

Korrigerte tall for nasjonale prøveresultater over tid

Elevers ferdigheter målt ved nasjonale prøver endret seg mer mellom 2014 og 2021 enn statistikken tidligere har vist. Dette fremgår av Statistisk sentralbyrås korrigeringer basert på nye estimater på skalapoeng fra Utdanningsdirektoratet.

Publisert: 13. november 2025

Forfatter: [Kjartan Steffensen](#)

Statistisk sentralbyrå (SSB) er ansvarlig for den offisielle statistikken over nasjonale prøver. Dataene hentes fra Utdanningsdirektoratet (Udir), og ferdighetsnivå måles i såkalte skalapoeng. En feil i beregningen av disse poengene for årene 2014–2021 ble kommunisert på Udirs nettsider i 2024, se [Feil i resultater fra nasjonale prøver 2014–2021 \(udir.no\)](#). Udir ble gjort oppmerksom på feilen av forskere ved Frischsenteret, som siden har publisert artikkelen [Skoleferdigheter i endring: Utviklingen over tid målt ved nasjonale prøver \(frischsenteret.no\)](#). Etter at feilen ble oppdaget har Udir beregnet nye skalapoeng for nasjonale prøver på 5. og 8. trinn. SSB har utnyttet disse nye dataene til å vurdere muligheten for å korrigere publisert statistikk, se notater 2025/32, [Reestimering av nasjonale prøveresultater 2014 – 2021](#). Som et resultat av dette arbeidet er statistikken nå oppdatert med korrigerte tall som gir et riktig bilde av hvordan elevene har prestert på nasjonale prøver over tid. I denne artikkelen går vi nærmere inn på hva som var feil i de tidligere beregningene for nasjonale prøveresultater og hvordan statistikken er blitt korrigert. Vi ser også på hva de korrigerte tallene sier om utviklingen over tid i perioden fra 2014 fram til de nasjonale prøvene ble endret i 2022.

Skalapoeng 2014-2021

Udir har valgt å betegne estimatet på ferdighetsnivået til en elev på nasjonale prøver (2014 – 2021) for skalapoeng. Ved refereanseåret til et prøveemne, ble skalapoengene standardisert slik at de hadde et snitt på 50, og standardavvik på 10. For regning og engelsk ble snittet og standardavviket satt i 2014, mens for lesing ble det satt i 2016. Skalapoeng er relativt til skalapoengene fra tidligere år. Dermed kan snittet og standardavviket (spredningen) til skalapoengene endres over tid, i takt med endringer i ferdighetsnivå.

Hva var feil i de tidligere beregningene for nasjonale prøveresultater?

Når nasjonale prøver ble innført i 2004 ble prøvene brukt for å gi et øyeblikksbilde av skoleelevenes ferdigheter i lesing, regning og engelsk. Resultatene for en gitt elev kunne ses i sammenheng med andre elevers skårer innenfor samme prøve, men kunne ikke si noe om ferdighetsutviklingen over tid. Fra og med 2014 (for prøvene i regning og engelsk) og fra og med 2016 (for prøvene i lesing), endret Udir metodikken for å kunne sammenligne prøveresultatene fra år til år, og dermed fange opp utvikling i elevenes ferdigheter over tid. For å kunne si noe om utvikling over tid selv om oppgavene i nasjonale prøver endret seg, benyttet Udir ankeroppgaver. Disse ankeroppgavene ble gitt til et tilfeldig utvalg blant elevene (ankerelevne) som tok de nasjonale prøvene. Ankeroppgavene var oppgaver som ble repetert i flere årganger. Ved å bruke resultater fra første år med ankeroppgaver i 2014 kjente man til vanskelighetsgrad og hvor godt ankeroppgavene skilte mellom elever på ulike ferdighetsnivå. Kombinert med hvordan ankerelever gjorde det på prøvene påfølgende år hadde man da et grunnlag for å si noe om ferdighetsnivået blant elevene hadde endret seg sammenlignet med 2014.

Analyseverktøyet Xcalibre som ble benyttet for å beregne skalapoeng (mål på ferdighetsnivå på nasjonale prøver) hadde imidlertid en feilspesifisering som ikke ble oppdaget. Dette gjorde at verktøyet gikk ut fra at elevkullene i ulike år var fordelt utover ferdighetsskalaen på samme måte fra år til år. Resultatet ble at gjennomsnittlig skalapoeng i statistikken feilaktig endret seg minimalt over tid i perioden fra 2014 til 2021.

På individnivå hadde ikke feilen lik effekt for alle elevene på nasjonale prøver. Kortfattet førte det til at forholdet mellom ankerelevne og de øvrige elevene ble feilestimert. Relativt til hverandre/gjennomsnittet fikk ankerelevne en overestimering i takt med positive endringer, mens de øvrige elevene fikk en underestimering. I tilfeller hvor det var en reduksjon i skalapoeng over tid så man den motsatte effekten. Siden det var langt færre ankerelever enn øvrige elever, var effekten mer markant for ankerelevne enn de øvrige elevene. I hovedsak balanserte disse effektene hverandre ut, slik at overestimeringen til ankerelevne og underestimeringen til de øvrige elevene utjevnet hverandre. Dermed var det liten endring i snittet for hele populasjonen, tross endringer blant ankerelevne og de øvrige elevene.

Det er viktig å notere seg at ankerelevne ikke er særlig over- eller underrepresentert i grupperinger som benyttes i den offisielle statistikken, slik som f.eks. fylke, foreldres utdanningsnivå, kjønn og innvandringskategori. Det var individuelle forskjeller mellom ankerelever og øvrige elever, samt endringer over tid som ble mest påvirket. Forholdet mellom grupperinger av elever innad samme år ble kun berørt i mindre grad.

Hvordan er statistikken blitt korrigert?

Som nevnt har Udir beregnet nye skalapoeng for nasjonale prøver på 5. og 8. trinn, og SSB har innhentet data fra Udir som har elevenes resultater på oppgavenivå og de nye skalapoengene. Udir har imidlertid ikke hatt juridisk hjemmel til å beholde personidentifiserende informasjon i datagrunnlaget som skalapoeng beregnes ut ifra. Dette betyr at dataene som SSB har hentet inn fra Udir med nye skalapoeng ikke inneholder kjennetegn som fødselsnummer. Dermed har vi ikke kunnet koble disse direkte til SSBs gamle data som danner grunnlaget for statistikken, og der SSB har andre kjennetegn om elevene som for eksempel foreldres utdanningsnivå, skolefylke, og innvandringskategori. Det har dermed vært nødvendig for SSB å se på mulige metoder for å knytte nye skalapoeng til våre gamle datasett selv om vi ikke kan gjøre en direkte kobling mellom datasettene. En mer teknisk og omfattende forklaring av det metodiske arbeidet SSB har gjort, fremgår av notater 2025/32, [Reestimering av nasjonale prøveresultater 2014 – 2021](#).

Notatet konkluderer med en anbefaling om at en enkel omskalering fra gamle til nye skalapoeng har tilstrekkelig presisjon til å korrigere de gamle resultatene, på grupperingene benyttet i de offisielle statistikkene. En slik omskalering av gamle skalapoeng gjøres slik at både gjennomsnittet (μ) og standardavviket (σ) for hele populasjonen blir likt de nye resultatene fra Udir, for hver prøve og hvert år. Formelen nedenfor viser hvordan dette er beregnet:

$$\hat{\theta}_n = \frac{\sigma_{\theta_n}}{\sigma_{\theta_g}} (\theta_g - \mu_{\theta_g}) + \mu_{\theta_n}$$

I formelen er estimatet på nye skalapoeng θ_n , våre gamle skalapoeng θ_g og de nye skalapoengene slik de er innhentet fra Udir θ_n . Videre er μ_{θ_n} og σ_{θ_n} gjennomsnittet og standardavviket til θ_n , og μ_{θ_g} og σ_{θ_g} er gjennomsnittet og standardavviket til θ_g . Dette antar at feilen i den opprinnelige beregningen av skalapoeng i hovedsak påvirket snittet og standardavviket til utvalget, og kun i mindre grad rangeringer av skalapoeng i dataene. Selv om det kan være en relativt stor forskjell mellom nye og gamle skalapoeng for et gitt individ, vil forskjellen mellom individer være relativt like ved både nye og gamle skalapoeng.

Det forventes systematiske forskjeller mellom ankerelever og øvrige elever, men ikke særlig endring i rangering innad disse. Forskjellene mellom ankerelevne og de øvrige elevene er derimot relativt små, og gir ikke nevneverdige utslag på statistikken. Det tidligere nevnte metodenotatet understøtter dette gjennom en analyse av hvordan skjevheter mellom ankerelevne og øvrige elever kan slå ut på samsvaret mellom våre nye omskalerte skalapoeng og de faktiske nye skalapoengene beregnet av Udir. Slike utslag vil primært kunne ha noe å si om man ser på små grupper med en håndfull elever, og selv da vil skjevhetene være av begrenset betydning.

En fordel ved å benytte den globale omskaleringsmetoden er at man ikke er avhengig av å ha tilgang til et nytt datasett fra SSB. En ekstern forsker kan for eksempel reestimere skalapoengene i sitt eget datasett, kun med tilgang til fire verdier for en gitt prøve et gitt år (μ_{θ_g} , μ_{θ_n} , σ_{θ_g} , σ_{θ_n}). Videre må ikke metoden anvendes på individnivå, men kan også anvendes på aggregerte data. De fire verdiene som trengs for omskalering er gjengitt i tabellen nedenfor.

→ **Tabell 1.****Omskaleringsgrunnlag fra gamle til nye skalapoeng**

SSB har altså nå publisert korrigerte tidsserier 2014-2021 for nasjonale prøver 5. og 8. trinn i vår statistikkbank og gir med tabellen over forskere og andre brukere muligheten til selv å gjøre omskalering på gamle data. Videre vil Nasjonal utdanningsdatabase bli oppdatert med korrigerte skalapoeng, og korrigerte mikrodata vil bli gjort tilgjengelig for forskere gjennom microdata.no og ordinært utlån av mikrodata.

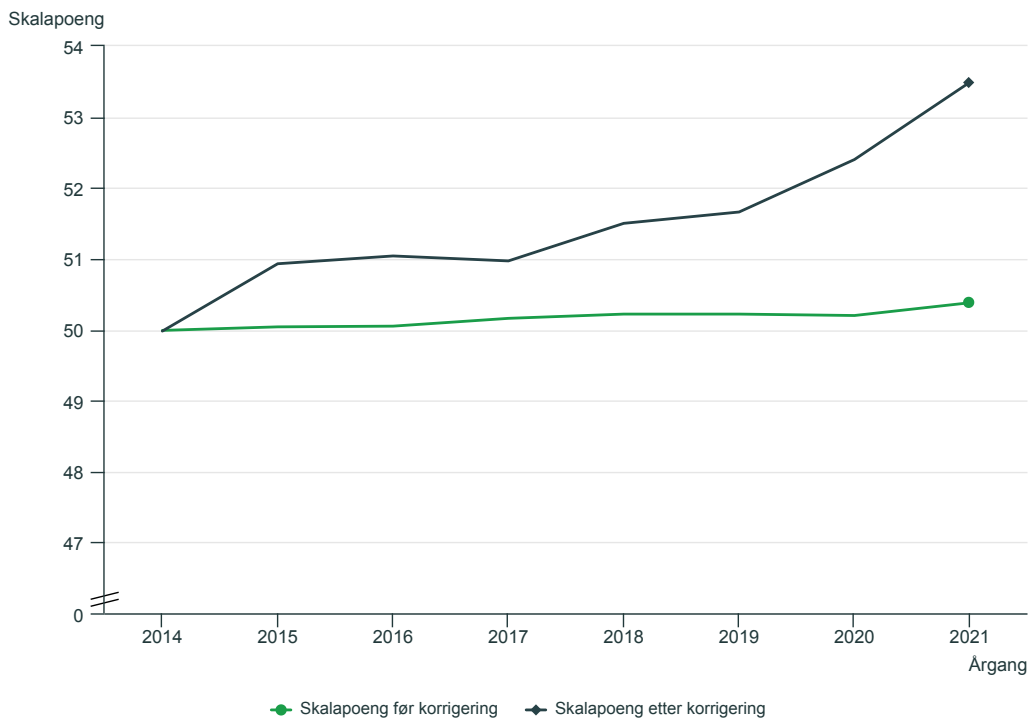
Hva kan vi nå si om utviklingen over tid i resultater på nasjonale prøver?

Ved å basere statistikken på de korrigerte skalapoengene for nasjonale prøver finner vi at det fra 2014 til 2021 har vært endringer over tid som ikke tidligere var fanget opp i statistikken. Korrigerte skalapoeng for perioden på alle de nasjonale prøvene for 5. og 8. trinn i perioden vises i statistikkbanktabellene [13718](#) og [13719](#), både for hele landet og for hvert fylke. Forskjellen mellom skalapoeng før og etter korrigering er størst for de nasjonale prøvene i engelsk.

Både på 5. og 8. trinn var ferdighetene i engelsk målt ved nasjonale prøver klart bedre mot slutten av perioden enn i 2014, en økning på over 3 skalapoeng for hver av prøvene.

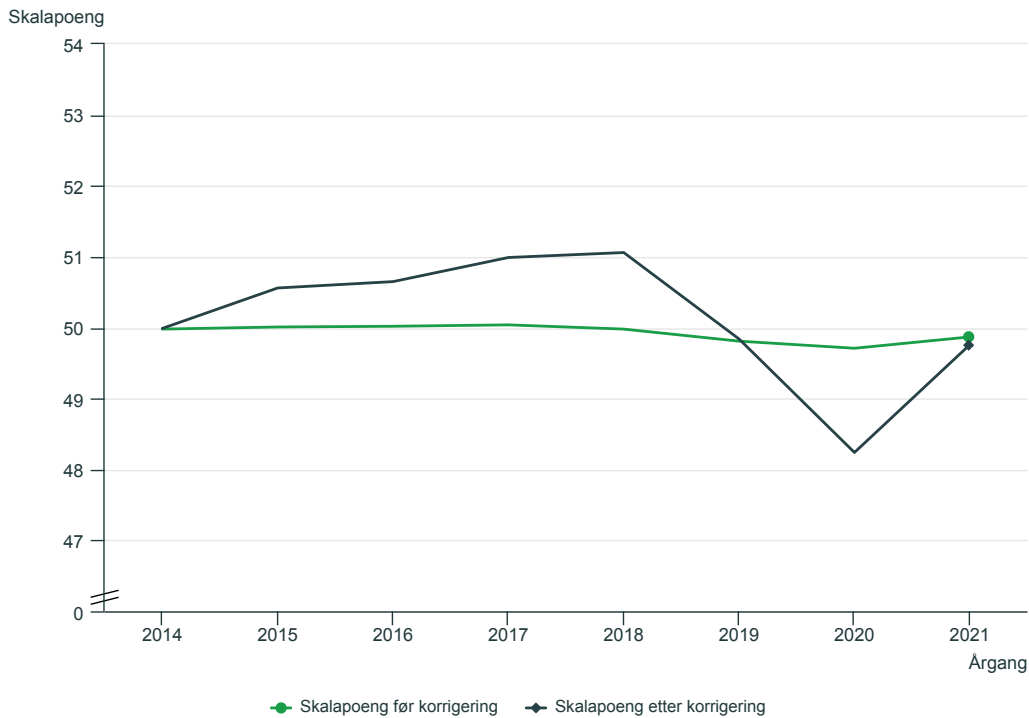
Figur 1. Gjennomsnittlig skalapoeng nasjonale prøver i engelsk 5. trinn før og etter korrigering. 2014-2021

Kilde: [Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå](#)

Figur 2. Gjennomsnittlig skalapoeng nasjonale prøver i engelsk 8. trinn før og etter korrigering. 2014-2021

Kilde: Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå

For regning er utslagene av korrigeringene i skalapoeng mindre enn for engelsk, men elevenes ferdigheter i regning har også variert noe mer over tid enn det statistikken tidligere har vist. Både på 5. og 8. trinn var ferdighetene i regning noe bedre i de første årene etter 2014, og i 2018 lå ferdighetsnivået til elevene i gjennomsnitt rundt 1 skalapoeng høyere enn i 2014. Fra 2018 til 2020 sank imidlertid elevens gjennomsnittlige ferdigheter i regning på 5. og 8. trinn til et nivå som lå lavere enn da målingene startet i 2014. For regning 5. trinn økte imidlertid ferdighetene fra 2020 til 2021, og lå da på omtrent samme nivå som i 2014.

Figur 3. Gjennomsnittlig skalapoeng nasjonale prøver i regning 5. trinn før og etter korrigering. 2014-2021

Kilde: Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå

Figur 4. Gjennomsnittlig skalapoeng nasjonale prøver i regning 8. trinn før og etter korrigering. 2014-2021

Kilde: Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå

Muligheten for å måle trend i nasjonale prøver i lesing ble først innført fra 2016. For lesing 5. trinn viser de korrigerte skalapoengene at elevenes gjennomsnittlig ferdigheter holdt seg på 2016-nivå i 2017 for deretter å synke over tid med nær 2 skalapoeng frem til

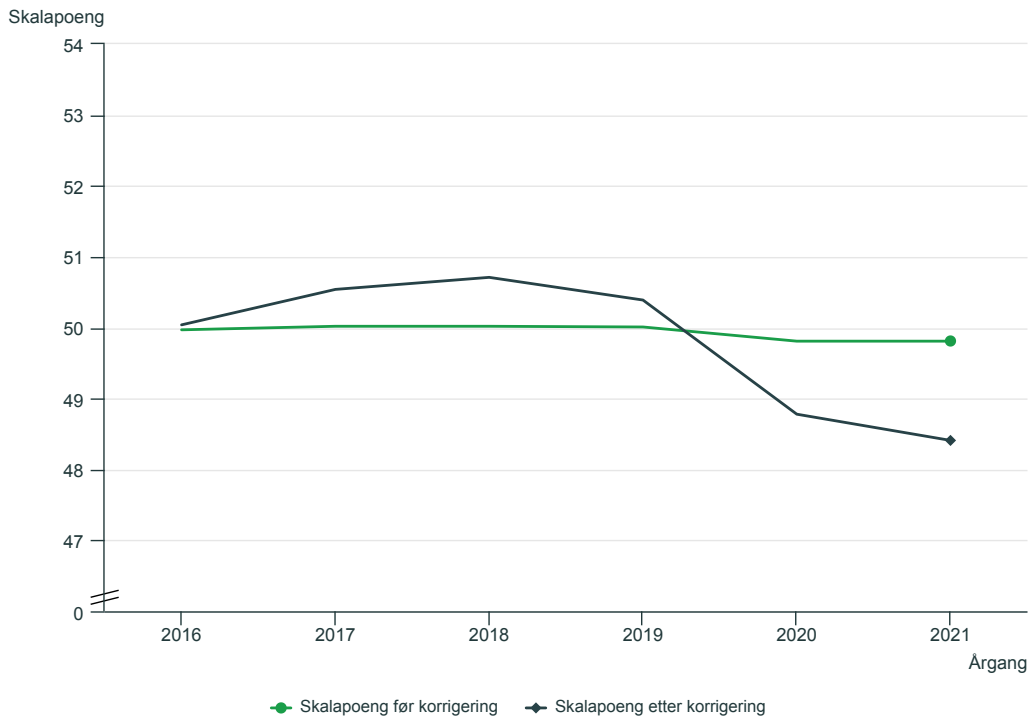
2020. Deretter var det en forbedring fra 2020 til 2021. Elevenes ferdigheter målt ved utgangen av perioden vi kan lage en sammenlignbar tidsserie for, altså 2016 til 2021, var likevel lavere enn ved starten av perioden.

I lesing 8. trinn hadde elevene et noe høyere ferdighetsnivå i perioden 2017 til 2019 enn i 2016, men ferdighetsnivået sank deretter tydelig til 2020 og videre til 2021 hvor gjennomsnittlig skalapoeng for elevene lå mer enn halvannet skalapoeng lavere enn i 2016.

Figur 5. Gjennomsnittlig skalapoeng nasjonale prøver i lesing 5. trinn før og etter korrigering. 2016-2021



Kilde: Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå

Figur 6 Gjennomsnittlig skalapoeng nasjonale prøver i lesing 8. trinn før og etter korrigering. 2016-2021

Kilde: Karakterer og nasjonale prøver i grunnskolen, Statistisk sentralbyrå

Fra 2022 startet måling av utvikling over tid på nytt i de nasjonale prøvene. Resultatene fra og med 2022 kan derfor ikke direkte sammenlignes med tidligere år. Korrigeringen av skalapoengene for perioden 2014-2021 har like fullt gitt oss en nyttig forståelse av at elevers ferdigheter målt ved nasjonale prøver endret seg mer mellom 2014 og 2021 enn statistikken tidligere har vist. Videre har erfaringene med feilen i nasjonale prøver ført til en rekke forbedringstiltak i arbeidet med statistikken. Udir har etter omlegging til ny beregningsmetode for nasjonale prøver fra 2022 gått over til å åpent tilgjengeliggjøre R-kode og teknisk dokumentasjon av beregningene som gjøres. For å sikre langtidslagring av grunnlagsdata har SSB fra og med fjorårets innhenting for nasjonale prøver også hentet inn data på oppgavenivå med personidentifiserende informasjon før disse slettes hos Udir. Dette bidrar til helhetlig lagring og dokumentasjon i produksjonskjeden for den offisielle statistikken.

Kontakt

Samuel Jones Gilje

samuel.jones.gilje@ssb.no

[41 37 27 58](tel:41372758)

Rachel Ekren

rachel.ekren@ssb.no

[95 83 89 67](tel:95838967)

Ingeborg Kool

ingeborg.kool@ssb.no

[94 46 42 45](#)