

Statistische Analyse Toegankelijkheid Stemlokalen Tweede Kamerverkiezing

Eindrapport



Datum	25 maart 2024
Auteurs	Dr. Karlijn Roex Pradeep Kumar, MSc. Drs. Marcia den Uijl
Versie	1.0

Uitgave

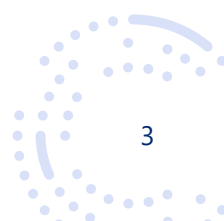
Centerdata
info@centerdata.nl
www.centerdata.nl

Contact

Karlijn Roex
karlijn.roex@centerdata.nl

© Centerdata, Tilburg, 2024

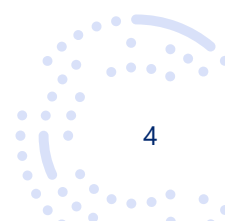
Alle rechten voorbehouden. Niets uit deze uitgave mag worden verveelvoudigd, opgeslagen in een geautomatiseerd gegevensbestand, of openbaar gemaakt, in enige vorm of op enige wijze, hetzij elektronisch, mechanisch, door fotokopieën, opnamen of op enige andere manier, zonder voorafgaande schriftelijke toestemming van de uitgever.





Inhoudsopgave

Samenvatting	5
1 Inleiding	8
1.1 Aanleiding en doel	8
1.2 Onderzoeksvragen	9
2 Data en methode	10
2.1 De datasets	10
2.2 Methodes	10
3 Statistische analyses	15
3.1 Verschillen tussen verkiezingen van maart en november 2023	15
3.2 Verschillen tussen provincies	19
3.3 Verschillen tussen gemeentegroottes	23
3.4 Verschillen in beoordeling tussen controleurs	26
3.5 Statistische verbanden data Ongehinderd	29
4 Steekproeftrekking	33
4.1 Het algoritme	33
4.2 De steekproeftrekking	34
Bijlagen	36
A Toegankelijkheidsscore 2 – Beschrijvende statistiek	36
B Toegankelijkheidsscore 2 – Scores per verkiezing	37
C Toegankelijkheidsscore 2 – Scorevariatie per provincie	38
D Toegankelijkheidsscore 2 – Beschrijvende statistiek van de drie gemeentegroottes	39
E Toegankelijkheidsscore 2 – Scorevariatie van de drie gemeentegroottes	40
F Toegankelijkheidsscore 2 – Scorevariatie van de controleurs	41
G Toegankelijkheidsscore 2 – Correlatieplots	42





Samenvatting

Tijdens de Tweede Kamerverkiezing van 22 november 2023 stemde 77,7% van de kiesgerechtigde bevolking in stemlokalen verspreid door het hele land. Volgens de Kieswet moeten alle stemlokalen toegankelijk zijn voor mensen met een beperking. De toegankelijkheidssituatie in en rondom stemlokalen wordt beoordeeld via steekproefonderzoeken door Ongehinderd. In maart 2023 is op verzoek van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) een adviesrapport uitgebracht door Prof. Ernst Wit over het trekken van steekproeven en het daarbij waarborgen van representativiteit.¹ Om de representativiteit van de steekproeven en de subjectiviteit van de beoordelingen te toetsen, is Centerdata in opdracht van BZK in de zomer van 2023 een onderzoek gestart naar de gemeenteraadsverkiezing uit 2022 en de Provinciale Staten- en Waterschapsverkiezing van maart 2023. Vervolgens verrichtte Centerdata begin 2024 een nieuw onderzoek, waarvan in dit huidige rapport verslag wordt gedaan.

In dit rapport buigen we ons opnieuw over de zes onderzoeksvragen die in het voorgaande onderzoek zijn gesteld, maar ditmaal kijken we naar de uitkomsten van de steekproefonderzoeken van de Tweede Kamerverkiezing van november 2023. Tevens wordt nagegaan hoe deze uitkomsten zich verhouden tot de uitkomsten van de Provinciale Staten- en Waterschapsverkiezing in maart 2023. Vanaf nu refereren we naar deze verkiezing in maart 2023 als 'PSWS verkiezing'. De samenvattende uitkomsten van de statistische analyses zijn hieronder uitgelicht.

Verschillen tussen de twee verkiezingen

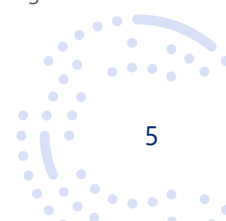
- Er is een statistisch significante toename gevonden in de gemiddelde toegankelijkheidsscore van de stemlokalen in de steekproefonderzoeken van de verkiezingen van maart en november 2023.
- In tegenstelling tot wat het geval was bij onze voorgaande rapportage,² waarin we de PSWS verkiezing vergeleken met de gemeenteraadsverkiezing in 2022, was er ditmaal sprake van eenzelfde soort steekproeftrekking tussen de twee onderzochte verkiezingen. Zowel bij de PSWS verkiezing als de Tweede Kamerverkiezing zijn er steeds 120 stemlokalen beoordeeld. In beide steekproeftrekkingen zijn de adviezen van prof. Ernst Wit opgevolgd om representativiteit te waarborgen.

Verschillen tussen provincies

- Er zijn duidelijke statistisch significante verschillen gevonden tussen de provincies in de gemiddelde toegankelijkheidsscores.
- Deze significante verschillen worden niet volledig verklaard door de gegroeide steekproeven van Friesland en Drenthe. Als de twee provincies worden weggelaten in de analyses blijft het beeld intact.

¹ Statistisch advies voor de steekproef methodologie met betrekking tot de toegankelijkheid van stemlokalen; Prof. Dr. Ernst C. Wit. Università della Svizzera italiana, 5 maart 2023.

² Höcük, S., Kumar, P. & Guimarães, J. (2023). Statistische Analyse Toegankelijkheid Stemlokalen. Eindrapport. Tilburg: Centerdata.





Verschillen tussen gemeentes

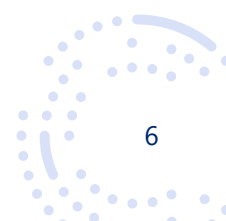
- Er zijn tevens statistisch significante verschillen gevonden in de gemiddelde toegankelijkheidsscores van kleine (inwonersaantal < 40.000), middelgrote (inwonersaantal 40.000 t/m 150.000), en grote (inwonersaantal > 150.000) gemeentes. Over het algemeen werden stemlokalen in kleine gemeentes positiever beoordeeld dan stemlokalen elders.

Verschillen tussen controleurs

- Er zijn statistisch significante verschillen gevonden tussen de controleurs in hun gemiddelde beoordelingen, maar enkel voor de Tweede Kamerverkiezing. Voor de PSWS verkiezing konden geen significante verschillen worden aangetoond.
- Bij al deze analyses geldt de kanttekening dat het gemiddeld aantal bezochte stemlokalen per controleur slechts 7,5 is. Ook bij de Tweede Kamerverkiezing is het nog steeds voorgekomen dat een aantal inspecteurs minder dan vijf stemlokalen beoordeeld heeft. Daardoor zijn de vergelijkingen tussen controleurs onzeker.
- Zoals ook in onze voorgaande rapportage is aangestipt, kunnen verschillen in de toegankelijkheidsscores tussen controleurs mogelijk zijn veroorzaakt door onderlinge verschillen tussen gemeentes. Iedere gemeente is immers door een andere controleur beoordeeld. Naar deze gemeentever verschillen kan helaas niet gekeken worden met de huidige steekproefomvang. Het aantal beoordeelde stemlokalen per gemeente is dan immers te laag (gemiddeld 3 per gemeente).

Verbanden tussen toetsingsonderdelen en eindbeoordeling

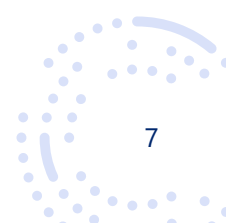
- Net als bij onze eerdere analyse op de data van de gemeenteraadsverkiezing en PSWS verkiezing troffen we ditmaal wederom géén sterke correlaties (onderlinge samenhang) aan tussen de verscheidene toetsingsonderdelen van toegankelijkheid. De onderdelen zijn dus onafhankelijk van elkaar.
- Er zijn ook géén sterke correlaties gevonden tussen de individuele beoordelingen van de 14 toetsingsonderdelen en de eindwaardering. Hoewel er uiteraard enige samenhang was tussen de onderdelen en de eindwaardering (de eindwaardering is immers per definitie afhankelijk van de beoordeling op de toetsingsonderdelen), is het niet zo dat de eindwaardering (sterk) afhangt van één bepaald toetsingsonderdeel.
- De gemiddelde toegankelijkheidsscore van enkele toetsingsonderdelen is gestegen gedurende 2023: de openbare weg (1.1), parkeergelegenheid (1.2), route entree (1.3), de route stembureau (2.1) en de leesloep (3.7). Daarentegen zijn de tussendeuren (2.2), verkeersruimte (3.2), kandidatenlijst (3.6) en juist gedaald in gemiddelde toegankelijkheidsscore. De variatie van de scores (ook wel aangeduid als de 'spreiding', 'variantie' of de 'standaarddeviatie') is constant gebleven voor het merendeel van de toetsingsonderdelen. Als we het hebben over de variatie of verdeling van de scores dan kijken we naar de mate waarin scores onderling verschillen: liggen de scores ver verspreid van elkaar over de gehele scoreschaal, of zien we juist één specifieke score bij veel stemlokalen?
- De daling in de score voor toetsingsonderdeel 2.2 (tussendeuren) kan mogelijk worden verklaard door het samenvoegen van toetsingsonderdelen voor dit onderdeel bij de Tweede Kamerverkiezing.
- Deze variatie in de scores nam iets af voor toetsingsonderdeel 3.7 (leesloep), terwijl deze juist sterk was gestegen in onze voorgaande rapportage. Terwijl de variatie 0,50 bedroeg in maart





2023, was deze nog maar 0,36 bij de verkiezing van november 2023. Ook daalde de variatie van scores voor toetsingsonderdeel 3.5 (stembus), namelijk van 0,18 naar 0,00 en van 0,58 naar 0,43 voor toetsingsonderdeel 2.1 (route stembureau).

Naast de statistische analyses is er opnieuw een steekproeftrekking van 120 stemlokalen gedaan met een goede afspiegeling van de landelijke situatie voor de Europese Parlementsverkiezing van 2024. De in juli 2023 aan het ministerie overgedragen steekproeflijst aan gemeentes is tussentijds gebruikt als steekproef voor de Tweede Kamerverkiezing van 2023. Omdat de wens bestaat om niet dezelfde steekproef te benutten voor achtereenvolgende verkiezingen, hebben wij op verzoek een nieuwe steekproef getrokken voor de Europese Parlementsverkiezing. Hierbij gebruikten wij het reeds ontwikkelde herbruikbare algoritme zoals omschreven in de voorgaande rapportage, aangepast om tegemoet te komen aan praktische verzoeken vanuit het ministerie van BZK. **Hoofdstuk 4** beschrijft het precieze algoritme. De resulterende lijst van gemeentes is op 19 maart 2024 opgeleverd aan het ministerie.





1 Inleiding

1.1 Aanleiding en doel

De gemeentes zijn verantwoordelijk voor het organiseren en toegankelijk maken van stemlokalen. Alle stemlokalen in Nederland moeten volgens de Kieswet toegankelijk zijn voor mensen met een beperking.³ Om de toegankelijkheidssituatie te beoordelen worden er data verzameld in opdracht van het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK).

Sinds 2022 voeren getrainde inspecteurs vanuit Ongehinderd⁴ gemeentes een beoordeling uit van de toegankelijkheidssituatie bij stemlokalen. Dat gebeurt vanwege praktische haalbaarheid niet bij alle stemlokalen in Nederland, maar bij een steekproef. In maart 2023 vond de Provinciale Staten- en Waterschapsverkiezing (vanaf nu afgekort als PSWS verkiezing) plaats, en diezelfde maand publiceerde Ongehinderd de bevindingen van diens inspecteurs.⁵ Om te komen tot een representatieve steekproef heeft BZK tevens in maart van dat jaar aan prof. Ernst Wit gevraagd om een statistisch advies uit te brengen.⁶ Op basis hiervan is in de zomer van 2023 een nieuwe steekproef genomen van 120 stemlokalen, met oorspronkelijk de bedoeling deze te benutten voor de Europese Parlementsverkiezing. Uiteindelijk is de steekproef ingezet voor de tussentijdse Tweede Kamerverkiezing in november 2023. Recentelijk is daarom de wens ontstaan, met het oog op de verkiezing voor het Europees Parlement (EP) in 2024, om alsnog een nieuwe representatieve steekproeftrekking van gemeentes te laten doen voor het toetsen van de toegankelijkheid van stemlokalen.

BZK heeft Centerdata gevraagd om de statistische analyses uit te voeren op de data die zijn verzameld tijdens de Tweede Kamerverkiezing van 2023⁷ en tevens deze te vergelijken met de data van de PSWS verkiezing van eerder dat jaar⁸. Hiervoor zijn een aantal onderzoeksvragen gesteld. In dit eindrapport worden de onderzoeksvragen van BZK beantwoord door middel van statistische analyses op de verzamelde data.

Eind augustus 2023 heeft Centerdata bovendien een eerdere rapportage uitgebracht op verzoek van BZK, waarin de door Ongehinderd verzamelde data zijn geanalyseerd voor het gericht beantwoorden van verschillende onderzoeksvragen. Daarin is de gemeenteraadsverkiezing van 2022 vergeleken met de PSWS verkiezing van 2023. In de huidige rapportage vergelijken we de Tweede Kamerverkiezing van november 2023 met de PSWS verkiezing van maart datzelfde jaar. Daarbij staan reeds dezelfde onderzoeksvragen centraal.

³ https://wetten.overheid.nl/BWBR0004627/2023-06-20/#AfdelingII_HoofdstukJ_Paragraaf1_ArtikelJ4

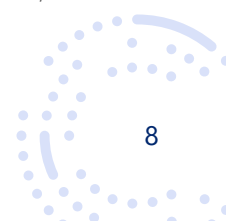
⁴ Eindrapport Steekproef Toegankelijkheid Stemlokalen Provinciale Staten- en Waterschapsverkiezing 15 maart 2023; Ongehinderd, gepubliceerd op 29 mei 2023.

⁵ <https://www.tweedekamer.nl/downloads/document?id=2023D38857>

⁶ Statistisch advies voor de steekproef methodologie met betrekking tot de toegankelijkheid van stemlokalen; Prof. Dr. Ernst C. Wit. Università della Svizzera italiana, 5 maart 2023.

⁷ Concept Steekproef Toegankelijkheid Stemlokalen. Tweede Kamerverkiezing 22 november 2023; Ongehinderd.

⁸ Eindrapport Steekproef Toegankelijkheid Stemlokalen Provinciale Staten- en Waterschapsverkiezing 15 maart 2023; Ongehinderd, gepubliceerd op 29 mei 2023.





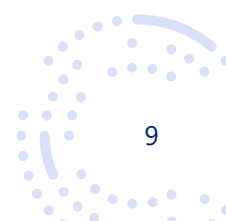
Belangrijk op te merken is dat we onze aanpak hebben doorontwikkeld. Hier wordt in dit hoofdstuk dieper op in gegaan. Door ditmaal rekening te houden met de verschillende zwaartes van elk toetsingsonderdeel van Ongehinderd zijn de resultaten van het huidige rapport niet één-op-één vergelijkbaar met die uit het voorgaande rapport. Wél zijn de cijfers over beide verkiezingen binnen dit rapport vergelijkbaar, omdat we dezelfde meetmethode toepassen op beide verkiezingen (PSWS verkiezing en Tweede Kamerverkiezing).

1.2 Onderzoeksvragen

Vanuit BZK zijn er zes onderzoeksvragen gesteld. Deze zijn:

1. Wat is de betrouwbaarheid en nauwkeurigheid van de uitkomsten van het onderzoek van Ongehinderd naar de toegankelijkheid van stemlokalen bij de Tweede Kamerverkiezing in 2023?
2. Welke statistische verbanden zijn er te leggen tussen de data van Ongehinderd?
3. Zijn er controleurseffecten waar te nemen (wordt de uitkomst beïnvloed door wie de controle op toegankelijkheid heeft gedaan)?
4. Zijn er verschillen in toegankelijkheid tussen gemeentes en/of provincies waar te nemen?
5. Hoe verhouden de uitkomsten van het onderzoek van Ongehinderd naar toegankelijkheid van stemlokalen bij de Tweede Kamerverkiezing in november 2023 zich tot de uitkomsten van hun onderzoek naar toegankelijkheid van stemlokalen bij de PSWS verkiezing in maart 2022, en hoe kunnen deze verschillen worden verklaard?
6. In welke gemeentes kunnen bij de volgende verkiezing stemlokalen worden onderzocht (en hoeveel per gemeente), rekening houdend met een goede afspiegeling van de landelijke situatie? BZK wil graag dat de steekproeftrekking wordt gedaan op de wijze die is omschreven in het rapport van prof. Ernst Wit. Om gemeentes niet te veel te belasten, wil BZK graag dat gemeentes die zijn onderzocht bij de Tweede Kamerverkiezing in november 2023 van de steekproef worden uitgesloten. En dat per provincie gemeentes worden gekozen die bij elkaar in de buurt liggen, zodat het voor de controleurs mogelijk is om alle gemeentes uit de steekproef te bezoeken op de verkiezingsdag.

De eerste vijf onderzoeksvragen worden behandeld in **hoofdstuk 3**. De laatste onderzoeksvraag wordt beantwoord in **hoofdstuk 4**, waarbij er een nieuwe lijst van gemeentes (inclusief het aantal stemlokalen) wordt gedeeld met BZK voor het steekproefonderzoek van de verkiezing voor het Europees Parlement in 2024.





2 Data en methode

2.1 De datasets

De data die zijn geanalyseerd in dit onderzoek komen van de steekproefonderzoeken die zijn verricht door Ongehinderd. Controleurs van Ongehinderd beoordelen de toegankelijkheid van stemlokalen aan de hand van de 'Criteria Integrale Toegankelijkheid Stembureaus' (CITS).⁹ Hiermee is in kaart gebracht in hoeverre mensen met een beperking gebruik kunnen maken van de ruimtes en functies in en rondom de stemlokalen die nodig zijn om een stem uit te brengen.

In dit rapport wordt er niet alleen gekeken naar de data van het onderzoek bij de Tweede Kamerverkiezing van november 2023, maar tevens teruggeblikt naar de PSWS verkiezing van maart 2023. Bij beide verkiezingen is een steekproef van 120 stemlokalen bezocht en beoordeeld.

De resultaten zijn afgeleid van de eindrapporten die zijn verstrekt door Ongehinderd. De achtergrondinformatie, zoals de naam van het stembureau, de gemeente, provincie en het controleursnummer, is ook door Ongehinderd verstrekt. Door gebruik te maken van een koppelvariabele genaamd 'Rapport ID', konden deze twee datasets aan elkaar worden gekoppeld.

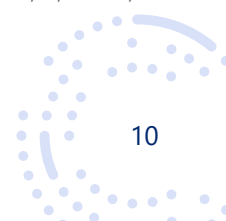
De datasets bevatten zowel numerieke als categorische beoordelingen. Dat wil zeggen dat een deel van de toetsingsonderdelen (onderdelen 1.1, 1.2, en 3.3 t/m 3.7 van CITS) aan de hand van punten zijn gewaardeerd en een ander deel van de onderdelen (1.3 t/m 3.2 van CITS) aan de hand van de categorieën *Goed* of *Fout*.

2.2 Methodes

Voor de Tweede Kamerverkiezing in november 2023 zijn twee CITS toetsingsonderdelen uit het voorgaande onderzoek naar de PSWS verkiezing samengevoegd. Dit betreft de toetsingsonderdelen 2.2 en 2.3. Toetsingsonderdeel 2.2 evalueerde het al dan niet toegankelijk en bruikbaar zijn van de eerste tussendeur in de route tussen hoofdentree en stemruimte. Toetsingsonderdeel 2.3 keek naar dit vraagstuk voor een eventuele tweede tussendeur. In de huidige data voor de Tweede Kamerverkiezing zijn de twee onderdelen samengevoegd tot één toetsingsonderdeel (2.2) voor *alle* tussendeuren op de route. Om consistent te blijven in dit onderzoek in de vergelijking tussen de steekproefonderzoeken van maart en november 2023 is het toetsingsonderdeel 2.3 voor de data uit maart 2023 uit de statistische analyses weggelaten. Dit had geen invloed op de resultaten. In de data van de PSWS verkiezing correleerden toetsingsonderdelen 2.2 en 2.3 maximaal.¹⁰

⁹ <https://www.rijksoverheid.nl/documenten/brochures/2020/09/30/criteria-integrale-toegankelijkheid-stembureaus>

¹⁰ De twee onderdelen correleerden maximaal bij alle stemlokalen. Zie pagina 19 in ons voorgaande rapport: Höcük, S., Kumar, P. & Guimarães, J. (2023). Statistische Analyse Toegankelijkheid Stemlokalen. Eindrapport. Tilburg: Centerdata: 19.





Nieuwe berekening van de toegankelijkheidsscore

Voor de huidige analyse hebben we besloten onze benadering beter aan te sluiten bij de oorspronkelijke weging van de toetsingsonderdelen in de primaire data van Ongehinderd. Die is in de voorgaande analyse van de steekproefdata van de gemeenteraadsverkiezing en PSWS verkiezing onbenut gelaten ten gunste van een efficiëntere somschaal. In de originele data van Ongehinderd hebben enkele CITS toetsingsonderdelen echter bewust meer zwaarte als criterium dan andere als het gaat om het bepalen van de toegankelijkheid van een stemlokaal. Een zekere subset aan individuele toetsingsonderdelen is zelfs onmisbaar voor een toegankelijk stemlokaal. Dit geldt voor de toetsingsonderdelen 1.3 t/m 3.2.¹¹ Dit zijn de toetsingsonderdelen die een categorische eindwaardering krijgen, *Goed* of *Fout*, in plaats van een cijfermatige ofwel numerieke waardering. Een *Fout* beoordeling betekent dat een stemlokaal in principe niet voldoet.

De beoordeling door Ongehinderd houdt sinds 2022 tevens rekening met een aantal uitzonderingen waarbij een stemlokaal op een cruciaal toetsingsonderdeel strikt genomen een *Fout* zou scoren, maar waarbij dit in de praktijk niet leidt tot een ontoegankelijke situatie. Hierdoor wordt een stemlokaal niet per definitie afgekeurd in de eindwaardering (in andere woorden: een 'Voldoet niet' toegekend als eindscore), maar krijgt daarentegen een aantal **strafpunten**. Het aantal strafpunten verschilt per toetsingsonderdeel en specifiek aangetroffen problemen. Naar dit soort situaties refereren we vanaf nu als 'twijfelgevallen'.

In dit onderzoek hebben we daarom onze voorgaande methodiek heroverwogen en gekozen voor een gecombineerde benadering. We streven net als bij ons voorgaand onderzoek naar een volledig cijfermatige beoordeling van de stemlokalen: *hoe* toegankelijk zijn ze? Dit betekent dat we nog steeds een cijfermatige toegankelijkheidsscore berekenen als eindresultaat. Deze score is net als in het vorige onderzoek een optelsom van alle individuele scores van de toetsingsonderdelen. De optelsom resulteert in een getal tussen de 0 en 10. Maar in tegenstelling tot onze voorgaande berekeningsmethode, pakken we voor het huidige onderzoek een aantal zaken anders aan om oog te houden voor verschillen in het relatieve belang van de verschillende toetsingsonderdelen. Hieronder zal worden besproken welke zaken we nu anders doen.

Ten eerste normaliseren we nu niet de scores van elk individueel toetsingsonderdeel. De toetsingsonderdelen verschillen onderling in de scorereikwijdte (de *range*), ofwel de variatie in het maximum aantal mogelijke punten per onderdeel. Die variatie houden we intact. Dit doen we omdat Ongehinderd de verschillende toetsingsonderdelen op die manier in verschillende mate bewust heeft laten meewegen in de eindwaardering: het ene toetsingsonderdeel dient inhoudelijk als zwaarwegender te worden behandeld dan andere. Hoe groter de numerieke variatie ofwel de *range* in te behalen scores per onderdeel, hoe crucialer het onderdeel.

¹¹ Ongehinderd. (2023). Toetsingsmethodiek. Den Haag: Ministerie van Binnenlandse Zaken en Koninkrijksrelaties, p. 1: *Scoort de locatie voor dit onderdeel onvoldoende op één van de criteria dan wordt deze locatie als niet toegankelijk beoordeeld. De waardering van de overige toegankelijkheidsonderdelen kan dan nog steeds goed worden bevonden, dit is dan echter niet meer relevant, omdat de locatie reeds is afgekeurd.*





Twee alternatieve toegankelijkheidsscores

Omdat het systeem van toetsingsonderdelen en strafpunten complex is, hebben we ervoor gekozen om twee alternatieve toegankelijkheidsscores te berekenen (**Toegankelijkheidsscore 1** en **Toegankelijkheidsscore 2** genoemd), die elk volgens een eigen benadering een specifiek verschil belichten. Dit biedt ons een vollediger inzicht in de toegankelijkheidssituatie bij de stemlokalen. Het enige aspect waarop de twee maten verschillen, betreft de omcodering van de categorische (*Goed* en *Fout*) scores naar numerieke waarden. Hieronder worden de twee toegankelijkheidsscores toegelicht.

1. Voor **Toegankelijkheidsscore 1** hebben we de som van alle cijfermatige toetsingsonderdelen genomen.¹² Stemlokalen die op minimaal één categorisch toetsingsonderdeel een *Fout* hebben staan, worden op een 5,5 geplaatst. Daarbovenop krijgen deze stemlokalen steeds een vast aantal strafpunten per toegekende *Fout*.¹³ We hebben daarnaast geconstateerd dat een strafpunt door een *Fout* score (een volledige afkeuring van een stemlokaal), nooit lager kan zijn dan een strafpunt uit een 'twijfelgeval'. Daarom hebben we van de twijfelgevallenstrafpunten niet de gehele waarde, maar twee-derde van die waarde genomen als aftrekpost van de somscore. Het resultaat van de somscore minus de strafpunten vormt de uiteindelijke Toegankelijkheidsscore 1.
2. In onze tweede maat (**Toegankelijkheidsscore 2**) krijgen stemlokalen die minstens één *Fout* score hebben, meteen een Toegankelijkheidsscore van 0 toegewezen. Het maakt daarbij niet uit bij hoeveel toetsingsonderdelen we een *Fout* aantreffen. Voor de overige stemlokalen wordt Toegankelijkheidsscore 2 de som van alle cijfermatige maten, minus eventuele strafpunten op de eerdergenoemde twijfelgevallen. Ditmaal zijn de volledige strafpunten meegerekend, dus niet slechts twee-derde van hun waarde. Net als bij Toegankelijkheidsscore 1 resulteert ook hier de optelsom van de numerieke toetsingsonderdelen in een totaalscore van 0 tot 10.

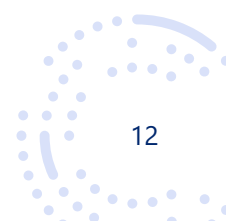
De toegankelijkheidsscore is een cijfermatige maat die door ons is berekend om extra kwantitatieve analyses te kunnen verrichten, en de stemlokalen te duiden op een soort scoreladder die de mate van toegankelijkheid aanduidt. Daarnaast is er de **Eindwaardering** die Ongehinderd zelf aan de stemlokalen heeft gegeven, gebaseerd op de beoordelingen van de controleurs op de individuele toetsingsonderdelen. De eindwaardering is niet numeriek, maar een categorische waardering waarop stemlokalen een 'Voldoet' of 'Voldoet niet' eindoordeel krijgen. **Figuur 1** vat alle verschillende toegankelijkheidsmaten samen.

Er zijn twee situaties waardoor een stemlokaal een 'Voldoet niet' krijgt als eindwaardering:

1. Stemlokalen die op minstens één van de categorische beoordelingen een *Fout* oordeel krijgen, hebben als eindwaardering per definitie een 'Voldoet niet'.

¹² De cijfermatige toetsingsonderdelen zijn: onderdeel 1.1, 1.2 en 3.3 t/m 3.7

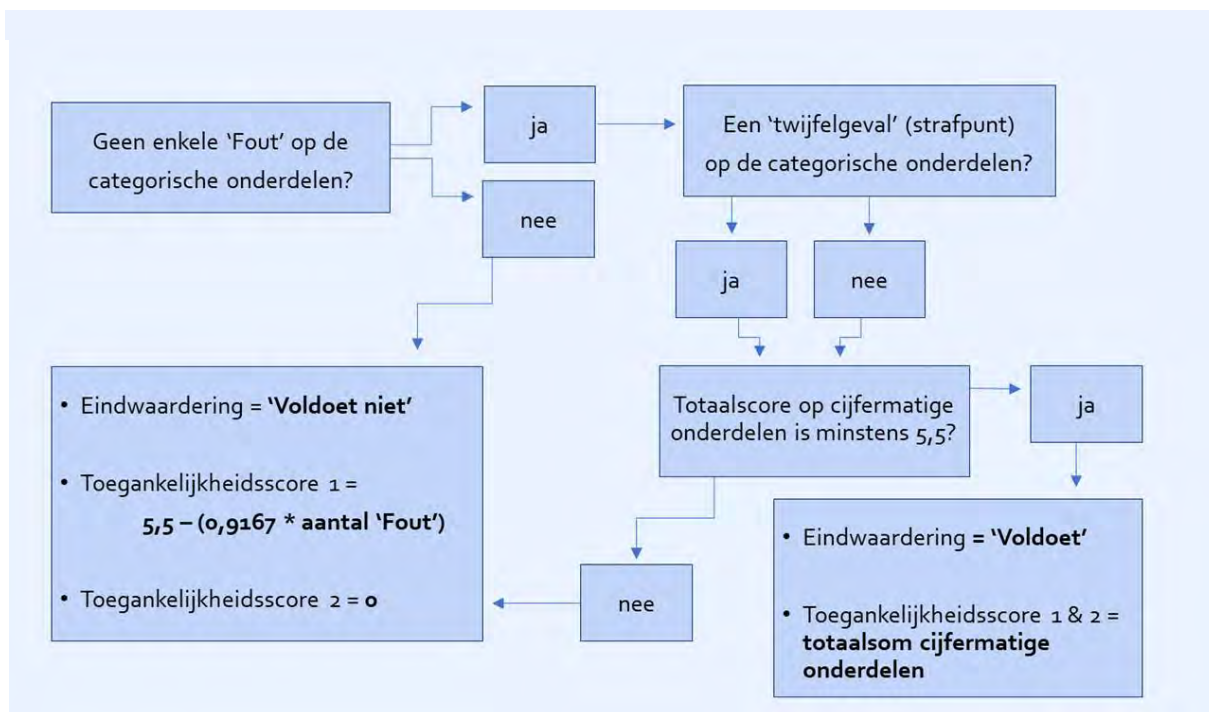
¹³ Het aantal strafpunten per "Fout"-gescoorde toetsingsonderdeel is door ons vastgesteld op $5,5 / 6 = 0,9167$. We houden daarbij rekening met het feit dat er zes categorische indicatoren zijn, en een Toegankelijkheidsscore van lager dan een 5,5 (op een rapportcijfer van 0 tot 10) moet worden gezien als een ontoegankelijk stemlokaal. Correspondentie met Ongehinderd (februari-maart 2024). Een stemlokaal met een *Fout* score op één onderdeel krijgt als toegankelijkheidscore dus een $5,5 - 0,9167 = 4,5833$. Een stemlokaal met een *Fout* score op twee onderdelen krijgt een $5,5 - 2 * 0,9167 = 3,6666$. Een stemlokaal met een *Fout* score op alle categorische onderdelen krijgt afgerond een 0.





2. Zelf hebben wij de stemlokalen toegevoegd die wellicht op alle categorische beoordelingen een *Goed* scoren, maar een totaalscore van lager dan 5,5 behalen op alle cijfermatige onderdelen bij elkaar opgeteld.

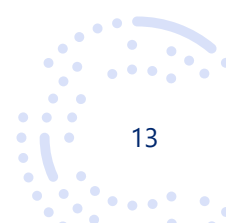
Figuur 1: Alle toegankelijkheidsmaten



Toegankelijkheidsscore 1 houdt rekening met het feit dat stemlokalen op meerdere essentiële punten ontoegankelijk kunnen zijn. Bij stemlokalen die op drie essentiële toetsingsonderdelen falen, heeft men meer problemen op te lossen dan bij een stemlokaal dat op één essentieel onderdeel tekortschiet. Vandaar dat het eerst omschreven stemlokaal een lagere toegankelijkheidsscore krijgt dan het tweede stemlokaal. Toegankelijkheidsscore 1 geeft aan hoe dicht alle stemlokalen is bij een eindwaardering 'Voldoet' staan (minimaal een 5,5 scoren).

Toegankelijkheidsscore 2, daarentegen, geeft meer inzicht in hoe ver de goedgekeurde stemlokalen verwijderd staan van een 'perfect grade'. Toegankelijkheidsscore 2 heeft tevens nog meer oog voor de doorslaggevende aard van elk individueel categorisch toetsingsonderdeel. Zodra een stemlokaal op één essentieel categorisch onderdeel onvoldoende is, is het een ontoegankelijke locatie en krijgt het een nul als eindscore. Alleen voor de toegankelijke stemlokalen is het vervolgens nog interessant te kijken *hoe* goed die locaties het doen. De bevindingen op deze alternatieve maat staan in de Bijlagen, en dienen als een extra robuustheidscheck.

Naast een categorische score die ons uiteindelijk het percentage van toegankelijke stemlokalen geeft ('Voldoet' versus 'Voldoet niet', hebben we twee alternatieve nieuwe cijfermatige scores gecreëerd. Hierbij maken we gebruik van alle individuele waarderingen die samen een totale toegankelijkheidsscore tussen 0 en 10 geven.





Cijfermatige beoordelingen zijn geschikter voor veel statistische analyses. De verschillende onderscheidingsdimensies binnen de data (verkiezing, provincie, gemeentegrootte, controleur) worden getoetst op statistisch significante verschillen in de gemiddelde waarden. Deze vergelijkingen van gemiddelden volgens een significantietoets bieden ons inzicht in de vraag of we kunnen spreken van daadwerkelijke verschillen tussen verkiezingen, provincies, gemeentes en/of controleurs. Vanwege de manier waarop de scores zijn verspreid (zie meer daarover in **Hoofdstuk 3**) vergelijken we daarnaast ook de medianen in plaats van alleen de gemiddelden. Tenslotte wordt er gekeken naar de scorevariatie aan de hand van meerdere statistische toetsen. De variatie laat zien hoe uiteenlopend de scores zijn, of er vaak voorkomende scores zijn en waar die veelvoorkomende scores liggen (bij hoge scores, in het midden, of bij lage scores). De toegepaste technieken zijn: *Shapiro-Wilk* normaliteitstoets¹⁴, *Levene's F*-toets¹⁵, tweezijdig *Welch t*-toets¹⁶, ANOVA¹⁷, de *Tukey's HSD*¹⁸, *McNemar exact test*¹⁹, *Wilcoxon rank-sum test*²⁰, *Kruskal-Wallis rank-sum test*²¹, en de *Spearman correlatie*²².

¹⁴ https://en.wikipedia.org/wiki/Shapiro-Wilk_test

¹⁵ https://nl.wikipedia.org/wiki/Toets_van_Levene

¹⁶ https://en.wikipedia.org/wiki/Welch%27s_t-test

¹⁷ <https://www.xlstat.com/en/solutions/features/welch-and-brown-forsythe-one-way-anova>

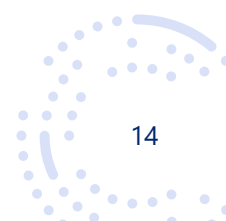
¹⁸ https://en.wikipedia.org/wiki/Tukey%27s_range_test

¹⁹ <https://measuringu.com/mcnemar-exact-test/>

²⁰ https://en.wikipedia.org/wiki/Mann%E2%80%93Whitney_U_test

²¹ <https://nl.wikipedia.org/wiki/Kruskal-Wallistoets>

²² https://en.wikipedia.org/wiki/Spearman%27s_rank_correlation_coefficient





3 Statistische analyses

In dit onderzoek richten de statistische analyses zich op verschillende onderscheiden: die tussen verkiezingen, provincies, gemeentegroottes, en controleurs. Op elk van die onderscheiden vergelijken we de gemiddelden, medianen en variatie van de scores. Vervolgens bepalen we of er sprake is van statistisch significante verschillen. De bevindingen worden weergegeven in grafieken, tabellen, en statistische toetsuitkomsten.

3.1 Verschillen tussen verkiezingen van maart en november 2023

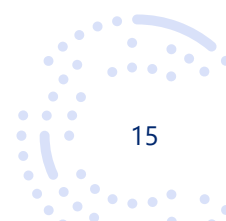
Kijkende naar de Eindwaardering ('Voldoet' versus 'Voldoet niet') is er een stijging zichtbaar tussen de verkiezingen van maart en november 2023 in het percentage toegankelijke stemlokalen. Tijdens de PSWS verkiezing was één-vijfde (21%) van de stemlokalen toegankelijk, terwijl dat aandeel bijna één-derde (31%) bedroeg voor de Tweede Kamerverkiezing (zie ook **Tabel 1**). Dat is precies tien procentpunten hoger. Deze toename werd ook al door Ongehinderd geconstateerd in hun rapportage.²³ In tegenstelling tot het voorgaande onderzoek, waarbij de PSWS verkiezing werd vergeleken met de gemeenteraadsverkiezing, is er nu sprake van een gelijke omvang van de twee steekproeven (120). Ook is de steekproefmethode hetzelfde. De toename kan dus niet worden verklaard uit steekproefverschillen. Desalniettemin blijft het percentage toegankelijke stemlokalen ver onder de norm die door de Kieswet wordt voorgeschreven, namelijk dat alle (100%) stemlokalen toegankelijk dienen te zijn.

Tabel 1 geeft de gemiddelde en mediane toegankelijkheidsscores aan voor de twee verkiezingen in 2023, voor Toegankelijkheidsscore 1. In **Bijlage A** staan de resultaten voor Toegankelijkheidsscore 2. Als gekeken wordt naar de gemiddelde toegankelijkheidsscores over de tijd heen, dan zien we een toename in toegankelijkheid. De mediaan biedt hier echter een doeltreffender indicatie van het typische beeld dan het gemiddelde. Dit heeft te maken met de variatie van de scores. Deze scores zijn niet meer normaal verdeeld zoals bij de toegankelijkheidsmaat gebruikt in de voorgaande rapportage, omdat we nu stemlokalen die een *Fout* score hebben op een cruciaal toetsingsonderdeel strenger 'afrekenen': zij scoren sowieso onder de 5,5 (Toegankelijkheidsscore 1) of zelfs een 0 (Toegankelijkheidsscore 2). Er ontstaat daardoor een tweedeling in de scores, vooral bij Toegankelijkheidsscore 2. Zoals eerder aangestipt hebben we voor deze strengere beoordeling gekozen om recht te doen aan de cruciale aard van enkele toetsingsonderdelen. Dit heeft gevolgen voor de variatie en de betekenis van een gemiddelde. Terwijl een gemiddelde sterk meegetrokken wordt naar extremen in een scheve of (meer specifiek hier) 'binomiale'²⁴ verdeling, blijft een mediaan hoe dan ook het midden van alle scores aangeven. Vandaar dat we tevens naar de mediaan kijken.

Zowel de gemiddelde als de mediane toegankelijkheidsscore geeft een toename aan tussen maart en november 2023. De toename lijkt iets kleiner wanneer we kijken naar de medianen in plaats van de gemiddelden. De variatie (standaarddeviatie) van de scores lijkt iets breder tijdens de Tweede

²³ Ongehinderd. (2024). *Concept Steekproef Toegankelijkheid Stemlokalen. Tweede Kamerverkiezing 22 november 2023*. Den Haag: Ongehinderd.

²⁴ https://nl.wikipedia.org/wiki/Binomiale_verdeling





Kamerverkiezing dan bij de PSWS verkiezing. De gemiddelde toegankelijkheidsscore in maart 2023 bedraagt 4,84 en is 5,54 in november dat jaar.²⁵

Op een schaal van 0 tot 10 zijn deze scores relatief mager wanneer we dit zien als een soort 'rapportcijfer'. Dit is niet verrassend, aangezien de meerderheid (69%) van de stemlokalen geen 'Voldoet' heeft behaald als Eindwaardering, en daarmee per definitie een toegankelijkheidsscore die lager is dan 5,5 (zie **Figuur 1**).²⁶ Zowel **Tabel 1** als **Bijlage A** laten zien dat de gemiddelden en medianen van de totale steekproeven in het gunstigste geval op de 5,5 liggen en meestal onder de 5,5.²⁷

Tabel 1: Eindwaarderingen en toegankelijkheidsscores (Toegankelijkheidsscore 1) verkiezingen 2023.

Maand	Eindwaardering "Voldoet"	Gemiddelde Toegankelijkheidsscore 1	Mediane Toegankelijkheidsscore 1
Mrt	21%	4,84 (2,31)	4,39
Nov	31%	5,54 (2,53)	4,58

De standaarddeviatie is tussen haakjes weergegeven.

Een histogram biedt een gedetailleerder beeld van de toegankelijkheidsscores voor de twee verkiezingen omdat de variatie van de scores van stemlokalen goed zichtbaar wordt (zie **Figuur 2** voor Toegankelijkheidsscore 1 en **Bijlage B** voor Toegankelijkheidsscore 2).

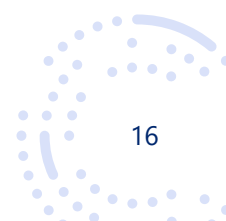
De variatie in toegankelijkheidsscores is tussen de twee verkiezingen hetzelfde gebleven. Wel is duidelijk te zien dat er relatief minder stemlokalen in de lagere scores zitten in november 2023 dan in maart. Tegelijkertijd zagen we in november 2023 meer stemlokalen onder de hogere scores dan in maart. De vraag blijft echter open of de verschillen statistisch significant zijn. Deze kwestie zal verder worden behandeld in de volgende paragraaf.

Het is belangrijk hier te herhalen dat we voor het huidige rapport andere toegankelijkheidsscores hebben berekend dan voor onze eerdere rapportage over de verkiezingen voor de PSWS en gemeenteraad. Daarom heeft de variatie die we hier bespreken een andere vorm dan de normale verdeling die we constateerden in het voorgaande rapport. In de volgende sectie bekijken we in hoeverre de verdeling statistisch afwijkt van een normaalverdeling.

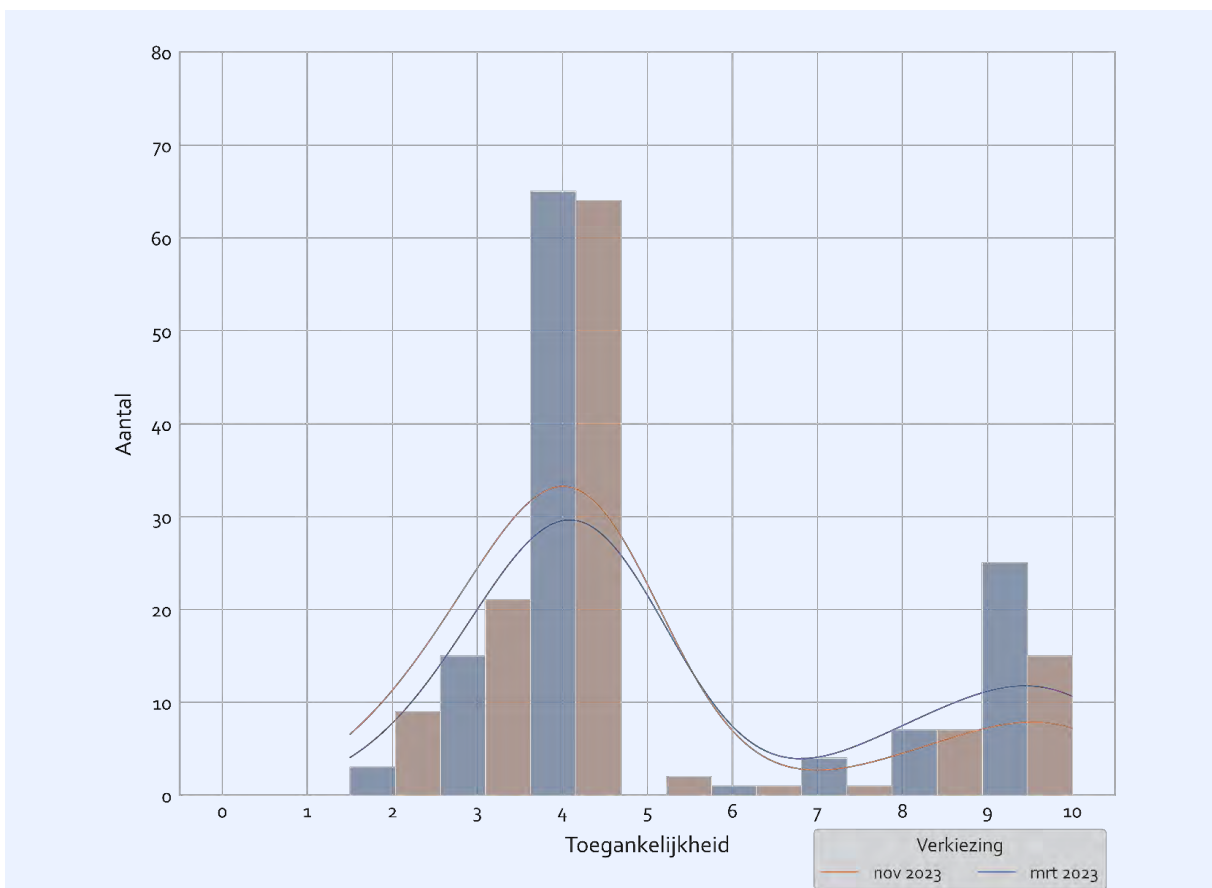
²⁵ Voor Toegankelijkheidsscore 2 steeg de gemiddelde toegankelijkheid van 1,95 naar 2,82 en steeg de Scorevariatie eveneens (zie Bijlage A).

²⁶ Zoals eerder besproken, leidt een tekortkoming op een cruciaal toetsingsonderdeel zowel bij Toegankelijkheidsscore 1 als 2 tot een eindscore onder de 5,5.

²⁷ Met het oog op het voorgaande is het tevens in de lijn der verwachting dat de mediane score op de Toegankelijkheidsscore 2 voor de gehele steekproef 0 bedraagt. De mediaan ligt immers in het midden van de verdeling, waarvan de meeste stemlokalen eindigden met een 0 omdat zij niet voldeden aan alle cruciale toegankelijkheidsvereisten. Kijken we alleen naar de 31% stemlokalen die wél voldeden tijdens de Tweede Kamerverkiezing, dan bedroeg de gemiddelde score op Toegankelijkheidsscore 2 een 9,14 en de mediaan een 9,5. Voor de 21% stemlokalen die toegankelijk waren tijdens de PSWS verkiezing, waren de gemiddelde en mediane scores op Toegankelijkheidsscore 2 iets lager: een 8,95 en 9,2 respectievelijk.



Figuur 2: Histogram van de scorevariatie van Toegankelijkheidsscore 1 voor de twee verschillende maanden.



Tests van de scorevariatie

Zoals te verwachten valt uit het beeld dat **Figuur 2** toont, laat de *Shapiro-Wilk* normaliteitstest zien dat de toegankelijkheidsscores van elke verkiezing afzonderlijk niet normaal verdeeld zijn. De normaliteitstest vertrekt vanuit de hypothese (de nulhypothese genoemd) dat de scores normaal verdeeld zijn. De kanswaarden (p-waarden) van de toetsen zijn extreem klein: $2,652 \times 10^{-11}$ voor de data uit maart 2023 en $4,431 \times 10^{-11}$ voor de data van november 2023.²⁸ De p-waarden zijn hiermee ruim onder de drempelwaarde van de 5% kans ($p < 0,05$) dat de afwijkingen door toeval komen. Dit betekent dat de nulhypothese van een normale verdeling is verworpen voor beide jaren.²⁹

Zoals gezegd is er grofweg een tweedeling in de scores door de manier waarop wij de toegankelijkheidsscores ditmaal hebben berekend, omdat we rekening wensen te houden met de doorslaggevende aard (wél of niet toegankelijk) van enkele toetsingsonderdelen. Daarom hebben we nog extra normaliteitstoetsen uitgevoerd voor elke zijde van de 'splitsing' van de verdeling afzonderlijk. Enerzijds voor de stemlokalen die als eindwaardering een 'Voldoet' kregen, versus de afgekeurde stemlokalen ('Voldoet niet'). Voor zowel de toegankelijke ('Voldoet') als niet-toegankelijke ('Voldoet niet') stemlokalen waren de p-waarden extreem klein. Dit was voor beide verkiezingen het

²⁸ Een p-waarde kleiner dan 0.01 wordt beschouwd als sterk significant en een p-waarde tussen de 0.01-0.05 als mild significant.

²⁹ De nulhypothese houdt in dat de verdeling volgens een normale verdeling is. Een p-waarde kleiner dan 0.05 verworpt de hypothese.



geval.³⁰ Ook voor de toegankelijke en niet-toegankelijke stemlokalen afzonderlijk weken de verdelingen aldus significant af van een normale verdeling.

Tenslotte blijkt dat ook alle scores van de twee verkiezingen tezamen geen normale verdeling vormen. De *Shapiro-Wilk* toetsuitkomst heeft een p-waarde van 2.461×10^{-16} , waarmee de verdeling van de scores statistisch significant afwijkt van een normaalverdeling.³¹

In het eerdere onderzoek naar de gemeenteraadsverkiezing en de PSWS verkiezing is een tweezijdige t-toets uitgevoerd om vast te stellen of de gemiddelde toegankelijkheidsscores van de stemlokalen tussen de twee jaren statistisch significant van elkaar verschillen. Een tweezijdige t-toets komt met een aantal aannames over de data waarover getoetst wordt. Een belangrijke aanname is dat de variatie van de scores een normaalverdeling laat zien. Een afwijking van die aanname was in dat voorgaande onderzoek een groter probleem geweest dan in het huidige, omdat de steekproefomvang binnen onze huidige data groter is. Deskundigen raden een steekproefomvang van minimaal 200 observaties aan om de mogelijke invloed op te vangen van het toetsen op niet-normaalverdeelde data.^{32,33} Onze huidige steekproef bedraagt 240 observaties: 120 stemlokalen per verkiezing.

Een tweede belangrijke aanname voor een t-toets is dat de verkiezingen die we vergelijken onderling ongeveer dezelfde variatie laten zien tussen de scores. Niet de hoogte van de scores, maar de wijze waarop die verspreid liggen, moet er redelijk hetzelfde uitzien. Voor het toetsen van deze aanname van gelijke scorevariatie is de *F-test of equal variances* zeer geschikt. Het testresultaat suggereert sterk dat de scores van de twee verkiezingen op eenzelfde manier verspreid liggen. De nulhypothese dat de verdelingen hetzelfde zijn, kon met een grote p-waarde als 0,32 (en een F-score die bijna gelijk is aan 1) niet worden verworpen.³⁴ Dit is ook zichtbaar in **Figuur 2**: de scoreverdeelheid heeft veelal dezelfde vorm. Dit biedt extra vertrouwen in de resultaten van een latere t-toets, die we in de volgende alinea bespreken.

Significantie van de verschillen over tijd

Een *Welch* tweezijdige t-toets geeft aan dat de *gemiddelde toegankelijkheidsscore van stemlokalen* tijdens de Tweede Kamerverkiezing significant hoger is dan van stemlokalen tijdens de PSWS verkiezing. De p-waarde zit onder de drempelwaarde van 0,05 ($p = 0,03$), waaronder de nulhypothese dat de gemiddelden hetzelfde zijn moet worden verworpen. Er is dus statistisch gezien een wezenlijk verschil tussen de verkiezingen dat niet te wijten is aan een toevallige steekproeftrekking.³⁵

³⁰ Voor toegankelijke stemlokalen bedroeg de p-waarde 0,0002 voor de Tweede Kamerverkiezing en 0,001 voor de PSWS verkiezing. Deze p-waarden bedroegen $2,808 \times 10^{-8}$ en $4,724 \times 10^{-7}$ voor de scorevariatie van niet-toegankelijke stemlokalen.

³¹ Ook voor de scores op Toegankelijkheidsscore 2 was de verdeling significant anders dan normaal ($p = 2,755 \times 10^{-16}$ voor de Tweede Kamerverkiezing en $2,2 \times 10^{-16}$ voor zowel de PSWS verkiezing als de gehele data).

³² Fagerland, M.W., Sandvik, L. (2009). Performance of five two-sample location tests for skewed distributions with unequal variances. *Contemp Clin Trials*, 30: 490-496. 10.1016/j.cct.2009.06.007.

³³ Fagerland M.W. (2012). T-tests, non-parametric tests, and large studies – a paradox of statistical practice? *BMC Med Res Methodol*, 12, 78: 10.1186/1471-2288-12-78.

³⁴ Dezelfde F-test op de verdeling van de scores op Toegankelijkheidsscore 2 gaf ook daar geen aanwijzing van een verschillende scorevariatie per verkiezing.

³⁵ De resultaten op Toegankelijkheidsscore 2 waren iets minder eenduidig. De t-toets gaf marginaal significantie als we de scores op Toegankelijkheidsscore 2 gebruikten ($p = 0,09$). De Wilcoxon rank-sum toets gaf echter wél een significant verschil tussen de medianen weer voor Toegankelijkheidsscore 2, hetgeen voor een niet-normaalverdeelde data zoals bij deze scores veelzeggender is. Toch gaf de vergelijkbare Kruskal-Wallis rank sum test een p-waarde van 0,12. Doordat op Toegankelijkheidsscore 2 veel scores zich aan de uiteinden concentreren (op de 0 of rond de 8 t/m 10), komt deze toets redelijk dichtbij een vergelijking van





Om extra zekerheid te krijgen in de resultaten is tevens een Wilcoxon rank-sum test (ook wel bekend als de Mann-Whitney U test) uitgevoerd. Dergelijke nonparametrische toetsen zijn geschikt voor het vergelijken van medianen (in plaats van gemiddelden) bij niet-normaalverdeelde scores.³⁶ We voeren een Wilcoxon rank-sum en een Kruskal-Wallis toets uit. Deze zijn als rank-sum methoden beiden geschikt voor onze data, waarbij de stemlokalen in onze twee steekproeven veelal niet hetzelfde zijn.

De testuitslag ondersteunt de hypothese dat de mediaan (toegankelijkheidsscore) significant toenam tussen de twee verkiezingen. De p-waarde ligt ruim onder de drempelwaarde ($p = 2.2 \times 10^{-16}$). Ook een vergelijkbare rank-sum toets, de Kruskal-Wallis, geeft aan dat het verschil tussen de verkiezingen significant is ($p = 0,02$).

Om te kijken of ook de *percentages toegankelijke stemlokalen* significant van elkaar verschillen tussen de verkiezingen is tevens een McNemar exact test uitgevoerd. Deze toets geeft een chi-square verschijscore tussen de geobserveerde uitkomsten en de uitkomsten die te verwachten zijn bij geen verschillen tussen de verkiezingen. Deze berekening houdt rekening met het feit dat onze steekproeven niet hetzelfde zijn, maar mogelijk wel enige overlap vertonen (zoals enkele grote steden). De chi-square score van 3,18 is marginaal significant: de bijbehorende p-waarde van 0,07 ligt boven de drempelwaarde van 0,05. Vooral omdat we een redelijk grote steekproef gebruiken (200), waarbij de kans op een significant verschil groter wordt, kan aldus worden geconcludeerd dat er niet overduidelijk sprake is van een daadwerkelijke toename in het percentage toegankelijke stemlokalen in Nederland. De kans dat onze geobserveerde toename komt door toevalsfactoren is te groot.

De gemiddelde en mediane toegankelijkheidsscores van stemlokalen zijn significant toegenomen bij de Tweede Kamerverkiezing ten opzichte van stemlokalen tijdens de PSWS verkiezing. Wat betreft de percentages toegankelijke stemlokalen die zijn ingezet verschillen de verkiezingen van maart en november 2023 echter niet statistisch significant van elkaar.

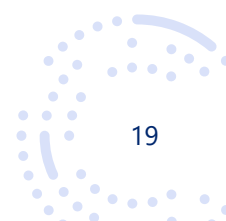
3.2 Verschillen tussen provincies

Er is bewijs voor een toename van wat? Tussen de verkiezingen in de gemiddelde en mediane toegankelijkheidsscore van stemlokalen. Bestaan er tevens verschillen tussen de Nederlandse provincies? Hier beantwoorden we die vraag met, wederom, zowel de toegankelijkheidsscores als de percentages toegankelijke stemlokalen.

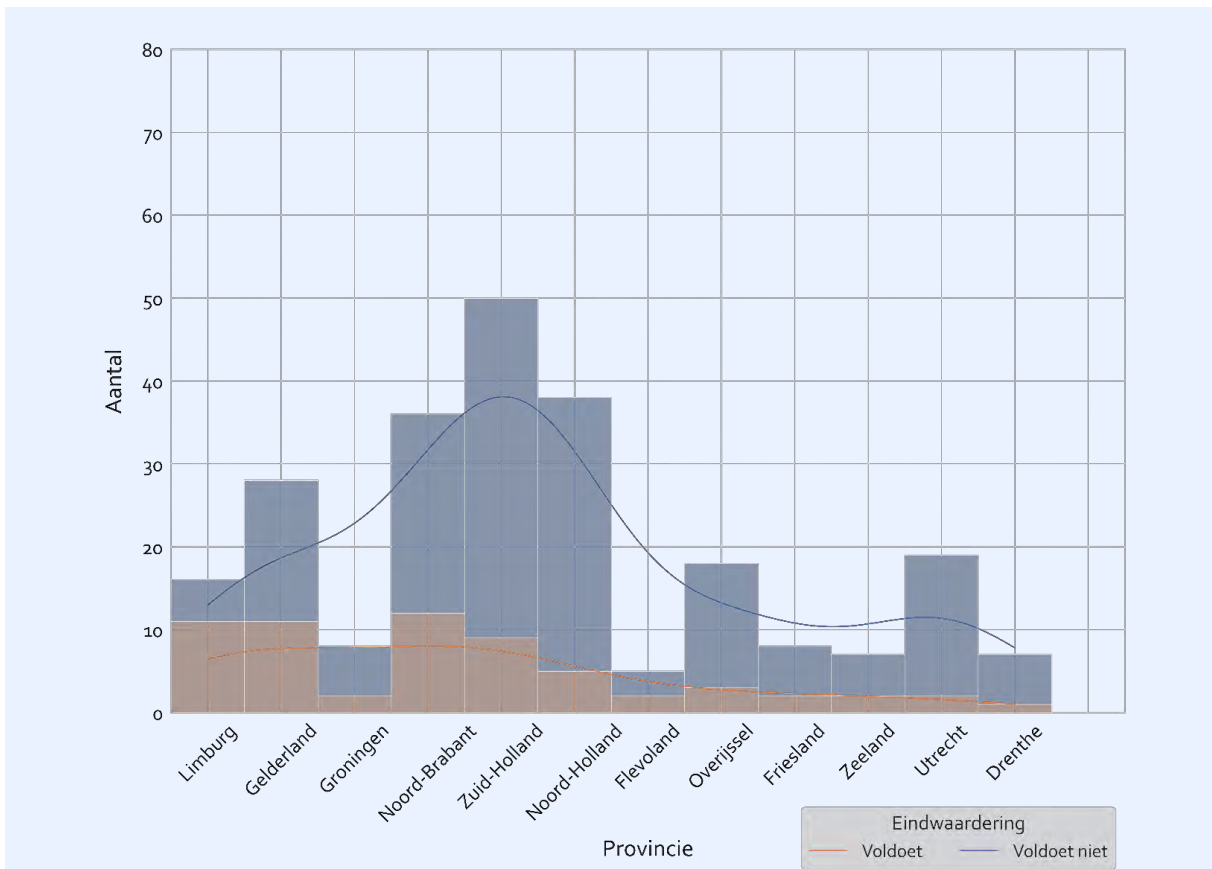
Eerst kijken we naar de stand van zaken voor heel 2023. **Figuur 3** Figuur toont het aantal toegankelijke en niet-toegankelijke stemlokalen per provincie voor de samengevoegde data van de verkiezingen van maart en november 2023. **Bijlage C** toont dezelfde informatie voor Toegankelijkheidsscore 2.

de percentages toegankelijke versus niet-toegankelijke stemlokalen. De resultaten daarmee waren inderdaad consistent (marginaal significant) als die we voor Toegankelijkheidsscore 2 aantreffen. Deze zin klopt niet grammaticaal

³⁶ Gu, C. & Ma, P. (2005). Generalized nonparametric mixed-effect models: Computation and smoothing parameter selection. *Journal of Computational and Graphical Statistics*, 14(2), 485-504.



Figuur 3: Toegankelijkheidspercentages van provincies van de steekproefonderzoeken.



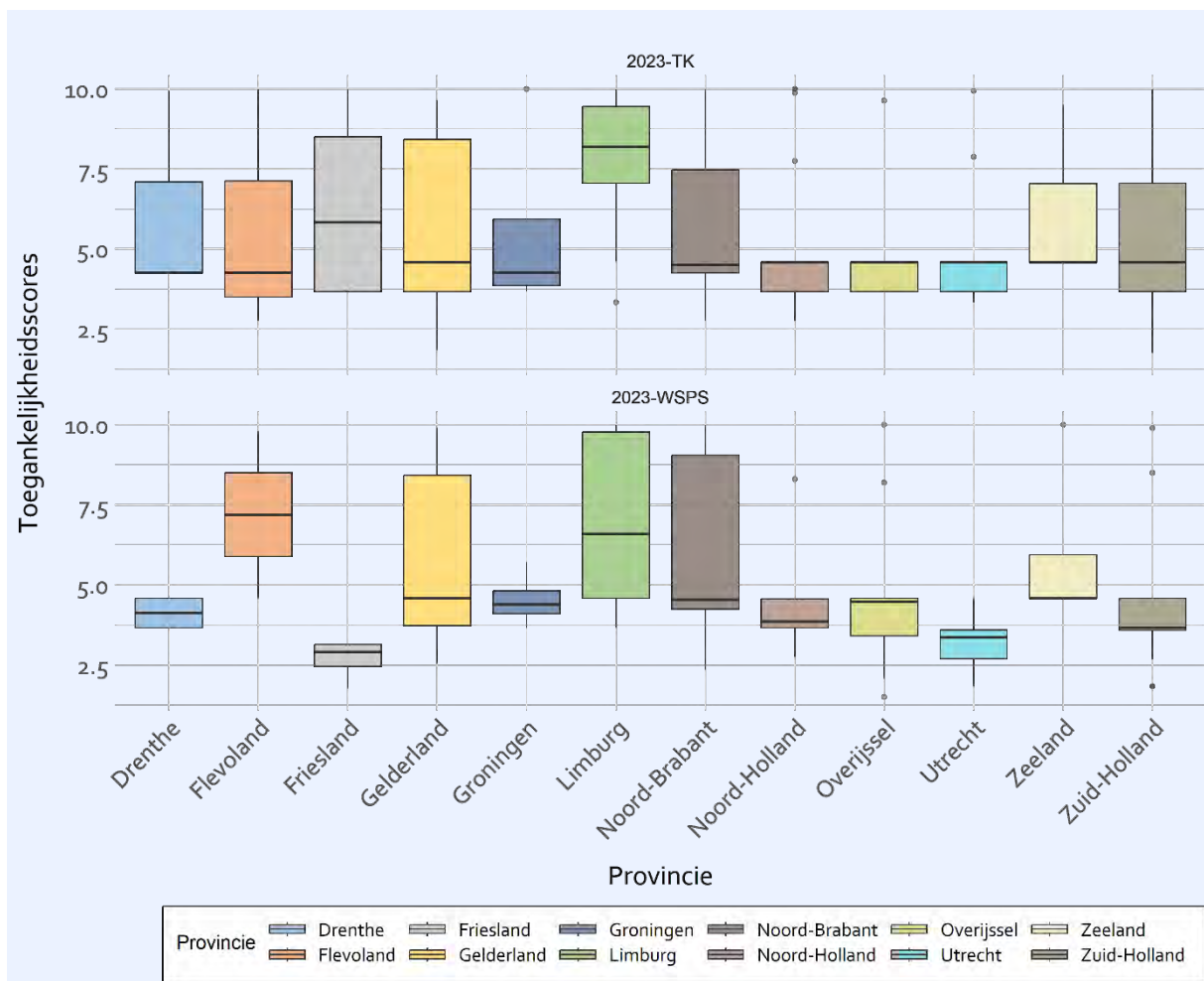
Er zijn redelijk grote verschillen te zien tussen de provincies in de verhouding van stemlokalen die wel en niet voldoen volgens de criteria van CITS. Een relatief laag percentage van toegankelijke stemlokalen is te vinden in Utrecht: slechts 11% van de beoordeelde stemlokalen was toegankelijk bevonden. De provincie Utrecht scoorde tevens laag in ons voorgaande rapport. Ook Noord- en Zuid-Holland hadden in ons voorgaande onderzoek relatief weinig toegankelijke stemlokalen, maar presteren dit keer verhoudingsgewijs beter. Toch zitten zij beide met 13% en 18% nog steeds onder de 20% toegankelijke stemlokalen. Friesland en Groningen lijken relatief laag te scoren in de figuur, maar hier moet rekening gehouden worden met het kleine aantal getoetste stemlokalen in deze provincies. Daarom biedt bij deze provincies het percentage meer inzicht dan de figuur met aantallen. Een kwart van de beoordeelde stemlokalen was toegankelijk in deze twee provincies. Hetzelfde geldt voor Zeeland en Flevoland, waar respectievelijk 29% en 40% van de bezochte stemlokalen toegankelijk bleek. De provincies Noord-Brabant (33%), Limburg (69%) en Gelderland (39%) doen het relatief goed. In deze provincies woont een relatief groter aandeel ouderen, waardoor mogelijk het bewustzijn over het belang van toegankelijkheid hoger ligt.

Via toegankelijkheidsscores kan een beter beeld worden verkregen over de *mate* van toegankelijkheid. Daarom kijken we hieronder naar de uitkomsten via de cijfermatige toegankelijkheidsscores. De provincies worden hierna per verkiezing van 2023 afzonderlijk beoordeeld.



Figuur 4 geeft de variatie van toegankelijkheidsscores van stemlokalen per provincie per jaar met boxplots.³⁷ De hoogte van een 'box' (de grootte van de 'doos' in verticale richting) staat in verhouding tot hoe breed gespreid de toegankelijkheidsscores van stemlokalen liggen.

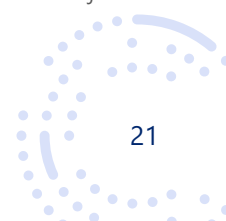
Figuur 4: Boxplot voor de variatie van de toegankelijkheidsscores van stemlokalen per provincie en jaar.



De boxplots laten grote verschillen zien tussen de provincies in de variatie van de toegankelijkheidsscores. Er zijn ook een aantal uitschieters, te zien aan de zwarte punten. Ook hier zijn de aantallen van beoordeelde stemlokalen echter van belang.

Voorals Utrecht valt door de hele periode op als provincie met een lage toegankelijkheidsscore, terwijl Noord-Brabant, Limburg en Gelderland het beter doen. Zuid-Holland doet het beter dan voorheen, hoewel dit net als in Friesland te maken kan hebben met een grotere (en gevarieerdere) steekproef. Of de verschillen ook statistisch significant zijn, zal een significantietoets uitwijzen. Die bespreken we hieronder.

³⁷ Een boxplot of een doosdiagram heeft drie horizontale strepen (onder, binnen, en boven) en twee verticale strepen. De drie horizontale strepen van de gekleurde dozen staan voor de drie kwartielen: 25%, 50%, en 75% (van onder naar boven). Het 50% kwartiel staat gelijk aan de mediaan. De verticale strepen buiten de dozen staan voor de inter quartile range (IQR) vermenigvuldigd met 1.5 en zijn begrensd door de uiterste datapunten die daar binnenin vallen. De zwarte punten voorbij de strepen zijn per definitie uitschieters in de data.





Tests van de scorevariatie

De scorevariatie binnen de provincies ziet er behoorlijk verschillend uit in **Figuur 4**, waardoor een belangrijke aanname voor een reguliere ANOVA toets mogelijk geschaad is: *homoscedasticiteit*.³⁸ Toch geeft een statistische toets aan dat de scorevariatie niet significant verschilt tussen provincies (een Levene's Test voor Gelijke Variantie gaf een p-waarde van 0,40).³⁹ Ook voor de Tweede Kamerverkiezing van november 2023 laat de variantietoets (Levene's Test) opnieuw geen significant verschil zien tussen de provincies in de mate van scorevariatie. Daarentegen waren de verschillen wél significant voor de PSWS verkiezing van maart 2023. De p-waarde van Levene's Test bedroeg 0,03.

Vanwege dit inconsistente beeld voeren we zowel een standaard ANOVA uit als een speciale variant ervan: de *Brown-Forsythe* test. Dit is een eenzijdige ANOVA test die rekening houdt met heteroscedastische data zoals we hier lijken te hebben volgens de boxplots (**Figuur 4**).

Significantie van de verschillen tussen provincies

Om de toegankelijkheidsscores in de provincies te vergelijken is een eenzijdige ANOVA toets uitgevoerd. Een eenzijdige ANOVA toets heeft niet noodzakelijk normaalverdeelde data nodig. De ANOVA resultaten voor allebei de verkiezingen tezamen laten zien dat de provincieverschillen statistisch significant zijn als het gaat om de gemiddelde toegankelijkheidsscore.⁴⁰ De reguliere ANOVA toets laat een p-waarde zien van 0,004 en de Brown-Forsythe toets een p-waarde van 0,03.

Tijdens de Tweede Kamerverkiezing zijn de provincieverschillen in de gemiddelde toegankelijkheidsscore van stemlokalen niet statistisch significant. Zowel de p-waarde van de gewone ANOVA ($p = 0,95$) als van de Brown-Forsythe toets ($p = 0,80$) waren te groot om significantie te ondersteunen.⁴¹

Tijdens de PSWS verkiezing waren de verschillen daarentegen wél significant. Voor de reguliere ANOVA toets bedraagt de p-waarde 0,001 en voor de Brown-Forsythe toets 0,006.⁴² Dit verschil tussen de twee verkiezingen heeft mogelijk te maken met het feit dat er bij de PSWS verkiezing een groter verschil was in de steekproefomvang tussen provincies. De steekproeven van Friesland en Drenthe zijn bijvoorbeeld sterk gegroeid tussen de verkiezingen in. Vooral Friesland scoort opvallend hoger bij de Tweede Kamerverkiezing dan eerder, dankzij de grotere en gevarieerdere steekproef. Toch wordt het beeld niet hierdoor verklaard. Laten we die twee provincies namelijk weg, dan blijft het beeld hetzelfde.⁴³ Het feit dat provincies significant van elkaar verschillen in de toegankelijkheid van

³⁸ <https://nl.wikipedia.org/wiki/Homoscedasticiteit>

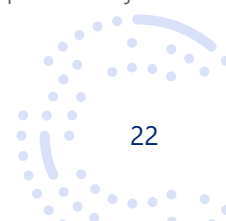
³⁹ De provincies verschilden in de variatie van hun scores ook niet significant op onze alternatieve maat (Toegankelijkheidsscore 2). De p-waarde van Levene's F-test bedroeg 0,14.

⁴⁰ Ook op onze alternatieve maat (Toegankelijkheidsscore 2), vinden we significant verschillen ($p = 0,003$).

⁴¹ Voor onze alternatieve maat (Toegankelijkheidsscore 2) vinden we dezelfde patronen voor de Tweede Kamerverkiezing. De scorevariatie verschilt niet significant tussen provincies ($p = 0,98$), en zowel de Welsh ANOVA ($p = 0,49$) als de Brown-Forsythe toets ($p = 0,68$) geven aan dat de provincies tijdens deze verkiezing niet significant verschilden in de gemiddelde toegankelijkheidsscore.

⁴² Ook hier ondersteunen de resultaten voor Toegankelijkheidsscore 2 dit patroon ($p = 0,31$ uit de ANOVA toets).

⁴³ De p-waardes voor de Tweede Kamerverkiezing zijn 0,54 voor de reguliere ANOVA en 0,67 voor de Brown-Forsythe toets. Voor de PSWS verkiezing blijft de gemiddelde toegankelijkheidsscore tussen provincies significant verschillen. De p-waardes zijn 0,002 en 0,02 respectievelijk.





stemlokalen bij de verkiezing in maart 2023 en niet in november dat jaar, komt niet door de afwijkende steekproeven van Drenthe en Friesland.

Voor de PSWS verkiezing onderzoeken we middels een Tukey's HSD toets *welke* provincieverschillen groot en significant zijn. Dan blijkt dat vooral Limburg substantieel hoger scoort dan veel andere provincies (Friesland, Utrecht, Zuid-Holland). Ook het verschil tussen Utrecht en Noord-Brabant valt op, waarbij Noord-Brabant significant hoger scoort.⁴⁴

Voor het statistisch kunnen vergelijken van de mediane toegankelijkheidsscores tussen provincies is er echter een probleem: het kleine aantal stemlokalen per provincie. We hebben bijvoorbeeld maar 5 stemlokalen voor Flevoland en 8 voor Friesland. Als we de verkiezingen afzonderlijk bekijken, is dit aantal nog kleiner. Dit is tevens een probleem voor het vergelijken van de percentages toegankelijke stemlokalen. Een zeer kleine groepsgrootte geeft een groot risico op vertekening van de resultaten bij chi-square toetsen. We houden het daarom hier alleen bij de ANOVA toets, ook omdat de ANOVA methodiek robuust is tegen non-normaliteit van de data en we via verschillende varianten van de toets rekening kunnen houden met mogelijk ongelijke scorevariatie tussen provincies.

Voor de gehele data bestaan er significante verschillen tussen provincies in de toegankelijkheidsscore van de stemlokalen. Dit beeld lijkt vooral te worden gedreven door de verschillen tussen provincies tijdens de PSWS verkiezing. Deze verschillen zijn statistisch significant, terwijl de verschillen die we zagen tijdens de Tweede Kamerverkiezing dat niet waren. Dit heeft niet te maken met de gegroeide steekproef van enkele provincies zoals Friesland.

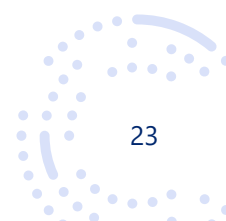
3.3 Verschillen tussen gemeentegroottes

Een adequate statistische analyse op het niveau van gemeentes is niet haalbaar vanwege de beperkte omvang van de steekproeven. De PSWS verkiezing omvat 41 gemeentes in de steekproef, terwijl de Tweede Kamerverkiezing 35 gemeentes omvat. Bovendien zijn slechts enkele stemlokalen per gemeente beoordeeld, wat resulteert in onvoldoende aantallen voor een betrouwbare statistische vergelijking.

In plaats van te kijken naar de verschillen tussen individuele gemeentes, wordt er hier gekeken naar de verschillen tussen de drie gemeentegroottes. De drie gemeentegroottes zijn: klein (inwoners < 40.000), middel (inwoners 40.000 t/m 150.000), en groot (inwoners > 150.000). De aantallen worden per jaar uitgesplitst en worden weergegeven in **Tabel 2**.

Er zijn relatief meer stemlokalen uit grote gemeentes in de steekproef van november (40 van de 120) dan in maart 2023 (30 van de 120). In november 2023 waren er juist minder stemlokalen uit middelgrote gemeentes vertegenwoordigd in de steekproef (48) vergeleken met maart (55).

⁴⁴ Alle p-waarden van deze verschillen zitten op of onder de 0,05 en de verschilcoëfficiënten waren relatief hoog voor deze verschillen (minimaal -1,7 of 1,7 en meestal rond de 3).





Tabel 2: Beschrijvende statistiek van de drie gemeentegroottes.

Gemeentegrootte (aantal inwoners)	Verkiezing	Aantal stemlokalen	Percentage voldoet	Toegankelijk- heidsscore	Standaard- deviatie	Mediaan
Klein	PSWS-2023	35	31,4%	5,60	2,68	4,52
(< 40.000)	TK-2023	32	50%	6,80	2,72	6,23
Middel	PSWS-2023	55	16,4%	4,46	1,98	4,39
(40.000 – 150.000)	TK-2023	48	22,9%	5,24	2,35	4,58
Groot	PSWS-2023	30	16,7%	4,67	2,29	3,80
(> 150.000)	TK-2023	40	25%	4,90	2,28	3,96

Zoals **Tabel 2** laat zien zijn er voldoende stemlokalen in de uitsplitsingen van gemeentegrootte met een minimum aantal van 10. In **Bijlage D** laat een tabel dezelfde informatie zien voor onze alternatieve maat: Toegankelijkheidsscore 2. Ook hier wordt gekeken naar de toegankelijkheidsscores met de scorevariatie (ofwel variantie) via een boxplot. Dit is weergegeven in **Figuur 5**. In **Bijlage E** staan corresponderende boxplots afgebeeld voor Toegankelijkheidsscore 2.

De boxplots laten grote verschillen zien in de variantie binnen de toegankelijkheidsscores tussen de gemeentegroottes. Vooral de kleine gemeentes vallen op, met een grotere variantie (ofwel scorevariatie). Wat verder opvalt is de hogere mediaan bij de kleine gemeentes in vergelijking met de rest tijdens de Tweede Kamerverkiezing. De grote gemeentes hebben bij beide verkiezingen een lagere mediaan. Hieronder bespreken we of deze verschillen statistisch significant zijn.

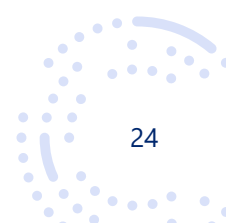
Tests van de scorevariatie

Zoals **Figuur 5** liet zien, tonen de verschillende gemeentegroottes elk een andere scorevariatie. Vooral de kleine gemeentes laten een veel grotere scorevariatie zien. Een Levene's Test of Equal Variances bevestigt dit beeld van verschillende mate van scorevariatie ($p = 0,01$).⁴⁵ Daarom is het belangrijk te kijken naar een ANOVA analyse die rekening houdt met deze *heteroscedasticiteit*.⁴⁶ We rapporteren daarom alleen de resultaten van een Brown-Forsythe toets. Daarnaast voeren we een Kruskal-Wallis chi-square test uit, omdat de groepen per gemeentegrootte ditmaal groot genoeg zijn om een grote vertekening te voorkomen.

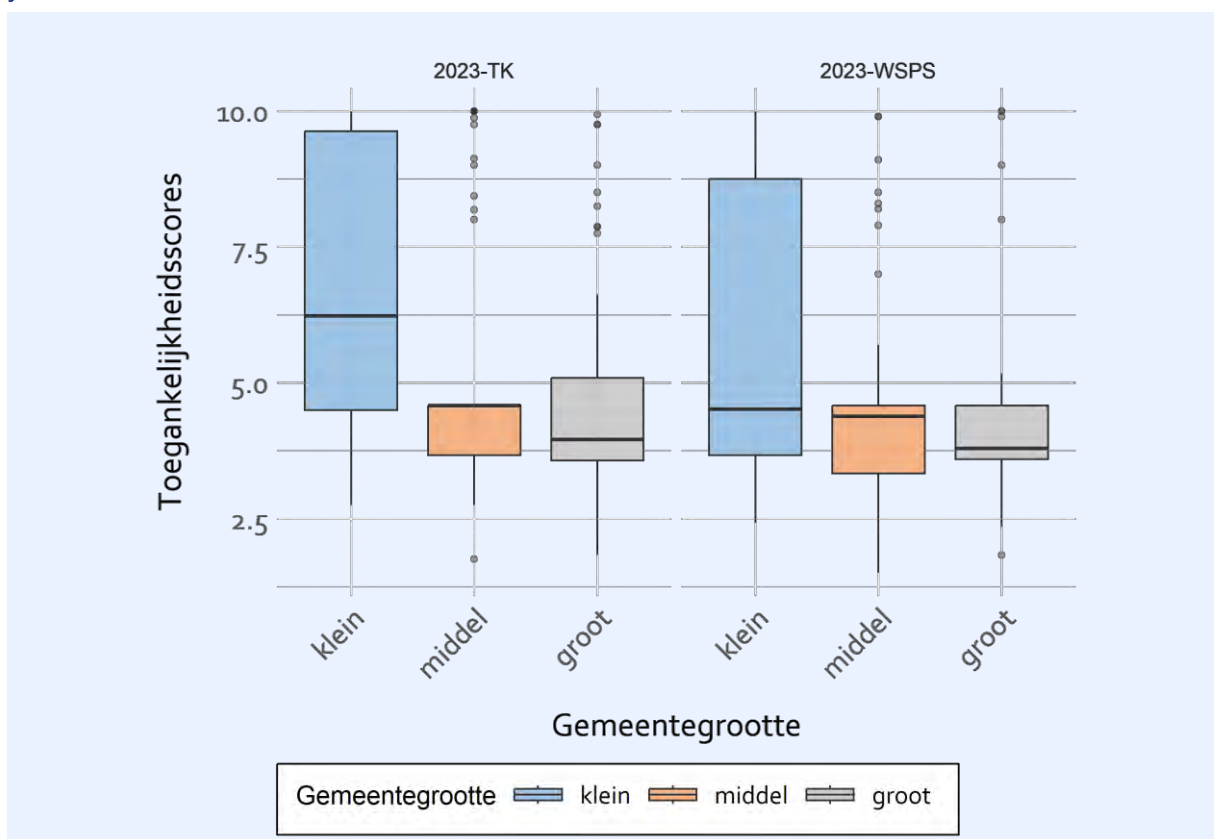
De kleine, middelgrote en grote gemeentes verschillen tevens significant in de scorevariatie als we alleen naar de Tweede Kamerverkiezing kijken. Levene's test gaf een p-waarde van 0,012. Dit is niet het geval voor de data van de PSWS verkiezing. Daarom voeren we voor de Tweede Kamerverkiezing een Brown-Forsythe toets uit en voor de PSWS verkiezing een reguliere ANOVA.

⁴⁵ Dit was tevens het geval voor de scores op Toegankelijkheidsscore 2, wat niet verrassend is, gezien de grote verschillen tussen kleine gemeentes en de rest in de scorevariatie (getoond in Bijlage E).

⁴⁶ https://en.wikipedia.org/wiki/Homoscedasticity_and_heteroscedasticity



Figuur 5: Boxplot voor de scorevariatie van de toegankelijkheidsscores van stemlokalen per gemeentegrootte en jaar.



Significantie van de verschillen tussen gemeentegroottes

Zowel de resultaten uit de Brown-Forsythe ANOVA als die uit de Kruskal-Wallis toets tonen aan dat de verschillen tussen gemeentegroottes statistisch significant zijn voor de gehele data ($p = 0,002$ voor beide toetsen).⁴⁷ Een Wilcoxon-toets is niet uitgevoerd, omdat deze voornamelijk bedoeld is voor het vergelijken van maximaal twee groepen.

Kijken we alleen naar de verkiezing van november 2023, dan blijkt dat de verschillen tussen gemeentegroottes tijdens de Tweede Kamerverkiezing statistisch significant zijn ($p = 0,008$). Ook de Kruskal-Willis toets van de medianen biedt ondersteuning voor het bestaan van significante verschillen tijdens die verkiezing ($p = 0,002$).

Tijdens de PSWS verkiezing zijn de verschillen slechts marginaal significant ($p = 0,07$). Ook de Kruskal-Willis chi-square score geeft een niet-significante coëfficiënt ($p = 0,18$). Dit bevestigt het beeld dat **Figuur 5** ons toonde: in vergelijking met de verkiezing in maart 2023, wijken de kleine gemeentes tijdens de Tweede Kamerverkiezing meer uitgesproken af van de rest. Stemlokalen in kleine gemeentes behaalden bij die laatste verkiezing over het algemeen hogere toegankelijkheidsscores dan elders. Ook de Tukey's HSD toets laat duidelijk zien dat het de verschillen zijn met de kleine gemeentes die

⁴⁷ Dit beeld wordt bevestigd door de resultaten op Toegankelijkheidsscore 2. De Brown-Forsythe ANOVA toets gaf een p-waarde van 0,006 en de Kruskal-Wallis rank-sum test een p-waarde van 0,001. Dit suggereert, kijkend naar de figuur in Bijlage E, dat vooral kleine gemeentes verschillen van de rest.



opvallen (zowel in verhouding tot middelgrote als grote gemeentes).⁴⁸ Zoals reeds aangestipt bij de bevindingen van provincieverschillen, kan dit mogelijk te maken hebben met de grotere vergrijzing in kleinere gemeentes. Deze bevolkingsfactor kan het bewustzijn hebben bevorderd van het belang van toegankelijke stemlokalen.

De toegankelijkheidsscores verschillen significant tussen kleine, middelgrote en grote gemeentes. Dit beeld is vooral gelegen in de grotere verschillen die we zien tussen kleine gemeentes en de rest tijdens de Tweede Kamerverkiezing. Stemlokalen gelegen in kleine gemeentes hebben over het algemeen een hogere toegankelijkheidsscore.

3.4 Verschillen in beoordeling tussen controleurs

Er waren in maart 2023 twaalf verschillende controleurs die de stemlokalen beoordeelden. In november dat jaar waren dat er zestien. Twaalf controleurs waren betrokken bij beide steekproefonderzoeken: dit waren alle inspecteurs die bij de verkiezing van maart 2023 reeds actief waren. De aantallen bezochte stemlokalen per controleur zijn weergegeven in **Tabel 3**.

Tabel 3: Aantal bezochte stemlokalen per controleur.

Verkiezing	Controleur															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2023-PSWS	12	12	14	10	10	10	9	9	10	9	8	7	0	0	0	
2023-TK	16	13	13	8	10	7	10	5	6	6	4	5	5	4	4	4

In het steekproefonderzoek van Ongehinderd zijn bij de verkiezing van maart 2023 drie inspecteurs in tweetallen gaan toetsen en negen alleen. Bij de Tweede Kamerverkiezing hebben er 12 van de 16 inspecteurs alleen getoetst. Er waren daarnaast 9 ervaringsdeskundige inspecteurs.⁴⁹ Alle inspecteurs hebben dezelfde training gevolgd waarbij uniform, betrouwbaar en neutraal toetsen het uitgangspunt is.

De variatie van de toegankelijkheidsscores wordt weergegeven door een boxplot in **Figuur 6. Bijlage F** toont dezelfde boxplots, maar dan voor Toegankelijkheidsscore 2.

In het voorgaande rapport hadden we reeds grote verschillen geconstateerd in de gemiddelde beoordelingsscores tussen inspecteurs. De huidige boxplots suggereren dat die verschillen kleiner zijn geworden na de PSWS verkiezing, kijkend naar de Tweede Kamerverkiezing in vergelijking met de data uit maart 2023. Toch lijken er nog een aantal verschillen te bestaan. Of deze verschillen statistisch significant zijn, bekijken we in de volgende alinea's.

⁴⁸ De verschilcoëfficiënt was -1,56 voor het verschil met middelgrote gemeentes ($p = 0,02$) en -1,90 voor het verschil met grote gemeentes ($p = 0,004$). In beide gevallen gaf de coëfficiënt aan dat kleine gemeentes hoger scoren.

⁴⁹ Correspondentie Ongehinderd, 26 februari 2024.





Tests van de scorevariatie

De verschillen in de variatie van de scores tussen inspecteurs is niet statistisch significant (Levene's Test, $p = 0,08$).⁵⁰

Voor de data van november 2023 was er sprake van een significant verschil in de variatie van scores tussen de inspecteurs ($p = 0,01$). Dit was niet het geval voor de data uit maart 2023 ($p = 0,26$).⁵¹ Hier houden we rekening mee in de keuze voor de analysetechniek. Voor de inspecteurverschillen van de Tweede Kamerverkiezing gebruiken we een Brown-Forsythe test die rekening houdt met de verschillende mate van scorevariatie. Voor de PSWS verkiezing gebruiken we een reguliere ANOVA en een Kruskal-Willis toets.

Significantie van de verschillen tussen controleurs

De verschillen tussen controleurs in de variantie binnen de door hen toegekende scores aan stemlokalen zijn verwaarloosbaar klein. Daarom doen wij een reguliere eenzijdige ANOVA analyse voor het vergelijken van de gemiddelde toegankelijkheidsscores. Voor de data van beide verkiezingen tezamen verschillen de controleurs statistisch significant in de gemiddelde beoordelingsscore ($p = 0,007$). Ook de Kruskal-Willis chi-square score is significant (p -waarde = $0,001$).

Een kanttekening die ook in onze voorgaande rapportage hierbij is geplaatst is het kleine aantal datapunten per controleur. Bij de Tweede Kamerverkiezing is het voorgekomen dat een aantal inspecteurs minder dan 5 stemlokalen bezocht heeft. Het gemiddeld aantal bezochte stemlokalen per controleur is 7,5. Dit is minder dan het aantal dat in het rapport van prof. Ernst Wit wordt geadviseerd. Vanwege dit lage aantal stemlokalen in een bepaald jaar voor enkele inspecteurs kan een Kruskal-Willis toets voor de analyses per verkiezing geen betrouwbaar beeld schetsen van de statistische significantie. We doen daarom voor de analyses die per verkiezing afzonderlijk worden gepleegd slechts een ANOVA analyse.

Wat betreft de Tweede Kamerverkiezing tonen de resultaten ook weer hier significante verschillen aan in toegankelijkheidsscores bij de verschillende controleurs onderling ($p = 0,008$). Voor de PSWS verkiezing is de p -waarde marginaal significant ($p = 0,07$).⁵² Er kan daar dus niet voldoende overtuigend worden aangetoond dat de toegankelijkheidsscores daadwerkelijk verschillen tussen inspecteurs. Mogelijk speelt hier een rol dat er bij die verkiezing minder inspecteurs waren en dus minder datapunten om over te toetsen.⁵³

⁵⁰ Voor de scores op de alternatieve maat (Toegankelijkheidsscore 2) was dit wél het geval ($p = 0,02$). Daarom is voor de gemiddelden op Toegankelijkheidsscore 2 een Brown-Forsythe ANOVA toets uitgevoerd. De verschillen in gemiddelde toegankelijkheidsscores tussen controleurs bleken hier significant ($p = 0,006$). De medianen verschilden marginaal significant ($p = 0,057$).

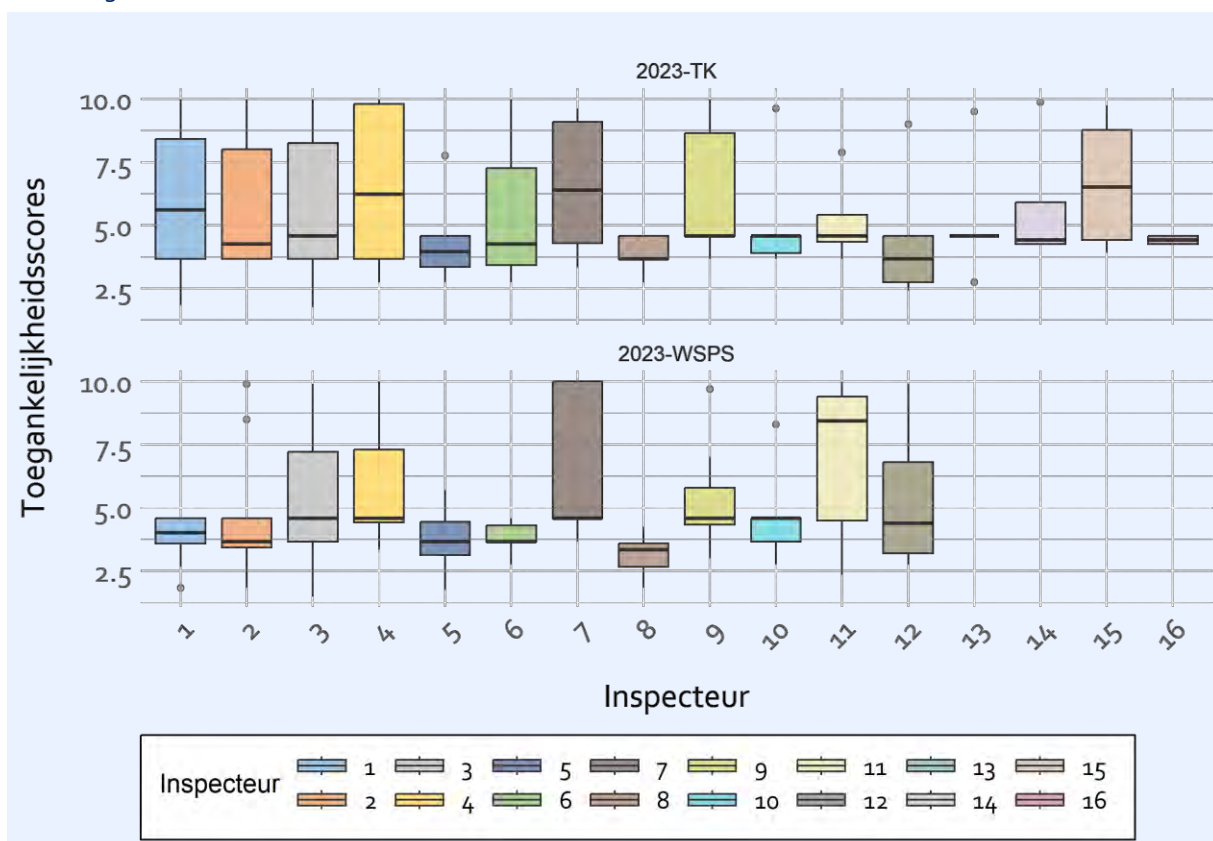
⁵¹ Ditzelfde patroon vonden we voor de alternatieve maat (Toegankelijkheidsscore 2): $p = 0,012$ voor november 2023 en $0,26$ voor maart 2023.

⁵² Dezelfde patronen vonden we ook hier weer voor de alternatieve maat (Toegankelijkheidsscore 2). Voor de Tweede Kamerverkiezing waren de verschillen significant ($p = 0,008$), maar voor de PSWS verkiezing slechts marginaal significant ($p = 0,067$).

⁵³ Deze bevinding over controleurseffecten tijdens de PSWS verkiezing wijkt af in vergelijking met die in onze voorgaande rapportage over de PSWS verkiezing (Höcük, S., Kumar, P. & Guimarães, J. (2023). Statistische Analyse Toegankelijkheid Stemlokalen. Eindrapport. Tilburg: Centerdata). Daar vonden we wel een significant verschil tussen controleurs. Dit lijkt op het eerste oog in tegenspraak met elkaar, maar heeft te maken met onze in **Hoofdstuk 2** uitgelegde veranderde meetaanpak. We voeren ditmaal een kritischere toets uit, omdat sommige toetsingsonderdelen nu als cruciaal worden behandeld en leiden tot



Figuur 6: Boxplot voor de scorevariatie van de toegankelijkheidsscores van stemlokalen per controleur en verkiezing.



De Tukey's HSD toets laat geen controleurs zien die uitgesproken verschillen van anderen. Dit heeft mogelijk te maken met het zeer lage aantal observaties per controleurs.

Een belangrijke nuancering bij de bevindingen over de inspecteursverschillen tijdens de Tweede Kamerverkiezing betreft de mogelijke rol van gemeentever verschillen. Omdat om praktische redenen iedere gemeente door een enkele controleur is beoordeeld kan het beeld mogelijk zijn beïnvloed door niet-onderzochte contextuele factoren op gemeenteniveau. We zagen al eerder dat kleine gemeentes een ander beeld laten zien dan de rest, wat mogelijk door kan spelen in de scoreverschillen die we hier tegenkomen tussen inspecteurs. Daarbij moet in beschouwing worden genomen dat de controleurs een meerdaagse training hebben gevolgd bij Ongehinderd, waarbij uitgebreid aandacht is besteed aan de criteria, meetmethoden en hoe afwijkende praktijksituaties te beoordelen. Dat kan de uniformiteit van de controleursbeoordelingen hebben bevorderd, waardoor in elk geval een deel van de hierboven besproken verschillen mogelijk gemeente-effecten zijn.⁵⁴

een zeer lage score. Mede omdat er weinig datapunten zijn per controleur, leiden dit soort strengere toetsen al relatief snel tot een verlies van significantie.

⁵⁴ Ongehinderd. (2024). *Steekproef Toegankelijkheid Stemlokalen. Tweede Kamerverkiezing 22 november 2023*. Den Haag: Ongehinderd.



Er zijn statistisch significante verschillen gevonden in de beoordelingen van controleurs tijdens de Tweede Kamerverkiezing. Een nuancering hier is dat de significante verschillen tussen controleurs mogelijk beïnvloed zijn door onderlinge verschillen tussen gemeentes, omdat om praktische redenen iedere gemeente door een enkele controleur is beoordeeld.

3.5 Statistische verbanden data Ongehinderd

We hebben gekeken of de onderdelen die de toegankelijkheid van een stemlokaal definiëren, sterk met elkaar samenhangen of juist elk een afzonderlijke factor zijn. We keken hiervoor naar de verbanden tussen de beoordelingen van de individuele toetsingsonderdelen van het CITS.⁵⁵ De verbanden worden weergegeven aan de hand van statistische gemiddelden en standaarddeviaties van de toegankelijkheidsscores en aan de hand van correlaties (samenhang) tussen de beoordelingen van de verschillende toetsingsonderdelen.

Omdat de kwalitatieve ('Voldoet' versus 'Voldoet niet') onderdelen tevens numerieke strafpunten bevatten (voor de in **Hoofdstuk 1** genoemde 'twijfelgevallen'), zijn deze voor deze specifieke correlatieanalyses aangepast. Het is immers lastig om per onderdeel afzonderlijk uit zo'n combinatie aan datatypes één eenduidig gemiddelde te halen, zonder hele arbitraire keuzes te maken. Omdat de kwalitatieve onderdelen elk over een noodzakelijk aspect van toegankelijkheid gaan, hebben we deze omgerekend tot een dichotome variabele waarbij een 0 staat voor een 'Voldoet niet' en een 1 (of een twijfelgevallenstrafpunt, waarbij een stemlokaal niet is afgekeurd) voor een 'Voldoet'. Het gemiddelde geeft dan de proportie (ofwel een percentage gedeeld door 100) van stemlokalen dat een 'Voldoet' scoorde op dat onderdeel.⁵⁶

Gedaald is de toegankelijkheid van de kandidatenlijst. Ongehinderd constateert in het steekproefonderzoek dat de kandidatenlijst 'niet aanwezig of niet leesbaar op de juiste hoogte' is en 'niet veel verbeterd' sinds de verkiezing in maart. Maar ook de verkeersruimte en de tussendeuren zijn minder toegankelijk bij de verkiezing in november 2023. In het onderzoek naar de Tweede Kamerverkiezing bestaat het 'Tussendeur' onderdeel uit twee samengevoegde eerdere onderdelen voor 'Tussendeur 1' en 'Tussendeur 2'. Dat kan mogelijk het verschil verklaren, **ook al** is de samenhang tussen de twee voorgaande onderdelen sterk. Vergeleken met de stijgingsomvang zijn de dalingen klein.

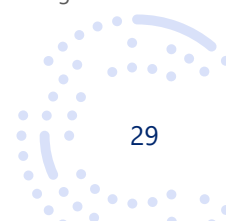
Tabel 4 geeft de gemiddelde toegankelijkheidsscores met de standaarddeviaties tussen haakjes weer.

Gedaald is de toegankelijkheid van de kandidatenlijst. Ongehinderd constateert in het steekproefonderzoek dat de kandidatenlijst 'niet aanwezig of niet leesbaar op de juiste hoogte' is en 'niet veel verbeterd' sinds de verkiezing in maart.⁵⁷ Maar ook de verkeersruimte en de tussendeuren zijn minder toegankelijk bij de verkiezing in november 2023. In het onderzoek naar de Tweede Kamerverkiezing bestaat het 'Tussendeur' onderdeel uit twee samengevoegde eerdere onderdelen

⁵⁵ <https://www.rijksoverheid.nl/documenten/brochures/2020/09/30/criteria-integrale-toegankelijkheid-stembureaus>

⁵⁶ Een gemiddelde van 0,94 op Score 3.2 betekent dus dat 94% van de stemlokalen hier een 'Voldoet' scoorde (of een strafpunt had door een twijfelgeval, maar niet werd afgekeurd), en de rest een 'Voldoet niet'.

⁵⁷ Ongehinderd. (2024). *Steekproef Toegankelijkheid Stemlokalen. Tweede Kamerverkiezing 22 november 2023*. Den Haag: Ongehinderd: 2.





voor 'Tussendeur 1' en 'Tussendeur 2'. Dat kan mogelijk het verschil verklaren, ook al is de samenhang tussen de twee voorgaande onderdelen sterk. Vergeleken met de stijgingsomvang zijn de dalingen klein.

Tabel 4: Gemiddelde toegankelijkheidsscores per toetsingsonderdeel van CITS per jaar.

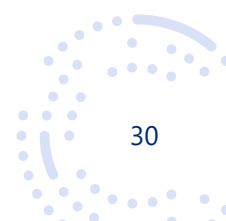
CITS onderdeel	Korte naamgeving	Gemiddelde score PSWS	Gemiddelde score TK
1.1	Openbare weg	1,29 (0,48)	1,37 (0,38)
1.2	Parkeergelegenheid	1,33 (0,95)	1,55 (0,84)
1.3	Route entree	0,53 (0,52)	0,68 (0,47)
1.4	Hoofdentree	0,31 (0,63)	0,34 (0,57)
2.1	Route stembureau	0,63 (0,58)	0,82 (0,43)
2.2	Tussendeuren	0,97 (0,22)	0,90 (0,30)
3.1	Deur stemruimte	0,72 (0,47)	0,76 (0,45)
3.2	Verkeersruimte	0,97 (0,18)	0,94 (0,24)
3.3	Stoelen	0,93 (0,69)	0,95 (0,71)
3.4	Stemhokje	0,91 (0,15)	0,94 (0,11)
3.5	Stembus	1,46 (0,18)	1,50 (0,00)
3.6	Kandidatenlijst	0,94 (0,08)	0,90 (0,20)
3.7	Leesloep	1,24 (0,50)	1,40 (0,36)

De standaarddeviatie, een mate van scorevariatie, lijkt over tijd heen betrekkelijk constant gebleven. Daar zijn een aantal uitzonderingen op. Voor de toegankelijkheid van de stembus is er bij de Tweede Kamerverkiezing geen scorevariatie meer overgebleven van de voorgaande 0,18. Ook voor de route stembureau en de leesloep nam de scorevariatie af, terwijl die voor de kandidatenlijst juist toenam.

Om de verbanden tussen de toetsingsonderdelen inzichtelijk te maken, wordt een correlatieanalyse uitgevoerd. Correlaties geven inzicht in de statistische samenhang tussen de beoordelingen waarbij de sterkte van de samenhang wordt uitgedrukt in een correlatiecoëfficiënt. De waarde daarvan kan variëren tussen de -1 en +1. Een coëfficiënt van 0 betekent absoluut geen samenhang en een coëfficiënt van +1 of -1 betekent een perfecte positieve of respectievelijk negatieve samenhang. In algemene zin wordt de grens getrokken bij een waarde van +0,70 of -0,70, waarboven of -onder er wordt gesproken over een matige tot sterke correlatie. Een voorbeeld van een perfecte negatieve samenhang (met coëfficiënt -1) is tussen de datasets van leeftijden en geboortejaren van mensen. Daar waar de een omhoog gaat, gaat de ander altijd omlaag. Via correlatie wordt er inzicht verleend in de (on)afhankelijkheid van de beoordelingen. Correlatie is echter geen bewijs van causaliteit, al kan het daar wel een aanwijzing voor zijn.

De *Spearman*⁵⁸ correlatie tussen de toetsingsonderdelen onderling en tussen de toetsingsonderdelen en de toegankelijkheidsscore (Variabele "Toegankelijkheid") wordt weergegeven in **Figuur 7** en **Figuur 8**. Corresponderende figuren voor Toegankelijkheidsscore 2 staan in **Bijlage G**. Er bestaan meerdere alternatieve correlatiematen, maar we hebben voor deze correlatiemaat gekozen omdat deze rekening houdt met het feit dat we onder andere met categorische "Voldoet" en "Voldoet niet" -waarden te maken hebben.

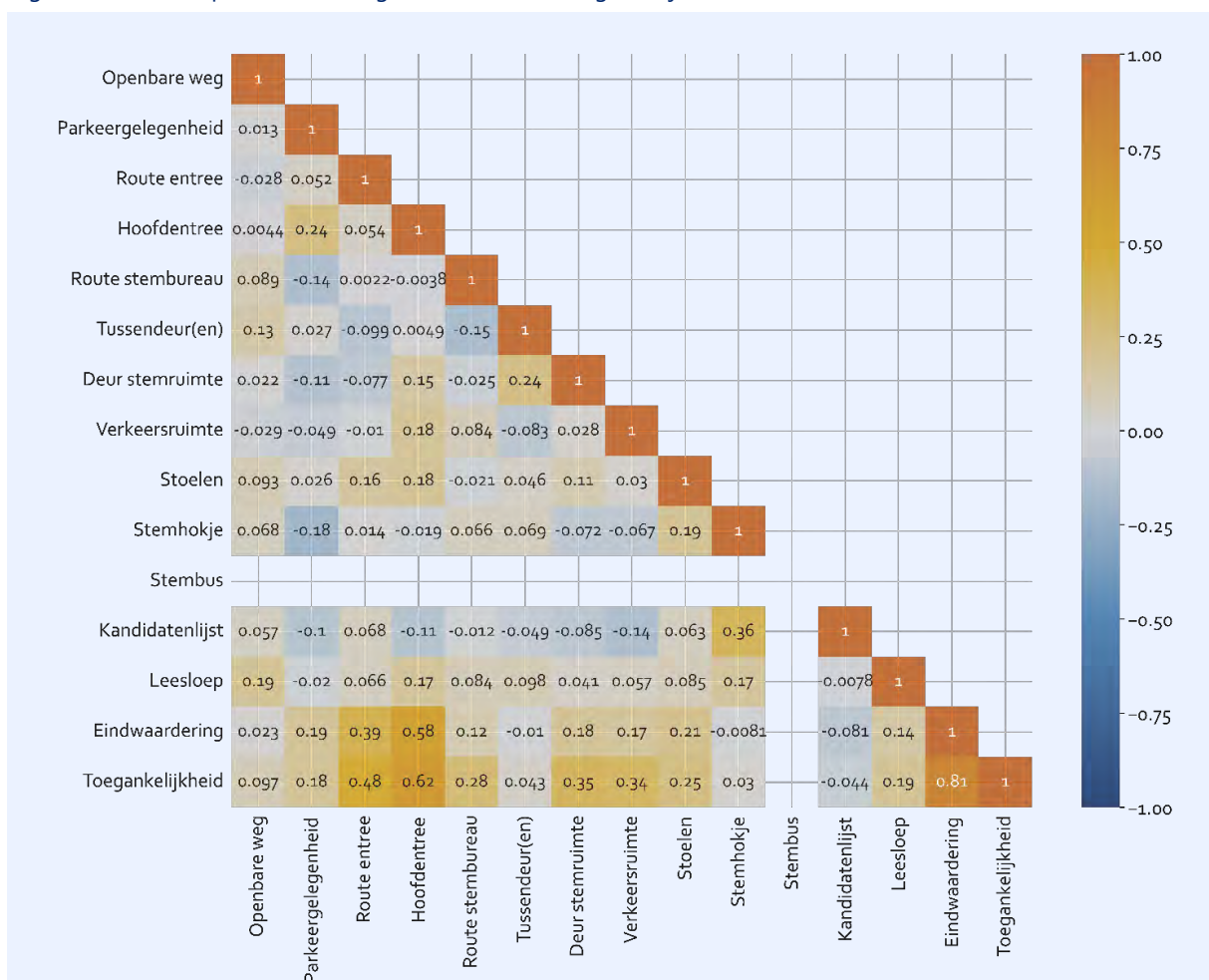
⁵⁸ <https://wikistatistiek.amc.nl/Correlatie>





In de figuren staat donkeroranje (+1) voor maximaal positief (consistent) gecorreleerd, donkerblauw (-1) voor maximaal negatief (tegengesteld) gecorreleerd. Wit (0) staat voor helemaal niet gecorreleerd. CITS onderdeel 3.5 ("Stembus") is niet meegenomen in de correlatieanalyse voor de Tweede Kamerverkiezing omdat alle stemlokalen daar dezelfde maximale score (1,5) op hadden (zie ook **Tabel 4**). Dit was ongeacht hoe zij scoorden op andere onderdelen. Die score was dus geheel onafhankelijk van andere scores en aldus niet daarmee gecorreleerd.

Figuur 7: Correlatieplot van toetsingsonderdelen en toegankelijkheidsscore in november 2023.



Zoals de vele lichte kleuren aangeven, zijn er amper sterke verbanden (van minimaal -0,70 of 0,70) te zien. Sterk is het verband tussen de Eindwaardering ('Voldoet' versus 'Voldoet niet') en de toegankelijkheidsscore, wat valt te verwachten. De toegankelijkheidsscore hangt immers net als de eindwaardering sterk af van het oordeel dat stemlokalen kregen op onderdeel 1.3 t/m 3.2. Om te kijken of er zaken zijn die veel voorkomen bij stemlokalen die een positieve eindwaardering krijgen, kijken we naar welke onderdelen sterk correleren met de Eindwaardering ('Voldoet'/'Voldoet niet'). Voor beide verkiezingen tezamen blijken de hoofdentree en de route naar de entree het belangrijkste voor een positief eindoordeel. Toch zijn de correlaties niet sterk: deze zitten steeds onder de 0,70.

Enkele van die voor toegankelijkheid noodzakelijk geachte onderdelen (de categorische onderdelen 1.3 t/m 3.2, zoals besproken in **Hoofdstuk 2**) correleren met de toegankelijkheidsscore, hoewel niet sterk (onder de 0,70). Dit zijn onderdeel 1.3 ('route entree'), 1.4 ('hoofdentree'), 2.1 ('route stembureau')

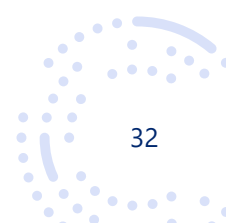
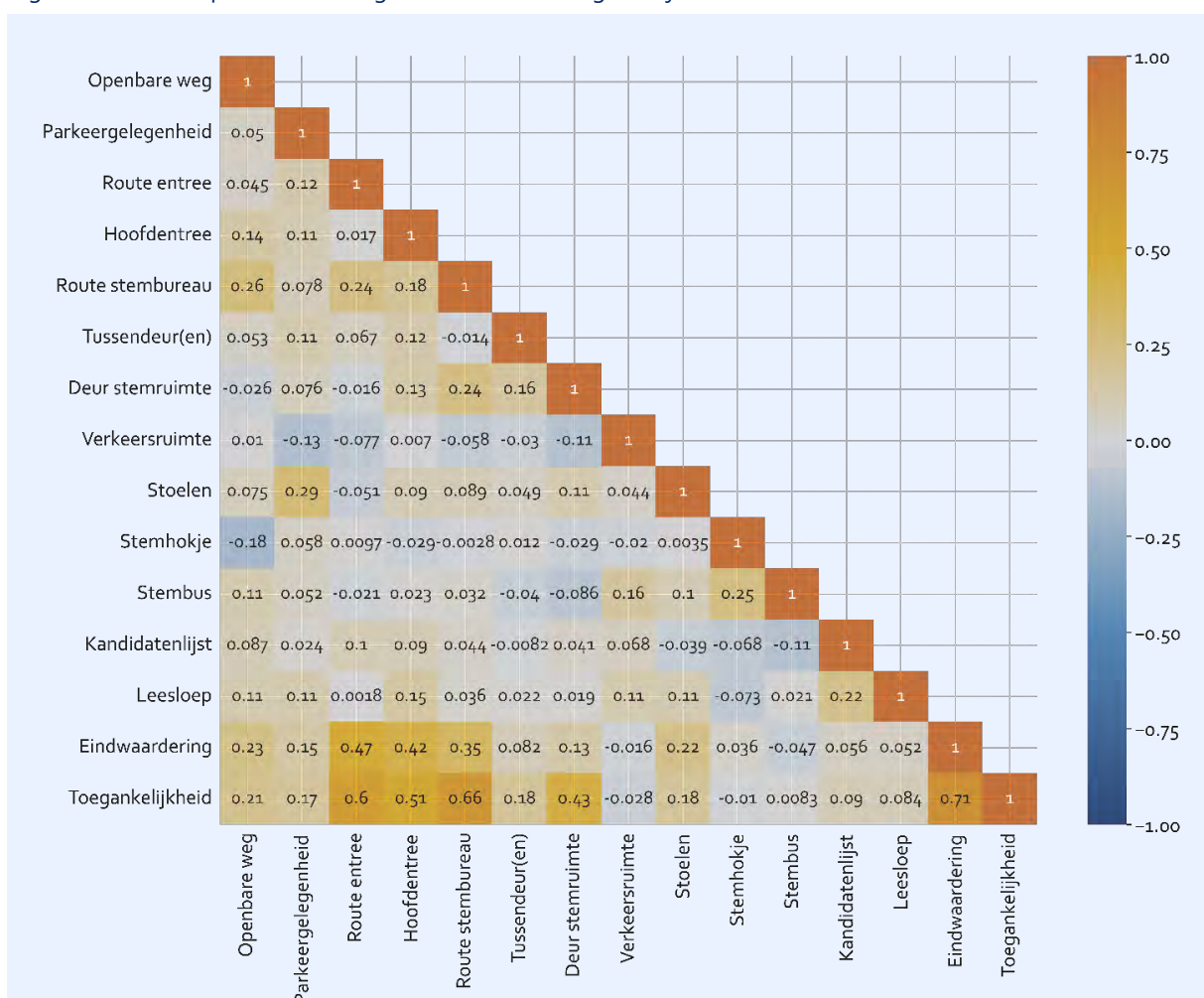


en 3.1 ('deur stemruimte'). Tijdens de verkiezing van november 2023 waren enkele van deze correlaties echter weer wat zwakker: de 'route stembureau' (0,28) en 'deur stemruimte' (0,43). De Toegankelijkheidsscore hangt bij de Tweede Kamerverkiezing net als de Eindwaardering het sterkst af van de beoordeling op "Hoofdentree" (met een correlatie van 0,62). "Route stembureau" was tijdens de PSWS verkiezing het onderdeel met de grootste samenhang met de toegankelijkheidsscore (met een correlatie van 0,66). Dit zijn bescheiden correlaties.

Ook is te zien dat de correlaties tussen de toetsingsonderdelen onderling niet sterk zijn. De sterkste correlatie is tussen "Stemhokje" en "Kandidatenlijst" in november 2023 met een coëfficiënt van 0,36. In maart 2023 is dit tussen "Route stembureau" en "Openbare weg" met een coëfficiënt van 0,26. Dit zijn zwakke correlaties.

Er zijn slechts bescheiden verbanden gevonden tussen de CITS toetsingsonderdelen en de eindbeoordeling en toegankelijkheidsscore. Er zijn ook geen sterke verbanden gevonden tussen de toetsingsonderdelen onderling.

Figuur 8: Correlatieplot van toetsingsonderdelen en toegankelijkheidsscore in maart 2023.





4 Steekproeftrekking

Het onderzoek naar de toegankelijkheid van stemlokalen wordt gedaan op een steekproef van gemeentes. Sinds 2023 worden voor de steekproeftrekking de adviezen van prof. Ernst Wit gevolgd.⁶ Dit betekent dat er voor het selecteren van de stemlokalen verhoudingsgewijs geselecteerd wordt, lettend op de inwonersaantallen binnen provincies en gemeentes. Hiermee wordt er rekening gehouden met een goede afspiegeling van de landelijke situatie.

Net als in de zomer van 2023 is Centerdata door het ministerie van BZK gevraagd om de steekproeftrekking voor de komende verkiezing te doen op de wijze die is omschreven in het rapport van prof. Ernst Wit. Net als de voorgaande keer zijn dezelfde twee beperkingen opgelegd. De eerste is dat de gemeentes die voor de steekproef van de afgelopen verkiezing waren geselecteerd uitgesloten moeten worden van herselectie. Dit om de last te minimaliseren voor gemeentes voor wat betreft het onderzoek naar toegankelijkheid. De tweede beperking is dat de gemeentes per provincie zo dicht mogelijk bij elkaar moeten liggen, zodat het voor controleurs haalbaar is om alle gemeentes op de verkiezingsdag te bezoeken.

Ook deze keer is er een uitzondering vastgesteld op de uitsluitingsregel voor grote gemeentes (met een inwonersaantal van meer dan 150.000). Als die gemeentes in de vorige steekproef zaten, worden zij niet uitgesloten van de huidige steekproef. Dit omdat het aantal grote gemeentes in Nederland relatief laag is. Als grote gemeenten zouden worden uitgesloten van de steekproef zou er een steekproefbias optreden.

Nieuw in deze steekproeftrekking is het verzoek vanuit het ministerie van BZK om tien specifiek genoemde gemeentes bij voorbaat buiten de steekproef te houden. Dit in verband met het feit dat deze gemeentes reeds meedoen aan een ander onderzoek en daarom niet de extra last van deze steekproef kunnen dragen gemeentes.⁵⁹

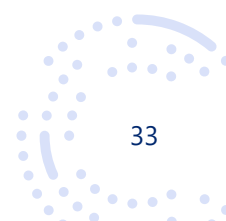
Er is een algoritme geschreven dat de steekproeftrekking automatisch verricht volgens de gespecificeerde wensen, beperkingen en uitzonderingsregels.

4.1 Het algoritme

Het algoritme is op dezelfde wijze gebouwd als bij de voorgaande steekproeftrekking door Centerdata. Een relatief simpel en herbruikbaar algoritme om de steekproeftrekking te doen is geschreven met de programmeertaal Python. Als input vraagt het algoritme het gewenste aantal stemlokalen aan te geven. Ook in dit geval gaat het om 120 stemlokalen. Dit aantal kan te allen tijde worden aangepast. De vervolgstap is dat het algoritme het aangegeven aantal stemlokalen verdeelt onder de 12 provincies naar rato van de populatie van de provincies.⁶⁰

⁵⁹ Het gaat hier om Alphen aan den Rijn, Boekel, Borne, Bladel, Leusden, Midden-Delftland, Oldenzaal, Tynaarlo, Westland en Zoeterwoude.

⁶⁰ De inwonersaantallen van 2023 zijn als basis genomen, maar kunnen in de toekomst aangepast worden door het veranderen van de ingelezen Excel-file met nieuwe inwonersaantallen.





Na het vaststellen van de aantallen per provincie gaat het algoritme verder met het selecteren van gemeentes. Eerst wordt een bestand ingelezen om eerder beoordeelde gemeentes uit te sluiten van de steekproeftrekking, waarbij latere aanpassing naar wens altijd mogelijk is. Grote gemeentes worden niet uitgesloten van deelname om te voorkomen dat er steekproefbias optreedt. In Nederland is er immers maar een beperkt aantal grote gemeentes.

Het algoritme begint met het selecteren van gemeentegroottes⁶¹ op basis van een kansberekening gebaseerd op inwonersaantallen. Zodra een gemeentegrootte is gekozen, wordt uit de subset van mogelijke gemeentes (rekening houdend met grootte, provincie en uitsluitingen) willekeurig een gemeente geselecteerd. Per kleine gemeente worden er twee stemlokalen toegewezen, per middelgrote gemeente vier, en per grote gemeente acht, zoals beschreven in het rapport van prof. Wit.

Een subtiel verschil is dat bij het selecteren van gemeentegroottes wordt gewerkt met waarschijnlijkheden in plaats van afrondingen, zoals in het rapport van prof. Ernst Wit. Dankzij dit principe bestaat er wel een kans op het selecteren van grote gemeentes, zelfs als deze kans klein is.

Een nieuwe toevoeging aan het huidige algoritme is dat binnen dezelfde provincie gemeentes die bij een reeds geselecteerde andere gemeente in de buurt liggen prioriteit krijgen (een hogere selectiekans).

4.2 De steekproeftrekking

De steekproeftrekking resulteert in 34 gemeentes. 12 hiervan zijn klein, 16 middelgroot, en 6 groot.

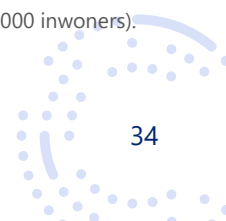
Tabel 5 toont de verdeling van de gemeentes. Eén gemeente komt voor zowel in de nieuwe steekproef als in de in **Hoofdstuk 3** is geanalyseerde steekproef voor de Tweede Kamerverkiezing: Den Haag. Dit is een grote gemeente, waarmee deze tweede verschijning niet in strijd is met de steekproefregels.

Tabel 5: De verdeling van de steekproeftrekking.

Provincie	Aantal stemlokalen	Klein (2x)	Middel (4x)	Groot (8x)	Provincie	Aantal stemlokalen	Klein (2x)	Middel (4x)	Groot (8x)
ZH	26	2	4	1	LB	8	1	2	0
NH	20	0	3	1	FR	4	0	1	0
NB	18	2	2	1	GR	4	0	1	0
GD	14	3	0	1	DR	3	1	1	0
UT	9	1	0	1	FL	3	0	0	1
OV	8	0	2	0	ZL	3	2	0	0

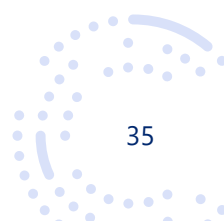
De tweede opgelegde beperking van het ministerie van BZK was dat de gemeentes per provincie zo dicht mogelijk in elkaars buurt moeten liggen, zodat het voor controleurs haalbaar is om alle geselecteerde gemeentes op de verkiezingsdag te bezoeken. Omdat per provincie het aantal niet-uitgesloten gemeentes beperkt is, is er weinig speling voor deze extra beperking. Algoritmisch gezien is er geen geprogrammeerde limitatie opgelegd qua afstand tussen de geselecteerde gemeentes.

⁶¹ Er zijn drie gemeentegroottes, klein (< 40.000 inwoners), middel (40.000 – 150.000 inwoners), en groot (> 150.000 inwoners).





De lijst met gemeentes van de steekproeftrekking voor de aankomende verkiezing is als dataset opgeleverd op 19 maart 2024.





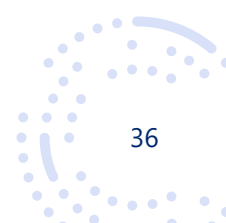
Bijlagen

A Toegankelijkheidsscore 2 – Beschrijvende statistiek

Hier worden enkele beschrijvende statistieken op de scores van onze alternatieve maat (Toegankelijkheidsscore 2) weergegeven.

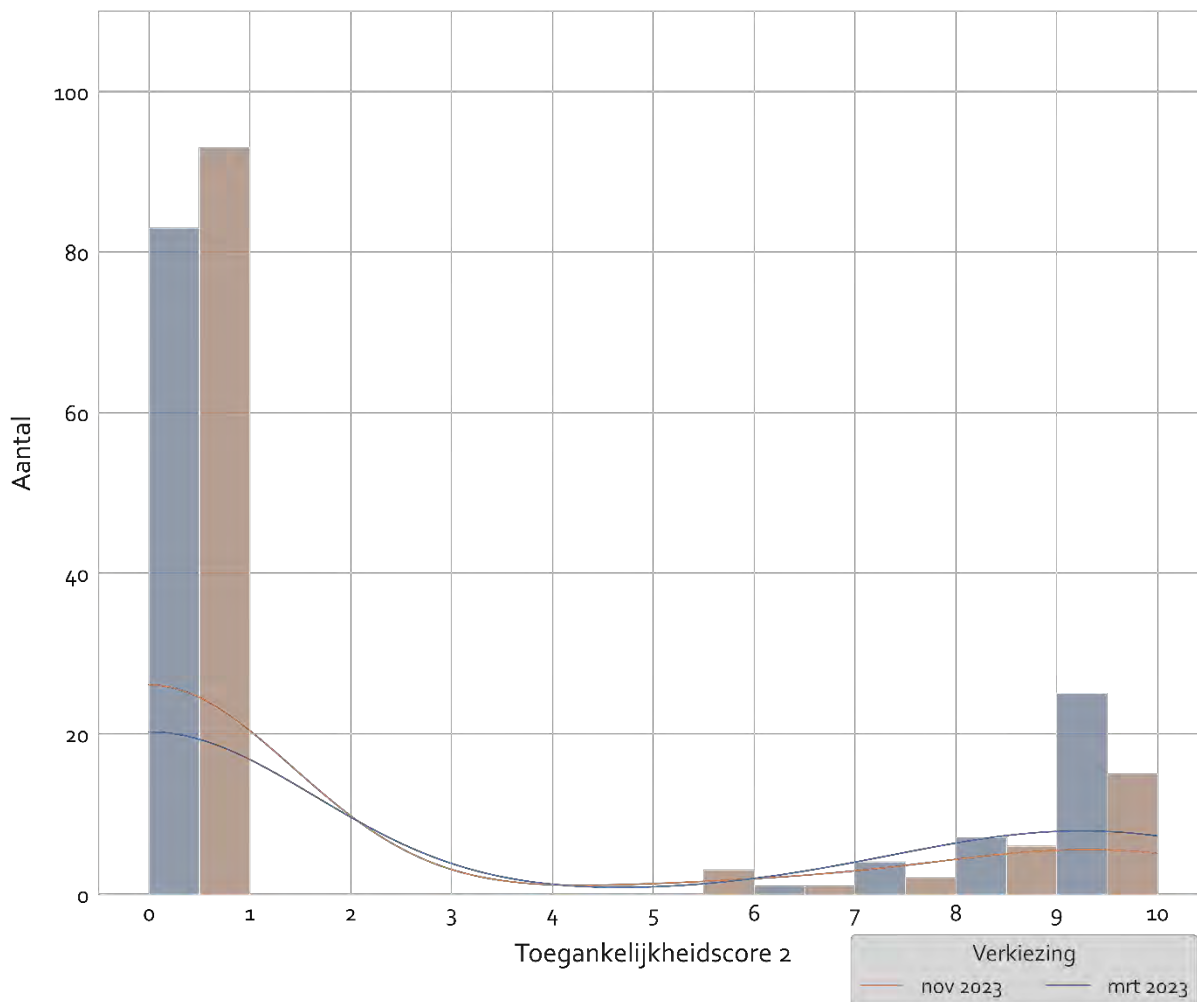
Maan d	Alle observaties	Alle observaties	Alleen 'Voldoet'	Alleen 'Voldoet'
	Gemiddelde Toegankelijkheidsscor e 2	Mediane Toegankelijkheidsscor e 2	Gemiddelde Toegankelijkheidsscor e 2	Mediane Toegankelijkheidsscor e 2
Mrt	1,95 (3,71)	0	8,95 (1,24)	9,2
Nov	2,82 (4,27)	0	9,14 (0,91)	9,5

De standaarddeviatie is tussen haakjes weergegeven.



B Toegankelijkheidsscore 2 – Scores per verkiezing

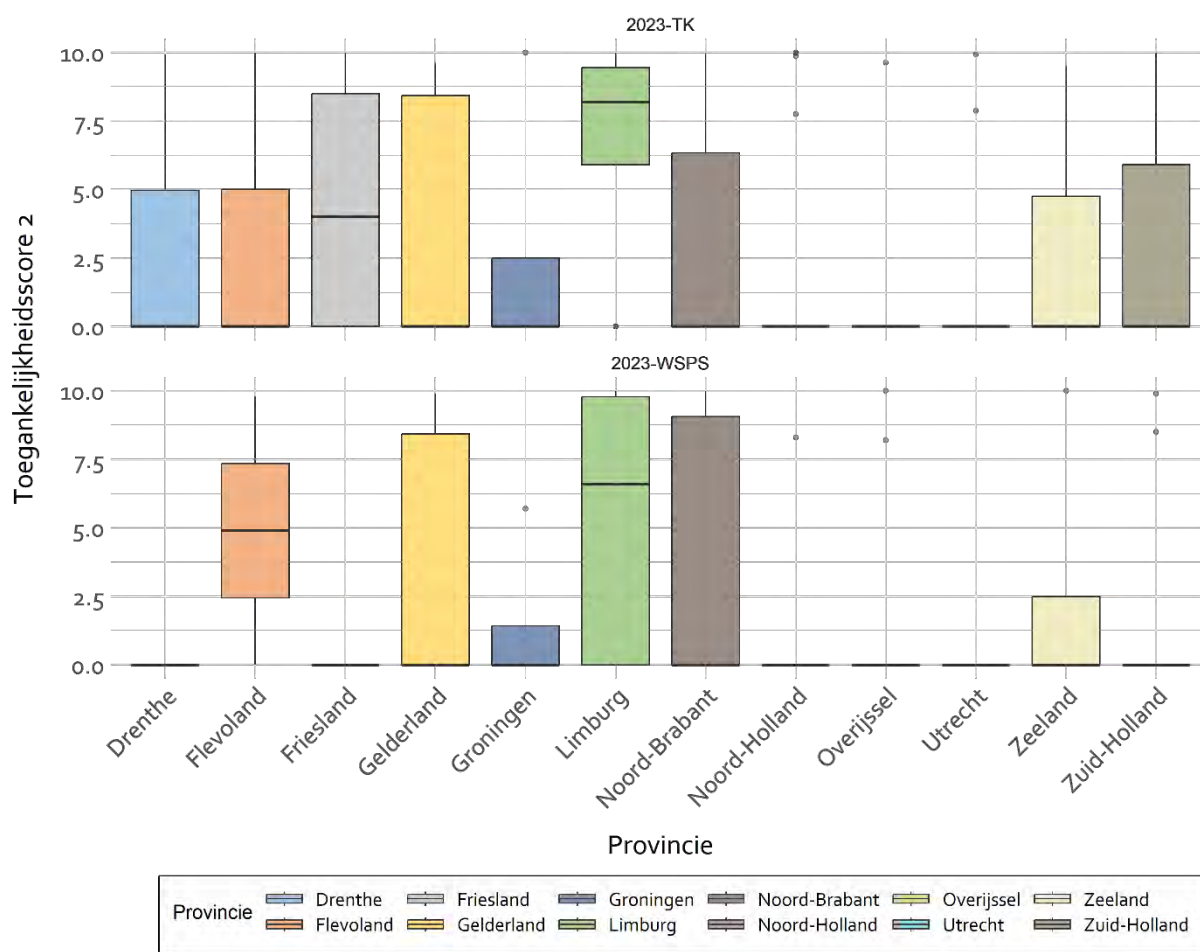
Hier wordt de verdeling van scores getoond op onze alternatieve maat (Toegankelijkheidsscore 2), per verkiezing in 2023.





C Toegankelijkheidsscore 2 – Scorevariatie per provincie

Hier wordt de verdeling van scores getoond op onze alternatieve maat (Toegankelijkheidsscore 2), per provincie en verkiezing in 2023.

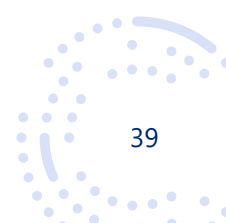




D Toegankelijkheidsscore 2 – Beschrijvende statistiek van de drie gemeentegroottes

Hier worden beschrijvende statistieken gegeven voor de drie gemeentegroottes op onze alternatieve maat (Toegankelijkheidsscore 2), per verkiezing in 2023.

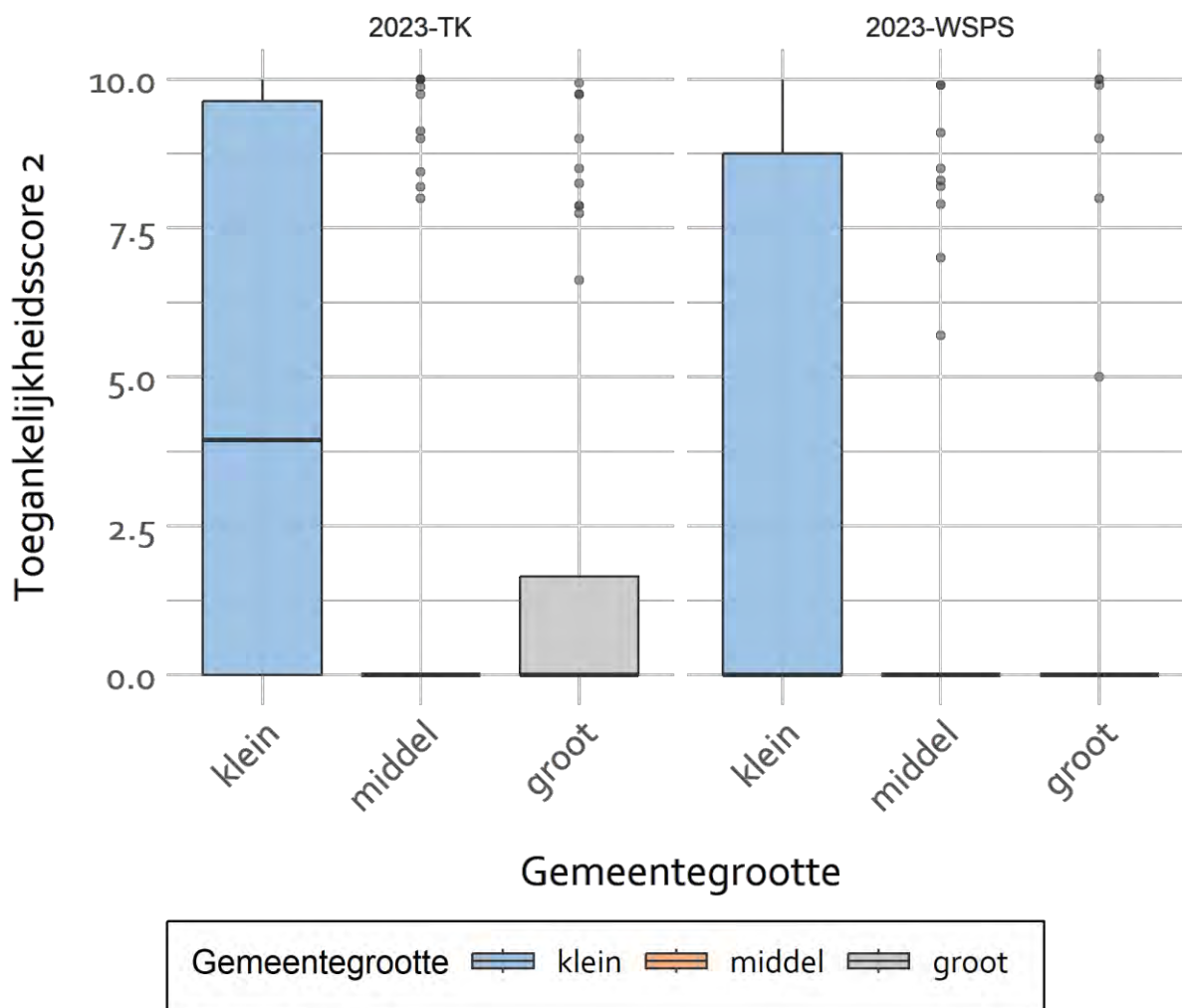
Gemeentegrootte (aantal inwoners)	Verkiezing	Aantal stemlokalen	Toegankelijk- heidsscore	Standaard- deviatie	Mediaan
Klein	PSWS-2023	35	3,08	4,81	0
(< 40.000)	TK-2023	32	4,70	3,97	3,94
Middel	PSWS-2023	55	1,36	3,78	0
(40.000 – 150.000)	TK-2023	48	2,13	4,41	0
Groot	PSWS-2023	30	1,73	3,13	0
(> 150.000)	TK-2023	40	2,13	3,61	0





E Toegankelijkheidsscore 2 – Scorevariatie van de drie gemeentegroottes

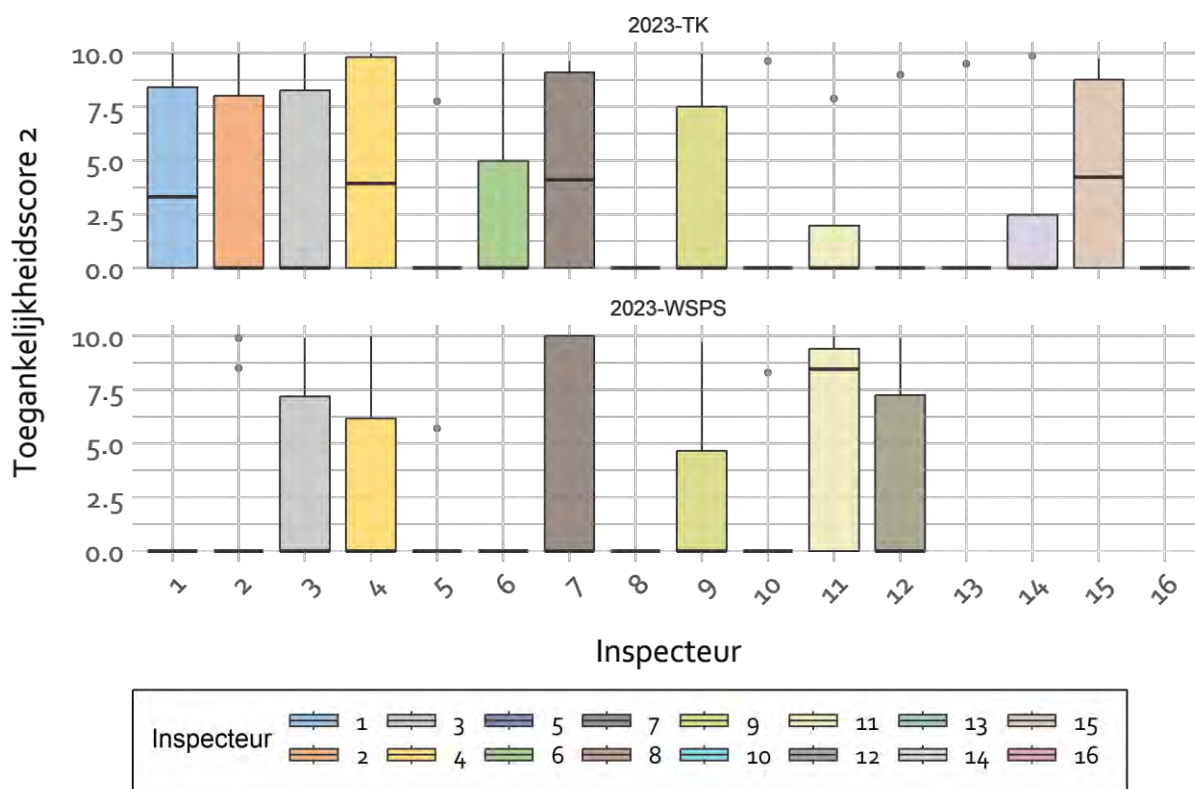
Hier wordt de verdeling van scores getoond op onze alternatieve maat (Toegankelijkheidsscore 2), per gemeentegrootte en verkiezing in 2023.





F Toegankelijkheidsscore 2 – Scorevariatie van de controleurs

Hier wordt de verdeling van scores getoond op onze alternatieve maat (Toegankelijkheidsscore 2), per controleur en verkiezing in 2023.

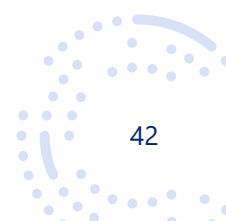
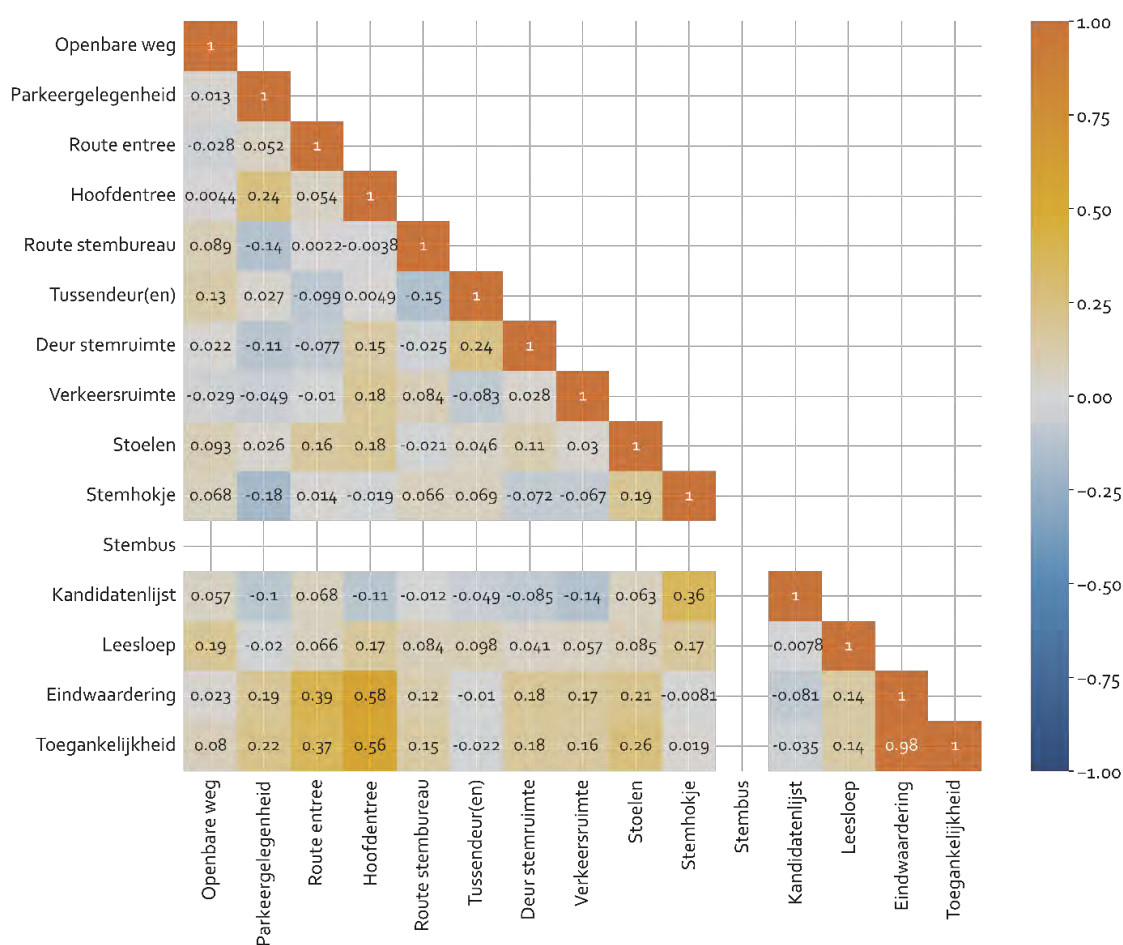




G Toegankelijkheidsscore 2 – Correlatieplots

Onderstaande geeft de correlaties tussen de CITS toetsingsonderdelen, alsmede tussen deze onderdelen en de Eindwaardering en Toegankelijkheidsscore 2. De correlaties tussen de toetsingsonderdelen zelf zijn gelijk aan die in Figuur 6 en 7, omdat de scores op de individuele toetsingsonderdelen waaruit de toegankelijkheidscores??? zijn berekend, hetzelfde zijn. De correlaties met Toegankelijkheidsscore 2 zijn het relevantst om hier te bekijken.

Tweede Kamerverkiezing





PSWS verkiezing

