

## Rammeverk for eksamen – LK20 og LK20S

Rammeverket for sentralt gitt eksamen er oppdatert i tråd med de nye læreplanene, som trådte i kraft høsten 2020. Rammeverket tas først i bruk for eksamen i grunnskolen og fellesfag i videregående opplæring.

---

ARTIKKEL | SIST ENDRET: 15.02.2021

---

Rammeverket skal brukes av oppgaveutviklere som lager eksamener etter nye læreplaner (LK20, LK20S), og skal bidra til å styrke den metodiske kvaliteten på eksamen og sensur.

Rammeverket er på overordnet nivå, og ikke spesifikt for hvert fag. I det nye rammeverket er kvalitetskravene til eksamen tydeligere beskrevet enn før.

Rammeverket er dynamisk og skal utvikles videre ut i fra erfaringer med bruk av rammeverket, og utvikling og utprøving av oppgaver til eksamen etter nye læreplaner.

Det gamle rammeverket gjelder eksamener etter LK06 og blir gradvis faset ut.

## Innhold

### Innledning

#### 1. Hva er formålet med rammeverket?

#### 2. Hva sier forskriften om eksamen?

#### 3. Læreplanen er utgangspunkt for det vi prøver til eksamen

#### 4. Kvalitetskriterier for utvikling av eksamen

#### 5. Retningslinjer for utvikling av eksamensoppgaver

- 5.1. Bruk flere oppgaver og oppgavetyper
- 5.2. Lag oppgaveinstruksjoner som er tydelige for kandidaten
- 5.3. Lag en eksamen som er så rettferdig som mulig
- 5.4. Gi tydelig informasjon om tidsrammer og hjelpemidler
- 5.5. Lag tydelige vurderingskriterier for hvordan et kandidatsvar skal vurderes

#### 6. Hvordan kan vi styrke reliabiliteten i eksamen?

#### 7. Oppsummering

## Innledning

Dette rammeverket er en oppfølging av eksamensgruppas anbefalinger knyttet til behovet for å styrke det metodiske grunnlaget for arbeidet med eksamen. Eksamensgruppa besto av forskere og representanter for organisasjonene i fagfornyelsen. Vi har lagt kunnskapsgrunnlaget og anbefalingene til eksamensgruppa til grunn for arbeidet.

Utkastet til rammeverk er vurdert bl.a. av vår eksterne kvalitetssikrer EKVA og CEMO ved UiO som har gitt viktige bidrag til denne versjonen. Fagmiljøer ved NTNU og UiB har benyttet en foreløpig versjon av rammeverket for arbeidet med å utvikle eksamen i fagfornyelsen, som har gitt verdifulle erfaringer fra blant annet utprøvinger av oppgaver med elever og lærere.

Utkastet til rammeverk er forankret i referansegruppen for videreutvikling av vurderingsordningen. Referansegruppen består av organisasjonene i fagfornyelsen som har drøftet rammeverkets innhold og som har kommet med innspill som er forsøkt innarbeidet i denne versjonen.

Rammeverket vil endres i lys av erfaringene våre oppgaveutviklere gjør seg med å bruke det i samarbeid med våre fagansvarlige. Vi vil også vurdere endringsbehov for å ta rammeverket i bruk i programfag.

## 1. Hva er formålet med rammeverket?

Dette rammeverket skal gi dere som lager eksamen hjelp til å utvikle eksamensoppgaver av god kvalitet. I rammeverket beskrives først føringer for eksamen og kvalitetskriterier for utvikling av eksamen. Deretter finner dere retningslinjer for å utvikle og kvalitetssikre oppgaver og vurderingskriterier.

Kvalitetskriteriene for eksamen er:

- Validitet

- Reliabilitet
- Rettferdighet
- Håndterbarhet

Hva disse kriteriene betyr finner du beskrevet lenger ned i teksten.

## 2. Hva sier forskriften om eksamen?

Forskrift til opplæringsloven legger føringer for eksamen som dere må følge når vi utvikler og vurderer eksamen. Her står det at:

- Sluttvurderinga skal gi informasjon om kompetansen til eleven, lærlingen, lære kandidaten eller praksisbrevkandidaten ved avslutninga av opplæringa i fag. (§3-14).
- Eksamen skal vere i samsvar med kompetansemåla i læreplanen (§ 3-22)
- Kompetansemåla skal forståast i lys av teksten om faget i læreplanen (§ 3-3)
- Eksamen skal gi eleven eller privatisten høve til å vise sin kompetanse i så stor del av faget som mogleg ut frå eksamensforma (§3-22)
- Ein eksamenskarakter skal vere uttrykk for den kompetansen kvar enkelt elev eller privatist viser på eksamen. (§3-22)

## 3. Læreplanen er utgangspunkt for det vi prøver til eksamen

En forutsetning for at en eksamen skal kunne gi riktig informasjon om kompetansen til kandidaten er at dere bruker læreplanen og lager en beskrivelse av kompetansen og det faglige innholdet som skal prøves til eksamen. Denne beskrivelsen kaller vi et konstrukt. Konstruktet skal i utgangspunktet ligge fast og følge læreplanen, men kan justeres ved behov. Beskrivelsen vil for eksempel redegjøre for den norskfaglige, matematikkfaglige eller samfunnsfaglige kompetansen det er aktuelt å prøve på eksamen.

Konstruktet vil, gitt de praktiske begrensningene som ligger i eksamen, være noe smalere enn den kompetansen som hele læreplanen og opplæringen legger opp til. Dette er også tydeliggjort i forskriften der det står at eksamen skal gi kandidaten mulighet til å vise sin kompetanse «i så stor del av faget som mulig ut fra eksamensformen». Hvor stor del av kompetansen i læreplanen som prøves til eksamen, kan variere blant annet ut fra eksamensform, tidsramme og hvordan kompetansemålene er formulert i læreplanene (Udir-2-2020). Dette er en viktig diskusjon dere må ta i det enkelte fag.

## 4. Kvalitetskriterier for utvikling av eksamen

Som oppgaveutviklere skal dere sikre at eksamen er av god kvalitet og i tråd med kvalitetskriteriene under.

Kriteriene er ikke absolutte og må alltid ses i forhold til hverandre. I retningslinjene lenger ned finner dere metoder for hvordan dere praktisk skal gå fram.

## **Reliabilitet**

Reliabilitet handler om i hvilken grad vi kan stole på at eksamensresultatene gir et presist bilde av den kompetansen kandidaten viser på eksamen og ikke andre faktorer, som for eksempel sprikende vurderinger blant sensorer. Fordi eksamen har store konsekvenser for den enkelte, må reliabiliteten være så høy som mulig. Litt forenklet kan vi si at vi øker reliabiliteten ved å øke antallet oppgaver, utvikle tydelige vurderingskriterier og la sensorer vurdere oppgavene uavhengig av hverandre. Når reliabiliteten er lav, kan vi ikke si sikkert hva det er vi måler til eksamen.

## **Validitet**

Validitet handler om at et resultat på en eksamen gjenspeiler den faglige kompetansen til en kandidat på en troverdig måte. Det betyr at oppgavene må være laget i tråd med beskrivelsen av kompetansen som eksamen er ment å måle, også kalt konstruktet. Oppgavene må også være innenfor beskrivelsen av fagets innhold slik det er definert i læreplanen. En forutsetning for å kunne si noe om at en eksamen er valid er at reliabiliteten er tilstrekkelig høy.

## **Rettferdighet**

Rettferdighet handler om at oppgaver, vedlegg og lignende verken skal være støtende for enkelte grupper, eller utilsiktet favorisere en gruppe framfor andre grupper for eksempel ved valg av tema, hvordan oppgavene er utformet eller hvordan vurderingskriteriene er utformet.

## **Håndterbarhet**

Håndterbarhet handler om at valg av oppgaver må være tilpasset tidsrammen for eksamen og i tråd med tilgjengelige ressurser, som for eksempel hvilke hjelpemidler kandidatene kan bruke på eksamen.

## **5. Retningslinjer for utvikling av eksamensoppgaver**

Når dere lager oppgaver til eksamen skal dere altså sikre at eksamen er mest mulig valid, reliabel, rettferdig og håndterbar innenfor rammene som til enhver tid gjelder for oppgaveutviklingen. Overordnet skal kandidaten få vise kompetansen sin på flere og varierte måter. Oppgavesettet skal samlet sett representere det som er mulig å vurdere til eksamen i læreplanen, altså den kompetansen og det faglige innholdet som er beskrevet i konstruktet.

Da må dere ta hensyn til disse prinsippene for oppgaveutvikling som er beskrevet nedenfor.

## 5.1. Bruk flere oppgaver og oppgavetyper

Å la kandidater få svare på flere forskjellige oppgaver, der det fremkommer tydelig hvilke områder av kompetansen som blir vurdert, er svært viktig for å styrke reliabiliteten og validiteten i eksamen.

For å ivareta dette anbefaler vi sterkt at dere i eksamener med åpne oppgaver har minimum tre oppgaver som utgangspunkt for å vurdere elevenes besvarelse.

Hvilke oppgavetyper dere velger henger sammen med hvilken kompetanse dere ønsker at kandidaten skal vise. Vi har to hovedkategoriene av oppgaver som brukes til eksamen:

- Åpne oppgaver der kandidater skriver et svar og
- Lukkede oppgaver der elevene velger fra gitte svaralternativer

Åpne oppgaver der kandidater skriver et svar er svært vanlige i eksamen. Et annet begrep for disse er «constructed response-oppgaver» (CR), for å markere at eleven konstruerer svaret selv, i motsetning til lukkede oppgaver med gitte svaralternativer (bl.a. flervalgsoppgaver), hvor eleven velger ett av flere alternativer. En åpen oppgave er å betrakte som en instruks til eleven: «Bruk faglig kompetanse for å utføre denne handlingen». Konkrete eksempler på slik fagkompetanse kan være en matematikkoppgave som skal prøve elevens matematikkfaglige kompetanse til å «lage, løse og forklare likningssett knytte til praktiske situasjoner», eller en norskoppgave som skal prøve elevens kompetanse til å «utforske språklig variasjon og mangfold i Norge og reflektere over holdninger til ulike språk og talespråkvarianter».

Lukkede oppgaver der eleven velger ett av flere alternativer kalles også «selected response» (SR). Det mest kjente lukkede formatet er flervalgsoppgaven som består av et spørsmål eller en annen instruerende tekst og vanligvis to til fire svaralternativ. Et av alternativene er korrekt (den såkalte nøkkelen), og de andre alternativene fungerer som «distraktorer». Det finnes også andre typer lukkede oppgaver der eleven velger sitt svar fra ett eller flere alternativer, og eksempler på dette ligger i det nye gjennomføringssystemet for eksamen og prøver.

Å lage lukkede oppgaver er generelt vanskeligere og mer tidkrevende enn å lage åpne oppgaver, og krever statistisk analyse. Gevinsten med disse er at de kan prøve kandidatens kompetanse uten at skriveferdighetene påvirker resultatet. Kandidaten får vist sin kompetanse på flere oppgaver og de kan skåres automatisk. Begge deler gjør resultatet mer pålitelig. Det må imidlertid vurderes nøye i hvert enkelt fag hvilken kompetanse i læreplanen som eventuelt kan egne seg til denne typen oppgaver. Dette må dere vurdere om dere lager flervalgsoppgaver:

- En oppgave skal fokusere på ett aspekt av kompetansen.
- Alle oppgaver må være uavhengige av hverandre, dvs. at det ikke skal være mulig å svare riktig på en oppgave fordi en har svart riktig på foregående.
- Oppgaven bør være så kortfattet og tydelig som mulig.
- Stammen skal ikke inneholde negasjoner.

- Alle svaralternativer skal være plausible.
- Svaralternativer skal være omtrent like lange.
- Svaralternativer skal ikke inneholde doble negasjoner.
- Svaralternativer skal presenteres i logisk rekkefølge (for eksempel i alfabetisk eller nummerert rekkefølge).
- Kun ett svaralternativ skal være korrekt.

## 5.2. Lag oppgaveinstruksjoner som er tydelige for kandidaten

Oppgavebestillingen må være forståelig for kandidaten og skal skrives slik at kandidaten ikke misforstår oppgaven. Det at oppgavebestillingen er tydelig, betyr likevel ikke at den ikke kan være kompleks/sammensatt. Dersom det er forberedelsesdel til eksamen, skal det fremgå tydelig hvordan denne skal forberede kandidatene på utfordringer de kan møte på eksamen. Det skal også fremgå tydelig i eksamensoppgavene hvordan kandidatene skal bruke det de har lært i forberedelsestida på eksamensdagen.

## 5.3. Lag en eksamen som er så rettferdig som mulig

Nedenfor er noen eksempler på hva dere må passe på for å unngå bias i oppgavene <sup>1</sup>. Begrepet bias kan oversettes med uheldig utfall, for eksempel at en eksamensoppgave utilsiktet favoriserer jenter, og dermed bidrar til at guttene skårer dårligere på den samme eksamen.

- Oppgaver bør formuleres så tydelig og forståelig som mulig: unødvendig vanskelig syntaks og språkbruk skal unngås (det inkluderer regionale varianter, faste uttrykk som ikke kan antas være kjent av alle elever, f.eks. «å hoppe etter Wirkola» og lignende)
- I eksamener som ikke krever det, bør emner, som det er grunn til å tro at visse elevgrupper vet mere om enn andre, unngås. Et eksempel er en spanskeksamen der alle oppgaver er knyttet til fotball.
- Det er også viktig å vurdere hvordan oppgaver, i fag der dette er relevant, introduserer sensitive temaer på en måte som ikke støter kandidater unødvendig.
- Oppgavene skal utformes på en slik måte at alle elever, uansett eventuelle spesialpedagogiske behov, skal kunne ta eksamen (universell utforming). Hvis oppgavene trenger tilpasninger skal disse være mulig å gjøre uten at oppgaven blir vesentlig forandret.

---

1) Listen er basert på Zieky (2016), som er et kapitel i Lane et al. (2016).

## 5.4. Gi tydelig informasjon om tidsrammer og hjelpemidler

En overordnet ramme for sentralt gitt skriftlig eksamen er at den skal vare i maksimalt fem timer. Det er viktig å spesifisere hvor lang tid en elev skal/bør/kan bruke på en oppgave, og hvilket omfang en oppgave må/kan ha. Slike tids- og omfangsfaktorer er en viktig del av instruksjonen fordi de skaper forutsigbarhet og trygghet for kandidaten, og de påvirker også hva som er mulig å si om kandidatens kompetanse.

Avhengig av hva en oppgave skal prøve, vil hjelpemidler og tilpasninger kunne påvirke hvordan svaret kan tolkes som et bevis på faglig kompetanse. Tilgang til eller begrensninger i digitale eller andre former for hjelpemidler til eksamen, bør begrunnes ut fra kompetansemålene i læreplanen, fagenes egenart, fagets nivå og eventuelle nye digitale oppgaveformater. Vurderingene må alltid ses opp mot gjeldende hjelpemiddelordning som rammer inn bruken av hjelpemidler på eksamen. Oppgavebestillingen skal bidra til at kandidatene bruker hjelpemidler på en hensiktsmessig måte for å løse oppgavene. Det må også fremgå i vurderingskriteriene hvordan bruk av hjelpemidler skal vurderes.

## 5.5. Lag tydelige vurderingskriterier for hvordan et kandidatsvar skal vurderes

For åpne oppgaver der elevene skal skrive et svar, skal det utvikles vurderingskriterier. Kandidaten og sensor må få kjennskap til hvilke kriterier som skal ligge til grunn for vurdering av kandidatens besvarelse.

Kandidaten og sensor må også være kjent med om noen sider ved svaret skal vektas mer enn andre. For lukkede oppgaver og flervalgsoppgaver trenger vi ikke vurderingskriterier fordi jobben allerede er gjort når en lager alternativene med fasit.

Vurderingskriteriene skal lages slik at analytisk vurdering er mulig. Med analytisk vurdering mener vi at vurderingen skal deles opp i ulike vurderingsområder for å vise tydelig hvilke deler som til sammen sier noe om kandidatens samlede kompetanse. Disse kriteriene skal så langt det er mulig være de samme fra år til år. Nedenfor presenterer vi noen retningslinjer for hvordan kriterier skal utvikles og brukes. I eksemplet nedenfor er kriteriene stilt opp som en matrise.

Vurderingskriterier eksamen (eksempel)

	Nivå 1	Nivå 2	Nivå 3	Nivå 4	Nivå 5	Nivå 6	Skår
<b>Vurderingsområde 1</b>							
<b>Vurderingsområde 2</b>							
<b>Vurderingsområde 3</b>							
<b>Vurderingsområde 4</b>							
<b>Vurderingsområde 5</b>							
Sumskår:							6-

I matrisen skal det fremkomme at eleven vil få et sumskår (samleskår) på oppgaven. Maksverdien vil være avhengig av antall vurderingsområder. I en matrise med fem områder og seks nivåer er maksimal sumskår 30. Med tre oppgaver er maksimalt eksamensresultat 90. Eksempelet her viser 6 nivåer, men det kan også være færre eller flere, avhengig av hvor mange kvalitetsnivåer det er mulig å skille mellom. Hvordan resultatet skal «oversettes» til en eksamenskarakter må vurderes av oppgaveutviklere sammen med Udir, og må tilpasses blant annet utformingen av læreplaner og fagenes egenart.

## 6. Hvordan kan vi styrke reliabiliteten i eksamen?

I oppgaveutvikling kan vi bidra til høy reliabilitet bl.a. ved å bruke flere oppgaver og oppgavetyper og ved å lage tydelige oppgaveinstruksjoner og vurderingskriterier. Dette kan minske effekten av at sensorer ikke vurderer helt likt. Vi kan også sjekke sensorreliabiliteten på forhånd i forbindelse med utprøving av vurderingskriterier og utprøving av oppgaver. Den evalueres også etter gjennomført eksamen i forbindelse med forberedelser til neste eksamensutvikling. Udir bestemmer hvem som deltar i dette arbeidet.

### Sensorreliabilitet

En av de største reliabilitetsutfordringene i oppgaver som krever vurdering er «sensor-effekter».

Sensoreffekter handler om at variasjon i eksamensresultatene er knyttet til sensors vurdering og ikke til kandidatens faktiske prestasjoner (Scullen, Mount, & Goff, 2000, p. 157). Lav sensorreliabilitet innebærer enten at én sensor gir ulik skår ved gjentatte vurderinger av samme oppgave, eller at to eller flere sensorer skårer samme oppgave ulikt.

Det finnes mange grunner til hvorfor to eller flere sensorer ikke ender opp med samme konklusjon om hvilket resultat en kandidatprestasjon bør ha. Noen av de vanligste grunnene er:

- Ulik skalabruk: sensorer har vist seg å bruke vurderingsskalaer ulikt. Noen tenderer mot å bruke midtdelen av skalaen og noen tenderer mot å bruke en av to ytterpunkter.
- Ulik strenghet: selv om sensorer er enige om hva som skal vurderes kan de være uenige i hvor strengt kandidatsvar skal vurderes.
- “Haloeffekt”: sensorer kan iblant la ett aspekt eller én dimensjon overskygge vurderingen av kandidatsvaret. For eksempel har det vist seg at sensorer kan gi lav skår på en kandidattekst som ellers er god, men preges av manglende tegnsetting.
- Tilnærming til vurdering: Det har vist seg at sensorer kan ha ulik forståelse av formålet med sensur; noen kan oppfatte det som viktigst å være kandidatens «advokat», mens andre kan oppfatte det som viktigst å være strikt i bruken av vurderingskriterier.

Andre grunner til uenighet mellom sensorer er ulik forståelse av konstruktet og/eller vurderingskriteriene. Når dette er tilfelle er uenigheten mer et validitetsspørsmål, som krever sensorskolering og/eller revisjon av kriterier slik at de blir mer brukbare.



## 7. Oppsummering

For at eksamen skal kunne møte kravene til validitet, reliabilitet, rettferdighet og håndterbarhet må oppgaveutviklingen gjøres ut fra den beskrivelsen dere har laget av kompetansen og det faglige innholdet (konstruktet), som baseres på læreplanen. Kandidatene skal ha mulighet til å vise kompetanse på flere og varierte måter, noe som innebærer at en eksamen må inneholde ulike typer oppgaver. Eksamen skal inneholde oppgaver som i størst mulig grad skal kunne løses av alle kandidater, inkludert kandidater med særskilte behov. Det er også viktig at eksamen legger til rette for at de som tar eksamen skal kunne bruke tillatte hjelpemidler på en hensiktsmessig og sammenlignbar måte. Ved utvikling av eksamen må dere i tillegg sørge for at oppgavene ikke har innhold som kan oppleves støtende. Videre må dere sørge for at vurderingskriterier, antall oppgaver og vurderingsprosedyrer bidrar til høy nok reliabilitet.

For å sikre validitet, reliabilitet, rettferdighet og håndterbarhet skal oppgaver også piloteres. Det betyr at de prøves ut i forkant av eksamen og analyseres med psykometriske metoder.

### Kildeliste

- Cizek, G. J. (2012). Setting performance standards: Foundations, methods, and innovations (2nd ed.). Routledge.

+ Vis mer