

# Karakterer – fins de fortsatt i 2030?

Guttorm Sindre

Norges teknisk-naturvitenskapelige universitet

**Sammendrag.** Karakterer er ofte betraktet som en nødvendighet i høyere utdanning, men er også gjenstand for kontroverser, og er en stressfaktor både for ansatte og studenter. Det er derfor interessant å spørre om fremtidens digitaliserte høyere utdanning fortsatt vil gi karakterer til studentene? Artikkelen skisserer to motsatte scenarier: (A) Sterk økning i bruk av karakterer. (B) Karakterer blir irrelevante og forsvinner. Begge scenarier kan muliggjøres av IT, ved at det blir billigere å lagre og analysere store informasjonsmengder, samt billigere å teste og vurdere kandidater og prestasjoner automatisk.

**Nøkkelord:** karakterer, vurdering, læringsutbytte, beslutningsstøtte

## 1 Innledning

Karakterer er det dominerende uttrykket ved summativ vurdering av studenters prestasjoner, både i Norge og internasjonalt. Ifølge UHR brukes bokstavkarakterer A-F for ca. 85% av de vurderingene som blir gjort i Norge, mens Bestått / Ikke Bestått (heretter: b/ib) brukes for ca. 15% [1]. Noen universiteter i utlandet har eksperimentert med mer utbredt bruk av b/ib (eller engelsk: pass/fail). McMorrان et al. [2] lister opp 9 universiteter som de er kjent med at benytter dette, men selv for disse har bruken av b/ib begrensninger, som at det bare gjelder enkelte studier (f.eks. medisin på Stanford), at det bare gjelder første årskurs (f.eks. MIT, to universiteter i England og ett i Singapore), eller at det fins flere graderinger av bestått (f.eks. Lund og Uppsala) slik at det egentlig likner mer på bokstavkarakterer.

Argumenter i favør av karakterer (dvs. A-F heller enn b/ib) er at det kan motivere studentene til innsats [1]. Noen studier fant dårligere prestasjoner ved overgang til b/ib [3, 4], dog må det nevnes at andre studier ikke har funnet signifikante forskjeller [5]. Det kan også sies at karakterer gir mer nyansert informasjon både til studentene selv [1] og til andre som trenger å vurdere en kandidats kompetanse, f.eks. ved opptak til studier eller søknad på jobb [1, 6]. Samtidig er det også problematiske sider ved karakterer:

- Karakterer skaper stress, overgang til b/ib har flere steder redusert dette, f.eks. [7, 8].

- Karakterer er en ytre motivasjon som kan motvirke den indre motivasjonen, gleden ved å lære i seg selv [9]. Studentene blir ensidig opptatt av det som er gjenstand for karaktersetting, heller enn nysgjerrighetsbasert utforskning av faget [10].
- En karakter er ikke spesielt nyansert informasjon etter et tresifret antall timeverk for å lære et emne. For å forbedre sin læring trenger studenter *formativ vurdering* [11] underveis i semesteret, og tidlige prestasjoner må kunne erstattes med senere, ikke nødvendigvis være tellende på karakteren [12].
- Faglærere som vil forbedre læringen i et emne, trenger å se *bak* karakterene på ulike delprestasjoner eller rubrikker – snitting og justering gjør at karakterer snarere kan skjule læringsproblemer enn fremheve dem [13].
- Med tanke på opptak til videre studier er det påvist systematiske forskjeller i karakterpraksis både for grunnskoler [14], videregående [15] og høyere utdanning [16], som kan gi urettferdige opptak til videre studier. «Karaktersettingspraksisen synes å variere så mye mellom institusjonene at karakternivået gir begrenset informasjon om studentenes ferdigheter.»[16](s.8)
- *Arbeidsgivere* er ofte mindre opptatt av karakterer som enkelte tror. En norsk intervju-studie blant firmaer som ansetter programmerere, fant for eksempel mye større vekt på engasjement, samarbeidsevner og personlighet enn på karakterer, selv ved ansettelse av nyutdannede [17].

Det ovenstående viser at karakterer knapt fungerer optimalt for noen av behovene de er ment å dekke. Da er det grunn til å etterstrebe bedre løsninger, kanskje relatert til digitalisering [18]?

Denne artikkelen trekker opp to mulige scenarier:

- Scenario A: Økt bruk av karakterer. Vitnemålene kan bli mye mer detaljerte enn de er nå, med mer informasjon om hva kandidatene har lært og mer fingranulære karakterer – for å gi mer presis beslutningsstøtte.
- Scenario B: Karakterer forsvinner fordi alle aktører som ellers kunne ha trengt dem, vil finne at andre typer informasjon er mer egnet for beslutningene som skal tas.

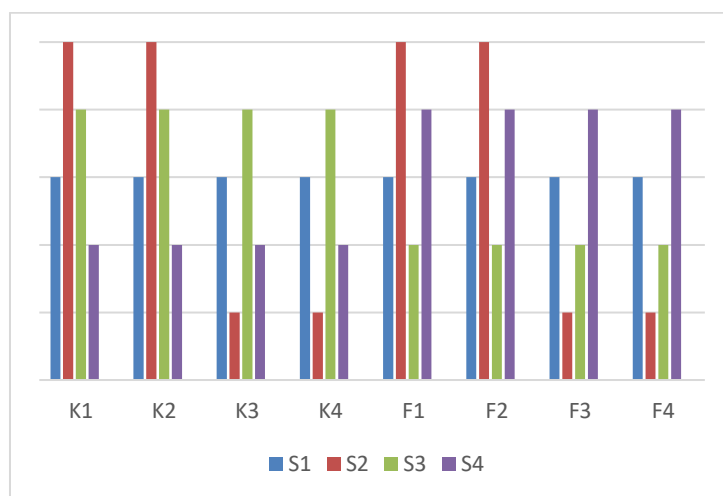
Resten av artikkelen er strukturert som følger: Seksjon 2 presenterer scenario A, mens seksjon 3 presenterer scenario B. Seksjon 3 gir en avsluttende diskusjon og konklusjon.

## 2 Scenario A: Mer karakterer

Mange av svakhetene ved dagens karaktersystem på lavt informasjonsinnhold. Etter 3-5 års studium sitter en kandidat med et vitnemål inkludert karakterutskrift som dokumentasjon. Karakterutskriften er en enkel liste over emner, med emnekode, navn, antall studiepoeng, oppnådd karakter, og hvilket semester emnet ble tatt. Informasjonen er mulig å få inn på et A4-ark, typisk med tanke på å skulle leses manuelt. «*Less is more*»? Bare hvis det «less» som man velger, treffer blink i forhold til informasjonsbehovet. Tre sentrale problemer med dagens karakterutskrift i så måte:

- Karakterer sier lite om studenters styrker og svakheter.
- Emnenavn sier lite om innhold og læringsutbytter i et emne.
- Relevante kompetanser går ofte både på kryss og tvers av emner.

Som vist i en case-studie av Gynnild [13] kan studenter prestere høyst ulikt på forskjellige pensumdelar eller læringsutbytter, men dette blir usynlig i karakterene etter snitting og justering. Figur 1 viser en hypotetisk situasjon hvor fire studenter S1- S4 har tatt samme emne, som er evaluert ut fra læringsutbytter K1-K4 om kunnskap og F1-F4 om ferdigheter. S1 (blå) har jevne C-prestasjoner. S2 (rød) har A på halve pensum (del 1 og 2) men E på den andre halvdel. S3 (grønn) er meget god på kunnskaper, svak på ferdigheter, mens S4 (lilla) motsatt. Med snitting kan alle disse få C, til tross for svært ulike læringsresultat fra det aktuelle emnet. Hvis det skulle være en gjennomgående tendens i mange emner (f.eks. S3 alltid bedre på kunnskap, S4 alltid bedre på ferdigheter), ville nok kandidatene passer til ulike typer jobber, men det blir usynlig i en serie karakterer på vitnemålet som ofte graviterer mot C.



Figur 1: Fiktivt eksempel, 4 studenters prestasjoner for 8 læringsutbytter i et emne.

Høyest søyle = A, lavest søyle = E.

Faglærere kan ha en tilsvarende informasjonsmangel relatert til karakterstatistikken for et helt emne. Den graviterer typisk mot C, som vil være forventet, men hvis klassen gjennomgående gjør det dårligere på visse læringsutbytter enn på andre, vil dette være skjult i karakterstatistikken. Tilsvarende vil det være skjult for studieprogramansvarlige hvis studentene på programmet tenderer til å være svakere på visse typer læringsmål, f.eks. svakere på ferdigheter enn på kunnskaper. Slik skjuler dagens karaktersystem mulige hint om forbedringsbehov.

Emnenavn er ikke alltid dekkende. Ulike læresteder kan ha emner med samme navn men ulikt innhold, eller ulikt navn men likt innhold. Det er tungvint å sjekke emnesider på nettet for å se hva en student har vært gjennom, dessuten misvisende hvis emnet har endret seg. Emnebeskrivelsene bør derfor være inkludert i vitnemålet – urealistisk for vitnemål på papir, men trivielt for digitale vitnemål. Hvis informasjonen struktureres slik at den kan analyseres automatisk, kan det potensielt være svært nyttig, f.eks. for å sammenligne vitnemål for kandidater på tvers av læresteder, til tross for ulike emnenavn og emneinndelinger.

For å illustrere poenget med at relevante kompetanser går på kryss og tvers av emner, anta at et firma skal ansette en programmerer. Firmaet utvikler sikkerhetskritisk programvare, og ønsker kompetanse innen programmering og informasjonssikkerhet, samt gode evner til kommunikasjon og samarbeid. La oss si at de får 20 søkere. I vitnemålene til disse søkerne er det typisk bare 2-4 emner med ordet «programmering» i tittelen, og 1-2 med «sikkerhet» i tittelen. Men begge deler kan inngå i mange andre emner. F.eks. kan «Databaser» inneholde en bolk om databasesikkerhet og noe om programmering av spørringer i *embedded SQL*. Eller kanskje ikke? Tilsvarende med mange andre emner. Hvis vitnemålet ikke bare lister opp emner, men læringsutbytter for hvert emne, med vekt og poengscore – ville man fort ha et vitnemål med hundrevis av poengscorer eller karakterer. Igjen ville dette primært være ment for automatisk analyse, ikke manuell lesning. Man ville da kunne aggregere dataene på nye måter, avhengig av behovet. Bedriften som trengte programmering og sikkerhet kunne be om oppsummering av aktuelle søkere etter disse to kriteriene: Hvor mye kan søkerne om dette? – plutselig mulig som en sammenligning på tvers av vitnemål fra ulike studiesteder, og på tvers av emner som ikke hadde programmering eller sikkerhet i navnet. En annen bedrift på jakt etter brukergrensesnittdesignere kunne aggregere vitnemål etter helt andre kriterier, og de som vil se det som før, med karakterer per emne, kan få dette.

Å sette mange karakterer eller poengscorer per kandidat per emne (typisk for 5-10 opplistede læringsutbytter) høres neppe særlig fristende ut. Men ofte er dette data fagstaben allerede har, f.eks. poeng på deloppgaver på eksamen eller delinnleveringer i løpet av semesteret som er

ment å dekke ulike læringsutbytter, eller poeng på ulike vurderingsrubrikker hvis karakteren settes mer holistisk bare basert på en enkelt prestasjon. Dessuten vil grensen for hva som kan testes og karaktersettes automatisk stadig flyttes [19, 20], så det er ikke sikkert at dette trenger gi mye merarbeid.

### **3 Scenario B: Karakterer forsvinner**

I stedet for å basere en vurdering av kandidaters kompetanse på karakterer fra tidligere utdanning, kan den baseres på tester som utføres direkte i forbindelse med søknadsprosessen. For studier ville dette innebære at opptak i økende grad baseres på opptaksprøver heller enn karakterer. For jobbsøknader kan man også se for seg at arbeidsgivere i økende grad vil bruke tester, i en tidlig fase over nettet for siling av en stor mengde søkere, og deretter grundigere in-situ tester for de mest aktuelle kandidatene. Tester brukes allerede i økende grad ved jobbintervjuer og kan gi større sikkerhet om kandidaters egnethet enn tradisjonelle intervjuer [21]. Både for studier og jobb har inngangstesting en fordel vs. avgangstesting med at det også tar høyde for realkompetanse, dvs. hva søkeren kan ha lært gjennom hobbyvirksomhet, deltidsjobber, verv i ideelle organisasjoner, osv. For studier har opptaksprøver også den fordel at det eliminerer problemene med at snille og strenge skoler gir urettferdighet ved karakterbasert opptak.

En annen mulighet for å komme til gode beslutninger uten karakterer, er å se direkte på kandidatens produkter og prestasjoner fra studietiden. Porteføljer med arbeider [22] kan være en bedre dokumentasjon enn karakterer med tanke på forbedring av læringen underveis, potensielt også ved søknad om jobb og videre studier. Porteføljer er allerede vanlig for jobbsøkere i noen bransjer – men typisk da slike hvor det som lages er av visuell natur (f.eks. arkitektur, grafisk design), som beslutningstakerne kan se på og ha en mening om ganske kjapt. Også i andre fagdisipliner kan produkter gi mer innsikt enn karakterer – vanligvis kan man si mye mer om hvor dyktig en kandidat er til å programmere ved å se på 10000 linjer programkode denne har skrevet i løpet av studietiden enn å se på karakterer, og et bedre inntrykk av en kandidats evner til rapportskrivning og forskning ved å lese hele masteroppgaven heller enn å bare se på dens karakter. Hinderet har vært at det tar for lang tid å se på slike produkter, særlig hvis man skal vurdere mange søkere, mens en karakterutskrift er kjapt å se på. Nå trenger det imidlertid ikke være mennesker som evaluerer produktene manuelt. Det fins allerede mange verktøy for å evaluere programkode i henhold til ulike kvalitetskriterier, og tilsvarende vil finnes for mange andre typer produkter. Det vil antageligvis lenge før 2030 finnes AI-verktøy som kan lese tekster i ulike sjangre (rapporter, vitenskapelige artikler, ...) og si noe om kvaliteten. I så fall kan

menneskelige beslutningstakere nøye seg med å se oppsummeringer av disse analyseresultatene, samt kanskje undersøke noen detaljer av spesiell interesse.

#### 4 Diskusjon og konklusjon

Hvilket scenario er mest sannsynlig, A, B eller for den del C: status quo? For C taler tradisjon, motvilje mot, og kostnader ved, endring, men også at mange synes at det nåværende systemet fungerer greit. Dagens karakterer er likevel ikke noen god spesielt løsning på noen av de behovene som skal dekkes. Karakterer er unyanserte og i verste fall tendensiøse, med urettferdige opptak til videre studier som resultat. Karakterutskrifter har en gammeldags papirepokestruktur som hindrer fleksible analyser og alternative måter å aggregere informasjonen på.

Gitt at karakterer skaper stress kan man frykte at Scenario A ville skape enda mer stress. En modererende faktor kan imidlertid være at hver enkelt karakter har mindre vekt. En gradvis mer fingranulær vurdering kan etter hvert komme til et punkt hvor delene blir så små at det virker mer fornuftig å gi dem b/ib – og forskjellen mellom sterke og svake kandidater kan heller være hvor mange læringsutbytter de klarer i løpet av et semester, heller enn om de får A eller E på dem – selvregulert studie hastighet á la Kellers PSI [23, 24].

Selv med Scenario B er det behov både for studenten og andre å kunne vite hva studenten kan. Hvis det i fremtiden blir slik at man kan se dette direkte fra kandidatens produkter og besvarelser, med støtte fra AI-verktøy, blir faglærers eller sensors meninger om studenten av mindre betydning enn før. Hvis ingen er særlig interessert i karakterene, kan b/ib være billigere og mindre stresskapende. Som McMorran et al. [2] observerer kan imidlertid fjerning av stress ett sted før økt stress et annet sted – her f.eks. i opptaksprøver eller jobbsøkningsprosesser.

Hvis man får mer fingranulære vurderinger per læringsutbytte, og hvis utbyttene kan formuleres ved hjelp av ontologier som muliggjør sammenligning på tvers av læresteder, kunne man se for seg at arbeidsgivere kunne spesifisere jobber ut fra samme ontologier som lærestedene bruker for å dokumentere kompetanse. Sjefen for LinkedIn, Jeff Weiner, sa ifølge [25]: *“We want to have a profile for every member of the global work force [...]. We want to have a profile for every company in the world [...] and digital representation of every job in the world [...] of every skill required to obtain those jobs, a digital presence for every university in the world and we want to make it easy for every individual company and university to share their professionally relevant knowledge.”* Hvis LinkedIn eller andre utvikler slike plattformer, vil studenter kunne bruke disse ikke bare ved jobbsøking nær slutten av studiet, men også under-

veis for å vurdere videre emnevalg. F.eks. kunne man tenke seg at studieløpet kunne visualiseres i et varmekart for sannsynlighet for å få jobb med en gitt kompetanse. Dette varmekartet kunne dynamisk oppdateres med nye fakta – hvilke nye jobber blir utlyst, hva trenger bedriftene de neste årene, hvor mange er underveis i utdanning for å skaffe seg slik kompetanse, hvem får hvilke jobber? Ikke bare studenter vil kunne bruke dette som beslutningsgrunnlag, men også f.eks. studieprogramansvarlige. Dagens vitnemålformat har for begrenset informasjonsgrunnlag til å understøtte denne typen vurderinger.

## Referanser

1. Universitets- og Høgskolerådet, *Retningslinjer for karakterskalaen bestått/ikke bestått*. 2015.
2. McMorran, C., K. Ragupathi, and S. Luo, *Assessment and learning without grades? Motivations and concerns with implementing gradeless learning in higher education*. *Assessment & Evaluation in Higher Education*, 2017. **42**(3): p. 361-377.
3. Gold, R.M., et al., *Academic achievement declines under pass-fail grading*. *The Journal of Experimental Education*, 1971. **39**(3): p. 17-21.
4. Lloyd, D., *Pass-fail grading fails to meet the grade*. *Academic Medicine*, 1992. **67**(9): p. 583-4.
5. Spring, L., et al., *Impact of pass/fail grading on medical students' well-being and academic outcomes*. *Medical education*, 2011. **45**(9): p. 867-877.
6. McKeachie, W.J., *College grades: A rationale and mild defense*. *AAUP Bulletin*, 1976. **62**(3): p. 320-322.
7. Reed, D.A., et al., *Relationship of pass/fail grading and curriculum structure with well-being among preclinical medical students: a multi-institutional study*. *Academic Medicine*, 2011. **86**(11): p. 1367-1373.
8. Bloodgood, R.A., et al., *A change to pass/fail grading in the first two years at one medical school results in improved psychological well-being*. *Academic Medicine*, 2009. **84**(5): p. 655-662.
9. Deci, E.L., R. Koestner, and R.M. Ryan, *Extrinsic rewards and intrinsic motivation in education: Reconsidered once again*. *Review of educational research*, 2001. **71**(1): p. 1-27.
10. Kohn, A., *Grading: The issue is not how but why*. *Educational Leadership*, 1994. **52**(2): p. 38-41.

11. Taras, M., *Assessment—summative and formative—some theoretical reflections*. British journal of educational studies, 2005. **53**(4): p. 466-478.
12. McTighe, J. and K. O'Connor, *Seven practices for effective learning*. Kaleidoscope: Contemporary and Classic Readings in Education, 2009. **174**.
13. Gynnild, V., "*Kriteriebasert vurdering*"-hva innebærer det i praksis. UNIPED, 2013. **36**(1).
14. Galloway, T.A., L.J. Kirkebøen, and M. Rønning, *Systematiske forskjeller i karakterpraksis i grunnskolen?*, in *Utdanning 2011 - veien til arbeidslivet*, N.P. Sandbu, et al., Editors. 2011, Statistisk Sentralbyrå: Oslo/Kongsvinger.
15. Eliassen, S., M.B. Reinart, and M.A. Nome, *Standpunkt karakterer i videregående skole - likebehandles elevene?* 2013, Oslo kommune, Kommunerevisjonen: Oslo.
16. Strøm, B., et al., *Karakterbruk og kvalitet i høyere utdanning*. 2013, Senter for Økonomisk Forskning: Trondheim.
17. Lauvås, P. and K. Raaen, *Passion, Cooperation and JavaScript: This is what the industry is looking for in a recently graduated computer programmer*, in *UDIT 2017*. 2017: Oslo.
18. Kunnskapsdepartementet, *Digitaliseringsstrategi for universitets- og høyskolesektoren 2017-2021*. 2017: Oslo.
19. Reilly, E.D., et al., *Evaluating the validity and applicability of automated essay scoring in two massive open online courses*. The International Review of Research in Open and Distributed Learning, 2014. **15**(5).
20. Burrows, S., I. Gurevych, and B. Stein, *The eras and trends of automatic short answer grading*. International Journal of Artificial Intelligence in Education, 2015. **25**(1): p. 60-117.
21. Moore, D.A., *How to Improve the Accuracy and Reduce the Cost of Personnel Selection*. California Management Review, 2017. **60**(1): p. 8-17.
22. Ellis, C., *The Importance of E-Portfolios for Effective Student-Facing Learning Analytics*, in *E-Portfolios in Higher Education*. 2017, Springer. p. 35-49.
23. Keller, F.S., *Engineering personalized instruction in the classroom*. Revista Interamericana de Psicología/Interamerican Journal of Psychology, 1967. **1**(3).
24. Eyre, H.L., *Keller's Personalized System of Instruction: Was it a Fleeting Fancy or is there a Revival on the Horizon?* The Behavior Analyst Today, 2007. **8**(3): p. 317.
25. Clayton, C., *Jeff Weiner Would Like to Connect With You on LinkedIn*, in *Sky*. 2014, Delta Airlines.