

HOGESCHOOL UTRECHT

Master of Informatics



MASTER'S THESIS

**Invloed van *big data analytics* op organisatorische prestaties
binnen Nederlandse overheidsorganisaties**

Auteur:

Arjen Davids

Begeleider:

Guido Ongena

Lector:

Pascal Ravesteijn

Datum:

12/07/2023

Hogeschool Utrecht | University of Applied Sciences
Postbus 182
3500 AD Utrecht
Netherlands

Voorwoord

Voor u ligt de uitkomst van een kwantitatief onderzoek naar de invloed van *big data analytics* op organisatorische prestaties binnen Nederlandse overheidsorganisaties. Deze thesis is uitgevoerd en geschreven als afronding van de opleiding *Applied Data Science* van de *Master of Informatics* (Mol) aan de Hogeschool Utrecht (HU).

Het schrijven van dit voorwoord betekent voor mij de afronding van een leerzame periode. Tijdens mijn werkervaring kwam ik erachter dat ik behoefte had aan meer verdieping rondom *data (science)*. In het eerste jaar van deze nieuwe opleiding ben ik gestart. Tijdens zes verschillende masterclasses heb ik veel theoretische én praktische ervaring opgedaan. Dit onderzoek is als afsluiting van de master. Tijdens de voorbereiding en vormgeving van de thesis is het meeste contact met mijn begeleider en medestudenten geweest. Zo hebben gezamenlijke overleggen voor de richting van het onderzoek en de totstandkoming van de vragenlijst gezorgd. Samenhangende thema's en data is hierin altijd de gemeenschappelijke deler geweest. Data kan veel zeggen, maar is nog geen informatie (Ponting, 2017). Hoe moet hiermee omgegaan worden om organisaties verder te helpen? Dit is mooi verwoord door Hal Varian, de hoofdeconoom van Google. Hieruit blijkt dat het vertalen van data naar inzichten de komende tientallen jaren een enorm belangrijke vaardigheid zal zijn:

The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it – that's going to be a hugely important skill in the next decades (McKinsey, 2009).

Mijn dank gaat uit naar mijn begeleider en programmadirecteur van de Master of Informatics; Guido Ongena. De fijne begeleiding en ondersteuning tijdens dit traject heeft bijgedragen aan dit onderzoek. Mijn dank gaat ook uit naar alle respondenten (medewerkers bij Nederlandse overheidsorganisaties) die mee hebben gewerkt aan dit onderzoek. Zonder mijn begeleider en de respondenten had dit onderzoek niet kunnen plaatsvinden.

Een woord van dank gaat ook nog uit naar de betrokken lectororen Pascal Ravesteijn en Marlies van Steenberghe voor de feedbackmomenten tijdens de thesisbijeenkomsten. Mijn medestudenten, mijn netwerk en mijn collega's wil ik ook bedanken voor alle suggesties en momenten om over mijn onderzoek te kunnen sparren en vertellen. Tot slot wil ik mijn vriendin en familie bedanken voor het vertrouwen in mij om deze studie en thesis af te ronden. De motiverende woorden van bovengenoemde personen hebben mij geholpen deze thesis inhoud en vorm te geven.

Ik wens u veel leesplezier toe.

Arjen Davids
Arnhem, 12 juli 2023

Samenvatting

Er is steeds meer data (of gegevens) beschikbaar binnen organisaties. De hoeveelheid neemt exponentieel toe. Zo worden Nederlandse overheidsorganisaties ook wel overspoeld door data. Maar, data *an sich* heeft weinig waarde en biedt geen informatie. Om impact te hebben, moet dit worden samengevoegd en geanalyseerd. Maar, kunnen deze overheidsorganisaties daar wel mee omgaan? Kunnen zij de (toenemende) hoeveelheid data en de verschillende vormen hiervan verwerken door hier analyses op uit te voeren en te vertalen in de praktijk?

Dit onderzoek gaat op zoek naar factoren die hieraan bijdragen. Centraal staat de vraag ‘*Welke factoren die bij big data en analyses komen kijken, zijn bij de medewerkers van Nederlandse overheidsorganisaties nodig zodat zij voldoende kennis en mogelijkheden hebben om analyses met grote databestanden te doen om organisatorische prestaties te behalen?*’ Op dit moment wordt voornamelijk reactief geacteerd op vragen uit de markt.

Via literatuuronderzoek zijn verschillende definities en modellen onderzocht. Zo zijn verschillende combinaties van *big data analytics capability (BDAC) models* aan bod gekomen. Bij big data zijn onder meer de 6 v's (*volume, velocity, variety, value, varacity en variability*) aan bod gekomen. Vanuit dit literatuuronderzoek is een conceptueel model opgesteld die leidt tot de hypothesen dat een hogere score op *big data analytics capabilities* leidt tot hogere interne en externe organisatorische prestaties.

Vanuit de literatuur is een vragenlijst opgesteld die uit is gezet onder medewerkers van Nederlandse overheidsorganisaties. Deze vragenlijst heeft betrekking op tastbare factoren, ontastbare factoren, menselijke vaardigheden en organisatorische prestaties. Uiteindelijk hebben 120 respondenten de vragenlijst volledig ingevuld. De resultaten hebben hier betrekking op. Twee derde van de respondenten is man. Daarnaast hebbende respondenten een gemiddelde leeftijd van 45 jaar en hebben zij gemiddeld 20 jaar werkervaring. Vier op de vijf respondenten hebben ervaring met (*big*) *data analytics*.

Als eerste blijkt uit dit onderzoek dat de gemiddelde score voor *big data analytics capabilities* een 4,1 is (op basis van een 7-puntschaal). De gemiddelde score op organisatorische prestaties is een 5,0. Vergeleken met de literatuur is het gemiddelde cijfer voor *big data analytics capabilities* een 4,4 en voor organisatorische prestaties een 4,8. Dit betekent dat de *capabilities* door Nederlandse overheidsmedewerkers lager ingeschat wordt dan de literatuur (4,1 vs. 4,4). Omgekeerd schatten zij de organisatorische prestaties juist hoger in (5,0 vs. 4,8). Bij deze vergelijking is geen rekening met verschillende contexten gehouden, zoals het aantal respondenten, achtergrondkenmerken van de respondenten en het land waarin de vragenlijst is uitgezet. Als er alleen gekeken wordt naar de gemiddelden, dan is er vooral ruimte voor verbetering op het gebied van basismiddelen, technologie en datagedreven cultuur.

Een PLS-SEM-analyse is (inclusief bootstrapping en PLSpredict) uitgevoerd om in kaart te brengen of *big data analytics capabilities* bijdragen aan de organisatorische prestaties. Hieruit blijkt dat dit zowel voor de interne ($\beta=0,680$, $t=12,714$, $p<0,001$) als externe ($\beta=0,495$, $t=9,053$, $p<0,001$) organisatorische prestaties een aanzienlijke positieve invloed heeft. Deze correlaties betekenen niet dat er een oorzakelijk verband is. De gevonden correlaties (hoe laag of hoog deze ook zijn) betekenen niet dat de *big data analytics capabilities* de oorzaak van de verandering zijn.

Daarnaast zijn via een post-hoc analyse extra inzichten verzameld. Hierin is gekeken welke subcategorieën van *big data analytics capabilities* voornamelijk bijdragen aan de interne en externe organisatorische prestaties. Hieruit komt naar voren dat de managementvaardigheden en lerende organisatie een significante positieve invloed hebben op de interne én externe organisatorische prestaties (managementvaardigheden intern $\beta=0,278$, $t=2,055$, $p<0,05$ en extern $\beta=0,529$, $t=4,007$, $p<0,001$; lerende organisatie intern $\beta=0,344$, $t=2,228$, $p<0,05$ en extern $\beta=0,335$, $t=2,740$, $p<0,05$).

Als aanbevelingen wordt onder meer vervolgonderzoek gesuggereerd. Via vervolgonderzoek kunnen de resultaten beter gevalideerd worden en kan voor een grotere onderzoekspopulatie gezorgd worden. Daarnaast kan het onderzoek internationaal, bijvoorbeeld in Europa, uitgezet worden zodat hier ook vergelijkingen mee gemaakt kunnen worden en de uitkomsten gevalideerd kunnen worden.

Voor Nederlandse overheidsorganisaties kunnen deze uitkomsten gebruikt worden om de bewustwording rondom (*big*) *data analytics* te vergroten. Mogelijk kan dit ook ingezet worden voor organisatorische prestaties. Doordat het speelveld van *big data analytics capabilities* een positieve invloed op de organisatorische prestaties heeft, kan het lonen om hier aandacht aan te besteden. Inzoomend op de onderliggende categorieën, gaat het volgens dit onderzoek voornamelijk om de managementvaardigheden en lerende organisatie. Deze twee categorieën hebben de meeste positieve invloed op interne en externe organisatorische prestaties. Overheidsorganisaties kunnen hier meer op inzetten, door medewerkers (o.a. de managers) op te leiden en ervoor te zorgen dat deze beschikken over managementvaardigheden en/of deze (verder) kunnen ontwikkelen. Door de managementvaardigheden te verbeteren, kunnen teams (en daarmee medewerkers) gestimuleerd worden om met hun werkzaamheden – al dan niet onbewust – organisatorische prestaties te behalen. Hiermee is het belangrijk dat binnen Nederlandse overheidsorganisaties de gedachten ingesteld zijn op het analyseren en visualiseren van data tot bruikbare (stuur)informatie, en de (zelf)leerzaamheid van de medewerkers en organisatie.

De uitkomsten zijn mogelijk ook relevant voor andere organisaties dan Nederlandse overheidsorganisaties. Te denken valt aan semi-overheidsorganisatie of profit-organisaties. Ondanks dat de organisatiedoelen hier anders zijn, is hier waarschijnlijk nog meer doel om organisatorische prestaties te behalen. Daaronder valt in grotere zin het behalen van financiële winst.

Inhoudsopgave

Voorwoord	2
Samenvatting	3
1. Inleiding.....	7
1.1. Achtergrond.....	7
1.2. Probleemomschrijving	8
1.3. Onderzoeksvraag	9
1.4. Relevantie en toegevoegde waarde	10
1.5. Leeswijzer	11
2. Theoretisch kader	12
2.1. Big data en big data analytics	13
2.2. Big data analytics capability (models)	15
2.3. Governmental performance.....	18
2.4. Big data analytics & government performance.....	20
2.5. Conceptueel model	21
2.6. Hypotheses	21
3. Methode.....	23
3.1. Aanpak	23
3.2. Instrument dataverzameling	23
3.3. Steekproef.....	26
3.4. Procedure dataverzameling	28
3.5. Data-analyse	29
4. Resultaten	31
4.1. Beschrijvende resultaten	31
4.2. Betrouwbaarheid en validiteit (measurement model)	35
4.3. Toetsing hypotheses (structural model)	38
4.4. Post-hoc analyse.....	40
5. Discussie en aanbevelingen	42
5.1. Bevindingen	42
5.2. Beperkingen.....	43
5.3. Vervolgonderzoek	44
5.4. Aanbevelingen	45
6. Conclusie.....	47
6.1. Beantwoording hoofdvraag	47

Literatuurlijst	49
Bijlagen	56
Bijlage 1. Criteria en zoektermen literatuuronderzoek	56
Bijlage 2. Vragenlijst.....	57
Bijlage 3. Syntax voor vragenlijstanalyse	66
Bijlage 4. Uitgebreide analyseresultaten	71
Bijlage 5: Figuren- en tabellenlijst	92

1. Inleiding

In dit inleidende hoofdstuk wordt gekeken naar de totstandkoming van de onderzoeksvraag. Als eerste wordt in paragraaf 1.1 de achtergrond en relevantie van dit onderzoek weergegeven. Vervolgens is in paragraaf 1.2 de probleemomschrijving te vinden, gevolgd door de onderzoeksvraag en deelvragen in paragraaf 1.3. Paragraaf 1.4 geeft de relevantie en toegevoegde waarde weer. Tot slot volgt de leeswijzer in paragraaf 1.5.

1.1. Achtergrond

Er is steeds meer data (of gegevens) beschikbaar binnen organisaties. De hoeveelheid neemt exponentieel toe (De Kruik et al., 2015). Zo ook bij steden en gemeenten, die overspoeld worden met data. Daarbij heeft data *an sich* weinig waarde. Om impact te hebben, moet dit worden samengevoegd en geanalyseerd. Het vergt mensen (medewerkers) die tijd hebben en over vaardigheden en middelen beschikken om data te interpreteren en hier inzichten uit te halen (Malomo & Sena, 2017). Maar, kunnen Nederlandse overheidsorganisaties daar wel mee omgaan? Kunnen zij de (toenemende) hoeveelheid data en de verschillende vormen hiervan verwerken door hier analyses op uit te voeren en te vertalen in de praktijk?

Vanuit de onderzoekspraktijk bij het Mulier Instituut (sportonderzoek) blijkt dat Nederlandse overheidsorganisaties (non-profit) de beschikbare data (nog) niet (altijd) goed weten te analyseren en zodoende in weten te zetten en/of te benutten (Mulier Instituut, 2020). Daarnaast verschilt het hoeveel tijd, middelen en aandacht per overheidsorganisatie beschikbaar is om dit op te pakken. Niet elke overheidsorganisatie, ongeacht de organisatiegrootte, kan hier dezelfde prioriteit aan geven. Als voorbeeld bestaat het gemeentelijk applicatielandschap nog uit veel applicaties die al dan niet aan elkaar gekoppeld of geïntegreerd zijn. Ieder gemeentelijk applicatielandschap is (vooralsnog) uniek (GEMMA Online, 2023). Gemeenten maken op die manier gebruik van tientallen of honderden verschillende applicaties. Het is realistisch gezien een ondenkbare gedachte om al deze data te ontsluiten en vertalen naar stuurinformatie (Gemeente Berg en Dal, 2023).

Dat er behoefte is om big data te gebruiken bij Nederlandse overheidsorganisaties, blijkt wel uit verschillende initiatieven die opgezet worden om het werken met data te bevorderen. Zo is de Vereniging Sport en Gemeenten (VSG) met een handreiking gekomen om met (lokale) data in de sport te werken (VSG, 2020). De databehoeftte onderschrijft Berenschot ook in opdracht van de Vereniging van Nederlandse Gemeenten (VNG). Bij gemeenten zijn hoge verwachtingen rondom datagedreven sturing. Hierbij is het besef dat een brede implementatie nog een lange en moeizame weg te gaan heeft. Daarbij zal voor kleinere gemeenten (overheidsorganisaties) een eerste kennismaking al een hele opgave zijn (Wesseling et al., 2018). Bij overheidsorganisaties is een hype over de kracht van big data. Tegelijkertijd is er een duidelijk gebrek aan empirisch onderzoek naar het gebruik van big data door de overheid en het potentieel hiervan (Chatfield et al., 2015). Door deze hoge verwachtingen, moeizame weg en opgave is de vraag of big data wel goed benut wordt om organisatorische prestaties te

verbeteren en behalen? Kunnen medewerkers en (sub)teams binnen Nederlandse overheidsorganisaties deze (complexe) analyses uitvoeren?

Uit onderzoek blijkt dat het moeilijk is om de enorme hoeveelheid beschikbare gegevens te verwerken en analyseren bij overheidsorganisaties. Eén van de belangrijkste uitdagingen waarmee overheidsorganisaties te maken hebben, is de implementatie van big data (Al-Sai & Abualigah, 2017). Dit bevestigen Malomo en Sena eveneens in 2017, dat de meeste lokale overheden worstelen met de vraag hoe ze de verschillende gegevens waarover ze beschikken kunnen exploiteren en hoe ze de nodige analytische capaciteiten kunnen opbouwen. Volgens Malomo en Sena – dat hun onderzoek uit hebben gevoerd in het Verenigd Koninkrijk – benutten de meeste lokale overheden hun big data niet optimaal dankzij de aanwezigheid van structurele datasilo's en andere contextuele factoren. Chatfield et al. (2015) komen met hun onderzoek over Amerikaanse steden tot de conclusie dat er een duidelijk gebrek is aan onderzoeken naar het gebruik van big data door de overheid. Wel zijn er vier soorten uitdagingen op het gebied van organisatorische capaciteiten gedefinieerd, namelijk; sociaal-politiek, strategisch, analytische en technisch. Belangrijk hierbij is om nu een grootschalige enquête op te zetten om dit verder te onderzoeken (Chatfield et al., 2015).

Naast de verschillende uitdagingen die hiervoor beschreven zijn over organisatorische capaciteiten, worden soortgelijke uitdagingen ook omschreven vanuit literatuuronderzoek op het gebied van big data (vaardigheden). Sivarajah et al. (2017) omschrijven drie big data uitdagingen vanuit hun literatuurstudie:

1. Data-uitdagingen hebben betrekking op de kenmerken van de data zelf (bijvoorbeeld de technische infrastructuur, het volume van de data en de snelheid van veranderende data).
2. Procesuitdagingen houden verband met het proces hoe data verzameld moet worden, deze geïntegreerd en getransformeerd wordt voordat het juiste analysemodel gekozen moet worden en de resultaten opgeleverd/gevisualiseerd kunnen worden.
3. Managementuitdagingen gaan over organisatorische aspecten rondom bijvoorbeeld privacy, beveiliging, *governance* en ethische aspecten.

Een vraag die hierbij naar voren komt, is welke vaardigheden binnen een organisatie van belang zijn om big data in te zetten en dat big data tot nieuwe inzichten leidt? Dit blijkt ook volgens Dremel et al. (2020) een onderwerp ter discussie om verder met kwantitatief onderzoek te onderzoeken.

1.2. Probleemomschrijving

Nu duidelijk is geworden dat er behoefte is om big data in te zetten, maar dat hiervoor capaciteiten nodig zijn, leidt dit tot de volgende vraag: *Welke factoren die bij big data en analyses komen kijken, zijn bij de medewerkers van Nederlandse overheidsorganisaties nodig zodat zij voldoende kennis en mogelijkheden hebben om analyses met grote databestanden te doen?* Op dit moment wordt voornamelijk reactief geacteerd op vragen uit de markt (Gemeente Berg en Dal, 2023). Als factoren kunnen helpen bij de realisatie om een proactieve werkwijze te creëren dan kunnen Nederlandse overheidsorganisaties meer uit (*big*) data halen.

Welke factoren bij kunnen dragen, kan omschreven worden als *Big Data Analytics Capabilities* (hierna uitgeschreven in kleine letters of BDAC). Dit zijn de mogelijkheden (*capabilities*) om big data te analyseren. In hoofdstuk 2.2 wordt hier verder op ingegaan. Uit de literatuur blijkt dat er geen eerder *peer-reviewed* onderzoek gevonden is dat gaat over de *big data analytics capabilities* bij Nederlandse overheidsorganisaties. Veelal gaan de onderzoeken over de profit sector, waar bedrijven als doel hebben om winst te behalen (Directiesecretaresse, z.d.). Rainey et al. (1976) hebben onderzoek gedaan naar de vergelijking tussen publieke en private organisaties (profit en non-profit). Zo zijn er belangrijke verschillen tussen omgevingsfactoren (o.a. de mate van marktblootstelling en politieke invloed), organisatie-omgeving transacties (o.a. openbare controle en breedte van de impact) en interne structuren en processen (o.a. de complexiteit van doelstellingen en persoonlijke kenmerken van medewerkers). Doordat er belangrijke verschillen tussen profit en non-profit organisaties zijn, is het relevant om dit onderzoek te richten op de non-profit sector, te weten Nederlandse overheidsorganisaties. Dit omdat er binnen andere sectoren (profit) en landen al wel onderzoek naar *big data analytics capabilities* gedaan is, welke in de literatuurstudie naar voren zijn gekomen.

De factoren die uit de literatuurstudie en dit onderzoek naar voren komen, kunnen door Nederlandse (overheids)organisaties (waaronder gemeenten, non-profit) ingezet worden om meer uit de huidige en toenemende hoeveelheid data te halen. Want op dit moment zijn Nederlandse overheidsorganisaties (waaronder gemeenten) nog niet altijd in staat om te verklaren welke factoren van invloed zijn om big data analyses uit te voeren en wat nodig is om organisatorische prestaties te behalen.

Het probleem zoals hiervoor weergegeven is, heeft betrekking op de grote (toenemende) hoeveelheid data en de verschillende vormen hiervan, die bij Nederlandse overheidsorganisaties (nog) niet (altijd) goed benut wordt. Dit probleem kan onderzocht worden en is in een onderzoeksvraag in de volgende paragraaf beschreven.

1.3. Onderzoeksvraag

Om tot een goede onderzoeksvraag te komen is gebruikt gemaakt van Oost en Markenhof (2002) om naar de onderzoeksfuncties te kijken. Wat moet de onderzoeksvraag beantwoorden? De inhoudelijke verankering gaat in op Nederlandse overheidsorganisaties en is afgebakend op *big data analytics capabilities*. De onderzoeksvraag is:

Welke big data analytics capabilities dragen bij aan organisatorische prestaties binnen Nederlandse overheidsorganisaties?

Deelvragen

Om de onderzoeksvraag te beantwoorden wordt gebruik gemaakt van deelvragen. Deze hangen samen met de hoofdvraag en zorgen ervoor dat de hoofdvraag beantwoord kan worden. De deelvragen zijn op de volgende pagina in tabel 1.1 weergegeven.

Tabel 1.1¹ Deelvragen onderzoeksvoorstel

Nr.	Deelvraag	Onderzoeks- functie	Methode beantwoording	tot
1	Wat is <i>big data</i> en <i>big data analytics</i> ?	Beschrijvend	Literatuurstudie	
2	Wat zijn <i>big data analytics capabilities</i> en welke bestaande modellen zijn er?	Beschrijvend	Literatuurstudie	
3	Wat zijn organisatorische (overheids)prestaties en hoe zijn deze te meten?	Definiërend	Literatuurstudie	
4	Wat zijn de verschillen en overeenkomsten tussen bestaande <i>big data analytics capability</i> modellen?	Vergelijkend	Literatuurstudie	
5	Welke <i>big data analytics capabilities</i> worden bij Nederlandse overheidsorganisaties ingezet?	Verklarend	Vragenlijst	
6	Welke verschillen zijn er tussen <i>big data analytics capabilities</i> bij Nederlandse overheidsorganisaties en de onderzochte organisaties vanuit de literatuur?	Vergelijkend	Vragenlijst en literatuurstudie	

1.4. Relevantie en toegevoegde waarde

De relevantie en toegevoegde waarde van dit onderzoek zijn onder te verdelen in de wetenschappelijke en praktische bijdrage, zoals hierna wordt toegelicht.

Wetenschappelijke bijdrage

De wetenschappelijke bijdrage komt voort uit de generaliseerbare vraagstelling. Het antwoord op de onderzoeksvraag moet nieuwe inzichten geven voor Nederlandse overheidsorganisaties, zoals gemeenten, om big data analyses (beter) uit te voeren en de organisatie hiervoor in te richten. Met als doel om de organisatorische prestaties te bevorderen. Het is mogelijk dat ook niet-overheidsorganisaties deze factoren kunnen gebruiken. Te denken valt aan (non-)profit organisaties in algemene zin. Welke *big data analytics capabilities* dragen in Nederland bij om organisatorische prestaties te behalen?

Praktische bijdrage

De praktische bijdrage is dat Nederlandse overheidsorganisaties (met name gemeenten) betere inzichten krijgen welke *big data analytics capabilities* bijdragen om organisatorische prestaties te behalen. Nederlandse overheidsorganisaties kunnen de factoren (randvoorwaarden) die voor big data analyses nodig zijn binnen de eigen organisatie in kaart brengen (aan welke factoren voldoet de

¹ In bijlage 5 is een figuren- en tabellenlijst te vinden waarin alle figuren en tabellen te vinden zijn met de bijbehorende paginanummers.

organisatie al?) en/of binnen de eigen (non-profit) organisatie uitvoeren. Vanuit verschillende invalshoeken is dit mogelijk (bijvoorbeeld IT-analytisch-organisatorisch).

Zowel Nederlandse overheidsorganisaties als andere organisaties die samen werken met deze organisaties, komen erachter welke *big data analytics capabilities* bijdragen om organisatorische prestaties te behalen. Hoe kunnen zij hun eigen big data beter analyseren en benutten? Mogelijk komen uit het onderzoek ook nog (praktische) punten naar voren waar non-profit organisaties mee worstelen in relatie tot de toenemende mate van het gebruik en de aanwezigheid van data.

(Lokale) overheden profiteren van dit onderzoek. Hierbij valt te denken aan gemeenten en provincies. Maar ook semi-overheidsorganisaties en/of andere organisaties binnen dezelfde sector of met dezelfde organisatiecultuur kunnen hiervan profiteren. Denk hierbij aan waterschappen, omgevingsdiensten en zelfstandige bestuursorganen (zbo's²: Rijksoverheid, z.d.a). Mogelijk is het niet sectorgerelateerd en kan het breder getrokken worden. Dit onderzoek lost de onderzoeksvraag op en kan in dit geval voor Nederlandse overheidsorganisaties *big data analytics capabilities* aangeven die bijdragen en/of vereist zijn om organisatorische prestaties te behalen. Het brengt overheidsorganisaties dicht bij de uitvoering en toepassing hiervan. Daarnaast kan dit onderzoek een aanzet voor vervolgonderzoeken zijn.

1.5. Leeswijzer

Het volgende hoofdstuk omvat het theoretisch kader, uitgevoerd met een literatuurstudie, waarin verschillende begrippen en concepten uitgelegd worden in relatie tot de onderzoeksvraag. Te denken valt aan big data, *big data analytics* en (overheids)prestaties. Hoofdstuk 3 geeft inzicht in de methode die gebruikt is voor dit onderzoek. Vervolgens zijn in hoofdstuk 4 de resultaten terug te vinden en in hoofdstuk 5 de daarbij behorende discussie en aanbevelingen. Tot besluit is in hoofdstuk 6 de conclusie te vinden.

² Zelfstandige bestuursorganen (zbo's) kunnen ook gezien worden als 'andere' overheidsorganisaties.

2. Theoretisch kader

In dit hoofdstuk wordt het theoretisch kader voor dit onderzoek geschreven in de vorm van een literatuurstudie. Hier komen theorieën, argumenten en bevindingen vanuit de literatuur die verband houden met dit onderzoek aan bod (Gratton et al., 2011). Dit zal geen opsomming zijn van alle beschikbare en gevonden literatuur, maar een kritische evaluatie van relevant bestaand werk. Hiermee worden verschillende begrippen en concepten uitgelegd die nodig zijn voor de beantwoording van de deelvragen. Voor deze scope is gekozen omdat de literatuur over termen als big data oneindig is en hier veel verschillende onderzoeken naar zijn gedaan en verschillende definities over bestaan.

In dit hoofdstuk zijn de volgende relevante begrippen en thema's nader omschreven, te weten *big data*, *big data analytics*, *big data analytics capability (models)* en *overheidsprestaties (governmental performance)*. Daarnaast is aan de hand van deze literatuur het conceptueel model opgesteld. Opgevolgd met de totstandkoming van de vragenlijst (onderzoeksmethode) en de hypothesen.

Voor het theoretisch kader en het opstellen van de vragenlijst is literatuur gezocht over big data, *big data analytics capabilities* en organisatorische prestaties. Dit is om wetenschappelijke achtergrondinformatie over deze onderwerpen op te halen, dit onderzoek af te bakenen en om een (gevalideerde) vragenlijst op te stellen. De vragenlijst bestaat – naast voorgaande thema's – uit twee extra thema's, te weten digitaal leiderschap (*digital leadership*) en datageletterdheid (data literacy). Deze thema's komen in dit theoretisch kader niet aan bod. Over deze extra thema's worden door andere studenten twee separate thesissen geschreven. In de methodesectie wordt nader ingegaan op de opgestelde vragenlijst.

Voor dit literatuuronderzoek wordt gebruik gemaakt van vier zoekmachines. Deze zoekmachines omvatten afzonderlijk en samen een grote hoeveelheid (wetenschappelijke) artikelen. Hierin wordt gezocht naar wetenschappelijk literatuur. De criteria (parameters) die gebruikt zijn in combinatie met de zoektermen zijn in bijlage 1 te vinden. De zoekmachines zijn:

1. HUGO: <https://hu-on-worldcat-org.hu.idm.oclc.org/discovery>
2. AIS eLibrary: <https://aisel.aisnet.org/>
3. Google Scholar: <https://scholar.google.nl/sci>
4. ScienceDirect: <https://www.sciencedirect.com/>

Binnen deze zoekmachines is door middel van verschillende zoektermen gezocht naar relevante literatuur. De zoektermen zijn zowel in de tekst als alleen in de titel opgezocht. De (gecombineerde) zoektermen zijn:

- *Big data*
- *Big data analytics*
- *Big data capability (of capabilities)*
- *Big data analytics capability (of capabilities)*

- *Big data analytics performance*
- *Governmental performance*
- *Organizational performance*
- *Competitive performance*
- *Governmental performance AND big data analytics capability (of capabilities)*
- *Organizational performance AND big data analytics capability (of capabilities)*
- *Competitive performance AND big data analytics capability (of capabilities)*

2.1. Big data en big data analytics

In deze paragraaf volgt de toelichting en definitie van *big data (analytics)*. Omdat hier veel verschillende definities voor zijn, worden alleen een aantal relevante en kwalitatief goede definities weergegeven. De relevantie en kwaliteit zit vooral in de begrijpbaarheid en volledigheid van de definities. Bijvoorbeeld door het gebruik van begrijpelijke definities, niet te lastige zinnen en de overeenkomst en gebruik van de toelichting en woordkeuzes in andere definities.

Big data

Voor *big data* zijn veel (verschillende) definities te vinden. De definities hebben vaak overeenkomsten, maar verschillen in terminologie en begrijpbaarheid. Hierna worden vier definities weergegeven die dit laten zien.

Ward en Barker (2013) hebben twaalf definities van onder andere (toonaangevende organisaties zoals) Gartner, Oracle en Microsoft onderzocht en definiëren *big data* als volgt:

“Big data is a term describing the storage and analysis of large and/or complex data sets using a series of techniques including, but not limited to: NoSQL, MapReduce, and machine learning.”

De TechAmerica Foundation (Federal Big Data Commission) (2012) geeft aan dat de term ‘big data’ steeds gebruikelijker wordt, maar de betekenis nog niet altijd duidelijk is. Daarom definiëren zij big data als volgt:

“Big data is a term that describes large volumes of high velocity, complex and variable data that require advanced techniques and technologies to enable the capture, storage, distribution, management, and analysis of the information.”

Het Nationaal instituut voor normen en technologie (NIST) uit Amerika omschrijft big data als volgt (2019):

“Big data is where the data volume, acquisition velocity, or data representation limits the ability to perform effective analysis using traditional relational approaches or requires the use of significant scaling (more nodes) for efficient processing.”

Als laatste hebben Provost en Fawcett (2013) tien jaar geleden big data omschreven als een uitdaging voor organisaties vanwege de enorme hoeveelheid aan gegevens. Ze onderschrijven hierbij de noodzaak om deze data effectief te kunnen analyseren om waardevolle inzichten te krijgen. De gehanteerde definitie is:

“Big data essentially means datasets that are too large for traditional data processing systems, and therefore require new processing technologies.”

Uit de onderzochte definities en toelichtingen blijkt dat de term big data verwijst naar de enorme hoeveelheid aan gegevens die gegenereerd wordt in de huidige samenleving. Deze gegevens zijn uit verschillende bronnen afkomstig. Vaak komen de volgende drie karaktereigenschappen naar voren: *volume*, *velocity* en *variety*. Dit worden ook wel de 3 v's genoemd. Tegenwoordig zijn de karaktereigenschappen die *big data* omschrijven uitgebreid tot 6 v's. Laney (2001) heeft de eerste drie karaktereigenschappen namens de Meta Group (later overgenomen door Gartner) gepubliceerd. Het gaat hierbij zoals hiervoor aangegeven om *volume*, *velocity* en *variety*. Later zijn hier door andere onderzoekers nog drie karaktereigenschappen aan toegevoegd. Het gaat om *value* (Marr, 2014), *veracity* (Marr, 2014) en *variability* (McNulty, 2014). De 6 v's houden het volgende in:

1. *Volume*: gaat over de omvang van de data en de relatie met de verwerkingscapaciteit.
2. *Velocity*: gaat over de voorwaarde dat data binnen minuten of seconden verwerkt moet (kunnen) zijn tot resultaten. Het gaat om de verwerkingssnelheid.
3. *Variety*: gaat over de verscheidenheid aan gegevens(bronnen) die opgeslagen worden en nog verwerkt en geanalyseerd moeten worden. De data kunnen verschillende typen zijn: gestructureerd, semigestructureerd en ongestructureerd.
4. *Value*: gaat over welke waarden uit de gegevens gehaald kunnen worden en hoe je met big data betere resultaten kunt krijgen uit de gegevens.
5. *Veracity*: gaat over de kwaliteit en oorsprong van de gegevens. Hiermee wordt aangetoond dat er kritisch naar de data gekeken is.
6. *Variability*: gaat over de voortdurende verandering van gegevens. Het gaat over het begrijpen en interpreteren van de juiste betekenissen van ruwe data.

Voor dit onderzoek is gekozen om de definitie van de TechAmerica Foundation (2012) aan te houden, omdat deze goed te begrijpen is en een duidelijke omschrijving geeft van big data. Deze omschrijving geeft aan dat big data een is term waarbij gesproken wordt over grote hoeveelheden snelle, complexe en variabele gegevens. Dit is in lijn met de beschikbare data bij Nederlandse overheidsorganisaties. Daarnaast zijn geavanceerde technieken en technologieën om met deze data te werken van belang. Zowel in de definitie van de TechAmerica Foundation als bij Nederlandse overheidsorganisaties is dit relevant.

Big data analytics

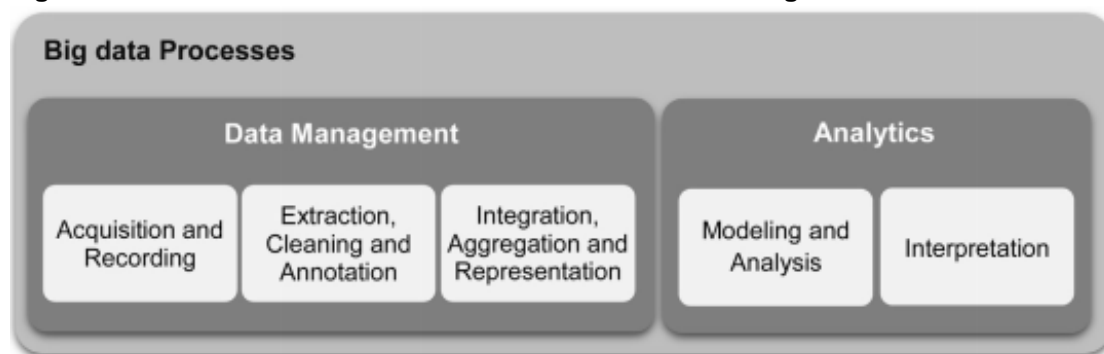
Uit voorgaande definities blijkt dat er overeenkomsten zijn in de verschillende *big-data*-definities. Ze benadrukken het grote volume, zoals de term ‘*big*’ al doet vermoeden. Daarnaast gaat het om de verscheidenheid en de snelheid van verandering. Maar wat is *big data analytics*? Dit staat eigenlijk in het verlengde van big data en gaat om het analyseren van deze (grote) hoeveelheid data. De gegevens moeten worden geanalyseerd om inzicht te krijgen in wat deze inhoudt. Hiervoor zijn bovendien nieuwe tools en enorme computercapaciteit nodig (Rajaraman, 2016).

Jelonek (2017) omschrijft *big data analytics* als volgt:

“Big data analytics is the process of examining large data sets to uncover hidden patterns, unknown correlations, market trends, customer preferences and other useful business information.”

Labrinidis en Jagadish (2012) hebben het proces van het verkrijgen van inzichten uit big data onderverdeeld in vijf fasen. Deze fasen hebben Gandomi en Haider (2014) gevisualiseerd in figuur 2.1. Hierin is inzichtelijk gemaakt dat *big data analytics* verder gaat met het analyseren en interpreteren van (big) data.

Figuur 2.1 Processen voor het extraheren van inzichten uit big data



Bron: Gandomi en Haider, 2014 (oorspronkelijk uit Labrinidis & Jagadish, 2012).

De omschrijvingen van *big data analytics* geven goed weer dat het doen van analyses (*analytics*) nodig is om tot waardevolle (bedrijfs)informatie te komen. De definitie van Jelonek (2017) omschrijft dit goed en begrijpelijk. Daarnaast sluit dit aan bij Ponting (2017), dat data veel kan zeggen maar nog geen informatie is. Dit is ook te vertalen naar big data en *big data analytics*, waar big data nog weinig zegt en *big data analytics* zorgt voor de (stuur)informatie.

2.2. Big data analytics capability (models)

Big data analytics capability (BDAC)

In de vorige paragraaf is aangegeven dat big data geanalyseerd moet worden om iets met deze grote hoeveelheid data te kunnen. Hiervoor zijn mogelijkheden (*capabilities*) nodig om dit te kunnen doen. Zoals aangegeven gaat het bijvoorbeeld om nieuwe tools en enorme computercapaciteit, ook wel

technische mogelijkheden genoemd. Maar, de organisatie en mens spelen hierin ook een belangrijke rol. Voor de capaciteiten kan ook breder gekeken worden dan alleen de *big data analytics capabilities*. Zo definiëren Eisenhardt en Martin (2000) dynamische capaciteiten. Dit zijn de processen bij bedrijven die middelen gebruiken, integreren, opnieuw configureren, verkrijgen en vrijmaken om marktverandering te evenaren en creëren.

De definitie volgens Eisenhardt en Martiin (2000) is:

“Dynamic capabilities thus are the organizational and strategic routines by which firms achieve new resource configurations as markets emerge, collide, split, evolve, and die.”

Akter et al. (2016) hebben onderzoek gedaan hoe bedrijfsprestaties in Amerika verbeterd kunnen worden met *big data analytics capabilities* en de afstemming op de bedrijfsstrategie. Hiervoor constateren Akter et al. (2016) dat bedrijven steeds meer uitgedaagd worden door big data. De afgelopen jaren is big data naar voren gekomen als grens van productiviteit en kansen. Dit zien zij terug in de manier waarop andere onderzoeken de *big data analytics capabilities* (BDAC) omschrijven. Zo wordt BDAC algemeen beschouwd als een verandering in de manier waarop bedrijven zaken doen (Davenport & Harris, 2007; Barton & Court, 2012). Eerdere onderzoeken brengen in kaart dat BDAC de volgende “*managementrevolutie*” is (McAfee & Brynjolfsson, 2012), “*het vierde paradigma van de wetenschap*” is (Strawn, 2012), het de “*next big thing in innovation*” is (Gobble, 2013) of “*het potentieel heeft om managementtheorie en praktijk te transformeren*” (George et al., 2014).

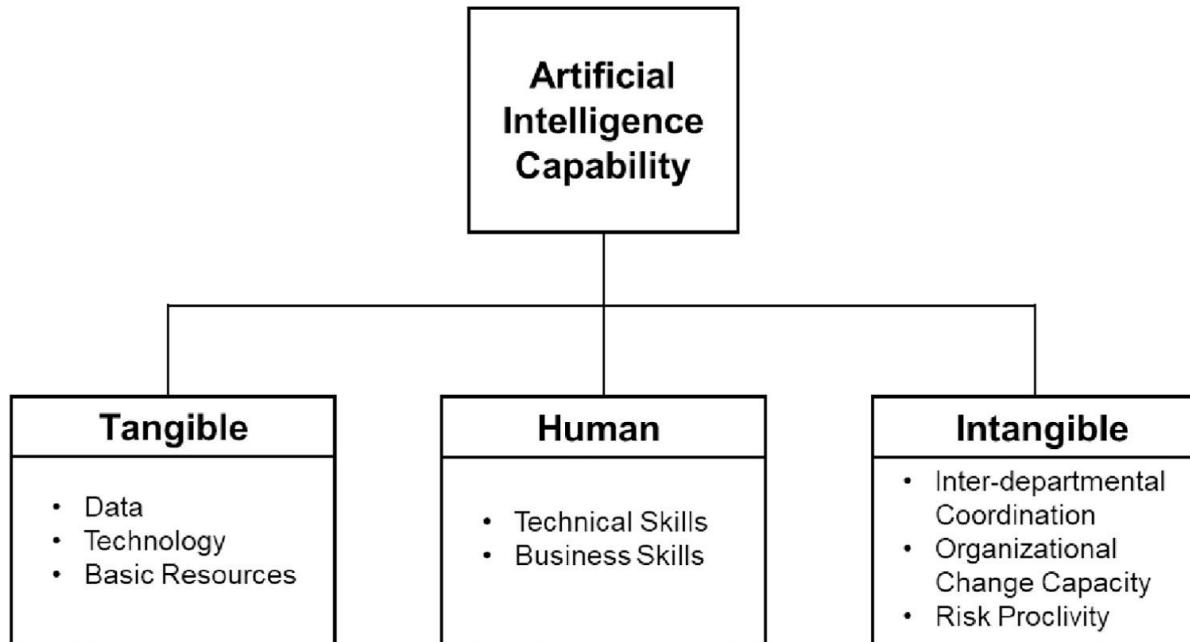
Akter et al. (2016) definiëren BDAC als het vormen van concurrentiekracht door middel van het bieden van zakelijke inzichten met behulp van gegevensbeheer, structuur en talent. Deze definitie heeft Akter et al. (2016) in eigen woorden opgesteld aan de hand van Kiron, Prentice en Ferguson (2014):

“Big data analytics capability (BDAC) is broadly defined as the competence to provide business insights using data management, infrastructure (technology) and talent (personnel) capability to transform business into a competitive force.”

Big data analytics capability models

Nu er meer zicht is op wat big data, *big data analytics* en *big data analytics capabilities* inhouden, verdiepen wij ons in dit onderdeel specifiek in de *big data analytics capability models*. In de literatuur zijn verschillende modellen ontwikkeld die hier – al dan niet gedeeltelijk – over gaan. Zo hebben Gupta en George (2016) een classificatie ontwikkeld voor *big data resources*. Dit leidde tot een model waarin de *big data analytics capabilities* geïdentificeerd werden. In navolging hiervan hebben Mikalef en Gupta in 2021 onderzoek gedaan naar de AI-capaciteit en categorisering van middelen (zie figuur 2.2). Daar waar in dit figuur *Artificial Intelligence Capability* staat, stond in het model van Gupta en George op deze plek ‘*big data analytics capability*’. Het model laat zien dat er 3 categorieën te onderscheiden zijn voor *Artificial Intelligence Capability*, te weten tangible (tastbaar), human (menselijk) en intangible (ongrijpbaar).

Figuur 2.2 AI-capaciteit en categorisering van middelen



Bron: Mikalef en Gupta, 2021.

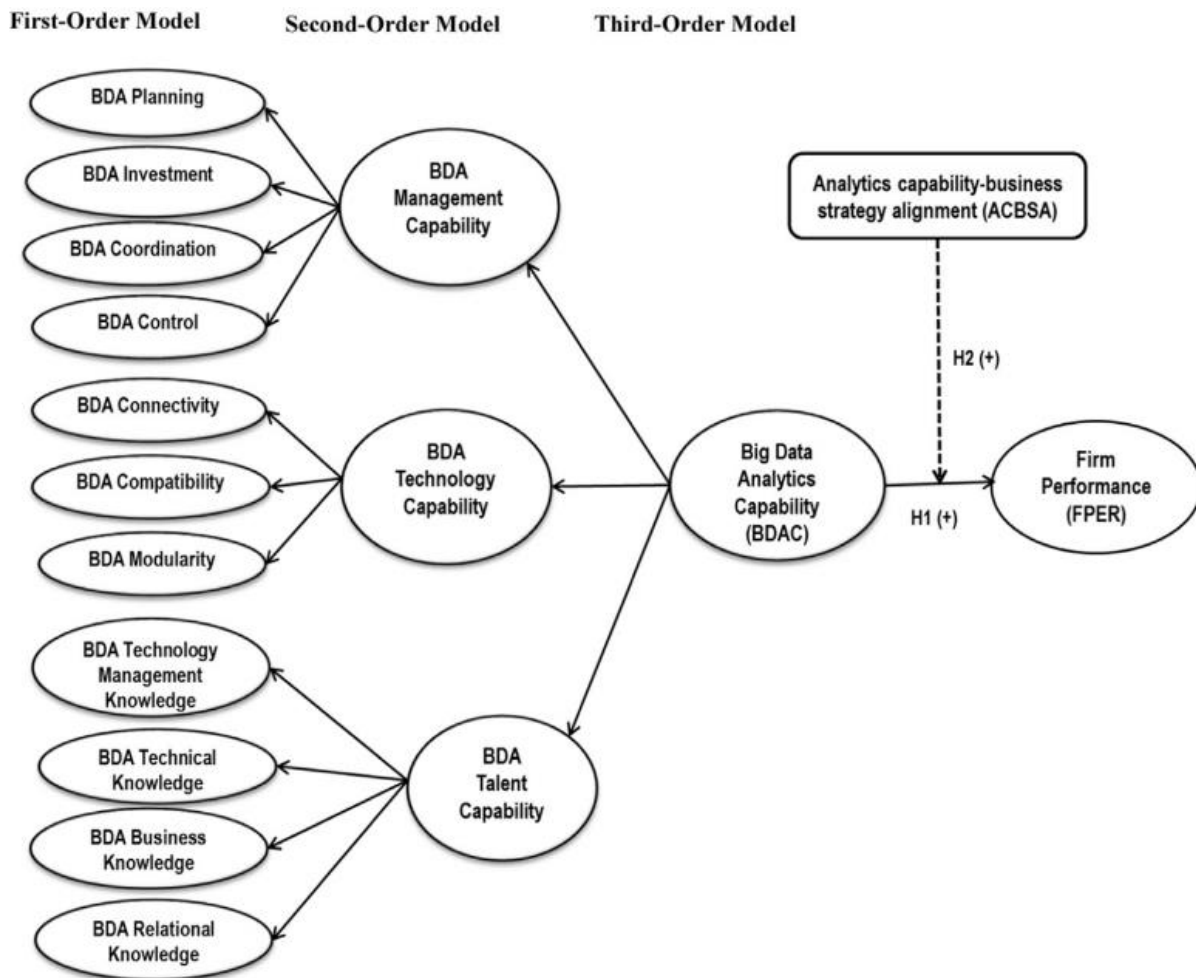
Om erachter te komen wanneer en hoe organisaties *big data analytics capabilities* opbouwen, hebben Dubey et al. (2018) dit onderzocht binnen de *supply chain agility* (keten) om concurrentievoordeel te behalen. Ten opzichte van dit onderzoek (non-profit sector) valt de *supply chain agility* in de profit sector. In de profit sector is het doel vaak om winst te maken. In de non-profit sector is dit niet het geval. Maar, de onderzoekers hebben vijf relevante vragen over *big data analytics capabilities* opgesteld. Deze vragen kunnen branche-onafhankelijk onderzocht en beantwoord worden.

De vragen kunnen op een vijfpuntschaal beantwoord worden, in hoeverre de respondenten het met de volgende stellingen eens zijn:

1. *We use advanced tools (like optimization/regression/simulation) for data analysis.*
2. *We use data gathered from multiple sources (like company reports, tweets, Instagram, YouTube) for data analysis.*
3. *We use data visualization techniques to assist decision makers in understanding complex information extracted from large data.*
4. *Our dashboards display information, which is useful for carrying out necessary diagnosis.*
5. *We have connected dashboard applications or information with the manager's communication devices.*

Naast Mikalef en Gupta (2021) hebben Akter et al. (2016) ook een model ontwikkeld om de organisatieprestaties te meten (figuur 2.3). Hiervoor hebben zij eerst twee Delphistudies uitgevoerd om de subdimensies van BDAC onder de geïdentificeerde primaire dimensies (*second-order model*, tweede 'kolom' in figuur 2.3) te verkennen. De drie primaire dimensies zijn managementbekwaamheid, infrastructuurcapaciteit en talentgeschiktheid. Dit model richt zich meer op de relatie tot de afstemming van de bedrijfsstrategie op managementniveau dan op de individuele medewerkers of teams.

Figuur 2.3 Onderzoeksmodel met de relatie tussen *big data analytics capabilities* en organisatieprestaties in relatie tot de afstemming van de bedrijfsstrategie



Bron: Akter et al., 2016.

2.3. Governmental performance

Overheidsprestaties omvatten verschillende interpretaties en zijn op verschillende manieren en vlakken te meten. Te denken valt aan financiële prestaties, waarbij de vraag is of de overheid niet meer geld uitgeeft dan er binnenkomt. Hierbij kan ook gekeken worden of de uitgaven legitiem zijn. Een andere manier is om de prestaties richting de burgers te bekijken, omdat overheden een verplichting naar burgers toe hebben. De Nederlandse overheid heeft een informatieplicht en geeft uitleg over de werking van de overheid (Rijksoverheid, z.d.b). Veel onderzoeken over overheidsprestaties gaan daarom ook over prestatiemetingen onder burgers. Het betrekken van burgers bij het proces van prestatiemetingen voegt waarde toe aan het algehele proces (Holzer & Kloby, 2005).

Yan en Ting (2018) hebben – als voorbeeld op de prestatiemetingen onder burgers volgens Holzer en Kloby (2005) – burgerevaluaties gebruikt om overheidsprestaties te meten. Als voorbeeld zijn de burgers (inwoners) gevraagd of het voor hen begrijpelijk is wat de functies en opdrachten van verschillende uitvoeringspraktijken (agencies) zijn. Dit is volledig vanuit het perspectief van de burger.

In eerder onderzoek gaat het niet alleen om de prestaties van de overheidsorganisaties (volgens burgers), maar ook de prestaties van medewerkers. Zo heeft Verbeeten (2008) in Nederland onderzoek gedaan op het gebied van de invloed van prestatiebeheer op de prestaties in organisaties in de publieke sector. Hieruit is vooral naar voren gekomen dat duidelijke en meetbare doelen geassocieerd worden met zowel kwantiteits- als kwaliteitsprestaties. Maar, dat bovendien het gebruik van (financiële) stimulansen (*incentives*) – positief hiermee relateert.

Voor dit onderzoek hebben overheidsprestaties betrekking op de organisatorische prestaties van de overheidsorganisaties zelf, zoals het (efficiënt en) goed uit kunnen voeren van de werkzaamheden. Het gaat hierbij niet om de interpretatie volgens burgers (burgeronderzoek) of het bevorderen van prestaties door hier (financiële) stimulansen tegenover te zetten. Brewer en Selden (2000) hebben een concept voor het meten van organisatieprestaties ingedeeld in interne en externe dimensies van efficiëntie, effectiviteit en eerlijkheid. Als vervolg op dit onderzoek heeft Kim (2004) de relatie tussen factoren op individueel niveau en organisatorische prestaties onderzocht door deze empirisch in de overheidsinstanties van Korea te testen. De organisatorische prestaties (*organizational performance*) worden in zes categorieën verdeeld. Deze categorieën komen overeen met en zijn overgenomen van de eerdere categorisatie van Brewer en Selden (2000). Dit is het uitgangspunt voor het huidige onderzoek, waar de focus ligt op het organisatieperspectief van medewerkers zelf. Door middel van vragen die aan respondenten (overheidsmedewerkers) voorgelegd zijn, wordt dit gemeten. De verschillende categorieën (dimensies) die Brewer en Selden (2000) onderschrijven, zijn in tabel 2.1 weergegeven.

Tabel 2.1 Categorieën (dimensies) van overheidsprestaties

		Administratieve waarden		
		<i>Efficiency</i>	<i>Effectiveness</i>	<i>Fairness</i>
Organisatorische focus	Intern	Interne efficiëntie	Interne effectiviteit	Interne eerlijkheid
	Extern	Externe efficiëntie	Externe effectiviteit	Externe eerlijkheid

Bron: Brewer & Selden, 2000.

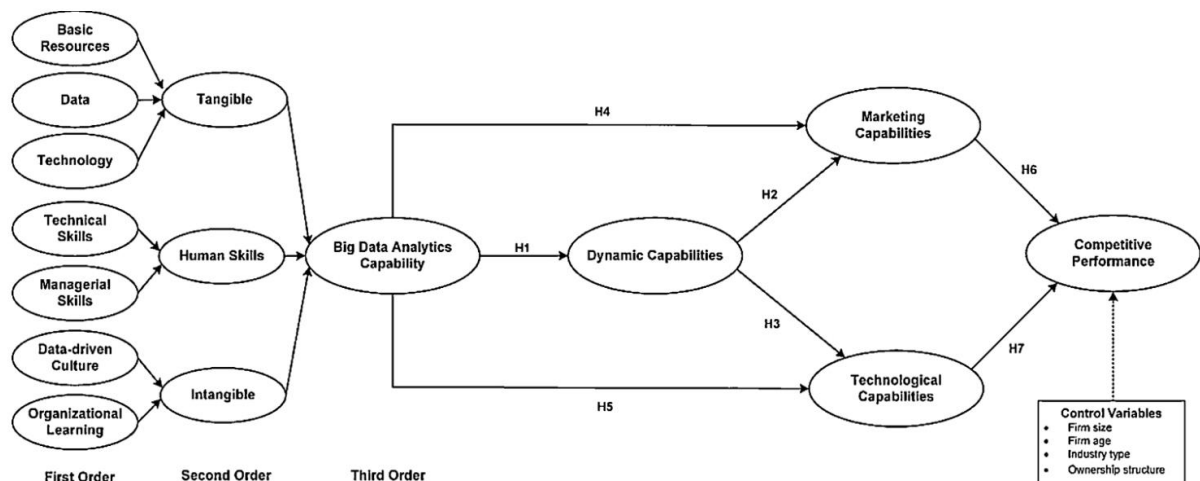
Een voorbeeldvraag uit de eerste categorie, waarbij de respondenten gevraagd zijn hun antwoord op een vijfpuntschaal te geven, is: “*My organization has made good use of my knowledge and skills in looking for ways to become more efficient.*” De efficiëntie verwijst naar de mogelijkheid van een organisatie om haar middelen effectief en efficiënt in te zetten. Te denken valt aan het maximaliseren van de productiviteit en minimaliseren van verspilling. De effectiviteit gaat over de mate waarin een organisatie haar beoogde doelen en doelstellingen bereikt. Als laatste gaat de eerlijkheid over het rechtvaardige belang, om onpartijdig te handelen. Hierbij wordt intern gekeken naar de organisatie zelf en extern wordt gekeken in relatie tot externe belanghebbenden (Brewer & Selden, 2000; Kim, 2004).

2.4. Big data analytics & government performance

In de voorgaande paragrafen zijn *big data analytics* en *government performance* los behandeld. De interessante relatie tussen deze twee is nauwelijks (empirisch) onderzocht. Mergel, Rethemeyer en Isett (2016) hebben wel in de publieke sector onderzocht wat het potentieel van big data analytics is. Het gaat hier echter om beschrijvend onderzoek zonder eigen data te verzamelen. De uitkomsten geven aan dat big data in de publieke sector contextspecifiek is en zinvol gecombineerd moet worden met verzamelde data om waarde voor het publieke programma (werkzaamheden) te creëren. Daarnaast wordt ook onderschreven dat er aandachtspunten zijn bij het gebruik van big data in de publieke sector, zoals ethische en privacy kwesties, beveiliging en de doeltreffendheid.

Buiten de publieke sector om hebben meerdere onderzoeken de relatie tussen *big data analytics* en organisatorische prestaties onderzocht. Recentelijk hebben Mikalef et al. (2020) laten zien hoe een model eruit ziet waarin *big data analytics capabilities* bij kunnen dragen aan competitieve prestaties. In figuur 2.4 is dit conceptueel model als (gevisualiseerd) voorbeeld weergegeven. Het onderzoek van Mikalef et al. (2020) heeft zich meer op de profit sector gericht. De linker twee 'kolommen' komen grotendeels overeen met de onderste 'rij' uit het *AI-capabilitiesmodel* van Mikalef en Gupta (2021; figuur 2.2). Daarnaast hebben Akter et al. (2016) een model ontwikkeld waarin zij onderschrijven dat *big data analytics capabilities* bijdragen aan organisatorische prestaties in relatie tot de bedrijfsstrategie. Dit bespreekt Teece (2014) ook aan de hand van zijn onderzoek naar de rol van *dynamic capabilities*, bij het verklaren van succes van multinationale ondernemingen en hun prestaties.

Figuur 2.4 Onderzoeksmodel met de relatie tussen *big data analytics capabilities* en competitieve prestaties



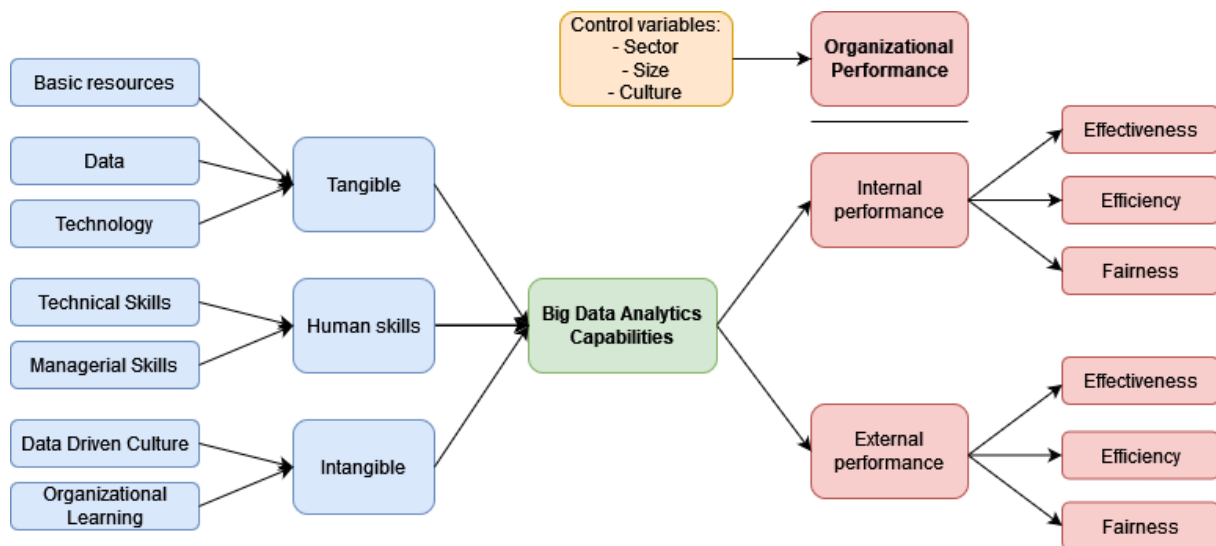
Bron: Mikalef et al., 2020.

Kortom, vanuit de wetenschappelijke literatuur is er weinig bekend over de combinatie van *big data analytics* (capabilities) en overheidsprestaties. Wel hebben meerdere onderzoeken aangetoond dat *big data analytics* bij kunnen dragen aan organisatorische prestaties in de commerciële sector (profit). Al gaat het vaak over onderzoeken die in andere landen hebben plaatsgevonden. Op die manier zijn de uitkomsten enigszins contextafhankelijk.

2.5. Conceptueel model

Voor het huidige onderzoek is een conceptueel model gemaakt op basis van het literatuuronderzoek. Dit conceptuele model beschrijft de verschillende *big data analytics capabilities* die bijdragen aan organisatorische prestaties in Nederlandse overheidsorganisaties en deze in kaart kunnen brengen (zie figuur 2.5). De categorieën en factoren in het conceptueel model sluiten aan bij de resultaten uit de literatuurstudie. Hierin is een combinatie gemaakt tussen de *big data analytics capabilities* en organisatorische prestaties bij Nederlandse overheidsorganisaties. De organisatorische prestaties zijn uitgesplitst naar interne en externe prestaties.

Figuur 2.5 Conceptueel model voor *big data analytics capabilities* binnen Nederlandse overheidsorganisaties



Het conceptueel model geeft de samenhang van mogelijke factoren weer die een rol kunnen spelen bij organisatorische prestaties binnen Nederlandse overheidsorganisaties. Hierin wordt gekeken welke invloed er is op de interne en externe prestaties. De mogelijke factoren zijn onderverdeeld in drie hoofdcategorieën, te weten: tastbaar (*tangible*), menselijk (*human skills*) en ongrijpbaar (*intangible*). Deze zijn onderverdeeld in specifiekere subcategorieën. Uiteindelijk wordt ook gekeken welke controlevariabelen er zijn die invloed op de organisatorische prestaties kunnen hebben. Dit zijn bijvoorbeeld de sector (branche), organisatiegrootte en organisatiecultuur.

2.6. Hypotheses

Voor dit onderzoek zijn twee hypothesen op te stellen. Deze zijn opgesteld aan de hand van de literatuur en het conceptuele model en worden met behulp van de resultaten later aangenomen of verworpen.

De hypothesen zijn:

1. Een hogere score op *big data analytics capabilities* leidt tot hogere interne organisatorische prestaties.
2. Een hogere score op *big data analytics capabilities* leidt tot hogere externe organisatorische prestaties.

Met deze hypothese wordt getoetst of *big data analytics capabilities* leiden tot hogere interne en/of externe organisatorische prestaties. Hierin zitten de verkapte subcategorieën van *big data analytics capabilities*. Via een post-hoc analyse wordt de invloed van de subcategorieën op de interne en externe organisatorische prestaties onderzocht. Het gaat hierbij om de basismiddelen, data, technologie, technische vaardigheden, managementvaardigheden, datagedreven cultuur en lerende organisatie.

3. Methode

Hoofdstuk 3 beschrijft de methodische aanpak van deze thesis. In paragraaf 3.1 zal een toelichting op de aanpak geschetst worden. In paragraaf 3.2 wordt de onderzoeksmethode toegelicht. Dit gaat over het instrument dat voor de dataverzameling gebruikt wordt, namelijk een vragenlijst. Vervolgens wordt in paragraaf 3.3 de steekproef bepaald en toegelicht. Paragraaf 3.4 omschrijft de procedure rondom de dataverzameling, zoals hoe en waar de vragenlijst uit wordt gezet. Tot slot wordt in paragraaf 3.5 de methode voor data-analyse toegelicht.

3.1. Aanpak

Om tot de beantwoording van de onderzoeksvraag te komen, wordt – naast het literatuuronderzoek – nog een onderzoeksmethode toegepast. Hoe dit onderzoek eruit ziet en wat de aanpak hiervoor is, wordt hierna beschreven.

Wanneer we het over de methodologie van dit onderzoek hebben dan komt het neer op theorie testing. Bestaande theorie wordt getoetst binnen een – voor dit onderzoek – nieuwe sector: overheidsorganisaties in Nederland. Omdat het voor dit onderzoek om een ‘nieuwe’ sector gaat, is hier ook gedeeltelijk sprake van theorie building. Het doel hierbij is om te achterhalen welke *big data analytics capabilities* bijdragen aan organisatorische prestaties binnen Nederlandse overheidsorganisaties. Vanuit wetenschappelijke literatuur wordt opgehaald hoe dit meetbaar is, uit bijvoorbeeld andere sectoren. Dit wordt hierna voorgelegd aan een nieuwe groep, namelijk medewerkers (ambtenaren) van Nederlandse overheidsorganisaties (bijvoorbeeld ministeries, gemeenten, provincies en zelfstandige bestuursorganen als DUO en Rijkswaterstaat (Nationale Ombudsman, z.d.; Overheid.nl, z.d.; Rijksoverheid, z.d.a)). Via een vragenlijst worden nieuwe inzichten opgehaald. Hierbij wordt theorie opgebouwd voor *big data analytics capabilities* binnen Nederlandse overheidsorganisaties. Met deze methode, om oorzaken en resultaten te koppelen, kunnen conclusies getrokken worden over hoe een oorzaak (of reeks oorzaken) bijdraagt aan het produceren van een uitkomst (Beach & Brun Pedersen, 2013). In dit geval *big data analytics capabilities* in relatie tot organisatorische prestaties. Als bewijstechniek gaat het om deductief onderzoek. De uit de literatuur naar boven gekomen factoren worden in de praktijk voorgelegd en getoetst. Hiervoor worden eerst hypothesen opgesteld (Bell et al., 2019).

3.2. Instrument dataverzameling

Om de onderzoeksvraag te beantwoorden wordt als instrument een vragenlijst (survey) gebruikt om data te verzamelen. De vragenlijst is kwantitatief opgesteld. Deze keuze is gemaakt om verschillende factoren in een relatief korte tijdsperiode bij een grote populatie te kunnen achterhalen en testen. Omdat het doel is om de resultaten te generaliseren naar een bredere doelgroep, is gekozen voor een empirisch onderzoek als vragenlijstmethode (survey; Verschuren & Doorewaard, 2007). Bewust is gekozen om geen kwalitatief onderzoek te doen, omdat de uitkomsten dan gebaseerd zijn op een kleinere steekproef, bijvoorbeeld over enkele gemeenten/respondenten. Via een vragenlijst wordt de

data bij een groter aantal Nederlandse overheidsorganisaties opgehaald zodat over een grotere populatie uitspraken gedaan kunnen worden. Op deze manier wordt ook aangesloten bij eerdere onderzoeken, zoals Kim (2004) en Mikalef et al. (2020) en kan hiermee vergeleken worden. De gebruikte vragen (items) kunnen daarnaast geanalyseerd worden op betrouwbaarheid en validiteit. Ook kunnen significante relaties getoetst worden.

Om deze vragenlijst uit te kunnen zetten, is voorafgaand literatuuronderzoek gedaan naar bestaande *big data analytics capabilities* en organisatorische prestaties (hoofdstuk 2). Dit literatuuronderzoek is gebruikt om het theoretisch kader vorm te geven. Hieruit worden relevante thema's en onderwerpen, in combinatie met de daaraan gekoppelde vragen, gehaald om hierna in een eigen online vragenlijst (ook wel enquête/survey genoemd) uit te zetten. De vragenlijst is van kwantitatieve aard (voornamelijk stellingen en gesloten vragen).

Met behulp van het literatuuronderzoek is de vragenlijst tot stand gekomen. Hierbij is gekeken naar de geschiktheid van relevante literatuur, modellen en de beschikbaarheid van reeds opgestelde vragen in relatie tot BDAC en overheidsprestaties. In de vragenlijst komen meerdere items aan bod. De hoofditens zijn *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties³. De organisatorische prestaties kunnen uitgesplitst worden naar interne en externe prestaties.

In tabel 3.1 zijn de bronnen weergegeven waaruit de stellingen afkomstig zijn voor de vragenlijst van dit onderzoek. De vragen sluiten aan bij vier verschillende dimensies; *tangible* (tastbaar), *human skills* (menselijke vaardigheden), *intangible* (ontastbaar) en *organizational performance* (organisatorische prestaties). De vragenlijst is in bijlage 2 te vinden.

Tabel 3.1 Overzicht bronnen voor totstandkoming vragenlijst

Dimensie	Bron vragen
<i>Tangible</i> (tastbaar)	Mikalef et al. (2020)
<i>Human skills</i> (menselijke vaardigheden)	Sternkopf en Mueller (2018)
<i>Intangible</i> (ontastbaar)	Mikalef et al. (2020)
<i>Organizational performance</i> (organisatorische prestaties)	Kim (2004)

Op de volgende pagina worden de keuzes voor de dimensies verder uitgelicht.

³ Digitaal leiderschap en datageletterdheid komen niet in het theoretisch kader en daarmee ook niet in dit onderzoek aan bod. Deze onderwerpen en stellingen zijn door medestudenten onderzocht in afzonderlijke thesissen.

Voor dit onderzoek zijn de items die vallen onder *big data analytics* van belang. Mikalef et al. (2020) hebben de relatie tussen *big data analytics capabilities* en organisatorische prestaties in Noorwegen onderzocht. Binnen dit item wordt onderscheid gemaakt tussen tastbare vaardigheden (*tangible*), menselijke vaardigheden (*human skills*) en immateriële vaardigheden (*intangible*). De stellingen over menselijke vaardigheden, onderverdeeld naar managementvaardigheden en technische vaardigheden, worden in dit onderzoek niet overgenomen. Daarvoor zijn passende vervangende stellingen vanuit datageletterdheid gebruikt. Deze zijn afkomstig uit een onderzoek van Sternkopf en Mueller (2018) en gaan over een volwassenheidsmodel voor datageletterdheid in niet-overheidsorganisaties. Hiervoor is gekozen om de vragenlijst niet onnodig lang te maken en geen dubbele vragen te stellen. Deze bron is verder toegelicht en uitgewerkt in een separate thesis over datageletterdheid. Met de vragen bij datageletterdheid kunnen de *human skills* items gevuld worden. In tabel 3.2 zijn de oorspronkelijke stellingen weergegeven (links) en door welke deze vervangen zijn (rechts).

Tabel 3.2 Vervangende stellingen t.b.v. het vragenblok ‘menselijke vaardigheden’ rondom BDAC

BDAC Mikalef et al. (2020)	Datageletterdheid Sternkopf en Mueller (2018)
MS1. <i>Our 'big data analytics' managers are able to understand the business need of (and collaborate with) other functional managers, suppliers, and customers to determine opportunities that big data might bring to our business.</i>	We zien data als een veilige bron en zien het als een aanjager voor vooruitgang en ondersteuning voor bestaande en geplande activiteiten
MS2. <i>Our 'big data analytics' managers are able to coordinate big data-related activities in ways that support other functional managers, suppliers, and customers.</i>	Hoger management en leiders ondersteunen data-initiatieven
MS3. <i>Our 'big data analytics' managers are able to understand and evaluate the output extracted from big data.</i>	1. We kunnen data output en resultaten vol vertrouwen en kritisch interpreteren 2. We kunnen een reeks tabellen, diagrammen en grafieken lezen en begrijpen
TS1. <i>Our 'big data analytics' staff has the right skills to accomplish their jobs successfully.</i>	We kunnen werken met geavanceerde statistieken (bijv. inferentiële weergaven van data, lineaire regressie, beslisbomen)
TS2. <i>Our 'big data analytics' staff is well trained</i>	We hebben goede kennis van verschillende databronnen en de meest relevante kunnen worden gekozen uit een selectie van databronnen

Toelichting: MS staat voor stellingen rondom managementvaardigheden (*managerial skills*) en TS staat voor technische vaardigheden (*technical skills*).

Voor de organisatorische prestaties zijn stellingen vertaald van Kim (2004). Deze stellingen hebben betrekking op individuele factoren voor organisatorische prestatie in overheidsorganisaties. In het

onderzoek van Kim is voor de antwoordmogelijkheden gebruik gemaakt van een likertschaal met 5 antwoordmogelijkheden. Mikalef et. al (2020) hebben van een 7-puntslikertschaal gebruik gemaakt. De meeste stellingen vanuit de literatuur zijn met een 7-puntsschaal uitgezet. Daarom is voor dit onderzoek gekozen om de 7-puntsschaal te hanteren. De resultaten voor *organisational performance* vanuit de literatuur worden omgerekend (van 5- naar 7-puntsschaal) zodat deze te vergelijken zijn met de resultaten uit het huidige onderzoek.

3.3. Steekproef

Bij het doen van uitspraken bij dit kwantitatieve onderzoek is het nodig om met een steekproef te werken. Door de omvang van de hele populatie is het niet haalbaar om deze allemaal te bevragen en een reactie (respons) van te ontvangen. Valide uitspraken kunnen wel over de populatie gedaan worden als de resultaten gaan over een representatieve steekproef. In 2019 waren er 350.000 medewerkers werkzaam bij Nederlandse overheidsorganisaties (in 4 sectoren), zie tabel 3.3. Met ongeveer 350.000 medewerkers binnen de Nederlandse overheidssector is een steekproef van 384 respondenten nodig (Kruisman, 2021). Als de omvang van de doelgroep boven enkele duizenden uitkomt, is de steekproef doorgaans 350 (maximaal 384, ongeacht de omvang). Dit is op basis van een 5 procent foutenmarge en 95 procent betrouwbaarheidsniveau.

Tabel 3.3 Aantal medewerkers bij Nederlandse overheidsorganisaties in 2019 naar sector

Sector	Aantal medewerkers
Rijk	126.986
Gemeenten	153.668
Provincies	11.050
Defensie	59.156
Totaal	350.860

Bron: BZK, Kerncijfers Personeel Overheid en Onderwijs, 2019.

Achtergrondkenmerken respondenten

De ingevulde vragenlijsten van 120 respondenten zijn voor dit onderzoek bruikbaar (afgerond en voldoen aan de voorwaarden, om bij non-profit organisaties in Nederland te werken). Van de respondenten is ongeveer twee derde (68 procent man), is de gemiddelde leeftijd 45 jaar, heeft de respondent gemiddeld 20 jaar werkervaring en is deze hoogopgeleid (hbo/wo; 94%). Vooral medewerkers (niet leidinggevend; 63%) hebben de vragenlijst ingevuld, gevolgd door autonome leidinggevend (24%). In bijlage 4 zijn alle achtergrondkenmerken te vinden.

In tabel 3.4 is de respondentenverdeling naar organisatiegrootte te vinden waar de respondent werkt. De respondenten zijn voornamelijk werkzaam bij grote organisaties, met meer dan 100 medewerkers (87%).

Tabel 3.4 Respondentenverdeling naar organisatiegrootte van Nederlandse overheidsorganisaties in 2021 (in aantallen en procenten)

Categorie	Aantal	Percentage
Minder dan 10	1	1%
10-50	5	4%
51-100	4	3%
101-1.000	72	60%
Meer dan 1.000	38	32%
Totaal	120	100%

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Wanneer de achtergrond van de respondenten verder bekeken wordt, dan zien we dat vier op de vijf respondenten (79%) ervaring heeft met (*big*) *data analytics* (tabel 3.5). Ruim een derde (38%) van de respondenten geeft aan hier meer dan 4 jaar ervaring mee te hebben.

Tabel 3.5 Aantal jaar ervaring van de respondenten bij Nederlandse overheidsorganisaties met (*big*) *data analytics* in 2021 (in aantallen)

Categorie	Aantal	Percentage*
Geen, ik heb er geen ervaring mee	25	21%
Minder dan 1 jaar	11	9%
1-2 jaar	16	13%
2-3 jaar	10	8%
3-4 jaar	12	10%
Meer dan 4 jaar	46	38%
Totaal	120	100%

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

* Door afronding telt deze kolom niet op tot 100 procent.

Bij de interpretatie van de data en daaruit volgende conclusies moet voorzichtigheid geboden worden. Als er naar de vereiste steekproef voor een representatief beeld gekeken wordt, blijkt dat er voor dit onderzoek te weinig respondenten zijn. Daarnaast lijkt het aantal jaar ervaring van de respondenten met (*big*) *data analytics* niet normaal verdeeld te zijn. Omdat hier geen cijfers over te vinden zijn, kan dit niet met zekerheid gezegd worden. De kans dat 38 procent van alle Nederlandse ambtenaren al meer dan 4 jaar met (*big*) *data analytics* werkt, is niet aannemelijk. Vanwege de moeilijkheid van de vragenlijst, specifieke thema's en lengte lijkt de steekproef daarom niet representatief voor alle Nederlandse ambtenaren.

3.4. Procedure dataverzameling

Deze paragraaf beschrijft de uitgangspunten en procedures voor de dataverzameling. Als eerste is gekozen voor een uitgangspunt op welk niveau de vragenlijst ingevuld moet worden. De ingestoken keuze is vanuit het 'teamperspectief'. De bevraging is hierop aangepast zodat de respondenten namens het team waarin zij werken kunnen reageren. Op die manier is voorkomen dat het perspectief van een individu als uitgangspunt wordt genomen, omdat het gaat om de prestaties van het bedrijf (team/afdeling). Er is bewust gekozen om dit niet op organisatieniveau uit te vragen, omdat bijvoorbeeld een medewerker van de gemeente op de sportafdeling niet (goed) zicht heeft op alle andere gemeenteafdelingen en een organisatie dit niet van al haar medewerkers weet.

De vragenlijst wordt geprogrammeerd in [Alchemer](#) (voorheen SurveyGizmo). Dit is één van de bekendste en grootste internationale vragenlijsttools. Deze vragenlijsttool is AVG-compliant. De data wordt opgeslagen in Europa (Frankfurt, Duitsland). Hierbij wordt de mogelijkheid om IP-adressen en andere achtergrondgegevens te achterhalen uitgezet. Daarnaast is een vragenlijst via Alchemer eenvoudig voor gebruikers in te vullen, zowel op een telefoon als computer (Alchemer, 2021). Hierdoor kunnen de respondenten zelf kiezen om de vragenlijst in te vullen en zijn de antwoorden in de analyse en rapportage niet herleidbaar tot een individu. Als een respondent de resultaten van het onderzoek wilt ontvangen, dan kan deze zijn of haar e-mailadres opgeven. Deze wordt in de dataset direct losgekoppeld zodat er geen koppeling met een individu meer mogelijk is.

Voor de dataverzameling is een open invullink aangemaakt die alle ambtenaren in kunnen vullen. Op deze manier kunnen medewerkers van Nederlandse overheidsorganisaties (waaronder gemeenten) de vragenlijst ook doorsturen naar collega's. Op deze manier is er geen beperking dat alleen de ontvanger van de link de vragenlijst in kan vullen (gesloten link). In de vragenlijst is ook een vraag opgenomen waar de respondent werkzaam is (binnen een Nederlandse overheidsorganisatie of niet). Op deze manier kunnen vragenlijsten uitgesloten worden wanneer respondenten de vragenlijst invullen die buiten de scope van het onderzoek vallen. Daarnaast is in de vragenlijst verwezen naar de privacyrichtlijnen, zoals het vertrouwelijk omgaan met de data en dat de antwoorden van de respondenten in publicaties nooit te herleiden zijn.

Het voornemen was om de vragenlijst uit te zetten via de Vereniging Sport en Gemeenten (VSG), waar ongeveer 70 procent van de gemeenten bij aan zijn gesloten (VSG, z.d.). Daarnaast was voornemens om de vragenlijst ook uit te zetten via contacten bij de Vereniging van Nederlandse Gemeenten (VNG), waar alle Nederlandse gemeenten bij aan zijn gesloten (VNG, 2021). In 2021 telt Nederland 352 gemeenten (CBS, 2020). Na aanvullend onderzoek bleek het helaas niet mogelijk te zijn om de vragenlijst via deze kanalen uit te zetten. Onder andere doordat recentelijk andere vragenlijsten al via deze kanalen verspreid waren. Hierdoor is afgeweken van deze kanalen en zijn de vragenlijsten verspreid via actieve benadering van overheidsorganisaties. Zo zijn alle gemeenten per e-mail of contactformulier benaderd met de vraag of ze de vragenlijst (inclusief voorbeelduitnodiging) binnen hun organisatie of enkele afdelingen uit konden zetten. Andere zbo's, zoals de brandweer, DUO en

Rijkswaterstaat, zijn ook benaderd. Daarnaast zijn persoonlijke contacten benaderd die werkzaam zijn bij Nederlandse overheidsorganisaties om de vragenlijst in te vullen en deze verder uit te zetten. Ook is LinkedIn gebruikt voor de verspreiding van de vragenlijst, zowel op openbare als gesloten pagina's (specifiek gericht op Nederlandse overheden en datagerelateerd).

Ook is geprobeerd de vragenlijst internationaal uit te zetten bij partners van de HU die deelnemen aan het DALICO-project (Data Literacy in Context). Het gaat hierbij over Duitsland, Hongarije en Spanje (DALICO, z.d.). Helaas heeft de toezegging op het uitzetten niet geleidt tot noemenswaardige respons uit deze landen om deze data mee te nemen in dit onderzoek. Een internationale vergelijking is in dit onderzoek daarom niet mogelijk.

3.5. Data-analyse

De vragenlijstdata (CSV/SPSS-data) zal in SPSS en met PLS-SEM worden geanalyseerd. SPSS is een statistisch analyseprogramma en wordt vaak door hogescholen, universiteiten en onderzoeksorganisaties gebruikt. Het is laagdrempelig, komt vanuit de verklarende hoek en is vanuit de wetenschap ontwikkeld. Hypotheses testen speelt hierin een belangrijke rol (Kromme, 2017). Hierin worden verschillende technieken gebruikt om de data te analyseren. Allereerst worden beschrijvende statistieken uitgedraaid (descriptives). Dit kunnen bijvoorbeeld rechte tellingen zijn, maar ook de standaarddeviatie⁴ (SD). Daarnaast wordt er dieper in de data gedoken door analyses/toetsen toe te passen om de opgestelde hypothesen te testen (inferentiële statistiek). Als eerste zal gekeken worden naar de mate van samenhang (interne consistentie) tussen meerdere vragen die samen een categorie vormen. De Cronbach's Alpha wordt gebruikt om te kijken of verschillende vragen hetzelfde meten en hetzelfde antwoord geven. Dit wordt daarnaast ook via de PLS-SEM methode⁵ (zie volgende alinea) geanalyseerd. De uitkomsten worden gevalideerd door deze te vergelijken met de uitkomsten uit eerdere onderzoeken, waar de vragen uit afkomstig zijn. De syntax voor de analyse is in bijlage 3 te vinden.

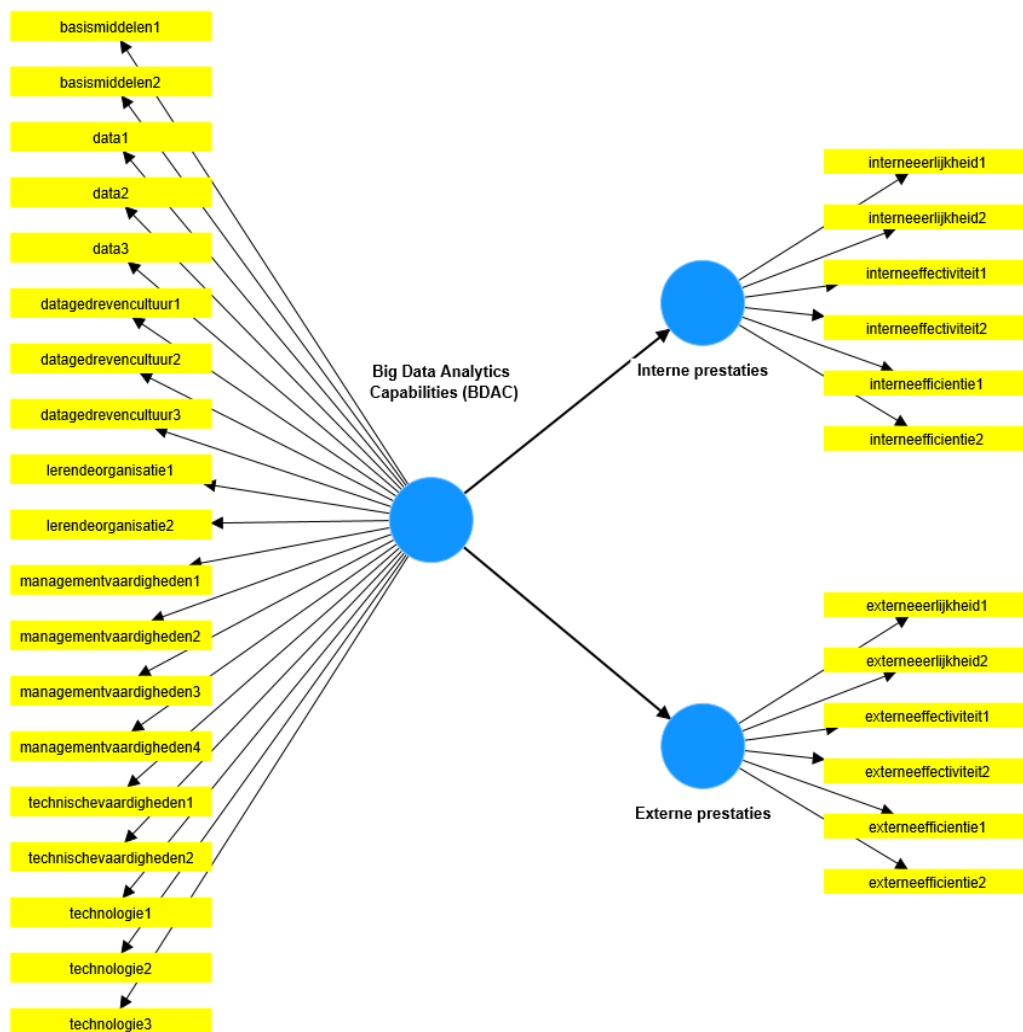
SPSS is vooral gebruikt vanwege het gebruiksgemak en de mogelijkheid om laagdrempelig beschrijvende statistiek uit te draaien. Daarnaast wordt hiermee (in combinatie met het gebruik van Smart PLS 4 en de mogelijkheid om Python te gebruiken) aangetoond dat via verschillende programma's data-analyses gedaan kunnen worden. De beschrijvende resultaten (paragraaf 4.2) zijn met behulp van analyses in SPSS tot stand gekomen.

⁴ De standaarddeviatie is de maat voor de spreiding en geeft aan in hoeverre de in kaart gebrachte waarden afwijken van het gemiddelde. Ervan uitgaande dat de data normaal verdeeld zijn, betekent dit dat 68 procent van de waarden tussen het gemiddelde en één standaarddeviatie liggen.

⁵ In bijlage 4.5 is een schematische weergave van PLS-SEM (*Partial Least Squares based Structural Equation Modeling*) vanuit de literatuur te vinden.

Naast de vergelijking met eerdere onderzoeken wordt verdiepend gekeken naar de relatie tussen variabelen. Hier wordt Partial Least Squares based Structural Equation Modeling (PLS-SEM) voor gebruikt. PLS-SEM test – net als de technieken in SPSS – ook de validiteit van statistische conclusies. Daarnaast biedt PLS-SEM – in tegenstelling tot de standaard statistische technieken, zoals regressie – de mogelijkheid om relaties tussen meerdere onafhankelijke en afhankelijke variabelen tegelijkertijd te testen. Meerdere relaties kunnen geanalyseerd worden, waar bijvoorbeeld lineaire regressieanalyses maar één relatie kunnen analyseren (Gefen et al., 2000). Ook kunnen latente variabelen worden meegenomen (niet direct te observeren variabelen; Haar et al., 2019). PLS-SEM-analyses kunnen bijvoorbeeld via Python of R uitgevoerd worden. Voor dit onderzoek is gebruik gemaakt van SmartPLS 4 om de PLS-SEM-analyse uit te voeren (Ringle et al., 2022). Daarnaast is de PLS-SEM-analyse ook in Python uitgevoerd. Dit om verschillende programma's voor hetzelfde doeleinde te gebruiken. In Python kan dit met de *package plspm*, die speciaal voor de *Partial Least Squares Path Modeling* (PLS-PM) analyse is ontwikkeld (Humble, 2020). Omdat uiteindelijk de uitkomsten van SmartPLS 4 uitgebreider en goed bruikbaar waren, zijn de code en uitkomsten van de Python-analyse niet meegenomen. In figuur 3.1 is de schematische weergave te vinden voor de PLS-SEM-analyse van dit onderzoek.

Figuur 3.1 Schematische weergave PLS-SEM-analyse in SmartPLS 4 t.b.v. de hypotheses



4. Resultaten

In dit hoofdstuk worden de resultaten gepresenteerd. Voorafgaand zijn in het vorige hoofdstuk – in paragraaf 3.3 – de respondenten omschreven (achtergrondkenmerken). Hierna volgen – in paragraaf 4.1 – als eerste de beschrijvende resultaten (statistiek), zoals de uitkomsten per vragenlijstcategorie (zowel voor het huidige onderzoek als resultaten uit de literatuur). Vanaf paragraaf 4.2 zijn verdiepende resultaten weergegeven. In paragraaf 4.2 is de betrouwbaarheid en validiteit van het model onderzocht via een PLS-SEM-analyse. In paragraaf 4.3 worden de hypothesen getoetst aan de hand van een bootstrapping analyse. Tot slot volgt in paragraaf 4.4 nog een post-hoc analyse naar subcategorieën van *big data analytics capabilities* die bijdragen aan de interne en externe organisatorische prestaties.

4.1. Beschrijvende resultaten

Met de categorisaties van *big data analytics capabilities* en organisatorische prestaties kunnen verdiepende analyses gedaan worden. Bij deze beschrijvende statistiek worden belangrijke kenmerken van de data beschreven en samengevat. Het gaat hierbij voornamelijk om centrummaten zoals het gemiddelde en spreidingsmaten zoals de standaarddeviatie (Sirkin, 2006).

De vragen uit het huidige onderzoek zijn afkomstig en vertaald vanuit bestaande wetenschappelijke literatuur. Hier zijn referentie uitkomsten voor beschikbaar. Deze referentie uitkomsten zijn niet in het huidige onderzoek verzameld, maar worden in deze paragraaf wel gepresenteerd om in vergelijking te zetten met de uitkomsten van het huidige onderzoek.

Als eerst worden de gemiddelde scores per categorie voor dit onderzoek op *big data analytics capabilities* en in vergelijking met uitkomsten uit de literatuur weergegeven en toegelicht. Vervolgens worden de uitkomsten en vergelijking voor organisatorische prestaties op eenzelfde manier gepresenteerd. Als laatste worden op een overzichtelijke manier de verschillen inzichtelijk gemaakt.

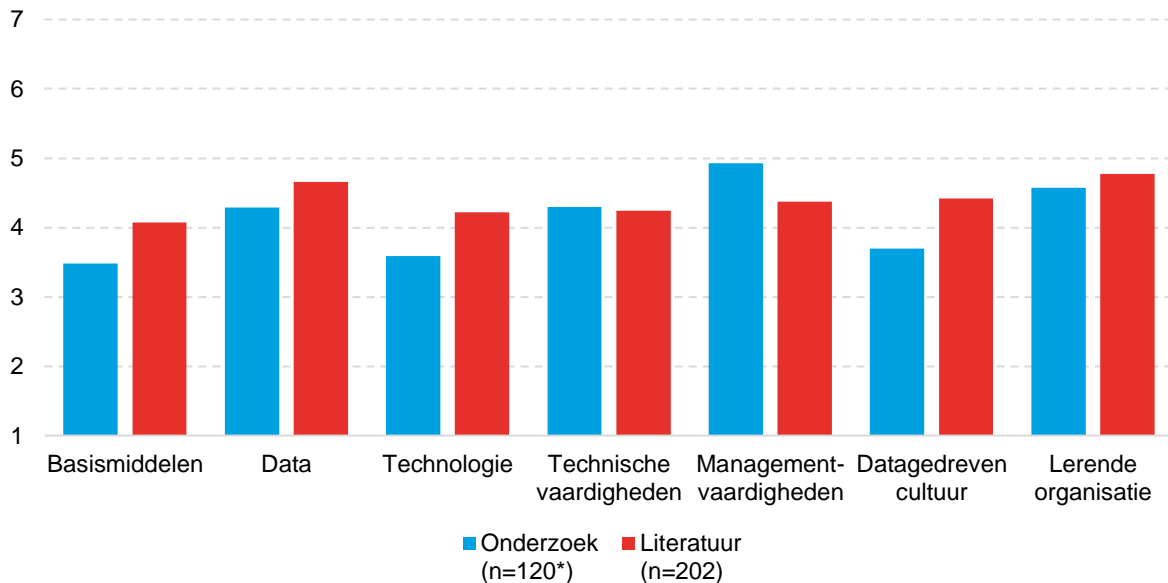
Uitkomsten *big data analytics capabilities*

De resultaten over *big data analytics capabilities* bij Nederlandse overheidsorganisaties zijn met gemiddelde scores in figuur 4.1 weergegeven (blauw). Deze zijn afgezet tegen de gemiddelden vanuit de literatuur (rood). De gemiddelde scores voor de clusters binnen *big data analytics capabilities* variëren bij dit onderzoek van 3,5 voor basismiddelen tot 4,9 voor managementvaardigheden. Het overal gemiddelde ligt op 4.1. In de literatuur varieert dit van 4,1 voor basismiddelen tot 4,8 voor lerende organisatie. Het gemiddelde is hier 4,4.

De standaarddeviatie voor elk cluster geeft aan hoeveel variatie er is tussen de antwoorden van de respondenten. De standaarddeviatie voor dit onderzoek ligt tussen de 1,2 en 1,6. In de literatuur ligt dit tussen de 1,4 en 2,0. Het valt op dat de standaarddeviaties relatief hoog zijn (in verhouding met de antwoordschaal en gemiddelden), wat betekent dat er veel variatie is in de scores van de respondenten.

Dit kan erop wijzen dat sommige organisaties al verder zijn op het gebied dan andere, en dat er meer werk nodig is om dit gelijk te trekken.

Figuur 4.1 Gemiddelde categorieën (o.b.v. stellingen) over *big data analytics capabilities* bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7)



Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021; Mikalef et al., 2020 (vragenlijst uitgezet in 2017).

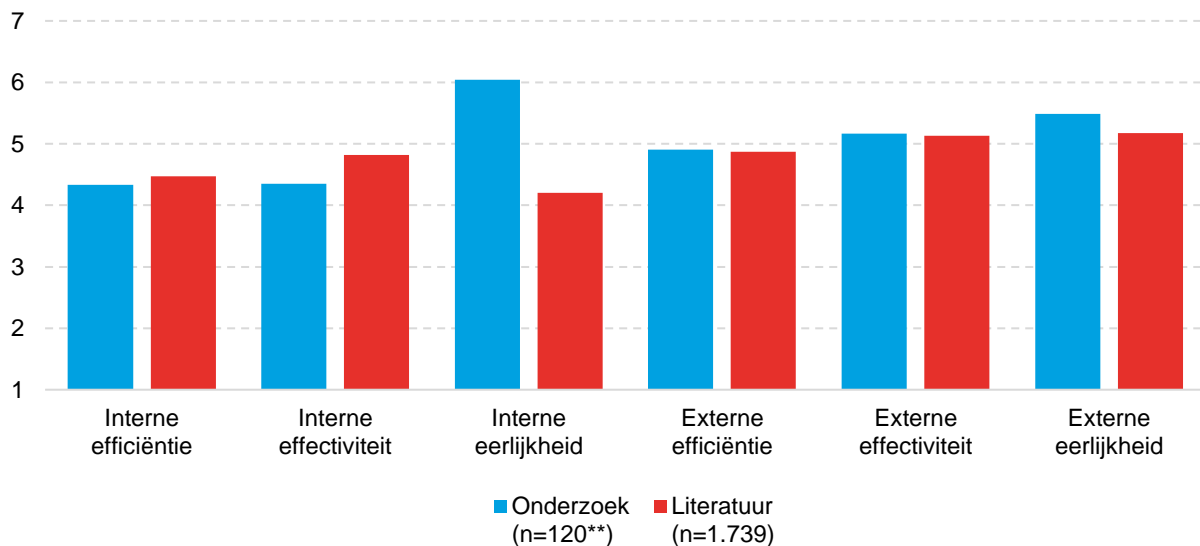
* Het aantal waarnemingen kan per stelling en categorie verschillen doordat de antwoordcategorie 'niet van toepassing' op missing gezet is. Het aantal waarnemingen ligt tussen de 94 en 120.

Uitkomsten organisatorische prestaties

De vragen uit het huidige onderzoek zijn afkomstig en vertaald vanuit bestaande wetenschappelijke literatuur. Deze zijn samen met de uitkomsten van het huidige onderzoek in figuur 4.2 weergegeven.

De gemiddelde scores voor organisatorische prestaties liggen hoger dan de *big data analytics capabilities* (figuur 4.1) en komen gemiddeld uit op 5,0. De variatie hierbij loopt van 4,3 voor interne efficiëntie tot 6,0 voor interne eerlijkheid. In de literatuur varieert dit van 4,2 voor interne eerlijkheid tot 5,2 voor externe eerlijkheid. Het gemiddelde is hier 4,8. De standaarddeviatie voor dit onderzoek ligt tussen de 1,0 en 1,4. In de literatuur ligt dit tussen de 1,1 en 1,4

Figuur 4.2 Gemiddelde categorieën (o.b.v. stellingen) over organisatorische prestaties bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7*)



Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021; Kim, 2004 (data verzameld in 2001).

* De stellingen over organisatorische prestaties zijn in de literatuur op een 5-puntsschaal uitgevraagd. Deze zijn omgerekend naar het gemiddelde op een 7-puntsschaal.

** * Het aantal waarnemingen kan per stelling en categorie verschillen doordat de antwoordcategorie 'niet van toepassing' op missing gezet is. Het aantal waarnemingen bij ligt tussen de 111 en 117.

Vergelijking uitkomsten huidige onderzoek en literatuur

Hiervoor zijn de uitkomsten van *big data analytics capabilities* en organisatorische prestaties bij Nederlandse overheidsorganisatie en uit het literatuuronderzoek weergegeven. In deze paragraaf worden de uitkomsten verder vergeleken.

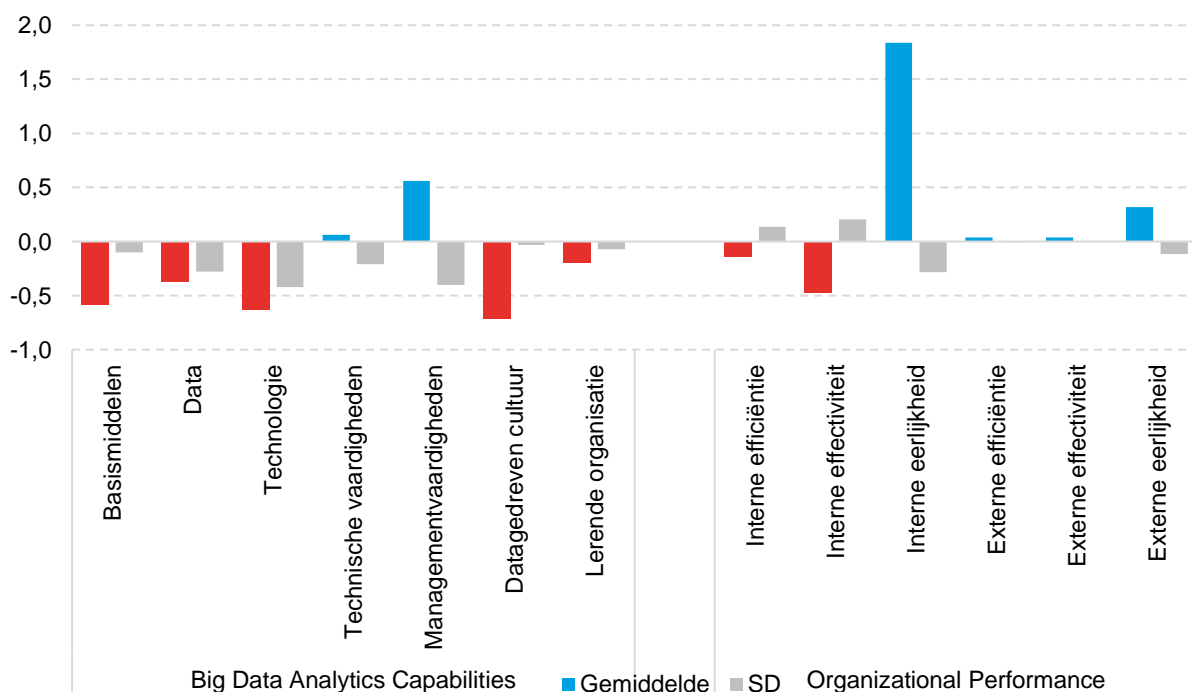
Figuur 4.3 geeft de verschillen in gemiddelde scores en standaarddeviatie voor *big data analytics capabilities* en organisatorische prestaties weer tussen het huidige onderzoek, bij Nederlandse overheidsorganisaties, en de resultaten uit de literatuur.

Over het algemeen is te zien dat de gemiddelde score voor *big data analytics capabilities* bij Nederlandse overheidsorganisaties, op management- en technische vaardigheden na, lager is dan het gemiddelde uit de literatuur. Het gaat gemiddeld om 0,5 punt. De gemiddelde score op managementvaardigheden is daarentegen 0,56 punt hoger. Dit kan mogelijk wijzen op het belang van management(- en technische) vaardigheden binnen een organisatie. Een andere verklaring kan zijn dat een kwart van de respondenten bestaat uit managers of leidinggevenden (24%), de respondenten voornamelijk bij grote (100 medewerkers of meer) organisaties werken en hun eigen (team)kwaliteiten hoger inschatten dan daadwerkelijk het geval is.

Kijkend naar de organisatorische prestaties (*organizational performance*) dan geeft figuur 4.3 aan dat de respondenten bij Nederlandse overheidsorganisaties, op interne eerlijkheid na, ongeveer even hoog scoren. Opvallend is het grote verschil bij interne eerlijkheid, welke 1,84 punt hoger ligt in vergelijking met de literatuur.

Over het algemeen lijken de scores voor organisatorische prestaties meer geclusterd te zijn dan die voor *big data analytics capabilities*. Dit suggereert dat er minder variabiliteit is in de perceptie van respondenten over de prestaties van hun organisatie dan over hun *big data analytics capabilities*.

Figuur 4.3 Verschil over *big data analytics capabilities* en organisatorische prestaties tussen Nederlandse overheidsorganisaties en literatuuronderzoek (gemiddelde en standaarddeviatie categorieën (o.b.v. stellingen), schaal op 1-7)



Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021; Mikalef et al., 2020; Kim, 2004.

Bij deze vergelijking is het belangrijk om rekening te houden met de verschillende contexten waarin de vragenlijsten zijn uitgezet. Zo zullen er andere meetinstrumenten (vragenlijsttools) gebruikt zijn. Daarnaast zijn de vragenlijsten in een andere periode uitgezet (huidig onderzoek in 2021, vanuit de literatuur voor BDAC in 2017 en organisatorische prestaties in 2001). Mogelijk dat het verschil van respectievelijk 4 en 20 jaar invloed heeft op de resultaten. Ook kan de land- en/of organisatiecontext van invloed zijn op de resultaten. Het onderzoek vanuit de literatuur voor *big data analytics capabilities* heeft plaatsgevonden onder de 500 grootste organisatorisch in Noorwegen. Hier is een digitale vragenlijst naartoe gestuurd, waaruit 202 bruikbare vragenlijsten terug zijn gekomen. Het onderzoek vanuit de literatuur voor organisatorische prestaties heeft plaatsgevonden onder 2.000 fulltime overheidsmedewerkers in Korea. Hiervan hebben 1.739 overheidsmedewerkers de vragenlijst ingevuld. Het kan zijn dat bijvoorbeeld in Korea andere waarden, normen en (organisatie)culturen zijn.

4.2. Betrouwbaarheid en validiteit (measurement model)

Voordat de complexe oorzaak-gevolgrelaties via padmodellering met latente variabelen inzichtelijk gemaakt worden, moet naar de interne consistentie gekeken worden. Verschillende beoordelingscriteria worden in de literatuur (Hair et al., 2019) beschreven voor de evaluatie van reflectieve (scores op een item) en formatieve constructen (gewicht toekennen aan construct). Omdat de hypothesen over BDAC en de interne en externe prestaties gaan, zijn er drie te meten variabelen.

Voor reflectieve latente constructies worden de betrouwbaarheid, convergente validiteit en discriminantvaliditeit beoordeeld (Hair et al., 2019). Voor de interne consistentiebetrouwbaarheid worden de *Cronbach's Alpha* (betrouwbaarheid) en *Composite Reliability* (convergente validiteit) gebruikt. Deze meten de mate of een aantal vragen hetzelfde meet. Deze waarden liggen tussen de 0,00 en 1,00. Hierbij is bij een 0,00 totaal geen samenhang tussen de vragen en bij 1,00 betekent dit dat de vragen elkaar volledig overlappen (Baarda en Van Dijkum, 2019).

In tabel 4.1 zijn de uitkomsten van de samenvoegingen tot categorieën weergegeven. Bij zowel de *Chronbach's Alpha* als *Composite Reliability* is de drempel van 0,7 gehaald. Dit suggereert dat de vragen voldoende samenhang hebben. De Average Variance Extracted (AVE) moet voor voldoende variatie boven de 0,5 liggen. In dat geval is deze acceptabel. Dit wordt niet bij alle categorieën gehaald, ondanks dat dit wel in de buurt van 0,5 ligt.

Tabel 4.1 Resultaten betrouwbaarheid en validiteit t.b.v. categorisatie stellingen *big data analytics capabilities*

	1	2	3
1 Big Data Analytics Capabilities	0,670		
2 Interne prestaties	0,473	0,734	
3 Externe prestaties	0,693	0,913	0,704
Cronbach's Alpha	0,931	0,832	0,798
Composite reliability (rho_a)	0,946	0,857	0,857
Composite reliability (rho_c)	0,938	0,875	0,848
Average Variance Extracted (AVE)	0,449	0,538	0,496

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: SmartPLS4.

Als laatste wordt de discriminantvaliditeit beoordeeld. Deze beoordeling gebeurt op drie manieren. Eerst wordt het Fornell-Larcker-criterium beoordeeld. De Fornell-Larcker stelt dat correlaties van het construct met andere constructen lager moeten zijn dan de vierkantswortel van de gemiddelde variantie die uit het construct is geëxtraheerd (Fornell & Larcker, 1981). De Fornell-Larcker-waarden worden weergegeven in de off-diagonaal van tabel 4.1 en de onderstaande correlaties. Dit zijn de schuingedrukte waarden. Alle waarden voldoen aan dit criterium dat discriminante validiteit suggereert.

Ten tweede moeten de ladingen van de reflectieve indicatoren op hun latente constructen aannemelijk groter zijn dan hun kruislingse ladingen op itemniveau. In bijlage 4.5 is te zien dat dit, op de twee items voor interne eerlijkheid na, het geval is.

Als derde en daarmee laatste manier om discriminantvaliditeit te beoordelen, wordt naar de HTMT-ratio gekeken (*heterotrait-monotrait*). Om duidelijk onderscheid te maken tussen twee factoren, moet de HTMT kleiner zijn dan 0,90 (Henseler et al., 2016). De verkregen correlaties in tabel 4.1 (vetgedrukte waarden) voldoen niet allemaal aan dit criterium. Dit geldt namelijk niet voor de interne prestaties in relatie tot de externe prestaties (0,913). Wel tonen deze uitkomsten de discriminantvaliditeit op interne prestaties of externe prestaties aan.

Ondanks de mogelijke effecten om te discrimineren is besloten deze items te behouden. Als eerste toont het Fornell-Larcker-criterium de discriminantvaliditeit aan. Daarnaast zijn alle vragen vanuit eerdere literatuur en op basis van theoretische onderbouwingen opgenomen. Het wordt daarom niet nodig geacht om items te verwijderen. De vragen voor *big data analytics capabilities* suggereren voldoende valide te zijn om als indicator af te zetten tegenover de interne en externe prestaties.

Validatiemaatregelen

Om multicollineariteitsproblemen te onderzoeken, worden de waarden voor de *Variance Inflation Factor* (VIF) geëvalueerd. De VIF wordt gebruikt om de collineariteit van de indicatoren te evalueren. Dit meet of verklarende variabelen gecorreleerd zijn (en op één lijn liggen). De VIF-waarden moeten idealiter lager zijn dan 3 (Hair et al., 2019). Hogere waarden duiden op problemen tussen de variabelen (indicatoren). In tabel 4.2 zijn de VIF-waarden van de *big data analytics capabilities* weergegeven in relatie tot de interne en externe prestaties. Te zien is dat bij *big data analytics capabilities* enkele VIF-waarden boven de 3 liggen. Dit suggereert dat deze uitkomsten met enige voorzichtigheid geïnterpreteerd moeten worden. De VIF-waarden voor interne en externe prestaties liggen allemaal onder de 3 en suggereren dat collineariteit geen belangrijk probleem is in dit onderzoek.

Tabel 4.2 Resultaten validatiemaatregelen voor formatieve constructen t.b.v. categorische stellingen *big data analytics capabilities*

Construct	Measures	VIF	Outer weights	Loadings
BDAC	basismiddelen1	3,317	0,053	0,651
	basismiddelen2	3,319	0,061	0,663
	data1	1,624	0,040	0,418
	data2	2,166	0,061	0,614
	data3	3,085	0,064	0,753
	technologie1	2,548	0,030	0,437
	technologie2	2,195	0,054	0,608
	technologie3	2,768	0,043	0,561
	technische-vaardigheden1	2,440	0,106	0,740
	technische-vaardigheden2	3,053	0,085	0,780
	managementvaardigheden1	1,924	0,088	0,639
	managementvaardigheden2	2,181	0,090	0,711
	managementvaardigheden3	3,852	0,106	0,709
	managementvaardigheden4	4,121	0,109	0,782
	datagedreven-cultuur1	3,217	0,080	0,715
	datagedreven-cultuur2	2,130	0,057	0,553
	datagedreven-cultuur3	2,821	0,092	0,725
	lerende-organisatie1	2,773	0,121	0,752
	lerende-organisatie2	3,178	0,106	0,759
Interne prestaties	interne-eerlijkheid1	2,429	0,101	0,466
	interne-eerlijkheid2	2,402	0,126	0,468
	interne-effectiviteit1	2,088	0,285	0,814
	interne-effectiviteit2	1,690	0,302	0,783
	interne-efficiëntie1	2,626	0,321	0,867
	interne-efficiëntie2	1,777	0,206	0,716
Externe prestaties	externe-eerlijkheid1	1,728	0,136	0,698
	externe-eerlijkheid2	1,767	0,240	0,757
	externe-effectiviteit1	1,841	0,208	0,757
	externe-effectiviteit2	1,734	0,329	0,798
	externe-efficiëntie1	1,564	0,181	0,673
	externe-efficiëntie2	1,632	0,256	0,713

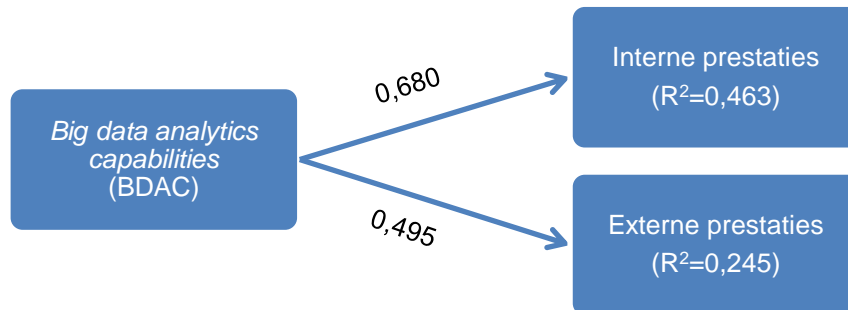
Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: PLS-SEM.

4.3. Toetsing hypotheses (structural model)

Via de PLS-SEM-analyse zijn de verklarende en voorspellende statistiek onderzocht. De padcoëfficiënten (*path coefficients*) zijn via de bootstrapanalyse verkregen. Daarnaast is de significantie meegenomen en weergegeven in figuur 4.4.

Figuur 4.4 Geschatte relaties BDAC op interne en externe prestaties (structural model)



Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: PLS-SEM.

Bij de bootstrapping wordt gekeken naar de verklarende kracht (R^2). R^2 meet de variantie die wordt verklaard in elk van de endogene constructen (in figuur 4.4 zijn dit de interne en externe prestaties). Dit zijn variabelen waarvan de maat wordt bepaald door het model. Als er gekeken wordt naar de bootstrapping in deze PLS-SEM-analyse, dan verklaart het structurele model een variantie van 46,3 procent voor de interne prestaties ($R^2=0,463$). Daarnaast verklaart het model 24,5 procent van de variantie voor externe prestaties ($R^2=0,245$). De determinatiecoëfficiënt voor interne prestaties vertegenwoordigt een gematigde verklarende kracht door *big data analytics capabilities* (Hair et al., 2019). De externe prestaties daarentegen vertegenwoordigen een zwakke verklarende kracht.

Naast de hierboven beschreven bootstrapping door de verklarende krachten (R^2) te meten, wordt ook de voorspellende kracht van het model beoordeeld. Dit wordt gedaan door het PLSpredict algoritme uit te voeren (Shmueli et al., 2016). Tabel 4.3 laat de voorspellingsfoutwaarden van de PLS-SEM zien. Deze zijn weergegeven in de *Root Mean Squared Error* (RMSE) en *Mean Absolute Error* (MAE). De tabel laat zien dat alle waarden (RMSE en MAE) vanuit de PLS-SEM lager zijn in vergelijking met de waarden van een lineair regressiemodel (LM). De waarden liggen wel bij elkaar in de buurt en wijken niet veel af. Deze uitkomsten geven aan dat het model een gemiddeld tot hoge voorspellende waarde heeft op een trainingssteekproef (*out-of-sample power*). Het model kan daarmee ingezet worden om nieuwe gevallen te voorspellen (Shmueli et al., 2016).

Tabel 4.3 Resultaten PLSpredict van manifeste variabelen* (organisatorische prestaties) in relatie tot *big data analytics capabilities*

Variabele	PLS-SEM		LM		PLS<LM	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Interne eerlijkheid1	1,262	0,941	1,365	1,027	Ja	Ja
Interne eerlijkheid2	1,056	0,808	1,147	0,890	Ja	Ja
Interne effectiviteit1	1,281	1,043	1,398	1,135	Ja	Ja
Interne effectiviteit2	1,148	0,872	1,212	0,906	Ja	Ja
Interne efficiëntie1	1,217	0,966	1,370	1,041	Ja	Ja
Interne efficiëntie2	1,292	0,977	1,439	1,083	Ja	Ja
Externe eerlijkheid1	1,155	0,828	1,251	0,933	Ja	Ja
Externe eerlijkheid2	1,007	0,788	1,052	0,866	Ja	Ja
Externe effectiviteit1	1,314	0,986	1,510	1,145	Ja	Ja
Externe effectiviteit2	1,027	0,757	1,059	0,806	Ja	Ja
Externe efficiëntie1	1,214	0,976	1,351	1,083	Ja	Ja
Externe efficiëntie2	1,240	0,977	1,344	1,078	Ja	Ja

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: PLSpredict.

* Manifeste variabelen zijn variabelen die direct kunnen worden gemeenten of waargenomen (Kenton, 2022).

Als laatste worden de hypothesen getoetst. De padcoëfficiënten (*path coefficients*) en de significantie van schattingen (t-statistieken) zijn via de bootstrapanalyse verkregen. De padcoëfficiënt toont de relatie tussen latente variabelen in het PLS-SEM model (het indirecte effect). De sterkte en richting van relaties worden hiermee aangegeven. Een positieve coëfficiënt geeft aan dat er een positieve relatie is. Hoe groter de coëfficiënt, hoe sterker de relatie (Hair et al., 2019). De bootstrapping is uitgevoerd met 5.000 *subsamples*.

De resultaten laten een significante invloed zien van de *big data analytics capabilities* op interne prestaties ($\beta=0,680$, $t=12,714$, $p<0,001$). Hiermee wordt hypothese 1 bekrachtigd. Daarnaast heeft *big data analytics capabilities* ook een aanzienlijk positieve invloed op de externe prestaties ($\beta=0,495$, $t=9,053$, $p<0,001$). Dit aandeel is wel lager ten opzichte van de interne prestaties, maar hiermee kan hypothese 2 ook bekrachtigd worden.

Belangrijke toevoeging is dat de correlaties geen oorzakelijk verband impliceren. De gevonden correlaties (hoe laag of hoog deze ook zijn) betekenen niet dat de *big data analytics capabilities* de oorzaak van de verandering zijn.

4.4. Post-hoc analyse

Naast de analyses om de hypothesen te toetsen, wordt in deze paragraaf de uitkomst van een post-hoc analyse weergegeven. Dit is een statistische analyse die wordt gespecificeerd nadat het onderzoek is afgerond en de gegevens zijn verzameld. In deze post-hoc analyse wordt door middel van eenzelfde PLS-SEM-analyse (inclusief bootstrapping en PLSpredict) gekeken welke (categorie(ën) van) *big data analytics capabilities* het meeste bijdragen aan interne en/of externe prestaties.

De betrouwbaarheid en validiteit zijn vanwege de additionele toevoegen van deze analyse, de overvloed aan cijfers en lengte van de tabellen in bijlage 4.5 (vanaf tabel 4.5b en figuur b4.5b) weergegeven (uitkomsten PLS-SEM en PLSpredict). In deze paragraaf worden alleen de uitkomsten van de verklarende statistiek weergegeven. Deze zijn via de bootstrapanalyse verkregen en worden in padcoëfficiënten inzichtelijk gemaakt. Deze coëfficiënten tonen de sterkte van een relatie en liggen doorgaans tussen de 1 en -1. De 1 staat voor een positieve relatie en een -1 voor een negatieve relatie (Hair et al., 2019).

Als eerste is gekeken naar de subcategorieën die *big data analytics capabilities* vormen in relatie tot interne organisatorische prestaties. De uitkomsten van deze bootstrapping zijn in tabel 4.4 te vinden. De belangrijkste uitkomsten zijn de significante ($p < 0,05$) positieve relaties tussen managementvaardigheden (0,278) en interne prestaties en tussen de lerende organisatie (0,344) en interne prestaties. De invloeden van deze relaties zijn matig, maar wel significant.

Tabel 4.4 Resultaten bootstrapanalyse *big data analytics capabilities* in relatie tot interne organisatorische prestaties

Categorie	Padcoëfficiënt (B)	T-toets (t)	P-waarden (p)
Basismiddelen	-0,034	0,346	0,729
Data	-0,108	0,926	0,354
Technologie	-0,023	0,230	0,818
Technische vaardigheden	0,194	1,587	0,113
Managementvaardigheden	0,278	2,055	0,040
Datagredreven cultuur	0,131	0,893	0,372
Lerende organisatie	0,344	2,228	0,026

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: bootstrapping.

Als tweede en laatste analyse is naar de relatie tussen de subcategorieën die *big data analytics capabilities* vormen en de externe organisatorische prestaties gekeken. De uitkomsten van deze bootstrapping zijn in tabel 4.5 (op de volgende pagina) te vinden. Hieruit is ook af te lezen dat de belangrijkste relaties liggen tussen de managementvaardigheden ($\beta = 0,529$, $t = 4,007$, $p < 0,001$) en de lerende organisatie ($\beta = 0,335$, $t = 2,740$, $p < 0,05$). Vooral de managementvaardigheden hebben een aanzienlijke positieve invloed op de externe prestaties.

Tabel 4.5 Resultaten bootstrapanalyse *big data analytics capabilities* in relatie tot externe organisatorische prestaties

Categorie	Padcoëfficiënt (B)	T-toets (t)	P-waarden (p)
Basismiddelen	-0,105	0,928	0,354
Data	-0,200	1,468	0,142
Technologie	-0,058	0,566	0,572
Technische vaardigheden	0,072	0,567	0,571
Managementvaardigheden	0,529	4,007	0,000
Datagredreven cultuur	-0,068	0,474	0,635
Lerende organisatie	0,335	2,740	0,006

Bron: Vragenlijst *big data analytics*, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021.

Analyse: bootstrapping.

5. Discussie en aanbevelingen

In dit hoofdstuk wordt gekeken naar de afbakening van het onderzoek. Als eerste wordt aangegeven hoe de bevindingen – die tijdens het onderzoek naar voren zijn gekomen en hierna nog gepresenteerd worden – zich tot de literatuur in het literatuuronderzoek (hoofdstuk 2) verhouden. Daarna wordt ingegaan op de beperkingen die tijdens het onderzoek zijn geconstateerd. Vervolgens worden op basis van deze beperkingen en uitkomsten suggesties voor vervolgonderzoek gegeven. Als laatste worden aanbevelingen weergegeven.

5.1. Bevindingen

Dit onderzoek is gestart met het schetsen van de achtergrond, waarna een literatuuronderzoek heeft plaatsgevonden. Vanuit dit literatuuronderzoek is het huidige onderzoek opgesteld, zijn er analyses uitgevoerd en resultaten gepresenteerd. Maar, hoe verhouden de bevindingen van dit onderzoek zich tot de literatuur, zoals gepresteerd in het literatuuronderzoek? In deze paragraaf wordt hier kort en bondig op teruggeblikt.

Uit de literatuur blijkt dat verschillende onderzoeken en onderzoekers (in verschillende hoedanigheden) modellen hebben ontwikkeld waarin factoren naar voren komen die vallen onder *big data analytics capabilities*. Zo heeft Akter et al. (2016) een model ontwikkeld waarin zij tot *big data analytics capabilities* zijn gekomen die bijdragen aan organisatorische prestaties in relatie tot de bedrijfsstrategie. In 2020 hebben Mikalef et al. verschillende *big data analytics capabilities* in een model samengevoegd om competitieve competenties in profit-organisaties te behalen en onderzoeken. Een jaar later (in 2021) heeft Mikalef samen met Gupta een nieuw model ontwikkeld voor *artificial intelligence capability*. Dit is een andere insteek, maar biedt wel dezelfde categorische uitsplitsing als Mikalef et al. in 2020 in hun model hebben ontwikkeld. Omdat bij de afbakening van dit onderzoek niet naar *artificial intelligence (AI)* en bedrijfsstrategieën gekeken is, zijn die onderzoeken verder niet relevant om met de bevindingen te vergelijken. De vergelijking is met name gericht op het model dat Mikalef et al. (2020) ontwikkeld heeft.

Naast de *big data analytics capabilities* heeft dit onderzoek zich ook tot de organisatorische prestaties gericht. In de voorgaande modellen is hier soms – in andere sectoren en met winstoogmerk – op ingegaan. Voor dit onderzoek is gebruik gemaakt van de vertalingen van de stellingen uit het model van Kim (2004). Hij heeft in Korea onderzoek gedaan en een model ontwikkeld dat gaat over organisatorische prestaties bij overheidsorganisaties. De vragen (stellingen) over organisatorische prestaties zijn vertaald en overgenomen. In dit onderzoek komen de organisatorische prestaties naar voren als onafhankelijke variabele waarop de *big data analytics capabilities* een positief effect hebben. Hierdoor kan gezegd worden dat het model uit de literatuur bruikbaar is en nu in een nieuw land is ingezet.

Wanneer er teruggekeken wordt naar de relatie tussen de *big data analytics capabilities* dan zien wij dat deze in zijn geheel bijdraagt aan de organisatorische prestaties, net als in het model van Mikalef et al. (2020) het geval is. Maar, als er verder ingezoomd wordt naar de subcategorieën, dan komt uit dit onderzoek alleen naar voren dat er een aanzienlijke positieve relatie is tussen de managementvaardigheden en lerende organisatie. Deze stonden ook in het model van Mikalef et al. (2020). Alleen, dragen de andere categorieën minder bij dan dat vanuit de literatuur verwacht zou mogen worden. Mikalef et al. (2020) hebben bij hun hypotheses vooral gekeken naar marketing *capabilities* en technische *capabilities* om tot de *competitive performance* te komen. Hierbij kijken ze veelal naar de *dynamic capabilities* zoals kansen, bedreigingen en groeimogelijkheden. Hierin is te zien dat deze uitkomsten gericht zijn op de profit sector en andere doelstellingen kent dan de Nederlandse overheidsorganisaties (zonder winstoogmerk). De bootstrappingwaarden zijn niet goed te vergelijken, omdat deze in het model van Mikalef et al. (2020) niet direct van de *big data analytics capabilities* op de organisatorische prestaties gemeten zijn. De verklarende kracht (R^2) op de *competitive performance* is bij Mikalef et al. 0,347. In het huidige onderzoek is dit voor interne prestaties 0,463 en voor externe prestaties 0,245. Daarmee ligt de waarde uit het literatuuronderzoek tussen de waarden van het huidige onderzoek in.

5.2. Beperkingen

Tijdens het uitvoeren van dit onderzoek zijn verschillende beperkingen naar voren gekomen. Hier wordt in deze paragraaf op teruggeblikt en verantwoording over afgelegd, hoe dit heeft kunnen gebeuren en wat in het vervolg beter zou kunnen.

Als eerste is de planning een beperking van het onderzoek geweest. Doordat de vragenlijst samen met twee medestudenten is opgesteld, was het lastig de planning hierop af te stemmen. Je kan immers zo snel gaan als de traagste persoon. Helaas heeft niet iedereen even adequaat zijn of haar actiepunten opgepakt en bleek het ontwikkelen van een onderzoeksvraag per student en de vragenlijst meer tijd in beslag te nemen. Hierdoor kon de vragenlijst pas in de zomervakantie van 2021 uitgezet worden. Uiteindelijk is gekozen om de vragenlijst toch in de zomervakantie (met vakantiebeperkende maatregelen door Corona) uit te zetten, zodat er geen verdere vertraging in de dataverzameling op zou lopen. Mogelijk dat het uitzetten buiten een vakantie om voor een hogere respons had gezorgd. Daarnaast hebben persoonlijke situaties en wijzigingen (zowel familie-, woon- als werkomstandigheden) voor extra vertraging gezorgd.

Een volgende beperking heeft betrekking op het uitzetten van de vragenlijst om een zo hoog mogelijke respons te behalen. Het uitzetten van de vragenlijst is door mijn medestudenten nauwelijks gedaan. Terwijl twee derde van de vragen specifiek voor hun onderzoek waren. Hierdoor is de vragenlijst extra lang en ingewikkeld geworden. De stellingen over digitaal leiderschap waren bijvoorbeeld lang en ingewikkeld. Uit de data blijkt dat veel respondenten hier afgehaakt zijn. Dit is ook terug te zien in de resultaten van digitaal leiderschap (uitgesloten van deze thesis), waar meer respondenten 'niet van toepassing' als antwoordoptie hebben gegeven, ten opzichte van de stellingen over '*big data analytics*' en 'prestaties'. Met meer inzet van mijn medestudenten had de respons hoger kunnen liggen. Of

omgekeerd, zonder mijn medestudenten was de vragenlijst korter en minder ingewikkeld geweest wat ten goede had kunnen komen aan de respons.

De voorlaatste beperking sluit aan op het vorige punt, waarin aan is gegeven dat de vragenlijst voor een deel uit (te) ingewikkelde stellingen bestond. De vragenlijst bleek achteraf toch vrij technisch van aard te zijn. Voor een gemiddelde overheidsmedewerker was dit lastig in te vullen. Van sommige begrippen hebben overheidsmedewerkers nog nooit gehoord en de stellingen waren soms erg lang. In het vervolg moet hier meer rekening mee gehouden worden. Zowel bij het opstellen van de vragenlijst als het uitzetten hiervan, onder welke doelgroep deze uitgezet wordt. Bij een kortere en makkelijkere vragenlijst was de respons waarschijnlijk hoger geweest. Dat is ook terug te zien aan het afhaakpercentage (69%).

Een laatste beperking heeft betrekking op de vragen. Deze zijn vanuit het Engels vertaald naar het Nederlands. Ook zijn de vragen (stellingen) aangepast zodat deze vanuit een teamperspectief ingevuld kunnen worden. Deze vertaling en aanpassingen kunnen invloed hebben op de uitkomsten en/of juiste vraagstelling.

Afgezien van deze beperkingen is het goed te zien dat er voldoende respons binnen is gekomen om te analyseren. Niet om uitspraken over een representatieve steekproef te kunnen doen, maar wel om eerste resultaten te kunnen weergeven. Met deze beperkingen in het achterhoofd kan hier in het vervolg of bij vervolgonderzoek rekening mee gehouden worden. Zo is het verstandig om vooraf de verwachtingen, planning en afspraken helder te hebben. Op die manier is het voor iedereen duidelijk wat er verwacht wordt en wanneer en wie waar verantwoordelijk voor is. Dit komt het gehele onderzoeksproces ten goede. Ook is het belangrijk om goed naar de vragenlijst te kijken, inhoudelijk (moeilijkheid en lengte) en praktisch (onder wie deze uitgezet wordt en op welke wijze de respondenten gevonden worden). Dit kan bij een vervolgonderzoek responsverhogend werken.

5.3. Vervolgonderzoek

Dit onderzoek heeft nieuwe inzichten opgeleverd over *big data analytics* in relatie tot organisatorische prestaties bij Nederlandse overheidsorganisaties, maar ook voor vervolgonderzoeken. Hierna zijn een aantal suggesties voor en om rekening te houden bij vervolgonderzoeken weergegeven.

Onderzoekspopulatie

Keuze

De vragenlijst is breed uitgezet bij overheidsorganisaties in Nederland. Doordat de vragenlijst vrij technisch was, is het aan te raden om te kijken of het nodig is om vervolgonderzoek uit te voeren en daarbij een specifiekere groep medewerkers (bijvoorbeeld IT-medewerkers of leidinggevendenden) te bevragen en/of een kleinere doelgroep te pakken (bijvoorbeeld alleen gemeenten). Ook is niet gevraagd naar de wijze waarop (teams van) overheidsorganisaties technische vragen oplossen. Wordt dit (altijd) intern opgevangen of worden hier externen voor ingehuurd (outsourcing)? Dit kan van invloed zijn op de analysevaardigheden van de medewerkers.

Vergroten

Voor vervolgonderzoek is het aan te raden om te zorgen voor een grotere onderzoekspopulatie. In dit huidige onderzoek is het aantal bruikbare respondenten te laag om voor de totale onderzoekspopulatie uitspraken te doen. Het vergroten van de onderzoekspopulatie kan mogelijk door een vereenvoudigde en kortere vragenlijst op te stellen die via interne kanalen bij overheidsorganisaties uitgezet kunnen worden.

Valideren

Vanwege de grote verscheidenheid in de onderzoekspopulatie is het goed om deze bij vervolgonderzoek te valideren. Te denken valt aan een normale verdeling naar achtergrondvariabelen, zoals leeftijd, geslacht en kennis van en ervaring met informatiesystemen en *big data analytics*.

Uitkomsten valideren

Een suggestie voor vervolgonderzoek is om de uitkomsten te valideren. Dit kan door een vergelijking met andere landen te maken (zie volgende punt, internationale vergelijking) of door de uitkomsten in de praktijk te testen. Door middel van interviews kunnen de uitkomsten gevalideerd worden en kunnen uitkomsten verduidelijkt en onderbouwd worden.

Een ander belangrijk aspect is de invloed van corona. Tijdens de dataverzameling was het coronavirus actief, waardoor veel overheidsmedewerkers thuis werkten. Dit kan van invloed zijn op de dataverzameling en uitkomsten. Ook na het coronatijdperk kan het zijn dat thuiswerken invloed heeft op de vaardigheden van medewerkers, doordat organisaties veranderen en vaker thuiswerken zal blijven (Rijksoverheid, 2021). De gevalideerde stellingen zijn afkomstig van de periode voor corona. Dit kan invloed hebben gehad op de vergelijking. Het is goed om hier vervolgonderzoek op uit te voeren of hier verschillen in zitten en de coronaperiode van invloed is op de antwoorden van de respondenten.

Internationale vergelijking

De uitkomsten van dit onderzoek hebben betrekking op Nederlandse overheidsorganisaties. Deze uitkomsten zijn vergeleken met internationale uitkomsten (uit eerder onderzoek), terwijl er in dit onderzoek onvoldoende internationale respondenten zijn voor een recente vergelijking. Hierdoor is de validatie ten opzichte van de internationale vragenlijst lastiger. Voor vervolgonderzoek is aan te raden om de vragenlijst in meer landen uit te zetten zodat een betere validatie mogelijk is.

5.4. Aanbevelingen

Het doel van dit onderzoek was om voorwaarden en eigenschappen in kaart te brengen die betrekking hebben op *big data analytics* die bij kunnen dragen aan de organisatorische prestaties van Nederlandse overheidsorganisaties. Vanuit de literatuur zijn *capabilities* opgehaald waar organisaties mee aan de slag kunnen. Daarnaast zijn deze *capabilities* onderzocht bij Nederlandse overheidsorganisaties om in de huidige context te valideren.

Bij veel (Nederlandse overheids)organisaties is veel data aanwezig, in toenemende mate. Het blijft de vraag of organisaties en medewerkers dit kunnen gebruiken om organisatorische prestaties te behalen. Mogelijk is dit geen doel van de organisaties of is het huidige datavolwassenheidsniveau te laag.

Uit dit onderzoek blijkt dat *big data analytics capabilities* van significante invloed zijn op interne en externe organisatorische prestaties. Voor Nederlandse overheidsorganisaties kunnen deze uitkomsten gebruikt worden om de bewustwording rondom (*big*) *data analytics* te vergroten. Doordat het speelveld van *big data analytics capabilities* een positieve invloed op de organisatorische prestaties heeft, kan het lonen om hier aandacht aan te besteden.

Daarnaast dragen – vanuit de *big data analytics capabilities* – de managementvaardigheden en de lerende organisatie bij aan de interne en externe organisatorische prestaties. Deze twee categorieën dragen met hun positieve invloed het meeste bij. Nederlandse overheidsorganisaties kunnen hier meer op inzetten, door bewuster bij de managementvaardigheden van medewerkers (o.a. managers) stil te staan, deze verder te (laten) ontwikkelen en aandacht aan te besteden. Datzelfde geldt voor de lerende organisatie, waarbij het belangrijk is dat deze gedachtegang er is en mag zijn. De organisatie en haar medewerkers mogen (en moeten misschien wel) voortdurend in ontwikkeling zijn en (verder) leren. Deze gedachtes dragen dan mogelijk positief bij aan de organisatorische prestaties. Mogelijk omdat deze vaardigheden en uitgangspunt niet hoeven te betekenen dat dit de oorzaak van de verandering is.

De uitkomsten zijn mogelijk ook relevant voor andere organisaties dan Nederlandse overheidsorganisaties. Te denken valt aan semi-overheidsorganisatie of profit-organisaties. Ondanks dat de organisatiedoelen hier anders zijn, is hier misschien nog wel meer als doel om organisatorische prestaties te behalen. Daaronder valt in grotere zin het behalen van financiële winst.

Bij de aanbevelingen is het goed om de beperkingen en aanbevelingen voor vervolgonderzoek in het achterhoofd te houden. Zonder deze kanttekeningen, kunnen de uitkomsten verkeerd geïnterpreteerd worden. Zo is het altijd noodzakelijk om per situatie (organisatie) te kijken naar de huidige stand van zaken, wensen, behoeften en mogelijkheden. Met de juiste inzet kunnen organisatorische prestaties gerealiseerd en/of verbeterd worden.

6. Conclusie

De achterliggende gedachte voor dit onderzoek is dat er steeds meer data (of gegevens) beschikbaar zijn binnen Nederlandse overheidsorganisaties. Maar, kunnen zij daar wel mee omgaan en de (toenemende) data en de verschillende vormen hiervan verwerken? Verwerken door analyses hierop uit te voeren en te vertalen in de praktijk zodat de organisatie hier voordelen uit kan behalen?

In deze thesis is onderzoek gedaan naar eigenschappen en voorwaarden om *big data analytics* toe te kunnen passen om organisatorische prestaties te behalen. Hiervoor zijn een hoofdvraag en deelvragen geformuleerd. De deelvragen zijn in hoofdstuk 2 met behulp van literatuuronderzoek beantwoord. Met behulp van dit literatuuronderzoek is een kwantitatief onderzoek opgezet door het uitzetten van een vragenlijst. Met de resultaten die uit dit onderzoek komen, kan de hoofdvraag opnieuw bekeken en beantwoord worden. Deze beantwoording vindt in dit hoofdstuk plaats.

6.1. Beantwoording hoofdvraag

Om de hoofdvraag te kunnen beantwoorden, is het goed om vooraf aan te geven dat de resultaten betrekking hebben op 120 respondenten die de vragenlijst in hebben gevuld. In het vorige hoofdstuk (5, discussie en aanbevelingen) is dieper ingegaan op de implicaties en suggesties voor vervolgonderzoek.

Als eerste blijkt uit dit onderzoek dat de gemiddelde score voor *big data analytics capabilities* een 4,1 is (op basis van een 7-punttschaal). De gemiddelde score op organisatorische prestaties is een 5,0. Vergeleken met de literatuur is het gemiddelde voor *big data analytics capabilities* een 4,4 en voor organisatorische prestaties een 4,8. Dit betekent dat de *capabilities* door Nederlandse overheidsmedewerkers lager ingeschat wordt dan de literatuur (4,1 vs. 4,4). Omgekeerd schatten zij de organisatorische prestaties juist hoger in (5,0 vs. 4,8). Bij deze vergelijking is geen rekening met verschillende contexten gehouden, zoals het aantal respondenten, achtergrondkenmerken van de respondenten en het land waarin de vragenlijst is uitgezet. Als er alleen gekeken wordt naar de gemiddelden, dan is er vooral ruimte voor verbetering op het gebied van basismiddelen, technologie en datagedreven cultuur.

De interne consistentie en betrouwbaarheid van dit onderzoek zijn op verschillende manieren onderzocht en gevalideerd. Zo zijn de bij de *Chronbach's Alpha* en *Composite Reliability* de drempel van 0,7 gehaald. Dit suggereert dat de vragen voldoende samenhang hebben. Daarnaast ligt de *Average Variance Extracted* (AVE) rond de 0,5 waardoor er met enige voorzichtigheid gezegd kan worden dat er voldoende variatie in de categorieën zit. Daarnaast zijn extra validaties gedaan om te stellen dat de gebruikte vragenlijst in dit onderzoek een goede (of acceptabele) interne consistentie en validiteit heeft. Dit is verder nog gedaan door het Fornell-Larcker-criterium te beoordelen, de HTMT-ratio (*heterotrait-monotrait*) te beoordelen en de VIF-waarden (*Variance Inflation Factor*) te onderzoeken. Deze waarden duiden in grote lijnen niet op problemen. Daar waar (kleine) afwijkingen gevonden zijn,

is in hoofdstuk 4 toegelicht waarom vragen toch behouden zijn en de antwoorden valide genoeg zijn om te presenteren.

Om tot de beantwoording van de hoofdvraag te komen, wordt deze hieronder nogmaals weergegeven:

Welke big data analytics capabilities dragen bij aan organisatorische prestaties binnen Nederlandse overheidsorganisaties?

In relatie tot de hoofdvraag zijn twee hypotheses opgesteld om te toetsen of *big data analytics capabilities* bijdragen aan organisatorische prestaties. Hiervoor is een PLS-SEM-analyse uitgevoerd en is door een bootstrapping uit te voeren onderzocht of er een verband is. De bootstrapping resulteert in padcoëfficiënten (*path coefficients*). Hieruit blijkt dat de *big data analytics capabilities* zowel op de interne ($\beta=0,680$, $t=12,714$, $p<0,001$) als externe ($\beta=0,495$, $t=9,053$, $p<0,001$) organisatorische prestaties een aanzienlijke positieve invloed hebben. Deze correlaties betekenen niet dat er een oorzakelijk verband is. De gevonden correlaties (hoe laag of hoog deze ook zijn) betekenen niet dat de *big data analytics capabilities* de oorzaak van de verandering zijn.

Daarnaast zijn via een post-hoc analyse extra inzichten verzameld. Hierin is gekeken welke *big data analytics capabilities* voornamelijk bijdragen aan de interne en externe organisatorische prestaties. Hierbij is uitgegaan van subcategorieën. Naar voren komt dat de managementvaardigheden en lerende organisatie een significante positieve invloed hebben op de interne én externe organisatorische prestaties (managementvaardigheden intern $\beta=0,278$, $t=2,055$, $p<0,05$ en extern $\beta=0,529$, $t=4,007$, $p<0,001$; lerende organisatie intern $\beta=0,344$, $t=2,228$, $p<0,05$ en extern $\beta=0,335$, $t=2,740$, $p<0,05$).

Literatuurlijst

Alchemer (2021). *About us*. Geraadpleegd op 30 november 2021, van <https://www.alchemer.com/company/about/>

Akter, S., Fosso Wamba, S., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, 182, 113-131. <https://doi.org/10.1016/j.ijpe.2016.08.018>

Al-Sai, Z. A., & Abualigah, L. M. (2017). Big data and E-government: A review. *8th International Conference on Information Technology (ICIT)*, 580-587. <https://doi.org/10.1109/ICITECH.2017.8080062>

Baarda, B., & Van Dijkum, C. (2019). *Basisboek Statistiek met SPSS. Handleiding voor het verwerken en analyseren van en rapporteren over (onderzoeks)gegevens* (6^e druk). Noordhoff Uitgevers bv.

Barton, D., & Court, D. (2012). Making advanced analytics work for you. *Harvard Business Review*, 90, 78.

Beach, D., & Brun Pedersen, R. (2013). *Process-Tracing Methods: Foundations and Guidelines*. University of Michigan Press.

Bell, E., Bryman, A., & Harley, B. (2019). *Business research methods* (fifth edition). Oxford University Press.

Brewer, G. A., & Selden, S. C. (2000). Why elephants gallop: Assessing and predicting organizational performance in federal agencies. *Journal of Public Administration Research and Theory*, 10(4), 685–711. <https://doi.org/10.1093/oxfordjournals.jpart.a024287>

BZK, Kerncijfers Personeel Overheid en Onderwijs (2019). *Omvang publieke sector*. Geraadpleegd op 13 mei 2021, van <https://kennisopenbaarbestuur.nl/thema/omvang-publieke-sector/>

CBS (Centraal Bureau voor de Statistiek) (2020). *Aantal gemeenten daalt in 2021 verder tot 352*. Geraadpleegd op 12 maart 2021, van <https://www.cbs.nl/nl-nl/nieuws/2020/53/aantal-gemeenten-daalt-in-2021-verder-tot-352>

Chatfield, A., Reddick, C., & Al-Zubaidi, W. (2015). *Capability Challenges in Transforming Government through Open and Big Data: Tales of Two Cities*. ICIS.

DALICO (z.d.). *Partners in the DaLiCo project*. Geraadpleegd op 30 november 2021, van <https://dalico.info/partners/>

Davenport, T. H., & Harris, J. G. (2007). *Competing on Analytics: The New Science of Winning*. Harvard Business School Press.

De Kruik, E., Schoenmakers, F., Dijksman, M., Gremmen, M., & Minderhoud, M. (2015). *Rol en positie van gemeenten in een samenleving van Big en Open Data*. Kwaliteitsinstituut Nederlandse gemeenten (KING)/BigData.Company/BigFellows.

Directiesecretaresse (z.d.). *Profit of non-profit bedrijf?* Geraadpleegd op 12 maart 2021, van <http://directie-secretaresse.weebly.com/profit-of-non-profit-bedrijf.html>

Dremel, C., Herterich, M. M., Wulf, J., & Brocke, J. (2020). Information & Management Actualizing big data analytics affordances: A revelatory case study. *Information & Management*, 57(1), 1–21. <https://doi.org/10.1016/j.im.2018.10.007>

Dubey, R., Gunasekaran, A., & Childe, S.J. (2018). Big data analytics capability in supply chain agility. *Management Decision*, 57(3). <https://doi.org/10.1108/MD-01-2018-0119>

Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic Capabilities: What Are They? *Strategic Management Journal*, 21(10-11), 1105-1121. [https://doi.org/10.1002/1097-0266\(200010/11\)21:10/11<1105::AID-SMJ133>3.0.CO;2-E](https://doi.org/10.1002/1097-0266(200010/11)21:10/11<1105::AID-SMJ133>3.0.CO;2-E)

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research*, 18(1), 39-50. <https://doi.org/10.2307/3151312>

Gandomi, A., & Haider, M. (2014). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>

Gefen, D., Straub, D., & Boudreau, M. (2000). Structural Equation Modeling and Regression: Guidelines for Research Practice. *Communications of the Association for Information Systems*, 4. <https://doi.org/10.17705/1CAIS.00407>

Gemeente Berg en Dal (2023). *Intern overleg t.b.v. de thesis over het datagebruik bij Nederlandse overheidsorganisaties*.

GEMMA Online (2023). *Informatie- en applicatiearchitectuur*. Geraadpleegd op 24 maart 2023, van https://www.gemmaonline.nl/index.php/GEMMA_2_Applicatiearchitectuur

- George, G., Haas, M. R., & Pentland, A. (2014). Big data and management. *Academy of Management Journal*, 57, 321–326.
- Gobble, M.M. (2013). Big data: the next big thing in innovation. *Research-Technology Management*, 56, 64–66.
- Gratton, C., Jones, I., & Robinson, T. (2011). *Onderzoeksmethoden voor Sportstudies* (tweede druk). Routledge.
- Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, 53(8), 1049-1064. <https://doi.org/10.1016/j.im.2016.07.004>
- Hair, J. F., Risher, J. J., Sarstedt, M. & Ringle, C. M. (2019). When to use and how to report the results of PLS-SEM. *European Business Review*, 31(1), 2-24. <https://doi.org/10.1108/EBR-11-2018-0203>
- Henseler, J. (2010). On the convergence of the partial least squares path modeling algorithm. *Computational Statistics*, 25(1), 107-120. <https://doi.org/10.1007/s00180-009-0164-x>
- Henseler, J., Hubona, G., & Ray, P. (2016). Using PLS path modeling in new technology research: updated guidelines. *Industrial Management & Data Systems*, 116(1), 2-20. <https://doi.org/10.1108/IMDS-09-2015-0382>
- Holzer, M., & Kloby, K. (2005). Public performance measurement. An assessment of the state-of-the-art and models for citizen participation. *International Journal of Productivity and Performance Management* 54(7), 517-532, <https://doi.org/10.1108/17410400510622205>
- Humble, J. (2020). *PLSPM: A library implementing Partial Least Squares Path Modeling PyPI version*. Geraadpleegd op 21 maart 2023, van <https://pypi.org/project/plspm/>
- Jelonek, D. (2017). Big Data Analytics in the Management of Business. *MATEC Web of Conferences* 125. <https://doi.org/10.1051/mateconf/201712504021>
- Kenton, W. (2022). *Manifest Variable*. Geraadpleegd op 12 mei 2023, van <https://www.investopedia.com/terms/m/manifest-variable.asp>
- Kim, S. (2004). Individual-Level Factors and Organizational Performance in Government Organizations. *Journal of Public Administration Research and Theory*, 15(2). <https://doi.org/10.1093/jopart/mui013>
- Kiron, D., Prentice, P. K., & Ferguson, R. B. (2014). The Analytics Mandate. *MIT Sloan Management Review*, 55(4), 1-25. <https://sloanreview.mit.edu/projects/analytics-mandate/>

Kromme, J. (2017). *Vergelijking van de meest gebruikte statistische software*. Geraadpleegd op 6 mei 2021, van <https://www.customertalk.nl/artikelen/onderzoek/vergelijking-van-de-meest-gebruikte-statistische-software/>

Kruisman, M. (2021). *Een betrouwbare steekproefgrootte berekenen*. Geraadpleegd op 13 mei 2021, van https://markkruisman.nl/steekproefgrootte_calculator/

Labrinidis, A., & Jagadish, H. V. (2012). Challenges and opportunities with big data. *Proceedings of the VLDB Endowment*, 5, 2032–2033. <https://doi.org/10.14778/2367502.2367572>

Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group Inc.

Malomo, F., & Sena, V. (2017). Data intelligence for local government? assessing the benefits and barriers to use of big data in the public sector. *Policy & Internet*, 9(1), 7–27. <https://doi.org.hu.idm.oclc.org/10.1002/poi3.141>

Marr, B. (2014). *Big data: The 5 vs everyone must know*. Geraadpleegd op 16 april 2021, van <https://www.linkedin.com/pulse/20140306073407-64875646-big-data-the-5-vs-everyone-must-know>

McAfee, A., & Brynjolfsson, E. (2012). Big data: the management revolution. *Harvard Business Review*, 60–66 (68), 128.

McKinsey (2009). *Hal Varian on how the Web challenges managers*. Geraadpleegd op 5 december 2021, van <https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/hal-varian-on-how-the-web-challenges-managers>

McNulty, E. (2014). *Understanding Big Data: The seven V's*. Geraadpleegd op 16 april 2021, van <http://dataconomy.com/seven-vs-big-data/>

Mergel, I., Rethemeyer, R.K., & Isett, K. (2016). Big Data in Public Affairs. *Public Administration Review*, 76(6), 928-937. <https://doi.org/10.1111/puar.12625>

Mikalef, P., & Gupta, M. (2021). Artificial intelligence capability: Conceptualization, measurement calibration, and empirical study on its impact on organizational creativity and firm performance. *Information & Management*, 58(3). <https://doi.org/10.1016/j.im.2021.103434>

Mikalef, P., Krogstie, J., Pappas, I. O., & Pavlou, P. (2020). Exploring the relationship between big data analytics capability and competitive performance: The mediating roles of dynamic and operational capabilities. *Information and Management*, 57(2). <https://doi.org/10.1016/j.im.2019.05.004>

Mulier Instituut (2020). *Intern thesisoverleg en probleemanalyse datagebruik overheidsorganisaties*.

National Institute of Standards and Technology (NIST) (2019). NIST Big Data Interoperability Framework: Volume 1, Definitions. *NIST Special Publication 1500-1*.
<http://dx.doi.org/10.6028/NIST.SP.1500-1>

Nationale Ombudsman (z.d.). *Overheidsinstanties*. Geraadpleegd op 30 november 2021, van <https://www.nationaleombudsman.nl/overheidsinstanties#:~:text=Hieronder%20vallen%20bijna%20alle%20overheidsorganisaties,%2C%20vinden%20wij%20een%20overheidsorganisatie>.

Oost, H., & Markenhof, A. (2002). *Een onderzoek voorbereiden*. HB Uitgevers.

Overheid.nl (z.d.). *Wie vormen de overheid?* Geraadpleegd op 30 november 2021, van <https://www.overheid.nl/wie-vormen-de-overheid>

Ponting, C. P. (2017). Big knowledge from big data in functional genomics. *Emerging topics in life sciences* 1(3), 245-248. <https://doi.org/10.1042/ETLS20170129>

Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.

Rainey, H., Backoff, R., & Levine, C. (1976). Comparing Public and Private Organizations. *Public Administration Review*, 36(2), 233-244. <https://doi.org/10.2307/975145>

Rajaraman, V. (2016). Big Data Analytics. *Resonance : Published by the Indian Academy of Sciences*, 21(8), 695-716. <https://doi.org/10.1007/s12045-016-0376-7>

Rijksoverheid (2021). *Onderzoek wijst uit: thuiswerken is een blijvertje*. Geraadpleegd op 20 juni 2022, van <https://www.rijksoverheid.nl/actueel/nieuws/2021/07/14/onderzoek-wijst-uit-thuiswerken-is-een-blijvertje#:~:text=Veel%20van%20deze%20maatregelen%20zijn,%3A%20%E2%80%9CDit%20is%20goed%20nieuws>

Rijksoverheid (z.d.a). *Zelfstandige bestuursorganen (zbo's)*. Geraadpleegd op 18 februari 2023, van <https://www.rijksoverheid.nl/onderwerpen/rijksoverheid/zelfstandige-bestuursorganen>

Rijksoverheid (z.d.b). *Communicatiebeleid van de Rijksoverheid*. Geraadpleegd op 29 april 2021, van [https://www.rijksoverheid.nl/onderwerpen/overheidscommunicatie/communicatiebeleid-van-de-overheid#:~:text=De%20overheid%20heeft%20een%20informatieplicht,de%20werking%20van%20de%20overheid.&text=Daarbij%20betreft%20de%20overheid%20burgers,Openbaarheid%20van%20Bestuur%20\(Wob\)](https://www.rijksoverheid.nl/onderwerpen/overheidscommunicatie/communicatiebeleid-van-de-overheid#:~:text=De%20overheid%20heeft%20een%20informatieplicht,de%20werking%20van%20de%20overheid.&text=Daarbij%20betreft%20de%20overheid%20burgers,Openbaarheid%20van%20Bestuur%20(Wob))

Ringle, C. M., Wende, S., & Becker, J.-M. (2022). *SmartPLS 4. Oststeinbek: SmartPLS*. Geraadpleegd in 2023, van <https://www.smartpls.com>

Shmueli, G., Ray, S., Estrada, J. M. V., & Chatla, S. B. (2016). The Elephant in the Room: Predictive Performance of PLS Models. *Journal of Business Research*, 69(10), 4552-4564. <https://doi.org/10.1016/j.jbusres.2016.03.049>

Sirkin, R. M. (2006). *Statistics for Social Sciences* (third edition). Sage Publications, Inc.

Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>

Sternkopf, H., & Mueller, R. M. (2018). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences*. <https://doi.org/10.24251/HICSS.2018.630>

Strawn, G. O. (2012). Scientific research: how many paradigms? *Educational Review*, 47, 26.

TechAmerica Foundation (2012). *Demystifying Big Data. A Practical Guide To Transforming The Business of Government*. TechAmerica Foundation.

Teece, D.J. (2014). A dynamic capabilities-based entrepreneurial theory of the multinational enterprise. *Journal of International Business Studies*, 45(1), 8–37. <https://doi.org/10.1057/jibs.2013.54>

Verbeeten, F. (2008). Performance Management Practices in Public Sector Organizations: Impact on Performance. *Accounting Auditing & Accountability Journal*, 21(3), 427-454. <https://doi.org/10.1108/09513570810863996>

Verschuren, P., & Doorewaard, H. (2007). *Het ontwerpen van een onderzoek* (vierde druk). Uitgeverij LEMMA.

VNG (Vereniging van Nederlandse Gemeenten) (2021). *Onze Leden*. Geraadpleegd op 12 maart 2021, van <https://vng.nl/artikelen/onze-leden#:~:text=Gerelateerde%20onderwerpen&text=Alle%20gemeenten%20van%20Nederland%20zijn,2020%20was%20het%20aantal%20355>

VSG (Vereniging Sport en Gemeenten) (2020). *Werken met lokale data in de sport*. Geraadpleegd op 2 oktober 2020, op <https://sportengemeenten.nl/werken-met-lokale-data-in-de-sport/>

VSG (Vereniging Sport en Gemeenten) (z.d.). *Netwerk & leden*. Geraadpleegd op 12 maart 2021, van <https://sportengemeenten.nl/netwerk-leden/>

Ward, J., & Barker, A. (2013). *Undefined By Data: A Survey of Big Data Definitions*. ArXiv, abs/1309.5821.

Wesseling, H., Postma, R.-M., Stolk, R., & Sabirovic, A. (2018). *Datagedreven sturing bij gemeenten Van data tot (gedeelde) informatie voor beter (samen) sturen*. Berenschot Groep B.V.

Yan, H., & Ting, Y. (2018). The Effectiveness of Online Citizen Evaluation of Government Performance: A Study of the Perceptions of Local Bureaucrats in China. *Public Personnel Management*, 47(4), 419-444. <https://doi.org/10.1177/0091026018767475>

Bijlagen

Bijlage 1. Criteria en zoektermen literatuuronderzoek

Voor het literatuuronderzoek zijn wetenschappelijke artikelen opgezocht. In vier verschillende zoekmachines zijn de termen opgezocht, te weten: HUGO, AIS eLibrary, Google Scholar en ScienceDirect. Hierbij is HUGO als belangrijkste zoekmachine gebruikt. Deze zoekmachine is van de Hogeschool Utrecht (HU) en hiermee worden tegelijkertijd meerdere wetenschappelijke bronnen doorzocht. De gebruikte parameters bij het zoeken zijn in tabel b1.1 weergegeven.

Tabel b1.1 Vereiste en ingestelde parameters voor het zoeken naar wetenschappelijke literatuur over *big data analytics capabilities* in 2021

Parameter	Waarde
Auteurs	Alle
Bibliotheken	Alle
Databases	Alle
Geografisch	Wereldwijd
Materiaalsoorten	Alle*
Publicatieperiode	Vanaf 2000**
Taal	Duits, Engels, Nederlands***
Type literatuur	Peer-reviewed (wetenschappelijk)
Vakgebied	Alle

* Voor dit literatuuronderzoek is alleen gebruik gemaakt van schriftelijke publicaties (voornaamste publicatietype in de gebruikte bibliotheken).

** De meeste literatuur is na 2010 gepubliceerd.

*** Duitse en Nederlandse relevante literatuur is niet gevonden of was niet relevant genoeg.

De gebruikte zoektermen zijn direct afgeleid van de onderzoeksvraag en deelvragen. Via deze zoektermen is naar relevante wetenschappelijke literatuur gezocht. In tabel b1.2 zijn de zoektermen te vinden, samen met het aantal gevonden artikelen.

Tabel b1.2 Zoektermen literatuuronderzoek *big data analytics capabilities* en aantal gevonden artikelen in de HUGO-zoekmachine in 2021

Categorie	Aantal gevonden artikelen*	Aantal gevonden artikelen met zoektermen in titel*
<i>Big data</i>	75.100	26.500
<i>Big data analytics</i>	7.100	2.400
<i>Big data capability</i>	2.500 (1.200)	64 (81)
<i>Big data analytics capability</i>	538 (392)	42 (56)
<i>Big data analytics performance</i>	1.900	115
<i>Governmental performance</i>	5.400	67
<i>Organizational performance</i>	39.700	6.400
<i>Competitive performance</i>	58.400	1.900
<i>Governmental performance AND big data analytics capability</i>	9 (4)	0 (0)
<i>Organizational performance AND big data analytics capability</i>	41 (48)	4 (3)
<i>Competitive performance AND big data analytics capability</i>	56 (40)	3 (3)

* Tussen haakjes is het aantal gevonden artikelen wanneer *capability* is aangepast naar *capabilities*.

Bijlage 2. Vragenlijst

Toelichting:

- De vragen met een * zijn verplicht.
- De vragen bij de vragenblokken *big data analytics* en prestaties zijn aangeduid met een ** en hebben betrekking op dit onderzoek.
- N.v.t. is in de vragenlijst uitgeschreven als 'niet van toepassing'. Vanwege ruimtebesparing is dat in deze bijlage ingekort.

Beginpagina

Welkom bij dit onderzoek binnen overheidsorganisaties!

De Hogeschool Utrecht doet onderzoek binnen overheidsorganisaties (bijvoorbeeld gemeenten) naar het werken met data. Om deze inzichten te verkrijgen, zijn wij benieuwd hoe uw team binnen uw organisatie werkt.

De thema's waar de vragen over gaan, zijn *big data analytics* (analyse van grote hoeveelheden data), digitaal leiderschap en datageletterdheid binnen overheidsorganisaties. De vragen gaan over (big) data en het team waarin u werkt, uw directe collega's.

Dit onderzoek maakt deel uit van het internationale DaLiCo-project (Data Literacy in Context) en wordt uitgevoerd in Nederland, Duitsland, Hongarije en Spanje. De inzichten uit deze landen wordt met elkaar vergeleken.

Het invullen van de vragenlijst duurt ongeveer 15 minuten. Uw antwoorden worden strikt vertrouwelijk behandeld door Hogeschool Utrecht en de resultaten worden niet persoonlijk gerapporteerd. Aan het einde van de vragenlijst heeft u de mogelijkheid om de uitkomsten van de onderzoeken te ontvangen.

Heeft u vragen over dit onderzoek? Neem dan contact op met Arjen Davids, via arjen.davids@student.hu.nl

Bedankt voor uw tijd en medewerking!

Guido Ongena (Program Director Master of Informatics and Business IT & Management)

Arjen Davids (student Master of Informatics - Applied Data Science)

Medestudent 1 (student Master of Informatics - Business & IT)

Medestudent 2 (student Master of Informatics - Business & IT)

Achtergrond

In welk land is uw organisatie gevestigd?*

- ☐ Nederland
- ☐ Duitsland
- ☐ Spanje
- ☐ Hongarije
- ☐ Anders, namelijk: ...

Onder welke sector valt uw organisatie?*

- ☐ Commercieel
- ☐ Non-profit overheid
- ☐ Non-profit anders

Hoeveel jaar werkervaring heeft u?*

...

Hoeveel medewerkers heeft uw organisatie?*

- ☐ Minder dan 10
- ☐ 10-50
- ☐ 51-100
- ☐ 101-1.000
- ☐ Meer dan 1.000

Wat is uw geslacht?*

- ☐ Man
- ☐ Vrouw
- ☐ Anders
- ☐ Wil ik niet zeggen

Wat is uw geboortjaar?*

...

Wat is uw hoogst afgeronde opleiding?*

- ☐ Basisschool
- ☐ Middelbare school
- ☐ Middelbaar beroepsonderwijs (mbo)
- ☐ Hoger beroepsonderwijs (hbo)
- ☐ Wetenschappelijk onderwijs (wo)
- ☐ Anders, namelijk: ...
- ☐ Wil ik niet zeggen

Hoeveel jaar ervaring heeft u met (big) data analytics?*

- ☐ Geen, ik heb er geen ervaring mee
- ☐ Minder dan 1 jaar
- ☐ 1 - 2 jaar
- ☐ 2 - 3 jaar
- ☐ 3 - 4 jaar
- ☐ Meer dan 4 jaar

Wat omschrijft uw rol het beste?*

- ☐ Medewerker, niet-leidinggevend
- ☐ Autonoom professional, niet-vallend onder een divisie, afdeling of team
- ☐ Leidinggevende/manager over een afdeling of team
- ☐ Eigenaar/directeur (zelfstandig ondernemer/zzp)
- ☐ CEO, bestuurder/directeur/manager van een bedrijf of organisatie
- ☐ Anders, namelijk: ...

Big data analytics**

Om erachter te komen hoe ver organisaties staan met data-analyses op grote datasets, willen we de analysevaardigheden in kaart brengen. Big data analytics is het vaak complexe proces van het onderzoeken van big data om informatie te ontdekken - zoals verborgen patronen, correlaties, markttrends en klantvoorkeuren - die organisaties kunnen helpen bij het nemen van weloverwogen zakelijke beslissingen.

In hoeverre zijn onderstaande uitspraken van toepassing op uw team?

1 = helemaal mee oneens, 7 = helemaal mee eens

Data*

	1	2	3	4	5	6	7	N.v.t.
We hebben toegang tot zeer grote, ongestructureerde en/of snel bewegende gegevens voor analyse	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We integreren gegevens uit meerdere interne bronnen in een datawarehouse of subset (datamart) voor gemakkelijke toegang	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We integreren externe gegevens met interne gegevens om hoogwaardige analyses van onze zakelijke omgeving mogelijk te maken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Basismiddelen*

	1	2	3	4	5	6	7	N.v.t.
Onze big data-analytics-projecten worden voldoende gefinancierd	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We krijgen voldoende tijd in onze big data analytics-projecten om onze doelstellingen te bereiken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Technologie*

	1	2	3	4	5	6	7	N.v.t.
We hebben parallelle computerbenaderingen (bijv. Hadoop) voor de verwerking van big data onderzocht of toegepast	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We hebben verschillende tools voor datavisualisatie onderzocht of toegepast	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We hebben nieuwe vormen van databases verkend of aangenomen, zoals Not Only SQL (NoSQL) voor het opslaan van gegevens	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Datagedreven cultuur*

	1	2	3	4	5	6	7	N.v.t.
We baseren onze beslissingen op data in plaats van instinct	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We zijn bereid om onze eigen intuïtie te negeren wanneer data onze standpunten tegenspreken	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We coachen onze medewerkers continu om beslissingen te nemen op basis van data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Lerende organisatie*

	1	2	3	4	5	6	7	N.v.t.
We zijn in staat om nieuwe en relevante kennis op te doen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We hebben gezamenlijke inspanningen geleverd voor het benutten van bestaande competenties en het verkennen van nieuwe kennis	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Digitaal Leiderschap

De essentie van digitaal leiderschap is het balanceren van de externe positionering van het bedrijf in lijn met de interne structuur van de organisatie. Het vinden van deze balans zorgt ervoor dat de organisatie in een passend tempo technologische ontwikkelingen in de markt signaleert en adopteert, inclusief de juiste innovatiecultuur en de benodigde vaardigheden. Met deze vragen willen we het effect van digitaal leiderschap op de prestaties van organisaties onderzoeken.

In hoeverre zijn onderstaande uitspraken van toepassing op uw team?

1 = helemaal mee oneens, 7 = helemaal mee eens

Business plan ontwikkeling*

	1	2	3	4	5	6	7	N.v.t.
We maken gebruik van specialistische kennis om een analyse te maken van de marktomgeving enz.	O	O	O	O	O	O	O	O
Wij bieden leiding aan bij het creëren van een informatiesysteemstrategie die voldoet aan de requirements van de business (bijv. Gedistribueerd, op mobiliteit gebaseerd) en die risico's en kansen omvat	O	O	O	O	O	O	O	O
We passen constant strategisch denken en organisatorisch leiderschap toe om het vermogen van informatietechnologie te benutten om de business te verbeteren	O	O	O	O	O	O	O	O

Architectuur ontwerp*

	1	2	3	4	5	6	7	N.v.t.
We benutten specialistische kennis om relevante ICT-technologie en specificaties te definiëren die kunnen worden ingezet bij de bouw van meerdere ICT-projecten, applicaties of infrastructuurverbeteringen	O	O	O	O	O	O	O	O
We handelen met een brede verantwoordelijkheid om de strategie te bepalen om ICT-technologie te implementeren in overeenstemming met de zakelijke behoeften en houden derhalve rekening met het huidige technologieplatform, verouderde apparatuur en de nieuwste technologische innovaties	O	O	O	O	O	O	O	O
We bieden constant strategisch ICT-leiderschap voor het implementeren van de bedrijfsstrategie. Wij passen strategisch denken toe om nieuwe patronen in enorme datasets en nieuwe ICT-systemen te ontdekken en te herkennen, om zo besparingen ten behoeve van de business te realiseren	O	O	O	O	O	O	O	O

Innoveren*

	1	2	3	4	5	6	7	N.v.t.
We passen voortdurend onafhankelijk denken en technologisch bewustzijn toe om de integratie van ongelijksoortige concepten voor het leveren van unieke oplossingen te leiden	O	O	O	O	O	O	O	O
We dagen voortdurend de status quo uit en bieden strategisch leiderschap bij de introductie van revolutionaire concepten	O	O	O	O	O	O	O	O

Business change management*

	1	2	3	4	5	6	7	N.v.t.
We evalueren voortdurend veranderingsvereisten en maken gebruik van specialistische vaardigheden om mogelijke methoden en standaarden te identificeren die kunnen worden ingezet	O	O	O	O	O	O	O	O
Wij bieden leiderschap bij het plannen, beheren en implementeren van belangrijke door ICT geleide bedrijfsveranderingen	O	O	O	O	O	O	O	O
We passen een diepgaande invloed toe om organisatieverandering te verankeren	O	O	O	O	O	O	O	O

Governance*

	1	2	3	4	5	6	7	N.v.t.
Wij geven invulling aan de governance strategie voor informatiesystemen door relevante processen over de gehele ICT-infrastructuur te communiceren, uit te dragen en te controleren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We definiëren en stemmen de governance strategie voor informatiesystemen voortdurend af en integreren deze in de governance strategie van de organisatie. We passen de governance strategie voor informatiesystemen aan, omdat we rekening houden met nieuwe belangrijke gebeurtenissen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Datageletterdheid

Kort gezegd is datageletterdheid het vermogen om zinvolle informatie uit data te halen. Het doel van de vragen over datageletterdheid is te onderzoeken welke vaardigheden nodig zijn om datageletterdheid te verbeteren.

In hoeverre zijn onderstaande uitspraken van toepassing op uw team?

1 = helemaal mee oneens, 7 = helemaal mee eens

Datacultuur*

	1	2	3	4	5	6	7	N.v.t.
We zien data als een veilige bron en zien het als een aanjager voor vooruitgang en ondersteuning voor bestaande en geplande activiteiten	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Hoger management en leiders ondersteunen data-initiatieven	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Ethiek en beveiliging van data*

	1	2	3	4	5	6	7	N.v.t.
We zijn ons bewust van de gevolgen van het gebruik van data	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We hebben richtlijnen opgesteld voor verantwoord omgaan met data en hebben deze intern in onze activiteiten verwerkt	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vragen stellen/definiëren*

	1	2	3	4	5	6	7	N.v.t.
We zijn in staat om vragen aan data nauwkeurig en doelgericht te formuleren om in de meeste gevallen zinvolle antwoorden te vinden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We zijn in staat om problemen in praktische situaties te identificeren en te beschrijven met behulp van een reeks databronnen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Vinden*

	1	2	3	4	5	6	7	N.v.t.
We hebben goede kennis van verschillende databronnen en de meest relevante kunnen worden gekozen uit een selectie van databronnen	O	O	O	O	O	O	O	O
We zijn op de hoogte van en gebruiken dataportalen voor specifieke onderwerpen	O	O	O	O	O	O	O	O

Verkrijgen*

	1	2	3	4	5	6	7	N.v.t.
We hebben toegang tot data met behulp van complexere dataformaten (bijv. JSON, XML)	O	O	O	O	O	O	O	O
We zijn in staat om API's te gebruiken om data te ontsluiten	O	O	O	O	O	O	O	O

Verifiëren*

	1	2	3	4	5	6	7	N.v.t.
We hebben meerdere lagen van datacontrole geïmplementeerd in standaardprocedures in de hele organisatie	O	O	O	O	O	O	O	O
We kunnen vaststellen of de data betrouwbaar is en indien nodig alternatieve bronnen zoeken	O	O	O	O	O	O	O	O

Opschonen*

	1	2	3	4	5	6	7	N.v.t.
We kunnen ongeldige records detecteren en verwijderen met behulp van programma's die het schonen van data ondersteunen (bijv. OpenRefine)	O	O	O	O	O	O	O	O
We zijn ons zeer bewust van de criteria voor datakwaliteit (bijv. bruikbaar voor machine learning, lege velden, duplicaatdetectie)	O	O	O	O	O	O	O	O

Analyseren*

	1	2	3	4	5	6	7	N.v.t.
We kunnen werken met geavanceerde statistieken (bijv. inferentiële weergaven van data, lineaire regressie, beslisbomen)	O	O	O	O	O	O	O	O
We kunnen de meest geschikte tool voor onze analysebehoefte onderzoeken en selecteren	O	O	O	O	O	O	O	O

Visualiseren*

	1	2	3	4	5	6	7	N.v.t.
We zijn in staat om interactieve grafieken en dashboards te maken	O	O	O	O	O	O	O	O
We kunnen onzekerheden die gepaard gaan met de data visualiseren	O	O	O	O	O	O	O	O

Communiceren*

	1	2	3	4	5	6	7	N.v.t.
Onze projecten worden ondersteund door interactieve visualisaties en meer verfijnde visuele weergaven in een bredere context (bijv. data storytelling, conferenties, lezingen, maandelijkse updates, blogposts)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We kunnen een reeks tabellen, diagrammen en grafieken lezen en begrijpen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Asset en interpreteren*

	1	2	3	4	5	6	7	N.v.t.
We kunnen data output en resultaten vol vertrouwen en kritisch interpreteren	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We hebben evaluatiecriteria opgesteld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Prestaties**

Met deze vragen willen we onderzoeken wat het effect is van de voorgaande aspecten (*big data analytics*, digitaal leiderschap en datageletterdheid) op de prestaties van organisaties.

In hoeverre zijn onderstaande uitspraken van toepassing op uw team?

1 = helemaal mee oneens, 7 = helemaal mee eens

Interne efficiëntie*

	1	2	3	4	5	6	7	N.v.t.
We hebben goed gebruik gemaakt van onze kennis en vaardigheden bij het zoeken naar manieren om efficiënter te worden	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
We proberen de kosten voor organisatiebeheer en prestaties te verlagen	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Interne effectiviteit*

	1	2	3	4	5	6	7	N.v.t.
In de afgelopen twee jaar hebben we een aanzienlijke productiviteitsverbetering vastgesteld	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
De kwaliteit van ons werk is over het algemeen hoog	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Interne eerlijkheid*

	1	2	3	4	5	6	7	N.v.t.
We bieden een eerlijke en rechtvaardige behandeling van werknemers en sollicitanten in alle aspecten van personeelszaken, ongeacht hun politieke overtuiging, geslacht, woonplaats, burgerlijke staat, leeftijd of beperking	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Over het algemeen behandelen we iedereen in onze organisatie met respect, ongeacht status en salaris	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Externe efficiëntie*

	1	2	3	4	5	6	7	N.v.t.
We onderhouden zeer snel zakelijke relaties met externe klanten	O	O	O	O	O	O	O	O
We maken zelden grote fouten bij het uitvoeren van onze werkzaamheden	O	O	O	O	O	O	O	O

Externe effectiviteit*

	1	2	3	4	5	6	7	N.v.t.
Ons werk biedt het publiek een waardevol rendement op hun belastingen	O	O	O	O	O	O	O	O
Het behalen van doelen is erg belangrijk voor ons	O	O	O	O	O	O	O	O

Externe eerlijkheid*

	1	2	3	4	5	6	7	N.v.t.
We bieden eerlijke en billijke diensten aan het publiek, ongeacht hun individuele achtergrond	O	O	O	O	O	O	O	O
Onze klanttevredenheid is zeer hoog	O	O	O	O	O	O	O	O

Afsluiting

U bent nu aangekomen bij de laatste vragen.

Wilt u aanvullende informatie over uw antwoorden of dit onderzoek delen, dan kunt u dat hieronder doen:

...

Indien u de resultaten van de onderzoeken wilt ontvangen, dan kunt u dit hieronder aangeven en uw e-mailadres opgeven.

- ☐ Big data analytics, e-mailadres: ...
- ☐ Digitaal leiderschap, e-mailadres: ...
- ☐ Datageletterdheid, e-mailadres: ...

Bedankt voor het invullen van de vragenlijst.

Als u nog vragen heeft, neem dan contact op met Arjen Davids, via arjen.davids@student.hu.nl

Bijlage 3. Syntax voor vragenlijstanalyse

De vragenlijst is als eerste in SPSS via de syntaxmogelijkheid geanalyseerd. Deze syntax is hieronder te vinden. Na deze eerste (SPSS-)analyse, heeft een PLS-SEM-analyse plaatsgevonden. Deze kan op meerdere manieren uitgevoerd worden. In dit onderzoek is gekozen om de PLS-SEM-analyse via SmartPLS (4) uit te voeren. De schematische weergave hiervan is later in dit hoofdstuk weergegeven. Het is ook mogelijk deze analyse via Python uit te voeren. De code (syntax) hiervoor is te vinden na de SPSS-syntax.

SPSS-syntax

* Encoding: windows-1252.

Dit is de syntax voor de analyse van de vragenlijst over *big data analytics capabilities* (BDAC) voor het afstudeeronderzoek van Arjen Davids

De vragenlijst is uitgezet via Alchemer op 7 juli 2021. De syntax is opgesteld door Arjen Davids.

** Vragenlijstdata inladen.

* LET OP: het verschilt per computer of deze link erkt.

get file "C:\Users\ArjenDavids\OneDrive - Mulier Instituut\Documenten\Master HU - applied data science\Thesis\Data BDAC.sav".

** Data opschonen en klaarzetten voor de analyse.

* Voor de analyse kijken hoeveel respondenten de vragenlijst in hebben gevuld.

fre vstatus.

* Variabelen verwijderen die niet relevant zijn voor dit onderzoek. Te beginnen met algemene (lege) variabelen, gevolgd door de variabelen die bij de vragen horen die buiten de scope van dit onderzoek vallen.

delete variables Vcomment Vlanguage Vreferer Vsessionid Vuseragent var51 var52 var53 var58 var59 var60 var65 var66 var74 var75 var76 var81 var148 var119 var120 var122 var123 var126 var128 var129 var131 var132

var134 var135 var138 var140 var141 var143 var147 var149O782 var149O783 var149O782Othr var149O783Othr.

* Variabelen datageletterdheid t.b.v. human skills (deze komen later aan bod).

* var116 var117 var125 var137 var144 var146.

* Alleen afgeronde vragenlijsten meenemen.

select if Vstatus='Complete'.

* Kijken hoeveel uit andere landen de vragenlijst in hebben gevuld.

fre var3.

* Alleen vragenlijsten meenemen met organisaties in Nederland.

select if var3=10002.

* Uitdraaien hoeveel respondenten werkzaam zijn bij andere organisaties (geen overheidsorganisaties) die de vragenlijst in hebben gevuld.

fre var4.

* Alleen respondenten meenemen die werkzaam zijn bij non-profit overheidsorganisaties.

select if var4=10009.

* Uitdraaien hoeveel respondenten daadwerkelijk voor dit onderzoek meegenomen kunnen worden.

fre Vrid.

* Bij alle antwoorden de optie 'n.v.t.' op missing zetten.

recode var15 var16 var17 (10049=sysmis).

recode var24 var25 (10057=sysmis).

recode var31 var32 var33 (10089=sysmis).

recode var39 var40 var41 (10131=sysmis).

recode var46 var47 (10163=sysmis).

recode var87 var88 (10373=sysmis).

recode var92 var93 (10397=sysmis).

recode var97 var98 (10421=sysmis).

recode var102 var103 (10445=sysmis).

recode var107 var108 (10469=sysmis).

recode var112 var113 (10493=sysmis).

recode var116 var117 (10519=sysmis).

recode var125 (10591=sysmis).

recode var137 (10687=sysmis).

recode var144 (10735=sysmis).

recode var146 (10759=sysmis).

* Check hoeveel respondenten de resultaten willen hebben + e-mailadressen uitdraaien voordat deze verwijderd worden. Deze e-mailadressen separaat opslaan om na afronding van de thesis de resultaten toe te kunnen sturen.

fre var149O781.

fre var149O781Othr.

delete variables var149O781 var149O781Othr.

* Thema's aanmaken en gemiddelden per thema berekenen.

```
compute bdac=1.
```

```
compute data=(var15+var16+var17)/3.
```

```
compute basismiddelen=(var24+var25)/2.
```

```
compute technologie=(var31+var32+var33)/3.
```

```
compute datagedreven_cultuur=(var39+var40+var41)/3.
```

```
compute lerende_organisatie=(var46+var47)/2.
```

```
compute managementvaardigheden=(var116+var117+var144+var146)/4.
```

```
compute technische_vaardigheden=(var137+var125)/2.
```

```
compute prestaties=1.
```

```
compute interne_efficientie=(var87+var88)/2.
```

```
compute interne_effectiviteit=(var92+var93)/2.
```

```
compute interne_eerlijkheid=(var97+var98)/2.
```

```
compute externe_efficientie=(var102+var103)/2.
```

```
compute externe_effectiviteit=(var107+var108)/2.
```

```
compute externe_eerlijkheid=(var112+var113)/2.
```

* Leeftijdscategorie maken o.b.v. geboortjaar en invuljaar.

```
compute leeftijd=2021-var8.
```

```
variable labels leeftijd 'leeftijd'.
```

* Check ander opleidingsniveau.

```
fre var9O779Othr.
```

* Eén ander opleidingsniveau (PhD) aanpassen naar WO.

```
recode var9 (10779=10030).
```

* Check andere rolomschrijving.

```
fre var11O777Othr.
```

* Andere rolomschrijvingen hercategoriseren.

```
do if var11O777Othr = 'Adviseur (Geo)Informatie'.
```

```
recode var11 (10777=10038).
```

```
end if.
```

```
do if var11O777Othr = 'Beleidsadvies en Projectmanager functioneel en technisch leidinggevende'.
```

```
recode var11 (10777=10041).
```

```
end if.
```

```
do if var11O777Othr = 'CIO Office member - consultant'.
recode var11 (10777=10038).
end if.
```

```
do if var11O777Othr = 'Geo Vastgoed Coordinator'.
recode var11 (10777=10041).
end if.
```

```
do if var11O777Othr = 'Policy advisor'.
recode var11 (10777=10038).
end if.
```

```
do if var11O777Othr = 'procesmanager'.
recode var11 (10777=10041).
end if.
```

```
* Databestand is nu klaar voor de analyse. Deze onder een andere naam opslaan.
save outfile "C:\Users\ArjenDavids\OneDrive - Mulier Instituut\Documenten\Master HU - applied data
science\Thesis\Data thesis BDAC.sav".
```

```
*****
```

```
*** Analyse.
```

```
* Achtergrondkenmerken uitdraaien.
```

```
* Aantal jaar werkervaring.
```

```
mean var5.
```

```
* Grootte organisatie.
```

```
fre var6.
```

```
* Geslacht.
```

```
fre var7.
```

```
* Leeftijd.
```

```
mean leeftijd.
```

```
* Opleidingsniveau.
```

```
fre var9.
```

* Aantal jaar ervaring met (big) data analytics.

fre var10.

* Omschrijving rol (anders=anoniem).

fre var11.

* Interne consistentie per categorie met Cronbach's Alpha uitdraaien.

reliability /var=var15 var16 var17 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var24 var25 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var31 var32 var33 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var39 var40 var41 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var46 var47 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var116 var117 var144 var146 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var137 var125 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var87 var88 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var92 var93 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var97 var98 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var102 var103 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var107 var108 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

reliability /var=var112 var113 /scale('all variables') all /model=alpha /statistics=scale /summary=total.

* Gemiddelden per vraag en categorie uitdraaien.

* BDAC.

mean var15 var16 var17 var24 var25 var31 var32 var33 var39 var40 var41 var46 var47 var116 var117 var137 var125.

mean data basismiddelen technologie datagedreven_cultuur lerende_organisatie managementvaardigheden technische_vaardigheden.

* Prestaties.

mean var87 var88 var92 var93 var97 var98 var102 var103 var107 var108 var112 var113.

mean interne_efficientie interne_effectiviteit interne_eerlijkheid externe_efficientie externe_effectiviteit externe_eerlijkheid.

* Open antwoorden uitdraaien.

fre var114.

Bijlage 4. Uitgebreide analyseresultaten

Bijlage 4.1 Uitkomsten analyse respondenten en opschonen databestand

Status vragenlijst

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Complete	144	30,2	30,2	30,2
Partial	333	69,8	69,8	100,0
Total	477	100,0	100,0	

In welk land is uw organisatie gevestigd?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Nederland	124	86,1	86,1	86,1
Duitsland	18	12,5	12,5	98,6
Hongarije	2	1,4	1,4	100,0
Total	144	100,0	100,0	

Onder welke sector valt uw organisatie?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Commercieel	1	,8	,8	,8
Non-profit overheid	120	96,8	96,8	97,6
Non-profit anders	3	2,4	2,4	100,0
Total	124	100,0	100,0	

Aantal respondenten voor analyse

N	Valid	120
	Missing	0

Aantal respondenten dat de resultaten na afloop wilt ontvangen

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Unchecked	83	69,2	69,2	69,2
Checked	37	30,8	30,8	100,0
Total	120	100,0	100,0	

Bijlage 4.2 Algemene uitkomsten

Hoeveel jaar werkervaring heeft u?

Mean	N	Std. Deviation
20,03	120	12,216

Hoeveel medewerkers heeft uw organisatie?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Minder dan 10	1	,8	,8	,8
10-50	5	4,2	4,2	5,0
51-100	4	3,3	3,3	8,3
101-1 000	72	60,0	60,0	68,3
Meer dan 1 000	38	31,7	31,7	100,0
Total	120	100,0	100,0	

Wat is uw geslacht?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Man	82	68,3	68,3	68,3
Vrouw	31	25,8	25,8	94,2
Wil ik niet zeggen	7	5,8	5,8	100,0
Total	120	100,0	100,0	

Leeftijd

Mean	N	Std. Deviation
45,3917	120	13,68757

Wat is uw hoogst afgeronde opleiding?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Basisschool	1	,8	,8	,8
Middelbaar beroepsonderwijs (mbo)	2	1,7	1,7	2,5
Hoger beroepsonderwijs (hbo)	51	42,5	42,5	45,0
Wetenschappelijk onderwijs (wo)	62	51,7	51,7	96,7
Middelbare school	2	1,7	1,7	98,3
Wil ik niet zeggen	2	1,7	1,7	100,0
Total	120	100,0	100,0	

Hoeveel jaar ervaring heeft u met (big) data analytics?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Geen, ik heb er geen ervaring mee	25	20,8	20,8	20,8
Minder dan 1 jaar	11	9,2	9,2	30,0
1 - 2 jaar	16	13,3	13,3	43,3
2 - 3 jaar	10	8,3	8,3	51,7
3 - 4 jaar	12	10,0	10,0	61,7
Meer dan 4 jaar	46	38,3	38,3	100,0
Total	120	100,0	100,0	

Wat omschrijft uw rol het beste?

	Frequency	Percent	Valid Percent	Cumulative Percent
Valid Medewerker, niet-leidinggevend	75	62,5	62,5	62,5
Autonoom professional, niet-vallend onder een divisie, afdeling of team	13	10,8	10,8	73,3
Leidinggevende/manager over een afdeling of team	29	24,2	24,2	97,5
Anders	1	,8	,8	98,3
CEO, bestuurder/directeur/manager van een bedrijf of organisatie	2	1,7	1,7	100,0
Total	120	100,0	100,0	

Bijlage 4.3 Interne consistentie met Cronbach's Alpha via SPSS

Data

Case Processing Summary

		N	%
Cases	Valid	109	90,8
	Excluded ^a	11	9,2
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,654	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We hebben toegang tot zeer grote, ongestructureerde en/of snel bewegende gegevens voor analyse	8,70	11,750	,300	,775
We integreren gegevens uit meerdere interne bronnen in een datawarehouse of subset (datamart) voor gemakkelijke toegang	8,26	9,489	,546	,443
We integreren externe gegevens met interne gegevens om hoogwaardige analyses van onze zakelijke omgeving mogelijk te maken	8,81	9,824	,577	,409

Basismiddelen

Case Processing Summary

		N	%
Cases	Valid	104	86,7
	Excluded ^a	16	13,3
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,888	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Onze big data-analytics-projecten worden voldoende gefinancierd	3,48	2,796	,799	.
We krijgen voldoende tijd in onze big data analytics-projecten om onze doelstellingen te bereiken	3,47	2,931	,799	.

Technologie

Case Processing Summary

		N	%
Cases	Valid	94	78,3
	Excluded ^a	26	21,7
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,790	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We hebben parallele computerbenaderingen (bijv. Hadoop) voor de verwerking van big data onderzocht of toegepast	8,15	11,763	,630	,716
We hebben verschillende tools voor datavisualisatie onderzocht of toegepast	6,06	13,243	,525	,819
We hebben nieuwe vormen van databases verkend of aangenomen, zoals Not Only SQL (NoSQL) voor het opslaan van gegevens	7,34	9,474	,756	,567

Datagedreven cultuur

Case Processing Summary

		N	%
Cases	Valid	118	98,3
	Excluded ^a	2	1,7
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,832	3

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We baseren onze beslissingen op data in plaats van instinct	7,34	9,029	,772	,693
We zijn bereid om onze eigen intuïtie te negeren wanneer data onze standpunten tegenspreken	7,34	9,970	,629	,826
We coachen onze medewerkers continu om beslissingen te nemen op basis van data	7,51	8,252	,688	,778

Lerende organisatie

Case Processing Summary

		N	%
Cases	Valid	120	100,0
	Excluded ^a	0	,0
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,843	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We zijn in staat om nieuwe en relevante kennis op te doen	4,43	1,844	,729	.
We hebben gezamenlijke inspanningen geleverd voor het benutten van bestaande competenties en het verkennen van nieuwe kennis	4,71	1,939	,729	.

Managementvaardigheden

Case Processing Summary

		N	%
Cases	Valid	115	95,8
	Excluded ^a	5	4,2
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,817	4

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We zien data als een veilige bron en zien het als een aanjager voor vooruitgang en ondersteuning voor bestaande en geplande activiteiten	14,72	15,922	,569	,801
Hoger management en leiders ondersteunen data-initiatieven	15,09	14,624	,568	,803
We kunnen een reeks tabellen, diagrammen en grafieken lezen en begrijpen	14,48	13,041	,700	,740
We kunnen data output en resultaten vol vertrouwen en kritisch interpreteren	14,83	13,198	,726	,727

Technische vaardigheden

Case Processing Summary

		N	%
Cases	Valid	114	95,0
	Excluded ^a	6	5,0
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,714	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We kunnen werken met geavanceerde statistieken (bijv. inferentiële weergaven van data, lineaire regressie, beslisbomen)	4,51	2,465	,561	.
We hebben goede kennis van verschillende databronnen en de meest relevante kunnen worden gekozen uit een selectie van databronnen	4,10	3,291	,561	.

Interne efficiëntie

Case Processing Summary

		N	%
Cases	Valid	112	93,3
	Excluded ^a	8	6,7
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,789	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We hebben goed gebruik gemaakt van onze kennis en vaardigheden bij het zoeken naar manieren om efficiënter te worden	4,46	2,070	,654	.
We proberen de kosten voor organisatiebeheer en prestaties te verlagen	4,21	2,440	,654	.

Interne effectiviteit

Case Processing Summary

		N	%
Cases	Valid	112	93,3
	Excluded ^a	8	6,7
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,684	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
In de afgelopen twee jaar hebben we een aanzienlijke productiviteitsverbetering vastgesteld	4,81	2,046	,522	.
De kwaliteit van ons werk is over het algemeen hoog	3,88	2,518	,522	.

Interne eerlijkheid

Case Processing Summary

		N	%
Cases	Valid	117	97,5
	Excluded ^a	3	2,5
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,857	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We bieden een eerlijke en rechtvaardige behandeling van werknemers en sollicitanten in alle aspecten van personeelszaken, ongeacht hun politieke overtuiging, geslacht, woonplaats, burgerlijke staat, leeftijd of beperking	6,13	1,164	,761	.
Over het algemeen behandelen we iedereen in onze organisatie met respect, ongeacht status en salaris	5,95	1,635	,761	.

Externe efficiëntie

Case Processing Summary

		N	%
Cases	Valid	111	92,5
	Excluded ^a	9	7,5
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,690	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We onderhouden zeer snel zakelijke relaties met externe klanten	4,86	1,743	,527	.
We maken zelden grote fouten bij het uitvoeren van onze werkzaamheden	4,97	1,627	,527	.

Externe effectiviteit

Case Processing Summary

		N	%
Cases	Valid	112	93,3
	Excluded ^a	8	6,7
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,663	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Ons werk biedt het publiek een waardevol rendement op hun belastingen	5,28	1,301	,508	.
Het behalen van doelen is erg belangrijk voor ons	5,07	1,995	,508	.

Externe eerlijkheid

Case Processing Summary

		N	%
Cases	Valid	111	92,5
	Excluded ^a	9	7,5
	Total	120	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	N of Items
,696	2

Item-Total Statistics

	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
We bieden eerlijke en billijke diensten aan het publiek, ongeacht hun individuele achtergrond	5,04	1,217	,535	.
Onze klanttevredenheid is zeer hoog	5,95	1,415	,535	.

Bijlage 4.4 Gemiddelden per categorie

BDAC

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
data	109	90,8%	11	9,2%	120	100,0%
basismiddelen	104	86,7%	16	13,3%	120	100,0%
technologie	94	78,3%	26	21,7%	120	100,0%
datagedreven_cultuur	118	98,3%	2	1,7%	120	100,0%
lerende_organisatie	120	100,0%	0	0,0%	120	100,0%
managementvaardigheden	115	95,8%	5	4,2%	120	100,0%
technische_vaardigheden	114	95,0%	6	5,0%	120	100,0%

	data	basismiddelen	technologie	Data gedreven_ cultuur	lerende_ organisatie	management vaardigheden	technische_ vaardigheden
Mean	4,2936	3,4760	3,5922	3,6977	4,5708	4,9261	4,3026
N	109	104	94	118	120	115	114
Std. Deviation	1,48568	1,60457	1,61249	1,44739	1,27894	1,21935	1,49574

Prestaties

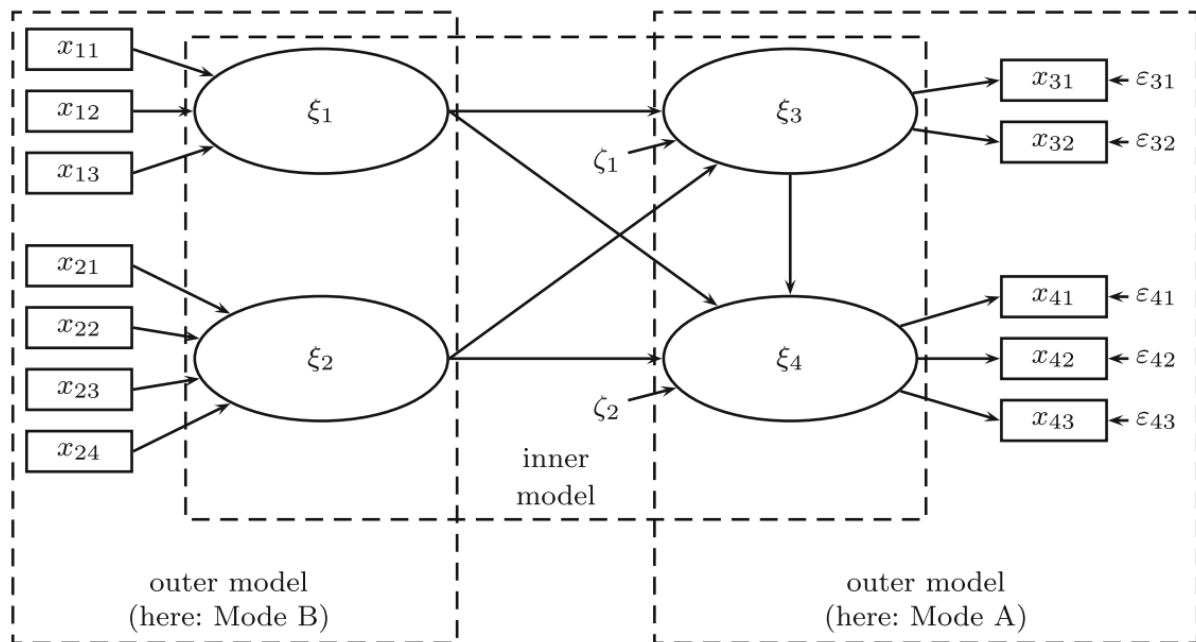
Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
interne_efficientie	112	93,3%	8	6,7%	120	100,0%
interne_effectiviteit	112	93,3%	8	6,7%	120	100,0%
interne_eerlijkheid	117	97,5%	3	2,5%	120	100,0%
externe_efficientie	111	92,5%	9	7,5%	120	100,0%
externe_effectiviteit	112	93,3%	8	6,7%	120	100,0%
externe_eerlijkheid	111	92,5%	9	7,5%	120	100,0%

	interne_ efficientie	interne_ effectiviteit	interne_ eerlijkheid	externe_ efficientie	externe_ effectiviteit	externe_ eerlijkheid
Mean	4,3348	4,3482	6,0385	4,9144	5,1741	5,4910
N	112	112	117	111	112	111
Std. Deviation	1,36459	1,31663	1,10671	1,13393	1,11037	1,00449

Bijlage 4.5 Extra uitkomsten PLS-SEM-analyse

Figuur b4.5a Schematische weergave Partial Least Squares based Structural Equation Modeling (PLS-SEM)

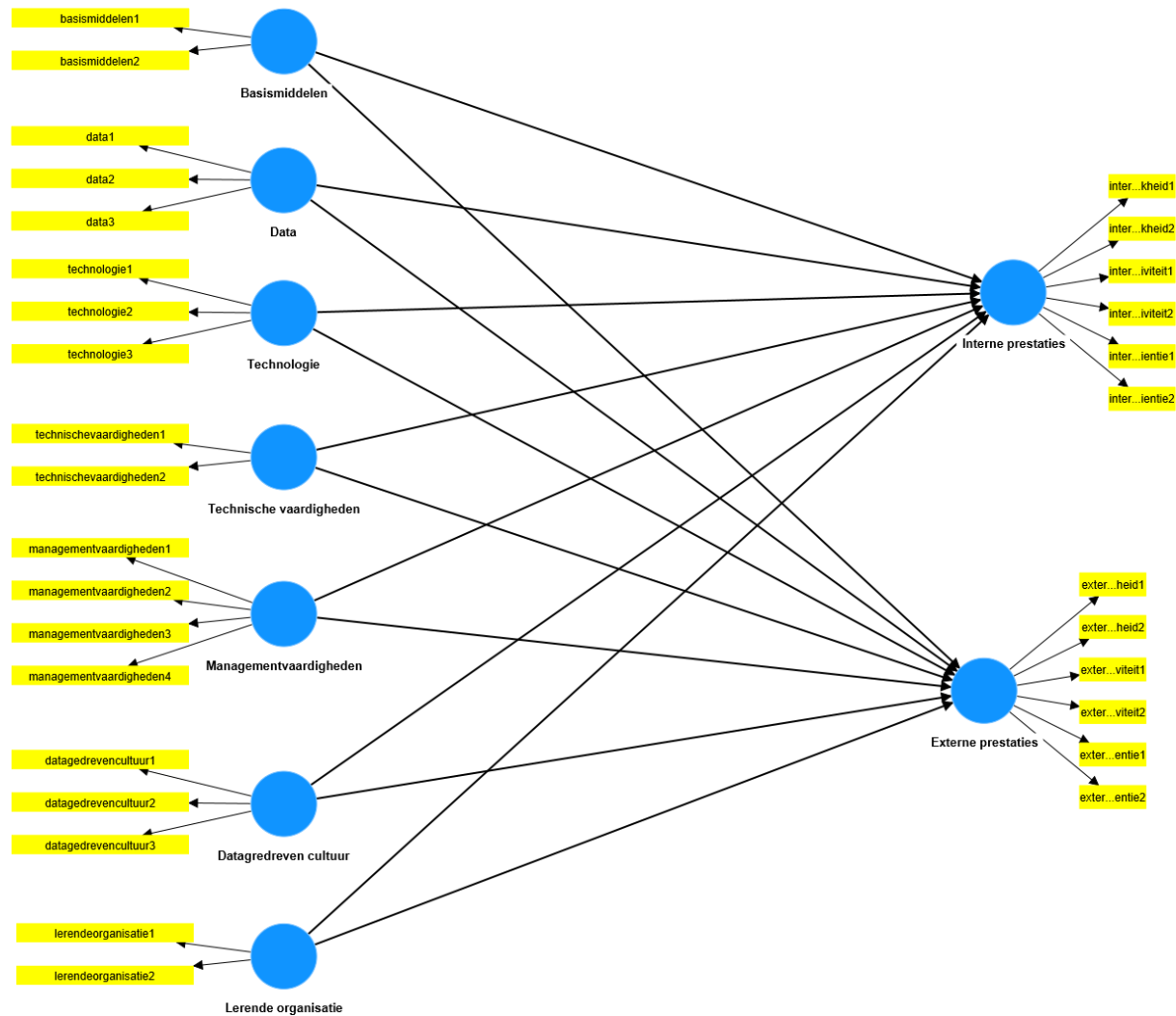


Bron: Henseler, 2010.

Tabel b4.5a Cross-loadings PLS-SEM-analyse in SmartPLS 4 t.b.v. de hypotheses

	BDAC	Interne prestaties	Externe prestaties
basismiddelen1	0,651	0,341	0,146
basismiddelen2	0,663	0,387	0,175
data1	0,418	0,257	0,113
data2	0,614	0,352	0,225
data3	0,753	0,374	0,233
technologie1	0,437	0,206	0,067
technologie2	0,608	0,319	0,188
technologie3	0,561	0,285	0,106
technische-vaardigheden1	0,740	0,556	0,465
technische-vaardigheden2	0,780	0,504	0,291
managementvaardigheden1	0,639	0,437	0,426
managementvaardigheden2	0,711	0,479	0,383
managementvaardigheden3	0,709	0,528	0,511
managementvaardigheden4	0,782	0,568	0,490
datagedreven-cultuur1	0,715	0,477	0,277
datagedreven-cultuur2	0,553	0,387	0,130
datagedreven-cultuur3	0,725	0,524	0,350
lerende-organisatie1	0,752	0,642	0,517
lerende-organisatie2	0,759	0,566	0,455
interne-eerlijkheid1	0,201	0,466	0,592
interne-eerlijkheid2	0,249	0,468	0,605
interne-effectiviteit1	0,564	0,814	0,470
interne-effectiviteit2	0,598	0,783	0,649
interne-efficiëntie1	0,635	0,867	0,492
interne-efficiëntie2	0,407	0,716	0,374
externe-eerlijkheid1	0,207	0,444	0,698
externe-eerlijkheid2	0,365	0,524	0,757
externe-effectiviteit1	0,315	0,512	0,757
externe-effectiviteit2	0,500	0,631	0,798
externe-efficiëntie1	0,275	0,423	0,673
externe-efficiëntie2	0,388	0,487	0,713

Figuur b4.5b Schematische weergave PLS-SEM-analyse in SmartPLS 4 voor post-hoc analyse



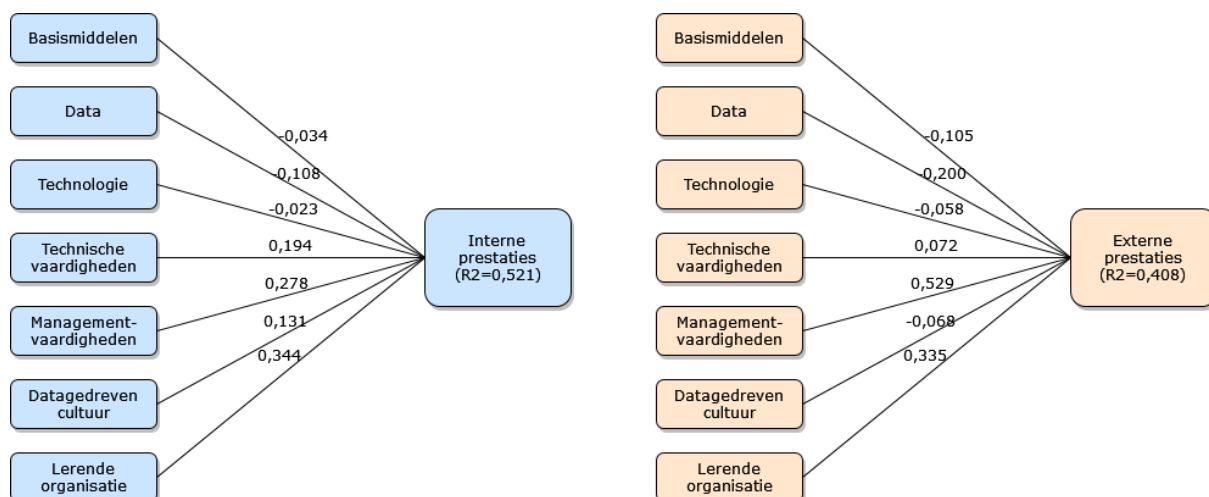
Tabel b4.5b Resultaten betrouwbaarheid en validiteit t.b.v. categorisatie stellingen *big data analytics capabilities* / *post-hoc analyse*

	1	2	3	4	5	6	7	8	9
1 Basismiddelen	0,943								
2 Data	0,741	0,777							
3 Datagredreven cultuur	0,696	0,592	0,866						
4 Externe prestaties	0,185	0,303	0,317	0,736					
5 Interne prestaties	0,402	0,515	0,566	0,913	0,707				
6 Lerende organisatie	0,580	0,677	0,706	0,593	0,770	0,930			
7 Managementvaardigheden	0,559	0,866	0,698	0,658	0,755	0,803	0,802		
8 Technische vaardigheden	0,649	1,053	0,690	0,522	0,732	0,869	0,974	0,880	
9 Technologie	0,687	0,835	0,595	0,212	0,408	0,440	0,580	0,698	0,825
Cronbach's alpha	0,877	0,657	0,835	0,832	0,798	0,844	0,813	0,711	0,776
Composite reliability (rho_a)	0,887	0,714	0,871	0,841	0,840	0,855	0,826	0,742	0,832
Composite reliability (rho_c)	0,942	0,815	0,900	0,876	0,853	0,927	0,878	0,872	0,864
Average variance extracted (AVE)	0,890	0,603	0,750	0,541	0,500	0,864	0,643	0,774	0,681

Tabel b4.5c Resultaten validatiemaatregelen voor formatieve constructen t.b.v. categorische stellingen *big data analytics capabilities / post-hoc analyse*

Construct	Measures	VIF	Outer weights	Loadings
BDAC	basismiddelen1	2,563	0,496	0,936
	basismiddelen2	2,563	0,563	0,951
	data1	1,098	0,308	0,565
	data2	1,684	0,465	0,854
	data3	1,709	0,492	0,872
	technologie1	1,814	0,267	0,733
	technologie2	1,481	0,520	0,851
	technologie3	2,320	0,409	0,884
	technische-vaardigheden1	1,439	0,637	0,911
	technische-vaardigheden2	1,439	0,496	0,847
	managementvaardigheden1	1,540	0,275	0,729
	managementvaardigheden2	1,587	0,283	0,742
	managementvaardigheden3	2,843	0,340	0,856
	managementvaardigheden4	2,968	0,343	0,871
	datagedreven-cultuur1	2,497	0,400	0,909
	datagedreven-cultuur2	1,715	0,292	0,788
	datagedrevencultuur3	2,081	0,454	0,895
	lerende-organisatie1	2,137	0,574	0,939
	lerende-organisatie2	2,137	0,501	0,920
Interne prestaties	interne-eerlijkheid1	2,429	0,145	0,530
	interne-eerlijkheid2	2,402	0,164	0,531
	interne-effectiviteit1	2,088	0,256	0,784
	interne-effectiviteit2	1,690	0,317	0,794
	interne-efficiëntie1	2,626	0,294	0,841
	interne-efficiëntie2	1,777	0,198	0,696
Externe prestaties	externe-eerlijkheid1	1,728	0,200	0,733
	externe-eerlijkheid2	1,767	0,251	0,763
	externe-effectiviteit1	1,841	0,166	0,740
	externe-effectiviteit2	1,734	0,285	0,777
	externe-efficiëntie1	1,564	0,200	0,687
	externe-efficiëntie2	1,632	0,253	0,711

Figuur b4.5c Geschatte relaties BDAC op interne en externe prestaties (structural model) / *post-hoc analyse*



Tabel b4.5d Resultaten PLSpredict van manifeste variabelen* (organisatorische prestaties) in relatie tot *big data analytics capabilities* / *post-hoc analyse*

Variabele	PLS-SEM		LM		PLS<LM	
	RMSE	MAE	RMSE	MAE	RMSE	MAE
Interne eerlijkheid1	1,262	0,941	1,365	1,027	Ja	Ja
Interne eerlijkheid2	1,056	0,808	1,147	0,890	Ja	Ja
Interne effectiviteit1	1,281	1,043	1,398	1,135	Ja	Ja
Interne effectiviteit2	1,148	0,872	1,212	0,906	Ja	Ja
Interne efficiëntie1	1,217	0,966	1,370	1,041	Ja	Ja
Interne efficiëntie2	1,292	0,977	1,439	1,083	Ja	Ja
Externe eerlijkheid1	1,155	0,828	1,251	0,933	Ja	Ja
Externe eerlijkheid2	1,007	0,788	1,052	0,866	Ja	Ja
Externe effectiviteit1	1,314	0,986	1,510	1,145	Ja	Ja
Externe effectiviteit2	1,027	0,757	1,059	0,806	Ja	Ja
Externe efficiëntie1	1,214	0,976	1,351	1,083	Ja	Ja
Externe efficiëntie2	1,240	0,977	1,344	1,078	Ja	Ja

Tabel bb4.5e Cross-loadings PLS-SEM-analyse in SmartPLS 4 voor post-hoc analyse

	1	2	3	4	5	6	7	8	9
basismiddelen1	0,936	0,551	0,535	0,129	0,326	0,456	0,444	0,462	0,528
basismiddelen2	0,951	0,525	0,594	0,161	0,363	0,480	0,427	0,482	0,512
data1	0,317	0,565	0,202	0,109	0,246	0,308	0,298	0,397	0,407
data2	0,390	0,854	0,358	0,213	0,341	0,343	0,534	0,585	0,456
data3	0,590	0,872	0,474	0,219	0,365	0,514	0,638	0,655	0,565
datagedreven-cultuur1	0,589	0,436	0,909	0,262	0,459	0,541	0,537	0,489	0,432
datagedreven-cultuur2	0,453	0,285	0,788	0,124	0,374	0,452	0,385	0,343	0,320
datagedreven-cultuur3	0,512	0,450	0,895	0,335	0,500	0,542	0,557	0,534	0,480
externe-eerlijkheid1	-0,017	0,033	0,085	0,733	0,486	0,266	0,316	0,178	0,015
externe-eerlijkheid2	0,028	0,175	0,282	0,763	0,544	0,409	0,425	0,324	0,062
externe-effectiviteit1	0,142	0,192	0,215	0,740	0,519	0,271	0,349	0,258	0,194
externe-effectiviteit2	0,246	0,263	0,372	0,777	0,651	0,494	0,544	0,408	0,183
externe-efficiëntie1	0,086	0,128	0,143	0,687	0,450	0,318	0,339	0,250	0,036
externe-efficiëntie2	0,158	0,225	0,140	0,711	0,504	0,447	0,429	0,414	0,137
interne-eerlijkheid1	-0,023	0,075	0,027	0,613	0,530	0,309	0,296	0,192	-0,023
interne-eerlijkheid2	0,041	0,104	0,079	0,619	0,531	0,346	0,332	0,221	0,010
interne-effectiviteit1	0,406	0,361	0,547	0,455	0,784	0,465	0,486	0,448	0,340
interne-effectiviteit2	0,273	0,358	0,454	0,644	0,794	0,611	0,577	0,565	0,227
interne-efficiëntie1	0,419	0,468	0,498	0,479	0,841	0,561	0,548	0,588	0,354
interne-efficiëntie2	0,252	0,217	0,370	0,368	0,696	0,392	0,338	0,338	0,292
lerende-organisatie1	0,425	0,481	0,535	0,512	0,650	0,939	0,632	0,605	0,346
lerende-organisatie2	0,505	0,461	0,574	0,449	0,565	0,920	0,604	0,636	0,352
managementvaardigheden1	0,377	0,442	0,468	0,410	0,435	0,431	0,729	0,488	0,401
managementvaardigheden2	0,497	0,474	0,520	0,382	0,481	0,591	0,742	0,502	0,442
managementvaardigheden3	0,286	0,536	0,367	0,508	0,537	0,546	0,856	0,685	0,318
managementvaardigheden4	0,349	0,629	0,523	0,483	0,565	0,564	0,871	0,714	0,400
technische-vaardigheden1	0,371	0,578	0,448	0,460	0,555	0,558	0,707	0,911	0,384
technische-vaardigheden2	0,534	0,702	0,511	0,278	0,491	0,626	0,613	0,847	0,558
technologie1	0,521	0,379	0,371	0,046	0,181	0,201	0,252	0,328	0,733
technologie2	0,424	0,614	0,370	0,183	0,310	0,370	0,498	0,487	0,851
technologie3	0,467	0,476	0,471	0,093	0,267	0,315	0,377	0,441	0,884

Toelichting: 1 = Basismiddelen, 2 = Data, 3 = Datagedreven cultuur, 4 = Externe prestaties, 5 = Interne prestaties, 6 = Lerende organisatie, 7 = Managementvaardigheden, 8 = Technische vaardigheden, 9 = Technologie.

Bijlage 4.6 Vergelijking uitkomsten per categorie en per stelling: onderzoek vs. Literatuur

Tabel b4.6a Gemiddelde en standaarddeviatie (SD) categorieën (o.b.v. stellingen) over *big data analytics capabilities* en organisatorische prestaties bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7*)

	Categorie	Onderzoek		Literatuur	
		Gemiddelde	SD	Gemiddelde	SD
Big Data Analytics Capabilities*	Basismiddelen	3,48	1,60	4,07	1,70
	Data	4,29	1,49	4,66	1,77
	Technologie	3,59	1,61	4,22	2,03
	Technische vaardigheden	4,30	1,50	4,24	1,71
	Managementvaardigheden	4,93	1,22	4,37	1,62
	Datagedreven cultuur	3,70	1,45	4,42	1,48
	Lerende organisatie	4,57	1,28	4,77	1,35
Organizational Performance*	Interne efficiëntie	4,33	1,36	4,47	1,23
	Interne effectiviteit	4,35	1,32	4,82	1,11
	Interne eerlijkheid	6,04	1,11	4,20	1,39
	Externe efficiëntie	4,91	1,13	4,87	1,13
	Externe effectiviteit	5,17	1,11	5,13	1,11
	Externe eerlijkheid	5,49	1,00	5,17	1,11

Bron: Vragenlijst *big data analytics*, *digitaal leiderschap*, *datageletterdheid* en *organisatorische prestaties* binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021; Mikalef et al., 2020; Gupta en George, 2016; Kim, 2004.

Toelichting: Het aantal waarnemingen verschilt voor het huidige onderzoek per stelling en categorie doordat de antwoordcategorie 'niet van toepassing' op missing gezet is. Het aantal waarnemingen bij dit onderzoek liggen voor BDAC tussen de 94 en 120. Bij organisatorische prestaties ligt dit tussen 111 en 117. Het aantal waarnemingen in de literatuur is respectievelijk 202 en 1.739.

* De stellingen over 'organizational performance' zijn op een 5-puntsschaal uitgevraagd. Deze zijn omgerekend naar het gemiddelde op een 7-puntsschaal.

Legenda: Groene arcering = uitkomst huidig onderzoek hoger t.o.v. literatuur.

Oranje arcering = uitkomst huidig onderzoek lager t.o.v. literatuur.

Tabel b4.6b Gemiddelde en standaarddeviatie (SD) stellingen over *big data analytics capabilities* bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7)

	Categorie	Stelling	Onderzoek		Literatuur	
			Gemiddelde	SD	Gemiddelde	SD
Big data analytics capabilities*	Basismiddelen	Onze big data-analytics-projecten worden voldoende gefinancierd	3,45	1,71	4,13	1,77
		We krijgen voldoende tijd in onze big data analytics-projecten om onze doelstellingen te bereiken	3,51	1,67	4,01	1,62
	Data	We hebben toegang tot zeer grote, ongestructureerde en/of snel bewegende gegevens voor analyse	4,19	1,99	4,91	1,74
		We integreren gegevens uit meerdere interne bronnen in een datawarehouse of subset (datamart) voor gemakkelijke toegang	4,66	1,93	5,05	1,76
		We integreren externe gegevens met interne gegevens om hoogwaardige analyses van onze zakelijke omgeving mogelijk te maken	4,10	1,83	4,011	1,80
	Technologie	We hebben parallelle computerbenaderingen (bijv. Hadoop) voor de verwerking van big data onderzocht of toegepast	2,62	1,86	3,69	2,12
		We hebben verschillende tools voor datavisualisatie onderzocht of toegepast	4,64	1,79	4,81	1,85
		We hebben nieuwe vormen van databases verkend of aangenomen, zoals Not Only SQL (NoSQL) voor het opslaan van gegevens	3,47	2,02	4,17	2,13
	Technische vaardigheden	We kunnen werken met geavanceerde statistieken (bijv. inferentiële weergaven van data, lineaire regressie, beslisbomen)	4,10	1,81	4,24	1,70
		We hebben goede kennis van verschillende databronnen en de meest relevante kunnen worden gekozen uit een selectie van databronnen	4,55	1,57	4,24	1,71
	Managementvaardigheden	We zien data als een veilige bron en zien het als een aanjager voor vooruitgang en ondersteuning voor bestaande en geplande activiteiten	4,97	1,33	4,40	1,65
		Hoger management en leiders ondersteunen data-initiatieven	4,62	1,55	4,18	1,63
		We kunnen data output en resultaten vol vertrouwen en kritisch interpreteren	4,89	1,54		
		We kunnen een reeks tabellen, diagrammen en grafieken lezen en begrijpen	5,26	1,61	4,52	1,58
	Datagedreven cultuur	We baseren onze beslissingen op data in plaats van instinct	3,76	1,58	4,39	1,47
		We zijn bereid om onze eigen intuïtie te negeren wanneer data onze standpunten tegenspreken	3,75	1,60	4,51	1,52
		We coachen onze medewerkers continu om beslissingen te nemen op basis van data	3,60	1,83	4,37	1,46
	Lerende organisatie	We zijn in staat om nieuwe en relevante kennis op te doen	4,71	1,39	5,02	1,29
		We hebben gezamenlijke inspanningen geleverd voor het benutten van bestaande competenties en het verkennen van nieuwe kennis	4,43	1,36	4,51	1,41

Bron: Vragenlijst *big data analytics*, *digitaal leiderschap*, *datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties*, uitgezet via sociale media en e-mails, 2021; Mikalef et al., 2020; Gupta en George, 2016.

Toelichting: Het aantal waarnemingen verschilt voor het huidige onderzoek per stelling en categorie doordat de antwoordcategorie 'niet van toepassing' op missing gezet is. Het aantal waarnemingen bij dit onderzoek ligt tussen de 94 en 120. Het aantal waarnemingen in de literatuur is 202.

Tabel b4.6c Gemiddelde en standaarddeviatie (SD) stellingen over organisatorische prestaties bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7*)

	Categorie	Stelling	Onderzoek		Literatuur	
			Gemiddelde	SD	Gemiddelde	SD
Organizational Performance*	Interne efficiëntie	We hebben goed gebruik gemaakt van onze kennis en vaardigheden bij het zoeken naar manieren om efficiënter te worden	4,25	1,56	4,35	1,25
		We proberen de kosten voor organisatiebeheer en prestaties te verlagen	4,46	1,44	4,59	1,20
	Interne effectiviteit	In de afgelopen twee jaar hebben we een aanzienlijke productiviteitsverbetering vastgesteld	3,88	1,59	4,72	1,12
		De kwaliteit van ons werk is over het algemeen hoog	4,84	1,40	4,93	1,11
	Interne eerlijkheid	We bieden een eerlijke en rechtvaardige behandeling van werknemers en sollicitanten in alle aspecten van personeelszaken, ongeacht hun politieke overtuiging, geslacht, woonplaats, burgerlijke staat, leeftijd of beperking	5,95	1,28	4,17	1,44
		Over het algemeen behandelen we iedereen in onze organisatie met respect, ongeacht status en salaris	6,14	1,08	4,23	1,34
	Externe efficiëntie	We onderhouden zeer snel zakelijke relaties met externe klanten	4,99	1,28	4,75	1,18
		We maken zelden grote fouten bij het uitvoeren van onze werkzaamheden	4,91	1,33	5,00	1,09
	Externe effectiviteit	Ons werk biedt het publiek een waardevol rendement op hun belastingen	5,07	1,41	5,24	1,16
		Het behalen van doelen is erg belangrijk voor ons	5,32	1,16	5,03	1,05
	Externe eerlijkheid	We bieden eerlijke en billijke diensten aan het publiek, ongeacht hun individuele achtergrond	5,96	1,18	5,36	1,13
		Onze klanttevredenheid is zeer hoog	5,04	1,10	4,98	1,09

Bron: Vragenlijst big data analytics, digitaal leiderschap, datageletterdheid en organisatorische prestaties binnen Nederlandse overheidsorganisaties, uitgezet via sociale media en e-mails, 2021; Kim, 2004.

Toelichting: Het aantal waarnemingen verschilt voor het huidige onderzoek per stelling en categorie doordat de antwoordcategorie 'niet van toepassing' op missing gezet is. Het aantal waarnemingen bij dit onderzoek ligt tussen de 111 en 117. Het aantal waarnemingen in de literatuur is 1.739.

** De vragen zijn in de literatuur op een 5-puntsschaal uitgevraagd. Deze zijn omgerekend naar het gemiddelde op een 7-puntsschaal.*

Bijlage 4.7 Open antwoorden

Uitkomsten op de open vraag 'Wilt u aanvullende informatie over uw antwoorden of dit onderzoek delen, dan kunt u dat hieronder doen:'

- BIG data is at the moment not necessarily the best key for rapid improvement: there is -in my opinion- a lot still to find in rather SMALL datasets. The same applies to: improving dataquality in the primary processes and systems. There are hardly any questions on improving dataquality and -richness at the basis. The thought 'there is enough data, and the quality is oke' is often wishfull thinking. Getting acces is difficult, data is incomplete, not standardised, has different reference dates, has not the same gradient, is to old, etc etc
- de vragen gaan over een grote range aan onderwerpen, waarvan de kennis en de verantwoordelijkheden verspreid over de organisatie liggen. Door te vragen naar hoe 'your team' hier mee omgaat, wordt de vragenlijst redelijk irrelevant. Aangezien het lang niet altijd over het team van de respondent gaat. Ik heb het gelezen als: jou organisatie.
- Geen
- I'm working for local goverment and i've been part of an informal datateam (datascientist, analist, datamanagers, geo). We seldom use large datasets, instead we use a variations of data to analyse. As a local goverment we intent to work datadriven , but it isn't part of the strategy yet.
- Ik vond de vragen soms best moeilijk te interpreteren.
- In ons werkveld speelt de AVG een belangrijke rol. BIG data is voor onze organisatie niet van toepassing.
- It wasn't always clear what the meaning is of 'we'. Is this my data team? Or the organisation?
- Sinds 1 jaar werkzaam bij mijn huidige werkgever. Datagedreven werken en (big) data analyses staan in de kinderschoenen (heb ik ontdekt). Klein groepje mensen wil graag en sneller. Data geletterdheid is beperkt; we starten volgend jaar met scholingstraject.
- We hebben de state of the art technologie, maar het college/raad willen niet en/of zien de nut niet in om in data gedreven werken te investeren. Bouw projecten gaan de interesse vooral naar toe en geven ze aan dat ze niet kunnen scoren op bijvoorbeeld innovatie, vanwege hun ambt termijn.
- Wij sturen vooral op de definitie dat datagedrevenheid vooral is gestoeld op vragen stellen, nieuwsgierigheid, verbindingen maken en samenwerking, maar niet op blind vertrouwen in data. De bedoeling staat voorop, data en technologie helpen ons maar mogen niet leidend zijn.
- You might want to look into the way the questions are formulated. The frequent use of the world 'always' is annoying. Also the concepts used in the questions are (often) too theoretical and difficult to grasp for most people. Also it seems that the questions are not targeted too local government organisations at all, but often more to large international businesses?

Bijlage 5: Figuren- en tabellenlijst

In tabel b5.1 zijn de gebruikte tabellen en figuren terug te vinden die in dit onderzoek gebruikt zijn. De tabellen in de bijlage zijn hierbij uitgesloten.

Tabel b5.1 Overzicht van gebruikte figuren en tabellen naar paginanummer

	Titel	Te vinden op pagina
Figuren	Figuur 2.1 Processen voor het extraheren van inzichten uit big data	15
	Figuur 2.2 AI-capaciteit en categorisering van middelen	17
	Figuur 2.3 Onderzoeksmodel met de relatie tussen <i>big data analytics capabilities</i> en organisatieprestaties in relatie tot de afstemming van de bedrijfsstrategie	18
	Figuur 2.4 Onderzoeksmodel met de relatie tussen <i>big data analytics capabilities</i> en competitieve prestaties	20
	Figuur 2.5 Conceptueel model voor <i>big data analytics capabilities</i> binnen Nederlandse overheidsorganisaties	21
	Figuur 3.1 Schematische weergave PLS-SEM-analyse in SmartPLS 4 t.b.v. de hypothesen	30
	Figuur 4.1 Gemiddelde categorieën (o.b.v. stellingen) over big data analytics capabilities bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7)	32
	Figuur 4.2 Gemiddelde categorieën (o.b.v. stellingen) over organisatorische prestaties bij Nederlandse overheidsorganisaties en literatuuronderzoek (schaal op 1-7)	33
	Figuur 4.3 Verschil over <i>big data analytics capabilities</i> en organisatorische prestaties tussen Nederlandse overheidsorganisaties en literatuuronderzoek (gemiddelde en standaarddeviatie categorieën (o.b.v. stellingen))	34
	Figuur 4.4 Geschatte relaties BDAC op interne en externe prestaties (<i>structural model</i>)	38
Tabellen	Tabel 1.1 Deelvragen onderzoeksvoorstel	10
	Tabel 2.1 Categorieën (dimensies) van overheidsprestaties	19
	Tabel 3.1 Overzicht bronnen voor totstandkoming vragenlijst	24
	Tabel 3.2 Vervangende stellingen t.b.v. het vragenblok 'menselijke vaardigheden' rondom BDAC	25
	Tabel 3.3 Aantal medewerkers bij Nederlandse overheidsorganisaties in 2019 naar sector	26
	Tabel 3.4 Respondentenverdeling naar organisatiegrootte van Nederlandse overheidsorganisaties in 2021	27

Vervolg tabel b5.1 Overzicht van gebruikte figuren en tabellen naar paginanummer

	Titel	Te vinden op pagina
Tabellen	Tabel 3.5 Aantal jaar ervaring van de respondenten bij Nederlandse overheidsorganisaties met (big) data analytics in 2021	27
	Tabel 4.1 Resultaten betrouwbaarheid en validiteit t.b.v. categorisatie stellingen <i>big data analytics capabilities</i>	35
	Tabel 4.2 Resultaten validatiemaatregelen voor formatieve constructen t.b.v. categorische stellingen <i>big data analytics capabilities</i>	37
	Tabel 4.3 Resultaten PLSpredict van manifeste variabelen (organisatorische prestaties) in relatie tot <i>big data analytics capabilities</i>	39
	Tabel 4.4 Resultaten bootstrapanalyse <i>big data analytics capabilities</i> in relatie tot interne organisatorische prestaties	40
	Tabel 4.5 Resultaten bootstrapanalyse <i>big data analytics capabilities</i> in relatie tot externe organisatorische prestaties	41

The ability to take data – to be able to understand it, to process it, to extract value from it, to visualize it, to communicate it – that's going to be a hugely important skill in the next decades (McKinsey, 2009).

