



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

## **HOE WORDT KWALITEITZORG BEOORDEELD?**

RESULTATEN VAN HET DERDE FAIRNESSONDERZOEK

Utrecht, juni 2020

## Voorwoord

Kunnen scholen, opleidingen en besturen erop vertrouwen dat de inspectie hen een fair oordeel geeft? Deze vraag is voor een toezichthouder, zoals de Inspectie van het Onderwijs, belangrijk. Het mag niet uitmaken welke inspecteur het onderzoek uitvoert. De inspectie doet daarom periodiek onderzoek naar de oordeelsvorming door inspecteurs in de zogeheten fairnessonderzoeken.

In dit rapport doen we verslag van het derde fairnessonderzoek naar de betrouwbaarheid en rechtvaardigheid van het inspecteursoordeel. We nemen de standaarden voor Kwaliteitszorg en ambitie (Kwaliteitszorg, Kwaliteitscultuur, en Verantwoording en dialoog) onder de loep. Deze standaarden vormen de kern van het huidige onderzoekskader en de kwaliteit van deze standaarden is van groot belang. We onderzoeken dan ook niet alleen of de oordelen van de inspecteurs fair waren, maar we kijken ook of de opbouw of inhoud van de standaarden te verbeteren is. Dit onderzoek is een onderdeel van de evaluatie van het onderzoekskader 2017 met het oog op de aanpassing van het onderzoekskader in 2021.

Voor dit onderzoek delen we de standaarden voor Kwaliteitszorg en ambitie op in onderliggende items die normaal gesproken niet apart beoordeeld worden. We analyseren de inspecteursoordelen op de onderliggende items en de relatie met inspecteursoordelen op de standaard als geheel. Ook verzamelen we gegevens over wat inspecteurs misten in de standaarden en over wat zij zagen als eigen aspecten van kwaliteit.

Met dit rapport laten we zien dat we blijvend aandacht houden voor fairness van onze oordelen. Op basis van dit rapport kunnen we intern het gesprek blijven voeren. Daarnaast worden de aanbevelingen gebruikt voor de herziening van het onderzoekskader vanaf augustus 2021. Ook de komende jaren zullen we aandacht besteden aan wat we nog kunnen verbeteren aan onze oordeelsvorming.

Esther Deursen  
Waarnemend Inspecteur-generaal van het Onderwijs

## INHOUD

Voorwoord 2  
Samenvatting 4

**1 Inleiding 6**  
1.1 Leeswijzer 6

**2 De opzet van het onderzoek 7**  
2.1 Onderzoeksdesign 7  
2.2 Dataverzameling en respons 7  
2.3 Data-analyse 8

**3 Oordelen op de standaarden voor Kwaliteitszorg en Ambitie 9**

**4 De betrouwbaarheid van het inspecteuroordeel 11**  
4.1 Betrouwbaarheid 11  
4.2 Itemscores en de oordelen 11  
4.3 Items die vaker en minder vaak voldoende scores 12  
4.4 De cesuur 14

**5 Opbouw en inhoud van de standaarden 16**  
5.1 Indeling van de standaarden 16  
5.2 Inhoud van de standaarden 16  
5.3 Wat misten inspecteurs in de standaarden 17  
5.4 Eigen aspecten van kwaliteit 18

**6 Conclusies 19**  
6.1 Vooruitblik 19

**7 Literatuurlijst 21**

**Bijlage I: Het instrument 22**

**Bijlage II: Voorbeeldresultaten factoranalyse 26**

**Bijlage III: Resultaten betrouwbaarheidsanalyse 29**

**Bijlage IV: Resultaten scorepatronen 30**

**Bijlage VI: Standaarden voor interbeoordelaarsbetrouwbaarheid (IBB) 32**

**Colofon 33**

## Samenvatting

De kern van een faire werkwijze van de Inspectie van het Onderwijs is dat scholen erop moeten kunnen vertrouwen dat zij eerlijk beoordeeld worden. Scholen met gelijke kwaliteit zouden hetzelfde oordeel moeten krijgen. Bovendien zouden verschillende (teams van) inspecteurs tot hetzelfde oordeel moeten komen na het onderzoeken van dezelfde school. Deze redenering geldt ook voor de waardering Goed.

In de fairnessonderzoeken onderzoekt de Inspectie van het Onderwijs periodiek de oordeelsvorming van de inspecteurs. Dit is het derde onderzoek in die reeks. We richten ons nu uitsluitend op de standaarden voor kwaliteitszorg en ambitie (KA1 t/m KA3: Kwaliteitszorg, Kwaliteitscultuur en Verantwoording en dialoog). Deze standaarden vormen de kern van het onderzoekskader 2017.

We onderzoeken de betrouwbaarheid van de inspecteursoordelen. Ook onderzoeken we of de inhoud en opbouw van de KA-standaarden verbeterd kunnen worden. Dit onderzoek is één van evaluaties van het onderzoekskader 2017 ten behoeve van de aanpassing van het onderzoekskader in 2021.

Net zoals in eerdere fairnessonderzoeken heeft de inspectie inspecteurs een tijd lang de standaarden meer gedetailleerd laten beoordelen door de standaard op te delen in losse items. Op basis van interviews met inspecteurs, literatuuronderzoek en juridisch advies zijn er ook een aantal items toegevoegd. Verder hebben inspecteurs vragen beantwoord over onderdelen die zij missen in de standaard en over eigen aspecten van kwaliteit.

### **Het beoordelen van standaarden**

Bij een regulier onderzoek bij scholen, instellingen of besturen beoordeelt de inspecteur de standaard enkel als geheel. De standaarden geven de essentie aan van goed onderwijs c.q. goede kwaliteitszorg/kwaliteitsborging, en het oordeel op de standaard is geen optelsom van oordelen op onderliggende tekstdelen. Door te kijken in hoeverre de onderliggende items zijn meegenomen in de onderbouwing kunnen we de onderlinge vergelijkbaarheid van de beoordelingen vaststellen.

De voornaamste conclusies en adviezen van het onderzoek volgen hieronder.

### **De betrouwbaarheid van het inspecteursoordeel**

De betrouwbaarheid van de items van elke standaard is voldoende tot zeer hoog. De scorepatronen laten zien dat een oordeel voldoende kan worden gegeven als niet alle items van de standaard voldoen. Er is een afwegingsgebied waarin het oordeel met hetzelfde aantal onvoldoendes zowel Voldoende als Onvoldoende kan zijn, maar dit gebied is maar klein. Dat is ook een gunstige indicatie voor de betrouwbaarheid van de oordelen.

De analyse van de scorepatronen laat ook zien dat er vrij veel items zijn, waar de meeste scholen en besturen aan voldoen. Er zijn relatief weinig items die waar een deel van de scholen en besturen met een Voldoende oordeel op de standaard nog niet aan voldoet. De wet stelt dus betrekkelijk weinig eisen op deze onderwerpen. Het is goed om te onderzoeken of de standaarden zo ingericht kunnen worden dat ze binnen de wettelijke mogelijkheden meer ruimte voor discussie bieden.

### **Indeling van de KA-standaarden behouden**

De items van elke KA-standaard meten meestal één dominant kenmerk van kwaliteitszorg en ambitie. Bij KA3 'Verantwoording en dialoog' blijkt dat minder sterk.

De standaarden vormen een voldoende consistent geheel en we zien dus op basis van de gegevens over dit kwaliteitsgebied geen redenen om de indeling van de standaarden aan te passen.

### **Apart beoordelen van items over wettelijk verplichte documenten of formele verplichtingen**

We zien dat inspecteurs de items behoorlijk consistent scoren en dat geeft vertrouwen in de standaarden. Toch kan de kwaliteit van de standaarden nog verbeterd worden door het toevoegen van goed functionerende aanvullende items die net als de huidige items gebaseerd zijn op de wet. De standaarden kunnen ook verbeterd worden door het schrappen of apart beoordelen van items die ruis veroorzaken in de beoordeling. Het gaat dan bijvoorbeeld om items die vragen om een wettelijke formaliteit of een document. Wij raden aan deze items apart te controleren en dus niet op te nemen in de standaarden.

### **Item over de relatie tussen kwaliteitszorg van het bestuur en de kwaliteitszorg op de school of de opleiding**

Inspecteurs geven vaak aan dat zij een item missen over de relatie tussen de beoogde kwaliteitszorg van het bestuur en de realisatie van kwaliteitszorg op scholen/opleidingen. Wij stellen voor dit onderwerp op te nemen in de standaarden.

### **Eigen aspecten van kwaliteit**

Wanneer de inspecteurs eigen aspecten van kwaliteit laten meewegen in hun waardering gaat het ofwel om een (zeer) goede uitvoering van een item van de standaard, ofwel om het nastreven van ambitieuze leerdoelen. Wij raden aan de voorbeelden van eigen aspecten van kwaliteit bij de afzonderlijke standaarden af te schaffen. Die leveren voor de inspecteur en het veld geen meerwaarde en zetten vooral op het verkeerde been.

### **Geef blijvende aandacht aan de professionele standaarden voor interbeoordelaarsbetrouwbaarheid**

We zien dat er al een goede betrouwbaarheid is. Dat is een verdienste die moet worden onderhouden en waar de inspectie blijvend aandacht aan moeten geven. Met dit doel voor ogen hebben we de richtlijnen voor interbeoordelaarsbetrouwbaarheid (IBB) geformuleerd (bijlage VI).

# 1 Inleiding

In het fairnessonderzoek onderzoekt de Inspectie van het Onderwijs periodiek de oordeelsvorming van de inspecteurs. Dit is het derde onderzoek in die reeks.<sup>1</sup> We richten ons nu uitsluitend op de standaarden voor Kwaliteitszorg en ambitie (KA1 t/m KA3: Kwaliteitszorg, Kwaliteitscultuur en Verantwoording en dialoog). Deze standaarden zijn van groot belang, omdat ze de kern vormen van het huidige onderzoekskader en niet alleen op school- of opleidingsniveau maar ook op bestuursniveau worden beoordeeld. Daarnaast is uit eerder fairnessonderzoek (Inspectie van het Onderwijs, 2018) gebleken dat met name KA1 lastig te beoordelen is.

In het Onderzoekskader 2017 van de inspectie wordt elke standaard voor de onderwijskwaliteit uitgewerkt met een beschrijving onderbouwd met elementen uit de wet. Deze beschrijving bestaat uit verschillende onderwerpen die gezamenlijk doorslaggevend kunnen zijn voor de beoordeling Onvoldoende of Voldoende. In geval van onvoldoende kwaliteit krijgt het bestuur een herstelopdracht. Deze beoordeling dient betrouwbaar en fair te zijn. De kern van een faire werkwijze van de Inspectie van het Onderwijs is dat scholen met dezelfde kwaliteit hetzelfde oordeel krijgen. Bovendien zouden verschillende (teams van) inspecteurs tot hetzelfde oordeel moeten komen na het onderzoeken van dezelfde school. Deze redenering geldt ook voor de waardering Goed.

Het is van belang om regelmatig onderzoek te doen naar de kwaliteit van het inspectietoezicht en de interbeoordelaarsbetrouwbaarheid. Dit onderzoek heeft daarom volgende onderzoeksvragen:

1. Zijn de oordelen op de KA-standaarden voldoende fair?
2. Hoe zou de omschrijving van het kwaliteitsgebied 'Kwaliteitszorg en ambitie' en van de standaarden Kwaliteitszorg (KA1), Kwaliteitscultuur (KA2) en Verantwoording en dialoog (KA3) verbeterd kunnen worden?

## 1.1 Leeswijzer

Dit onderzoek is één van evaluaties van het onderzoekskader 2017 ten behoeve van de aanpassing van het onderzoekskader in 2021. Hieronder bespreken we eerst de opzet van het onderzoek. Daarna gaan we in op hoe fair de oordelen op de KA-standaarden. Ook bespreken we de bevindingen en adviezen over de inhoud en opbouw van de standaarden. Tot slot bespreken we de conclusies, mogelijkheden voor vervolgonderzoek en de richtlijnen voor fairness van het inspecteursoordeel, gebaseerd op het huidige en eerder fairnessonderzoek.

---

<sup>1</sup> In 2016 en 2018 zijn de rapporten verschenen van de twee voorgaande fairnessonderzoeken (Inspectie van het Onderwijs, 2016, 2018).

## 2 De opzet van het onderzoek

### 2.1 Onderzoeksdesign

Voor dit onderzoek hebben we de standaarden voor Kwaliteitszorg en ambitie in afzonderlijke onderdelen (items) gesplitst. Daarnaast zijn aanvullende items toegevoegd die mogelijk een inhoudelijke verbetering zijn. De aanvullende items zijn gekozen op basis van interviews met inspecteurs, literatuuronderzoek en juridisch advies. We vroegen aan inspecteurs die onderzoeken doen bij besturen of scholen om naast de reguliere beoordeling van de standaard als geheel, dit instrument in te vullen (in de periode mei t/m oktober 2019). Het betreft dus steeds reguliere bezoeken voor het toezicht. Op die manier verzamelden we informatie over de beoordeling van afzonderlijke onderdelen van de kwaliteitszorgstandaarden, die normaal gesproken als geheel worden beoordeeld. Ook stelden we open vragen over welke onderdelen inspecteurs misten in de standaard en over welke eigen aspecten van kwaliteit een rol speelden in de beoordeling voor de waardering Goed. Met de analyse gaan we na of de oordelen op de standaard fair zijn. Ook gaan we na of de standaarden consistent zijn en welke onderdelen minder goed bij de standaard passen en daardoor mogelijk de beoordeling verstoren.

De analysemethoden die we hier toepassen worden vaak gebruikt voor het beoordelen van toetsinstrumenten met losse items. Bij een regulier onderzoek bij scholen, instellingen of besturen beoordeelt de inspecteur de standaard echter enkel als geheel. We splitsen de standaarden hier dus uitzonderlijk op in losse onderdelen omdat dit ons inzicht geeft in het tot stand komen van de oordelen op de standaard als geheel en de evaluatie van afzonderlijke onderdelen. Het kan dan ook voorkomen dat een school het oordeel Voldoende krijgt op een standaard terwijl niet alle losse items een voldoende scoren. Een school, opleiding of bestuur laat dan een positief beeld zien op het met de standaard te beoordelen kenmerk van kwaliteitszorg, terwijl aan een bepaald wettelijk element van de standaard niet wordt voldaan. Als er een zwaarwegend argument is, dan kan de inspecteur toch een Voldoende voor de standaard geven. Dit is de discretionaire ruimte van een inspecteur.

In Bijlage I staat het instrument dat is gebruikt voor dit fairnessonderzoek voor po, vo en so.

### 2.2 Dataverzameling en respons

In totaal is informatie ontvangen over 223 beoordelingen (Tabel 2.2a). De verdeling over de sectoren is als volgt.

Tabel 2.2a. De respons (in aantallen)

Sector	Bestuur	School/opleiding
Primair onderwijs (po)	39	53
Speciaal onderwijs (so)	2	6
Voortgezet onderwijs (vo)	21	28
Mbo	15	46
Samenwerkingsverband	13	-
Totaal	90	133

Bij de analyse nemen we de sectoren po, vo en so samen (in totaal 62 besturen en 87 scholen), omdat de standaarden voor deze sectoren vrijwel hetzelfde zijn. Mbo en Samenwerkingsverbanden hebben een meer afwijkende omschrijving van de kwaliteitsstandaarden en worden apart geanalyseerd.

### 2.3 Data-analyse

We onderzoeken de relatie tussen itemscores en oordelen op de standaard. Daarbij letten we ook op het aantal onvoldoendes voor items dat leidt tot een Onvoldoende oordeel op de standaard en of er items zijn die bij een Voldoende op de standaard nog vaak een onvoldoende scoren.

Verder onderzoeken we of de items van de standaarden goed bij de standaarden passen, of de elke standaard één of meerdere kenmerken van kwaliteitszorg meet en of er items zijn die ruis vertonen en dus niet consistent worden beoordeeld door de inspecteurs. Dit doen we met een factoranalyse en een betrouwbaarheidsanalyse.<sup>2</sup> Met de analyses bekijken we of de verschillende onderdelen van elke standaard een consistente eenheid vormen en of het toevoegen van de aanvullende onderdelen of juist het weglaten van bepaalde onderdelen de standaard zou kunnen verbeteren. Naast deze kwantitatieve analyses is er een analyse gedaan van de antwoorden op de open vragen over wat inspecteurs misten in de standaard en over eigen aspecten van kwaliteit die inspecteurs hebben laten meewegen in hun oordeel. We deden de analyses eerst bij de po-vo-so besturen en daarna bij de scholen. Dezelfde analyses zijn gedaan voor mbo en voor samenwerkingsverbanden. Aangezien er voor deze laatste twee minder gegevens beschikbaar waren dan voor po, vo en so, zijn de analyses minder betrouwbaar en daarom leunen we daar meer op inhoudelijke argumenten.

---

<sup>2</sup> Met een factoranalyse kun je voor een groter aantal variabelen een kleiner aantal achterliggende factoren identificeren (zie bijvoorbeeld Field et al., 2014). Dat wil in ons geval zeggen dat we op basis van de scorepatronen op de items kunnen bepalen welke groepen van items dezelfde achterliggende factor oftewel hetzelfde kenmerk van kwaliteitszorg en ambitie vertegenwoordigen. Met de betrouwbaarheidsanalyse bepalen we de betrouwbaarheid (Cronbachs alfa) van de set items van een standaard. Hiermee kunnen we zien of er op elke standaard consistent gescoord wordt en of er items zijn die het patroon verstoren.



### 3 Oordelen op de standaarden voor Kwaliteitszorg en Ambitie

De oordelen op de standaarden voor Kwaliteitszorg en Ambitie werden gegeven binnen de dataverzamingsperiode. Het betreft dus niet een heel schooljaar en ook geen representatief beeld.<sup>3</sup> Tabellen 3a t/m 3c tonen de oordelen op de standaarden op bestuursniveau en opleidingsniveau of schoolniveau (indien van toepassing). Tabel 3a laat zien dat er voor po, vo en so enige differentiatie in de oordelen zit. Dat geldt vooral voor KA1. Bij KA2 zijn er al een stuk minder Onvoldoendes gegeven, maar werd wel vrij vaak de waardering Goed gegeven. Bij KA3 zijn er nauwelijks Onvoldoendes en is er weinig differentiatie in de oordelen. KA3 werd op schoolniveau een stuk minder vaak beoordeeld dan KA1 en KA2.

Tabel 3a. Beoordelingen po vo en so per standaard

Standaard		Scholen		Besturen		Totaal	
		n	%	n	%	n	%
KA1 Kwaliteitszorg	Onvoldoende	16	23	13	21	29	22
	Voldoende	51	72	43	70	94	71
	Goed	4	6	6	10	10	8
<b>Totaal</b>		71	100	61	100	132	100
KA2 Kwaliteitscultuur	Onvoldoende	5	7	3	5	8	6
	Voldoende	42	63	41	69	83	66
	Goed	20	30	15	25	35	28
<b>Totaal</b>		67	100	59	100	126	100
KA3 Verantwoording en dialoog	Onvoldoende	1	3	1	2	2	2
	Voldoende	29	88	51	86	80	87
	Goed	3	9	7	12	10	11
<b>Totaal</b>		33	100	59	100	92	100

Voor de sector mbo zijn er wat minder gegevens beschikbaar dan voor de sectoren po, vo en so samen. Op bestuursniveau ontbreken vooral beoordelingen van KA2 omdat deze standaard bij niet bekostigde instellingen niet wordt beoordeeld. Omdat we bij KA2 en KA3 soms verschillende items hebben toegevoegd voor bestuur en opleiding beperkt ook dat de analysemogelijkheden. Er is wel variatie in de oordelen (Tabel 3b). Er wordt duidelijk onderscheid gemaakt in de kwaliteitszorg van de verschillende besturen en opleidingen.

<sup>3</sup> Een volledig beeld van de oordelen van alle onderzochte besturen en scholen is te vinden in de Voortgangsrapportage 2018/2019 (Inspectie van het Onderwijs 2020).

Tabel 3b. Beoordelingen mbo per standaard

Standaard		Opleidingen		Besturen		Totaal	
		n	%	n	%	n	%
KA1 Kwaliteitszorg	Onvoldoende	14	32	3	20	17	29
	Voldoende	27	61	12	80	39	66
	Goed	3	7	0	0	3	5
<b>Totaal</b>		44	100	15	100	59	100
KA2 Kwaliteitscultuur	Onvoldoende	4	10	2	25	6	12
	Voldoende	31	76	6	75	37	76
	Goed	6	15	0	0	6	12
<b>Totaal</b>		41	100	8	100	49	100
KA3 Verantwoording en dialoog	Onvoldoende	8	25	2	13	10	21
	Voldoende	23	72	11	73	34	72
	Goed	1	3	2	13	3	6
<b>Totaal</b>		32	100	15	100	47	100

De samenwerkingsverbanden voor passend onderwijs hebben een aantal bij wet vastgelegde taken om voor leerlingen met een extra ondersteuningsbehoefte (zorgleerlingen) een optimaal passend onderwijsaanbod te realiseren. Samenwerkingsverbanden beschikken daarvoor over een budget. In de onderzoekskaders van de sectoren po, vo en so is een apart waarderingskader samenwerkingsverbanden beschreven.

In totaal zijn er 13 ingevulde beoordelingen met onderbouwing teruggekomen. Dit lijkt niet zoveel, maar in de periode dat de dataverzameling liep, zijn over een zeer groot deel van de bezoeken de beoordelingen ingevuld (er zijn 24 onderzoeken vermeld in het systeem van de inspectie in de periode mei tot november 2019). Bij KA1 is er variantie in de oordelen en bij KA2 enigszins. KA3 kreeg altijd het oordeel Voldoende (Tabel 3c).

Tabel 3c. Beoordelingen Samenwerkingsverbanden per standaard

Standaard		n	%
KA1 Kwaliteitszorg	Onvoldoende	5	38
	Voldoende	7	54
	Goed	1	8
<b>Totaal</b>		13	100
KA2 Kwaliteitscultuur	Onvoldoende	2	15
	Voldoende	10	77
	Goed	1	8
<b>Totaal</b>		13	100
KA3 Verantwoording en dialoog	Onvoldoende	0	0
	Voldoende	13	100
	Goed	0	0
<b>Totaal</b>		13	100

We zien bij po, vo, so, mbo en samenwerkingsverbanden dat met name KA1 discrimineert tussen een onvoldoende, voldoende en goede kwaliteit. KA2 en KA3 maken iets minder onderscheid, met name bij po, vo en so.

## 4 De betrouwbaarheid van het inspecteuroordeel

Scholen/instellingen en besturen moeten erop kunnen vertrouwen dat de inspecteuroordelen fair zijn. Om te onderzoeken hoe fair en betrouwbaar de gegeven oordelen waren analyseerden we:

- De betrouwbaarheid van de itemsets van de KA-standaarden (KA1: Kwaliteitszorg, KA2: Kwaliteitscultuur, en KA3: Verantwoording en dialoog).
- De relatie tussen de itemscores en de oordelen over de standaarden.

### 4.1 Betrouwbaarheid

De betrouwbaarheid van de standaarden werd bepaald voor de itemset voor en na toevoeging van aanvullende items, en voor de itemset van het eindvoorstel. Het eindvoorstel bevatte de op basis van dit onderzoek geselecteerde set van te behouden en toe te voegen items. Meestal vonden we hoge scores van >.90 voor de betrouwbaarheid van de standaarden (Tabel 1 t/m 3 in Bijlage III) (Evers et al., 2009). Hoewel inspecteurs in de praktijk geen afzonderlijke onderdelen scoren, maar alleen een standaard in zijn geheel, geeft deze onderzoeksbevinding vertrouwen in de oordelen. Toch is een woord van relativering is op zijn plaats. In de eerste plaats hebben inspecteurs de neiging om de items conservatief en daardoor homogeen te oordelen. We bedoelen daarmee dat een inspecteur pas een item een onvoldoende geeft als hij er volledig zeker van is en dat twijfel meestal als voldoende uitpakt. Dit kan de betrouwbaarheid kunstmatig ophogen. Daarnaast is het mogelijk dat inspecteurs de standaard eerst globaal beoordelen en daarna pas de afzonderlijke items. Ook dit kan de betrouwbaarheid verhogen.

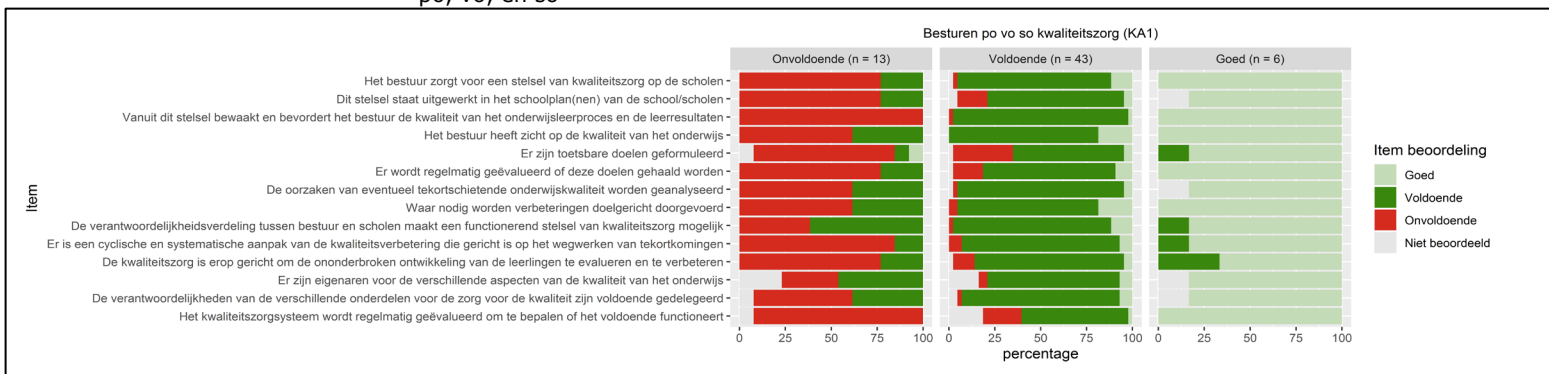
### 4.2 Itemscores en de oordelen

De analyses van de relatie tussen itemscores en scores op de standaard ondersteunen het beeld dat de oordelen voldoende betrouwbaar zijn. Bij deze analyses keken we naar het aantal items met de score onvoldoende, voldoende of goed in relatie tot het eindoordeel. In Figuur 4.1a en 4.1b hieronder zien we twee voorbeelden van de scorepatronen bij po, vo en so. Figuur 4.1a toont het scorepatroon voor KA1 op bestuursniveau bij po, vo en so, en Figuur 4.1b toont het scorepatroon voor KA2 op schoolniveau bij po, vo, en so. De figuren laten per item zien hoe vaak elke score voorkwam bij een standaardbeoordeling Onvoldoende of Voldoende, of bij de waardering Goed. In de linker kolom staan de itemscores wanneer de standaard het oordeel Onvoldoende kreeg. In de middelste kolom de scores op de items bij de Voldoende standaarden en in de rechterkolom de itemscores bij de waardering Goed.<sup>4</sup>

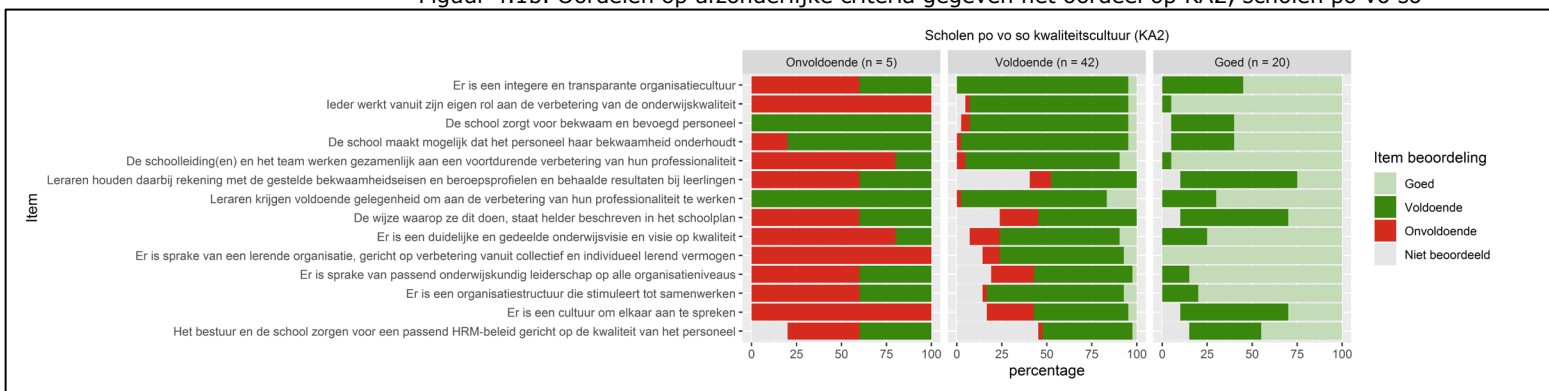
---

<sup>4</sup> De overige figuren met scorepatronen staan in Bijlage IV. De gegevens zijn in tabelvorm op te vragen bij de Inspectie van het onderwijs.

Figuur 4.1a. Oordelen op afzonderlijke items gegeven het oordeel op standaard KA1, Besturen po, vo, en so



Figuur 4.1b. Oordelen op afzonderlijke criteria gegeven het oordeel op KA2, scholen po vo so



Figuur 4.1a en 4.1b laten zien dat wanneer de inspecteurs de standaard een Onvoldoende geven, er vaak veel onderdelen niet voldoen (minimaal 3). Wanneer een standaard de waardering Goed krijgt dan zijn er vaak meerdere onderdelen goed. Dit geldt ook voor de andere KA-standaarden bij po, vo, so, mbo en samenwerkingsverbanden op zowel school/instillings- als bestuursniveau. Het komt bovendien vrijwel niet voor dat de waardering Goed is gegeven wanneer onderliggende items onvoldoende zijn.

### 4.3 Items die vaker en minder vaak voldoende scores

Het komt voor dat de standaard een Voldoende krijgt terwijl er één of meerdere items niet voldoen, en het valt op dat sommige items bij het oordeel Voldoende op de standaard vaker een onvoldoende scoren dan andere items. Mogelijke verklaringen hiervoor zijn:

- Sommige items veroorzaken een ruis in de beoordeling en krijgen vaker een afwijkende score, bijvoorbeeld omdat zij minder goed passen bij de standaard of moeilijker te beoordelen zijn.
- We gaan ervanuit dat een standaard voor kwaliteitszorg en ambitie een mate van vaardigheid van het bestuur of de school op het gebied van kwaliteitszorg meet. Die vaardigheid kan groter of kleiner zijn. De items meten gezamenlijk de vaardigheid van een bestuur of school op het onderwerp van de standaard. Om een voldoende te scoren voor een item is een bepaalde mate van vaardigheid in kwaliteitszorg nodig, en voor het behalen van een voldoende op sommige items is er meer vaardigheid nodig dan bij andere items.

Items die het scorepatroon verstoren kunnen vaak beter worden geschraapt of verplaatst naar een andere standaard (zie ook paragraaf 5.2), maar een goede balans tussen items die vaker en minder vaak een onvoldoende score is wenselijk om ruimte te geven voor discussie tussen inspecteurs over het Oordeel en daarmee goed te kunnen discrimineren tussen een onvoldoende en voldoende kwaliteitszorg. Items die bij een Voldoende op de standaard meer dan 10% van de keren een onvoldoende krijgen beschouwen we als opvallend. Dit zijn er maar weinig. Voorbeelden zijn: 'Er zijn toetsbare doelen geformuleerd' en 'Er wordt regelmatig geëvalueerd of deze doelen gehaald worden' uit KA1 po, vo, en so (zie Figuur 4.1a). Het formuleren van concrete toetsbare doelen en het evalueren van het behalen van die doelen blijken aspecten van kwaliteitszorg te zijn waaraan niet snel wordt voldaan. Ook bij mbo op bestuursniveau bleek een item over toetsbare doelen te behoren bij de groep items met >10% onvoldoendes bij een Voldoende standaard. Bij mbo op opleidingsniveau was 'De oorzaken van eventueel tekortschietende onderwijskwaliteit worden geanalyseerd en waar nodig worden verbeteringen doorgevoerd' (KA1) een voorbeeld van een item dat nog relatief vaak een onvoldoende kreeg als de standaard het oordeel Voldoende kreeg. Bij KA2 (po, vo, so op schoolniveau en bestuursniveau) kregen een aantal aanvullende items minder snel een voldoende (Figuur 4.1b):

- Er is een duidelijke en gedeelde visie op onderwijskwaliteit
- Er is sprake van passend onderwijskundig leiderschap op alle organisatieniveaus
- Er is een cultuur om elkaar aan te spreken

Deze aanvullende items toetsen de concretisering en de realisatie van de kwaliteitscultuur en vragen gericht naar uitwerking. Dit lijken de onderdelen van de standaard te zijn waar minder vaak aan wordt voldaan. Bij mbo is bij KA2 'Op alle niveaus is men aanspreekbaar op gemaakte afspraken en wordt dit ook van anderen gevraagd' een voorbeeld dit type item.

De meerderheid van de items voldoet wel vaak, soms ook als de standaard als geheel een Onvoldoende oordeel krijgt (er zijn dan andere items die niet voldoen). 'De school zorgt voor bekwaam en bevoegd personeel' (KA2, po vo so) is bijvoorbeeld een dergelijk item (Figuur 4.1b). Nog een voorbeeld van een item dat relatief vaak voldoet is: 'De leraren (die onder dit bestuur vallen) krijgen voldoende gelegenheid om aan de verbetering van hun professionaliteit te werken' (KA2). Het kreeg bij po, vo, en so maar een enkele onvoldoende op school- en bestuursniveau. Dit type items discrimineert niet sterk tussen een Onvoldoende en Voldoende kwaliteitszorg, zij krijgen immers ook bij een Onvoldoende oordeel voor de standaard vaak een voldoende.

Met name op het gebied van kwaliteitscultuur (KA2) en verantwoording en dialoog (KA3) zijn er betrekkelijk weinig wettelijke eisen. Het is goed om te onderzoeken of de wet ruimte biedt om in de standaarden meer balans te brengen tussen items over kwaliteitszorg, kwaliteitscultuur en verantwoording en dialoog waaraan vaker en minder vaak wordt voldaan. Met name de toegevoegde items bij KA2 zoals 'Er is een cultuur om elkaar aan te spreken', kunnen een stap in de goede richting te zijn. De aanvullende items moeten net als de huidige onderdelen van de standaard gebaseerd zijn op de wet en volgen uit de algemene kwaliteitszorgplicht die voor de besturen geldt. Het gaat er hier uitdrukkelijk niet om dat er meer Onvoldoendes voor de standaard komen, maar wel dat een standaard meer onderdelen bevat waaraan niet alle besturen en scholen voldoen zodat er meer ruimte voor discussie ontstaat discussie met het veld over de beoordeling en ook tussen inspecteurs onderling. Hierdoor wordt het toezicht stimulerender en het komt de betrouwbaarheid ten goede.

#### 4.4

#### De cesuur

Het komt voor dat een standaard het oordeel Voldoende krijgt als enkele onderliggende items niet voldoen. Bij een bepaald aantal onvoldoendes op de onderliggende items is er een gebied waarbij het oordeel met hetzelfde aantal onvoldoendes zowel Onvoldoende als Goed kan zijn. Dit afwegingsgebied betreft een klein deel van de beoordelingen. Bij mbo is dat ongeveer in 10% van de gevallen, bij po, vo en so is dat zelfs nog lager ca. 6%.<sup>5</sup> Bij de samenwerkingsverbanden kwam het niet voor dat er met een gelijk aantal onvoldoendes verschillende oordelen werden gegeven.

Het afwegingsgebied is klein, maar we zien bij een aantal standaarden wel dat er een fors aantal onvoldoende items moet zijn (gemiddeld 6) alvorens het oordeel onvoldoende wordt. Ter illustratie, Tabel 4.1a toont het aantal oordelen Onvoldoende, Voldoende en Goed dat werd gegeven per aantal onvoldoende itemscores voor KA1 bij scholen po, vo, en so. Pas vanaf 7 onvoldoendes voor de items werd er een oordeel onvoldoende gegeven voor de standaard. Tabel 4.1b toont voor KA1, besturen po, vo, en so, een groter gebied waarin met een zelfde aantal onvoldoendes het oordeel zowel voldoende als goed kan zijn. Het oordeel Onvoldoende op de standaard is één keer gegeven bij 3 onvoldoendes op de items. De andere Onvoldoende oordelen waren echter pas bij 9 of 10 onvoldoende items.

Tabel 4.1a. Aantal onvoldoendes voor de items gegeven het oordeel op de standaard KA1, scholen po vo so

Aantal onvoldoendes	Oordeel op de standaard KA1			Totaal
	Onvoldoende	Voldoende	Goed	
0	0	28	4	32
1	0	10	0	10
2	0	7	0	7
3	0	1	0	1
5	0	3	0	3
6	0	1	0	1
7	3	1	0	4
8	2	0	0	2
9	3	0	0	3
10	2	0	0	2
11	3	0	0	3
13	2	0	0	2
14	1	0	0	1
	16	51	4	71

Tabel 4.1b. Aantal onvoldoendes voor de items gegeven het oordeel op de standaard KA2, besturen po vo so

Aantal onvoldoendes	Oordeel op de standaard KA2 bestuursniveau			Totaal
	Onvoldoende	Voldoende	Goed	
0	0	12	15	27
1	0	13	0	13
2	0	7	0	7
3	1	4	0	5
4	0	3	0	3
5	0	1	0	1
6	0	1	0	1
9	1	0	0	1
10	1	0	0	1
	3	41	15	59

<sup>5</sup> Dit valt binnen een acceptabele marge, want een dergelijke omvang van het onzekere gebied zie je ook in andere takken van de besliskunde en beoordeling (bijvoorbeeld testbetrouwbaarheid, selecties, en dergelijke) (Zie bijvoorbeeld Dousma & Horsten, 1980, p. 139; Drenth & Sijtsma, 2006).

Het aantal onvoldoende items dat leidt tot een oordeel Onvoldoende op de standaard is dus gemiddeld hoog. Dit roept de vraag op of we de cesuur soms niet te laag leggen. Ook de bevinding dat er bij KA2 en KA3 vaak een gering aantal Onvoldoende oordelen voor de standaarden werden gegeven wijst in deze richting. Bij po, vo en so was dit bij KA2 6% en bij KA3 2%, versus 22% bij KA1 (zie Tabel 3a in hoofdstuk 3). Dit kan echter ook verklaard worden doordat de wet betrekkelijk weinig eisen stelt op de met KA2 en KA3 beoordeelde kenmerken. Ook kan het zijn dat deze standaarden minder aandacht krijgen. Bij mbo is er meer differentiatie in de oordelen; het percentage onvoldoendes voor KA2 en KA3 is wat hoger dan bij po, vo, so. Dit kan het gevolg zijn van inhoudelijke verschillen in de standaarden voor mbo en voor po, vo en so, maar ook door een verschil in de cesuur. Dit leidt tot het idee dat in een volgend fairnessonderzoek vooral inhoudelijk naar de cesuur gekeken moet worden: Wanneer zijn argumenten doorslaggevend om een onvoldoende te geven en is iedereen het eens over deze grenzen?

## **5 Opbouw en inhoud van de standaarden**

We vatten een standaard op als een kenmerk van onderwijskwaliteit dat onderbouwd wordt met elementen uit de wet waarbij dit kenmerk in de praktijk van de school of het bestuur goed ontwikkeld, minder goed ontwikkeld of zelfs onvoldoende kan zijn. Een standaard moet verder ook consistent en eenduidig zijn. We bedoelen daarmee dat alle onderdelen van de standaardomschrijving bij moeten dragen aan de beoordeling van een kenmerk van kwaliteitszorg en geen ruis bij de beoordeling mogen veroorzaken. In dit hoofdstuk gaan we in op de indeling van de standaarden en de inhoud ervan.

### **5.1 Indeling van de standaarden**

Om te onderzoeken of de items van elke afzonderlijke standaard een eigen kenmerk van kwaliteitszorg en ambitie meten, hebben we waar mogelijk een factoranalyse van alle items van alle standaarden samen gedaan. Deze factoranalyses tonen aan dat de onderdelen van elke KA-standaard meestal één dominant kenmerk van kwaliteitszorg en ambitie meten. Dit is een gunstige indicatie dat de standaarden elk een eigen onderdeel van de zorg voor de kwaliteit uitdrukken. KA1 (Kwaliteitszorg) gaat vooral over het stelsel van kwaliteitszorg, en KA2 (Kwaliteitscultuur) over aansturing van de vormgeving van kwaliteitszorg. Bij KA3 (Verantwoording en dialoog) is dat minder duidelijk. Daar heeft men bij po, vo en so op bestuursniveau te maken met twee onderdelen, namelijk verantwoording en rapportage versus het organiseren van dialoog en tegenspraak (Zie Bijlage II voor een voorbeeld van de analyse). Bij mbo zagen we niet meer dan één dimensie. We zien geen doorslaggevende reden om de standaard op te splitsen.

De standaarden vormen een voldoende consistent geheel. We zien dan ook geen redenen om de indeling van de standaarden in de drie gebieden (kwaliteitszorg, kwaliteitscultuur en verantwoording en dialoog) aan te passen.<sup>6</sup>

### **5.2 Inhoud van de standaarden**

Met de analyses konden we items identificeren die minder goed bij de betreffende standaard passen of die ruis veroorzaken in de beoordeling. Het schrappen van deze items kan de kwaliteit van de standaarden verhogen. Het gaat hier bijvoorbeeld om items die inhoudelijk sterk op elkaar lijken. Uit de betrouwbaarheidsanalyse (Bijlage III) blijkt dat de onderdelen van de standaard meestal heel consistent worden beoordeeld (de betrouwbaarheid van de itemsets van elke standaard was voldoende tot zeer hoog). Dat is een gunstige indicatie van de betrouwbaarheid van de standaarden, maar het kan er ook op duiden dat de items te beperkt geformuleerd zijn en inhoudelijk te veel op elkaar lijken (Streiner, 2003). Het is daarom goed om items die nagenoeg hetzelfde vragen als een ander item te schrappen.

Verder was opvallend dat onderdelen die gaan over de aanwezigheid van een wettelijk verplicht document of over wettelijk verplichte formaliteiten vaak minder goed bij het kenmerk dat wordt beoordeeld met de standaard passen. Voorbeelden van dit type onderdelen (bij po, vo en so) zijn:

---

<sup>6</sup> In dit onderzoek zijn enkel de standaarden van het kwaliteitsgebied 'Kwaliteitszorg en ambitie' onder de loep genomen. Er kunnen dus geen uitspraken worden gedaan over het eventueel samenvoegen met standaarden van een ander kwaliteitsgebied.



- Het bestuur handelt volgens de Code Goed Bestuur (KA2);
- De wijze waarop de leraren die onder dit bestuur vallen de bekwaamheid onderhouden staat helder beschreven in het schoolplan (KA2).

Bij deze onderdelen is het vaak lastig om het doel van het item (bijvoorbeeld de zorg voor kwaliteit) te scheiden van de formaliteit (het hebben van een schoolgids). Wanneer dergelijke onderdelen in de standaard staan leiden ze vaker tot verwarring in de oordeelsvorming. Er ontstaan dan verschillen tussen inspecteurs en ruis in de beoordeling. Een inspecteur wordt gevraagd om een standaard te beoordelen op een continuüm van kwaliteit en deze zeer feitelijk elementen verstoren die beoordeling. Inspecteurs moeten dan afwegen of het oordeel vanwege de aan- of juist afwezigheid van dat formele document nu de ene of de andere kant op moet. De Kwaliteitszorg kan bijvoorbeeld zeer goed zijn uitgevoerd terwijl het schoolplan niet voldoet aan alle wettelijke eisen. Meer passend in het bestuursgericht toezicht is wanneer men alleen apart toetst of dergelijke documenten aanwezig zijn. Uiteraard volgt uit het ontbreken van een dergelijk document dat het er moet komen. Wanneer men een standaard moet beoordelen dan komt indien nodig de inhoud van een document (een schoolgids, een schoolplan, enz.) vanzelf aan bod, maar de aanwezigheid van het document op zich is geen item voor de waardering van de standaard. Ter illustratie, items als 'Het bestuur zorgt voor een stelsel van kwaliteitszorg op de scholen' of 'Er zijn toetsbare doelen geformuleerd' kunnen beoordeeld worden zonder dat er een formeel document aanwezig is waarin het stelsel van kwaliteitszorg en de toetsbare doelen staan beschreven.

Naast items die geschrapt konden worden waren er ook juist aanvullende items die konden worden toegevoegd aan de standaard. Voorbeelden hiervan zijn:

- 'Het kwaliteitszorgsysteem wordt regelmatig geëvalueerd om te bepalen of het voldoende functioneert' (KA1: kwaliteitszorg)
- 'De verantwoordelijkheden van de verschillende onderdelen voor de zorg voor de kwaliteit zijn goed gedelegeerd' (KA1: Kwaliteitszorg)
- Er is een cultuur om elkaar aan te spreken (KA2: Kwaliteitscultuur)
- Er is sprake van een lerende organisatie, gericht op verbetering vanuit collectief en individueel lerend vermogen (KA2: Kwaliteitscultuur)
- Er is sprake van passend onderwijskundig leiderschap op alle organisatieniveaus (KA2: Kwaliteitscultuur)
- De interne dialoog over het te voeren beleid functioneert naar behoren (KA3: Verantwoording en dialoog)

Het toevoegen van deze items werd veelal door de analyses ondersteund en de items voegen ook inhoudelijk iets toe aan de standaard. De aanvullende items bij KA2 voegen bijvoorbeeld inhoud toe over de concrete realisatie van kwaliteitscultuur. Met het schrappen of toevoegen van items kunnen we de kwaliteit van de standaarden, en daarmee de betrouwbaarheid van de standaarden, verhogen.

### 5.3 Wat misten inspecteurs in de standaarden

Naast itemscores en oordelen op de standaarden gaven inspecteurs antwoord op de vraag welk onderdeel zij lieten meewegen in hun oordeel dat nog niet in de standaarden stond. In alle sectoren werd vaak de relatie tussen de kwaliteitszorg op bestuursniveau en de praktijk op de school of de opleiding genoemd. Voorbeelden van antwoorden van inspecteurs zijn:

*'Ik mis de vraag in hoeverre het beleid van het bestuur bekend is en doorwerkt bij de scholen. Kwaliteitszorg is meer dan een mooi plaatje op papier en een goede structuur. Het moet ook zichtbaar en voelbaar zijn op de werkvloer.'* (po)

*'Ik mis de doorwerking van het beleid.'*(vo)

Dit type item past goed bij bestuursgericht toezicht en wij adviseren dan ook dit onderwerp op te nemen in de beschrijving van standaard KA1.

#### **5.4 Eigen aspecten van kwaliteit**

In het huidige onderzoekskader kunnen naast de onderdelen van de standaard voor de waardering Goed ook eigen aspecten van kwaliteit worden meegewogen. We hebben inspecteurs ook de vraag gesteld of zij eigen (niet wettelijk verplichte) aspecten van kwaliteit hebben gewaardeerd en welke die aspecten waren. De meeste antwoorden op de vraag over eigen aspecten die zijn meegewogen voor de waardering Goed gingen over een uitzonderlijk goede uitvoering van de standaard of een onderdeel daarvan. Dit werd gezien als een eigen aspect van kwaliteit. Bij KA1 gaat het bijvoorbeeld om een goede kwaliteitscyclus of externe feedback, bij KA2 om een zeer goed vormgegeven cultuur of sterk onderwijskundig leiderschap en bij KA3 om goede voorbeelden van de interne dialoog en de dialoog met de omgeving. Daarnaast worden vooral bij KA1 het hebben van eigen ambities of ambitieuze leerdoelen genoemd. Deze ambities gaan meestal over ambitieuze leerdoelen zoals talentontwikkeling, meer kansen geven aan leerlingen met een achterstand of 21<sup>e</sup> - eeuwse leerdoelen.

Bijvoorbeeld:

*'Er zijn ambities en doelen voor het welbevinden van de leerlingen, doelen voor de 21<sup>e</sup> - eeuwse vaardigheden, sociale cohesie en actief burgerschap en men heeft geconcretiseerde hoge verwachtingen van leerlingen.'* (po)

Verder zijn er ook ambities op onderwijskundig terrein, zoals het zorgdragen voor een ononderbroken leerlijn, maar ook dat heeft uiteindelijk tot doel om te komen tot een beter doorstroomresultaat. Uit de antwoorden blijft vaak onduidelijk of die ambitieuze doelen ook daadwerkelijk worden gerealiseerd.

De conclusie is niet anders dan in het vorige fairnessonderzoek uit 2018. Bij eigen aspecten van kwaliteit gaat het ofwel om een zeer goede en hoogwaardige invulling van de standaard of om ambitieuze leerdoelen. De standaardbeschrijving omvat al voldoende om dit mee te laten wegen. De voorbeelden van eigen aspecten van kwaliteit in het huidige onderzoekskader zetten vooral op het verkeerde been. Men zou wel een aparte niet-wettelijk verplichte standaard kunnen formuleren over die aanvullende doelen en ambities: De school/het bestuur kent aanvullende ambitieuze leerdoelen. Als hiervoor niet wordt gekozen, bevelen wij aan om in ieder geval afspraken te maken over een gelijke behandeling van dit onderwerp door de inspecteurs. Daarnaast kan men ook afspraken maken om dit onderwerp op een gestandaardiseerde gelijkvormige manier in de rapportage weer te geven.

## 6 Conclusies

In dit derde fairnessonderzoek brengen we in kaart hoe fair de oordelen van inspecteurs op de KA-standaarden zijn. Daarnaast gaan we na hoe we de standaarden van het kwaliteitsgebied Kwaliteitszorg en ambitie kunnen verbeteren.

- De analyses tonen aan dat de inspectie voldoende betrouwbaar oordeelt en dat er een voldoende scherpe cesuur is (Hoofdstuk 4). Dat is een verdienste die worden onderhouden en waar we blijvend aandacht aan moeten geven.
- Het viel op dat de standaarden vrij veel tekstdelen bevatten die vaak een voldoende scoorden en dat ook de standaard als geheel vaak al snel een voldoende kreeg. De wettelijke basis is dus laag, met name bij KA2. Wij raden aan om te onderzoeken of de wet mogelijkheden biedt om de standaarden zo op te stellen dat ze meer ruimte bieden voor discussie over het oordeel en daarmee stimulerender zijn voor het onderwijs. Ook zou er in toekomstig fairnessonderzoek aandacht moeten zijn voor de cesuur (Hoofdstuk 4).
- Elke standaard vertegenwoordigt in voldoende mate een eigen kenmerk van kwaliteitszorg. We zagen geen redenen om standaarden samen te voegen of te splitsen, maar we doen op basis van de bevindingen wel suggesties om bepaalde tekstdelen te schrappen of juist toe te voegen. Zo zagen we dat tekstdelen waarin gevraagd wordt naar een wettelijke formaliteit vaak minder goed bij de standaard passen en adviseren we deze los van de standaard te beoordelen (Hoofdstuk 5).

Het is belangrijk om zorg te blijven dragen voor fairness zodat de goede interbeoordelaarsbetrouwbaarheid wordt behouden. Op basis van de literatuur en de verschillende fairnessonderzoeken hebben we een aantal richtlijnen voor interbeoordelaarsbetrouwbaarheid (IBB) geformuleerd, die we ook al in eerder fairnessonderzoek hebben genoemd (Bijlage VI). Wij hopen dat ze uitnodigen om uit te zoeken wat men nog kan doen om de IBB te verbeteren. De kern van IBB is dat inspecteurs met elkaar in gesprek gaan over de oordeelsvorming.

### 6.1 Vooruitblik

Het opsplitsen van de standaarden in onderliggende items blijkt een handige methode om onderzoek te doen naar de betrouwbaarheid van het inspecteursoordeel. Wanneer het kader wordt aangepast, is het nuttig om dit type onderzoek te herhalen zodat we kunnen zien hoe de relatie is tussen scores op onderliggende items en scores op de standaard. Wij denken dat het vooral van belang blijft om ons de vraag te stellen of inspecteurs de cesuur op de dezelfde plaats leggen. Het is goed om een komend onderzoek te wijden aan de cesuur tussen onvoldoende en voldoende. Dat zou bijvoorbeeld met vignetstudies waarbij inspecteurs en mogelijk ook andere deskundigen, casuïstiek krijgen voorgelegd om er een oordeel over te geven.

Een tweede onderwerp dat van belang is voor verder onderzoek is de onderbouwing van de oordelen in de rapporten. Ook dat kan een bron van verschillen zijn die in ieder geval bij de gebruikers in het veld opvalt maar die met een studie zoals we die nu hebben gedaan niet wordt opgemerkt. Eerder onderzoek heeft al laten zien dat hiervoor meer aandacht moet komen (Inspectie van het Onderwijs, 2019). Verder zullen we de oordelen regelmatig moeten blijven monitoren. Het blijft van belang om

trends in de oordelen te bekijken en de vraag te stellen waarom er veel of juist weinig onvoldoendes zijn.

## 7 Literatuurlijst

- Dousma, T., & Horsten, A. (1980). *Tentamineren*. Het Spectrum.
- Drenth, P. J. D., & Sijtsma, K. (2006). *Testtheorie: inleiding in de theorie van de psychologische test en zijn toepassingen*. (4e dr.). Bohn Stafleu Van Loghum.
- Evers, A., Lucassen, W., Meijer, R., & Sijtsma, K. (2009). *COTAN beoordelingssysteem voor de kwaliteit van tests (geheel herziene versie)*. Amsterdam NIP.
- Field, A., Miles, J., & Field, Z. (2014). Exploratory factor analysis. In *Discovering Statistics Using R* (pp. 749–811). Sage publications.
- Inspectie van het Onderwijs. (2016). *Kan het nieuwe waarderingskader de scholen en besturen goed onderscheiden en is de uitvoering hiervan voldoende fair. Rapportage fairnessonderzoek*.
- Inspectie van het Onderwijs. (2018). *Betrouwbaarheid en fairness van het inspecteursoordeel. Rapportage fairnessonderzoek 2017*.
- Inspectie van het Onderwijs. (2019). *Kwaliteitszorg en ambitie 2017/2018* (p. 48).
- Inspectie van het Onderwijs. (2020). *Voortgangsrapportage 2018/2019*
- Streiner, D. L. (2003). Being inconsistent about consistency: when coefficient alpha does and doesn't matter. *Journal of Personality Assessment*, 80(3), 217–222.

## Bijlage I: Het instrument

Deze bijlage bevat het onderzoeksinstrument gebruikt voor po, vo en so. De instrumenten voor mbo en samenwerkingsverbanden zijn op te vragen bij de Inspectie.

<b>Po, so en vo kader</b>					
<b>Standaard KA 1: Kwaliteitszorg</b>					
Wat is het oordeel op de standaard KA1: Kwaliteitszorg?					
<b>Huidige itemdelen van de standaard</b>					
1	Het bestuur zorgt voor een stelsel van kwaliteitszorg op de scholen	o	o	o	o
2	Dit stelsel staat uitgewerkt in het schoolplan(nen) van de school/scholen	o	o	o	o
3	Vanuit dit stelsel bewaakt en bevordert het bestuur de kwaliteit van het onderwijsleerproces	o	o	o	o
4	Vanuit dit stelsel bewaakt en bevordert het bestuur de kwaliteit van de leerresultaten	o	o	o	o
5	Het bestuur en de scholen hebben zicht op de kwaliteit van het onderwijs	o	o	o	o
6	Er zijn toetsbare doelen geformuleerd	o	o	o	o
7	Er wordt regelmatig geëvalueerd of deze doelen gehaald worden	o	o	o	o
8	De oorzaken van eventueel tekortschietende onderwijskwaliteit worden geanalyseerd	o	o	o	o
9	Waar nodig worden verbeteringen doelgericht doorgevoerd	o	o	o	o
10	De verantwoordelijkheidsverdeling tussen bestuur en scholen maakt een functionerend stelsel van kwaliteitszorg mogelijk	o	o	o	o
<b>Items voor opname in het onderzoeksinstrument</b>					
A.1	Er is een cyclische en systematische aanpak van de kwaliteitsverbetering die gericht is op het wegwerken van tekortkomingen.	o	o	o	o
A.2	De kwaliteitszorg is erop gericht om de ononderbroken ontwikkeling van de leerlingen te evalueren en te verbeteren	o	o	o	o
A.3	Er zijn eigenaren voor de verschillende aspecten van de kwaliteit van het onderwijs	o	o	o	o
A.4	De verantwoordelijkheden van de verschillende onderdelen voor de zorg voor de kwaliteit zijn goed gedelegeerd	o	o	o	o
A.5	Het kwaliteitszorgsysteem wordt regelmatig geëvalueerd om te bepalen of het voldoende functioneert	o	o	o	o

### **Aanvullende (open) vraag:**

Is er een onderdeel geweest dat niet bij de boven beschreven items staat maar dat je wel hebt laten meewegen bij je **beoordeling van de basiskwaliteit** van standaard KA1?

Nee

Ja

Zo Ja: Welk onderdeel is dit bij KA1? Kun je uitleggen waarom je dit onderdeel ook hebt laten meewegen?

Zijn er nog **eigen aspecten** van kwaliteit geweest die je hebt laten meewegen bij de waardering van de standaard KA1?

Nee

0 Ja

Zo Ja: Welke **eigen aspecten** van kwaliteit bij standaard KA1 zijn dit? Kun je uitleggen waarom je deze ook hebt laten meewegen bij KA1?

<b>Po, so en vo Kader</b>					
<b>Standaard KA 2: Kwaliteitscultuur</b>					
		O	V	G	Ntb
	Wat is het oordeel op de standaard KA2: Kwaliteitscultuur?	o	o	o	o
<b>Huidige itemdelen van de standaard</b>					
1	Het bestuur handelt volgens de code van goed bestuur en legt uit wanneer zij daarvan afwijkt	o	o	o	o
2	Deze handelwijze leidt tot een integere en transparante organisatiecultuur	o	o	o	o
3	Ieder werkt vanuit zijn eigen rol aan de verbetering van de onderwijskwaliteit	o	o	o	o
4	Het bestuur zorgt voor bekwaam en bevoegd personeel op alle scholen	o	o	o	o
5	Het bestuur maakt mogelijk dat het personeel haar bekwaamheid onderhoudt	o	o	o	o
6	De schoolleiding en het team werken gezamenlijk aan een voortdurende verbetering van hun professionaliteit	o	o	o	o
7	Leraren houden daarbij rekening met de gestelde bekwaamheidseisen en beroepsprofielen en behaalde resultaten bij leerlingen	o	o	o	o
8	Leraren krijgen voldoende gelegenheid om aan de verbetering van hun professionaliteit te werken.	o	o	o	o
9	De wijze waarop ze dit doen, staat helder beschreven in het schoolplan	o	o	o	o
<b>Items voor opname in het onderzoeksinstrument</b>					
A.1	Er is een duidelijke en gedeelde onderwijsvisie en visie op kwaliteit	o	o	o	o
A.2	Er is sprake van een lerende organisatie, gericht op verbetering vanuit collectief en individueel lerend vermogen	o	o	o	o
A.3	Er is sprake van passend onderwijskundig leiderschap op alle organisatieniveaus	o	o	o	o
A.4	Er is een organisatiestructuur die stimuleert tot samenwerken	o	o	o	o
A.5	Er is een cultuur om elkaar aan te spreken	o	o	o	o
A.6	Het bestuur en de school zorgen voor een passend HRM-beleid gericht op de kwaliteit van het personeel	o	o	o	o

**Aanvullende (open) vraag:**

Is er een onderdeel geweest dat niet bij de boven beschreven items staat maar dat je wel hebt laten meewegen bij je **beoordeling van de basiskwaliteit** van standaard KA2?

0 Nee

0 Ja

Zo Ja: Welk onderdeel is dit bij KA2? Kun je uitleggen waarom je dit onderdeel ook hebt laten meewegen?

Zijn er nog **eigen aspecten** van kwaliteit geweest die je hebt laten meewegen bij de waardering van de standaard KA2?

0 Nee

0 Ja

Zo Ja: Welke **eigen aspecten** van kwaliteit bij standaard KA2 zijn dit? Kun je uitleggen waarom je deze ook hebt laten meewegen bij KA2?

	Po, so en vo Kader				
	<b>Standaard KA 3: Verantwoording en dialoog</b>				
	Wat is het oordeel op de standaard?				
	<b>Huidige itemdelen van de standaard</b>	O	V	G	Ntb
1	Het bestuur en de scholen hebben tegenspraak georganiseerd	o	o	o	o
2	De tegenspraak gebeurt onder meer door het intern toezicht te betrekken bij beleids- en besluitvorming	o	o	o	o
3	De tegenspraak gebeurt onder meer door de (G)MR te betrekken bij beleids- en besluitvorming				
4	Het bestuur en de school (scholen) brengen minimaal jaarlijks verslag uit over hun doelen	o	o	o	o
5	Het bestuur en de school (scholen) brengen minimaal jaarlijks verslag uit over de resultaten die zij behaalt	o	o	o	o
6	Zij doen dit op toegankelijke wijze	o	o	o	o
7	Het bestuur verantwoordt zich aan de intern toezichthouder	o	o	o	o
8	Het bestuur en de scholen verantwoorden zich aan de overheid en belanghebbenden	o	o	o	o
	<b>Items voor opname in het onderzoeksinstrument</b>				
A.1	De interne dialoog over het te voeren beleid functioneert naar behoren	o	o	o	o
A.2	Het schoolplan voldoet aan de eisen	o	o	o	o
A.3	De school rapporteert in de schoolgids richting ouders en leerlingen in hoeverre de doelen die zij hebben opgesteld worden bereikt	o	o	o	o
A.4	Het bestuur heeft een cultuur en houding waarin de intern toezichthouder onafhankelijk kan functioneren ( <i>alleen voor het bestuur</i> )	o	o	o	o
A.5	Het bestuur en de school betrekken belanghebbenden bij het onderwijsbeleid	o	o	o	o
A.6	Het bestuur houdt rekening met de context waarin de school opereert ( <i>alleen voor het bestuur</i> )	o	o	o	o
A.7	De school houdt rekening met de omgeving en de doelgroep ( <i>alleen voor de school</i> )				

**Aanvullende (open) vraag:**

Is er een onderdeel geweest dat niet bij de boven beschreven items staat maar dat je wel hebt laten meewegen bij je **beoordeling van de basiskwaliteit** van standaard KA3?

0 Nee

0 Ja

Zo Ja: Welk onderdeel is dit bij KA3? Kun je uitleggen waarom je dit onderdeel ook hebt laten meewegen?

Zijn er nog **eigen aspecten** van kwaliteit geweest die je hebt laten meewegen bij de waardering van de standaard KA3?

0 Nee

0 Ja

Zo Ja: Welke **eigen aspecten** van kwaliteit bij standaard KA3 zijn dit? Kun je uitleggen waarom je deze ook hebt laten meewegen bij KA3?

**Slotvraag**



Als je nog vragen of opmerkingen hebt, dan kun je ze hieronder kwijt.

## Bijlage II: Voorbeeldresultaten factoranalyse

Tabel 1 t/m 3 tonen een voorbeeld van een factoranalyse van de inspecteursscores op alle items van de drie KA-standaarden op bestuursniveau bij de sectoren po, vo en so. Voor de analyse zijn alle items samengenomen. De kolommen in Tabel 1t/m3 staan voor de factoren die uit de analyse naar voren kwamen. De getallen geven de sterkte van de relatie tussen de items en de factor aan. Deze kunnen variëren tussen -1 en 1. De factoren vertegenwoordigen een kenmerk van Kwaliteitszorg en ambitie en waar mogelijk koppelen we dat kenmerk aan een standaard(voor verdere toelichting zie de volgende pagina).

Tabel 1. Voorbeeld resultaten factoranalyse po, vo, en so op bestuursniveau, items KA1

	KA1	KA2	KA3A	KA3B	Overige factoren		
<b>Criteria aan KA1 ontleend</b>							
Het bestuur zorgt voor een stelsel van kwaliteitszorg op de scholen	0,88						
Dit stelsel staat uitgewerkt in het schoolplan(nen) van de school/scholen	0,72		0,51				0,34
Vanuit dit stelsel bewaakt en bevordert het bestuur de kwaliteit van het onderwijsleerproces en de leerresultaten	0,88						
Het bestuur heeft zicht op de kwaliteit van het onderwijs	0,66	0,47					
Er zijn toetsbare doelen geformuleerd	0,72						
Er wordt regelmatig geëvalueerd of deze doelen gehaald worden	0,84						
De oorzaken van eventueel tekortschietende onderwijskwaliteit worden geanalyseerd	0,76				0,33		
Waar nodig worden verbeteringen doelgericht doorgevoerd	0,61	0,45					
De verantwoordelijkheidsverdeling tussen bestuur en scholen maakt een functionerend stelsel van kwaliteitszorg mogelijk	0,61			0,56			
Er is een cyclische en systematische aanpak van de kwaliteitsverbetering die gericht is op het wegwerken van tekortkomingen	0,80	0,36					
De kwaliteitszorg is erop gericht om de ononderbroken ontwikkeling van de leerlingen te evalueren en te verbeteren	0,83		0,36				
Er zijn eigenaren voor de verschillende aspecten van de kwaliteit van het onderwijs	0,62			0,39			0,49
De verantwoordelijkheden van de verschillende onderdelen voor de zorg voor de kwaliteit zijn voldoende gedelegeerd	0,81						
Het kwaliteitszorgsysteem wordt regelmatig geëvalueerd om te bepalen of het voldoende functioneert	0,74				0,38		

Tabel 2. Voorbeeld resultaten factoranalyse po, vo, en so op bestuursniveau, items KA2

	KA1	KA2	KA3A	KA3B	Overige factoren	
<b>Criteria aan KA2 ontleend</b>						
Het bestuur handelt volgens de code van goed bestuur en legt uit wanneer zij daarvan afwijkt		0,39	0,50		0,60	
Deze handelwijze leidt tot een integere en transparante organisatiecultuur		0,55		0,45	0,49	
Ieder werkt vanuit zijn eigen rol aan de verbetering van de onderwijskwaliteit	0,40	0,61		0,40		
Het bestuur zorgt voor bekwaam en bevoegd personeel op alle scholen		0,72		0,34		
Het bestuur maakt mogelijk dat het personeel haar bekwaamheid onderhoudt		0,83				
De schoolleiding(en) en de teams die onder dit bestuur vallen werken gezamenlijk aan een voortdurende verbetering van hun professionaliteit		0,78				
De leraren die onder dit bestuur vallen houden daarbij rekening met de gestelde bekwaamheidseisen en beroepsprofielen en behaalde resultaten bij leerlingen	0,46	0,60				0,35
De leraren die onder dit bestuur vallen krijgen voldoende gelegenheid om aan de verbetering van hun professionaliteit te werken.		0,82				
De wijze waarop de leraren die onder dit bestuur vallen dit doen, staat helder beschreven in het schoolplan		0,33	0,53			0,58
Er is een duidelijke en gedeelde onderwijsvisie en visie op kwaliteit	0,35	0,68				
Er is sprake van een lerende organisatie, gericht op verbetering vanuit collectief en individueel lerend vermogen		0,72		0,36		0,37
Er is sprake van passend onderwijskundig leiderschap op alle organisatieniveaus		0,65				-0,36
Er is een organisatiestructuur die stimuleert tot samenwerken		0,57	0,44			
Er is een cultuur om elkaar aan te spreken	0,41	0,67				
Het bestuur zorgt voor een passend HRM-beleid gericht op de kwaliteit van het personeel		0,30			0,80	

Tabel 3. Voorbeeld resultaten factoranalyse po, vo, en so op bestuursniveau items KA3

	KA1	KA2	KA3	Overige factoren			
<b>Criteria aan KA3 ontleend</b>							
Het bestuur heeft tegenspraak georganiseerd			0,41	0,59	0,44		
De tegenspraak gebeurt onder meer door het intern toezicht te betrekken bij beleids- en besluitvorming		0,48	0,35	0,55	0,32		
De tegenspraak gebeurt onder meer door de (G)MR te betrekken bij beleids- en besluitvorming		0,45	0,42	0,65			
Het bestuur brengt minimaal jaarlijks verslag uit over hun doelen			0,83				
Het bestuur brengt minimaal jaarlijks verslag uit over de resultaten die zij behalen	0,33		0,74				
Zij doen dit op toegankelijke wijze			0,83				
Het bestuur verantwoordt zich aan de intern toezichthouder			0,72	0,54			
Het bestuur verantwoordt zich aan de overheid en belanghebbenden			0,84	0,32			
De interne dialoog over het te voeren beleid functioneert naar behoren			0,52	0,60		0,36	
Het schoolplan voldoet aan de eisen			0,81				
Het bestuur betreft belanghebbenden bij het onderwijsbeleid	0,32	0,38	0,40	0,31		0,58	
Het bestuur heeft een cultuur en houding waarin de intern toezichthouder onafhankelijk kan functioneren		0,36	0,32	0,49	0,53		
Het bestuur houdt rekening met de context waarin de school opereert	0,34	0,31				0,64	

Het volgende valt op:

- De standaarden KA1 en KA2 zijn goed terug te zien in het resultaat (kolom 1 en 2). De items passen meestal goed bij het kenmerk dat correspondeert met de eigen standaard. Met grijze arcering is dat aangegeven.
- Bij KA3 zie je dat de standaard in twee subkenmerken uiteen loopt (ook grijs gearceerd onder de kolommen KA3A en KA3B. Het gaat (1) om verantwoordten en (2) om tegenspraak.
- Met geel de items aangegeven die wel bij de standaard horen maar voor een deel ook onder een andere factor vallen.
- Met rood zijn items aangegeven die op basis van de scores niet goed in de standaard lijken te passen. Met name hier moeten we kijken of er redenen zijn om deze te verwijderen of elders te plaatsen.

## Bijlage III: Resultaten betrouwbaarheidsanalyse

Tabel 1 t/m 3 hieronder bevatten de betrouwbaarheidscoëfficiënten (alfa) voor de oorspronkelijke items van elke standaard, de items met aanvullingen en het eindvoorstel. Het eindvoorstel is een voorstel voor de items in elke standaard gebaseerd op de uitkomsten van dit onderzoek.

Tabel 1. Uitkomsten van de betrouwbaarheidsanalyse (Cronbachs alfa) op de standaarden van po, vo en so <sup>7</sup>

		KA1	KA2	KA3
Bestuursniveau	Alleen oorspronkelijke items	.93	.92	.86
	Met aanvullingen	.95	.94	.91
	Eindvoorstel	.95	.94	.90
Schoolniveau	Alleen oorspronkelijke items	.93	.90	.93
	Met aanvullingen	.95	.94	.94
	Eindvoorstel	.94	.93	.93

Tabel 2. Uitkomsten van de betrouwbaarheidsanalyse (Cronbachs alfa) op de standaarden van mbo

		KA1	KA2	KA3
Bestuursniveau	Alleen oorspronkelijke items	.94	.60	.94
	Met aanvullingen	.97	.81	.96
	Eindvoorstel	.97	.95	.95
Opleidingsniveau	Alleen oorspronkelijke items	.92	.87	.82
	Met aanvullingen	.96	.93	.90
	Eindvoorstel	.95	.93	.78

Tabel 3. Uitkomsten van de betrouwbaarheidsanalyse (Cronbachs alfa) op de standaarden van samenwerkingsverbanden

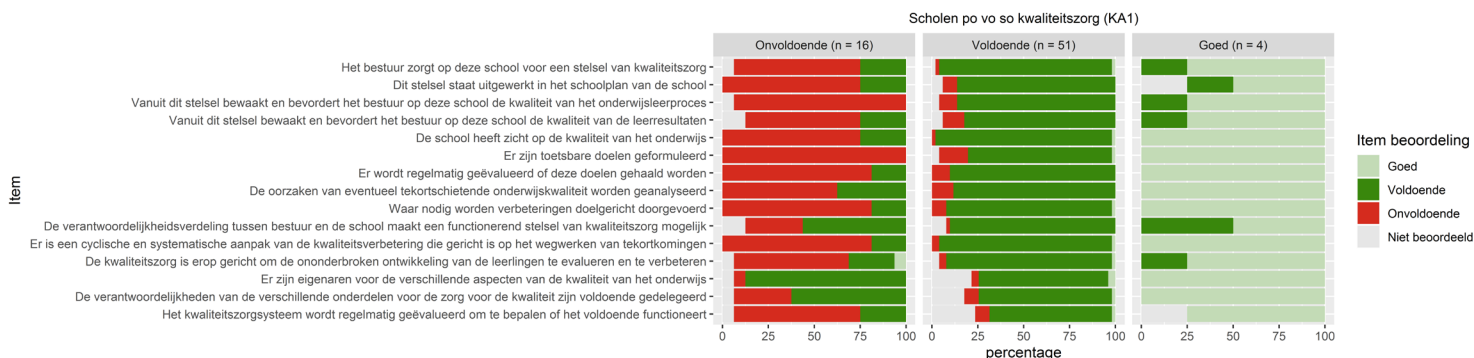
	KA1	KA2	KA3
Alleen oorspronkelijke items	.91	.89	.84
Met aanvullingen	.93	.95	.89
Eindvoorstel	.93	.98	.86

<sup>7</sup> Behandeling van niet beoordeelde criteria: Op voldoende gezet vanuit het idee dat de inspecteur bij het betreffende item geen belemmeringen zien. Dit is de procedure die het laagste resultaat voor alfa geeft.

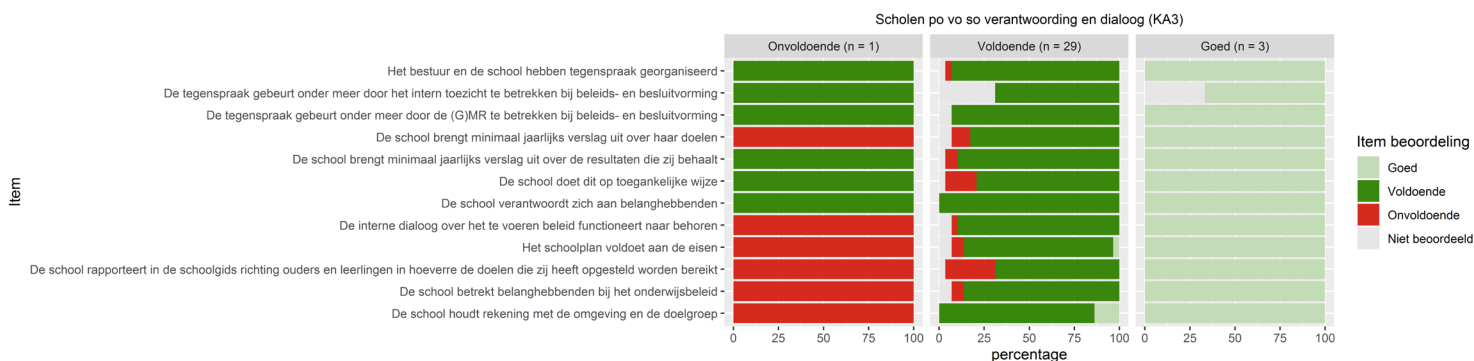
## Bijlage IV: Resultaten scorepatronen

Figuur 1 t/m 4 tonen de itemscores per oordeel op de standaard voor po, vo en so op schoolniveau en bestuursniveau. De resultaten voor besturen po vo en so KA1 en voor scholen po vo en so KA2 zijn te vinden in Figuur 4.1a en 4.1b in hoofdstuk 4.

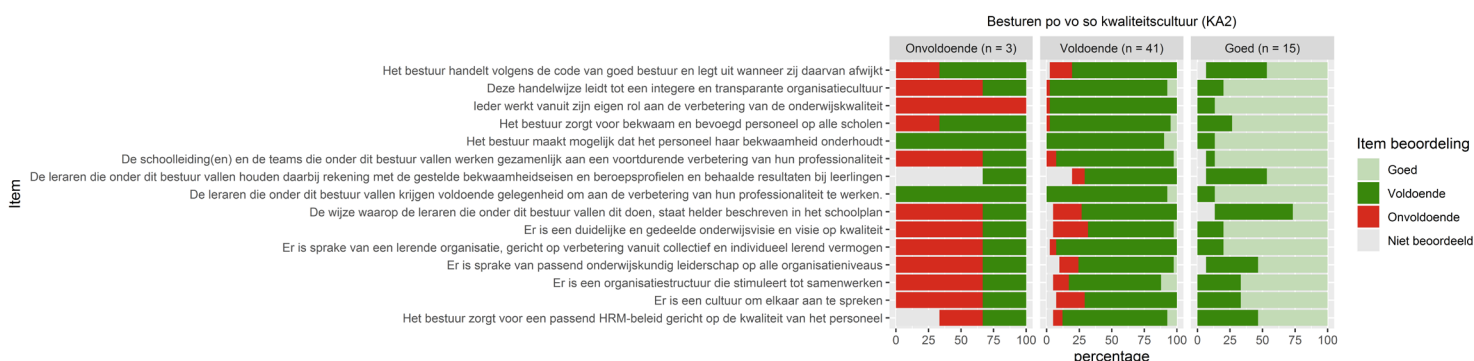
Figuur 1. Oordelen op afzonderlijke items gegeven het oordeel op standaard KA1, Scholen po, vo, en so



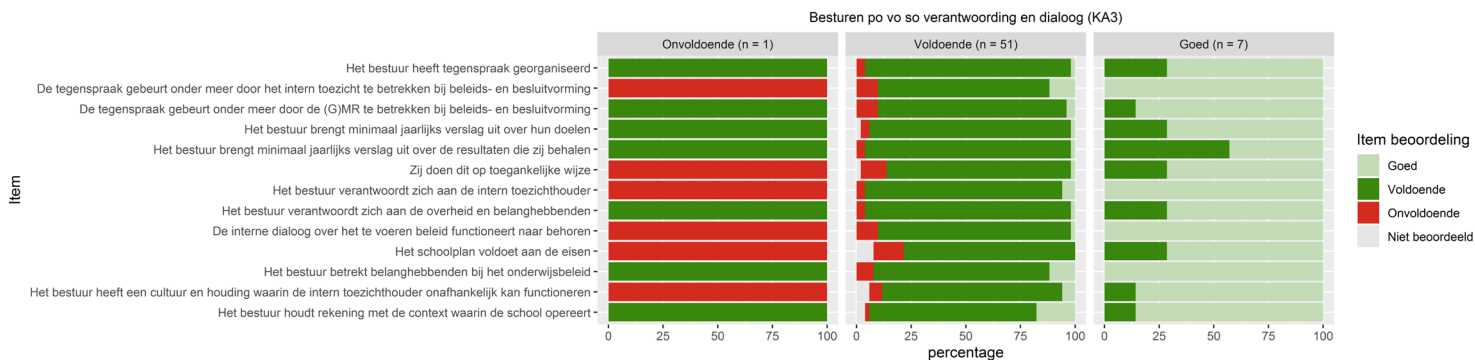
Figuur 2. Oordelen op afzonderlijke items gegeven het oordeel op standaard KA3, Scholen po, vo, en so



Figuur 3. Oordelen op afzonderlijke items gegeven het oordeel op standaard KA2, Besturen po, vo, en so



Figuur 4. Oordelen op afzonderlijke items gegeven het oordeel op standaard KA3, Besturen po, vo, en so.



## **Bijlage VI: Standaarden voor interbeoordelaarsbetrouwbaarheid (IBB)**

Een voldoende zorg voor de IBB dient naast procedures als het meelezen van rapporten en escalatie bij de oordelen zwak en zeer zwak, het volgende te omvatten:

- Alle beoordelingen van zowel de besturen als de scholen/opleidingen worden (voor ze definitief in een rapport worden geplaatst) collegiaal doorgesproken in een consensusoverleg.
- Wanneer onduidelijkheden worden geconstateerd bij de interpretatie van onderdelen van de standaard wordt ervoor gezorgd dat er centrale afspraken zijn.
- Indien teams van inspecteurs onderzoeken doen, rouleer de samenstelling van de teams regelmatig.
- Inspecteurs doen jaarlijks minstens één of meer onderzoek met collega's bij andere kantoren of andere afdelingen zodat er ook aandacht is voor interbeoordelaarsbetrouwbaarheid over de kantoren en afdelingen heen.
- Organiseer deze activiteiten ook over de sectoren heen, zodat inspecteurs gevoeligheid ontwikkelen voor beoordelingsvraagstukken bij andere sectoren en voor een breed draagvlak van de IBB-afspraken.
- Investeer in regelmatige scholing in het beoordelen van standaarden.
- Koppel informatie over de oordelen en uit onderzoeken zoals deze regelmatig terug aan inspecteurs.
- Door ieder afdelingshoofd wordt voor zijn of haar team de zorg voor de IBB aangestuurd en wordt hierover gerapporteerd.



## Colofon

Inspectie van het Onderwijs  
Postbus 2730 | 3500 GS Utrecht  
[www.onderwijsinspectie.nl](http://www.onderwijsinspectie.nl)

Een exemplaar van deze publicatie is te downloaden vanaf de website van de  
Inspectie van het Onderwijs: [www.onderwijsinspectie.nl](http://www.onderwijsinspectie.nl).

© Inspectie van het Onderwijs | juni 2020