



## Utforming av vurderingskriterier og sensorveiledninger – sentrale spørsmål og avveininger

Hege Hermansen



UiO **LINK** – Senter for læring og utdanning  
Universitetet i Oslo

## Innhold

Introduksjon	2
Sensorveiledninger og vurderingskriterier: noen grunnleggende begreper	3
Sensorveiledninger og vurderingskriterier i sensurprosessen	5
Sensorveiledninger og vurderingskriterier som del av studentenes læringsprosess	10
Digitale verktøy ved UiO til bruk i vurdering	13

## 1. Introduksjon

Fra høsten 2018 er bruk av sensorveiledninger obligatorisk ved Universitetet i Oslo. Dette notatet skal være en støtte til undervisere ved UiO i arbeidet med å utvikle sensorveiledninger og vurderingskriterier, ved å skissere noen grunnleggende prinsipper og sentrale spørsmål og avveininger knyttet til denne tematikken. Notatet er basert på norsk og internasjonal vurderingsforskning, og referer også til relevante dokumenter og retningslinjer utarbeidet ved UiO.

Sensorveiledninger bidrar til å styrke studenters rettssikkerhet ved å gjøre grunnlaget for vurderingen eksplisitt, og gi undervisere, sensorer og studenter et felles referansepunkt for hva som skal ligge til grunn for sensurprosessen (Reddy & Andrade, 2010; Sadler, 2012). Sensorveiledninger er derfor et sentralt verktøy for å håndtere eventuell uenighet mellom sensorer, og for å sikre at en klagesensur behandles på like vilkår som den opprinnelige sensuren. På denne måten støtter også sensorveiledninger opp om vår juridiske forpliktelse til å "sørge for at kandidatenes kunnskaper og ferdigheter blir prøvet og vurdert på en upartisk og faglig betryggende måte" (Universitets- og høyskoleloven, § 3-9., første ledd).

Sensorveiledninger kan også bidra til å skape helhet og sammenheng mellom læringsmålene for et emne/program, eksamensoppgavene og sensuren (Biggs & Tang, 2011). Mye forskningslitteratur har påpekt mangel på samsvar mellom eksamensformer, læringsmål og undervisningsaktiviteter i høyere utdanning (Havnes & Raaheim, 2009). Sensorveiledninger er et av flere verktøy som kan bidra til å styrke disse sammenhengene, ved å fungere som et bindeledd mellom læringsmålene i emne/programbeskrivelser og eksamensformer og sensur. Sensorveiledninger og vurderingskriterier som er tilgjengelige for studenter i forkant av eksamen kan også fungere som støtte i deres læringsprosess, ved å tydeliggjøre de sentrale aspektene av faget.

Veilederen er organisert i tre deler. Den første delen presenterer noen grunnleggende begreper som er relevante for sensorveiledninger og kriteriebasert vurdering. Den andre delen fokuserer på selve sensurprosessen. Sentrale temaer her er hvordan sensorveiledninger kan støtte opp under denne prosessen, og hvordan ulike egenskaper ved sensorveiledninger kan påvirke sensurarbeidet. Den tredje delen setter sensorveiledninger i sammenheng med de andre lærings- og vurderingsaktivitetene som foregår innenfor rammen av et gitt emne/utdanningsprogram. Sentrale temaer her er hvordan bruk av sensorveiledninger og vurderingskriterier kan være med på å styrke helhet og sammenheng mellom læring, undervisning og vurdering (Solbrekke & Karseth, 2016), og dermed støtte opp om studenters læring.

## 2. Sensorveiledninger og vurderingskriterier:

### Noen grunnleggende begreper

#### 2.1. Kriteriebasert og normbasert vurdering (Havnes & Raaheim, 2009; Sadler, 1987)

Historisk har mye vurdering innenfor høyere utdanning vært *normbasert* (Brookhart *m. fl.*, 2016). I praksis betyr dette at man, eksplisitt eller implisitt, har fordelt karakterer etter en normalfordelingskurve som har indikert omtrent hvor mange studenter som burde få A, hvor mange som burde få B, og så videre. En grunnleggende svakhet ved normbasert vurdering er at karakteren en student får avhenger av kvaliteten på besvarelsene for resten av studentkullet.

I kontrast til dette er et grunnleggende premiss ved *kriteriebasert* vurdering at studenten kun skal vurderes opp mot et sett vurderingskriterier som beskriver utvalgte kvaliteter ved studentens arbeid. Utvikling av slike vurderingskriterier vil ofte handle om å sette ord på standarder som allerede eksisterer innenfor et gitt fagfelt. Historisk har slike standarder ofte vært implisitte, og nykommere har blitt gradvis sosialisert inn en forståelse av hva som kjennetegner faglig kvalitet (for eksempel slik den kommer til uttrykk i en eksamensbesvarelse). Ved å eksplisere dem blir disse kriteriene tilgjengelig som et tydelig referansepunkt hva som skal vektlegges i sensurprosessen (Reddy & Andrade, 2010; Sadler, 2012). Eksplisitte kriterier kan også gjøre det enklere for studenter å forstå hva som vektlegges i vurdering av eksamen.

#### 2.2. Summativ og formativ vurdering (Boud & Molloy, 2013; Black & Wiliam, 1998; 2009)

I vurderingsforskningen trekkes man et overordnet skille mellom *summativ* og *formativ* vurdering. Summativ vurdering har som hensikt å gi en endelig kvalitetsvurdering av et sluttprodukt, for eksempel ved å sette en karakter på en eksamensbesvarelse. Formativ vurdering er vurderingssituasjoner som skjer underveis i læringsløpet, og som har som hovedhensikt å hjelpe studentene videre i læringsprosessen.

Summativ vurdering	Formativ vurdering
<ul style="list-style-type: none"><li>- Vurdering av kvaliteten på et sluttprodukt («eksamen»)</li><li>- Inntreffer på slutten av et undervisningsforløp</li><li>- Har som formål å kvalifisere/sertifisere/rangere studenter</li></ul>	<ul style="list-style-type: none"><li>- Vurdering for å øke kvaliteten på et (fremtidig) sluttprodukt</li><li>- Skjer underveis i undervisningsforløpet</li><li>- Har som formål å støtte studenters læring, og å foreta justeringer av undervisningen ved behov</li></ul>

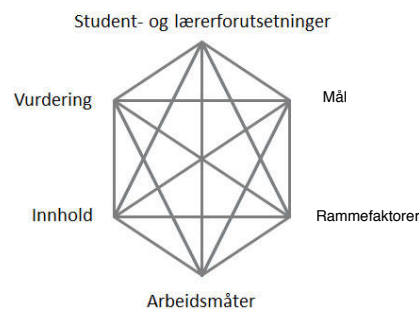
Forskning viser at mange studenter strukturerer sine egne læringsprosesser basert på hva de tror er viktig for eksamen (Havnes & Raaheim, 2016). Det er derfor god grunn til å se formativ og

summativ vurdering i sammenheng. En måte å gjøre dette på kan være at vurderingskriteriene som legges til grunn for eksamen diskuteres med studentene underveis i undervisningsløpet. En annen tilnærming er å gi studentene formative vurderingssituasjoner underveis i læringsprosessen som tester de samme kunnskapene og ferdighetene som etterspørres på eksamen.

#### 2.4. Helhet og sammenheng mellom læringsmål, vurdering og undervisning (Biggs, 1989; Biggs & Collis, 1989; Solbrekke & Karseth, 2016).

Et funn som går igjen i forskningslitteraturen er at studenter opplever manglende sammenheng mellom læringsmålene, vurderingsformene og undervisningsaktivitetene i et emne (Havnes & Raaheim, 2016; Wijngaards-de Meij & Merx, 2018). Dette kan for eksempel handle om det som måles på eksamen er noe annet enn det som står i fokus for undervisningen. Samtidig viser forskning at en tydelig sammenheng mellom læringsmål, vurderingsformer og undervisningsaktiviteter kan styrke studenters læring (Biggs & Collis, 1989). Mer generelt kan det i planlegging av undervisning og vurdering være nyttig å tenke gjennom det gjensidige samspillet mellom (Engelsen, 1992, s. 38):

- læringsmål
- vurderingsformer
- det faglige innholdet
- studentenes faglige forutsetninger
- arbeidsformene studentene må delta i gjennom undervisningen
- generelle rammebetingelser for lærings- og vurderingsaktivitetene.



Sensorveiledninger og vurderingskriterier kan fungere som et bindeledd mellom målene for undervisningen (slik de er definert på emnesiden), summative og formative vurderingsformer, og de læringsprosessene som studentene engasjeres i gjennom studieløpet. Flere studier tyder også på at aktiv bruk av vurderingskriterier gjennom hele undervisningsløpet bidrar til studentenes læring og deres forståelse av sammenhengene mellom læringsmål, undervisningsaktiviteter og eksamen (Ball, *m. fl.*, 2014, Boud & Molloy, 2013).

### 3. Sensorveiledninger og vurderingskriterier i sensurprosessen

#### 3.1. Overordnede tilnæringer til utforming av sensorveiledninger

Arbeidsgruppen for innføring av sensorveiledninger ved UiO vektlegger i sin rapport at sensorveiledninger ikke er et ensbetydende begrep, og at utformingen av veiledninger dermed vil avhenge av kontekst. Et overordnet skille kan trekkes mellom *generiske* og *spesifikke* sensorveiledninger (Jonsson & Svingby, 2007; Ludvigsen *m. fl.* 2018).

Generiske sensorveiledninger inneholder vurderingskriterier som beskriver generelle egenskaper ved en studentbesvarelse, og som kan være relevante for mer enn en eksamen. Eksempler på slike kriterier kan være:

- "Kandidaten viser sterk forståelse av relevant pensumlitteratur, og evner å analysere og kritisk diskutere denne litteraturen".
- "Kandidaten har formulert en presis problemstilling, og argumentert for den forsknings- og samfunnsmessige relevansen av problemstillingen".

Spesifikke sensorveiledninger vil i større grad være aktuelle på eksamensformer hvor det forventes et korrekt svar, for eksempel ved matematiske utregningsoppgaver eller faktabaserte spørsmål. Slike sensorveiledninger vil typisk inkludere en fasit eller rettenøkkel, samt en veiledning til sensorene om hvordan man skal forholde seg til delvis riktige svar.

Generiske sensorveiledninger (Vurderingskriterier)	Kombinasjon av generiske og spesifikke sensorveiledninger	Spesifikke sensorveiledninger (Fasit/Løsningsforslag)
F.eks. masteroppgaver, større semesteroppgaver mm	F.eks. tekstoppgaver, langvaroppgaver mm	F.eks. eksamener med konkrete svaralternativ
Utarbeides i forkant, tar utgangspunkt i emnebeskrivelsen	Utarbeides i forkant, underveis eller i etterkant avhengig av eksamensform og –oppgaver	Utarbeides underveis eller i etterkant, tar utgangspunkt i eksamensoppgavene
Generelle retningslinjer for vurdering av besvarelsene	Kan inneholde både generelle retningslinjer for vurdering av besvarelsene og konkrete momenter på enkeltoppgaver	Fasit/rettenøkkel med angivelser av riktige svar og hvordan alternative/delvis riktige svar skal vurderes
Stabile over tid, gjenbrukes flere semestre	Vesentlige elementer er stabile over tid, men tilpasses den enkelte eksamen	Utarbeides for hver eksamen og tilpasses oppgavene
Kan være tilgjengelige for kandidatene før eksamen	Offentliggjøre normalt for kandidatene etter eksamen	Offentliggjøres for kandidatene først etter eksamen

Disse to tilnærmingene til sensorveiledninger er ikke gjensidig utelukkende, og kan kombineres (Ludvigsen *m. fl.*, 2018). Videre vil fasitsvar eller rettenøkler som regel gjenspeile mer generiske kriterier. For eksempel kan en utregning i eksamensoppgave som kun har et gyldig svar reflektere det mer generiske vurderingskriteriet "Kandidaten kan sette opp, gjennomføre og tolke kompliserte beregninger både analytisk og numerisk".

Arbeidsgruppen understreker at denne tabellen ikke skal begrense mulighetene for utformingen av sensorveiledninger, men illustrere hva slags problemstillinger fagmiljøene bør ta utgangspunkt i og drøfte. Andre sentrale utgangspunkt for utformingen av sensorveiledninger inkluderer:

- læringsutbyttebeskrivelsene for emnet og mål for utdanningsprogrammet
- fagområdets egenart
- eksamensformen
- pensumlitteraturen
- karakterskalaen ved UiO
- undervisningsaktivitetene i emnet
- informasjon om vektingen mellom ulike deloppgaver i eksamenssettet
- informasjon om vektingen mellom de ulike vurderingskriteriene i sensorveiledningen
- informasjon om hvordan avvik fra formelle krav skal håndteres (for eksempel oppgaver som i vesentlig grad avviker fra normert sideantall).

### **3.2 Utforming og bruk av generiske vurderingskriterier**

Siden innføringen av Kvalitetsreformen i 2003 har læringsutbyttebeskrivelser vært en integrert del av emne- og programbeskrivelser i høyere utdanning (Caspersen & Frølich, 2017). Læringsutbyttebeskrivelsene redegjør for hva slags kunnskap, ferdigheter og generelle kompetanser studenter skal ha oppnådd etter et gitt utdanningsløp. Eksamensformer og vurderingskriterier utledes fra disse læringsutbyttebeskrivelsene. Vurderingskriterier beskriver hva slags nivå av kunnskap, ferdigheter eller generell kompetanse som kjennetegner eksamensbesvarelsen. Eksempler på vurderingskriterier i vurdering av en masteroppgave kan være:

- "Oppgaven har en problemstilling som er faglig interessant, relevant, begrunnet og presist formulert."
- "Det er en sterk grad av sammenheng mellom problemstillingen og valg av metode"
- "Kandidaten vurderer kritisk det empiriske materialet som presenteres"

Hovedformålet med vurderingskriterier er at de skal gjøre det eksplisitt for undervisere, sensorer og studenter hva som vektlegges i sensurprosessen (Jonsson & Svingby, 2007; Sadler 1987, 2012). Denne ekspliseringen av grunnlaget for vurderingen fungerer som et felles referansepunkt for alle som er involvert i sensurprosessen: interne og eksterne sensorer, undervisere/emneledere, studieprogramledere og studenter. Vurderingskriterier er også et sentralt verktøy for å håndtere uenighet mellom sensorer, og for å sikre at en klagesensur behandles på like vilkår som den opprinnelige sensuren (Jonsson & Svingby, 2007).

I oppgaveformer hvor det ikke finnes et korrekt svar, som for eksempel masteroppgaver eller essay-oppgaver på bachelornivå, vil det verken være mulig eller ønskelig å utvikle et uttømmende sett med vurderingskriterier som dekker alle svaralternativer (Bloxham *et al.*, 2011; Sadler, 1987, 2007; Wyatt-Smith & Klenowski 2012). Vurderingskriterier skal altså ikke erstatte sensorenes skjønnsette i sensurprosessen, men tydeliggjøre kriteriene som denne skjønnsette skal basere seg på.

Samtidig viser forskning også at menneskers vurderingsevne er sårbar for ulike kontekstuelle faktorer (Brookhart, 2012; Kahnemann 2011; Kirkebøen, 2013; Sadler, 1987). I vurderingssammenheng kan dette handle om at man blir påvirket av egenskaper ved de oppgavene man allerede har vurdert, at man ubevisst lar seg påvirke av inntrykket man har av studenten fra undervisningssituasjoner, eller at man lar seg personlig provosere av faglige argumenter som gjøres i en oppgave. Videre kan fravær av sensorsamsvar handle om at man vektet forskjellige aspekter av en besvarelse ulikt, eller at man anlegger forskjellige argumenter for hvordan et aspekt av oppgaven skal tolkes opp mot kriteriene. Det kan derfor være nyttig å angi i en sensorveiledning om alle vurderingskriteriene skal tillegges lik vekt.

Vurderingskriterier, sammen med den øvrige informasjonen som inkluderes i sensorveiledninger, bidrar til å motvirke noe av vilkårligheten som vi vet kan kjennetegne menneskelig dømmekraft (Reddy & Andrade, 2010; Van der Schaaf *et al.*, 2011). Flere studier rapporterer at undervisere i høyere utdanning opplever at sensorveiledninger bidrar til økt objektivitet, reliabilitet og konsistens i karaktersetningen (*ibid.*). Samtidig er ikke sensorveiledninger i seg selv en garanti for sensorsamsvar (Gynnild, 2013). Forskningslitteraturen fremhever blant annet betydningen av at sensorveiledninger og vurderingskriterier diskuteres på et kollektivt nivå i fagmiljøet, for at man skal oppnå en felles forståelse av hvordan veiledningene skal tolkes (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). Sensurmøter i forkant av sensurperioden er et tiltak som allerede er i bruk ved UiO for å styrke slike felles forståelser.

### 3.3. Differensiering av generiske vurderingskriterier

Differensierte vurderingskriterier gjør det enklere for sensorer å vurdere hva som skiller en A fra en B, en B fra en C, og så videre. I forskningslitteraturen finnes det ulike tilnærminger til differensiering (Krathwohl, 2002; Biggs & Collis, 1989; Sadler, 1987). To overordnede tilnærminger kan beskrives som:

- gradsdifferensiering av de samme kriteriene på tvers av karakterskalaen, og
- bruk av ulike kriterier for å beskrive besvarelser på forskjellige nivåer i karakterskalaen.

Det første tilfellet, gradsdifferensiering, ligger til grunn for den offisielle karakterskalaen ved UiO. Der brukes for eksempel "vurderingsevne" og "selvstendighet" som kriterier for å beskrive alle nivåene på skalaen fra A til F, men de graderes ved bruk av termer som "svært god", "veldig god", "en viss grad" eller "lite". For eksempel:

- Kandidaten syner særs god vurderingsevne og stor grad av sjølvstende. (A)



- Kandidaten syner god vurderingsevne og sjølvstende på dei viktigaste områda. (C)

Det andre tilfellet, der ulike kriterier brukes for å beskrive ulike karakter, har som et grunnleggende premiss at kriteriene skal beskrive en økende grad av kognitiv og faglig kompleksitet når man beveger seg oppover karakterskalaen. Dette kan for eksempel komme til uttrykk ved at karakterer mot toppen av skalaen beskrives med begreper som "analysere", "tolke", "kritisk reflektere" eller "evne til å se sammenhenger", mens karakterer mot den lavere delen av skalaen beskrives med begreper som "redegjøre for", "formidle" eller "oppsummere". For eksempel:

- Kandidaten kan redegjøre for de viktigste elementene i teoretisk perspektiv X, og beskrive hvorfor denne teorien er sentral for fagområde Y. (nedre del av skalaen)
- Kandidaten kan bruke teoretisk perspektiv X til å tolke og analysere en empirisk problemstilling, og sette denne teoretiske tilnærmingen i sammenheng med historiske utviklingstrekk innenfor fagområde Y (øvre del av skalaen).

Formulering av vurderingskriterier og utforming av eksamensoppgaver må sees i tett sammenheng. Et viktig aspekt er at eksamensoppgaven må teste det vi faktisk ønsker å prøve studentene i, slik dette er reflektert i vurderingskriterier og overordnede læringsmål. Dersom vurderingskriteriene representerer andre kunnskaper, kompetanser eller ferdigheter enn de som faktisk testes i eksamensoppgaven, vil vi få lav validitet i sensur av eksamen. Et annet aspekt er at eksamensoppgavene bør utformes på en måte som gjør det sannsynlig at man får eksamensbesvarelser som dekker hele karakterskalaen. Med andre ord bør vurderingskriteriene utformes på en måte som hjelper oss å skille svært gode besvarelser fra de som er ganske gode eller mindre gode.

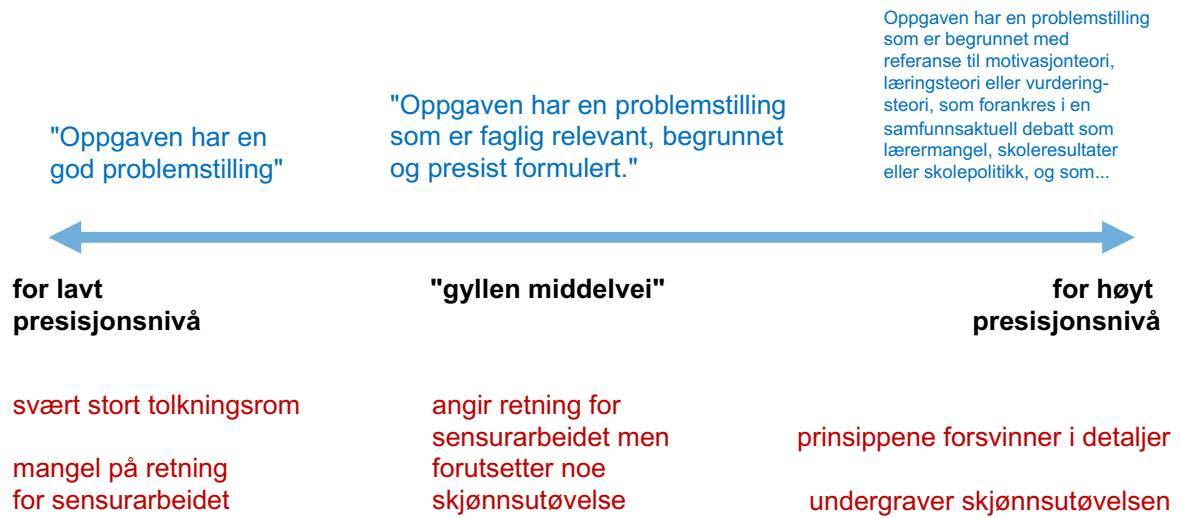
### 3.4 Kjente utfordringer ved utforming og bruk av generiske vurderingskriterier

Noen utfordringer ved bruk av vurderingskriterier er "gjengangere" i forskningslitteraturen, og det kan være nyttig å være oppmerksom på disse når man utformer vurderingskriterier.

*Kriteriene er for detaljerte og spesifikke.* Mange empiriske studier har vist at lengre lister med svært detaljerte vurderingskriterier kan invitere til overfladiske og instrumentalistiske læringsstrategier blant studenter (Torrance 2007, Sadler, 2007). En praktisk implikasjon er at listen med vurderingskriterier ikke bør bli alt for lang, eller ha et alt for høyt spesifikasjonsnivå. Samtidig er det viktig at sensorveiledningen inkluderer alle aspektene som skal ligge til grunn for vurderingsprosessen. Detaljnivået vil også variere avhengig av fagfelt og eksamensform.

*Kriteriene er for vage.* I motsatt ende kan generiske kriterier også bli for åpne og vage. Når kriterier blir for åpne gir de ikke lenger retning for sensurarbeidet, og de forutsetter så stor grad av skjønnsutøvelse at sensorveiledningen har lite praktisk nytte. En praktisk implikasjon er at kriteriene bør være spesifikke nok til at de peker mot observerbare kvaliteter ved studentarbeidet. Dette betyr ikke at meningsinnholdet er fullt ut definert - i eksemplet ovenfor kan man for eksempel fremdeles diskutere tolkningen av "forskningsmessig- og samfunnsmessig relevans".

## Presisjonsnivå på vurderingskriterier



*Manglende validitet.* Dette problemet oppstår når vurderingskriteriene ikke beskriver det man egentlig vil måle, og/eller når eksamensformen og vurderingskriteriene ikke samsvarer med læringsmålene for emnet (Jonsson & Svingby, 2007; Reddy & Andrade, 2010). En praktisk implikasjon er at kriterier spesielt, og sensorveiledninger mer generelt, bør valideres opp mot eksamensformen og læringsutbyttebeskrivelsene i emnebeskrivelsen før de tas i bruk.

## 4. Sensorveiledninger og vurderingskriterier som del av studentenes læringsprosess

Arbeidsgruppen for sensorveiledninger ved UiO (Ludvigsen *m. fl.*, 2018) konstaterer at sensorveiledninger som i stor grad inneholder generelle vurderingskriterier, kan gjøres tilgjengelig når som helst. Dette åpner for at vurderingskriteriene kan brukes som et pedagogisk verktøy i undervisningen, for eksempel gjennom bruk av formativ vurdering der studentene får kriteriebaserte tilbakemeldinger på eget arbeid (Ball, *m. fl.*, 2014; Boud & Molloy, 2013; Hattie & Timperley, 2007; Reddy & Andrade, 2010).

Forskning tyder på at systematisk bruk av kriterier i undervisningen kan støtte studenters læring på ulike måter:

### 4.1. Tydeliggjøring av faglige krav og forventninger

For studenter kan meningsinnholdet i generiske kriterier ofte være uklart (Jonsson & Svingby, 2007). De kan for eksempel ha utfordringer med å forstå hva "selvstendighet" og "vurderingsevne" betyr i konteksten av en semesteroppgave. Dersom studenter får mulighet til å jobbe systematisk med kriteriene som en integrert del av undervisningsløpet øker det sannsynligheten for at de forstår hva kriteriene krever av dem (Andrade, 2000; Reddy & Andrade, 2010). Hvordan dette arbeidet skjer vil avhenge av fagområdet, men en fellesnevner er at generiske vurderingskriterier kobles til mer spesifikt faglig innhold. Dette kan for eksempel skje gjennom

- diskusjon av pensumlitteratur, der tekstene brukes til å illustrere hvordan sentrale aspekter ved vurderingskriteriene (f.eks. «drøfte, analysere, tolke, sammenlikne») utføres innenfor det bestemte fagområdet.
- analyser av tidligere eksamensbesvarelser, der studentene f.eks. blir bedt om å vurdere hvilke vurderingskriterier de tror ligger til grunn for sensuren etter å ha lest en A-oppgave og en C-oppgave.
- formativ vurdering fra underviser til student, der studentene får tilbakemeldinger på eget arbeid underveis i studieløpet basert på et sett med vurderingskriterier.
- formativ vurdering mellom studenter, der studenter gir tilbakemeldinger på hverandres arbeid basert på et sett vurderingskriterier.

Systematisk arbeid med vurderingskriterier i undervisningen gjør at eksamensresultatet ofte kommer mindre "overraskende" på studenter. Dette er fordi de har fått mulighet til å utvikle en dypere forståelse av meningsinnholdet i vurderingskriteriene, og fordi studentarbeidet allerede er blitt vurdert opp mot kriteriene gjennom formativ vurdering. Når studenter forstår hva vurderingskriteriene representerer, forstår de også bedre hva som forventes av dem på eksamen. Dette kan bidra til færre klager på eksamenskarakterer, og til at studenter opplever vurderingen som mer rettferdig (Reddy & Andrade, 2010).

## 4.2 Innvielse i fagspesifikke tenkemåter

Mange studenter strever med å forstå hva det betyr å "tenke som" en matematiker/historiker/antropolog (sett inn navn på disiplin/fagområde/profesjon). Studenter kan for eksempel pugge definisjoner og pensumlitteratur, men mangle helhetsforståelsen av hva slags rolle definisjonene og litteraturen spiller for fagets oppbygging og arbeidsmetoder. De kan også mangle en forståelse av de grunnleggende «byggesteinene» som inngår i et faglig resonnement innenfor et bestemt fagområde. Kriterier som formuleres på en måte som gjenspeiler fagets sentrale tenkemåter og metodiske tilnæringer kan derfor være et nyttig verktøy for å støtte opp under studenters læring og deres forståelse av fagets egenart. For eksempel vil verb som «analysere», «sammenlikne», «drøfte», «undersøke», «modellere», «beregne» og «tolke» stå sentralt i mange disipliner. Samtidig vil måten disse begrepene operasjonaliseres være forskjellig innen ulike fagområder. Å koble slike kriterier til fagspesifikt innhold som del av undervisningsaktiviteter (jamfør punkt 4.1 ovenfor) kan derfor hjelpe studentene å utvikle en forståelse av fagspesifikke måter å resonnerer på.

### 4.35 Styrke helhet og sammenheng mellom læringsmål, undervisning og vurdering

Et gjentakende funn i forskningslitteraturen er at studenter opplever svak grad av sammenheng mellom undervisningen, læringsmålene for et emne, og det de blir bedt om å gjøre på eksamen (Havnes & Raaheim, 2016). Motsatt er det slik at når studenter klarer å se disse sammenhengene, så bidrar dette positivt til deres opplevelse av læring. Vurderingskriterier kan representere et konkret bindeledd mellom a) læringsmålene slik de er definert i emnebeskrivelsen, b) det som skjer i undervisningen og b) eksamensformene. Dersom kriterier i tillegg brukes underveis i læringsprosessen vil dette kunne styrke studenters opplevelse av helhet og sammenheng. To eksempler kan være å gi studentene tilbakemeldinger på arbeidskrav underveis i studiet, eller å la studenter kommentere på hverandres tekst/oppgaveutkast eller løsningsforslag. I begge tilfeller kan tilbakemeldingene ta utgangspunkt i et sett med vurderingskriterier. Undervisere som jobber systematisk med vurderingskriterier rapporterer også at de selv opplever økt bevissthet om hvordan de legger opp undervisningen og hva de ønsker å oppnå med de ulike undervisningsaktivitetene.

### 4.4 Styrke studenters bevissthet om egen læring

Vurderingskriterier er også et verktøy for å styrke studenters evne til selv-regulering og metakognisjon (Andrade, 2005; Reddy & Andrade, 2010). Begrepene selv-regulering og metakognisjon (Bransford, Brown, & Cocking; Schunk & Zimmerman, 2012) referer til studenters evne til å overvåke sine egne læringsprosesser og tilpasse sine studiestrategier underveis i studieløpet. Dette anses som viktig fordi forskning tyder på at det å innta et bevisst forhold til egne læringsprosesser øker sannsynligheten for et høyt læringsutbytte (ibid.).

Mer konkret handler dette om å kontinuerlig vurdere hvor man står faglig, og hvordan man kan

best jobbe videre for å nå et bestemt mål. Læringsmål og vurderingskriterier er sentrale verktøy i denne prosessen fordi de gir studentene informasjon om hvor de skal og hva som kjennetegner læring innenfor et bestemt fag. Med andre ord representerer vurderingskriteriene et sett med faglige standarder som kan brukes for å styrke studenters forståelse av sin egen faglige progresjon. Flere studier vektlegger betydningen av at studenter introduseres til vurderingskriterier på et tidlig stadium i læringsprosessen for at disse prosessene skal bli fruktbare (Andrade 2000; Osana & Seymour 2004; Tierney & Simon 2004; Song 2006).

## 5. Digitale verktøy ved UiO til bruk i vurdering

*Inspira* er en digital eksamensløsning som brukes av mange fagmiljøer ved UiO. En veiledning for utvikling av digitale sensorveiledninger i *Inspira* er tilgjengelig her:

<https://www.uio.no/for-ansatte/arbeidsstotte/sta/eksamen/digital-eksamen/veiledninger/eksamensadministrator/sensorveiledning.html>

*Canvas* har en test-funksjon og en oppgave-funksjon som kan brukes til formativ og summativ vurdering underveis i studiet. Veiledninger er tilgjengelig her:

<http://no.guides.instructure.com/m/68501>

*Mentimeter* er et studentrespons-verktøy som blant annet kan brukes til formativ vurdering underveis i studieløpet. Informasjon om *mentimeter* er tilgjengelig her:

<https://www.uio.no/tjenester/it/utdanning/srs/>

## Relevant litteratur

- Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13-19.
- Andrade, H., & Du, Y. (2005). Student perspectives on rubric-referenced assessment. *Practical Assessment, Research & Evaluation*, 10(3), 1-11.
- Biggs, J., & Collis, K. (1989). Towards a model of school-based curriculum development and assessment using the SOLO taxonomy. *Australian Journal of Education*, 33(2), 151 – 163.
- Biggs, J. B. (1989). Approaches to the Enhancement of Tertiary Teaching. *Higher Education Research and Development*, 8(1), 7-25.
- Biggs, J., & Tang, C. (2011). Train-the-Trainers: Implementing Outcomes-Based Teaching and Learning in Malaysian Higher Education. *Malaysian Journal of Learning and Instruction*, 8, 1-19.
- Black, P., & Wiliam, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74.
- Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31.
- Bloxham, S., Boyd, P., & Orr, S. (2011). Mark my words: the role of assessment criteria in UK higher education grading practices. *Studies in Higher Education*, 36(6), 655-670.
- Boud, D., & Molloy, E. (2013). Rethinking models of feedback for learning: The challenge of design. *Assessment & Evaluation in Higher Education*, 38(6), 698-712.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: brain, mind, experience, and school (expanded ed)* The National Academies Press. Washington, D. C.: National Academy Press.
- Brookhart, S. M. (2012). The use of teacher judgement for summative assessment in the USA. *Assessment in Education: Principles, Policy & Practice*, 1-22.
- Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F. et al. (2016). A Century of Grading Research. *Review of Educational Research*, 86(4), 803-848.
- Caspersen, J. & Frølich, N. (2017). Higher education learning outcomes - transforming higher education? *European Journal of Education*, 52(1), 3-7.
- Engelsen, B. U. (1992). *Kan læring planlegges?* Oslo: Ad Notam Gyldendal.
- Fraile, J., Panadero, E., & Pardo, R. (2017). Co-creating rubrics: The effects on self-regulated learning, self-efficacy and performance of establishing assessment criteria with students. *Studies in Educational Evaluation*, 53, 69-76.
- Frølich, N., Opheim, V., Brandt, S., & Prøitz, T. S. (2009). *Hva er viktige kvaliteter ved god sensur? En kartlegging av bruk av ekstern sensor på lavere grad med fokus på kvalitet, økonomi, vurdering og læring*. Oslo: NIFU
- Gibbs, G., & Simpson, C. (2005). Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education*, 1), 3-31.
- Gynnild, V. (2013). Kriteriebasert vurdering – hva innebærer det i praksis. *Uniped*, 36(1).
- Handley, K., & Williams, L. (2011). From copying to learning: using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, 36(1), 95-108.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1).
- Havnes, A. & Raaheim, A. (2016). Vurdering og eksamen. I Strømsø, H., Lycke, K. og Lauvås, P. (red.) *Når læring er det viktigste. Undervisning i høyere utdanning*. Oslo: Cappelen Akademisk Forlag.
- Imrie, B. W. (1995). Assessment for Learning: quality and taxonomies. *Assessment & Evaluation in Higher Education*, 20(2), 175-189.
- Jakobsen, A. & Lauvås, P. (2002). *Exit eksamen – eller? Former for summativ evaluering i høyere utdanning*. Oslo: Cappelen Damm Akademisk
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130-144.
- Kahneman, D. (2011). *Thinking, fast and slow*. Basingstoke: Macmillan.
- Kirkebøen, G. (2013). Kan vi stole på fagfolks skjønn? I A. Molander & J.-C. Smeby (Eds.), *Profesjonsstudier II* (pp. 27-44). Oslo: Universitetsforlaget.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory Into Practice*, 41(4), 212-218.
- Ludvigsen, S. m. fl./ Arbeidsgruppen for sensorveiledninger. (2018). *Notat. Sensorveiledninger ved Universitetet i Oslo*. Notat av 15. januar 2018.
- O'Donovan, B., Price, M., & Rust, C. (2008). Developing student understanding of assessment standards: a nested hierarchy of approaches. *Teaching in Higher Education*, 13(2), 205-217.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129-144.
- Reddy, Y. M., & Andrade, H. (2010). A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, 35(4), 435-448.
- Sadler, D. R. (1987). Specifying and Promulgating Achievement Standards. *Oxford Review of Education*, 13(2), 191-209.
- Sadler, D. R. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, 30(1), 1-11.

- Evaluation in Higher Education*, 30(2), 175-194.
- Sadler, D. R. (2009). Indeterminacy in the use of preset criteria for assessment and grading. *Assessment & Evaluation in Higher Education*, 34(2), 159-179.
- Sadler, D. R. (2010). Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5), 535-550.
- Sadler, D. R. (2012). Assuring academic achievement standards: From moderation to calibration. *Assessment in Education: Principles, Policy & Practice*, 20(1), 5-19.
- Schunk, D. H., & Zimmerman, B. J. (red.). (2012). *Motivation and self-regulated learning: Theory, research, and applications*. New York & London: Routledge.
- Solbrekke, T. & Karseth, B. (2016). Kvalifikasjonsrammeverk og læringsutbytte: Til nytte eller besvær? I. I Strømsø, H., Lycke, K. og Lauvås, P. (red.) *Når læring er det viktigste. Undervisning i høyere utdanning*. Oslo: Cappelen Akademisk Forlag.
- Torrance, H. (2007). Assessment as learning? How the use of explicit learning objectives, assessment criteria and feedback in post-secondary education and training can come to dominate learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281-294.
- Universitets- og Høgskolerådet. (2016). Karakterbruk i UH- sektoren 2016. Oslo: UHR
- Van der Schaaf, M., Baartman, L., & Prins, F. (2011). Exploring the role of assessment criteria during teachers' collaborative judgement processes of students' portfolios. *Assessment & Evaluation in Higher Education*, 1-14.
- Wijngaards-de Meij, L., & Merx, S. (2018). Improving curriculum alignment and achieving learning goals by making the curriculum visible. *International Journal for Academic Development*, 23(3), 219-231.
- Wollenschläger, M., Hattie, J., Machts, N., Möller, J., & Harms, U. (2016). What makes rubrics effective in teacher-feedback? Transparency of learning goals is not enough. *Contemporary Educational Psychology*, 44-45, 1-11.