

Trenger vi læringsassistenter? Hvordan kan hverandrevurdering og/eller egenvurdering gi god læringsutbytte.

Omid Mirmotahari and Yngvar Berg

Department of Informatics, University of Oslo, Norway

Sammendrag Denne studien undersøkes effekten av lærerassistentvurdering (LV), egenvurdering (EV) og hverandrevurdering (HV) på læringsutbyttet til 528 studenter i et introduksjonskurs i informatikk. Studentene ble tilfeldig fordelt i tre grupper, hvor hver gruppe mottok en annen type tilbakemelding på en obligatorisk oppgave. Resultatene viste at alle intervensjonene førte til en signifikant forbedring i studentenes prestasjoner fra første utkast til endelig innlevering ($p < 0.05$). HV-gruppen hadde den største gjennomsnittlige økningen i poeng ($M = 0.52$), etterfulgt av LV ($M = 0.45$) og EV ($M = 0.36$). På eksamen oppnådde HV-gruppen den høyeste gjennomsnittlige poengsummen (67.08%), sammenlignet med EV (62.85%) og LV (62.59%). Disse funnene fremhever potensialet til alternative vurderingsmetoder, spesielt hverandrevurdering, som effektive verktøy for å fremme læring og prestasjoner. Studien bidrar til forskningen på vurdering i høyere utdanning og gir innsikt i fordelene og utfordringene ved hver tilnærming. Resultatene har viktige implikasjoner for vurderingspraksis og understreker behovet for å kombinere egenvurdering og hverandrevurdering med tradisjonell lærerassistentvurdering for å skape et mer helhetlig og studentsentrert vurderingsmiljø.

Keywords: Hverandre-vurdering · Egen-vurdering · tilbakemeldingsferdighet · førsteårsstudent · informatikk · Feedback Fruits

1 Innledning

Vurdering spiller en avgjørende rolle i undervisning og læring, både for å fremme studentenes utvikling og måle deres prestasjoner [Sembey et al., 2024, Irons and Elkington, 2021]. Formative vurderingsmetoder, som gir tilbakemelding for å forbedre læring, har fått økt oppmerksomhet i de senere år [Wanner and Palmer, 2018]. Blant disse har egen- og hverandrevurdering dukket opp som verdifulle verktøy for å engasjere studenter som aktive deltakere i læringsprosessen [Misiejuk and Wasson, 2021]. Egenvurdering innebærer at studenter reflekterer over kvaliteten på sitt eget arbeid, mens hverandrevurdering får dem til å gi tilbakemelding på medstudenters arbeid [Wanner and Palmer, 2018]. Disse metodene kan fremme ferdigheter som selvregulering, kritisk tenkning og samarbeid [Rawlusyk, 2018]. Samtidig reiser de spørsmål om pålitelighet og gyldighet sammenlignet med tradisjonell lærerassistentvurdering [Iglesias Pérez et al., 2022].

Tidligere forskning som sammenligner egen-, hverandre- og lærervurdering har gitt blandede resultater. Noen studier finner samsvar mellom vurderingsmetoder [Sadler and

Good, 2006, Sung et al., 2005], mens andre avdekker avvik [Ayyoub, 2017, Bouzidi and Jaillet, 2009]. For å undersøke dette nærmere gjennomfører denne studien en grundig analyse av effekten av egenvurdering, hverandrevurdering og lærerassistentvurdering på læringsutbytte i et stort innføringskurs i datamaskinarkitektur. Over 500 studenter fullførte en obligatorisk oppgave og mottok tilbakemelding gjennom en av de tre intervensjoner, som de brukte til å revidere oppgaven for endelig innlevering. Ved å sammenligne studentprestasjoner på tvers av intervensjonene søker denne studien å gi ny innsikt i de relative fordelene med hver vurderingsmetode innen informatikkutdanning.

Denne forskningen bygger på tidligere studier [Sadler and Good, 2006, Sung et al., 2005, Bouzidi and Jaillet, 2009, Cho et al., 2006, Lin et al., 2001, Mirmotahari et al., 2003, Mirmotahari and Berg, 2023], men utvider dette arbeidet ved å fokusere på informatikkstudie og undersøke bredere effekter på læring. Forskningsspørsmålene i denne studien er som følger:

- *Blir det signifikant forbedring av innleveringsoppgaver ved å få tilbakemelding underveis?*
- *Hvilke tilbakemelding, LV-EV-HV, gir mest signifikant forbedring?*

Artikkelen er strukturert som følger: I avsnitt 2 forankrer vi vårt arbeid i relevant litteratur og etablerer forskningsverdien av vårt studie. I avsnitt 3 på neste side presenterer vi emnet og intervensjonene i detalj. Det er i avsnitt 4 på side 4 hvor metoden og forskningsdesignet beskrives. Resultatene, statistiske metodene og analyser presenteres i avsnitt 5 på side 5. Artikkelen avsluttes med en diskusjon, refleksjon av funnene i avsnitt 6 på side 10, samt forslag til videre arbeid.

2 Teoretisk rammeverk

For å forstå effektene av egen-, hverandre- og lærerassistentvurdering er det viktig å forankre analysen i relevante pedagogiske teorier. Banduras teori om mestringstro antyder at egenvurdering kan bygge studentenes selvtillit og fremme motivasjon [Bandura, 1977]. Vygotskys sosiokulturelle teori fremhever verdien av sosial interaksjon, som hverandrevurdering, for kognitiv utvikling [Vygotsky, 1978]. Kognitive læringsmodeller som Blooms taksonomi tyder på at lærerassistentvurdering kan være godt egnet til å fremme høyere ordens tenkning [Bloom et al., 1956].

Teorier om tilbakemelding understreker viktigheten av at vurdering er rettidig og spesifikk for å være effektiv [Shute, 2008]. Teorier om selvregulert læring antyder at egenvurdering og hverandrevurdering kan fremme studenters metakognitive ferdigheter og autonomi [Zimmerman, 2000, Mirmotahari et al., 2019]. Metakognitiv teori fremhever at EV kan fremme studenters utvikling av selvregulerende læringsferdigheter. Ved å reflektere over sitt eget arbeid, kan studenter følge med på og kontrollere sin egen læring mer effektivt [Nicol and MacFarlane-Dick, 2006]. EV kan imidlertid lide under kognitive bias, for eksempel at lavtpresterende studenter overvurderer sine egne prestasjoner [Kruger and Dunning, 1999].

Ved å trekke på disse teoretiske perspektivene, søker denne studien å belyse fordelene og ulempene ved EV, HV og LV i en stor innføringsinformatikkurs på universitetsnivå. Målet er å belyse de pedagogiske prosessene som ligger til grunn for hver vur-

deringsmetode og dermed informere utformingen av effektive vurderingsinngrep for å støtte studenters læring innen sentrale datavitenskapelige emner.

3 Kasusstudie

Denne studien er gjennomført i et introduksjonskurs i informatikk, IN1020 - Introduksjon til datateknikk, ved Universitetet i Oslo. Som et felles utgangspunkt for studenter som er innmeldt i de seks studieprogrammene tilbudt av Institutt for informatikk (IFI), har kurset som mål å gi en omfattende forståelse av informatikkens bredde. Emnet dekker fagområder som sikkerhet, nettverk, operativsystemer, programmering, maskinkode, dataarkitektur og boolsk algebra. Historisk sett har emnet over 800 påmeldte studenter, hvorav ca 550 studenter tar eksamen. Emnet gir 10 studiepoeng og evalueres med en bestått/ikke bestått avsluttende digital eksamen.

IN1020s undervisningsstruktur består av fire timer forelesninger og to timer gruppeundervisning per uke. For å kvalifisere seg til avsluttende eksamen må studentene ha bestått tre obligatoriske oppgaver innen samme semester. I tillegg får studentene ukentlige oppgaver med tilhørende løsningsforslag, som primært gjennomgås i gruppetimene - som undervises av læringsassistenter. Denne pedagogiske tilnærmingen er utformet for å tilrettelegge for aktiv læring og gi jevnlig tilbakemelding til studentene - noe som er nøkkelelementer i å fremme studentengasjement og suksess [Prince, 2004, Hattie, 2009].

Gitt de ulike studentbakgrunnene i IN1020, er det avgjørende å ta i bruk en pedagogisk tilnærming som er både adaptiv og inkluderende [Bransford et al., 2000]. Tradisjonelle obligatoriske oppgaver og eksamener kan føles begrensende for noen studenter, spesielt i lys av deres varierte læringsstiler og personlige forpliktelser [Kolb and Kolb, 2005]. For å imøtekomme dette mangfoldet og fremme en mer elevsentrert tilnærming, utforsker vi en alternativ pedagogisk strategi gjennom integrasjon av vurdering i den siste obligatoriske oppgaven. I denne studien tar vi sikte på å undersøke effektiviteten av alternative vurderingsmetoder (lærerassistentvurdering (LV), egenvurdering (EV) og hverandrevurdering (HV)) og deres innvirkning på studentenes læringsutbytte og engasjement i konteksten av et introduksjonskurs i informatikk.

Den aktuelle obligatoriske oppgaven har tre hovedoppgaver. Den første oppgaven hadde som mål å vurdere studentenes teoretiske og konseptuelle forståelse ved å la dem gi tilbakemelding på et tidligere gitt eksamensspørsmål. Studentene ble bedt om å evaluere eksamensspørsmålets klarhet, vanskelighetsgrad og samsvar med læringsmålene for kurset. De var også pålagt å lage en utvidet løsning, som forklarte begrunnelsen bak de riktige svarene. Denne tilnærmingen oppmuntrer studentene til å engasjere seg i reflekterende praksis og utvikle sine metakognitive ferdigheter, som er avgjørende for effektiv læring [Schön, 1983, Flavell, 1979].

Den andre oppgaven involverte kretsanalyse fra et eksamensspørsmål med flervalgsvar. Studentene skulle gi en grundig forklaring og dokumentasjon av en krets' funksjonalitet, ved hjelp av verktøy som sannhetstabeller og funksjonsuttrykk som de hadde lært i kurset. Hovedformålet med denne delen var å få studentene til å argumentere for sine valg, noe som fremmer kritisk tenkning og dyp forståelse av fagstoffet [Paul and Elder, 2006].

Den tredje oppgaven fokuserte på kretsforbedring og presenterte studentene for en krets som kunne forbedres på ulike måter. Studentene ble bedt om å gjøre forbedringer på kretsen og begrunne sine valg, noe som viser deres forståelse av boolsk algebra og kretsoptimalisering. Denne delen oppfordrer studentene til å anvende sin kunnskap i en praktisk sammenheng og utvikle problemløsningsferdigheter, som er avgjørende innen informatikk [Jonassen, 2000].

4 Metode og forskningsdesign

Studien benytter et eksperimentelt forskningsdesign, der deltakerne ble tilfeldig fordelt i tre ulike grupper: hverandrevurdering (HV), egenvurdering (EV) og lærerassistentvurdering (LV). Dette designet muliggjorde evaluering av effektiviteten til forskjellige tilbakemeldings- og vurderingsmetoder på studentenes prestasjoner og læringsutbytte.

Deltakerne i studien var studenter som hadde fullført de to foregående obligatoriske oppgavene i kurset. Studentene ble tilfeldig fordelt i de tre nevnte gruppene (HV, EV og LV) for å sikre en uhildet fordeling av deltakere. Studentpopulasjonen var mangfoldig, og den tilfeldige grupperingen sørget for at eventuelle demografiske forskjeller ble jevnt fordelt mellom gruppene.

Den siste obligatoriske oppgaven besto av tre faser - (i) første innlevering, (ii) vurdering og tilbakemelding og (iii) tilslutt endelig innlevering. I den første fasen leverte studentene sitt første utkast. En uke senere måtte de gjennomføre vurderingsprosessen, som varierte avhengig av hvilken gruppe de tilhørte. Fristen for den reviderte innleveringen ble satt til en uke etter vurderingsprosessen. Det er verdt å merke seg at det ikke var noen begrensning på lengden av innleveringene, noe som resulterte i stor variasjon. De tre gruppene fulgte ulike prosedyrer i løpet av tilbakemeldingsfasen:

- *Hverandrevurdering (HV)-gruppen*: Umiddelbart etter fristen for første innlevering fikk studentene i denne gruppen tilgang til tre av medstudentenes innleveringer. De hadde én kalenderuke på seg til å gi tilbakemelding til sine medstudenter. I motsetning til tilnærmingen som ofte finnes i forskningslitteraturen, ble det ikke gitt spesifikk veiledning eller rubrikker for tilbakemeldingsprosessen. Alle studentene mottok imidlertid følgende instruksjon: "I denne obligatoriske oppgaven ønsker vi å gi deg en gyllen mulighet - å utvikle deg selv som en dyktig akademisk kommunikator. God akademisk tilbakemelding er avgjørende i et samarbeidende profesjonelt liv. Å være en effektiv kommunikator som kan gi god akademisk og konstruktiv tilbakemelding er en viktig og ettertraktet ferdighet. I tillegg vil du ha tilgang til og kan benytte tilbakemeldinger i en endelig innlevering." Dette bevisste valget ble gjort for å evaluere den kvalitative verdien av studenttilbakemeldinger og vurdere effekten av å bli eksponert for medstudenters arbeid. Etter å ha mottatt tilbakemeldingene hadde studentene ytterligere en kalenderuke på seg til å innarbeide forslagene og levere sitt endelige, reviderte arbeid.
- *Egenvurdering (EV)-gruppen*: Etter fristen for første innlevering fikk studentene i denne gruppen tilgang til en nettbasert rubrikk for å evaluere og vurdere sitt eget arbeid basert på spesifikke kriterier. Rubrikken besto av en blanding av Likert-skala-spørsmål, binære spørsmål (f.eks. om visse elementer var inkludert i innleveringen)

og åpne spørsmål som krevde skriftlige svar for å oppmuntre til metakognitiv læring. Rubrikken ble utformet ved hjelp av en sokratisk tilnærming, der studentene ikke fikk de riktige svarene, men i stedet ble bedt om å vurdere inkludering eller utelatelse av ulike akademiske begreper og metoder.

- *lærerassistentvurdering (LV)-gruppen*: Studentene i denne gruppen mottok tilbakemelding fra en læringsassistent innen en uke etter den første innleveringen. Denne gruppen var ment å representere en kontrollgruppe, da denne tilnærmingen har blitt brukt tidligere for alle studenter. Det var 18 undervisningsassistenter (LA-er) involvert i å gi tilbakemelding, og de hadde full frihet til å gi tilbakemelding i sin egen stil.

Prosesen var delt inn i tre deler: (i) Første innlevering, der studentene leverte sitt første utkast; (ii) Vurderingsprosess, som varierte mellom gruppene; og (iii) Revisjon, der studentene brukte tilbakemeldingene til å forbedre sin første innlevering og lage en endelig innlevering. Hovedforskjellen mellom gruppene var vurderingsprosessen: HV involverte hverandrevurdering, EV involverte egenvurdering og LV involverte lærerassistenttilbakemelding. Studentene i HV- og EV-gruppene var aktive deltakere i tilbakemeldingsprosessen, mens de i LV-gruppen var passive mottakere.

Feedback Fruits (FF), en nettbasert plattform som integreres med Canvas (LMS), ble brukt til å administrere tilbakemeldings- og vurderingsprosessen. Alle innleveringer og tilbakemeldinger ble gjort direkte gjennom FF i Canvas.

Data ble eksportert fra FF, og kvalitative vurderinger ble utført på alle innleveringer og tilbakemeldinger. For å sikre pålitelighet og gyldighet på tvers av grupper ble det utført en objektiv tredjeparts vurdering for å kvantifisere kvaliteten på hver innlevering. Flere detaljer om denne prosessen er gitt i avsnitt Avsnitt 5.

Alle data har blitt konvertert til dataframes i R-studio, og all statistisk analyse er oppnådd gjennom R-programmeringsspråk og støttede biblioteker.

Studien fikk godkjenning fra SIKT - Kunnskapssektorens tjenesteleverandør, som leverer produkter og tjenester for utdanning og forskning i Norge. Meldeskjemaet for personopplysninger, et nettbasert skjema fra SIKT, ble fylt ut og godkjent, noe som sikret at prosjektet oppfylte kravene til personvern. Tiltak ble iverksatt for å beskytte deltakernes personvern og konfidensialitet gjennom hele studien.

5 Resultater

For å sikre en grundig og pålitelig evaluering av alle innleveringene, både første utkast og endelig versjon, ble de gjennomgått av to uavhengige sensorer. Sensorene benyttet en poengskala fra -1 til 3 for både første utkast og endelig innlevering. For å forenkle diskusjonen av funnene har vi valgt å oversette denne poengskalaen til en skala fra 0 til 100, og deretter til en karakterskala fra A til F. Selv om kurset ikke benytter karakterer, mener vi at dette gjør det enklere å visualisere og diskutere resultatene. Oversettelsen fra poengskalaen [-1, 3] til [0, 100] gjøres ved hjelp av formelen: $X = Y/4 * 100$, der X er poengskalaen [0, 100] og Y er poengskalaen [-1, 3]. Ettersom dette er en obligatorisk innleveringsoppgave, kan karakteroversettelsen til A-F oppleves som noe mild eller til fordel for studentene. Men siden det er den samme terskelverdien som brukes for både

	A	B	C	D	E	F
Terskel	85	66	41	21	6	
Spennvidde	15	19	25	20	15	6

Tabell 1. Oversettelse av poeng til A-F karakter.

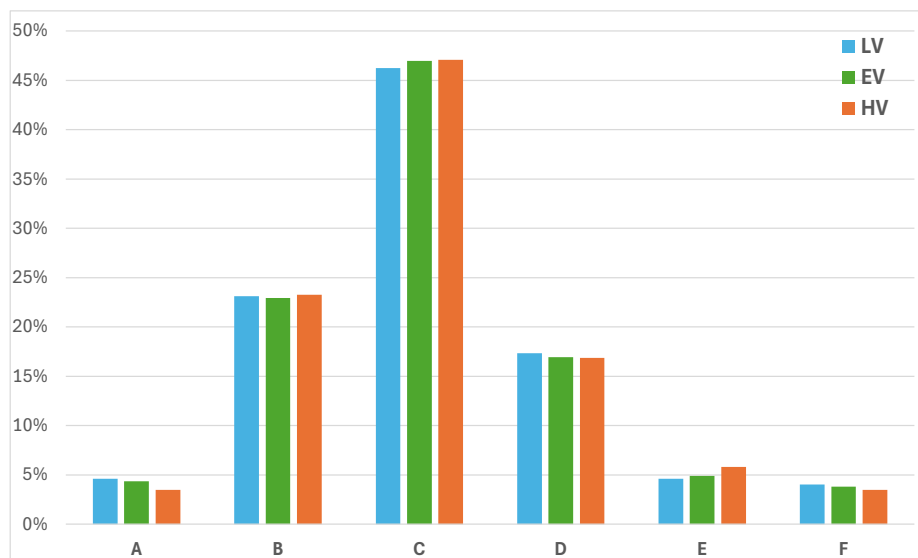
	LV	EV	HV
Gjennomsnitt	▲0.45	▲0.36	▲0.52

Tabell 2. Endringen i gjennomsnittspoeng (skala [-1,3]) mellom første utkast og endelig utkast.

første utkast og endelig innlevering, er det den relative endringen som er avgjørende, ikke nødvendigvis de enkelte punktene/verdiene. Terskelverdier og spennvidde for de valgte karakterene er vist i Tabell 1.

Etter validering av data for innleveringene endte vi opp med totalt 528 innleveringer for både første utkast og endelig versjon. Disse var fordelt på de ulike gruppene som følger: lærerassistentvurdering (LV) = 173, egenvurdering (EV) = 183 og hverandre-vurdering (HV) = 172.

Karakterfordelingen for de ulike gruppene ved første utkast er vist i Figur 1. Her ser vi at fordelingene er nær identiske, noe som er forventet når det er et stort antall deltakere i hver gruppe. Denne fordelingen viser at gruppenes demografiske sammensetning er nøytralisert, og at alle gruppene har samme utgangspunkt før intervensjonen iverksettes. Selv om fordelingen ikke har en normalfordeling, er det den relative endringen mellom første utkast og endelig innlevering som er i fokus.

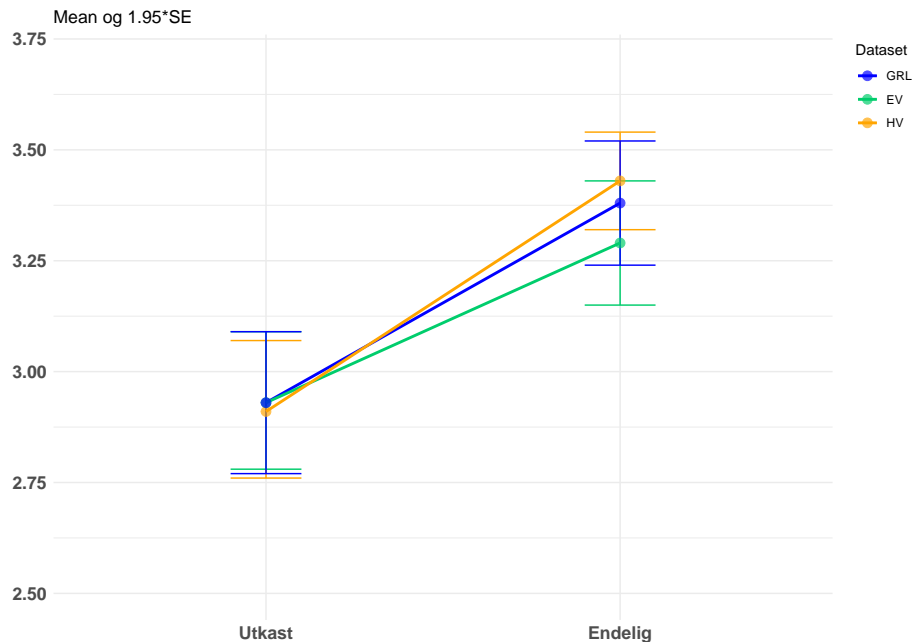


Figur 1. Karakterfordelingen ved første innlevering.

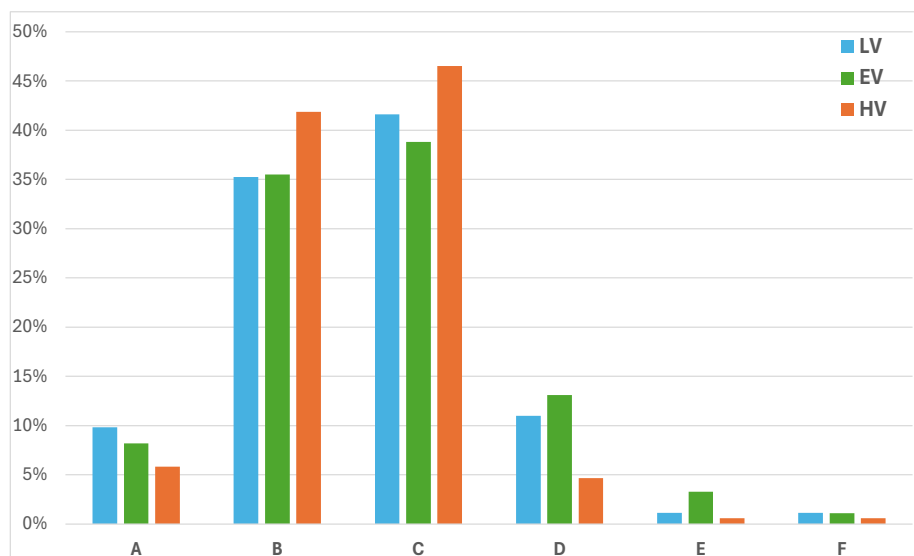
Når vi ser på endringene mellom første utkast og endelig innlevering, observerer vi en positiv økning i poeng for alle gruppene. Tabell 2 på forrige side viser hvordan gjennomsnittspoengene i de ulike gruppene har endret seg på poengskalaen [-1, 3]. For LV-gruppen tilsvarer en økning på 0.45 poeng et hopp på 11.25 poeng på 0-100 skalaen. Effekten av denne endringen avhenger av hvor på karakterskalaen gjennomsnittet ligger, og hvordan endringen slår ut. Dette kan vi se nærmere på i Figur 2, som viser utgangspunktet og endringen på en karakterskala der $C = 3$. I figuren representerer prikkene gjennomsnittskarakteren, og y-aksen viser bokstavkarakteren gjort om til tallverdi ($A = 5, B = 4, C = 3, D = 2, E = 1$). Grensestreke (errorbars) viser $1.95 * SE$, der SE er standardfeilen (standard error). Når bunnen av grensestreken (altså $-1.95 * SE$) for endelig innlevering er høyere enn toppen av grensestreken ($1.95 * SE$) for første utkast, indikerer dette statistisk signifikans for endringen. Med andre ord har intervensjonen for hver av gruppene bidratt til en signifikant forbedring.

Karakterfordelingen for de endelige besvarelsene, vist i Figur 3 på neste side, viser at de fleste studentene har forbedret seg, spesielt de som opprinnelig lå i de lavere karaktersjiktene. Vi ser at LV-gruppen har den største økningen i andelen studenter som oppnår karakteren A, mens HV-gruppen har den laveste økningen for karakteren A.

For å undersøke detaljene nærmere har vi i Figur 4 på side 9 vist hvor de ulike intervensjonene har størst innvirkning. X-aksen representerer poengsummen på første innlevering, mens y-aksen viser effekten av intervensjonene. Grafen er tilnærmet med en polynomisk funksjon for å balansere utliggerne i datasettet.



Figur 2. Mean og $2xSE$. Hvor karakteren $C = 3.0$



Figur 3. Karakterfordelingen ved endelig innlevering.

Frafallet mellom den obligatoriske innleveringsoppgaven og eksamen for de ulike gruppene er som følger: LV = -2.89%, EV = -5.46% og HV = -3.57%. Den høyeste enkeltpoengsummen, uavhengig av intervensjonsgruppe, var 95.5 poeng av 100 mulige. For hver gruppe var den høyeste poengsummen: LV = 95.5, EV = 89.6 og HV = 90.5.

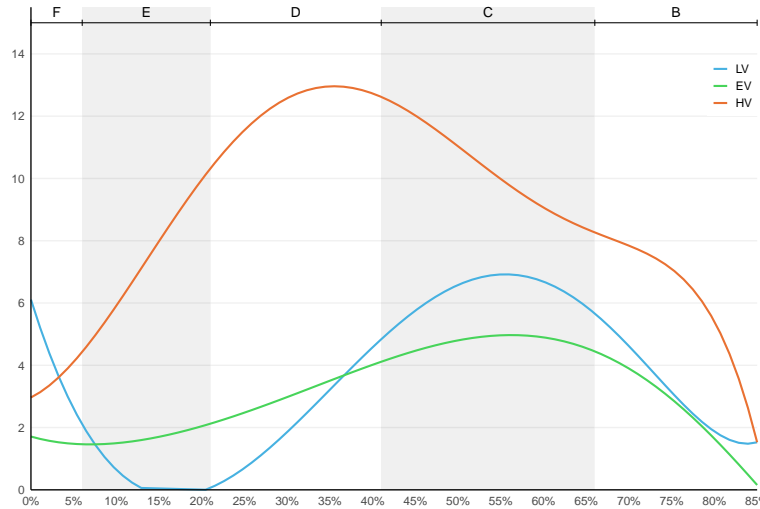
Når det gjelder gjennomsnittlig poengsum for hele eksamen, som har en annen karakter-/poengoversettelse, er resultatene som følger for de ulike gruppene: LV = 62.59%, EV = 62.85% og HV = 67.08%. På eksamen anser vi 63 poeng som terskelen for karakteren C. Dette betyr at gjennomsnittskarakteren for LV og EV tilsvarer en sterk D, mens HV ligger på en svak C.

Hvis vi fokuserer på maskinvaredelen av eksamen, var den gjennomsnittlige poengsummen: LV = 43.84, EV = 46.40 og HV = 49.67. Med en terskel på 48 poeng for karakteren C, tilsvarer LV en svak D, EV en sterk D og HV en svak C.

Det er viktig å merke seg at alle studentene må bestå alle fire delene av eksamen for å få bestått. Hver del vurderes individuelt, og ideelt sett burde de hatt de samme karakter-/poengterskler, men på grunn av uklarheter i noen oppgaver er karakter-/poengfordelingen ulik mellom delene. Dette er årsaken til at vi her må gjøre analyser basert på tre ulike karakter-/poengoversettelser.

5.1 Statistisk analyse

Deskriptiv statistikk for variablene EV, LV og HV viser at skjevheten er signifikant forskjellig fra 0 for alle tre variablene, noe som indikerer at fordelingene ikke er symmetriske, vist i Tabell 3 på neste side. For variabelen LV er både skjevhet og kurtose signifikant forskjellig fra 0, noe som tyder på at fordelingen ikke er symmetrisk og har



Figur 4. Antall studenter som får en positiv effekt av de ulike intervensjonene, samt hvilke karakterområde de tilhører opprinnelig.

tyngre haler/skarpere topper enn en normalfordeling. For å vurdere normaliteten til dataene brukes Shapiro-Wilk-testen, ettersom den er kjent for å være en kraftig test for normalitet, spesielt for små til mellomstore utvalg. Shapiro-Wilk-testen viser at dataene ikke er normalfordelte, med høy antall frihetsgrader (df) og en p-verdi < 0,05 for alle variablene, som vist i Tabell 4.

Ettersom dataene ikke er normalfordelte, brukes den ikke-parametriske Kruskal-Wallis-testen for å sammenligne gruppene. Kruskal-Wallis-testen er et godt alternativ til den parametriske ANOVA-testen når antakelsen om normalitet ikke er oppfylt. Kruskal-

Gruppe	Gj.snitt	SD	Skjevhet	Kurtose
EV	1,14	0,74	-0,65	0,14
LV	1,27	0,79	-1,01	1,22
HV	1,22	0,85	-0,75	0,02

Tabell 3. Deskriptive statistikk.

	Shapiro-Wilk	df	p-verdi
EV	0,96	183	0,000
LV	0,92	173	0,000
HV	0,92	172	0,000

Tabell 4. Resultatene for hver av intervensjonene gjennom Shapiro-Wilk testen.

Parvis	Z	P.unadj	P.adj
HV - LV	0,61051	0,5415218	1,00000
HV - EV	3,47769	0,0005058	0,00152
LV - EV	2,86295	0,0041971	0,01259

Tabell 5. Resultatene fra Dunns test for de parvise sammenligningene.

Wallis-testen viser at det er signifikante forskjeller mellom gruppene (Kruskal-Wallis chi-squared = 13,925, df = 2, p-verdi = 0,0009466). Ettersom p-verdien er < 0,05, utføres Dunns test for parvise sammenligninger. P-verdiene fra Dunns test er justert ved hjelp av Bonferroni-metoden.

Resultatene fra Dunns test, vist i Tabell 5 viser at det er signifikante forskjeller mellom HV og EV (p-verdi = 0,001517277) og mellom LV og EV (p-verdi = 0,012591395), mens det ikke er signifikant forskjell mellom HV og LV (p-verdi = 1,0). Dette indikerer at forbedringen er signifikant forskjellig mellom HV og EV, samt mellom LV og EV, men ikke mellom HV og LV.

6 Diskusjon

Denne studien har undersøkt effekten av tre ulike vurderingsintervensjoner - lærerassistentvurdering (LV), egenvurdering (EV) og hverandrevurdering (HV) - på læringsutbyttet til studenter i et stort introduksjonskurs i informatikk. Resultatene viser at alle tre intervensjonene førte til en signifikant forbedring i studentenes prestasjoner fra første utkast til endelig innlevering. HV-gruppen hadde den største gjennomsnittlige økningen i poeng, etterfulgt av LV- og EV-gruppene. Disse funnene antyder at alternative vurderingsmetoder, spesielt hverandrevurdering, kan være effektive verktøy for å fremme studenters læring og prestasjoner.

Forbedringen i studentenes prestasjoner etter intervensjonene kan tilskrives flere faktorer. For det første gir både EV og HV studentene mulighet til å reflektere over sitt eget arbeid og få innsikt i medstudentenes perspektiver [Nicol and MacFarlane-Dick, 2006]. Denne refleksjonsprosessen kan hjelpe studentene med å identifisere styrker og svakheter i sitt eget arbeid og dermed foreta informerte revisjoner [Boud and Molloy, 2013]. For det andre kan tilbakemeldinger fra medstudenter og lærerassistenter gi verdifull veiledning og støtte studentene i å forbedre sitt arbeid [Hattie, 2009]. Tilbakemeldinger kan hjelpe studentene med å klargjøre misforståelser, utfordre deres tenkning og foreslå nye tilnærminger [Shute, 2008].

Det er interessant å merke seg at HV-gruppen hadde den største gjennomsnittlige forbedringen i poeng, noe som antyder at hverandrevurdering kan være spesielt gunstig for studenters læring. Dette funnet er i tråd med tidligere forskning som har vist at hverandrevurdering kan fremme dypere læring, kritisk tenkning og metakognitive ferdigheter [Topping, 1998, van Zundert et al., 2010]. Ved å vurdere medstudentenes arbeid må studentene anvende sin forståelse av fagstoffet og gi konstruktiv tilbakemelding. Denne prosessen kan hjelpe studentene med å utvikle en dypere forståelse av læringsmålene og forbedre sine egne prestasjoner [Nicol et al., 2014].

Samtidig er det viktig å erkjenne potensielle utfordringer og begrensninger ved hverandrevurdering. Studenter kan være motvillige til å gi ærlig tilbakemelding til medstudenter på grunn av sosiale forhold eller bekymringer for gjengjeldelse og/eller mistillit til medstudenters kompetanse. Studentenes begrenset erfaring og fagkunnskap kan også påvirke kvaliteten og påliteligheten av tilbakemeldingene de gir [Kaufman and Schunn, 2010]. Forskningslitteraturen peker på disse utfordringene og fremmer at undervisere bør gi opplæring og veiledning i hvordan man gir konstruktiv tilbakemelding, etablere klare vurderingskriterier og skape et støttende og inkluderende læringsmiljø [Gielen et al., 2010]. I denne studien har vi bevisst valg å ikke ha noe veiledning for å se hvordan studentenes responderer til hverandre og om de i felleskap kan løfte hverandre. I tillegg så vil denne studien også kaste nytt lys på eksponeringseffekten - altså det å eksponere studentene for ulike løsninger og hva det gjør med dem i sin egen meta-kognitive læringsyklus.

Selv om EV-gruppen viste den minste gjennomsnittlige forbedringen i poeng, bør verdien av egenvurdering ikke undervurderes. Egenvurdering kan fremme studentenes autonomi, selvregulering og metakognitive bevissthet [Andrade, 2018]. Ved å reflektere over sitt eget arbeid og vurdere det i forhold til etablerte kriterier, kan studentene utvikle en dypere forståelse av sine egne læringsprosesser og bli mer selvstendige i sin læring [Boud, 2000]. Egenvurdering kan også hjelpe studentene med å sette mål, vurdere sin egen fremgang og tilpasse sine læringsstrategier [Panadero et al., 2017].

LV-gruppen, som fungerte som kontrollgruppe i denne studien, viste også en signifikant forbedring i prestasjoner. Dette understreker den viktige rollen undervisere og lærerassistenter spiller i å gi tilbakemelding og støtte studentenes læring [Hattie and Timperley, 2007]. Lærere, som fageeksperter, kan gi innsiktsfull og faglig tilbakemelding som hjelper studentene med å forbedre sine prestasjoner og nå læringsmålene [Shute, 2008]. Samtidig bør lærere være oppmerksomme på begrensningene til lærervurdering, for eksempel tidspres, høyt antall studenter og potensiell ensidighet i tilbakemeldingen [Carless and Winstone, 2023]. Vi kan stille spørsmål om hvor signifikant læringsutbytte er satt opp mot det økonomiske bilde med lønnskostnader. Vi ser tydelig fra Figur 4 på side 9 at LV gir best utslag for studenter som hovedsakelig ligger i mellom bestått/ikke bestått området og at det løfter heller gode studenter enn de svake.

Når det gjelder eksamensresultatene, viste HV-gruppen den høyeste gjennomsnittlige poengsummen og prosentandelen, etterfulgt av EV- og LV-gruppene. Disse funnene antyder at de ferdighetene og kunnskapene som studentene tilegnet seg gjennom hverandrevurdering, kan ha overføringsverdi til andre vurderingssituasjoner. Ved å delta aktivt i vurderingsprosessen og utvikle sin vurderingsevne, kan studentene bli bedre rustet til å møte de akademiske kravene i eksamen og andre vurderinger [Sadler, 2010].

Studien har viktige implikasjoner for vurderingspraksis i høyere utdanning. Funnene antyder at å inkorporere egenvurdering og hverandrevurdering i vurderingsprosessen kan fremme studenters læring og prestasjoner. Undervisere bør vurdere å bruke disse vurderingsmetodene sammen med tradisjonell faglærervurdering for å skape et mer helhetlig og studentsentrert vurderingsmiljø. Dette innebærer å gi studenter opplæring og støtte i å gi og motta tilbakemelding, utvikle klare vurderingskriterier og skape muligheter for dialog og refleksjon.

Videre understreker studiens funn viktigheten av å utvikle studentenes vurderings- evne og evne til selv- og hverandrevurdering. Ved å utstyre studentene med disse fer- dighetene kan vi fostre selvstendige og selvregulerte studenter som er i stand til å følge utviklingen av og styre sin egen læring. Dette er avgjørende ferdigheter ikke bare for akademisk suksess, men også for livslang læring og faglig utvikling.

Avslutningsvis gir denne studien verdifull bidrag i forskningslitteraturen i effekten av ulike vurderingsintervensjoner på studenters læringsutbytte i et introduksjonskurs i informatikk. Resultatene fremhever potensialet til egenvurdering og hverandrevurde- ring som effektive verktøy for å fremme studenters læring og prestasjoner, spesielt når de brukes i kombinasjon med faglærervurdering. Ved å implementere disse vurderings- metodene på en gjennomtenkt og systematisk måte kan vi skape et mer inkluderende og styrkende vurderingsmiljø som fosterer studenters akademiske og personlige vekst.

7 Videre arbeid

Studien bidrar til den voksende mengden forskning som utforsker effekten av ulike vurderingsmetoder på studenters læringsutbytte. Implikasjonene av denne studien for vurderingspraksis i høyere utdanning er betydelige. Undervisere og utdanningsinstitu- sjoner bør vurdere å implementere alternative vurderingsmetoder som komplementerer tradisjonell lærervurdering. Til tross for studiens lovende funn er det viktig å anerkjen- ne dens begrensninger. Fremtidig forskning bør ta sikte på å replikere disse funnene på tvers av ulike kontekster, disipliner og studentpopulasjoner. I tillegg bør fremtidige stu- dier utforske langtidseffektene av vurderingsintervensjoner og undersøke hvordan disse ferdighetene overføres til andre akademiske og profesjonelle settinger - helt i tråd med tilbakemeldingsferdighet.

Studiet er gjennomført med signifikat støtte fra Helge E. Kleivane og Andreas Tand- berg fra ”IT i utdanning” seksjonen på fakultetet og økonomiske støtte fra SK-ITU sen- tralt på UiO. Studiet har også godkjenning fra SIKT.

Referanser

- Andrade, 2018. Andrade, H. L. (2018). Feedback in the context of self-assessment. *Part III - Context and Sources of Feedback, The Cambridge Handbook of Instructional Feedback*.
- Ayyoub, 2017. Ayyoub, A. (2017). An action research approach for using self/peer assessment to enhance learning and teaching outcomes. *Journal of Teaching and Teacher Education*, 5:33–42.
- Bandura, 1977. Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral chan- ge. *Psychological Review*, 84(2):191–215.
- Bloom et al., 1956. Bloom, B., Engelhart, M., Furst, E., Hill, W., and Krathwohl, D. (1956). *Taxonomy of educational objectives: The classification of educational goals*. Longmans, Green.
- Boud, 2000. Boud, D. (2000). Sustainable Assessment: Rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2):151–167.
- Boud and Molloy, 2013. Boud, D. and Molloy, E. (2013). *Feedback in Higher and Professional Education: Understanding it and doing it well*. Routledge.
- Bouzidi and Jaillet, 2009. Bouzidi, L. and Jaillet, A. (2009). Can online peer assessment be trusted? *Journal of Educational Technology & Society*, 12(4):257–268.

- Bransford et al., 2000. Bransford, J. D., Brown, A. L., and Cocking, R. R. (2000). *How People Learn; Brain, mind, experience, and school*. National Academy Press.
- Carless and Winstone, 2023. Carless, D. and Winstone, N. (2023). Teacher feedback literacy and its interplay with student feedback literacy. *Teaching in Higher Education*, 28(1):150–163.
- Cho et al., 2006. Cho, K., Schunn, C. D., and Wilson, R. W. (2006). Validity and reliability of scaffolded peer assessment of writing from instructor and student perspectives. *Journal of Educational Psychology*, 98(4):891.
- Flavell, 1979. Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry. *American Psychologist*, 34(10):906–911.
- Gielen et al., 2010. Gielen, S., Peeters, E., Dochy, F., Onghena, P., and Struyven, K. (2010). Improving the effectiveness of peer feedback for learning. *Learning and Instruction*, 20(4):304–315. Unravelling Peer Assessment.
- Hattie, 2009. Hattie, J. (2009). *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement*. Routledge, London.
- Hattie and Timperley, 2007. Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.
- Iglesias Pérez et al., 2022. Iglesias Pérez, M. C., Vidal-Puga, J., and Pino Juste, M. R. (2022). The role of self and peer assessment in higher education. *Studies in Higher Education*, 47(3):683–692.
- Irons and Elkington, 2021. Irons, A. and Elkington, S. (2021). *Enhancing learning through formative assessment and feedback*. Routledge.
- Jonassen, 2000. Jonassen, D. H. (2000). *Toward a design theory of problem solving*, volume 48. Springer.
- Kaufman and Schunn, 2010. Kaufman, J. H. and Schunn, C. D. (2010). Students' perceptions about peer assessment for writing: their origin and impact on revision work. *Instructional Science*, 39(3):387–406.
- Kolb and Kolb, 2005. Kolb, A. Y. and Kolb, D. A. (2005). Learning styles and learning spaces: Enhancing experiential learning in higher education. *Academy of Management Learning & Education*, 4(2):193–212.
- Kruger and Dunning, 1999. Kruger, J. and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6):1121–1134.
- Lin et al., 2001. Lin, S. S., Liu, E. Z.-F., and Yuan, S.-M. (2001). Web-based peer assessment: feedback for students with various thinking-styles. *Journal of computer assisted Learning*, 17(4):420–432.
- Mirmotahari and Berg, 2023. Mirmotahari, O. and Berg, Y. (2023). Utvikling av førsteårsstudenters tilbakemeldingskompetanse gjennom egen-vurdering. In *Norsk IKT-konferanse for forskning og utdanning*, number 4 in 1.
- Mirmotahari et al., 2019. Mirmotahari, O., Berg, Y., Fremstad, E., and Damsa, C. (2019). Student engagement by employing student peer reviews with criteria-based assessment. In *IEEE Global Engineering Education Conference (EDUCON)*, pages 1152–1157.
- Mirmotahari et al., 2003. Mirmotahari, O., Holmboe, C., and Kaasbøll, J. (2003). Difficulties learning computer architecture. In *Proceedings of the 8th annual conference on Innovation and technology in computer science education*, pages 247–247.
- Misiejuk and Wasson, 2021. Misiejuk, K. and Wasson, B. (2021). Backward evaluation in peer assessment: A scoping review. *Computers & Education*, 175:104319.
- Nicol and MacFarlane-Dick, 2006. Nicol, D. and MacFarlane-Dick, D. (2006). Formative assessment and selfregulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2):199–218.

- Nicol et al., 2014. Nicol, D., Thomson, A., and Breslin, C. (2014). Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1):102–122.
- Panadero et al., 2017. Panadero, E., Jonsson, A., and Botella, J. (2017). Effects of self-assessment on self-regulated learning and self-efficacy: Four meta-analyses. *Educational Research Review*, 22:74–98.
- Paul and Elder, 2006. Paul, R. and Elder, L. (2006). *Critical thinking: Tools for taking charge of your learning and your life*. Pearson Prentice Hall.
- Prince, 2004. Prince, M. (2004). Does active learning work? a review of the research. *Journal of Engineering Education*, 93(3):223–231.
- Rawlasyk, 2018. Rawlasyk, P. E. (2018). Assessment in higher education and student learning. *Journal of Instructional Pedagogies*, 21.
- Sadler, 2010. Sadler, D. R. (2010). Beyond feedback: developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, 35(5):535–550.
- Sadler and Good, 2006. Sadler, P. and Good, E. (2006). The impact of self and peer-grading on student learning. *Educational Assessment*, 11:1–31.
- Schön, 1983. Schön, D. A. (1983). *The Reflective Practitioner*. Basic Books.
- Sembey et al., 2024. Sembey, R., Hoda, R., and Grundy, J. (2024). Emerging technologies in higher education assessment and feedback practices: A systematic literature review. *Journal of Systems and Software*, page 111988.
- Shute, 2008. Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research*, 78(1):153–189.
- Sung et al., 2005. Sung, Y.-T., Chang, K.-E., Chiou, S.-K., and Hou, H.-T. (2005). The design and application of a web-based self- and peer-assessment system. *Computers & Education*, 45(2):187–202.
- Topping, 1998. Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3):249–276.
- van Zundert et al., 2010. van Zundert, M., Sluijsmans, D., and van Merriënboer, J. (2010). Effective peer assessment processes: Research findings and future directions. *Learning and Instruction*, 20(4):270–279.
- Vygotsky, 1978. Vygotsky, L. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press.
- Wanner and Palmer, 2018. Wanner, T. and Palmer, E. (2018). Formative self-and peer assessment for improved student learning: the crucial factors of design, teacher participation and feedback. *Assessment and Evaluation in Higher Education*, 43(7):1032–1047.
- Zimmerman, 2000. Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25(1):82–91.