

*Bidragsformat: Muntlig presentasjon*

# Hvorfor karakterer ikke gir mening

Per-Odd Eggen

<sup>1</sup>IFY, NTNU, Norge

---

**Sammendrag:** Presentasjonen tar utgangspunkt i offentlige dokumenter om vurdering fra regjeringen.no, NOKUT og universitetene. Det argumenteres for at inndeling av karakterer på en seksdelt skala ikke kan være et presist mål for kandidatens kompetanse og at karakterene har liten informasjonsverdi, gitt at vurderingen skjer i tråd med kvalitetsrammeverket. NIFUs arbeidsgiverundersøkelse fra 2022 viser liten sannsynlighet for at karakterene har særlig nytteverdi ved ansettelse. Til slutt drøftes konsekvensene av vurderingsordningen for kvaliteten av utdanningen og det blir delt noen erfaringer fra utprøving av karakterfri utdanning ved NTNU. Konklusjonen er at karaktergivingen delvis er misvisende, har liten bruksverdi, at den i stor grad påvirker utdanningen negativt og at det finnes gode, karakterfrie alternativer.

---

Nøkkelord:

Vurdering, karakterer, karakterfri

# Hvorfor karakterer ikke gir mening

## 1 Presentasjon

### 1.1 Hvor presis er karakterenes gradering av kompetanse?

Det er nå 25 år siden St.meld. nr. 27, 2000-2001, punkt 5.3.6 foreslo å fase ut den gamle karakterskalaen med 31 ulike bestått-trinn, fra 1,0 til 4,0. De påpekte at det særlig var «relevant å stille spørsmål ved om man kan vurdere med tidels nøyaktighet prestasjoner i fag der det nødvendigvis er høy grad av skjønn i vurderingen.» [St.meld.nr. 27 (2000-2001) del 2]. Det er rimelig å anta at dagens fem beståttnivåer gir høyere reliabilitet i vurderingen enn tilfellet var før år 2000, men det er fortsatt mange spørsmål som kan og bør stilles om universitetenes vurderingspraksis.

Etter UH-loven og Norsk kvalitetsrammeverk for høyere utdanning (NKR), skal alle UH-emner ha læringsutbyttebeskrivelser som viser at en kandidat har et sett med kunnskaper, ferdigheter og generell kompetanse etter å ha fullført et emne: «Det forutsettes at læringsutbyttet for rett nivå (6 for bachelor, 7 for master, 8 for doktorgrad) brukes ved utarbeidelse av læringsutbyttebeskrivelser i studieplanene ved institusjonene og ved akkreditering av studier<sup>1</sup>.»

Vurderingen skal si «hvorvidt studenten har oppnådd det intenderte læringsutbyttet og eventuelt hvor god kompetanse hun eller han viser»<sup>2</sup>.

Det er flere grunner til at dette er vanskelig å oppnå. Et grunnleggende kjennetegn på læringsutbyttebeskrivelsene, er at de ikke angir noe *nivå* for det som skal oppnås. Det står for eksempel at en kandidat på bachelor-nivå «kan anvende faglig kunnskap og relevante resultater fra forsknings- og utviklingsarbeid på praktiske og teoretiske problemstillinger og treffe begrunnede valg». Hva dette betyr, blir det opp til den enkelte emneansvarlige å finne ut av. Det samme gjelder hva som skiller en testsituasjon som viser disse kvalitetene, og dermed skal gi bestått (karakter E eller bedre) fra en som ikke er bestått.

En skriftlig eksamen gir begrenset informasjon om kompetansen den skal gjenspeile. Besvarelsene kan neppe si noe om studenten kan «oppdatere sin kunnskap innenfor fagområdet» eller kan «reflektere over egen faglig utøvelse og justere denne under veiledning», noe som skal være med i læringsutbyttebeskrivelsene<sup>3</sup>. I en travel undervisningshverdag, der mange kandidater skal vurderes, har klagerett og karakteren skal settes innen tre uker etter eksamen, blir løsningen oftest (i følge NOKUTs skriv om *Kvalitetssikring av eksamensoppgaver*<sup>4</sup>) en tradisjonell skoleeksamen.

Den europeiske ECTS-standard angir at karakterene fordeles slik at de 10% dårligste skal ha E, deretter 25% D, 30% C, 25% B og 10% A. I Norge skal *ikke* karakterene normalfordeles slik det gjøres i resten av Europa, men i stedet er det et mål at de skal

---

<sup>1</sup> Hentet fra [regjeringen.no](http://regjeringen.no)

<sup>2</sup> Fra NOKUT om [Vurdering og eksamen](#)

<sup>3</sup> Disse beskrivelsene er hentet fra «Nivåbeskrivelser i Nasjonalt kvalifikasjonsrammeverk for livslang læring (NKR), Nivå 6.2

<sup>4</sup> [https://www.nokut.no/globalassets/nokut/rapporter/nasjonaleksamen7/sykepleier/kvalitetssikring-av-eksamensoppgaver\\_eks-fra-sykepleierutd.pdf](https://www.nokut.no/globalassets/nokut/rapporter/nasjonaleksamen7/sykepleier/kvalitetssikring-av-eksamensoppgaver_eks-fra-sykepleierutd.pdf)

angi kvaliteten på studentens prestasjon. "Sensor skal vurdere en prestasjon basert på fastsatte kriterier og ikke det relative resultatet innenfor en gruppe eller et kull" (Mld.st. 16 (2016-2017)). NOKUTs egne undersøkelser reiser tvil om vurderingen i realiteten er kriteriebasert, blant annet i *Er karakterer rettferdige? Et notat om klagesensur September 2017*<sup>5</sup>

Det er opp til den enkelte sensor å vurdere om kompetansekravene er oppfylt som f.eks. en «akseptabel prestasjon med noen vesentlige mangler» noe som beskriver karakteren D, eller om det er en «Jevnt god prestasjon som er tilfredsstillende på de fleste områder», og dermed gir karakteren C.

Det finnes ikke klare retningslinjer for hvordan en prestasjon i en vurderingssituasjon skal vurderes kvalitativt. Karakterene må dermed settes basert på skjønn, med utgangspunkt i læringsutbyttebeskrivelser som ikke angir nivå for det som skal presteres. Et alternativ til bare å gjøre en slik skjønnsmessig vurdering, er å basere den på et poengsystem, slik Figur 1 viser:

Karakterskala for prosentvurderingsmetoden	
A:	89-100 poeng
B:	77-88 poeng
C:	65-76 poeng
D:	53-64 poeng
E:	41-52 poeng
F:	0-40 poeng
Skalaen benyttes først og fremst på teknologiemner, masteroppgaver innenfor matematikk, naturvitenskap, teknologi, lektorutdanning i realfag og 2-årige masterprogram ved Det medisinske fakultetet. Sensor er ikke pålagt å bruke prosentvurderingsmetoden.	
Karaktergrensene er veiledende.	

Figur 1 «Prosentvurderingsmetoden», hentet fra nettsidene til ntnu<sup>6</sup>

Selv om karakteren settes på grunnlag av poengskår, vil vurderingssituasjonen avgjøre om vurderingen i realiteten er kriteriebasert. Har man en slutteksamen som består av utelukkende dikotomiske (rett/feil) flervalgsoppgaver, vil vurderingen være enkel rent matematisk, og den vil framstå som kvantitativ. En slik eksamen vil være i dårlig samsvar med læringsutbyttebeskrivelsene, siden disse skal inneholde elementer som ikke kan vurderes dikotomisk. En vurderingssituasjon som er mer åpen kan også vurderes med poengskala, men da vil poengberegningen i realiteten oftest være skjønnsmessig. Dermed framstår vurderingen objektiv og kvantitativ, mens den i realiteten er skjønnsbasert.

Siden det ikke er utarbeidet et etterprøvbart krav til nivå for universitetsemner, er grunnlaget for å gradere kompetanse mangelfullt. Karakteren blir dermed et resultat av en enkeltpersons<sup>7</sup> inntrykk fra en vurderingssituasjon, basert på den kunnskapen og de tradisjonene sensor har med seg. Ofte er læringsutbyttebeskrivelsene generelle, og

<sup>5</sup> [https://www.nokut.no/globalassets/nokut/notater/2017/er\\_karakterer\\_rettferdige\\_2017.pdf](https://www.nokut.no/globalassets/nokut/notater/2017/er_karakterer_rettferdige_2017.pdf)

<sup>6</sup> <https://i.ntnu.no/wiki/-/wiki/norsk/prosentvurderingsmetoden>

<sup>7</sup> I en del tilfeller skal det være to sensorer, se UH-loven §§11-6 og 11-7.

eksamen begrenset til konkrete problemstillinger som er enkle for en sensor å vurdere på en rett-feil-skala. I så fall sier karakteren lite om kompetansen slik den er beskrevet i NKR. Eksemplet vist i Figur 3 under viser utdrag fra NOKUTs retningslinjer for vurdering i UH-sektoren, der de bruker *Nasjonal deleksamen i anatomi, fysiologi og biokjemi* for sykepleiere (utarbeidet av NOKUT) som utgangspunkt. Fra NOKUTs analyse av oppgavene i Figur 2, ser vi at de bruker den statistiske sammenhengen mellom andelen rett svar på oppgaven og skår på hele prøvesettet til å konkludere med at oppgave 5.12 er en «svært god eksamensoppgave» og at oppgave 5.18 er «litt mer urettferdig».

De to flervalgsoppgavene nedenfor (oppgave 5.12 og 5.18) er hentet fra nasjonal deleksamen for sykepleier desember 2016. Analysene viser at 64 % svarte riktig på 5.12, og at disse studentene i gjennomsnitt fikk 57 poeng ( $z = 0.36$ ) av totalt 100 poeng på hele eksamenssettet. 36 % svarte galt på 5.12, og disse studentene i gjennomsnitt fikk 35,5 poeng ( $z = -0,64$ ) av totalt 100 poeng på hele eksamenssettet. Avstanden i z-skår mellom de som svarte riktig og de som svarte galt viser at oppgaven totalt sett diskriminerer eller skiller godt mellom studenter med høy og lav dyktighet, og at dette er en svært god eksamensoppgave. Vi ser på tilsvarende måte at oppgave 5.18 skiller svakere mellom studenter med høy og lav dyktighet, og at oppgaven dermed er en urettferdig eksamensoppgave. Oppgave 5.12 har dermed bedre egenskaper som eksamensoppgave enn oppgave 5.18:

**5.12 Hvilket utsagn om nyrene er riktig?**

- A. Filtrasjon av blodet foregår i distale tubuli
- B. Råurin/preurin inneholder erythrocytter og store proteiner
- C. Reabsorpsjon foregår i tubuli og samlerør
- D. Sekresjon er transport av urin fra samlerør til nyrebekken

*Riktig svar: C*

**5.18 Hvilket utsagn om temperaturregulering ved feber er riktig?**

- A. Ved stigende feber er blodstrømmen i huden nedsatt
- B. Ved stigende feber øker blodstrømmen i huden
- C. Den forhåndsinnstilte temperaturen i hypotalamus senkes når feberen stiger
- D. Det er typisk å få skjelvinger når feberen synker

*Riktig svar: A*

Figur 2 Utsnitt fra NOKUTs *Kvalitetssikring av eksamensoppgaver: eksempler fra sykepleierutdanning*<sup>8</sup>

At denne prøven stort sett tester reprodutiv kunnskap, gjør det vanskelig å hevde at oppgavene er uavhengige av hverandre. Dette svekker også påstandene om oppgavenes kvalitet. I tillegg er denne nasjonale prøven ikke bare brukt som deleksamen. Det finnes eksempler på at den gir 100% av karaktergrunnlaget i et emne. Når en eksamen i hovedsak tester reprodutiv kunnskap og i tillegg teller hundre prosent, vet vi at vurderingsformen ikke er dekkende for læringsutbyttebeskrivelsene slik de skal være ifølge Kvalitetsrammeverket. En tilsvarende undersøkelse av andre emner viser at eksemplet over ikke er et unntak, men derimot et vanlig fenomen.

### 1.2 Hvor stor informasjonsverdi har karakteren ved ansettelse?

Universitets og høyskoleloven angir ikke noe formål med vurderingen i UH-sektoren. Ved et søk ble det heller ikke funnet noen beskrivelse av formål hos de ulike UH-institusjonene, mens NOKUT beskriver sluttvurderingen som «en kontrollmekanisme på oppnådd læring». Det er interessant å merke seg at NOKUT relaterer sluttvurderingen til læring og ikke til «fastsatte kriterier», slik Mld.st. 16 (2016-2017)) angir.

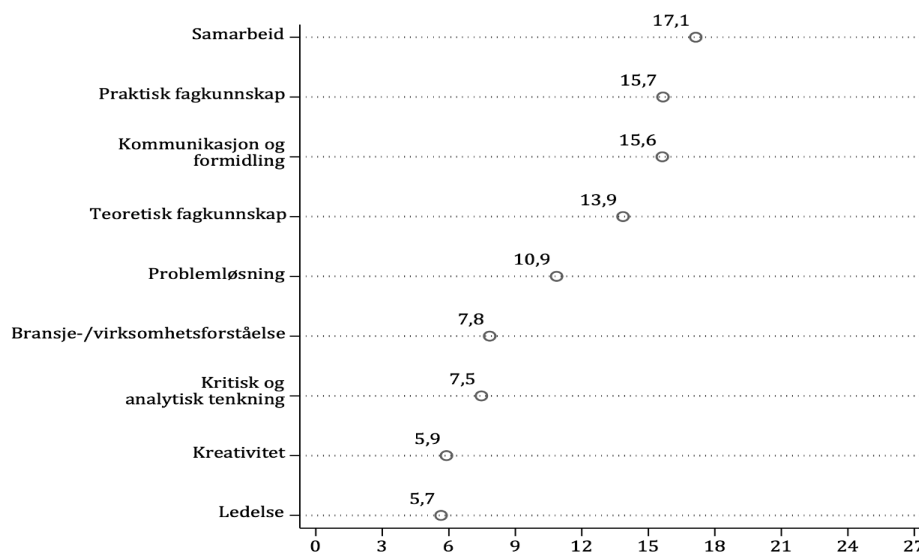
Den årlige investeringen i flere hundremillionersklassen som vanligvis går med til å rangere studenter med karakter må ha en antatt bruksverdi. Et mulig formål med

<sup>8</sup> [https://www.nokut.no/globalassets/nokut/rapporter/nasjonal-deleksamen7/sykepleier/kvalitetssikring-av-eksamensoppgaver\\_eks-fra-sykepleierutd.pdf](https://www.nokut.no/globalassets/nokut/rapporter/nasjonal-deleksamen7/sykepleier/kvalitetssikring-av-eksamensoppgaver_eks-fra-sykepleierutd.pdf)

seksdelt karakterskala kan være at universiteter kan ha nytte av karakterstatistikk til f.eks. markedsføring, en annen at studenten kan ha interesse av å vite hvordan studieprestasjonen deres er vurdert. Her drøftes i stedet et tredje punkt - karakterene som redskap for rangering ved ansettelse:

Vitnemålet kan fortelle om en søker til en stilling er kvalifisert i forhold til utlysningen. Til dette trenger man ikke karakterer, men kan greie seg med bestått/ikke bestått. Vurdering fra A-E kan brukes for å rangere søkerne, for eksempel med et karaktergjennomsnitt av hele eller deler av vitnemålet. Ønsker man å gå i dybden, må man sjekke læringsutbyttebeskrivelsene for det aktuelle emnet det aktuelle året de ble gjennomført og i tillegg vurderingsformen, noe som kan være vanskelig tilgjengelig informasjon. Det er derfor rimelig å anta at karakteren og emnetittelen er den eneste informasjonen som er tilgjengelig om et emne i et vitnemål. Vi vet lite om vitnemålets rolle i ansettelsesprosesser, men lovverket gir vitnemålene større verdi i offentlig sektor enn i privat. I offentlig sektor kan rangering klages inn og fagorganisasjonene har en rolle. I privat sektor står virksomhetene friere.

*Arbeidsgiverundersøkelsen 2022: utdanning i et kunnskapsintensivt samfunn*<sup>9</sup> viser at «en stor andel av arbeidsgiverne» er mer opptatt av arbeidserfaring enn utdanningsnivå. Dette varierer likevel mye, og noen yrker har strenge krav til formell utdanning, som f.eks. helse- og sosialfag. Rapporten sier ikke noe eksplisitt om betydningen av karakternivå, men viser likevel at bransjene rangerer evnen til samarbeid på topp av de ni undersøkte egenskapene en søker bør ha, og deretter praktisk fagkunnskap, slik Figur 3 viser. Teoretisk fagkunnskap kommer på fjerdeplass og problemløsning på 5.



**Figur 5.1 «Tenk deg en situasjon der du kan velge hvilke egenskaper en nyutdannet søker med fagskole/bachelor/masterutdanning skal inneha. Vennligst alloker 100 poeng mellom egenskapene nedenfor, der du gir mest poeng til de egenskapene du anser som mest viktig i vurderinger av nyutdannede søkere» (Q5.1). Gjennomsnitt.**

Figur 3 viser en av figurene fra NIFU-rapport 2024:1, *Arbeidsgiverundersøkelsen 2022: utdanning i et kunnskapsintensivt samfunn*, virksomheters vurderinger knyttet til fagskole-, bachelor- og masterutdannede kandidater.

<sup>9</sup> <https://nifu.brage.unit.no/nifu-xmlui/bitstream/handle/11250/3114706/NIFUrapport2024-1.pdf?sequence=1&isAllowed=y> (side 75)

Ut fra rangeringen av egenskaper ved ansettelse, der teoretisk kunnskap vektes sånn midt på treet, er det betimelig å stille spørsmål ved at det brukes såpass store ressurser til å seksgradere teoretisk kunnskap/reproduksjon i sluttvurderingen, mens samarbeidsevne ser ut til å vektlegges i liten grad.

Det finnes lite objektiv informasjon fra rekrutteringsbyråene som bidrar til ansettelse. I sin egen markedsføring hevder de at de vurderer personlige egenskaper og profil i tillegg til ulike kartlegginger. Det er rimelig å anta at intervju, selvrappotering og innhenting av referanseuttalelser er sentrale ved alle ansettelse i denne kategorien. Det er lite tilgjengelig informasjon om vektingen av karakterer i slike prosesser.

### *1.3 Hvordan påvirker karakterene studentenes arbeids- og læringsprosess?*

Det viktigste argumentet for å unngå seksdelt karakterskala er ikke at karakterene er uegnet for å rangere basert på kompetanse, men påvirkningen den har på opplæringen.

Flere studier viser hvordan vurderingen påvirker opplæringen (Andrews et al., 2002; McKinley & Thompson, 2018). Det er også kjent at den påvirker hvordan studenter og undervisere prioriterer, deres arbeidsmetoder og dermed også læringsutbyttet (Boud & Dochy, 2010; Boud & Soler, 2016; Wiliam, 2011). Vurderingsordningen har også betydning for holdningene til det å studere (Dahlgren, 2009; Gorichanaz, 2022; Harlen et al., 2002).

Norsk UH har en hel rekke reguleringer for å sikre kvaliteten til en vurdering: Emneansvarlige må revidere læringsutbyttebeskrivelser og bestemme vurderingsordning til bestemte tidspunkter, studentene har klagerett og rett til begrunnelser, vurderingen må skje i det digitale systemet Inspira og vurderes innen en gitt frist, skoleeksamen kan ikke være en del av en mappevurdering, avhengig av vurderingssystemet skal det være én eller to sensorer og så videre, alt for å sikre kvaliteten i sluttvurderingen.

Men kontrollmekanismene for eksamensordningen har også hull. Det er ingen kontroll av verken innhold eller validitet av det faglige innholdet i en eksamen. Det finnes heller ikke noe formelt formål med vurderingen, noe som ville gjort det mulig å sjekke om formålet faktisk ble overholdt. Det er ikke noe system for å finne ut om det faglige omfanget i et emne korresponderer med antall studiepoeng eller med studentenes startkompetanse. Og det finnes ikke noe system for å måle effekten vurderingsordningen har på studentenes læring og arbeid med studiet.

For vurderingen er ikke bare en sjekk av resultatet av en utdanning, den påvirker også den samme utdanningen. Et sentralt spørsmål burde da være: Siden vurderingsordningen påvirker opplæringen, hvilken vurderingsordning bidrar til best mulig læring, og hva bidrar negativt? Begrepet konsekvensvaliditet (consequential validity) brukes om konsekvensen av vurderingsordningen på det som skal vurderes (Cizek et al., 2010; Cizek et al., 2008).

Alle reguleringene og de store kostnadene forbundet med eksamensavvikling tyder på at karakterene er ansett som viktige. Det er ikke tilsvarende oppmerksomhet rundt konsekvensene av karaktervurderingen på undervisningen.

Til slutt noen erfaringer fra utprøving av karakterfri eksamen ved IFY, NTNU:

- Studentene er ofte uvant med vurderingsformen Bestått/ikke bestått.

- En del av studentene ønsker gradert vurdering, spesielt før de har er faring med alternative vurderingsformer
- Gradert vurdering gir studentene motivasjon til en ekstra innsats fram mot vurderingssituasjonen. Ved Bestått/ikke bestått som vurderingsform bør det innføres et tilsvarende motiverende element. Dette kan for eksempel være at studentene får presentere sine resultater for sensor og hverandre.
- Ved å erstatte slutteksamen med arbeidskrav fordelt gjennom semesteret, oppnås det en jevnere arbeidsinnsats for en del av studentene.
- Vurderingsformen er bestått/ikke bestått forskyver prioriteringene til studenter fra det å imponere sensor i tema de behersker godt til arbeid med tema der de er usikre.

## Referanser

- Andrews, S., Fullilove, J., & Wong, Y. (2002). Targeting washback—a case-study. *System*, 30(2), 207-223.
- Boud, D., & Dochy, F. (2010). Assessment 2020. Seven propositions for assessment reform in higher education.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400-413.
- Cizek, G. J., Bowen, D., & Church, K. (2010). Sources of validity evidence for educational and psychological tests: A follow-up study. *Educational and Psychological Measurement*, 70(5), 732-743.
- Cizek, G. J., Rosenberg, S. L., & Koons, H. H. (2008). Sources of validity evidence for educational and psychological tests. *Educational and Psychological Measurement*, 68(3), 397-412.
- Dahlgren, P. (2009). *Media and political engagement: Citizens, communication and democracy*. Cambridge University Press.
- Gorichanaz, T. (2022). Relating information seeking and use to intellectual humility. *Journal of the Association for Information Science and Technology*, 73(5), 643-654.
- Harlen, W., Crick, R. D., Broadfoot, P., Daugherty, R., Gardner, J., James, M., & Stobart, G. (2002). A systematic review of the impact of summative assessment and tests on students' motivation for learning.
- McKinley, J., & Thompson, G. (2018). Washback effect in teaching English as an international language. *The TESOL encyclopedia of English language teaching*, 1-12.
- Wiliam, D. (2011). What is assessment for learning? *Studies in educational evaluation*, 37(1), 3-14.