

Studiebarometeret på programnivå: effekt av vurderingsformer og obligatoriske øvinger?

Guttorm Sindre^[0000-0001-5739-8265]

Institutt for datateknologi og informatikk, NTNU
guttorm.sindre@ntnu.no

Sammendrag. Avsluttende skriftlig eksamen er den dominerende vurderingsformen ved norske universiteter, særlig i tidlige studieår – gjerne i kombinasjon med obligatoriske øvinger. Dette gjelder også IT-studier, men det er betydelig variasjon mellom studieprogrammer. Et interessant spørsmål er hvorvidt studenters fornøydhetsnivå påvirkes av variasjon i vurderingsformer. Denne artikkelen ser på den vektende sammensetningen av vurderingsformer for teknisk-naturvitenskapelige studieprogram ved NTNU og undersøker korrelasjon mellom dette og studenters tilfredshet med de ulike programmene i Studiebarometeret. Resultatene indikerer at studenter jevnt over er noe mer positive til programmer med variert bruk av vurderingsformer, men det er få av spørsmålene i Studiebarometeret hvor forskjellen er signifikant. Særlig for obligatoriske øvinger er det likevel påfallende mange negative korrelasjoner med Studiebarometeret for 2020, som indikerer at dette er en praksis utdanningsinstitusjonene må tenke bedre gjennom. Problemet er ikke nødvendigvis obligatoriske aktiviteter som sådan, men den typiske måten disse gjøres på, hvor studentene får liten følelse av jobbrelevans, refleksjon og stimulering til nytenkning.

Nøkkelord: undervisning, vurdering, læringsutbytter, studenttilfredshet.

1 Innledning

God bruk av vurderingsformer er avgjørende både for studenters læring [1] og trivsel [2]. I forbindelse med prosjektet Fremtidens teknologistudier (FTS)¹ ved NTNU ble det gjort en kartlegging av bruk av vurderingsformer i studieprogrammer innenfor dette prosjektets skop. Kartleggingen fant blant annet følgende:

Eksamen er den dominerende vurderingsformen, men det er stor variasjon mellom studieprogram. Dette illustreres i Figur 1 for 5-årige masterprogrammer innen FTS-porteføljen. Hver rad viser et studieprogram med sin bokstavkode (f.eks. MTKOM = Kommunikasjonsteknologi, MTDI = Datateknologi). Blå strek viser andel av programmets obligatoriske emner hvor karakter settes kun basert på avsluttende eksamen, grønn viser tilsvarende andel hvor karakter settes på bakgrunn av arbeider (ingen eksamen). Rød + gul viser andelen emner i programmet hvor vurderingen er en kombinasjon av eksamen og annet (ofte kalt «mappe»), der rød viser eksamens andel av mapper, gul

¹ <https://www.ntnu.no/fremtidensteknologistudier>

viser andre vurderinger i mapper. Studieprogrammene er dermed sortert fra lavest til høyest andel eksamen i karaktersettingen av obligatoriske emner (blå + rød), hvor andelen eksamen varierer fra 29% (Arkitekt) til 83% (Materialteknologi). **Eksamen er enda mer dominerende i 1.årskurs, men igjen med stor variasjon mellom programmer.** Her går spennet fra 43% til 95% eksamen. Omfattende bruk av eksamen er neppe spesielt for NTNU; avsluttende skriftlig eksamen er nok den dominerende vurderingsformen i de fleste studieprogrammer ved norske universiteter, selv om andre vurderingsformer ofte anses som foretrukne blant pedagogisk ekspertise når det gjelder egenhet for å understøtte studenters læring [3].

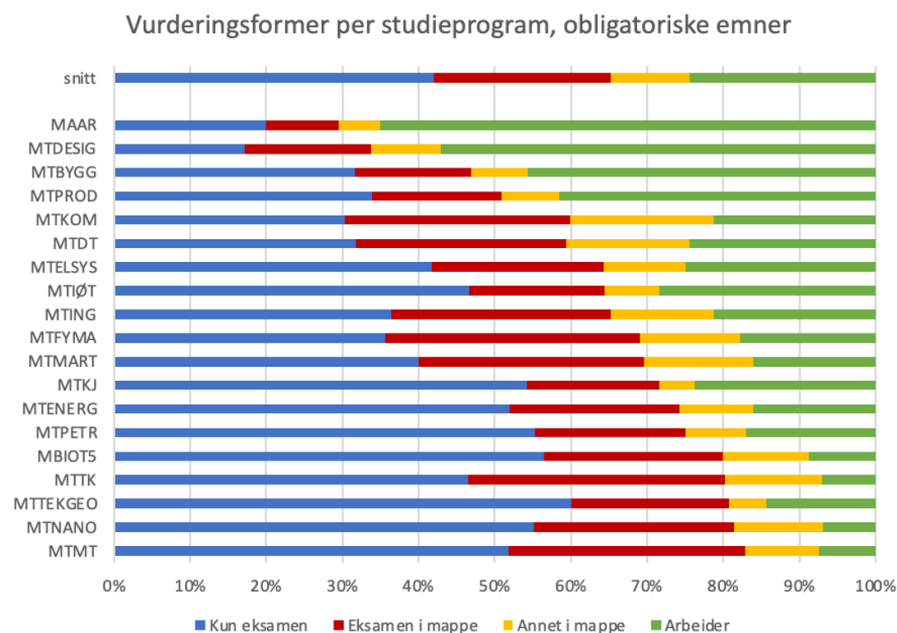


Fig. 1. Andel av ulike vurderingsformer ved 5-årige studieprogram i FTS-porteføljen.

Den dominerende karakterregelen er A-F (93%), det er få emner (7%) som bruker Bestått / Ikke bestått. Også her er det noe variasjon mellom studieprogram, men over et mye smalere spektrum enn for vurderingsformer. 18% av studieprogrammene bruker kun A-F, 62% har 90-99% A-F, 16% har 80-89% A-F, og bare 4% av FTS-studieprogrammene har bokstavkarakterer i mindre enn 80% av emnene.

Obligatoriske arbeidskrav (typisk øvinger) er mye brukt, men igjen med betydelig variasjon mellom programmer. Figur 2 illustrerer dette for bachelorprogrammene i FTS-porteføljen, hvor andelen av emner i studieprogrammet som har obligatoriske øvinger går fra 50% til 100%. For 5-årige masterprogrammer er spennet noe smalere, fra 57% til 78%. Obligatoriske øvinger er vanlig også for tekniske studier ved

andre norske læresteder, som indikert f.eks. av Lauvås som så på bruk av obligatoriske arbeid i IT-fagene [4].

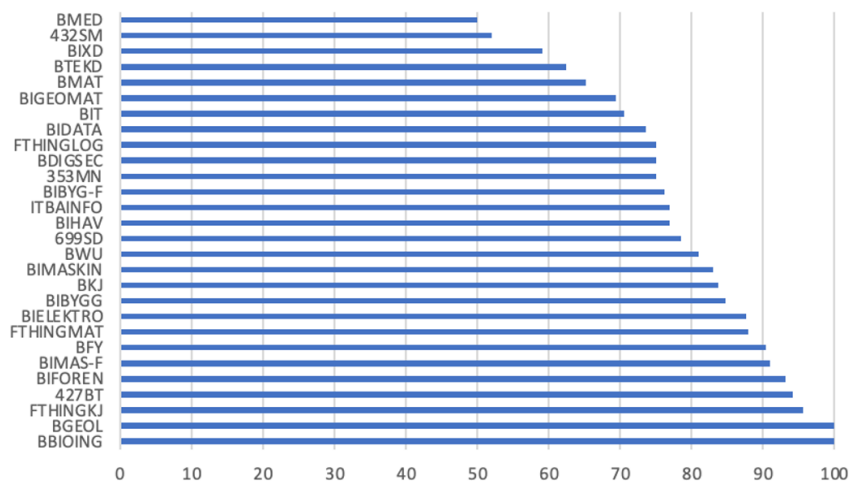


Fig. 2. Andel av emner med obliger ved 3-årige studieprogram i FTS-porteføljen.

Gitt den antatte viktigheten av både formativ og summativ vurdering for studenters læring og læringsmiljø, er det naturlig å anta at variasjoner i bruk av obligatoriske øvinger og eksamen versus andre vurderingsformer vil påvirke studenters tilfredshet i Studiebarometeret², som NOKUT hver høst sender til alle studenter i 2. og 5. årskurs av studieprogrammer ved norske universiteter og høyskoler. Spesifikt for Studiebarometeret 2020 var det i tillegg til de vanlige spørsmålene også tatt med 25 spørsmål som omhandlet universitetenes og studentenes håndtering av koronapandemien. Koronaspørsmålene er ikke inkludert i analysen vår siden de i mindre grad sier noe om langsiktige egenskaper ved studieprogrammene. Tabell 1 viser en oversikt over spørsmål som har inngått i analysen vår.

Forskningsspørsmålene for denne artikkelen er som følger:

FS1: Er det sammenheng mellom eksamensandel, karakterregel og score i Studiebarometeret, enten på indekser (1a) eller ikke-indekserte spørsmål (1b)?

FS2: Er det sammenheng mellom andel emner med obligatoriske øvinger og score i Studiebarometeret, enten på indekser (2a) eller ikke-indekserte spørsmål (2b)?

FS3: Fins det andre lett identifiserbare faktorer som påvirker Studiebarometeret mer enn vurderingsformer eller obligatoriske øvinger, f.eks. kullstørrelse, kvinneandel eller opptaksgrenser?

² <https://www.studiebarometeret.no/no/>

Tabell 1. Spørsmålsgrupper i Studiebarometeret som inngikk i analysen.

Indekser	Temaer	# spm
Undervisning	Engasjerende, forståelig, god dekning, aktiviserende	4
Tilbakemelding og veiledning	Fra fagstab: Antall, konstruktiv, veil.; Fra medstudenter	4
Miljø	Sosialt, faglig, mellom stud. og ans.	3
Organisering	Tilgj.info, kvalitet info, faglig smheng, adm. tilrettel.	4
Vurderingsformer	Smh.pensum, klare kriterier, forståelse, faglig utvikling	4
Inspirasjon	Stimulerende, utfordrende, motiverende	3
Engasjement	Motivasjon, aktivitet, innsats, forberedt	4
Forventning	Klare forventninger, forberedt, deltagelse, ambisjoner	4
Digitale verktøy	Aktiviserende, kompetanse, opplæring, læringsplattform	4
Yrkesrelevans	Eksempelbruk, bransje, formidling, bidrag, prosjekt	5
Spørsmålsgrupper uten indeks		# spm
Tilbakemelding og veiledning	Før innlev., etter, ikke-skriftl, andre stud., faglig utv., ant.	6
Medvirkning	Mulighet til å medvirke i innhold og underv.metoder	1
Læringsutbytter	Teori, forskningsmetode, FoU-erfaring, fag- / yrkes-spes. ferdigheter, refleksjon, samarbeid, muntlig, skriftlig, tenke nytt, selvstendig arbeid	10
Tidsbruk	Org. læringsaktiviteter, selvstudium, jobb utenom	3
Overordnet tilfredshet	Studium du helst ville gå på? Alt i alt tilfreds?	2

Resten av artikkelen er strukturert som følger: Seksjon 2 presenterer noe tidligere forskning relatert til Studiebarometeret og sammenhenger mellom dette og andre utdanningsrelaterte metrikker. Seksjon 3 gjør rede for forskningsmetode, metrikker og hypoteser. Seksjon 4 presenterer resultatene, og seksjon 5 gir en avsluttende diskusjon og konklusjon.

2 Relatert arbeid

NOKUT selv publiserer hvert år flere rapporter basert på dataene i Studiebarometeret, f.eks. [5] for en detaljert gjennomgang av resultatene for 2020. Andre har også forsket på data fra Studiebarometeret. Mye av forskningen har handlet om interne sammenhenger i Studiebarometeret. For eksempel analyserte Hovdhaugen et al. [6] korrelasjoner internt i ulike indekser i spørreskjemainstrumentet slik det forelå ved undersøkelsen høsten 2015. Munthe [7] så spesifikt på resultater på spørsmålene om studentengasjement, for diverse ulike studietyper. Abrahamsen et al. [8] så tilsvarende på sammenhengen mellom motivasjon, undervisning og konstruktive tilbakemeldinger, med spesifikt fokus på resultater for studenter ved UiS. Wiers-Jensen og Hovdhaugen [9] analyserte resultater knyttet til studieinnsats, hvordan denne varierer mellom ulike programmer, og hvordan den henger sammen med andre faktorer.

En forskjell mellom denne artikkelen og mange av de ovennevnte er at vi *ikke* har analysert data i form av studenters individuelle svar, *kun* data på programnivå. Fokus er ikke på sammenhenger internt i Studiebarometeret, men hvordan resultater fra Studiebarometeret henger sammen med data hentet fra andre kilder – med tanke på å finne hvilke faktorer som i større eller mindre grad kan ha effekt på resultatene i Studiebarometeret. Videre ser vi kun på data fra bachelorprogrammer (28) og 5-årige masterprogrammer (19) innen den såkalte FTS-porteføljen ved NTNU (dvs. $N = 28+19 = 47$),

mens andre undersøkelser enten har sett på andre studieprogrammer, eller alle slags studieprogrammer.

3 Forskningsmetode

Grunnen til at vi velger å se på kun bachelorprogrammer og 5-årige masterprogrammer innen FTS-porteføljen, og ikke også de 2-årige masterprogrammene (63) i FTS-porteføljen, er at de 2-årige masterprogrammene spriker mye mer når det gjelder rekrutteringsgrunnlag og undervisningsopplegg. Noen masterprogram er nasjonale, andre internasjonale; noen rekrutterer primært rett fra bachelor, andre er erfaringsbaserte og ment videreutdanning for personer som har vært noen år i yrkeslivet. Bachelor og 5-årige masterprogram har derimot til felles at de har opptak direkte fra videregående studier og stort sett rekrutterer fra Norge. Det store antallet 2-årige masterprogram gjør dessuten at disse ville blitt dominerende i datagrunnlaget hvis de ble tatt med, samtidig som mange av disse programmene har forholdsvis få studenter, hvorav igjen ikke alle har svart på Studiebarometeret slik at data for 2-årige masterprogrammer også ville være mer usikre. Det fremsto dermed som metodisk mest solid å begrense undersøkelsen til kun bachelor og 5-årig master. Som det også står på NOKUTs egen webside om publisering av data fra Studiebarometeret: «Vi gjør oppmerksom på at det er systematiske forskjeller mellom studieprogram i ulike fagområder, en sammenligning bør derfor helst gjøres av like, eller relativt like, studieprogram. Jo færre svarende et studieprogram har, jo mindre robuste er dataene.»³

Data på programnivå fra Studiebarometeret 2020 ble lastet ned som et Excel-ark fra Studiebarometerets nettside³. Data knyttet til studieprogrammer og emner ligger lagret i FS, disse ble hentet ut av personer i FS-gruppa ved NTNU ved spørringer i FS, dels ved hjelp av dataanalyseverktøyet Tableau. Vi mottok disse dataene i form av et stort Excel-ark på 7924 rader og 12 kolonner. Det store antallet rader skyldes blant annet at emner som inngår i mange studieprogram er gjentatt for hvert program. Regnearket med rådata ble omgjort til en lettere analyserbar form, for eksempel ved at tekstlige brøker som «30/100» ble omgjort til tall. En del hjelpekolonner ble også lagt til for å lette påfølgende telling og aggregering av data som vi var spesielt interessert i. Data ble så aggregert per studieprogram, for å finne fram til følgende metrikker som er brukt videre i forskningen til denne artikkelen:

- **%EKS:** for studieprogrammets obligatoriske emner, hvor stor prosentandel av karaktersettingen er basert på eksamen. Dette er tallet som fremkommer som blå + rød i bjelkene for hvert studieprogram i Figur 1.
- **%A-F:** prosentandel av emner i programmet som bruker bokstavkarakterer.

³ <https://www.studiebarometeret.no/no/artikkel/5>. (En del av) de data som er benyttet her er hentet fra Studiebarometeret (2020). Undersøkelsen er finansiert av Kunnskapsdepartementet (KD). Data i anonymisert form er stilt til disposisjon av NOKUT. Verken NOKUT, KD eller NSD er ansvarlig for analysen av dataene, eller de tolkninger som er gjort her.

- **%OBL:** prosentandel av obligatoriske emner i studieprogrammet som har obligatoriske øvingsaktiviteter, jfr. Figur 2.

Det er ikke bare interessant å se hvorvidt disse metrikkene har sammenheng med resultater i Studiebarometeret, men også hvorvidt sammenhengen er sterkere eller svakere enn andre metrikker på studieprogramnivå. Vi så derfor også på:

- **%KV:** kvinneandel på studieprogrammene, som oppgitt i kontekstdata til Studiebarometeret (som igjen har hentet disse fra DBH).
- **KST:** kullstørrelse per studieprogram, grovt estimert ved å ta antall studenter på studieprogrammet, som oppgitt i kontekstdata til Studiebarometeret, og så dele dette på antall årskurs (3 eller 5).
- **ORDF:** opptaksgrense med førstegangsvitnemål for hvert studieprogram i 2020, som oppgitt på nettsidene til Samordna opptak. Enkelte studieprogram hadde opptak i flere byer (Gjøvik / Trondheim / Ålesund) – her ble et vektet snitt av grensene brukt. For studieprogram uten oppgitt grense («Alle») ble grensen satt til 30, siden de laveste med grenser oppgitt innen FTS-porteføljen lå lavt på 30-tallet.

Nullhypotesene for studien var som følger:

- H_{1,0}: %EKS har ingen signifikant effekt på svarene i Studiebarometeret
- H_{2,0}: %A-F har ingen signifikant effekt på svarene i Studiebarometeret
- H_{3,0}: %OBL har ingen signifikant effekt på svarene i Studiebarometeret
- H_{4,0}: Andre faktorer som ORDF, %KV eller KST har ingen signifikant effekt på svarene i Studiebarometeret

For hver indeks i Studiebarometeret, og for hvert enkeltspørsmål, analyserte vi korrelasjon med de ovennevnte metrikkene. Dette ble gjort med den vanlige Excel-funksjonen CORREL() som gir Pearson-korrelasjon mellom to datasett (f.eks. studieprogrammets eksamensprosent versus de samme studieprogrammets score på indeksen «Vurdering» i Studiebarometeret). For korrelasjoner $\geq 0,3$ ble det utført regresjonsanalyse for å se om de var signifikante, dette ble gjort ved dataanalysefunksjonen Regresjon i Excel. Et problem når man gjør mange ulike sammenligninger, er imidlertid at det alltid vil være noe som kommer opp med høy korrelasjon og dermed viser seg å være signifikant. Med et konfidensintervall på 95% er det 5% sjanse for at konklusjonen kan være en falsk positiv. Dette er til å leve med hvis man gjør bare én eller noen få hypotesetester, men her har vi 3 FS-baserte metrikker (%EKS, %A-F, %OBL) og 3 andre metrikker (%KV, KST, ORDF) som testes mot 10 indekser fra Studiebarometeret, altså 60 ulike tester. Da vil man forvente å få noe i retning av $60 \times 0,05 = 3$ signifikante resultater bare ved ren tilfeldighet. For å kompensere for dette, ble resultatene korrigert i henhold til Benjamini-Hochberg-metoden [10]. For de som da fortsatt gjensto som signifikante, ble det også gjort en analyse av Studiebarometerdata for 2019 for å se om samme relasjon fantes der. 5 av bachelorprogrammene i FTS-porteføljen var nye i 2019 og hadde kun studenter i 1.årskurs, dermed ingen respondenter. Dette var BMED, BIBYG, BIBYG-F, BIMAS og BIMAS-F. Disse ble utelatt fra analysen, som så ble utført spesifikt for de indeksene som var signifikante for 2020.

4 Resultater

4.1 Resultater for indekser i Studiebarometeret

Figur 3 gir en oversikt over resultatene av analyse av hvordan programmets score på ulike indekser i Studiebarometeret 2020 korrelerer med metrikkene nevnt i forrige seksjon. Hver gruppe av seks søyler i ulike farger viser korrelasjoner for én indeks, med %OBL blå, %EKS rød, %A-F grå, KST gul, %KV rosa og ORDF grønn. Søylegruppen helt til høyre viser snittet av korrelasjonene, %OBL mest negativ (-0,26) og ORDF mest positiv (+0,26). De 9 indeksene er fra venstre mot høyre: Undervisning (UND), Tilbakemelding og veiledning (TILB), Læringsmiljø (MILJ), Organisering (ORG), Vurderingsformer (VURD), Inspirasjon (INSP), Engasjement (ENG), Forventninger (FORV), Digitale verktøy (DIGI) og Yrkesrelevans (YRKE).



Fig. 3. Korrelasjoner med indekser i Studiebarometeret 2020.

En kjapp oppsummering av resultatene i Figur 3 er som følger:

- %OBL korrelerer negativt med alle indeksene, og signifikant for TILB ($r=-0,37$; $p=0,02$), ENG ($r=-0,39$; $p=0,007$) og YRKE ($r=-0,42$; $p=0,003$). Det virker dermed som høyere andel emner med obligatoriske øvinger gir lavere tilfredshet blant studentene.
- %EKS og %A-F korrelerer stort sett negativt, men mindre tydelig enn %OBL. Signifikante for %EKS er TILB ($r=-0,29$; $p=0,04$), FORV ($r=-0,31$; $p=0,04$) og YRKE ($r=-0,37$; $p=0,01$), og for %A-F TILB ($r=-0,37$; $p=0,03$) og FORV ($r=-0,45$;

$p=0,004$). Høyere andel eksamen og bokstavkarakterer ser dermed også ut til å gi lavere tilfredshet på enkelte indekser.

- For KST og %KV fremstår resultatene mer blandet, med noen positive og noen negative korrelasjoner. KST heller noe i negativ retning, men den eneste signifikante er TILB ($r=-0,39$; $p=0,007$). %KV slår ut noe mer i positiv retning, og eneste signifikante er den positive korrelasjonen for ENG ($r=+0,32$; $p=0,03$).
- ORDF har gjennomgående positive korrelasjoner, og hele fem av indeksene er signifikant positive: MILJ ($r=+0,32$; $p=0,03$), ORG, ($r=+0,41$; $p=0,004$), VURD ($r=+0,34$; $p=0,02$), INSP ($r=+0,50$; $p=0,000$), DIGI ($r=+0,42$; $p=0,003$). Med andre ord: Jo høyere opptaksgrense et studieprogram har, jo mer fornøyde ser studentene ut til å være. Samme analyse ble forsøkt også med ORD (opptaksgrense utenom førstegangskvoten), men her var korrelasjonene vesentlig lavere og ikke signifikante (MILJ: $r=+0,05$; ORG: $r=+0,22$; VURD: $r=+0,02$; INSP: $r=+0,19$; DIGI: $r=+0,14$).

Etter Benjamini-Hochberg-korreksjon gjensto kun fire resultater som signifikante: %A-F korrelerer negativt med opplevde forventninger fra fagstaben, OBL% korrelerer negativt med opplevd yrkesrelevans, og ORDF korrelerer positivt med inspirasjon og tilfredshet med digitale verktøy.

Det ble undersøkt om disse fire korrelasjonene var signifikante også for Studiebarometeret 2019. To av dem viste seg å være signifikante i 2019-dataene (%OBL vs. yrkesrelevans, ORDF vs. inspirasjon) mens de to andre ikke var signifikante for 2019.

4.2 Tilbakemelding og veiledning

Indeksen med flest signifikante resultater er Tilbakemelding og veiledning, hvor det er betydelig negativ korrelasjon med fire av metrikkene våre: %OBL, %EKS, %A-F og KST. Figur 4 gir et mer detaljert innblikk i resultatene her, med samme fargebruk som tidligere (%OBL: blå, %EKS: rød, %A-F: grå, KST: gul).

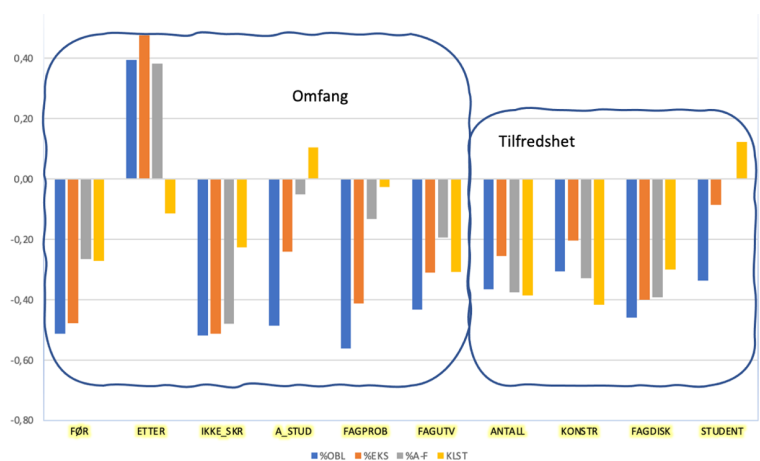


Fig. 4. Korrelasjoner med spørsmål om tilbakemelding og veiledning

Som vist i Figur 4 er det 10 spørsmål i SB som handler om tilbakemelding og veiledning, hvorav 6 handler om *omfanget* av tilbakemelding og veiledning (til venstre i figuren) og 4 handler om *tilfredshet* med tilbakemelding og veiledning (til høyre). Spørsmålene om omfang har stammen «Hvor ofte hittil i studiet har du: ...» med enkeltspørsmål:

- «Fått tilbakemeldinger fra faglig ansatte på skriftlig arbeid før endelig innlevering?» (kalt FØR i Figur 5)
- «Fått tilbakemeldinger fra faglig ansatte etter endelig innlevert skriftlig arbeid?» (ETTER)
- «Fått tilbakemeldinger fra faglig ansatte på ikke-skriftlig arbeid?» (IKKE_SKR)
- «Fått tilbakemeldinger fra andre studenter på skriftlig eller ikke-skriftlig arbeid?» (A_STUD)
- «Diskutert faglige problemstillinger med faglig ansatte?» (FAGPROB)
- «Diskutert din faglige utvikling/resultater med faglig ansatte?» (FAGUTV)

Spørsmålene om tilfredshet har stammen «Hvor tilfreds er du med ...» og enkeltspørsmål «Antall tilbakemeldinger du får fra faglig ansatte på arbeidet ditt?» (ANTALL), «De faglig ansattes evne til å gi konstruktive tilbakemeldinger på arbeidet ditt? [...]», «Faglig veiledning og diskusjoner med faglig ansatte?» (FAGDISK) og «Medstudenters evne til å gi konstruktive tilbakemeldinger [...]?» Det er disse fire spørsmålene om tilfredshet med tilbakemelding og veiledning som inngår i indeksen for tilbakemelding og veiledning som inngikk i Figur 4, mens spørsmålene om omfang er holdt utenfor indeksen.

For spørsmålene om omfang har %OBL, %EKS og %A-F tydelig positive korrelasjoner på ETTER, men negative korrelasjoner på omfang av all annen tilbakemelding. Når det gjelder tilfredshet er det gjennomgående negativt, bortsett fra en svak (ikke-signifikant) positiv korrelasjon for klassestørrelse når det gjelder tilfredshet med medstudenters evne til konstruktiv tilbakemelding.

4.3 Læringsutbytter

Studiebarometeret har også en del andre spørsmål som ikke inngår i noen indeks. En betydelig gruppe av slike spørsmål angår studenters opplevde læringsutbytte, hvor korrelasjoner er vist i Figur 5. Disse har stammen «Hvor tilfreds er du med eget læringsutbytte hittil i studiet, når det gjelder: ...», og med kompletteringer «Teoretisk kunnskap» (TEORI i figuren), «Kunnskap om vitenskapelig arbeidsmetode og forskning» (FORSK), «Egen erfaring med forsknings- og utviklingsarbeid» (FoU), «Yrkes- og fagspesifikke ferdigheter» (FAGSPES), «Evne til refleksjon og kritisk tenkning» (REFLEKS), «Samarbeidsevne» (SAMARB), «Muntlig kommunikasjonsevne» (MUNKOM), «Skriftlig kommunikasjonsevne» (SKRKOM), «Evne til å tenke nytt» (TENKE), «Evne til å arbeide selvstendig» (SELVST). Fargekodingen i Figur 5 er den samme som før, og sortert etter gjennomsnittlig korrelasjon (fra mest negativ til mest

positiv) kommer metrikkene ut i samme rekkefølge på disse spørsmålene om læringsutbytte som de gjorde for indeksene i Figur 3 – med %OBL, %EKS, %A-F, KST som negative, %KV og ORDF som positive.

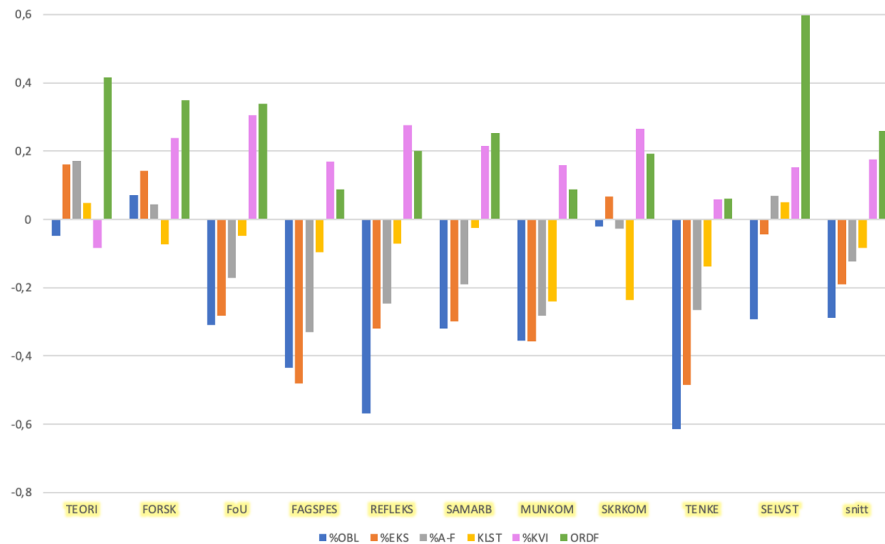


Fig. 5. Korrelasjoner med spørsmål om læringsutbytter.

For kvinneandel er de positive utslagene stort sett for små til å være signifikante, mens det for ORDF er signifikante utslag særlig for evne til selvstendig arbeid ($r=+0,60$; $p=0,000$) og teoretisk kunnskap ($r=+0,42$; $p=0,004$). For %OBL er korrelasjonene nær null for teoretisk kunnskap, forskningsmetode og skriftlig kommunikasjon, men ellers går de klart i negativ retning, med særlig store utslag for evne til å tenke nytt ($r=-0,61$; $p=0,000$), evne til refleksjon og kritisk tenkning ($r=-0,56$; $p=0,000$), og yrkes- og fagspesifikke ferdigheter ($r=-0,43$; $p=0,002$). På sistnevnte er den negative korrelasjonen med %EKS enda større ($r=-0,48$; $p=0,001$), og %EKS har også en tilsvarende negativ korrelasjon med evne til å tenke nytt ($r=-0,49$; $p=0,001$).

4.4 Tidsbruk, medvirkning, overordnet tilfredshet

En annen gruppe spørsmål som ikke inngår i noen indeks, men som ofte tillegges en viss vekt, har resultater illustrert i Figur 6. Dette gjelder studentenes selvrapporterte tidsbruk – på henholdsvis organiserte læringsaktiviteter (TID_ORG), egeninnsats på læring (TID_EGEN) og på betalt arbeid utenom studiene (TID_JOB). Et annet interessant spørsmål er hvorvidt studentene føler de har mulighet til medvirkning på avgjørelser om studiets innhold og opplegg (MEDVIRK). Avslutningsvis i undersøkelsen

fins det to spørsmål om overordnet tilfredshet: Hvorvidt man går på det studieprogrammet man helst ville ha gått på (HELST), og hvorvidt man alt i alt er fornøyd med studieprogrammet (ALTIALT)..

For tidsbruk ser vi at %EKS (rød) har en betydelig positiv korrelasjon med tidsbruk på organiserte læringsaktiviteter, men tilsvarende negativ på tidsbruk til egeninnsats. Lignende mønster ses for %OBL, bare at korrelasjonen for TID_ORG er litt for liten til å være signifikant. %OBL har også klart negative korrelasjoner med følelse av medvirkning, samt til de to spørsmålene om overordnet tilfredshet. Den eneste faktoren som korrelerer signifikant positivt med de to spørsmålene om overordnet tilfredshet, er ORDF – dvs., jo høyere opptaksgrense, jo mer tilfredse er studentene med programmet.

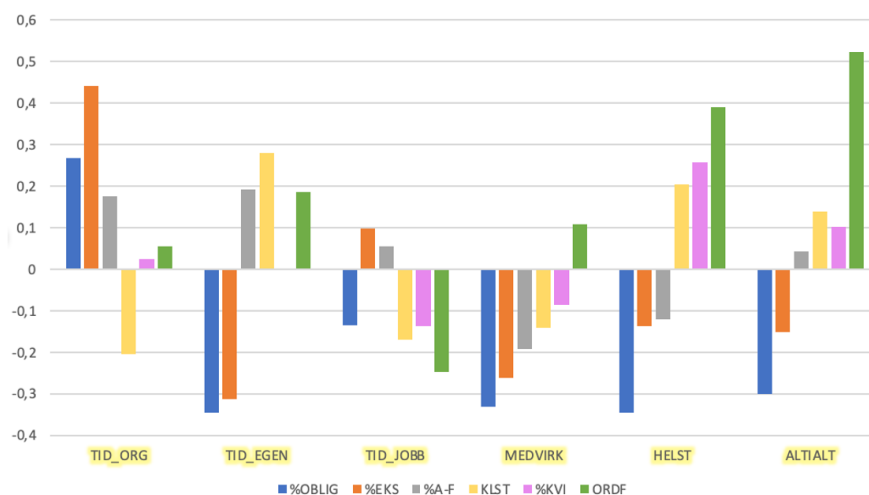


Fig. 6. Korrelasjoner med spørsmål om tidsbruk, medvirkning, tilfredshet.

5 Diskusjon og konklusjon

Gitt resultatene i forrige seksjon blir svarene på forskningsspørsmålene som følger:

- FS1: Til tross for hovedsakelig negative korrelasjoner mellom %EKS og indekser i studiebarometeret, ble ingen av funnene knyttet til %EKS stående igjen som signifikante etter Benjamini-Hochberg-korreksjon – kun ett funn knyttet til negativ korrelasjon mellom %A-F og opplevelse av forventninger fra fagstaben. Denne ene er for så vidt nok til å avvise nullhypotesen, men funnene er ikke veldig tydelige, særlig siden samme korrelasjon ikke var signifikant for 2019. Funnet kan dermed være relatert til spesielle omstendigheter i 2020 snarere enn å indikere permanente relasjoner mellom %EKS, %A-F og tilfredshet i Studiebarometeret.
- FS2: Nullhypotesen må avvises. Andel emner med obligatoriske øvinger har tydelige korrelasjoner i negativ retning for de fleste aspekter av Studiebarometeret. Etter Benjamini-Hochberg-korreksjon er det kun én indeks som har signifikant resultat

(negativ korrelasjon mellom %OBL og opplevd yrkesrelevans – som også ble funnet for 2019), men dessuten er det sterke negative korrelasjoner med opplevde læringsutbytter særlig mhp evne til refleksjon og nytenkning.

- FS3: ORDF – altså opptaksgrense med førstegangsvitnemål – viste seg å ha gjennomgående positiv korrelasjon med svar i Studiebarometeret, og for to indekser var den signifikant selv etter Benjamini-Hochberg-korreksjon, hvorav én var signifikant også for 2019 (positiv korrelasjon med inspirasjon). Andre input-variable hadde mindre tydelig effekt.

Påvisning av signifikante korrelasjoner er ikke nødvendigvis konstruktivt i seg selv, det er mer interessant å diskutere *hvorfor*, og *hvordan* disse innsiktene kan brukes til å forbedre studieprogrammene. Vårt datagrunnlag kan i liten grad si noe om årsakssammenhenger. For eksempel:

Hvorfor korrelerer ORDF positivt med tilfredshet? Er det fordi studieprogrammer med høy tilfredshet gradvis har tiltrukket seg studenter med høyere poengsum (dvs. høy ORDF er effekten av høy tilfredshet), eller at mer skoleflinke studenter er lettere å tilfredsstille fordi de klarer seg bra enten undervisningen er god eller dårlig (dvs. høy ORDF er årsaken til høy tilfredshet) – eller en mer kompleks vekselvirkning mellom flere slike årsakssammenhenger? Det faktum at ORDF bl.a. har signifikant positiv korrelasjon med tilfredsheten med digitale verktøy – som stort sett vil være de samme for alle disse NTNU-programmene (Blackboard, Inspira, ...) – kan tyde på at ORDF er mer årsak enn virkning: skoleflinke studenter lykkes i større grad, og er dermed mer fornøyd enn svakere studenter som eksponeres for de samme verktøyene.

Hvorfor korrelerer %OBL negativt med såpass mange aspekter av Studiebarometeret? En mulig forklaring er at obligatoriske øvinger i seg selv har en negativ effekt på studenters tilfredshet – for eksempel når det gjelder tilbakemelding (at studenter stort sett oppfatter tilbakemeldingene på øvinger og eksamen som mangelfull). En annen mulig forklaring kan være at obligatoriske øvinger og eksamen tar tid (for fagstab og studenter), og at studieprogrammer som har lavere %OBL og %EKS i stedet inkluderer andre læringsmetoder og vurderingsformer med mer tilbakemelding. Haugan et al. argumenterer i [11, 12] for at bruken av obligatoriske arbeidskrav bør reduseres for å stimulere studenters indre motivasjon heller enn ytre kontroll, og beskriver en vellykket pilot med et emne hvor obligatoriske øvinger ble erstattet med rent formativ vurdering. Lignende motivasjon lå bak et forsøk fra Hellem og Lorås, hvor en forsøksgruppe studenter slapp obligatoriske øvinger i et emne hvor resten av studentene hadde dette [13]. Forsøksgruppa hadde ikke dårligere læringsutbytte enn kontrollgruppa, men heller ikke bedre. Lauvås og Sandnes [14] utførte en spørreundersøkelse blant studenter, kombinert med intervjuer, for å finne ut hva studentene selv synes om obligatoriske øvinger. Jevnt over var studentene positive til å ha obligatoriske øvinger, dog under visse forutsetninger (engasjerende og motiverende, passende vanskegrad og omfang, høy eksamensrelevans). Det behøver altså ikke være eksistensen av obligatoriske øvinger som sådan som gjør at dette korrelerer negativt med tilfredshet, det kan i stedet være *hvor* obligatoriske øvinger tenderer til å bli brukt i de aktuelle studieprogrammene – for eksempel at vanskegraden ikke oppleves som passende, at det blir for høyt arbeidspress, at det blir lite tilbakemelding og refleksjon rundt øvingene, eller at de oppleves som lite

relevante. Et kjent fenomen ved NTNU er dessuten «koking» [15] - at studenter etter hvert sliter med å overholde tidsfrister og ender opp med å kopiere svar på obligatoriske øvinger fra medstudenter heller enn å faktisk gjøre øvingen. Dette gir lavt læringsutbytte, og det er ikke overraskende at en slik praksis vil gi negativ korrelasjon med spørsmål og læringsutbytter knyttet til refleksjon og nytenkning.

Sutherland et al. [16] ser på hvilke faktorer som påvirker studenters tilfredshet med emner. Deres hovedfunn er at direkte kontaktid mellom fagstab og studenter er en driver for tilfredshet, mens høy opplevd vanskegrad i et emne er en driver for misnøye. Dette kan tilsi at det ikke nødvendigvis er høy prosentandel eksamen eller obligatoriske øvinger som er årsak til misnøye som sådan, men at dette er typisk opplegg i emner med mange studenter, hvor det gjerne blir mindre tid til individuell kontakt og lite bruk av alternative vurderingsformer. En internasjonal studie av Garcia-Aracil [17] så på alumni sin fornøydhet med studieprogrammer, både med hensyn på faktorer i utdanningene selv (f.eks. undervisningskvalitet) og utenforliggende faktorer (som kjønn, økonomi, familiebakgrunn etc.). Faktorer i utdanningene som typisk forårsaker misnøye blant studenter var manglende mulighet til å delta i forskningsprosjekt, mangel på gode læringsressurser, dårlige muligheter for praksis eller arbeidsrelevante erfaringer gjennom studiet, vurderingsformer og karaktersystemer, og få muligheter for kontakt med undervisere utenom organiserte læringsaktiviteter. Vurderingsformer og karaktersystemer er således en viktig faktor også internasjonalt. Obligatoriske øvinger nevnes ikke eksplisitt i Garcia-Aracils konklusjoner, men kan til dels ses som eksempel på en vurderingsform (om de ikke gir karakter, må man iallfall få godkjent for å få adgang til eksamen), og dels også som en læringsressurs – som kan være av variabel kvalitet. Konklusjonen er derfor ikke nødvendigvis at man må slutte med obligatoriske øvinger og avsluttende eksamen, men at man trenger mer variasjon – og *hvis* man har f.eks. obligatoriske øvinger, bør man sørge for at de fungerer bra som læringsressurs for studentene.

Den studien vi har gjennomført her, har klare metodiske svakheter. Spesielt innebærer tråling etter høye korrelasjoner i en stor mengde sammenlikninger på kryss og tvers, at man ikke kan konkludere med signifikante funn basert på et vanlig 95% konfidensintervall. Ved korreksjon for dette, gjensto bare fire funn som signifikante: opplevd yrkesrelevans påvirkes negativt av obligatoriske øvinger, opplevde forventninger fra fagstaben påvirkes negativt av bokstavkarakterer, og ORDF påvirket positivt inspirasjon og tilfredshet med digitale verktøy. Dette kan innebære at vi har utelukket funn som ville ha vært – og burde ha vært – signifikante med bruk av en mer målrettet metodikk. En mulig vei for fremtidig forskning vil dermed være å se på tilsvarende data for årene videre framover, og da med målrettede hypoteser f.eks. om at %OBL og %EKS vil påvirke visse indekser i Studiebarometeret negativt – slik at man kan benytte vanlige konfidensintervaller. Tilsvarende analyser de neste årene vil også bote på en annen viktig utfordring ved denne undersøkelsen, nemlig at de andreårsstudentene som responderte på Studiebarometeret i 2020 har studert under dels eksepjonelle forhold, med undervisningsopplegg som brått måtte tilpasses smittevernstiltak. Tilsvarende analyser flere år i strekk vil være interessant for å se om f.eks. %OBL fortsetter å ha like stor negativ korrelasjon med Studiebarometeret – som kan tyde på et iboende problem med obligatoriske øvinger – eller om denne negative korrelasjonen forsvinner, som

snarere vil kunne tyde på at studieprogrammer med mye obligatoriske øvinger hadde vanskeligere for å tilpasse seg pandemien, men kan fungere greit under normale forhold.

Referanser

1. Wiliam, D., What is assessment for learning? *Studies in educational evaluation*, 2011. 37(1): 3-14.
2. Jones, E., et al., *Student wellbeing and assessment in higher education: the balancing act*. *Assessment & Evaluation in Higher Education*, 2021. 46(3): 438-45.
3. Raaheim, A. (2019) Eksamensrevolusjonen: Råd og tips om eksamen og alternative vurderingsformer. 2. utgave, Oslo: Gyldendal.
4. Lauvås Jr, P.; Sandnes, T. (2019) Mandatory coursework in higher Norwegian IT education. Norsk IKT-konferanse for forskning og utdanning.
5. Pedersen, L.F.A.; Wiggen, K.S. (2021) Studiebarometeret 2020: Dokumentasjonsrapport. 2021, NOKUT.
6. Hovdhaugen, E., et al., Indikatorer på kvalitet i høyere utdanning. 2016, NIFU.
7. Munthe, E., Studentengasjement i høyere utdanning: Hvilken innsikt gir Studiebarometeret? *Uniped*, 2020. 43(02): 04-116.
8. Abrahamsen, E.B., et al. (2020) Om sammenhengen mellom motivasjon, studentaktiv undervisning og konstruktive tilbakemeldinger. En studie fra Universitetet i Stavanger. *Uniped*, 43(04): 298-311.
9. Wiers-Jenssen, J.; Hovdhaugen, E. (2019) Studieinnsats på lavere grad–hva kan Studiebarometeret fortelle oss? *Uniped*, 42(03): 274-289.
10. Thissen, D., Steinberg, L. Kuang, D. (2002) Quick and easy implementation of the Benjamini-Hochberg procedure for controlling the false positive rate in multiple comparisons. *Journal of educational and behavioral statistics*, 27(1): 77-83.
11. Haugan, J., Lysebo, M.; Lauvås, P. (2017) Mandatory coursework assignments can be, and should be, eliminated! *European Journal of Engineering Education*, 42(6): 1408-1421.
12. Haugan, J.; Lysebo, M. (2018) Hvorfor antall arbeidskrav bør reduseres. *Uniped*, 41(03): 347-360.
13. Hellem, V.; Lorås, M. (2020) The effect of mandatory assignments on students learning outcome and performance in introductory programming courses. 2020 IEEE Global Engineering Education Conference (EDUCON).
14. Lauvås, P.; Sandnes, T. (2020) IT students' perceptions of mandatory coursework. Norsk IKT-konferanse for forskning og utdanning.
15. Rønning, F., et al., (2017) Bruk av digitale øvinger i grunnutdanningen i matematikk. *Læring om læring*, 1: 113-119.
16. Sutherland, D.; Warwick, P.; Anderson, J. (2019) What factors influence student satisfaction with module quality? A comparative analysis in a UK business school context. *The International Journal of Management Education*, 17(3): 100312.
17. García-Aracil, A., European graduates' level of satisfaction with higher education. *Higher Education*, 2009. 57(1): 1.