



# **Berekenwijze Sociaal Economische Status scores**

Koos Arts

Ruben van Gaalen

Jan van der Laan

Frank Linder

Jan Mol

Johan van Rooijen

Clemens Siermann

projectnummer

PR001119  
SQS/SAL/SET  
15-10-2021

# Inhoudsopgave

<b>1.</b>	<b>Inleiding</b>	<b>4</b>
1.1	Aanleiding en achtergrond	4
1.2	Beoogde publicatievariabelen	4
<b>2.</b>	<b>Bronbestanden</b>	<b>7</b>
2.1	Welvaart	7
2.2	Opleidingsniveau	8
2.3	Recent arbeidsverleden	9
<b>3.</b>	<b>Berekenwijze</b>	<b>11</b>
3.1	Deelscore welvaart per huishouden	11
3.2	Deelscore opleidingsniveau per huishouden	11
3.3	Deelscore recent arbeidsverleden per huishouden	12
3.4	Totaalscore per huishouden	14
3.5	Deelscores en totaalscores per wijk/buurt/NL	15
3.6	Ongelijkheidsmaat	16
3.7	Betrouwbaarheidsmarges	16
	<b>Bijlage 1: Samenstelling klankbordgroep</b>	<b>18</b>

# 1. Inleiding

## 1.1 Aanleiding en achtergrond

In het voorjaar van 2021 heeft het ministerie van VWS aan het CBS gevraagd om een project te starten om een breed toepasbare sociaaleconomische statusscore (SES-score) per wijk/buurt te berekenen. Dit om het gat op te vullen dat is ontstaan toen het Sociaal Cultureel Planbureau (SCP) is gestopt met het samenstellen van SES-scores per postcodegebied. SES-scores op laagregionaal niveau worden namelijk door vele partijen voor allerlei doeleinden ingezet. Uiteraard om laagregionaal zicht te krijgen op kwetsbaarheid zijn zodat daar landelijk en regionaal gericht beleid en inzet op ontwikkeld kan worden. Maar SES wordt ook veel gebruikt ook als verklarende variabelen in onderzoeken naar verschillen tussen wijken op het gebied van leefbaarheid, brede welvaart, participatie, welzijn, (sociaal-economische) gezondheid(verschillen), gebruik van gezondheidsvoorzieningen, onderwijs, gedrag, leefstijl. Ook zijn SES-scores in het verleden gebruikt in verdeelmodellen voor budgetten zoals bijvoorbeeld bij de toekenning van extra budgetten aan huisartsen in achterstandswijken en de verdeling van de 'Gezond In de Stad' gelden. De SES-scores van het SCP werden gemiddeld zo'n 60 keer per maand gedownload door een zeer diverse groep gebruikers, waaronder ggd/ggz, scholen, gemeenteambtenaren, onderzoekers van universiteiten en medische centra en kennisinstituten als PBL, RIVM, TNO, Nza, Trimbos en Nivel.

Om te zorgen dat de door het CBS gepubliceerde statusscores breed toepasbaar zullen zijn, heeft het ministerie van VWS aan het CBS gevraagd om een klankbordgroep in te richten met diverse externe experts.

In de eerste fase van het project heeft het CBS, in afstemming met de klankbordgroep, bepaald welke bronbestanden gebruikt gaan worden voor de scores en hoe de berekenwijze er op hoofdlijnen uit komt te zien. In deze rapportage worden de bronbestanden en berekenwijze beschreven<sup>1</sup>.

In de volgende fase van het project zal het CBS de scores daadwerkelijk gaan berekenen en publiceren.

## 1.2 Beoogde publicatievariabelen

Het doel van het project is het samenstellen en publiceren van een breed toepasbare objectieve SES-score per wijk en buurt. Deze breed toepasbare objectieve SES-score geeft aan wat de relatieve status van de wijk of buurt is in termen van de drie gangbare pijlers van een SES/score; te weten inkomen/vermogen, opleidingsniveau en arbeid<sup>2</sup>. Deze SES-score wordt niet gecorrigeerd voor eventuele achtergrondkenmerken die de verschillen tussen wijken, of de ontwikkelingen in de tijd, deels zouden kunnen verklaren. Denk daarbij aan zaken als regionale verschillen in kosten van levensonderhoud of veranderende bevolkingssamenstelling van de wijk. Dit zijn zeer interessante onderwerpen voor nader onderzoek naar statusverschillen, maar horen niet thuis in een feitelijke en eenvoudig te interpreteren SES-score van elke wijk.

---

<sup>1</sup> In de eerste fase van het project heeft het CBS daarnaast een vooronderzoek uitgevoerd naar het gebruik van het opleidingsniveaubestand van het CBS voor de SES-scores. De resultaten van dat vooronderzoek zijn in een aparte rapportage beschreven.

<sup>2</sup> Zie <https://www.volksgezondheidenzorg.info/onderwerp/sociaaleconomische-status/cijfers-context/samenhang-met-gezondheid#definities>

De door mensen ervaren sociale status van een wijk is echter meer dan de optelsom van inkomen, vermogen, opleidingsniveau en arbeidssituatie van deze inwoners. Een wijk is immers meer dan de som der delen en de status van een wijk wordt ook bepaald door aspecten als sociale cohesie, veiligheid, groenvoorzieningen etc. Het verdient dus aanbeveling om naast deze objectieve SES-score ook een breed afgestemde set van aanvullende indicatoren te ontwikkelen om een wijk vollediger te kunnen beschrijven. Het ontwikkelen van zo'n set met aanvullende indicatoren is geen onderdeel van dit project<sup>3</sup>.

In het kader van dit project zullen per wijk en buurt<sup>4</sup>, en voor het totaal van Nederland, de volgende gegevens gepubliceerd worden:

- Deelscore welvaart
  - Zowel in termen van een categorale uitsplitsing in 3 tertielen<sup>5</sup>, een gemiddelde percentielscore én één numerieke deelscore.
  - Daarnaast zal ook een categorale uitsplitsing gegeven worden van alleen inkomen in 3 tertielen en een gemiddelde percentielscore en van alleen vermogen in 3 tertielen en een gemiddelde percentielscore.
- Deelscore opleidingsniveau
  - Zowel in termen van een categorale uitsplitsing in de driedeling hoog/midden/laag als in één numerieke score
- Deelscore recent arbeidsverleden
  - Zowel in termen van een categorale uitsplitsing in de driedeling volledig/gedeeltelijk/niet werkzaam als in één numerieke score.
- Totalscore (numerieke waarde)
- Betrouwbaarheidsmarge van deelscore opleidingsniveau en van totalscore
- Ongelijkheidsmaat voor elk van de 3 numerieke deelscores en voor de totalscore.

Omdat de sociaaleconomische status van een wijk door meer kenmerken wordt bepaald dan alleen welvaart, opleidingsniveau en arbeidsverleden, past de term 'SES-score' minder goed bij de feitelijke inhoud van de nieuwe score. Met name de sociale component is onderbelicht. Daarom wordt voorgesteld om te zoeken naar een alternatieve naam voor de score, die beter aangeeft wat er mee bedoeld wordt. In de klankbordgroep is de term 'WOA-score' (Welvaart, Opleiding en Arbeid) als optie genoemd. In de rest van dit document zal deze term dan ook gehanteerd worden.

De scores van wijken of buurten met te weinig inwoners en/of te grote betrouwbaarheidsmarges zullen niet gepubliceerd worden.

Bij alle scores en categorale uitsplitsingen per wijk/buurt zal aangegeven worden hoeveel unieke huishoudens de betreffende score/categorie hebben én hoeveel unieke personen (huishoudensleden) deze huishoudens in totaal bevatten.

Daarnaast zal er een microdatabestand met de deelscores en totalscore per huishouden worden gemaakt, voor gebruik door externe onderzoekers op de beveiligde microdata-omgeving van CBS.

---

<sup>3</sup> Hierbij kan ook gekeken worden naar bestaande Initiatieven zoals de leefbarometer ([www.leefbarometer.nl](http://www.leefbarometer.nl)) en de regionale monitor brede welvaart (<https://dashboards.cbs.nl/rmbw/regionalemonitorbredewelvaart/>)

<sup>4</sup> En mogelijk ook per 4-cijferig postcodegebied

<sup>5</sup> De keuze voor tertielen is vooral ingegeven door het feit dat een uitsplitsing in decielen of kwintielen op wijkniveau tot te lage aantallen leidt om nog te kunnen publiceren.

In eerste instantie zullen bovenstaande variabelen/bestanden gepubliceerd worden over de jaren 2014 t/m 2019. Daarna kan jaarlijks een nieuw jaar toegevoegd worden, waarbij de hele tijdreeks wordt omgezet naar de wijk/buurt indeling van het toegevoegde jaar. Het is de daarbij de wens dat de eerder gepubliceerde uitkomsten uit de oude tijdreeks beschikbaar zullen blijven<sup>6</sup>.

---

<sup>6</sup> De exacte wijze waarop is nog niet bekend. Om verwarring te voorkomen zal er op StatLine altijd maar één uitkomst per wijk/buurt per jaar staan. Zodra de nieuwe reeks op StatLine komt, zal de oude reeks dus in de andere verschijningsvorm beschikbaar gesteld worden (bv als Excel-bestand op de website van het CBS)

## 2. Bronbestanden

Voor het berekenen van de breed toepasbare WOA-score wordt gebruikt gemaakt van drie microdatabestanden van het CBS: een bestand met informatie over welvaart, een bestand met informatie over opleidingsniveau en een bestand met informatie over het recente arbeidsverleden. Deze drie bestanden worden hieronder kort beschreven.

### 2.1 Welvaart

De welvaartsmaat bevat informatie over zowel het gestandaardiseerd besteedbaar inkomen als het vermogen van huishoudens. De welvaart van een huishouden is vastgesteld als de som van het cumulatieve aandeel in het totale inkomen van alle huishoudens en het cumulatieve aandeel in het totale vermogen. Op grond van de optelling zijn huishoudens opnieuw geordend en vervolgens uitgesmeerd over de welvaarts ladder. Het resultaat hiervan is dat huishoudens in de laagste welvaartsgroep een laag inkomen én een laag vermogen hebben. Naarmate het inkomen of vermogen hoger is, wordt een huishouden in een hogere groep ingedeeld. Huishoudens in de hoogste welvaartsgroep hebben een hoog inkomen én een hoog vermogen. Met de welvaartsmaat wordt een realistischer beeld van de welvaartsverdeling gepresenteerd dan wanneer alleen inkomen in beschouwing wordt genomen.

#### **Gestandaardiseerd besteedbaar inkomen**

Het besteedbare inkomen van een huishouden bestaat uit het bruto-inkomen (inkomen uit arbeid, eigen onderneming en vermogen, uitkeringen en ontvangen overdrachten) verminderd met betaalde overdrachten (bijv. partneralimentatie), premies en belasting op inkomen en vermogen. Het besteedbaar inkomen wordt gecorrigeerd voor verschillen in grootte en samenstelling van het huishouden. Dit gebeurt door het huishoudensinkomen te delen door een equivalentiefactor<sup>7</sup> die uitdrukt hoe groot het schaalvoordeel is bij het voeren van een gemeenschappelijke huishouding, gebaseerd op de uitgaven voor levensonderhoud (bestedingen). Hierbij is het eenpersoonshuishouden als norm gekozen.

#### **Vermogen**

Het vermogen is het saldo van bezittingen en schulden. De bezittingen omvatten bank- en spaartegoeden, effecten, eigen woning en ander onroerend goed, ondernemingsvermogen, aanmerkelijk belang en overige bezittingen. De schulden betreffen de hypotheekschuld eigen woning, studieschulden en overige schulden. De hypotheekschuld is de stand van de schuld waarover rente is verschuldigd. Opgebouwde tegoeden voor de aflossing van de hypotheek via spaar- en beleggingshypotheken zijn voor een deel waargenomen en in mindering gebracht op de hypotheekschuld. Pensioenaanspraken worden niet tot het vermogen gerekend, omdat ze niet vrijelijk beschikbaar en overdraagbaar zijn. Voor het vermogen is geen standaardisatie toegepast voor verschillen in grootte en samenstelling van het huishouden. De correctiefactoren die worden gebruikt voor de standaardisatie van inkomen zijn niet direct toepasbaar bij een eventuele standaardisatie van het vermogen, omdat dan impliciet de veronderstelling wordt gemaakt dat het vermogen door de huidige huishoudensamenstelling wordt opgebruikt voor de bestedingen. Dat kan voor sommige toepassingen een redelijke veronderstelling zijn, maar niet in het algemeen en ook niet voor het hele vermogen.

---

<sup>7</sup> Meer informatie over deze equivalentiefactoren is te vinden op <https://longreads.cbs.nl/materiele-welvaart-in-nederland-2020/bijlagen/>

## Populatie

De populatie voor de inkomens- en vermogensstatistiek bestaat uit (personen in) particuliere huishoudens, waarvan het inkomen is waargenomen. Het gaat om de bevolking van Nederland op 1 januari van het verslagjaar met het inkomen in het gehele verslagjaar en de stand van het vermogen op 1 januari. De bevolking in instellingen, inrichtingen en tehuizen wordt buiten beschouwing gelaten, omdat de besteding van hun inkomen grotendeels vastligt (denk aan de verzorgingsbijdrage van tehuisbewoners). Van circa 100 duizend particuliere huishoudens is het inkomen niet bekend. Dit gaat vaak om migranten, expats en personeel van buitenlandse ondernemingen/instellingen.

## 2.2 Opleidingsniveau

Het Opleidingsniveaubestand bevat data met hoogst behaalde en hoogst gevolgde opleidingsniveau van de Nederlandse bevolking op peilmoment (1 oktober van het verslagjaar). Het bestand is gebaseerd op gegevens uit diverse registers met onderwijsgegevens en de Enquête BeroepsBevolking (EBB). Door het gebruik van meerdere (jaargangen van) bronnen heeft het Opleidingsniveaubestand een zeer hoge dekkingsgraad (2019: ruim 11,9 miljoen personen) die bovendien jaarlijks toeneemt.

Alhoewel de dekkingsgraad hoog is, vertegenwoordigt het bestand niet de gehele doelpopulatie. De informatie is voor een deel van de bevolking integraal en voor de overige personen op steekproefbasis (EBB). De waarneming is bovendien sterk selectief naar leeftijd. Dit komt doordat ouderen veelal hun opleidingen hadden afgerond voordat de registers tot stand kwamen die voor het Opleidingsniveaubestand gebruikt worden. Om informatie te krijgen over het opleidingsniveau van de hele Nederlandse bevolking en om te corrigeren voor selectiviteit, wordt opgehoogd en gewogen. Daarbij wordt onderscheid gemaakt tussen waarnemingen vanuit registers en waarnemingen op basis van de EBB.

De eerste zijn niet gebaseerd op een aselechte steekproef. Daarom mogen zij uitsluitend zichzelf vertegenwoordigen. Met andere woorden, registerwaarnemingen krijgen per definitie een gewicht gelijk aan 1 en worden dus feitelijk uitgesloten van de weging.

Daarentegen zijn de EBB-waarnemingen wel gebaseerd op een aselechte steekproef, zij worden onderworpen aan een genuanceerde weging waarin rekening wordt gehouden met de achtergrondkenmerken leeftijd, geslacht, herkomst, burgerlijke staat, type huishouden, regio, inkomen en inkomensbron.

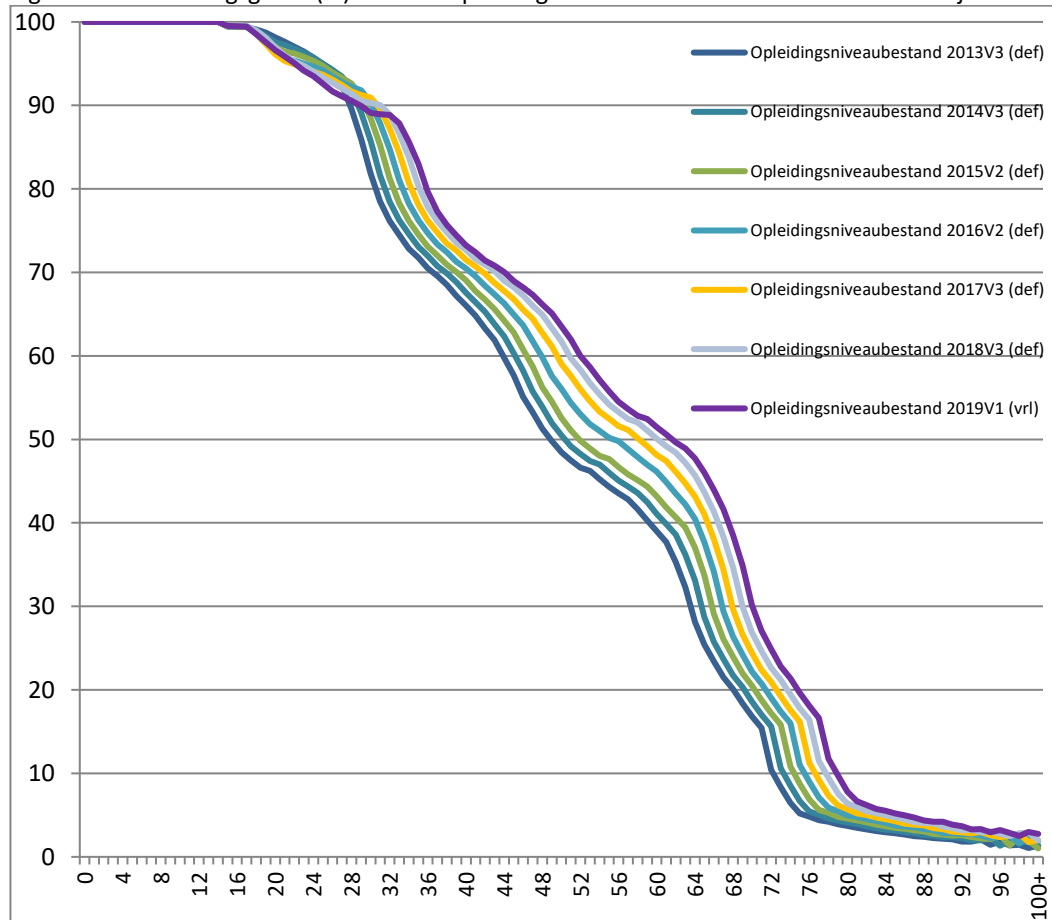
Na weging is de selectiviteit opgeheven en vertegenwoordigt het Opleidingsniveaubestand de gehele Nederlandse bevolking. Er is echter nog sprake van een kleine vertekening in de waarneming van het opleidingsniveau. Via de registers die gebruikt worden voor het Opleidingsniveaubestand wordt al het door overheid bekostigd onderwijs waargenomen maar onderwijs dat wordt gevolgd aan particuliere instellingen of in het buitenland wordt niet volledig door deze registers waargenomen. De EBB neemt (op steekproefbasis) opleidingen in het buitenland en particulier onderwijs wel waar, en compenseert daarvoor, zij het niet volledig. Per saldo onderschat het Opleidingsniveaubestand dus in een klein deel van de gevallen het opleidingsniveau. Hiervoor wordt gecorrigeerd door te imputeren. Daartoe wordt op basis van registerwaarnemingen en EBB-waarnemingen een model geschat dat vervolgens wordt gebruikt om in een klein deel van de records het opleidingsniveau naar boven toe bij te stellen.



Het Opleidingsniveaubestand is deels gebaseerd op steekproefwaarneming. Om die reden is er sprake van een steekproeffout. Die steekproeffout kan sterk variëren al naar gelang de deelpopulatie die in ogenschouw wordt genomen. Een deelpopulatie die uit veel ouderen bestaat zal gepaard gaan met een verhoudingsgewijs grote steekproeffout omdat de waarneming onder ouderen sterker leunt op steekproefwaarneming.

De dekkinggraad van het Opleidingsniveaubestand neemt ieder jaar toe omdat er steeds meer waarnemingen bijkomen. Voor oplopende leeftijden wordt de dekking lager, maar waar de dekkinggraad van bijvoorbeeld 60-jarigen in 2013 41% bedroeg is dat al opgelopen tot 51% in 2019.

Figuur 2.2.1. Dekkingsgraad (%) van het Opleidingsniveaubestand 2013-2019 naar leeftijd



### 2.3 Recent arbeidsverleden

Het recente arbeidsverleden geeft aan in welke mate iemand langdurig deelneemt aan het arbeidsproces. De aanname is dat als iemand langere tijd ononderbroken betaald werk heeft, dat die persoon een hogere sociaaleconomische status heeft dan als iemand niet of een onderbroken arbeidsverleden heeft. Langdurige arbeidsdeelname geeft niet alleen aan dat iemand participeert, maar ook dat dit op langere termijn succesvol is. Personen die langdurig aan het werk zijn, hebben blijkbaar competenties die nodig zijn om aan het werk te blijven.

Voor het recente arbeidsverleden wordt voor elke maand in de afgelopen 4 kalenderjaren

bekeken of iemand minimaal 1 uur in de week werkzaam was. Het werk hoeft daarbij niet de voornaamste inkomensbron te zijn, al is dat meestal wel het geval. De populatie bestaat uit alle personen van 15 jaar en ouder die op 31 december van het verslagjaar deel uit maken van de Nederlandse bevolking. Deze populatie bevat ook de zelfstandigen. Van de zelfstandigen is de arbeidssituatie echter alleen per kalenderjaar bekend, waar het voor de werknemers per maand bekend is. Voor zelfstandigen worden de verdiensten gelijkelijk over alle kalendermaanden van dat jaar verdeeld.

In het standaard microdatabestand van het CBS over recent arbeidsverleden worden de volgende categorieën onderscheiden<sup>8</sup>:

1. Voortdurend werkzaam  
(werkzaam in laatste maand en minimaal 46 van 47 maanden daarvoor)
2. Meer dan 3 jaar werkzaam maar niet voortdurend in afgelopen 4 jaar
3. Meer dan 2 jaar werkzaam in afgelopen 4 jaar
4. Meer dan 1 jaar werkzaam in afgelopen 4 jaar
5. Minimaal 1 maand tot en met maximaal 1 jaar werkzaam in afgelopen 4 jaar
6. Niet werkzaam in afgelopen 4 jaar

In de categorie 'voortdurend werkzaam' is de eis opgenomen dat de persoon werkzaam moet zijn in de laatste maand (de maand december van het verslagjaar). Zonder deze eis zou het begrip 'voortdurend' verwateren, omdat dan ook personen in deze groep vallen waarvan later in de tijd zal blijken dat die laatste maand de start was van een kortere of langere periode waarin ze niet werkzaam zijn geweest. In 2019 was de verdeling over deze categorieën als volgt:

Arbeidsverleden	Percentage van bevolking van 15 jaar en ouder	Percentage van bevolking van 25-65 jaar	Percentage van mannen van 25-65 jaar
Voortdurend werkzaam	45 %	65%	71%
Gedeeltelijk werkzaam - meer dan 3 jaar	9 %	10%	9%
Gedeeltelijk werkzaam - tussen 2 en 3 jaar	6 %	5%	5%
Gedeeltelijk werkzaam - tussen 1 en 2 jaar	5 %	4%	3%
Gedeeltelijk werkzaam - tussen 1 maand en 1 jaar	6 %	3%	3%
Niet werkzaam	28 %	13%	9%

<sup>8</sup> De categorieën zijn hier voor de eenvoud alleen omschreven in termen van personen die 48 maanden tot de Nederlandse bevolking behoorden. Bij personen waar dat niet het geval is, worden de categorieën naar rato vastgesteld.

### 3. Berekenwijze

De scores per wijk en buurt in jaar T worden berekend op basis van de scores van de niet-institutionele huishoudens (de particuliere huishoudens) in de wijk/buurt op 1 januari van jaar T. Van circa 100 duizend particuliere huishoudens is het inkomen en vermogen niet bekend (en meestal het arbeidsverleden en opleidingsniveau ook niet), deze huishoudens worden niet meegenomen in de berekeningen. Net zoals bij de reguliere statistieken van het CBS over inkomen en vermogen wordt voor deze ontbrekende huishoudens niet gecorrigeerd.

Omdat de huishoudenscores worden berekend op basis van gegevens op persoonsniveau, wordt daarbij ook uitgegaan van de samenstelling van het huishouden op 1 januari van jaar T.

#### 3.1 Deelscore welvaart per huishouden

De categorale welvaartscore van een huishouden in jaar T (als input voor de WOA-score) is het landelijke welvaartsdeciel waar het huishouden toe behoort in jaar T. Dit welvaartsdeciel wordt vastgesteld op basis van de welvaartspercentielen in de reguliere welvaart-statistieken van het CBS.

De categorale inkomensscore van een huishouden in jaar T is het inkomensdeciel waar het huishouden toe behoort in jaar T.

De categorale vermogensscore van een huishouden in jaar T is het vermogensdeciel waar het huishouden toe behoort in jaar T.

#### 3.2 Deelscore opleidingsniveau per huishouden

Voor het opleidingsniveau van het huishouden wordt gekeken naar het hoogst behaalde opleidingsniveau van de hoofdkostwinner en diens eventuele partner. Het meenemen van het hoogst behaalde opleidingsniveau van de eventuele partner is geen triviale zaak. Het opleidingsniveau is namelijk niet van alle Nederlanders bekend; er moet dan ook altijd gewerkt worden met een weging om te komen tot representatieve schattingen van het opleidingsniveau in een wijk of buurt. Omdat de bestaande gewichten op persoonsniveau berekend zijn, zijn deze niet a priori bruikbaar voor representatieve uitkomsten op basis van scores per huishouden. Complicerende factor daarbij is het feit dat van sommige huishoudens alleen het opleidingsniveau van de hoofdkostwinner bekend is en in andere huishoudens alleen dat van diens partner (zie onderstaande tabel).

Tabel 3.2.1. Aantal huishoudens waarvan opleidingsniveau bekend is.

Huishoudenssamenstelling	Aantal huishoudens in 2019 (in mln.)	Aantal huishoudens in 2019 waarvan opleidingsniveau bekend is (in mln.)
Alleen hoofdkostwinner	3,6	2,2
Hoofdkostwinner + partner	4,2	1,8 van beiden bekend 0,5 alleen van hoofdkostwinner bekend 0,6 alleen van partner bekend

Van de huishoudens waar het opleidingsniveau van zowel hoofdkostwinner als partner bekend zijn, is bij een 3-deling van het opleidingsniveau in hoog-midden-laag in 59% van de huishoudens het opleidingsniveau gelijk, in 26% heeft de hoofdkostwinner een hoger niveau en in 15% heeft de partner een hoger niveau.

In de eerste fase heeft het CBS al een kort vooronderzoek gedaan naar het gebruik van het opleidingsniveaubestand bij het berekenen van een WOA-score. De oplossingsrichting wordt vooralsnog gezocht in de methode van multiple imputatie van alle onbekende opleidingsniveaus, waarbij de imputatiemethodiek rekening houdt met een aantal randtotalen en verhoudingen (zoals onder ander de verhouding tussen het opleidingsniveau van hoofdkostwinner en partner in huishoudens waar deze wél allebei bekend zijn). Een groot voordeel van imputatie boven weging is dat op die manier voor de WOA-score de welvaart en arbeidsgeschiedenis van de ongeveer 40% huishoudens waarvan het opleidingsniveau niet bekend is, toch meegenomen kan worden. Een ander bijkomend voordeel van multiple imputatie ten opzichte van weging is dat vervolgens de berekening van betrouwbaarheidsmarges van de totaal- en deelscores eenvoudiger te implementeren is<sup>9</sup>. Bij andere projecten in het verleden is bovendien gebleken dat imputatie of weging in de praktijk hetzelfde resultaat oplevert (uiteraard onder de voorwaarde dat in het imputatiemodel en in het weegmodel rekening wordt gehouden met dezelfde set van randtotalen).

Bij het toekennen van een score aan het huishouden voor het hoogst behaalde opleidingsniveau wordt gebruik gemaakt van de betreffende indeling in de Standaard Onderwijs Indeling (SOI) 2016, zie de tabel hieronder.

Tabel 3.2.2. Onderwijsniveau indeling in SOI 2021

1 <sup>e</sup> digit	2 <sup>e</sup> digit	3 <sup>e</sup> digit
1 Laag	11 Basisonderwijs	111 Basisonderwijs
	12 Vmbo, avo onderbouw, mbo1	121 Vmbo-b/k, mbo1
		122 Vmbo-g/t, avo onderbouw
2 Midden	21 Havo, vwo, mbo	211 Mbo2 en mbo3
		212 Mbo4
		213 Havo, vwo
3 Hoog	31 Hbo-, wo-bachelor	311 Hbo-, wo-bachelor
	32 Hbo-, wo-master, doctor	312 Hbo-, wo-master, doctor

De categorale score van het huishouden (als input voor de WOA-scores) wordt bepaald door het maximum te nemen van de score van de hoofdkostwinner en diens partner op 2<sup>e</sup> digit van de SOI-indeling.

### 3.3 Deelscore recent arbeidsverleden per huishouden

De verreweg eenvoudigste manier om het recente arbeidsverleden van het huishouden te bepalen is om uit te gaan van het arbeidsverleden van de hoofdkostwinner van het huishouden. In de klankbordgroep is de wens uitgesproken om na te gaan of het mogelijk en zinvol zou zijn om dit uit te breiden met het arbeidsverleden van de eventuele partner van deze hoofdkostwinner. Hierbij zijn in theorie twee varianten te onderscheiden:

<sup>9</sup> Zeker gezien de beperkte capaciteit voor de 2<sup>e</sup> fase, is dit een niet triviaal voordeel.

- Neem het maximum van de score van de hoofdkostwinner en de score van diens eventuele partner (waarbij volledig werkzaam gezien wordt als de hoogste score en niet werkzaam als de laagste score)
- Bereken een 'gecombineerde' score van zowel de score van de hoofdkostwinner én diens partner ('beiden volledig werkzaam', 'één volledig werkzaam en de ander gedeeltelijk werkzaam' etc.).

De tweede variant is conceptueel lastiger. Hoe moet zo'n gecombineerde score vergeleken worden met de score van een huishouden met alleen een hoofdkostwinner? Hoe verhoudt de status van een huishouden met een volledig werkzame hoofdkostwinner en een niet werkzame partner zich tot de status van een huishouden met twee gedeeltelijk werkzame partners? Vanwege deze conceptuele uitdagingen, in combinatie met de beperkte capaciteit in de 2<sup>e</sup> fase, zal het CBS in de 2<sup>e</sup> fase dan ook alleen de eerste variant nader onderzoeken op haalbaarheid en zinvolheid.

De categorale score van het huishouden (als input voor de WOA-score) is de score volgens de onderstaande indeling in zes categorieën

- Voortdurend werkzaam
- Gedeeltelijk werkzaam - meer dan 3 jaar
- Gedeeltelijk werkzaam - tussen 2 en 3 jaar
- Gedeeltelijk werkzaam - tussen 1 en 2 jaar
- Gedeeltelijk werkzaam - tussen 1 maand en 1 jaar
- Niet werkzaam

Twee groepen personen verdienen mogelijk extra aandacht bij de berekening van een WOA-deelscore voor het recente arbeidsverleden:

- Gepensioneerden, die per definitie niet meer (hoeven te) werken.  
In de 2<sup>e</sup> fase zullen de volgende varianten worden bekeken:
  - Deze groep wordt niet anders behandeld dan alle andere personen in Nederland (dus na hun pensioen krijgen ze 3 jaar lang de score 'gedeeltelijk werkzaam' en vervolgens de rest van hun leven de score 'niet werkzaam'). Deze keuze reflecteert het idee dat een persoon na zijn pensioen (iets van) zijn status verliest op de arbeidsdimensie, omdat hij het netwerk en steunbronnen vanuit zijn werkomgeving verliest.
  - Deze groep personen krijgt voor de berekening van de WOA-score per definitie de score 'volledig werkzaam'. Deze keuze reflecteert het idee dat het bij de deelscore over arbeid vooral gaat om mensen die niet werken maar wel tot de werkzame beroepsbevolking behoren. Personen die met pensioen zijn, hoeven niet meer te werken en het zou onwenselijk zijn als zij de WOA-score van de wijk zouden verlagen.
  - Deze groep personen wordt toegekend aan een nieuwe 4<sup>e</sup> categorie (naast volledig/gedeeltelijk/niet werkzaam), namelijk 'gepensioneerd'. In deze keuze wordt het aan de MCA-techniek overgelaten om de gepensioneerden te 'waarderen' ten opzichte van de mensen die volledig/gedeeltelijk/niet werkzaam zijn.
- Jongeren die nog geen vier jaar werkzaam kunnen zijn geweest.  
Omdat het in de praktijk soms lastig is om vast te stellen vanaf welk moment een jongere werkzaam had kunnen zijn, wordt hier bij de berekening van de WOA-score geen rekening mee gehouden. Een jongere van 19 die, sinds hij op zijn 17<sup>e</sup> van school kwam, al de volledige twee jaar werkt, wordt als gedeeltelijk werkzaam in de afgelopen

vier jaar gezien. In vergelijking met een oudere werknemer die al wel een arbeidsverleden van (minstens) vier jaar heeft, is de stabiliteit van de arbeidsrelatie aan het begin van de loopbaan ook wat minder zeker, wat tot uiting mag komen in een 'lagere' deelscore op arbeidsverleden.

De klankbordgroep heeft twee andere suggesties gedaan voor mogelijke uitbreiding/verfijning van de deelscore over arbeid. Dit ging om:

- Verruiming van categorie 'niet werkzaam' naar 'niet of nauwelijks werkzaam' en verruiming van de categorie 'volledig werkzaam' naar '(bijna) volledig werkzaam'.
- Rekening houden met het aantal onderbrekingen in de afgelopen 4 jaar.

Omwille van de volgende praktische overwegingen, worden deze twee onderwerpen niet verder uitgewerkt in de 2<sup>e</sup> fase:

- Van de zelfstandigen is alleen op jaarbasis bekend of ze werkzaam waren; dit is niet verder uit te splitsen naar de maanden in een jaar. Beide suggesties zijn voor de zelfstandigen dan ook niet met dezelfde betrouwbaarheid uit te voeren dan voor werknemers. Omdat de zelfstandigen een substantiële groep zijn (in 2020 17% van de werkzame beroepsbevolking), zouden deze uitbreidingen de representativiteit van de deelscore verslechteren.
- Met name het 2<sup>e</sup> punt verhoogt de complexiteit van berekenwijze van de deelscore op arbeid aanzienlijk, waar vermoedelijk onvoldoende tijd voor zal zijn in de 2<sup>e</sup> fase.

### 3.4 Totaalscore per huishouden

De totaalscore per huishouden wordt berekend met behulp van een Multiple Correspondentie Analyse (MCA).

Multiple Correspondentie Analyse (MCA) is een methode die de onderliggende samenhang tussen categoriale variabelen modelleert. Het is in feite een soort principale componenten analyse (PCA), maar dan voor categoriale variabelen. Waarnemingen worden beschreven met behulp van een zogenaamde indicator matrix. Hierin is iedere rij een huishouden en de matrix heeft voor iedere categorie voor elk van de 3 deelvariabelen (welvaart, opleiding, arbeid) een dummyvariabele die aangeeft of het huishouden een waarde heeft gelijk aan die categorie.

Input voor de MCA is voor alle huishoudens in Nederland:

- Het landelijke welvaartsdecil waar het huishouden in valt (10 categorieën)
- De categorale opleidingsniveauscore van het huishouden volgens de **2<sup>e</sup> digit** van de SOI-indeling (5 categorieën)
- De categorie voor recent arbeidsverleden waar het huishouden in valt (in 6 of mogelijk 7 categorieën)

Elke huishouden is te zien als een punt in een hoog-dimensionale ruimte. Deze punten worden op een lager dimensionale ruimte geprojecteerd net zoals dat bij PCA gebeurt. De eerste dimensie bevat hierbij het grootste deel van de samenhang tussen de dummyvariabelen en kan dus worden gezien als de gewenste WOA-indicator.

De MCA-methode levert naast een totaalscore per huishouden ook numerieke scores per huishouden voor welvaart, opleidingsniveau en arbeidsverleden. Dit is de projectie van het huishouden op de 'welvaarts-dimensie', 'opleidingsniveau-dimensie' en arbeidsverleden-dimensie' in de multidimensionale ruimte van de MCA.

De totaalscore van het huishouden is daarbij het gemiddelde van de 3 numerieke deelscores.

In de 2<sup>e</sup> fase zal nog nader bekeken worden of het gewenst en uitvoerbaar is als de scores in de tijd ‘meebewegen’ als Nederland gemiddeld welvarender of hoger opgeleid of werkzamer wordt. De belangrijkste doelen van de scores zijn om in een bepaald jaar wijken en buurten onderling te kunnen vergelijken met elkaar én om door de jaren heen de relatieve positie van één wijk/buurt te kunnen volgen. Een kleine complicerende factor is het feit dat in de beoogde methodiek van drie deelscores niet allemaal dezelfde aard hebben en de totaalscore dus al een wat hybride karakter heeft op dit punt. De deelscore over welvaart is op zichzelf namelijk al een relatieve score ten opzichte van alle andere huishoudens. De gemiddelde percentielscore voor heel Nederland is elk jaar per definitie 50,5 en zal dus niet meebewegen als Nederland welvarender wordt. Maar de inputdata over opleidingsniveau en arbeidsverleden bewegen wél mee als Nederland gemiddeld hoger opgeleid of werkzamer wordt. Op hoofdlijnen zijn er de volgende opties:

1. Voeg alle jaren samen en voer één MCA-analyse uit over alle jaren.
2. Voer de MCA analyse uit in 1 jaar (het basisjaar) en houdt de berekende MCA-factoren vast in alle andere jaren.
3. Voer de MCA voor elk jaar afzonderlijk uit. Het gemiddelde van de totaalscore is dan elk jaar per definitie 0.

### 3.5 Deelscores en totaalscores per wijk/buurt/NL

De categorale uitsplitsing van het welvaart, inkomen en vermogen van de huishoudens in een wijk is het aantal huishoudens in de landelijke welvaarts- c.q. inkomens- c.q. vermogenstertielen<sup>10</sup>.

De categorale uitsplitsing van het opleidingsniveau van de huishoudens in een wijk is het aantal huishoudens in elke van de groepen op 1<sup>e</sup> **digit** van de SOI-indeling (te weten: hoog, midden, laag).

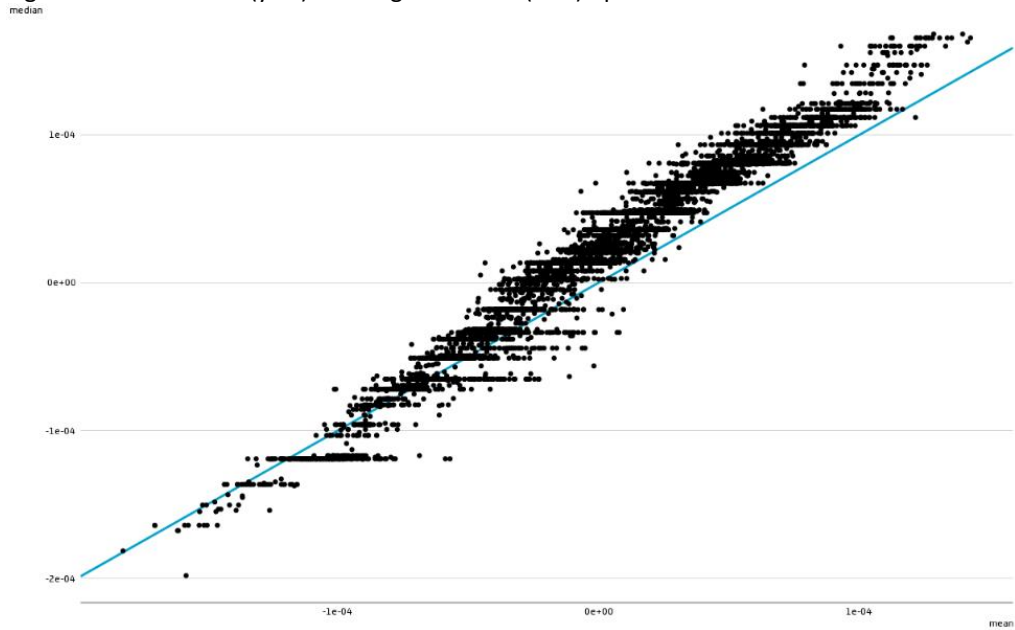
De categorale uitsplitsing van het arbeidsverleden van de huishoudens in een wijk is het aantal huishoudens in elk van de volgende groepen: volledig werkzaam, gedeeltelijk werkzaam, niet werkzaam (mogelijk nog uitgebreid met een vierde groep ‘gepensioneerd’).

De drie numerieke deelscores (voor welvaart, opleidingsniveau en arbeidsverleden) van een wijk/buurt/NL én de numerieke totaalscore van een wijk/buurt/NL worden berekend als het gemiddelde van de betreffende numerieke scores van de huishoudens in de wijk/buurt. Een alternatief voor het gemiddelde zou de mediaan kunnen zijn. Echter, een snelle en daardoor ook op onderdelen beperkte analyse op gegevens over welvaart, opleidingsniveau en arbeidsverleden in 2013, laat zien dat op buurtniveau het gemiddelde en de mediaan van de totaalscores op huishoudniveau doorgaans dicht bij elkaar liggen (zie de onderstaande figuur). Het gemiddelde heeft in dat geval de voorkeur, omdat daardoor ook op wijk/buurt/NL-niveau de totaalscore het gemiddelde is van de drie deelscores.

---

<sup>10</sup> De keuze voor tertielen is vooral ingegeven door het feit dat een uitsplitsing in decielen of kwintielen op wijkniveau tot te lage aantallen leidt om nog te kunnen publiceren.

Figuur 3.5.1. Mediaan (y-as) versus gemiddelde (x-as) op buurtniveau.



Voor wijken/buurtten met minder dan 100 huishoudens, worden de scores niet gepubliceerd.

### 3.6 Ongelijkheidsmaat

Voor elke wijk en buurt wordt het aantal huishoudens uitgesplitst naar welvaart (in 3 tertielen), opleidingsniveau (in de driedeling hoog-midden-laag) en arbeidsverleden (in de driedeling volledig/gedeeltelijk/niet werkzaam). Dit geeft al een eerste beeld van de ongelijkheid binnen de wijk. Een gedetailleerd beeld wordt verkregen door voor elk van de 3 deelscores en voor de totaalscore de ongelijkheid in de numerieke scores per huishouden te publiceren. Als maatstaf is het gebruik van een Gini-coëfficiënt te overwegen. De Gini-coëfficiënt is een maatstaf van de ongelijkheid van een verdeling. De Gini-coëfficiënt wordt vaak gebruikt om de ongelijkheid in inkomen en vermogen aan te geven, maar is geschikt voor het in beeld brengen van de ongelijkheid in elke verdeling. Doorgaans heeft de Gini-coëfficiënt een waarde tussen nul en één. Bij de waarde nul hebben alle huishoudens exact dezelfde WOA-score, bij de waarde één is de ongelijkheid maximaal. In de 2<sup>e</sup> fase zal nog kort geïnventariseerd worden of mogelijke andere maatstaven zinvoller zouden zijn dan een Gini-coëfficiënt.

### 3.7 Betrouwbaarheidsmarges

Omdat van een aanzienlijk deel van de Nederlandse bevolking het opleidingsniveau niet bekend is, zijn een aantal uitkomsten omgeven met betrouwbaarheidsmarges. Dit betreft de uitsplitsing van het aantal huishoudens in de wijk/buurt naar hoog/midden/laag opleidingsniveau, de WOA-deelscore voor opleidingsniveau, de WOA-totaalscore, de Gini-coëfficiënt voor de deelscore op opleidingsniveau en de Gini-coëfficiënt voor de WOA-totaalscore.

Zoals in paragraaf 3.2 is beschreven, zullen de ontbrekende opleidingsniveaus met multiple imputatie worden berekend. Dat wil zeggen dat de imputatieslag niet één keer, maar meerdere keren wordt uitgevoerd. Meestal in de orde van enkele tientallen keren. Uit elke imputatie-iteratie kunnen vervolgens alle uitkomsten uitgerekend worden en uit de spreiding tussen de verschillende imputatie-iteraties kunnen de betrouwbaarheidsintervallen van de uitkomsten bepaald worden.





## Bijlage 1: Samenstelling klankbordgroep

Het CBS is grote dank verschuldigd aan de leden van de klankbordgroep. De klankbordgroep heeft meegedacht met het CBS over de inhoudelijke keuzes in de berekenwijze van de scores om zo te borgen dat de WOA-scores zo breed mogelijk toepasbaar zijn.

De klankbordgroep bestond uit de volgende personen:

Aldien Poll (Pharos)

Annemarie Ruijsbroek (RIVM)

Harry ten Caten (gemeente Eindhoven en Vereniging Statistiek en Onderzoek)

Hidde Boonstra (VNG Realisatie)

Jan Mol (CBS)

Jeroen Boelhouwer (SCP)

Joline v.d. Rijt (Ministerie van VWS)

Laurens Zwakhals (RIVM)

Nannah Tak (GGD-GHOR)

Robbert van Bokhoven (Pharos)

Ruben van Gaalen (CBS/UvA)

Tanja Houweling (Erasmus MC)

Wim van der Zanden (gemeente Rotterdam en Vereniging Statistiek en Onderzoek)

De klankbordgroep is drie keer bij elkaar gekomen, op 17 juni, 15 juli en 9 september.