

# Sammendrag av analyserapporter fra nasjonale prøver i 2012

---

Dette er et sammendrag av de tre analyserapportene fra gjennomføringen av nasjonale prøver høsten 2012. Det ble gjennomført nasjonale prøver i lesing og regning på 5., 8. og 9. trinn og i engelsk på 5. og 8. trinn i grunnskolen.

Hver av de nasjonale prøvene ble gjennomført av omtrent 56 000 elever.<sup>1</sup>

Prøvene i engelsk og regning er elektroniske, mens prøvene i lesing er på papir. De elektroniske prøvene er i tre versjoner, mens leseprøvene er like for hele trinnet.

Alle oppgavene er testet grundig på forhånd.

I analyserapportene fra selve prøvegjennomføringen blir oppgavene presentert med en p-verdi som forteller oss prosentandelen av elevene som klarte oppgaven. Dersom en oppgave har en p-verdi på 0,6, var det 60 prosent av elevene som svarte riktig. Det er også regnet ut en gjennomsnittlig p-verdi for hver prøve, som sier noe om prøvens totale vanskelighetsgrad.

## Lesing 5. trinn

Nasjonale prøver i lesing på 5. trinn måler elevenes evne til å finne og tolke informasjon og til å reflektere over innhold og form i teksten. Prøven er forankret i kompetansemål for fag etter 4. trinn der grunnleggende ferdighet i lesing er integrert.

*Sammendrag av analysen av leseprøven på 5. trinn:*

- Prøven besto av et hefte med fem tekster og 28 oppgaver, og prøvetiden var 90 minutter.
- Den gjennomsnittlige p-verdien på hele prøven var 62. 2012-prøven har samme vanskegrad som prøven i 2009. Den er noe vanskeligere enn 2011-prøven (p-verdi 66), men lettere enn prøven i 2010 (p-verdi 57).
- Andre tekniske data: Cronbachs Alpha (reliabilitet) 0.86.

---

<sup>1</sup> Lesing 5. trinn: 54 296. Lesing 8. trinn: 59 003. Lesing 9. trinn: 57 662. Regning 5. trinn: 54 964. Regning 8. trinn: 59 327. Regning 9. trinn: 58 056. Engelsk 5. trinn: 52 327. Engelsk 8. trinn: 53 645.

- Prøven består av oppgaver som måler tre aspekter ved lesing – å finne informasjon, å tolke, og å reflektere over tekstens form og innhold. Den består av 5 finne-, 19 tolke- og 4 reflektere-oppgaver.
- Oppgavetyperne måles i tre ulike formater – åpne oppgaver, flervalgsoppgaver, og komplekse flervalgsoppgaver. Prøven består av 3 åpne-, 19 flervalgs- og 6 komplekse flervalgsoppgaver.
- I årets prøve er det liten forskjell i vanskegrad mellom de åpne oppgavene (p-verdi 67) og flervalgsoppgavene. De sammensatte flervalgsoppgavene er vanskeligere (p-verdi 52) enn tradisjonelle flervalgsoppgaver (p-verdi 66).
- Alle oppgavene i prøvene har god diskrimineringsevne ( $d \geq 0,21$ ), også de letteste oppgavene. Det ser ut til at også de letteste oppgavene holder høy kvalitet i et testteknisk perspektiv.
- Jentene gjør det gjennomgående litt bedre enn guttene på prøven. Gjennomsnittlig skåre for jentene er 18.9 (p-verdi= 63) mens gjennomsnittlig skåre for guttene er 18.2 (p-verdi= 61).
- Den gjennomgående tendensen til at jentene gjør det litt bedre enn guttene, gjør seg også gjeldende for de tre ulike aspektene ved lesing og de ulike oppgaveformatene.
- Fordelingen mellom kjønnene på mestringsnivåene samsvarer med resultatene på leseprøven i 2010 og 2011. 25 % av jentene og 30 % av guttene er på mestringsnivå 1. 48 % av jentene og 46 % av guttene er på mestringsnivå 2. 26 % av jentene og 24 % av guttene er på mestringsnivå 3.
- For å få et inntrykk av hvor store kjønnsforskjellene er har Lesesenteret beregnet effektstørrelse<sup>2</sup>. Effektstørrelsen på forskjellen mellom gutter og jenters gjennomsnittlige skåre er 0.11, en forskjell som kan betegnes som liten og som også er mindre enn i 2011 (0.17), 2010 (0,15) og 2009 (0,19).
- Forskjellene er også mindre enn rapporterte kjønnsforskjeller for samme aldersgruppe i PIRLS 2006<sup>3</sup> og mindre enn rapporterte kjønnsforskjeller i lesing for eldre elever, både i nasjonale prøver i lesing på 8. trinn<sup>4</sup> og PISA 2003, 2006 og 2009<sup>5</sup>.
- Det er flere gutter enn jenter som lar være å besvare oppgaver. Dette er en tendens vi også har sett tidligere år. Det er imidlertid ingen sterk tendens til flere blanke svar mot slutten av prøven, noe som indikerer at arbeidsomfanget i prøven ikke har vært for stort.
- Gjennomgangen av oppgavene på 5. trinn viser at oppgavene der elevene skulle finne informasjon i teksten var lettere enn oppgaver som krever tolkning og refleksjon.

<sup>2</sup> Effektstørrelsene i denne rapporten er beregnet som Cohen's  $d$ . Cohen's  $d = M_1 - M_2 / \sigma_{\text{pooled}}$  where  $\sigma_{\text{pooled}} = \sqrt{[(\sigma_1^2 + \sigma_2^2) / 2]}$

<sup>3</sup> van Daal, V. et al., 2007

<sup>4</sup> Vagle, W. et al., 2009

<sup>5</sup> Kjærnsli, M. et al. 2007; Kjærnsli, M. & Roe, A. (Red.), 2010

## Lesing 8. og 9. trinn

Nasjonale prøver i lesing på 8. og 9. trinn måler elevenes evne til å finne og tolke informasjon og til å reflektere over innhold og form i teksten. Prøven er forankret i kompetansemål for fag etter 7. trinn der grunnleggende ferdighet i lesing er integrert.

*Sammendrag av analysen av leseprøven på 8. og 9. trinn:*

- Prøvetiden var 90 minutter, og prøven besto av et hefte med åtte forskjellige tekster og 46 oppgaver.
- De 46 oppgavene er sammensatt av 19 forstå og tolke-oppgaver, 15 finne-oppgaver og 12 reflektere og vurdere-oppgaver.
- Oppgavetyperne måles i tre ulike formater – åpne oppgaver (14 stykker), flervalgsoppgaver (30 stykker), og sammensatte flervalgsoppgaver (2 stykker).
- Resultatene fra den nasjonale leseprøven for 8. og 9. trinn i 2012 er i tråd med tidligere nasjonale prøver.
- Den gjennomsnittlige p-verdien på hele leseprøven er 60,8 på 8. trinn og 67,8 på 9. trinn. Det innebærer at elevene på 9. trinn ligger litt lenger foran elevene på 8. trinn enn de har gjort de to foregående årene. I 2011-prøven var p-verdiene 62,1 på 8. trinn og 67,9 på 9. trinn.
- Prøven har en spredning i vanskegrad, målt i gjennomsnittlige p-verdier for hver oppgave, fra 15,2 til 82,8 på 8. trinn, og fra 20,6 til 87,8 på 9. trinn.
- Standardavvikene for oppgavene i 2012-prøven er 15,6 på 8. trinn og 14,7 på 9. trinn. Differansen på 7,0 prosentpoeng mellom 8. og 9. trinn tilsvarer dermed 0,45 standardavvik. I 2011 var differansen mellom 8. og 9. trinn 0,32 standardavvik.
- Kjønnsforskjellene er i tråd med tidligere resultater. Jentenes gjennomsnittlige prestasjon på prøven ligger markant høyere enn guttenes på begge trinn. Den gjennomsnittlige p-verdien på 8. trinn er 63,7 for jenter og 58,1 for gutter. På 9. trinn 70,7 for jenter og 64,9 for gutter.
- Effektstørrelsen av kjønnsforskjellene (0,30 på 8. trinn og 0,32 på 9. trinn) er beregnet som prosentandeler av standardavvikene, og de er på omtrent samme nivå som foregående år på både 8. og 9. trinn.
- Prosentfordeling på mestringsnivåer (1-5) på 8. trinn: Gutter 10-24-38-20-9. Jenter: 5-17-40-26-12.
- Prosentfordeling på mestringsnivåer (1-5) på 9. trinn: Gutter 6-16-34-26-17. Jenter: 2-10-32-31-24.
- Tallene viser at jentene ligger nesten ett årstrinn foran guttene i utviklingen av lesekompetanse.
- Som i tidligere år presterer jentene bedre enn guttene på alle de tre lesemåtene og på alle de tre oppgaveformatene. Jentene hevder seg særlig godt på refleksjonsoppgavene og på de åpne oppgavene.
- Begge kjønn har problemer med tekster og oppgaver som krever konsentrert og nøyaktig lesing. *Reflektere og vurdere* kommer ut som det vanskeligste aspektet.

Dette leseaspektet har alltid, med unntak av fjorårets prøve, vist seg å være det vanskeligste for elevene på 8. trinn.

- Det er en tendens til at andelen blanke svar øker mot slutten av oppgaveheftet, men denne økningen dreier seg bare om 2 – 3 prosentpoeng på flervalgsoppgavene. Dette tyder på at elevene har hatt nok tid på seg.

## Regning 5. trinn

Nasjonale prøver i regning er ikke en prøve i faget matematikk, men i regning som grunnleggende ferdighet i alle fag. Prøven er forankret i kompetansemål for fag etter 4. trinn der grunnleggende ferdigheter i regning er integrert. Den tar for seg områdene tall, måling og statistikk.

*Sammendrag av analysen av regneprøven på 5. trinn:*

- Prøven på femte trinn består av tre versjoner. Oppgavene er de samme, men rekkefølgen på de ti første og de ti siste oppgavene er forskjellig. Analyserapporten tar kun utgangspunkt i versjon 1. Det var 18 412 elever som gjennomførte denne versjonen.
- Prøven på 5. trinn består av 45 oppgaver og prøvetiden er på 90 minutter. Hver av oppgavene er pilotert på omtrent 1000 elever før gjennomføringen.
- 30 av oppgavene er flervalgsoppgaver, mens 15 er åpne oppgaver. 20 av oppgavene omhandler området tall, 17 omhandler måling, mens 8 omhandler statistikk.
- Mange av oppgavene er relevante for flere fag. Oppgavene på 5. trinnsprøven kan knyttes til følgende fag: Matematikk, norsk, naturfag, samfunnsfag, mat og helse, kroppsøving, kunst og håndverk, og religion, livssyn og etikk.
- P-verdien for hele prøven er 59,4. Det er statistikkområdet som har høyest gjennomsnittlig p-verdi: 65,6. For tall- og målingsoppgavene er løsningsprosenten henholdsvis 59,6 og 56,2.
- Prøven har god reliabilitet<sup>6</sup>: Chronbachs Alpha er 0,90.
- Alle oppgavene diskriminerer<sup>7</sup> godt, med unntak av to stykker.
- Elevene som har løst en oppgave riktig har i gjennomsnitt høyere poengsum på hele prøven enn de som ikke har funnet riktig svar. Dette gjelder samtlige av prøvens oppgaver.
- P-verdien for jenter på hele prøven er 57,8, mens p-verdien for gutter på hele prøven er 60,9.
- Guttene gjør det signifikant bedre enn jentene, men effekten av kjønn er liten. Forskjellen mellom guttene og jentene er imidlertid større i 2012-prøven på 5. trinn enn 2011-utgaven.
- Dette samsvarer med resultatene fra TIMSS 2011, som viser at det er en liten, signifikant kjønnsforskjell på prestasjoner i guttenes favør på 4. trinn i

<sup>6</sup> Reliabilitet – pålitelighet – et mål for kvalitet, målt i form av Chronbachs alpha  $> 0,80$

<sup>7</sup> Diskriminering: Diskr (D) betyr oppgavens diskriminering, dvs. hvordan oppgaven samsvarer med summen av alle oppgavene.

matematikk (Grønmo mfl. 2012). I PISA 2006 (Kjærnsli mfl. 2007) gjør guttene i Norge det litt bedre enn jentene, og forskjellen er statistisk signifikant. Også resultatene etter PISA 2009 viser at guttene skårer bedre enn jentene i matematikk.

- Det er prosentvis flere jenter enn gutter på mestringsnivå 1 og 2, mens det er omvendt på mestringsnivå 3. Omtrent 29 prosent av guttene i utvalget oppnår en poengsum tilsvarende mestringsnivå 3, mens omtrent 21 prosent av jentene oppnår det samme.
- Guttene gjør det bedre enn jentene på fem av de seks vanskeligste oppgavene på prøven. Forskjellen er større i guttefavør enn i jentefavør.
- Guttene har bedre resultat enn jentene på seks av de ti letteste oppgavene. Jentene gjør det bedre enn guttene på statistikkoppgavene, der det handler om å lage og bearbeide informasjon i søylediagram.
- I likhet med tidligere år, ser vi en klar tendens til at guttene i gjennomsnitt gjør det bedre enn jentene i oppgaver med omgjøring mellom enheter i masse og lengde (spesielt kg og g, km og m).
- Det ser ut til at jentene i gjennomsnitt skårer høyere enn guttene i subtraksjonsoppgaver i kontekst og i ferdigoppstilte oppgaver.
- I gjennomsnitt er 0,9 prosent av oppgavene i prøven ubesvart. Det er en tendens til at andelen ubesvarte oppgaver øker mot slutten av prøven. Analysen viser at det er litt høyere andel ubesvarte blant de åpne oppgavene enn blant flervalgsoppgavene. Forskjellen mellom jenter og gutter når det gjelder andel ubesvarte på åpne oppgaver og flervalgsoppgaver er veldig liten.

## Regning 8. og 9. trinn

Nasjonale prøver i regning er ikke en prøve i faget matematikk, men i regning som grunnleggende ferdighet i alle fag. Prøven er forankret i kompetansemål for fag etter 7. trinn der grunnleggende ferdigheter i regning er integrert. Den tar for seg områdene tall, måling og statistikk.

*Sammendrag av analysen av regneprøven på 8. og 9. trinn:*

- Prøven på femte trinn består av tre versjoner. Oppgavene er de samme, men rekkefølgen på de ti første og de ti siste oppgavene er forskjellig. Analyserapporten tar kun utgangspunkt i versjon 1. Det var 19 609 elever på 8. trinn og 19 370 elever på 9. trinn som gjennomførte denne versjonen.
- Prøven på 8. trinn består av 58 oppgaver og prøvetiden er på 90 minutter. Hver av oppgavene er pilotert på omtrent 1000 elever før gjennomføringen.
- 34 av oppgavene er flervalgsoppgaver, mens 24 er åpne oppgaver. 23 av oppgavene omhandler området tall, 21 omhandler måling, mens 14 omhandler statistikk.

- Mange av oppgavene er relevante for flere fag. Oppgavene på 8. trinnsprøven kan knyttes til følgende fag: Matematikk, norsk, naturfag, samfunnsfag, mat og helse, kroppsøving, kunst og håndverk, musikk, engelsk og religion, livssyn og etikk.
- P-verdien for hele prøven på 8. trinn er 49,1. P-verdien for hele prøven på 9. trinn er 55,9.
- Det er området statistikk som har høyest gjennomsnittlig løsningsprosent på denne prøven: 59,9 (66,1 på 9. trinn). Måling har den laveste løsningsprosenten: 40,4 (48,0 på 9. trinn). Vi ser at tendensen er den samme for 9. som for 8. trinn, men at elevene på 9. trinn i gjennomsnitt har løst flere oppgaver riktig enn elevene på 8. trinn.
- Prøven har god reliabilitet<sup>8</sup>: Cronbachs Alpha er 0,93 på 8. trinn og 0,94 på 9. trinn.
- 54 av de 58 oppgavene fungerer meget bra ( $D > 0,30$ ). De fire andre har diskriminering<sup>9</sup> fra 0,17 til 0,28. På 9. trinn er det bare to av disse oppgavene som har  $D < 0,30$  (på henholdsvis 0,18 og 0,25). Begge disse oppgavene har svært høy p-verdi.
- Elevene som har løst en oppgave riktig har i gjennomsnitt høyere poengsum på hele prøven enn de som ikke har funnet riktig svar. Dette gjelder samtlige av prøvens oppgaver.
- Det er en tendens til at elever på 9. trinn behersker sammenhengen mellom brøk og prosent, og regning med desimaltall, prosent og tid, bedre enn elevene på 8. trinn. I tillegg løser elevene på 9. trinn sammensatte oppgaver i større grad enn 8. trinns elevene.
- P-verdien for jenter på hele 8. trinnsprøven er 48,0, mens p-verdien for gutter på hele 8. trinnsprøven er 50,2.
- P-verdien for jenter på hele 9. trinnsprøven er 54,3, mens p-verdien for gutter på hele 9. trinnsprøven er 57,4.
- Guttene gjorde det signifikant<sup>10</sup> bedre enn jentene på prøven som helhet. Forskjellen er 2,2 prosentpoeng på 8. trinn og 3,1 prosentpoeng på 9. trinn i guttenes favør. Effekten av kjønn er allikevel liten.
- Analysen viser at betydningen av et års ekstra skolegang er litt større for guttene enn for jentene.
- Den gjennomsnittlige kjønnsforskjellen på prøven for 2012 er litt mindre enn i prøven for 2011, men stabil i guttefavør som den har vært alle år siden 2007, og på begge trinn siden 2010.
- Guttene gjør det signifikant bedre enn jentene i mer enn dobbelt så mange flervalgsoppgaver som åpne oppgaver. Jentene gjør det signifikant bedre enn guttene i mer enn dobbelt så mange åpne oppgaver som flervalgsoppgaver
- Guttene gjør det bedre enn jentene i sju av de ti vanskeligste oppgavene på både 8. og 9. trinn. Åtte av de ti vanskeligste oppgavene på begge trinn, er innenfor områdene tall og måling. Dette stemmer med tendensen fra tidligere års prøver.

<sup>8</sup> Reliabilitet – pålitelighet – et mål for kvalitet, målt i form av Chronbachs alpha  $> 0,80$

<sup>9</sup> Diskriminering: Diskr (D) betyr oppgavens diskriminering, dvs. hvordan oppgaven samsvarer med summen av alle oppgavene.

<sup>10</sup> Signifikant – forskjellen er så stor at den ikke kan skyldes tilfeldigheter. 5 % signifikansnivå.

- Blant de ti letteste oppgavene er det like mange oppgaver i jente- som i guttefavør. Oppgavene er de samme på begge trinn, og de består av seks statistikk- og fire talloppgaver. Alle oppgavene kommer i første halvdel av oppgavesettet, og sju av oppgavene er flervalgsoppgaver.
- I gjennomsnitt er det 4,0 prosent ubesvarte oppgaver for 8. trinn og 3,9 prosent ubesvarte for 9.trinn. Største andel ubesvarte finner vi blant de åpne oppgavene.
- Det er en tendens til at andelen ubesvarte oppgaver øker mot slutten av prøven. Det er også en tendens til at andelen jenter som ikke svarer, øker mer i siste del av prøven enn andel gutter som ikke svarer.

## Engelsk 5. trinn

Nasjonale prøver i engelsk prøver deler av faget engelsk. Prøven måler delferdighetene vokabular, leseforståelse og grammatikk og tar utgangspunkt i kompetansemålene etter 4. trinn. Det er tre ulike versjoner av prøven i tillegg til en ankerprøve. Versjonene kan sammenlignes når det gjelder innhold og vanskelighetsgrad.

*Sammendrag av analysen av engelskprøven på 5. trinn:*

### *Ankerprøven*

- I 2013 ble det utviklet en ankerprøve på 5. trinn. Ankerprøven er gjennomført av 2255 elever. Gjennomsnittlig  $p$ -verdi er 0.57.
- Ankerprøven har tilfredsstillende mål for reliabilitet og diskriminering.

### *Hovedprøven*

- De tre ulike versjonene av prøven på 5. trinn består alle av 50 oppgaver. Ca 17,5 tusen elever har gjennomført hver av de tre versjonene.
- 5. trinnsprøvene har gjennomsnittlige  $p$ -verdier på ca 0.545, og fordelingen av resultatene på 5. trinn er tilnærmet normalfordelte.
- Prøvene har tilfredsstillende og svært like reliabilitetsmål: Fra 0,936 til 0,937.
- Sett under ett diskriminerer prøvene mellom sterke og svake elever på en fornuftig måte.
- Den gjennomsnittlige poengsummen og standardavviket viser at forskjellen mellom gutter og jenter på 5. trinn er svært små, men signifikant. Jentene skårer gjennomsnittlig litt høyere enn guttene.

## Engelsk 8. trinn

Nasjonale prøver i engelsk prøver deler av faget engelsk. Prøven måler delferdighetene vokabular, leseforståelse og grammatikk og tar utgangspunkt i kompetansemålene etter 7. trinn. Det er tre ulike versjoner av prøven i tillegg til tre ulike ankerprøver. Versjonene kan sammenlignes når det gjelder innhold og vanskelighetsgrad.

## *Sammendrag av analysen av engelskprøven på 8. trinn:*

### *Ankerprøvene*

- Det ble utviklet tre ankerprøver på 8. trinn. Hver ankerprøve er gjennomført av omtrent 1700 elever. Gjennomsnittlig  $p$ -verdi varierer fra 0.520 til 0.584.
- Alle de tre ankerprøvene har tilfredsstillende mål for reliabilitet og diskriminering.

### *Hovedprøven*

- De tre ulike versjonene av prøven på 8. trinn består alle av 56 oppgaver. 17,5-18 tusen elever har gjennomført hver av de tre versjonene.
- 8. trinnsprøvene har gjennomsnittlige  $p$ -verdier på ca 0.547. Fordelingen av elevresultatene på poengskalaen for 8. trinn er ikke helt normalfordelte.
- Prøvene har tilfredsstillende og svært like reliabilitetsmål: Fra 0,952 til 0,953.
- Sett under ett diskriminerer prøvene svært godt mellom sterke og svake elever.
- Forskjellen mellom gutter og jenter er svært liten, og ikke signifikant. De svakeste guttene er svakere enn de svakeste jentene, og samtidig er de sterkeste guttene litt sterkere enn de sterkeste jentene. Det betyr at det er større spredning blant guttene enn hos jentene. Det er imidlertid viktig å huske på at dette kun er tendenser, og at forskjellene er meget små.
- Guttene gjør det best i oppgaver der utgangspunktet er en kort tekst, mens jentene gjør det bedre når elevene må forholde seg til litt lengre tekster.