

Språkrådets test av skriveprogrammer på norsk

Rapport

Språkrådet

Språkrådets konklusjoner, kort oppsummert:

- Ingen av de mest brukte skriveprogrammene i norsk skole innfrir kravene som opplæringsloven (2024) stiller til skriveprogrammer.
- Skriveprogrammene fungerer (jevnt over) dårligere på nynorsk enn på bokmål.
- Skriveprogrammene overser mange feil og markerer mange korrekt skrevne ord som feil.
- Skriveprogrammene foreslår ofte alternativer som ikke er riktige, ved røde og blå markeringer.
- Skriveprogrammene har, med få unntak, ikke blitt bedre i tidsrommet mellom Språkrådets to tester (i januar og juni 2025).

Introduksjon

Språkrådet er statens forvaltningsorgan i språkspørsmål og følger opp den norske språkpolitikken på oppdrag fra Kultur- og likestillingsdepartementet (KUD).

I den nye opplæringsloven, som trådte i kraft 1. august 2024, er det tatt inn et eget punkt om skriveprogrammer, § 15-4 andre ledd (Opplæringsloven 2024):

Skolen skal bruke skriveprogram som støttar både bokmål og nynorsk, og som følger offisiell rettskriving.

Loven stiller imidlertid ingen krav om formell godkjenning av skriveprogrammer som skal brukes i skolen, slik tilfellet er for norskspråklige ordlister og ordbøker. Slike ressurser skal etter lovens § 15-4 første ledd godkjennes av Språkrådet.

For 2025 har Språkrådet fått følgende oppdrag fra KUD:

[...] bidra til at aktører som utvikler og tilbyr språkteknologiske verktøy og tjenester, er kjent med de kravene som stilles til slike verktøy og tjenester i offentlig bruk (Tildelingsbrev 2025)

[...] bidra til at de får innspill og dokumentasjon som bidrar til at produktene de tilbyr, har en tilfredsstillende språklig kvalitet (Tildelingsbrev 2025)

På bakgrunn av kravet i opplæringsloven og Språkrådets oppdrag slik det er formulert i tildelingsbrevet fra KUD, har Språkrådet per august 2025 gjennomført testing av stave- og grammatikkontrollen i de tre skriveprogrammene som til sammen har om lag 90 prosent av markedet i den norske skolen (LNK-undersøkelsen 2024). Hensikten med

testen har vært å finne ut om skriveprogrammene oppfyller opplæringslovens krav. Målet med å gjennomføre testingen har vært todelt:

1. Språkrådet bidrar til at programmene blir bedre ved å opplyse leverandørene av programmene om lovkravene og hva de innebærer, samt om utfordringene de enkelte programmene har.
2. Språkrådet opplyser skoler og myndigheter (de som lovkravene er rettet mot) om hvor godt eller dårlig programmene fungerer, slik at det kan tas opplyste valg som ikke resulterer i at opplæringsloven brytes. Dermed kan både bokmåls- og nynorskelever få best mulige skriveprogrammer.

Spesifisering av kravet i opplæringsloven § 15-4 andre ledd

Språkrådet legger følgende forståelse av ordlyden i lovbestemmelsen til grunn:

- At skriveprogrammene støtter både bokmål og nynorsk, betyr at funksjonene i skriveprogrammene skal virke like godt på begge språkene. Både bokmål og nynorsk må være med i lista over korrekturspråk.
- At skriveprogrammene følger offisiell rettskriving, betyr at de
 - følger reglene for staving og bøyning av ord
 - følger formelle skriveregler om tegnsetting, bruk av store og små bokstaver m.m.
 - følger norske regler for skriving av sammensatte ord

For stavekontrollen i skriveprogrammene betyr dette nærmere bestemt

- at kontrollen må markere alle feil i staving og bøyning, alle brudd på formelle skriveregler om tegnsetting og bruk av store og små forbokstaver m.m. og alle brudd på norske regler for skriving av sammensatte ord
- at kontrollen *ikke* skal markere skrivemåter som er godtatt i normen, som feil, heller ikke skrivemåter som er uvanlige
- at nynorskkontrollen ikke skal akseptere bokmålsord i nynorsktekst, og vice versa
- at minst ett av de alternative forslagene som kommer opp når programmene markerer noe som feil, skal være korrekt

Verken lovteksten eller forarbeidene til loven presiserer nærmere hvor godt et skriveprogram må fungere for at opplæringslovens krav skal være oppfylt. Språkrådet oppfatter ordlyden «følger offisiell rettskriving» som et absolutt krav, det vil si at alle feil skal oppdages og markeres, og at programmene ikke skal feilmarkere der det ikke er noen feil. Når Språkrådet vurderer ordbøker og ordlister til godkjenning for skolebruk,

påtaler vi alle feil vi finner, og forventer at de blir rettet opp. Det er en forutsetning for at disse ressursene skal få godkjenning. Nå ser det ut til at skriveprogrammene langt på vei er i ferd med å erstatte ordbøkene som elevenes «fasit» for rettskriving. Vi mener derfor det ikke er rimelig å stille lavere krav til kvaliteten på skriveprogrammer enn til kvaliteten på ordbøker, selv om det ikke stilles formelle krav om godkjenning av skriveprogrammer som skal brukes i skolen.

Metode

Vi har testet stavekontrollen i tre skriveprogrammer, fordi disse er de mest brukte i skolen i dag (LNK-undersøkelsen 2024). Disse er utviklet og levert av Microsoft, Apple og Google (heretter kalt *leverandørene*):

- Microsoft Word
- Apple Pages
- Google Docs

Basert på kriteriene for en vellykket stavekontroll slik de er skissert ovenfor, ville vi måle

- hvor god stavekontrollen er til å identifisere faktiske feil
- hvor ofte forslagene fra stavekontrollen er riktige
- hvor god stavekontrollen er til å ikke feilmarkere korrekt språk

Vi laget korte testtekster som vi matet inn i skriveprogrammene. Noen av tekstene hadde med hensikt høy konsentrasjon av feil, slik at vi kunne se om stavekontrollene identifiserte feilene og kom med gode forslag til alternativer. Andre tekster inneholdt korrekte former som vi mistenkte kunne være utfordrende for stavekontrollen å identifisere, fordi formene er mindre vanlige i bruk. Vi kaller disse for «feller».

Resultatene i rapporten her er sammenlikninger av markeringene fra stavekontrollene opp mot fasit.

Vi kunne testet flere aspekter ved stavekontrollene, for eksempel hastighet eller brukertilfredshet. Det har vi valgt å ikke inkludere i denne undersøkelsen, blant annet fordi det ikke er relevant for lovkravet.

Automatisk og menneskelig evaluering

Språkteknologiske verktøy kan evalueres både automatisk og manuelt av et menneske. Fordelen med automatiske evalueringsmetoder, såkalte *referansemålinger* (benchmarks), er etterprøvbareheten. I tillegg er ressursbruken lav når det automatiske systemet først er laget. En referansemåling består av datasett og utvalgte evalueringsmål.

Forskere ved Universitetet på Island (Háskóli Íslands) publiserte i 2025 et datasett ment for referansemåling av stavekontroller og grammatikkontroller for islandsk språk

(Ármannsson et al. 2025). Testsettet består av parallellstilte setninger. Setningene er identiske med det unntaket at den ene inneholder feil, mens den andre ikke gjør det (fasit). En stavekontroll eller språkmodell testes ved at man mater inn setningen som inneholder feil, og sammenlikner det stavekontrollen eller språkmodellen gir tilbake, med fasitsetningen. I teorien kan da en ny stavekontroll eller språkmodell evalueres med kun et tastetrykk. Denne automatiske evalueringsmetoden har likevel et grunnleggende problem ved seg: Den gir bare rom for én fasit per setning, og all utputt som skiller seg fra fasiten, vil trekke ned skåren. Referansemålingen egner seg derfor dårlig for den norske språksituasjonen, der det finnes flere alternative og likestilte skrivemåter for mange ord.

Det foreligger for øyeblikket ingen formålstjenlig referansemåling for norsk stavekontroll, verken for bokmål eller nynorsk. Norsk andrespråskorpus (ASK) (en samling tekster på norsk skrevet av andrespråksbrukere, med merknader fra en lærer) har blitt konvertert til et parallelldatasett som likner på det i Ármannsson et al. (Jentoft 2023), men a) det er bare én fasit per setning, b) datasettet er ikke gjennomgått i detalj av et menneske, og c) feilene er ikke markert på ordnivå, men på setningsnivå. Derfor kan vi ikke bruke dette datasettet i denne undersøkelsen.

I påvente av at noen lager en velfungerende og formålstjenlig referansemåling for stavekontroller på norsk som blant annet tar høyde for at det finnes flere korrekte fasiter, har vi denne gangen valgt å evaluere manuelt. Stavekontroller er litt mindre kompliserte å evaluere enn enkelte andre språkteknologiske verktøy, og evalueringen kan gjennomføres på små kunstige datasett. Allerede ved å se på stavekontrollens ytelse på korte tekster (med høy konsentrasjon av feil og feller) kan man se om verktøyet gjør jobben sin eller ikke. Derfor kreves det ikke manuell gjennomgang av et stort antall tekster for å få et relevant resultat.

Det var ikke veldig ressurs- eller tidkrevende å gjennomføre evalueringen manuelt med Språkrådets egen fagkompetanse som fasit, og vi mener testen vår egner seg godt til å identifisere utfordringene i stavekontrollsystemene som er integrert i skriveprogrammene norske skoleelever bruker hver dag.

Testrundene

Vi har gjort tester på to tidspunkter. Første testrunde ble gjennomført på slutten av 2024 (Word) og i begynnelsen av 2025 (Pages og Docs). Andre testrunde ble gjennomført i juni 2025 for alle tre skriveprogrammene. Begge testrundene foregikk altså etter at den nye opplæringsloven trådte i kraft. Til testrunde 2 ble alle tekstene, feilene og fellene fra testrunde 1 byttet ut. Testtekstene fra runde 1 ble delt med leverandørene og kunne derfor ikke gjenbrukes. Vi gjorde i tillegg enkelte forbedringer basert på erfaringer fra første testrunde.

Testtekstene

For å sjekke kvaliteten på skriveprogrammene opp mot kravet i opplæringsloven har vi laget et sett med tekster som blir brukt til å se hvor godt stavekontrollene presterer i møte med feil og feller.

For hvert skriftspråk er det to tester:

Test A: «full av feil»

Her har vi tekster som inneholder ulike typer feil. Eksempel (nynorsk) med feil markert i **rødt**:

- Ho hadde **altid** ein **kassett spelar** i **sine lommer** om trongen til dans skulle melde seg.

Test/tekst A	bokmål:	nynorsk:
testrunde 1:	36 innlagte feil	43 innlagte feil
testrunde 2:	36 innlagte feil	37 innlagte feil

Med test A tar vi sikte på å sjekke om kontrollen fanger opp det som ikke er i tråd med rettskrivingen. Her skal kontrollen markere alle innlagte feil i tekstene med farget strek. I tillegg sjekker vi at minst ett av de alternative forslagene som kommer opp ved feilmarkeringene, er korrekt.

De innlagte feilene i tekstene er av ulike slag: stavefeil, bøyingsfeil, samsvarsfeil, samskrivingsfeil, og/å-feil, feil i bruk av stor eller liten forbokstav, kommafeil, bruk av feil preposisjon, inkonsekvent infinitivsending (nynorsk), foranstilt eiendomspronomen (nynorsk), bokmålsord (nynorsk) og nynorskord (bokmål).

Test B: «full av feller»

Her har vi feilfrie tekster som inneholder en del former og skrivemåter som alle er tillatt innenfor rettskrivingen, men som er mindre brukt enn andre alternativer. Disse kaller vi *feller*. Eksempel (bokmål) med feller markert i **grønt**:

- En **dølehest brøyt** seg i går inn på en **jassfestival** i Kongo.

Test/tekst B	bokmål:	nynorsk:
testrunde 1:	24 feller	23 feller
testrunde 2:	28 feller	22 feller

Med test B tar vi sikte på å forsikre oss om at kontrollen ikke markerer feil der det ikke er gjort feil. Her skal programmet ikke markere noe i det hele tatt, og det skal følgelig heller ikke komme opp alternative forslag.

En av grunnene til å gjøre denne delen av testen er at vi mistenker at stavekontrollene kan favorisere hyppigere brukte former av ord, noe som vil kunne føre til en ensretting av språket. Enkelte brukere av skriveprogrammene mener å ha sett en negativ utvikling på dette området de siste årene. Dette kan være et tegn på at KI og maskinlæring er involvert, med den konsekvensen at statistisk frekvens i datasett styrer programmenes atferd.

Evalueringsmål (metrikker)

Stavekontroll som oppgave kan sees på som en type informasjonsgjenfinning (information retrieval), en prosess der det gjelder å finne relevant informasjon i strukturerte eller ustrukturerte data. I vårt tilfelle er testtekstene ustrukturerte data, og informasjonsgjenfinningen går ut på å markere eller ignorere ord eller elementer. Alle evalueringsmålene vi bruker, er *feilbasert*. Det vil si at vi bare ser på feilene eller fellene som er lagt inn i tekstene, når vi teller, og ignorerer ordene uten feil. Dette står i kontrast til *leksikalske* evalueringsmål, der alle ordene i teksten ville vært med i tellingen.

De vanligste evalueringsmålene i informasjonsgjenfinning er *gjenfinningsgrad* (recall), *presisjon* (precision) og ulike varianter av aggregeringsmålet *f-skår* (Huyssteen et al. 2004). Alle evalueringsmålene vi bruker, går fra 0 til 1, der 1 er best. I denne rapporten visualiserer vi ofte flere av disse skårene i samme graf, men det er viktig at leseren husker på at det er snakk om ulike mål på samme y-akse. Alle disse målene kan omgjøres til prosenter, der 0,5 tilsvarer 50 prosent osv.

Test A (full av feil)

For å finne hvor mange av de faktiske feilene stavekontrollen klarer å markere, bruker vi evalueringsmålet *gjenfinningsgrad for feil* (error recall). Vi er ute etter andelen feil som stavekontrollen har fått med seg, sett opp mot det totale antallet feil i teksten. Vi deler antallet korrekte feilmarkeringer (true positive, TP) på det totale antallet feil i teksten (TP + false negative FN). Skåren går fra 0 til 1, der 1 er best.

- *Gjenfinningsgrad for feil*, eksempel: 10 feil i teksten, 8 av disse markert av stavekontrollen = $8/(8+2) = 0,8$.

Test A – forslagene

For å si noe om kvaliteten på forslagene fra stavekontrollen bruker vi evalueringsmålet *presisjon*, heretter kalt *forlagspresisjon*. Til dette formålet finnes det også mer finmaskede og strenge evalueringsmål (se Huyssteen et al. 2004), men i vår variant ser vi bare på om minst ett av forslagene fra stavekontrollen er korrekt. Vi deler antallet markeringer som inneholder minst ett korrekt forslag (TP), på det totale antallet feilmarkeringer (TP + false positive FP). Skåren går fra 0 til 1, der 1 er best.

- *Forslagspresisjon*, eksempel: 10 feilmarkeringer, 5 av disse har relevante og korrekte forslag til alternativ = $5/(5+5) = 0,5$.

Test B (full av feller)

For B-testene bruker vi en tilpasset variant av evalueringsmålet presisjon kalt *feilpresisjon* (error precision) (Huyssteen et al. 2004). Vi er ute etter antall feilmarkeringer, sett opp mot det totale antallet feller i teksten. Vi deler antallet feller stavekontrollen korrekt har ignorert, altså antall feller minus antallet markeringer (TP), på det totale antallet feller (TP + FP). Skåren går fra 0 til 1, der 1 er best.

I evaluering av stavekontroller er det vanlig å bruke en annen type presisjon, der man ser på stavekontrollens feilmarkeringer og hvor mange av disse som er korrekte. Men dette fungerer ikke i vårt tilfelle, fordi tekstene våre ikke inneholder noen feil som skulle vært markert, og da måtte man hatt 0 i telleren.

Selv om vi her kun fokuserer på de fellene vi bevisst har lagt inn, vil vi påpeke at flere av stavekontrollene i test B også feilmarkerer ord som ikke var ment å være feller.

Sistnevnte feilmarkeringer er også regnet med i TP-utregningen. Denne usikkerheten er en av grunnene til at vi har vektet denne skåren lavere enn gjenfinningsgrad for feil (test A).

- *Feilpresisjon* (error precision), eksempel: 10 feller i teksten, (minimum) 7 av disse er unngått av stavekontrollen = $7/(7+3) = 0,7$.

Språkrådet-skår

For å få et helhetsbilde og en totalskår for hver enkelt stavekontroll tar vi et vektet gjennomsnitt av gjenfinningsgrad for feil (test A, 60 prosent), forslagspresisjon (test A, 25 prosent) og feilpresisjon (test B, 15 prosent). Vi vekter førstnevnte høyere enn de to andre fordi vi mener feilmarkering (test A) er viktigere enn både forslagene (test A) og det å unngå feilmarkering av mindre brukte, men korrekte former (test B). Vi tror også denne vektingen samsvarer med den allmenne oppfatningen av hva de viktigste formålene med en stavekontroll er. Siden stavekontrollene i B-testen også har feilmarkert ord som ikke var ment å være feller, vil vi være forsiktige med å vektlegge denne delen av testen like mye, og vi vekter den dermed kun til 15 prosent. F-skår egner seg i vårt tilfelle ikke til å kombinere presisjon og gjenfinningsgrad fordi målingene ikke er gjort på de samme dataene (test A og test B bruker ulike tekster). Skåren går fra 0 til 1, der 1 er best.

- *Språkrådet-skår*, eksempel: gjenfinningsgrad for feil (test A) = 0,8, forslagspresisjon (test A) = 0,3, feilpresisjon (test B) = 0,4. Språkrådet-skår = $(0,6 \times 0,8) + (0,25 \times 0,3) + (0,15 \times 0,4) = 0,62$.

Det praktiske rundt testingen

Testrunde 1:

- test av Microsoft Word: utført på pc av oss i Språkrådet i desember 2024 på en skrivebordsversjon av Word
- test av Apple Pages: utført på iPad i lokalene til Utdanningsetaten i Oslo i januar 2025 av en lærer i Oslo-skolen, instruert og observert av Språkrådet
- test av Google Docs: utført på pc av en medarbeider i Øygarden kommune i et Teams-møte i januar 2025, instruert og observert av Språkrådet. Vi testet Docs gjennom verktøypakken «Google for Education», som er pakken skoleelever i Norge bruker Docs gjennom.

Testrunde 2:

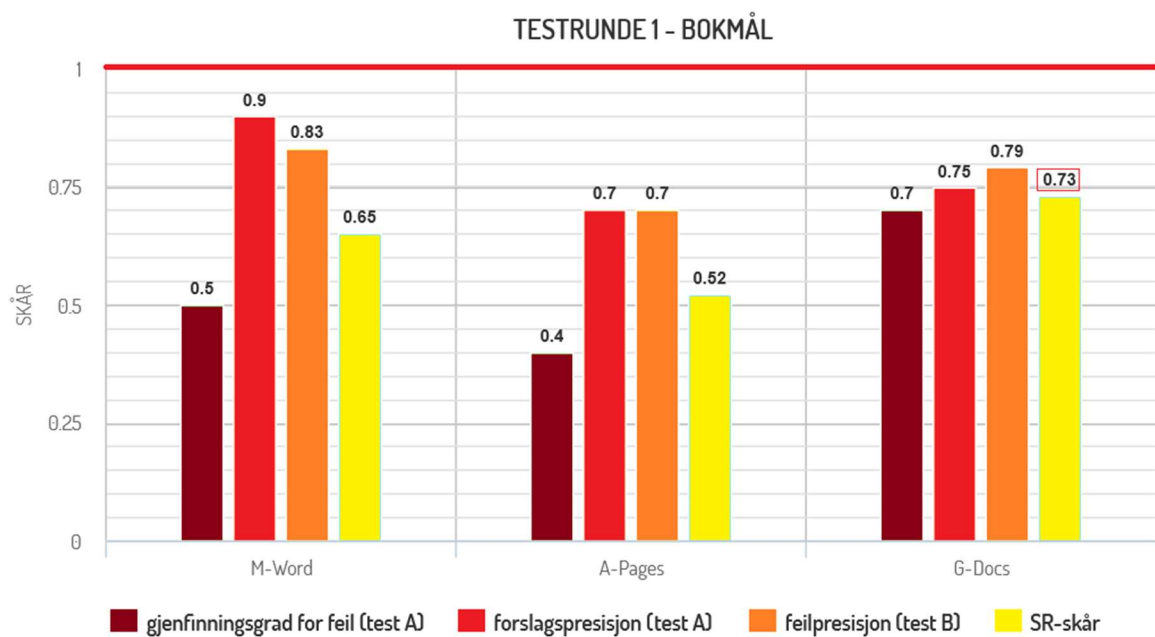
- test av Microsoft Word: utført på pc av oss i Språkrådet 16. juni 2025 på en nettversjon av Word (Office 365)
- test av Apple Pages: utført på iPad av oss i Språkrådet 13. juni 2025
- test av Google Docs: utført på pc av oss i Språkrådet 13. juni 2025 med en tilgang til programmet «Workspace for Education» som vi fikk fra Google

I testrunde 1 ble testtekstene limt inn i dokumentene, i testrunde 2 ble de skrevet inn. Vi har gjort kontrolltester som viser at denne forskjellen ikke påvirker resultatene. Videre hadde vi langt bedre kontroll over testsituasjonen i runde 2 enn i runde 1, fordi vi utførte all testingen selv. En effekt av dette er at vi lettere kan korrigere resultatene for autokorrektur som programmet eventuelt gjør. En annen effekt er at vi gjorde sikrere observasjoner av de alternative forslagene som kommer opp ved feilmarkeringene. Tallene som sier noe om disse forslagene i runde 1, er langt mer usikre enn i runde 2.

Resultatene¹

Testrunde 1 – bokmål

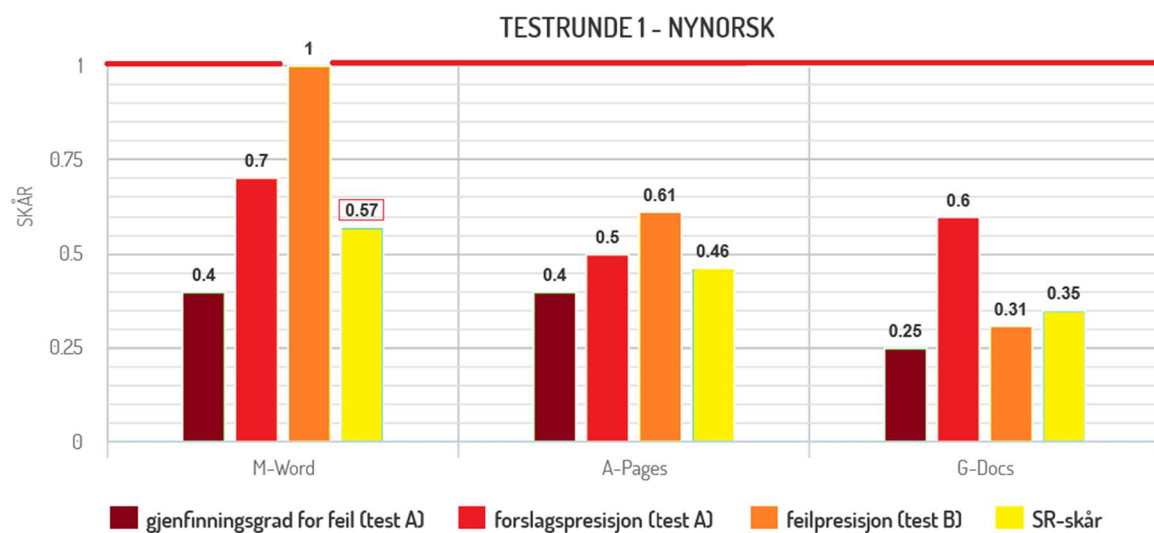
	Test A: full av feil		Test B: full av feller	Total
	Gjenfinningsgrad for feil	Forslagspresisjon	Feilpresisjon	Språkrådet-skår
M-Word	0,5	0,9	0,83	0,65
A-Pages	0,4	0,7	0,7	0,52
G-Docs	0,7	0,75	0,79	0,73



¹ Desimaltallene i grafene vises her feilaktig med punktum istedenfor komma. Det skyldes en teknisk feil ved verktøyet vi har valgt til å lage grafer.

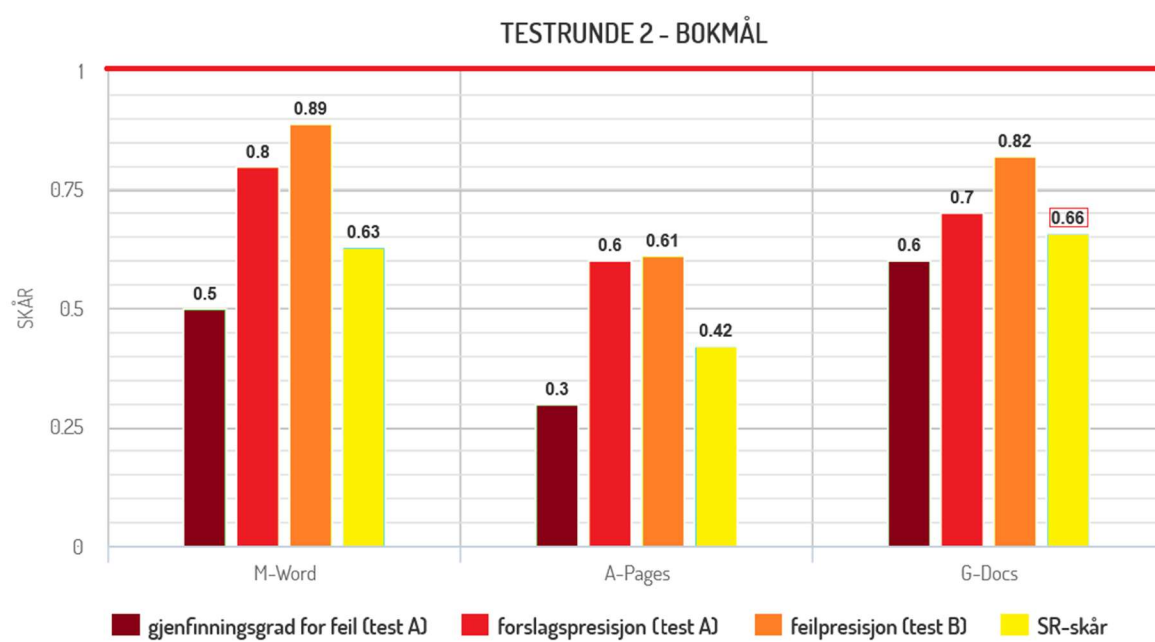
Testrunde 1 – nynorsk

	Test A: full av feil		Test B: full av feller	Total
	Gjenfinningsgrad for feil	Forslagspresisjon	Feilpresisjon	Språkrådet-skår
M-Word	0,4	0,7	1	0,57
A-Pages	0,4	0,5	0,61	0,46
G-Docs	0,25	0,6	0,31	0,35



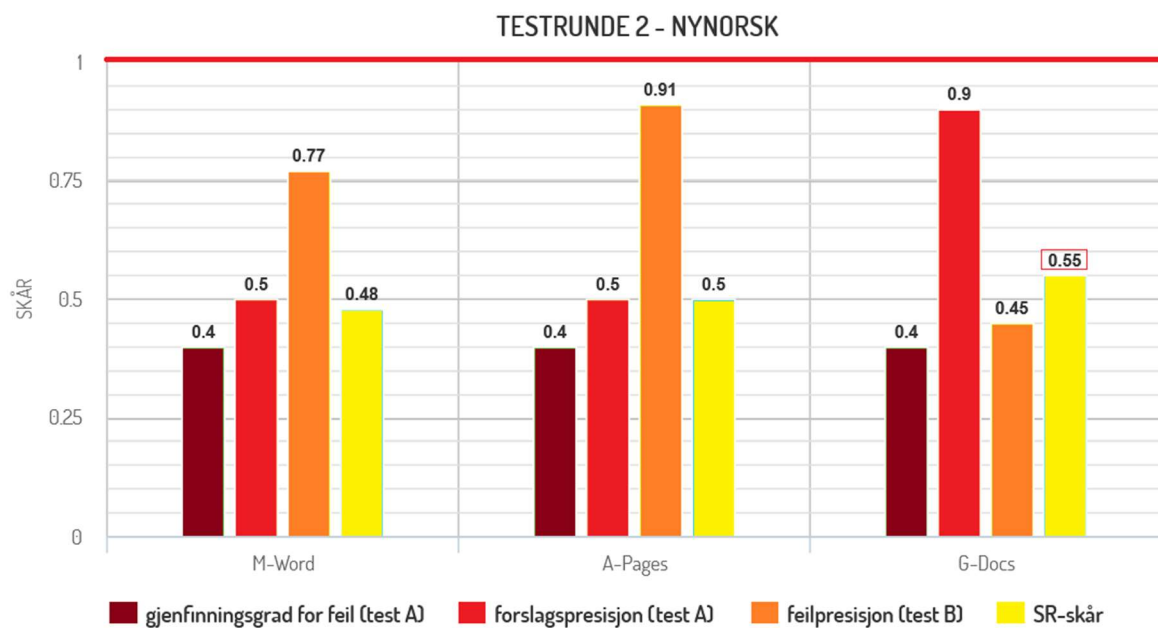
Testrunde 2 – bokmål (*endring i skår fra første testrunde i kursiv*)

	Test A: full av feil		Test B: full av feller	
	Gjenfinningsgrad for feil	Forslagspresisjon	Feilpresisjon	Språkrådet-skår
M-Word	0,5	0,8	0,89	0,63 (<i>-0,02</i>)
A-Pages	0,3	0,6	0,61	0,42 (<i>-0,1</i>)
G-Docs	0,6	0,7	0,82	0,66 (<i>-0,07</i>)



Testrunde 2 – nynorsk (*endring i skår fra første testrunde i kursiv*)

	Test A: full av feil		Test B: full av feller	
	Gjenfinningsgrad for feil	Forslagspresisjon	Feilpresisjon	Språkrådet-skår
M-Word	0,4	0,5	0,77	0,48 (−0,09)
A-Pages	0,4	0,5	0,91	0,50 (+0,04)
G-Docs	0,4	0,9	0,45	0,55 (+0,2)



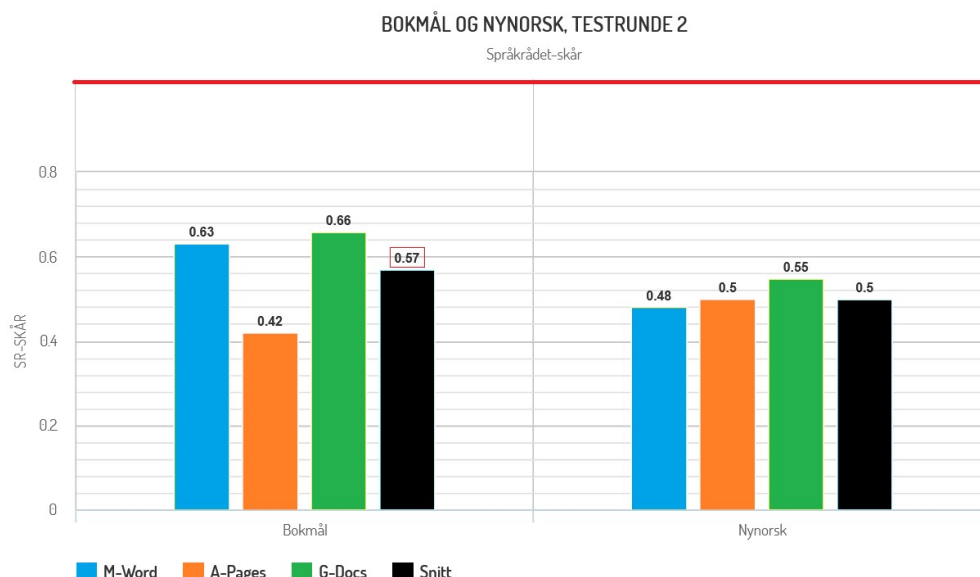
Konsekvenstest

I norsk rettskrivingstradisjon heter det at en tekst bør ha *konsekvens*, altså at normvalget bør være enhetlig gjennom hele teksten. Det vil for eksempel si at man på bokmål ikke skriver *sola* og *solen* i samme tekst, og heller ikke *ei jente* og *en jente*. Vi utformet en test også for dette, men fordi resultatene tydelig viste at ingen av stavekontrollene så ut til å ha funksjonalitet for å oppdage inkonsekvens, rapporterer vi ikke mer detaljert om disse resultatene.

Analyse av resultatene

Resultatene av testrunde 2 bekrefter langt på vei resultatene vi fikk på de første testene. Den mest merkbare endringen er at den nynorske stavekontrollen i Google Docs gjør det tydelig bedre i den siste testen enn i den første. For nynorsk er Google Docs nå på høyde med de andre programmene. Bokmålskontrollen til Apple Pages gjør det tydelig dårligere enn i den første testen. Ellers er resultatene omtrent de samme eller noe svakere.

Vi velger nå å konsentrere oss om resultatene vi fikk på den siste testrunden. Resultatene fra den sier mest om hvor gode stavekontrollene er per i dag. Tallene fra den første testen er dessuten mer usikre, siden vi hadde mindre kontroll på testsituasjonene enn vi hadde i testrunde 2.



Hovedfunn testrunde 2

- Resultatene er dårligere for de nynorske stavekontrollene enn for bokmålskontrollene. Det gjelder både hvor mange av de innlagte feilene som blir funnet, og hvor mange av stavekontrollens forslag som er korrekte. Forskjellen er stor for Microsoft Word og Google Docs, mens Apple Pages faktisk er bedre på nynorsk enn bokmål. Dette reduserer den gjennomsnittlige differansen mellom bokmål og nynorsk noe.
- I bokmålstekst varierer andelen av innlagte feil som blir funnet, mellom 30 prosent og 60 prosent. Selv i det beste av programmene er altså 4 av 10 feil uoppdaget. Det dårligste av programmene overser 7 av 10 feil. I nynorsktekst er 6 av 10 feil uoppdaget i alle tre programmene.
 - Det er vanskelig å se noe mønster i hvilke feiltyper som blir oppdaget og ikke, selv innenfor ett og samme program. Men ingen av de nynorske

stavekontrollene fanger opp kommafeil, samskrivingsfeil, og/å-feil, bruk av feil preposisjon og blanding av a- og e-infinitiv. Kommafeil og bruk av feil preposisjon blir heller ikke fanget opp i noen av bokmålskontrollene.

- Ved en relativt høy andel av feilmarkeringene i A-tekstene kommer det ikke opp noe forslag som er korrekt. Ingen av programmene har full skår her, og i nynorskkontrollene til Word og Pages mangler det et korrekt forslag ved halvparten av feilmarkeringene.
- Ingen av stavekontrollene som er testet, unngår helt å markere feil i feilfri tekst. Her spriker resultatene en del både mellom språkene og programmene. Totalt sett kommer Word best ut. Bokmålskontrollen i Apple Pages er merkbart dårligere enn nynorskkontrollen. For Google Docs er det omvendt: Her er resultatene desidert dårligst for nynorskkontrollen.
 - Ved de 19 feilmarkeringene som blir gjort i feilfri tekst av nynorskkontrollene, blir det foreslått bokmålsvarianter eller mer konvensjonelle skrivemåter som alternativ til 13 av dem (ca. 70 prosent).
 - Ved de 19 feilmarkeringene som blir gjort i feilfri tekst av bokmålskontrollene, blir det foreslått mer konvensjonelle skrivemåter som alternativ til 9 av dem (snaue 50 prosent). Om vi ser på resultatene for Microsoft Word og Google Docs isolert, blir det foreslått mer konvensjonelle skrivemåter ved 6 av 8 feilmarkeringer (75 prosent). Pages har gjort hele 11 feilmarkeringer, men bare ved få av dem er det foreslått mer konvensjonelle skrivemåter. Det er dermed i mange tilfeller vanskelig å se hva som har utløst feilmarkeringene i Pages.

Konklusjon og implikasjoner

Resultatene av testingen Språkrådet har foretatt, viser at Microsofts, Apples og Googles skriveprogrammer, som brukes av 9 av 10 elever i den norske skolen, per juni 2025 ikke oppfyller kravene i opplæringsloven § 15-4 andre ledd, verken på bokmål eller nynorsk. Alle aktørene tilbyr nynorske stave- og grammatikkontroller i skriveprogrammene sine, men når disse fungerer dårligere enn kontrollene for bokmål, innebærer det at nynorskelevene har tilgang på dårligere språklige verktøy enn bokmåselevene. Dette er i strid med intensjonen bak formuleringen i § 15-4 andre ledd om at skriveprogrammene skal støtte både bokmål og nynorsk. Når det gjelder lovens krav om at skriveprogrammene skal følge rettskrivingen, viser testresultatene klart at ingen av programmene oppfyller dette kravet. De er ikke engang i nærheten.

Det er tankevekkende at skriveprogrammene jevnt over ikke har blitt nevneverdig bedre i tidsrommet mellom de to testene Språkrådet har gjort. Etter testene tidlig i 2025 hadde Språkrådet separate møter med representanter fra Microsoft, Apple og Google. I

møtene gav Språkrådet tydelig informasjon om hva kravet i opplæringsloven § 15-4 andre ledd innebærer. Aktørene ble også orientert om hva testresultatene viste for deres skriveprogram. Dette ble fulgt opp med mer detaljert skriftlig dokumentasjon i etterkant av møtene.

Det er også tankevekkende at nynorskkontrollen i Google Docs er den eneste der vi kan dokumentere en viss forbedring fra test 1 til test 2. Nynorskkontrollen til Google fikk hard medfart i en NRK-artikkel på slutten av 2024 (NRK 2024) etter en tidligere test Språkrådet hadde gjort. Vi kan bare spekulere i om den negative omtalen har hatt betydning for resultatforbedringen. Resultatforbedringen for Google Docs viser uansett at det er mulig å gjøre merkbare forbedringer, selv om også Google fortsatt har en lang vei å gå. Teknisk burde det ikke være vanskelig, det er trolig mest et spørsmål om prioriteringer og ressursbruk.

Opplæringsloven stiller ikke krav om formell godkjenning av skriveprogrammer som skal brukes i skolen. Det er likevel viktig at skolemyndigheter, skoleeiere og lærere holder seg orientert om kvaliteten på skriveprogrammene som blir brukt, slik at de kan unngå å bryte loven ved å ta i bruk programmer som det er dokumentert at ikke oppfyller lovens krav.

Referanser

Ármansson et al. 2025

- *Playing by the Rules: A Benchmark Set for Standardized Icelandic Orthography* (Ármansson et al., NoDaLiDa 2025) ([lenke](#)).

Huyssteen et al. 2004

- *Evaluating Evaluation Metrics for Spelling Checker Evaluations*] (Huyssteen et al. 2004) ([lenke](#)).

Jentoft 2023

- *Grammatical Error Correction with byte-level language models*. Masteroppgave Matias Jentoft. Universitetet i Oslo, 2023 ([lenke](#)).

LNK-undersøkelsen 2024

- *Rapport: Opplæringslova og skriveprogram, ei undersøking blant rektorane i grunnskulen*, LNK (Landssamanslutninga av nynorskkommunar), 25. september 2024.

NRK 2024

- *Stavekontroll i norsk skole: Gotteri, låv og kjærleg-heit er nynorsk.* Journalist Aslak Øgrim Borgersund ([lenke](#)).

Opplæringsloven 2024

- § 15-4 *Ordlister, ordbøker og skriveprogram* ([lenke](#)).

Tildelingsbrev 2025

- *Statsbudsjettet 2025 – Språkrådet – tildelingsbrev* ([lenke](#)).