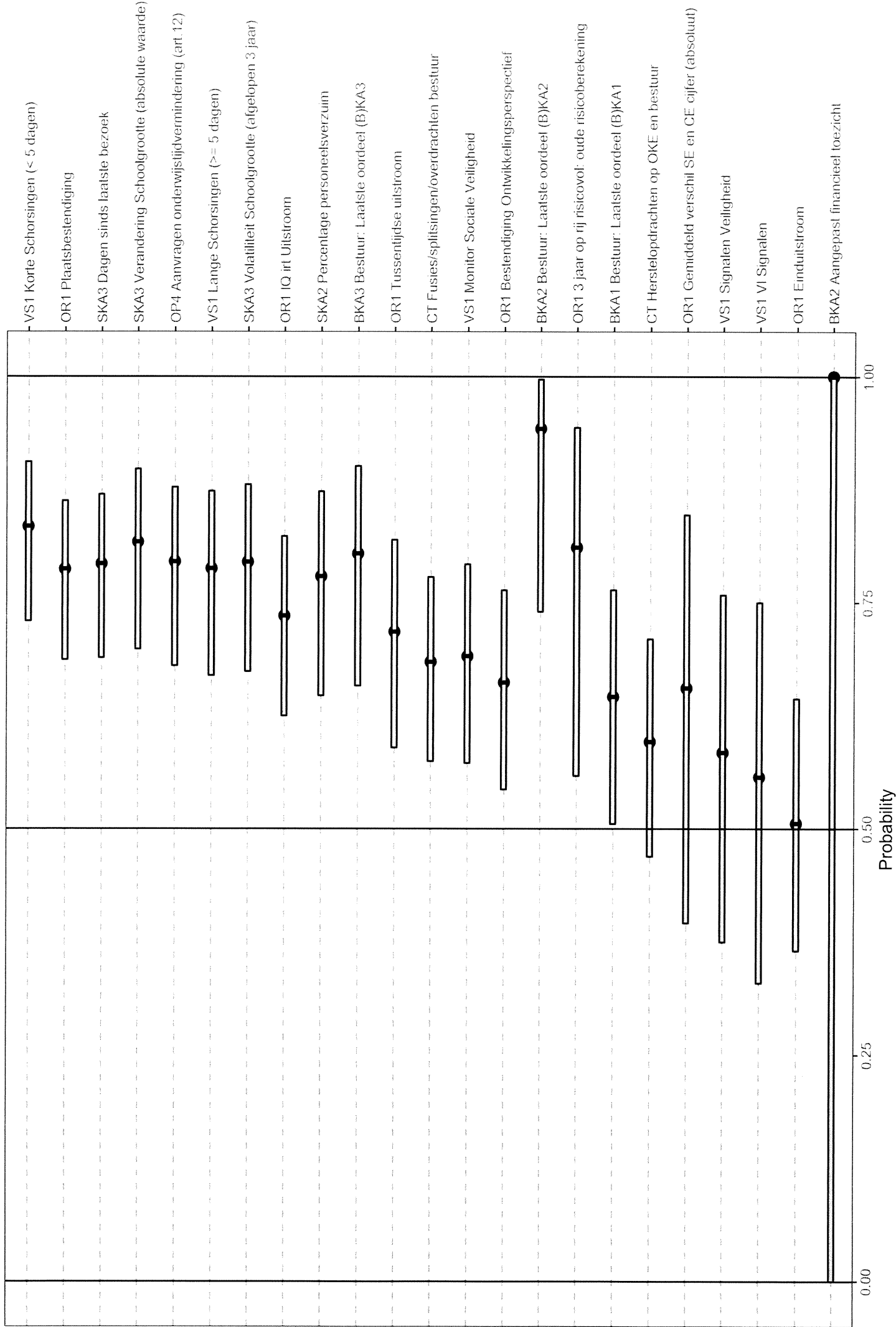


# PM VSO 2021 - PM rang en uitkomst EA of oordeel

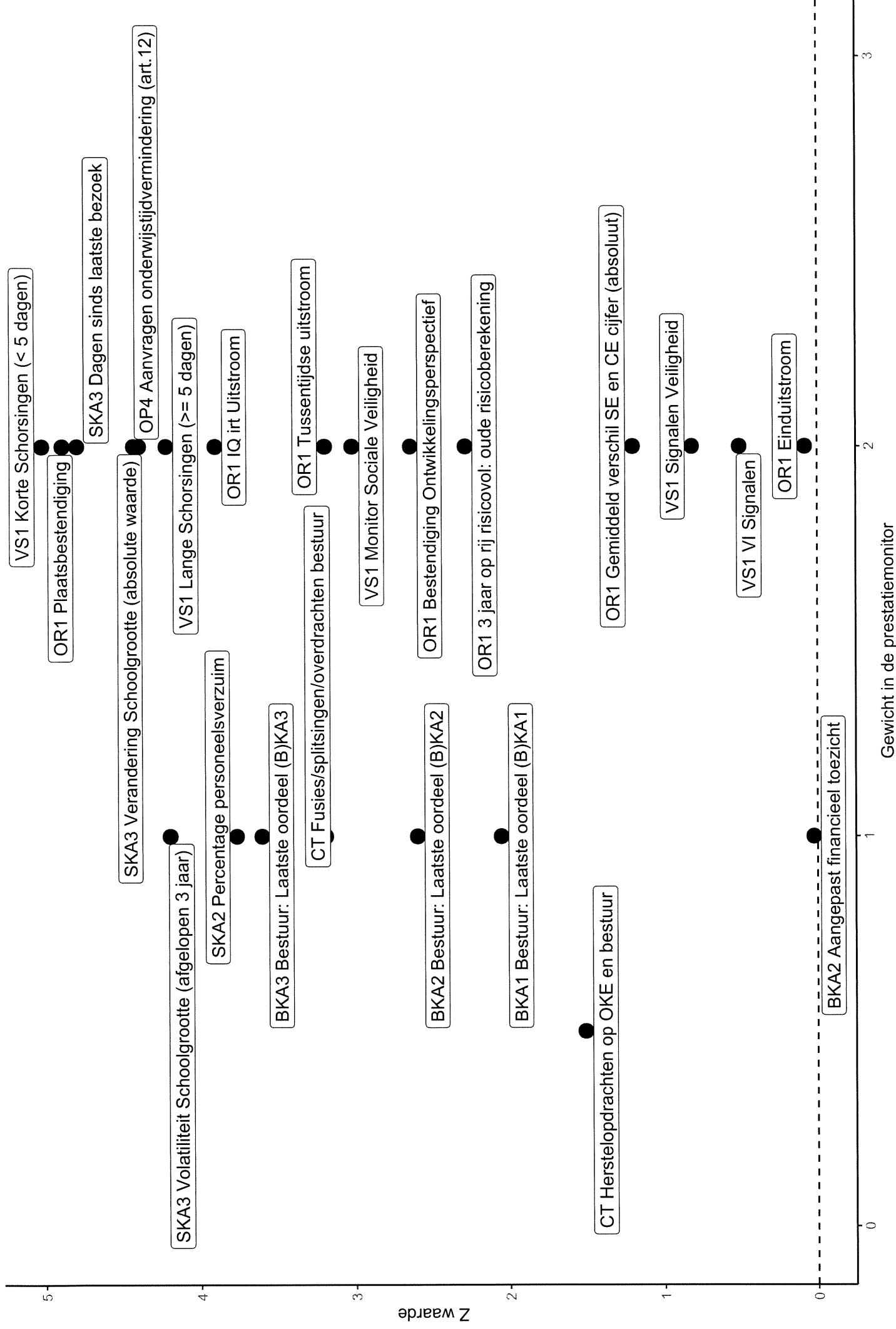
Per kwaliteitsgebied



# PM VSO 2021 - Voorspelkracht indicatoren uitkomst expertanalyse of oordeel



# PM VSO 2021 - Relatie voorspelkracht en gewicht





Prestatiemonitor VSO 2021 - Gevoeligheid

Gevoeligheid	Besturing, kwaliteitszorg en ambitie	Context	Onderwijsresultaten	Sturen, kwaliteitszorg en ambitie	Veiligheid en Schoolklimaat	Totaal
Sensitiviteit	0.43	0.48	0.51	0.53	0.45	0.70
Specifiteit	0.81	0.83	0.84	0.85	0.82	0.93
Precisie	0.49	0.55	0.59	0.61	0.52	0.80



Inspectie van het Onderwijs  
*Ministerie van Onderwijs, Cultuur en  
Wetenschap*

20.

# **Projectplan**

## **Prestatiemonitors 2023**

Opdrachtgever:  
Programmamanager TIP:  
Projectleider prestatimonitors:

Directeur Kennis  
**5.1.2e**

Versie 1.0 januari 2023



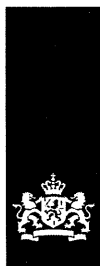
## 6 Conclusie en aanbevelingen

Zoals beschreven in de inleiding is het belangrijk om inzicht te hebben in 'de voorkeuren' van de risicomodellen van de IvhO. Uit de bovenstaande analyses blijkt dat in de Prestatiemonitor VO geldt dat scholen met een hoog percentage leerlingen uit apc-gebied, kleine scholen en scholen met een vmbo-afdeling relatief vaker een hoge risicoscore krijgen. Met name de factoren schoolgrootte en percentage leerlingen uit een apc-gebied vertonen relatief sterke samenhang met de risicoscores. Het voornaamste doel van deze analyse is het inzichtelijk maken van deze voorkeuren, om het mogelijk te maken hier op te reflecteren.

Het is belangrijk te benoemen dat de aanwezigheid van indirecte bias op zich geen probleem hoeft te zijn. Het is mogelijk dat de oververtegenwoordiging van sommige groepen scholen goed te beredeneren valt, en aansluit op de ervaringen van inspecteurs en analisten. De indicatoren waarmee de vormen van bias samenhangen hangen dan, naar mening van de sector VO en de ontwikkelaars van de prestatie-monitor, duidelijk samen met risico's voor de onderwijskwaliteit. In dit geval zou een aanpassing van de prestatie-monitor niet nodig zijn, omdat de prestatie-monitor zorgvuldig en beredeneerd tot stand is gekomen. Het feit dat bepaalde typen scholen dan vaker naar voren komen in de groep scholen met hoge risicoscores is dan te verklaren. Wel kan het wenselijk zijn om meer bewustwording te creëren, bijvoorbeeld door gebruikers van de prestatie-monitor te informeren dat enkele typen scholen vaker hoog scoren dan andere typen scholen. Het zou in bredere zin ook aanleiding kunnen geven om te onderzoeken *waarom* er grotere risico's zijn op dergelijke scholen.

Het is ook mogelijk dat, nu de bias inzichtelijk is gemaakt, de oververtegenwoordiging van bepaalde typen scholen toch niet goed te beredeneren valt. Wellicht worden de voorkeuren van de Prestatiemonitor VO t.a.v. deze groepen niet herkend, dan wel als onwenselijk gezien. In dat geval is het zinvol om de prestatie-monitor aan te passen, bijvoorbeeld door vormen van bias corrigeren in de berekening van de risicoscore. Bijvoorbeeld door het aanpassen/weglaten specifieke indicatoren die sterke samenhang vertonen met het betrokken kenmerk. Na een dergelijke aanpassing zouden de bovenstaande analyses wederom uitgevoerd kunnen worden om te bekijken of de vormen van bias aanhouden.

Middels deze rapportage willen we de sector VO vragen om te reflecteren op de indirecte biases van de Prestatiemonitor VO. Zijn de voorkeuren zoals aangetoond verassend, of juist herkenbaar met de beelden die in de praktijk worden opgehaald? En acht de sector het wenselijk om te corrigeren voor (een deel van) de factoren waarop de Prestatiemonitor bias vertoont? Graag gaan wij hierover verder in gesprek.



aantal leerlingen op een vestiging.

Alle indicatoren lijken significant verband te houden met het aantal leerlingen op de vestigingen, alhoewel ze wel verschillen wat betreft de richting van dit verband: sommige hebben een positief verband, daar waar andere een negatief verband hebben. In het bijzonder valt op dat de indicatoren OR1\_verschil\_sece (verschil schoolexamen en centraal eindexamen), VS1\_meldingen\_sv (percentage meldingen schorsingen en verwijderingen) en CT\_trend\_leerlingen (trend leerlingontwikkeling) het sterkste negatieve verband vertonen met het aantal leerlingen op de vestiging. De indicatoren OR1\_orm\_aandachtsgebied (onderwijsresultaten vallen in het aandachtsgebied) en OR1\_orm\_dalend (dalende onderwijsresultaten) hangen daarentegen positief samen met een grotere vestiging in termen van leerlingaantallen.

**Tabel 4.1: Samenhang indicatoren en grootte vestiging (aantal leerlingen)**

	<b>coefficient</b>	<b>p_value</b>
BKA_historische_score	-0,303	0,000
CT_prognose_leerlingen	-0,286	0,000
CT_trend_leerlingen	-0,832	0,000
OP_historische_score	-0,403	0,000
OP_meldingen	-0,209	0,000
OP6_examensignalen	-0,191	0,000
OR1_orm_aandachtsgebied	0,505	0,000
OR1_orm_dalend	0,521	0,000
OR1_orm_geen_oordeel	-0,788	0,000
OR1_verschil_sece	-1,073	0,000
SKA_meldingen	-0,133	0,000
SKA2_tijdelijk_personeel_perc	-0,288	0,000
SKA2_ziekteverzuim_perc	-0,175	0,000
SKA3_extern_rendement_vmbo	-0,401	0,000
SKA3_niet_melders	-0,363	0,000
VS_meldingen	-0,452	0,000
VS1_meldingen_sv	-0,841	0,000
OP6_art5_eb	-0,123	0,000
CT_bg_financieel_toezicht	-0,238	0,000
VS_monitor_ontvangst	0,079	0,000
CT_meldingen	-0,085	0,000
VS_historische_score	-0,698	0,000
OR_historische_score	-0,395	0,000
OR1_orm_onvoldoende	0,192	0,000
OP2_vsv	-0,099	0,000
SKA_historische_score	0,107	0,000
SKA2_fusie	0,245	0,000
OR_meldingen	0,201	0,000
OR1_CE_duim	-0,019	0,000

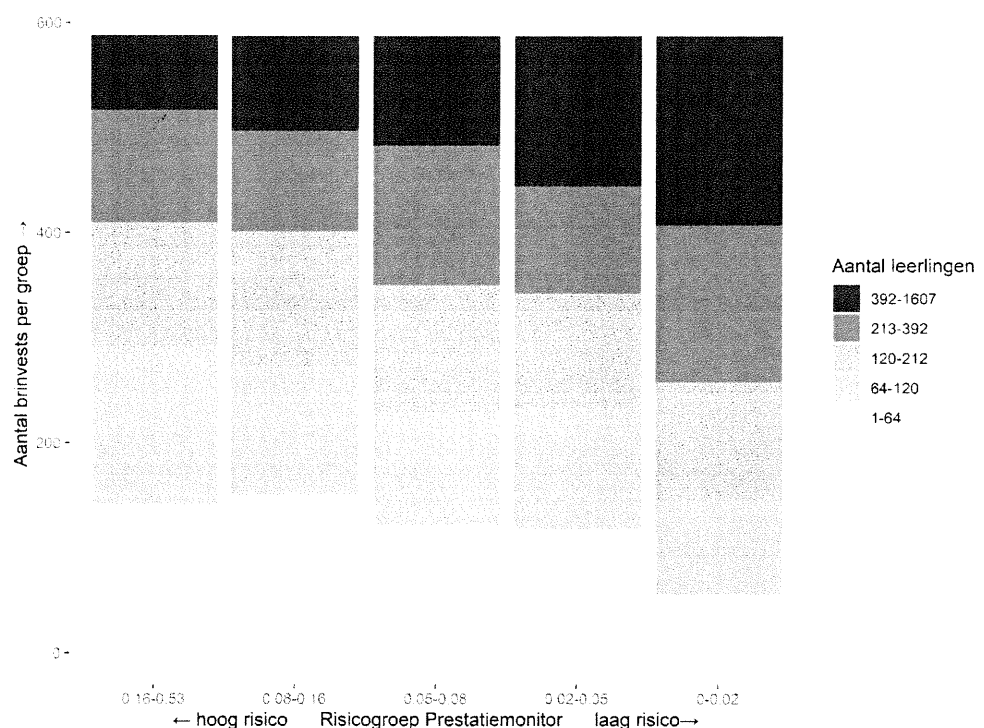


## 5 Grootte

Het derde aspect waar mogelijk bias plaatsvindt en dat is onderzocht is de grootte van de vestiging. Hier wordt de grootte van de vestiging gedefinieerd als het aantal leerlingen op de betreffende vestiging. Deze variabele is ontwikkeld uit de 1-cijferbestanden en bevat het aantal leerlingen dat ingeschreven was op de vestiging op de peildatum.

Wederom zijn de vestigingen ingedeeld in groepen met ongeveer een gelijke grootte wat betreft leerlingaantallen. De groepen zijn zo gekozen dat er in elke groep ongeveer evenveel vestigingen zitten. Deze groepen zijn vervolgens in onderstaande visualisatie afgezet tegen de gegroepeerde risicoscores.

Uit onderstaande visualisatie blijkt duidelijk dat in de groep vestigingen met de laagste risicoscores vaker vestigingen voorkomen met een grote leerlingpopulatie dan in de groep met de hoogste risicoscores. Grofweg kan aan de hand van deze visualisatie geconcludeerd worden dat met name vestigingen met een kleine leerlingpopulatie een hoge risicoscore krijgen. Het aandeel vestigingen met een kleine leerlingpopulatie in een groep met ongeveer gelijke risicoscores neemt toe naar mate de risicoscores hoger worden.



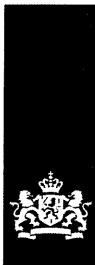
Vervolgens is ook voor de grootte van de vestiging in leerlingaantallen gekeken naar de samenhang met de indicatoren. Hierbij is het leerlingaantal als afhankelijke variabele genomen en zijn de indicatoren telkens als onafhankelijke variabele gebruikt. Aangezien de afhankelijke variabele een telling is, is gekozen voor een Poisson-regressiemodel. Vervolgens zijn de coëfficiënten met elkaar vergeleken om een uitspraak te kunnen doen over de samenhang tussen de indicatoren en het



Kijkend naar de samenhang tussen de indicatoren en de aanwezigheid van vmbo op een vestiging lijken er enkele indicatoren een positieve samenhang te vertonen en enkele een negatieve samenhang. Bij een significantieniveau van 0.01 lijken VS1\_meldingen\_sv, CT\_trend\_leerlingen, OP6\_art5\_eb, CT\_prognose\_leerlingen, OR1\_orm\_geen\_oordeel, VS\_meldingen een positieve samenhang te vertonen, gesorteerd op afnemende mate van samenhang. Bij eenzelfde significantieniveau vertonen SKA3\_niet\_melders, OR1\_orm\_aandachtsgebied, OR1\_orm\_dalend, VS\_monitor\_ontvangst, BKA\_historische\_score en CT\_meldingen juist een negatief verband met de aanwezigheid van vmbo op een vestiging, wat inhoudt dat risico's op deze indicatoren minder waarschijnlijk zijn als het vmbo op de vestiging vertegenwoordigd is.

**Tabel 3.1: Samenhang indicatoren en aanwezigheid vmbo op vestiging**

	<b>coefficient</b>	<b>p_value</b>
SKA3_niet_melders	-1,509	0,000
OR1_orm_aandachtsgebied	-1,417	0,000
VS1_meldingen_sv	3,072	0,000
OR1_orm_dalend	-1,265	0,000
CT_trend_leerlingen	1,002	0,000
VS_monitor_ontvangst	-0,517	0,000
OP6_art5_eb	0,760	0,000
CT_prognose_leerlingen	0,681	0,000
OR1_orm_geen_oordeel	0,521	0,000
BKA_historische_score	-0,614	0,001
VS_meldingen	0,473	0,003
CT_meldingen	-0,331	0,008
OR1_orm_onvoldoende	-0,894	0,014
SKA2_ziekteverzuim_perc	0,333	0,031
SKA_historische_score	-0,569	0,032
OR1_CE_duim	0,335	0,062
OP_meldingen	0,263	0,066
OP_historische_score	-0,515	0,093
CT_bg_financieel_toezicht	-0,359	0,093
SKA_meldingen	0,229	0,108
SKA2_fusie	0,609	0,196
OR_meldingen	-0,838	0,226
OR_historische_score	0,742	0,328
OP6_examensignalen	0,157	0,328
OR1_verschil_sece	0,621	0,400
SKA2_tijdelijk_personeel_perc	-0,025	0,858
OP2_vsv	-0,016	0,944
VS_historische_score	15,233	0,982
SKA3_extern_rendement_vmbo	-0,000	1,000

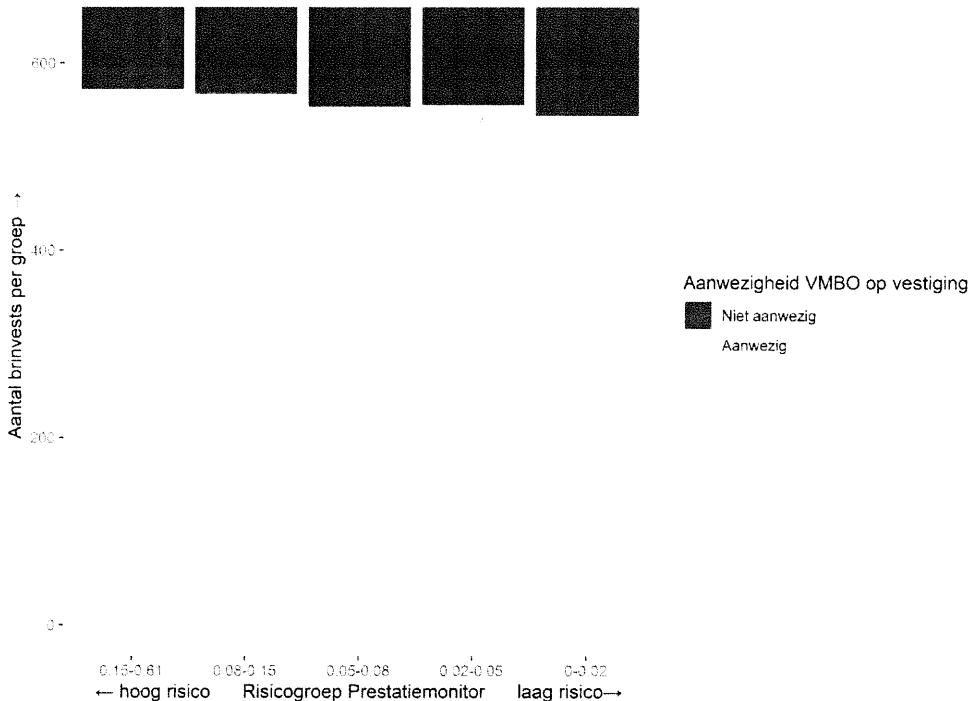


## 4 VMBO

Een tweede mogelijke vorm van bias zoals deze samen met de sector is gedefinieerd is de aanwezigheid van een vmbo-afdeling op de vestiging. Deze variabele is dichotoom gescoord, wat betekent dat de variabele de waarde 0 had op het moment dat een vestiging geen vmbo-afdeling had en dat de variabele de waarde 1 had op het moment dat er op een vestiging wel een vmbo-afdeling aanwezig was.

Wederom is de variabele waar mogelijk onbedoeld op wordt geselecteerd afgezet tegen de risicoscores uit de prestatie-monitor. De groepen voor de risicoscores zijn op een vergelijkbare manier samengesteld als bij het percentage leerlingen uit apc-gebied: de groepen zijn in absolute zin ongeveer even groot.

Uit de visualisatie blijkt dat vestigingen die hoog scoren wat betreft de risicoscore iets vaker een vmbo-afdeling hadden dan vestigingen met een relatief lage risicoscore. Naarmate de risicoscores toenemen, neemt ook het aandeel vestigingen met een vmbo-afdeling toe. De verschillen lijken echter wel minder uitgesproken vergeleken met de zeer duidelijke verschillen die zichtbaar waren bij het percentage leerlingen uit apc-gebied.



Voor het aspect vmbo is eveneens gekeken naar de samenhang tussen de indicatoren en de aanwezigheid van het vmbo op een vestiging. Hier is de aanwezigheid van het vmbo (0 = niet aanwezig, 1 = wel aanwezig) gebruikt als een afhankelijke variabele in een logistische regressie-model. De indicatoren zijn individueel iteratief meegenomen in dit model en de coëfficiënten zijn vervolgens met elkaar vergeleken om een uitspraak te kunnen doen over de samenhang.



**Tabel 2.1: Samenhang indicatoren en pct. leerlingen uit apc-gebied**

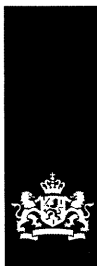
	<b>coefficient</b>	<b>p_value</b>
SKA2_tijdelijk_personeel_perc	0,858	0,000
VS1_meldingen_sv	0,590	0,000
VS_meldingen	0,338	0,000
CT_meldingen	0,310	0,000
SKA_meldingen	0,303	0,000
OP2_vsv	0,451	0,000
OR1_CE_duim	0,309	0,000
OP_meldingen	0,259	0,000
OR1_orm_geen_oordeel	-0,282	0,000
CT_prognose_leerlingen	0,250	0,000
OP6_art5_eb	0,240	0,000
BKA_historische_score	-0,343	0,000
OR1_orm_aandachtsgebied	0,344	0,001
VS_monitor_ontvangst	-0,144	0,003
SKA3_extern_rendement_vmbo	0,234	0,020
CT_trend_leerlingen	0,143	0,025
OP6_examensignalen	0,126	0,046
SKA3_niet_melders	-0,224	0,065
OP_historische_score	0,317	0,067
SKA2_ziekteverzuim_perc	0,079	0,204
OR1_verschil_sece	0,299	0,209
CT_bg_financieel_toezicht	0,131	0,234
OR_meldingen	-0,404	0,309
OR_historische_score	0,278	0,325
OR1_orm_dalend	0,099	0,429
OR1_orm_onvoldoende	0,147	0,473
SKA2_fusie	0,185	0,619
VS_historische_score	-0,183	0,637
SKA_historische_score	-0,059	0,664





Vervolgens is er gekeken naar welke indicatoren in de prestatiemonitor sterk samenhangen met het percentage leerlingen uit apc-gebied. Dit is gedaan door middel van een beta-regressie, waarbij het percentage leerlingen uit apc-gebied als afhankelijke variabele fungeerde en de telkens een van de indicatoren uit de prestatiemonitor als onafhankelijke variabele. Dit is iteratief toegepast, waardoor dit uiteindelijk per indicator één regressiemodel opleverde. Uiteindelijk zijn de coëfficiënten van de modellen met elkaar vergeleken om een uitspraak te kunnen doen over welke indicatoren het sterkst samenhangen met het percentage leerlingen uit apc-gebied. Voor inhoudelijke duiding van de verschillende indicatoren in deze Prestatiemonitor verwijzen wij naar de toelichting van de *Prestatiemonitor VO*

Gesorteerd op p-waarden wordt duidelijk dat met name de indicatoren die gebaseerd zijn op meldingen sterke positieve samenhang vertonen met het percentage leerlingen uit apc-gebied. Dit geldt eveneens voor een significantieniveau van 0.01 voor het percentage tijdelijk personeel (SKA2\_tijdelijk\_personeel\_perc ; deze heeft de grootste coëfficiënt), het percentage vroegtijdige schoolverlaters, de centrale examencijfers zonder duimvakken, de prognose van de leerlingontwikkeling, het percentage meldingen op basis van artikel 5 van het examenbesluit, en onderwijsresultaten die vallen in het aandachtsgebied. Ter illustratie; voor de indicator 'percentage tijdelijk personeel' betekent dit dat er op scholen met veel leerlingen uit een apc-gebied veel vaker tijdelijk personeel wordt ingezet. Voor het ontbreken van een eindoordeel ORM, het ontvangst van de monitor sociale veiligheid en de historische score besturing, kwaliteitszorg en ambitie geldt juist dat zij met eenzelfde significantieniveau juist een negatieve samenhang vertonen met het percentage leerlingen uit apc-gebied.

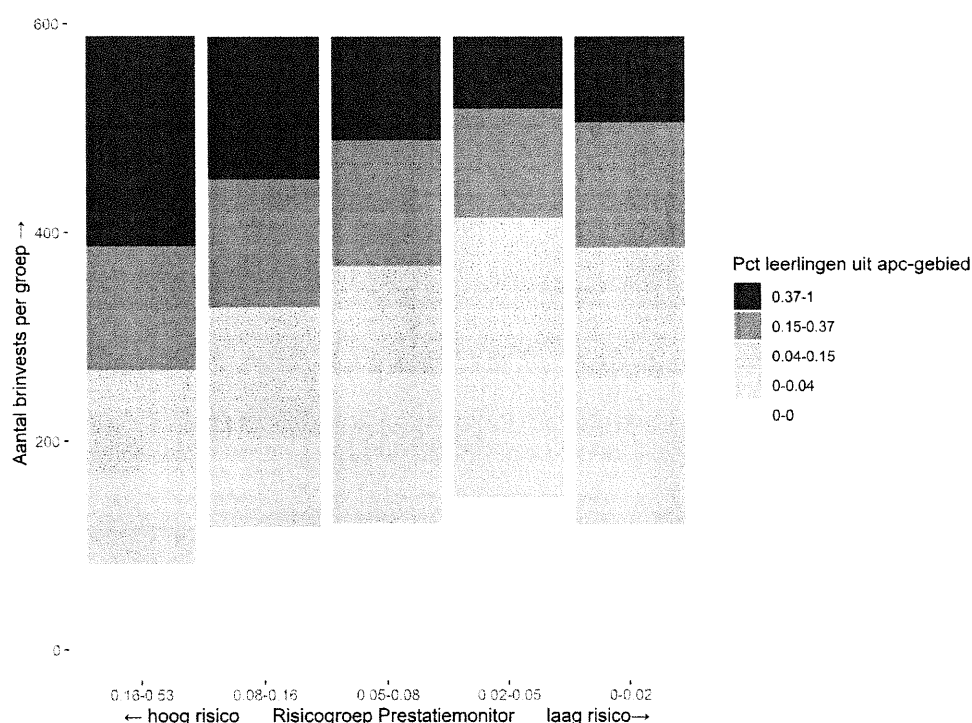


### 3 Percentage APCG

Een eerste mogelijke vorm van bias is het percentage leerlingen dat afkomstig is uit een armoedeprobleemcumulatiegebied ("apc-gebied"). Een viercijfer-postcodegebied wordt als een apc-gebied beschouwd als er een relatief hoog aandeel huishoudens met een inkomen onder de lage inkomensgrens en een relatief hoog aandeel huishoudens met een uitkering en een relatief hoog aandeel huishoudens met een hoofdkostwinner met een niet-westerse migratie-achtergrond. Hierbij worden studentenhuishoudens buiten beschouwing gelaten. Het Centraal Bureau voor de Statistiek ("CBS") berekent de apcg-status van postcodegebieden in opdracht van het Ministerie van Onderwijs, Cultuur en Wetenschappen. Het percentage leerlingen afkomstig uit apc-gebied wordt berekend door te kijken naar de inschrijvingen op vestigingsniveau.

Onderstaande visualisatie groepeerde de brin-vestigingen naar vergelijkbare risicoscores in groepen van gelijke grootte. Vervolgens wordt de verdeling van het percentage leerlingen uit apc-gebied (eveneens gegroepeerd naar vergelijkbare scores in groepen van gelijke grootte) over de verdeling van de risicoscores gevisualiseerd.

Duidelijk is dat in de groep brin-vestigingen met de hoogste risicoscores veel meer brin-vestigingen zitten met een hoog percentage leerlingen uit apc-gebied. In de groepen met lage risicoscores zijn de groepen met hoge percentages apcg-leerlingen juist kleiner, en de groepen met lage percentages apcg-leerlingen juist groter. Derhalve lijkt het er op dat OVT's met een hoge risicoscore vaker OVT's zijn met een hoog percentage leerlingen uit apc-gebied.





## 2 Inleiding

De Inspectie van het Onderwijs gebruikt prestatie-monitors binnen de risico-analyses. Binnen deze prestatie-monitors wordt een risicoscore berekend aan de hand van verschillende indicatoren. Bij overschrijding van de grenswaarde voor een indicator worden er risicopunten toegekend. De uiteindelijke risicoscore is het behaalde aantal risicopunten gedeeld door het totaal behaalbare risicopunten.

De indicatoren die in de prestatie-monitor zijn meegenomen behelzen factoren die (mogelijk) samenhangen met de kwaliteit van het betreffende schoolbestuur of schoolvestiging. Deze indicatoren zijn expliciet geprogrammeerd en dragen daarom direct bij aan de berekening van de risicoscore.

Naast de indicatoren die direct bijdragen aan een hogere score, is het ook mogelijk dat er factoren indirect samenhangen met een hogere risicoscore. Dat zijn kenmerken die niet expliciet in de prestatie-monitor zijn meegenomen. In dat geval spreken we in dit stuk van *indirecte biases*. Het is mogelijk dat indirecte bias zich voordoet omdat deze factor samenhangt met een van de indicatoren die wel in het model zit, of omdat meerdere indicatoren tezamen samenhangen met de impliciete factor. Om te voorkomen dat risicomodellen bepaalde groepen scholen naar voren halen zonder dat we ons daar van bewust zijn is het belangrijk om indirecte vormen van bias inzichtelijk te maken.

Het doel van het voorliggende memo is om enkele mogelijke indirecte biases zichtbaar te maken binnen de Prestatie-monitor Voortgezet Onderwijs. Samen met de sector VO zijn er drie mogelijke vormen van bias geïdentificeerd, die in onderstaande hoofdstukken nader worden toegelicht: het percentage leerlingen uit apc-gebied, de aanwezigheid van vmbo op een vestiging, en de grootte van een vestiging gemeten in het aantal leerlingen. Voor elk van deze mogelijke biases geldt dat zij niet als indicator in de prestatie-monitor zijn meegenomen en daarom dus niet direct bijdragen aan de totstandkoming van de risicoscore zoals hierboven beschreven.

Per hoofdstuk zal de samenhang tussen de mogelijke vorm van bias en de risicoscores worden toegelicht. Daarnaast zal er per hoofdstuk een analyse gegeven worden over welke indicatoren (die expliciet gecodeerd zijn) samenhangen met de betreffende vorm van bias.



## **1      Inhoud**

<b>1</b>	<b><i>Inhoud.....</i></b>	<b>2</b>
<b>2</b>	<b><i>Inleiding.....</i></b>	<b>3</b>
<b>3</b>	<b><i>Percentage APCG.....</i></b>	<b>4</b>
<b>4</b>	<b><i>VMBO.....</i></b>	<b>7</b>
<b>5</b>	<b><i>Grootte.....</i></b>	<b>9</b>
<b>6</b>	<b><i>Conclusie en aanbevelingen.....</i></b>	<b>11</b>



## **EVALUATIE INDIRECTE BIASES IN DE PRESTATIEMONITOR VO**

PM-VO

September 2022

## Overige zaken

### ***Onkoppelbare OKE's***

Ten tijde van het opstellen van dit document is er voor een aantal OKE's geen mogelijkheid tot het berekenen van een aantal indicatoren. Dit is het geval bij alle indicatoren die gebruik maken van leerlingaantallen. Voor deze OKE's kan de informatie niet betrouwbaar worden berekend, omdat een BRIN-VST onderdeel van twee OKE's is. Het is dan onmogelijk om te bepalen op basis van de inschrijvingsbestanden welke leerling van de vestiging op welk OKE zit. De prestatiemonitor is voor deze OKE's dan minder gevuld dan de andere OKE's. Het is dan maar de vraag of bij deze scholen de risicodetectie volledig werkt. Daarmee moet rekening worden gehouden gedurende expert-analyse

De OKE's waarbij hier sprake van is, voor het schooljaar 2021-2022:

SO:

OKE	naam	bgnr
5.1.2.e		

VSO:

OKE	naam	bgnr
5.1.2.e		

## **Sturing Kwaliteitszorg en Ambitie**

### **Percentage ziekteverzuim personeel**

Het bij Duo gerapporteerde ziekteverzuim van personeel. Dit is het verzuim gerapporteerd voor het hele BRIN. Data op het niveau van OKE is helaas niet beschikbaar. Voor schooljaar 2019-2020 dient er rekening te worden gehouden met verschillende registratiepraktijken wegens de eerste corona lockdown. Sommige besturen hebben corona besmetting wel en andere niet als ziekteverzuim geregistreerd.

Verandering leerlingaantal (absolute waarde) De absolute mate van verandering in schoolgrootte ten opzichte van het voorgaande jaar wordt per school per jaar berekend. Over de totale dataset wordt een gewogen gemiddelde van de veranderingen over drie jaar genomen. De grenswaarde wordt berekend door over dit gewogen gemiddelde het 85<sup>e</sup> percentiel te nemen. Dit betekent dat de 15% scholen met de meest extreme absolute veranderingen in leerlingpopulatie als risico scoren op deze indicator.

### **Volatiliteit Schoolgrootte (afgelopen 3 jaar)**

Deze indicator laat de volatiliteit van de leerlingontwikkeling (procentuele groei of krimp) op OKE niveau over drie jaar zien. Risico wordt berekend op de som van absolute procentuele verandering schoolgrootte (zie bovenstaande indicator). Volatiliteit drukt uit hoe 'grillig' het verloop van de schoolgrootte is. Scholen die het ene jaar sterk stijgen en het andere jaar weer sterk dalen scoren risicovol op deze indicator.

### **Dagen sinds laatste bezoek**

Aantal dagen sinds het laatste onderzoek heeft plaats gevonden. Bij meer dan  $365 \times 5 + 1 = 1826$  dagen is er sprake van een risico. Alleen onderzoeken die een centrale activiteit 'Uitvoeren onderzoek op Locatie' hebben tellen mee in de analyse. Met andere woorden, we moeten fysiek over de drempel zijn geweest. Onderzoeken op afstand tellen dus niet mee

## **Context**

### **Aantal fusies/splitsingen/overdrachten (school)**

Deze indicator bekijkt bij hoeveel fusies, overdrachten of splitsingen de school de afgelopen 5 jaar betrokken is geweest. De fusies, splitsingen en overdrachten tellen allemaal even zwaar mee. Wanneer er sprake is geweest van minstens 1 verandering op het schoolniveau, is er sprake van een risico voor de desbetreffende school.

## **Overig Onderwijsresultaat**

### ***Drie jaar op rij risicovol (oude risicoberekening)***

Is de OKE volgens de oude risicoberekening 3 jaar op rij risicovol? Deze indicator bouwt voort op de oude prestatie-analyse. Wanneer een school drie jaar op rij als risicovol naar voren komt middels de oude analyse, wordt 'Drie keer rood' risico. Dit betekent dat de school dan een score 1 heeft op de indicator. Voor een volledige uitleg zie:

<https://proza.ocw.local/otcs/cs.exe/app/nodes/23492093>



### **Lege indicatoren in de vragenlijst?**

Vragenlijst indicatoren kunnen om meerdere redenen leeg zijn bij een school in een bepaald jaar. Dat is niet perse een datafout. Hieronder de mogelijkheden:

- Wanneer een school geen vragenlijst inlevert, worden er geen indicatoren gevuld.
- Wanneer een school niet alle verplichte velden van onderdeel van de vragenlijst niet heeft gevuld, kan de indicator niet worden berekend.
- Voor residentiële locaties worden geen indicatoren berekend.
- Bij scholen in IQ-groep 2 worden de indicatorscore *einduitstroom* en *tussentijdse uitstroom* niet gevuld. Dit zou de indruk wekken dat wij verwachten van scholen dat kinderen met een laag IQ kunnen uitstromen naar regulier onderwijs.
- Wanneer een school onvoldoende IQ-gegevens invult komt deze in IQ-groep 3: onbekend terecht. Hierdoor kan sowieso de indicator *IQ irt uitstroom* niet worden berekend. Wanneer er onvoldoende scholen in deze groep vallen (minder dan 5), worden de percentielgrenzen niet opgesteld voor deze groep en kunnen de indicatoren *einduitstroom*, *tussentijdse uitstroom* en *plaatsbestending* ook niet worden berekend. Er zijn dan onvoldoende scholen om mee te vergelijken.

## Onderwijsresultaten

### Vragenlijst

Onder onderwijsresultaten staan een aantal indicatoren die worden berekend op basis van de gegevens die ophalen met onze jaarlijkse vragenlijst uitstroomgegevens. De volgende indicatoren zijn hetzelfde gebleven:

- Einduitstroom
- IQ irt Uitstroom
- Plaatsbestendinging
- Tussentijdse uitstroom
- Ontwikkelingsperspectief

De berekening van deze indicatoren staat uitvoerig beschreven op de website [Uitwerking berekening prestatieanalyse \(voortgezet\) speciaal onderwijs 2021 | Publicatie | Inspectie van het onderwijs \(onderwijsinspectie.nl\)](#) en in Proza: <https://proza.ocw.local/otcs/cs.exe/app/nodes/31510383>. Voor al deze indicatoren geldt dat de score 1 gelijk staat aan risico op de indicator. Hieronder volgen korte beschrijving

### **Einduitstroom**

De verhouding tussen einduitstroom naar regulier vs speciaal (so) of onderwijs vs arbeid (vso). De 25% scholen met laagste percentage einduitstroom naar regulier onderwijs (i.p.v. speciaal, so) vervolgonderwijs (i.p.v. arbeid, vso) scoren risicovol op de indicator. In jaren dat meer dan 25% van de scholen 0% uitstroom naar regulier onderwijs realiseren, krijgen alle scholen met de score 0% risico op deze indicator.

### **IQ irt Uitstroom**

Gemiddelde uitstroompunt min gemiddelde IQ-punt levert het verschilpunt. De 25% scholen met het laagste verschilpunt worden onvoldoende. Om de uitstroom te relateren aan het IQ van de uitgestroomde leerlingen is er een waarde toegekend aan zowel de uitstroomniveaus als de IQ score (in categorieën). Deze waarden worden met elkaar vergeleken.

### **Plaatsbestendinging**

Obv percentage bestendinging: leerlingen die in 2017-2018 zijn uitgestroomd en op 1-10-2019 nog steeds op een uitstroombestemming van hetzelfde niveau zitten. De 25% scholen met laagste bestendinging scoren risicovol op de indicator. In jaren dat meer dan 25% van de scholen 0% van de leerlingen bestendigen, krijgen alle scholen met de score 0% risico op deze indicator.

### **Tussentijdse uitstroom**

De verhouding tussen tussentijdse uitstroom naar regulier- en speciaal (voortgezet) onderwijs. De 25% scholen met laagste percentage uitstroom naar regulier (voortgezet) onderwijs (i.p.v. speciaal) scoren risicovol op de indicator. In jaren dat meer dan 25% van de scholen 0% tussentijdse uitstroom naar regulier onderwijs hebben, krijgen alle scholen met de score 0% risico op deze indicator.

### **Uitstroom op ontwikkelingsperspectief**

Percentage leerling dat uitstroomt op OPP. Wanneer dit op 75% of hoger ligt is er geen sprake van risico. Bij scores onder de 75% is er sprake van risico op de indicator. Dit wordt in de prestatie-monitor al vertaald naar een score van 0 of 1, waarbij de score '1' een risico aangeeft.

### ***Veiligheidsmonitor: niet of te laat ingeleverd***

Deze indicator meet of door een school is voldaan aan de levering van data omtrent monitoring van sociale veiligheid per schooljaar. De data over die monitoring kan de inspectie op twee manieren bereiken. Ten eerste zijn er de scholen die de data aanleveren via een van de instrument aanbieders. Ten tweede is er voor scholen de mogelijkheid om documenten te leveren via het ISD.

Wanneer een OKE via één of via beide kanalen informatie heeft ingeleverd, is er géén sprake van risico. Dus scholen die niks aanleveren ontvangen risico voor de indicator op dit schooljaar. Dit wordt in de scores vertaald naar een 0=niet geleverd, 1 = niet geleverd.

De wettelijke verplichting geldt op het niveau van BRIN-VEST. Omdat sommige scholen de data van de scholen (onterecht) samenvoegen voor zij het leveren, wordt er risico gedetecteerd op het niveau van OKE. Dat betekent wanneer een OKE twee vestigingen onder zich heeft hangen, en maar op basis van één van de twee vestigingen wij een levering ontvangen, de andere BRIN-VEST (en daarmee het OKE) voordeel van de twijfel krijgt. Zelfs deze coulante manier van benaderen levert al genoeg risicoscores op.

In de toekomst wordt het mogelijk om via het nieuw te starten project Verdiepingsinformatie de inhoud van de veiligheidsmonitoren zichtbaar wordt.

### ***Schorsingen***

Er zijn twee indicatoren m.b.t. schorsingen:

Korte Schorsingen (< 5 dagen) Lange Schorsingen (>= 5 dagen)

Beide indicatoren worden berekend door het aantal geregistreerde schorsingen te delen door het aantal leerlingen op het OKE, keer 100. Hierbij wordt de 15% 'slechtst' presterende OKE's (dus met het hoogste percentage schorsingen) een risico toegekend. Schorsingen van minstens 5 dagen vallen buiten het wettelijk kader. Hierdoor is hier een aparte indicator van gemaakt.

Elke schorsing is als een individuele schorsing behandeld. Dat wil zeggen dat er niet wordt gekeken naar of de leerling al eerder is geschorst.

# Indicatoren uitleg

## Onderwijsproces

### **Aantal aanvragen onderwijstijdvermindering (art.12)**

Deze indicator bekijkt het aantal aanvragen onderwijstijdvermindering conform art.12. Per OKE wordt het aantal aanvragen onderwijstijdonderschreidingen bekeken. Dit aantal wordt gedeeld door het aantal leerlingen op het OKE, om eerlijke vergelijking te kunnen maken tussen scholen van verschillende groottes. Enkel aanvragen t.a.v. WEC art. 12 worden meegenomen in de berekening (dus niet op basis van art. 39). Alleen toegewezen of afgewezen aanvragen tellen mee. De 15% scholen met de meeste aanvragen (rekening houdend met de schoolgrootte) lopen risico. Dit leidt ertoe dat alle scholen met een aanvraag tot onderwijstijdonderschreiding risico scoren op de indicator, aangezien momenteel minder dan 15% van de scholen in het VSO een onderwijstijdonderschreiding aanvraagt.

## Veiligheid en Schoolklimaat

### **Signalen indicatoren**

Er zijn twee indicatoren m.b.t. signalen:

- Signalen veiligheid
- VI signalen

De veiligheidssignalen telt hoeveel signalen ten aanzien van het onderwerp veiligheid er op een OKE zijn binnengekomen. Deze worden uitgesplitst naar urgente en niet-urgente signalen. Urgente signalen tellen dan 3 keer zo zwaar als een niet-urgent signaal

De VI signalen zijn een iets complexer geval, omdat slechts een klein deel van de signalen die bij de VI's binnen komen worden geregistreerd, en de registratie meestal later plaatsvindt (deels doordat de betrokkenen lang niet altijd toestemming geven te registreren en deels doordat er lange tijd slecht geregistreerd werd door de VI's).

Al deze indicatoren worden berekend door het aantal signalen in die categorie te delen door het aantal leerlingen op het OKE, keer 100. Dat levert een percentage op. In principe is besloten dat we de 15% scholen met het hoogste percentage signalen risico toekennen. Dit nu voor alle sectoren gelijk, maar wanneer er goede redenen voor zijn om hiervan af te wijken is dit mogelijk. Uit de data blijkt echter dat minder dan 15% van de OKE's signalen binnen krijgt. Daarom is het in de praktijk zo dat elk OKE waar een relevant signaal binnen komt, hier een risico scoort op de bijbehorende indicator.

Update PROZA 2020: Sinds de update van de signalen registratie naar Proza dient voorzichtig te worden omgegaan met deze indicator. In principe gaat alles goed en worden de dubbele signalen verwijderd. Er is echter nooit te wapenen tegen registratiefouten.

### In het kort:

Wanneer de gegevens per indicator bekend zijn, worden deze gebruikt om de totale risicoscore te berekenen. De indicatoren worden eerst gewogen. Het gewicht wat elke indicator krijgt wordt afgestemd tussen Kennis en de desbetreffende sector. Deze gewogen indicatoren worden omgezet naar de totale risicoscore. De totale risicoscore is een zeer abstract begrip. Het enige wat het doet, is de scholen binnen één prestatiemonitor rangschikken ten opzichte van elkaar. De totale risicoscore kan niet voor meer dan dat doel gebruikt worden.

In de prestatiemonitor is ook altijd een beknopte uitleg te vinden onder de Toelichting. Vergeet niet dat je kan scrollen in dit veld.

Toelichting indicatoren SO &amp; VSO 2022 -3

# Inleiding

In dit document geven wij een toelichting op de indicatoren van de Prestatiemonitor VSO 2022. Dit document is een aanvulling op de algemene [Toelichting prestatieindicator](#) en de toelichting per indicator zoals die in de Prestatiemonitor zelf te vinden is. Ook staat veel informatie over de indicatoren in de [Bijsluiter toelichting prestatieanalyse SO 2018](#), waar naar we in dit document zullen verwijzen.

## Inhoud

Inleiding.....	2
Scores berekening .....	3
Indicatoren uitleg .....	4
Onderwijsproces.....	4
Bestendiging Ontwikkelingsperspectief .....	6
Schoolklimaat en Veiligheid.....	4
Onderwijsresultaten .....	6
Kwaliteitszorg en Ambitie.....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Financieel Beheer .....	<b>Fout! Bladwijzer niet gedefinieerd.</b>
Context .....	9
Kanttekening.....	<b>Fout! Bladwijzer niet gedefinieerd.</b>