



UiT Noregs arktiske universitet

Fakultet for humaniora, samfunnsvitenskap og lærerutdanning

Kjønnsvurderingsgapet – Hvorfor gjør gutter det relativt bedre på eksamen enn jenter?

En studie om kjønnsvurderingsgapet i videregående opplæring

Herman Hovda Lien

SOS-3981, våren 2023

Innholdsfortegnelse

Sammendrag	2
Forord.....	3
1.1 Innledning.....	4
1.2 Hvorfor er det viktig å forske på dette?.....	5
1.3 Oppgavens struktur	6
2. Tidligere forskning på kjønnsvurderingsgapet	7
3. Teori.....	12
3.1 Forskningslogikk og paradigme.....	12
3.2 Begreper som brukes i analysen.....	14
4. Metode	19
4.1 Kvantitativ del.....	20
4.2 Kvalitativ del.....	22
4.3 Forskningsetiske hensyn.....	29
4.4 Metodens troverdighet.....	30
5. Resultater.....	32
5.1 Presentasjon av kvantitative data	32
5.2 Presentasjon av kvalitative data	33
5.2.1 Oppførsel og vurdering.....	34
5.2.2 Om eksamen.....	37
5.2.3 Kjønnsvurderingsgapet i skolen	40
5.2.4 Forklaringer på kjønnsvurderingsgapet	42
5.2.5 Presise vurderinger	47
6. Analyse av data.....	51
6.1 Er det kjønnsforskjeller i prestasjonsavvik mellom standpunkt og eksamen i Norge for studiespesialiserende linje?	52
6.2 Hva er årsakene til kjønnsvurderingsgapet?	52
6.3 Hva er konsekvensene av kjønnsvurderingsgapet for?.....	64
7. Avslutning	68
Referanser	72
Vedlegg 1: Informasjonsskriv	75
Vedlegg 2: Vurdering av behandling av personopplysninger fra NSD.....	79
Vedlegg 3: Intervjuguide.....	81

Sammendrag

Temaet for denne oppgaven er kjønnsulikhet i skolen. Mer konkret er problemstillingen min *kjønnsvurderingsgapet – hvorfor gjør gutter det relativt bedre på eksamen enn jenter?.* Jenter gjør det bedre i skolen uansett vurderingsform, men guttene minker altså forspranget på eksamen når de vurderes av eksterne sensorer. Dette fenomenet kaller jeg *kjønnsvurderingsgapet*. Det som foreligger av forskning på *kjønnsvurderingsgapet* til nå, er kvantitativ og har vektlagt å vise at fenomenet forekommer i flere land. Jeg har derfor valgt en kvalitativ tilnærming, hvor jeg har gjennomført dybdeintervjuer med lærere for bedre å forstå *hvorfor* gutter minker jentenes forsprang på eksamen. Skolens legitimitet som meritokratisk fordeler av muligheter i samfunnet gjennom karakterer, hviler på at alle vurderes likt på faglig grunnlag. Derfor er det viktig å utforske om *kjønnsvurderingsgapet* peker på et brudd med denne universelle vurderingen av elevene.

Jeg trekker ut flere mulige årsaksforklaringer på *kjønnsvurderingsgapet* fra intervjuene mine. Disse analyseres ved hjelp av både tidligere forskning på fenomenet og teoretiske begreper, for å forsøke å holde opp noen forklaringer som mer plausible enn andre. De teoretiske analyseverktøyene jeg benytter meg av er hovedsakelig kausale mekanismer som kan belyse hvorfor gutter gjør det relativt bedre på eksamen enn jentene, over tid og i ulike sammenhenger – eksempelvis matteuseffekten eller selvoppfyllende profetier.

Ut fra datamaterialet mitt ser det ut til at guttene gjør det relativt bedre enn jentene på eksamen fordi jentene får uforholdsmessig gode standpunkt karakterer, sett i forhold til faglig kompetanse. Dette avsløres deretter på eksamen, hvor guttene minker jentenes forsprang fra standpunkt. Særlig trekker jeg fram hvordan lærernes fagvurderinger påvirkes av jentenes gode oppførsel og sosiale ferdigheter. I tillegg viser jeg også hvordan tradisjonell vurderingspraksis som vektlegger puggekunnskap passer jentene bedre enn guttene, og hvordan lærernes vurderinger preges av en kumulativ forutinntatthet, hvor de elevene som gjør det godt tidlig -som oftest jenter- får uforholdsmessig gode vurderinger senere i skoleåret. I tillegg utforsker jeg konsekvensene av *kjønnsvurderingsgapet* for gutters rolle i den meritokratiske skolen. Her drar jeg særlig fram at fenomenet bidrar til å øke de allmenne kjønnsforskjellene i skolen, og argumenterer derigjennom for behovet for mer forskning på området.

Forord

Denne oppgaven er på flere måter kulminasjonen av det jeg har jobbet med de siste årene. Først og fremst fordi den markerer slutten på lærerutdanningen min, men også fordi kjønn i skolen har vært litt min greie gjennom studieløpet. Særlig har akkurat kjønnsvurderingsgapet fascinert meg, fordi det er et relativt ubelyst område innenfor temaet som det ikke foreligger etablerte forklaringer på – og nå har jeg altså fått levert masteroppgave om fenomenet. I den forbindelse er det flere å takke.

Tusen takk til alle informantene mine. Samtlige av dere var reflekterte, åpne og engasjerte i samtalene våre. Livet som lærer er travelt, så jeg setter veldig stor pris på at dere satt dere ned med meg og fortalte om erfaringene deres i klasserommet. Oppgaven hadde aldri blitt til uten dere!

Tusen takk til veileder Gunnar C. Aakvaag for god, givende og morsom veiledning dette året. Det er passende at den som lokket meg til å velge sosiologi på et informasjonsmøte for fem år siden, også ble den som loset meg gjennom masteråret. Du har en unik evne til å sette deg inn i og engasjere deg for det studentene dine kommer med, som har vært uvurderlig i forskningsprosessen. Hele veien har du drevet fram en klarhet i oppgaven gjennom gode råd, påpekninger og spørsmål. Jeg setter stor pris på innsatsen. Virkelig.

En takk må også rettes til Unn-Doris Bæck, som først kanaliserte interessen min for kjønnsforskning inn på kjønnsvurderingsgapet. En forutsetning for at masterskriving blir lystbetont, er jo at en interesserer seg for og er genuint nysgjerrig på det en forsker på. Så takk for tips og veiledning der!

Til sist må jeg takke alle de flotte medstudentene mine som har gjort masteråret til det beste av mange gode år på universitetet. Det har vært en smal sak å møte opp på leshuset hver dag når dere har vært der. Ingen nevnt, ingen glemt. Nå er jeg bare glad og stolt over å være ferdig, og gleder meg til å begynne i jobb som lærer til høsten.

Herman Hovda Lien,

14.05.2023

1.1 Innledning

Temaet for denne masteravhandlingen er ulikhet mellom kjønnene i skolen. Mer spesifikt ønsker jeg se på diskrepansen mellom gutter og jenters oppnådde karakterer på henholdsvis standpunkt og eksamen. Det er en kjensgjerning at jenter gjør det bedre på skolen enn gutter (Krumsvik, 2021, 49); de siste årene har jentenes karaktersnitt i grunnskolen ligget omtrent en halv karakter over guttenes. Samtidig, både i Norge og andre land, minker forskjellen mellom kjønnene markant idet elevene vurderes av eksterne sensorer på eksamen (se f.eks Protivínsky & Münich, 2018 for en kunnskapsoversikt om forskning på fenomenet). For å illustrere med et eksempel: Skoleåret 2018/2019 fikk jenter og gutter på VG1 studiespesialiserende i snitt henholdsvis 4,0 og 3,8 i standpunkt i T-matte, mens eksamenskarakteren på skriftlig eksamen ble 3,3 for begge kjønn (Udir.no). Det er vanlig å gjøre det dårligere på skriftlig eksamen enn på standpunkt for begge kjønn, men guttene gjør det altså *relativt* bedre enn jentene. Hvor mye guttene relativt sett gjør det bedre på eksamen varierer mellom fag, eksamenstype og skolenivå, men trenden er klar og tydelig alt sett under ett.

Grunnen til at jeg ønsker å forske på dette er at utdanningssystemet i Norge i utgangspunktet skal være meritokratisk. Gode resultater gir gode muligheter senere – og motsatt. Av den grunn er det viktig at karakterene som settes i skolen er presise. Det er to typer karakterer i skolen i dag: internt vurderte standpunktkarakter, og eksternt vurderte eksamenskarakterer. Begge vurderingene skal måle elevens beherskelse av læreplanen i faget. Følgelig burde avstanden mellom jentenes og guttenes snittkarakter holde seg stabil mellom standpunkt og eksamen. Slik er det likevel ikke. Når jentene tydelig taper mer på eksamen enn guttene, må en undersøke hva det kommer av. Er det slik at diskrepansen stammer fra et klasserom hvor lærere systematisk gir et kjønn overbetalt eller underbetalt, og at dette deretter synliggjøres på eksamen når den «korrigerer» karakteren elevene har fått på standpunkt? Eller skapes diskrepansen *på* eksamen, gjennom at den passer guttene bedre på et vis? I masteren har jeg valgt å kalle dette fenomenet for kjønnsvurderingsgapet, i tråd med at det ofte benevnes som *the gender grading gap* i eksisterende forskning.

Jeg kommer til å gå eksplorasende til verks i denne oppgaven, som innebærer at jeg har et forskningsdesign som vektlegger å få innsikt om og forståelse for et relativt ubelyst område. Det er nemlig forsket heller lite på kjønnsvurderingsgapet. Det som *er* blitt gjort av arbeid - som jeg skal komme tilbake til - er kvantitativt og har vektlagt å vise at gutter gjør det relativt

bedre på eksamen (Se f.eks Lavy, 2008; Falch & Naper, 2013; Protivínský & München, 2018; Terrier, 2020 eller Casula et al., 2022). Derimot har det blitt skrevet lite om årsakene til fenomenet. Derfor har jeg valgt en problemstilling som supplerer dette hullet i den eksisterende forskningen – *hvorfor gjør guttene det relativt bedre på eksamen?* Jeg har også valgt en mer kvalitativ tilnærming av samme grunn; jeg ønsker bedre å forstå hvordan lærere vurderer og setter karakterer både på standpunkt og eksamen, slik at jeg kan komme til bunns i hva det er som forårsaker kjønnsvurderingsgapet. Da er jeg nødt til å snakke med de som faktisk står i klasserommet. Jeg har derfor gjennomført kvalitative dybdeintervjuer om kjønnsvurderingsgapet med lærere. I analysen min tar jeg utgangspunkt i intervjuene, som jeg belyser og vurderer ved hjelp av tidligere forskning og teoretiske begreper. Gjennom det forsøker jeg å finne noen plausible forklaringer på fenomenet.

Problemstillingen min er satt som *kjønnsvurderingsgapet i skolen – hvorfor gjør gutter det relativt bedre på eksamen enn jenter?* Ut fra problemstillingen har jeg laget tre forskningsspørsmål – ett beskrivende, ett årsaksforklarende, og ett som er rettet mot en konsekvensene av fenomenet:

1. Er det kjønnsforskjeller i prestasjonsavvik mellom standpunkt og eksamen i Norge i dag?
2. Hva er årsakene til kjønnsvurderingsgapet?
3. Hvilke konsekvenser har kjønnsvurderingsgapet for gutters posisjon i den meritokratiske skolen?

I tråd med problemstillingen legger jeg mest vekt på forskningsspørsmålet om fenomenets årsaker.

1.2 Hvorfor er det viktig å forske på dette?

«Retten til likestilling uavhengig av kjønn er en klar rettsregel med et forbud mot direkte og indirekte diskriminering på grunn av kjønn. Å fastholde praksis og tradisjoner i undervisningen som systematisk resulterer i at det ene kjønn kommer dårligere ut enn det annet (resultater som blant annet kan leses av statistikk mange år senere), er en lovstridig handling». (Aune, 2014, 3).

Jurist Helga Aune skrev dette i en artikkel i tidsskriftet *Bedre skole* for noen år siden, og treffer den forskningsbegrunnende spikeren (min) på hodet. Her begir jeg meg ikke inn på noen diskusjon om kjønnsforskjellene i skolen generelt. Det jeg sikter til i denne sammenhengen er snarere at skolens legitimitet som meritokratisk fordeler av muligheter i

samfunnet gjennom karakterer, hviler på at alle vurderes likt på faglig grunnlag. Når karakterstatistikk viser at avstanden mellom guttene og jentene minker markant fra standpunkt til eksamen, kan det se ut til at det ikke er tilfelle. Dette er alarmerende uansett for «feilen» ligger. Dersom eksamen på et vis begunstiger gutter i større grad enn jenter, må en kanskje se på eksamensform eller gjennomføring. På samme vis trengs det også opprydding i vurderingspraksis i klasserommet dersom «[...] Teachers favor female students by consciously or unconsciously “inflating” their scores on non-blind tests» - altså interne karakterer, som er det som gjør opp standpunkt karakteren (Lavy, 2008: 2090).

1.3 Oppgavens struktur

Før jeg går videre, skal jeg kort kommentere tekstforløpet i oppgaven. Først skal jeg redegjøre for forskningen som er gjort på kjønnsvurderingsgapet til nå. Deretter beveger jeg meg videre til det teoretiske grunnlaget for oppgaven, beskriver hvilke begreper jeg skal bruke i analysekapitlet og hvorfor disse er relevante for å svare på problemstillingen min. Forskningens metode blir deretter gjort rede for i kapittel 4. Som jeg kommer tilbake til da, har jeg også samlet inn noen kvantitative data; metodekapitlet er derfor delt i en kortere kvantitativ del og en lengre kvalitativ del. Til sist i metodekapitlet vil jeg også kommentere hvilke forskningsetiske hensyn jeg har tatt, og drøfte metodens troverdighet. Brorparten av oppgavens omfang brukes så til å vise resultatene av intervjuene mine (kap.5), samt analysen av disse (kap.6). I analysekapitlet vurderer jeg forklaringene på kjønnsvurderingsgapet som framgikk i intervjuene, opp mot den tidligere forskningen og begrepene jeg går gjennom i kapittel 2 og 3, og holder fram noen som mer plausible enn andre. I tillegg kommenterer jeg hva konsekvensene av kjønnsvurderingsgapet er for gutters posisjon i den meritokratiske skolen. Til sist samler jeg trådene i oppgaven i en avslutning (kap.7), og svarer på problemstillingen min.

2. Tidligere forskning på kjønnsvurderingsgapet

I det følgende kapitlet skal jeg gi leseren et innblikk i det som er gjort av forskning på kjønnsvurderingsgapet til nå. Et sentralt aspekt av fenomenet som det er viktig å ha klart for seg, er at guttene gjør det *relativt* bedre på eksamen sammenlignet med jentene. Guttene tar altså igjen noe av jentenes forsprang fra standpunkt til eksamen, men jentene gjør det fortsatt bedre i absolutte tall. Siden kjønnsvurderingsgapet peker på at kjønnsforskjellene i skolen kanskje er uforholdsmessig store, er det relevant for leseren å vite noe om hvorfor jenter i det store og hele gjør det bedre på skolen enn gutter. Jeg skal derfor kort si litt om dette før jeg går videre til tidligere forskning på kjønnsvurderingsgapet.

Internasjonale data viser at jentene har hatt bedre karakterer enn guttene - især i språkfagene - i godt over hundre år (NOU, 2019:3, 12). Jentene har med andre ord vært i førersetet i skolen helt siden de ble inkludert (i alle fall på karakterfronten). I dag ligger jentenes karaktersnitt omtrent en halv karakter over guttene i grunnskolen (Krumsvik, 2021, 49). For videregående opplæring på studiespesialiserende linje er forskjellen noe mindre – 0,35 karakterer dersom en samlet ser på de tre siste skoleårene før covid-19 inntraff (Utdanningsdirektoratet, 2023).

Jenter fullfører også videregående i høyere grad enn gutter, og på høyskole - og universitetsnivå er jentene i et klart flertall. Dette er et vanskelig problem å hankses med, all den tid «[...] årsaksfaktorene bak kjønnsforskjellene i skolen er komplekse, mangfoldige og ligger på ulike nivåer» (Wollscheid, 2021, 121). Noen av årsaksforholdene som dras frem i skoleforskning er eksempelvis at jenter modnes raskere enn gutter ((NOU,2019:3, 76-77; Solli, 2021, 215-227; Grøgaard & Arnesen, 2016,42-68), at de sosiale ferdighetene som knyttes til mestring av skolen- slik som evne selvregulering, samarbeid og ansvarlighet- i større grad er å finne hos jenter (Gustavsen, 2018, 16; Krumsvik, 2021, 43-45, Nordahl & Gustavsen, 2021), eller at gutter er mer sårbare for risikofaktorer knyttet til liten grad av skolemestring, slik som lavt utdannede foresatte, dårlig undervisning, mye egenarbeid og svake rammer i klasserommet (NOU, 2019: 3,14, 49; Nordahl, 2011, 75).

Selv om kjønn i skolen er et godt belyst tema, er det heller lite forskning som eksplisitt adresser kjønnsvurderingsgapet. Lavy (2008) er på mange måter foregangsstudien på området. Han skulle i utgangspunktet undersøke om stereotyping og diskriminering av kvinner i arbeidslivet i Israel også var å finne i skolen – og fant det motsatte. Dette undersøkte han ved hjelp av en «difference in differences»-metode (også kalt double

difference i noe litteratur). Israelske skolebarn har to sluttvurderinger i hvert fag: en internt vurdert av faglærer, og en eksternt (og anonymt) vurdert. Begge disse vurderingene skal måle det samme. Tanken i metoden er at den anonyme vurderingen er fri for eventuell påvirkning fra lærer eller klasseromsinteraksjonen. Lavy brukte forskjellen i kjønnsforskjellene mellom den interne og eksterne vurderingen som et mål for potensielle kjønnsbias i lærernes vurderinger.

Jentene gjorde det bedre enn guttene uansett vurderingsform. Lavy fant likevel at forskjellen mellom kjønnene minket markant på den eksterne vurderingen, som han tolket dithen at guttene ble feilaktig vurdert internt av lærerne sine. Videre fant han at «The direction of the bias was replicated [...] in humanities and science subjects alike, at various levels of curriculum of study, among underperforming and best-performing students, in schools where girls outperform boys and in schools where boys outperform girls» (Lavy, 2008, 2103).

Det som ellers er gjort av forskning på kjønnsvurderingsgapet benytter i stor grad samme metode som Lavy og er gjennomgående kvantitativ. Samme funn er gjort i blant annet Tsjekkia (Protivínský & München, 2018,), Norge (Falch & Naper, 2013), Sverige (Berg, Palmgren og Tyrefors, 2022; Lindahl, 2007), Frankrike (Terrier, 2020), Italia (Casula et al., 2022) og USA (Cornwell et al., 2013). Det er med andre ord ganske godt grunnlag for å si at gutter gjør det relativt bedre på eksterne vurderinger som eksamen. Hvorfor det er slik, er imidlertid mer uvisst. De fleste studier på området vektlegger først og fremst å kartlegge om fenomenet er *tilstede* også i deres respektive land. Deretter har de fortrinnsvis sett på hvorvidt forskjellige hypoteser kan styrkes eller svekkes ved å teste om ulike variabler eller utvalg øker eller minker fenomenet. De hyppigst undersøkte årsaksforklaringene er knyttet til forskjellen i kontekst og vurderingspraksis mellom interne og eksterne vurderinger (mellom standpunkt og eksamen, altså), eller til interaksjonen i klasserommet hvor oppførsel, relasjoner og trekk ved lærere og elever får mer spillerom (Protivínský & Munich, 2018, 142-143).

Hva forskjellen mellom interne og eksterne vurderinger gjelder, er det særlig konkurranseintensitet forskerne skriver om. Med det siktes det til at det er mer press og stress før og rundt eksamen enn ellers i semesteret. Dersom jenter er mer utsatt for stress enn gutter, eller guttene «trigges» av konkurranse i større grad, kan det forklare at guttene minker jentenes forsprang på eksamen. I så måte skyldes ikke kjønnsvurderingsgapet upresis vurdering fra lærernes side. Dette en forklaring som flere av studiene gir liten støtte til, dog;

kjønnsvurderingsgapet er blant annet også til stede på lavterskelprøver som nasjonale prøver hvor det er lite trolig at gutter kaver seg opp eller jentene stresser, til den grad at det kan forklare fenomenet (Falch & Naper, 2013,19-21; Lindahl, 2007, 17-18). Videre finner Protivínsky & München ingen forskjell i kjønnsvurderingsgapets omfang mellom elever som rapporter liten grad av stress før og under store prøver, og hele elevpopulasjonen (Protivínsky & München, 2018, 146-147).

Et annet moment som problematiseres i litteraturen er at standpunkt og eksamen kanskje måler ulike ting, og at det er *der* årsaken til fenomenet ligger, heller enn i unøyaktige vurderinger fra lærere (Casula et al., 2022, 484-485). Terrier (2020) tar eksempelvis opp at standpunktkarakterer settes ut fra et bredt grunnlag som også kan vektlegge ferdigheter som ikke går på det kognitive eller faglige, slik som punktlighet eller arbeidsmoral etc., mens dette ikke vurderes på eksamen (Terrier, 2020, 7).

Med lærer-elev relasjonen sikter forskerne til at ikke-kognitive ferdigheter kanskje reflekteres i karakterene elevene gir. Dette problematiseres også av NOU-rapporten om kjønnsforskjeller i den norske skolen fra 2019:

Selv om fravær, orden og atferd i prinsippet ikke skal påvirke den faglige vurderingen av elevene, er det likevel trolig at elevenes karakterer reflekterer mer enn fagkunnskap. For eksempel kan standpunktkarakterer i større grad enn eksamenskarakterer reflektere samhandlingen i klassen, og dermed forskjeller i atferd og sosiale ferdigheter mellom gutter og jenter (NOU, 2019:3, 46)

Casula et.al (2022) viser til at elever som utøver problematisk og forstyrrende atferd oftere er gutter, og at lærerne dermed [...] might sanction/reward these gender non-neutral skills by giving to students lower/higher marks and this would result in a difference between blind and non-blind scores in favour of girls (Casula et al., 2022, 484). Norsk skoleforskning gir støtte til dette forklaringsområdet. Nordahl & Gustavsen (2021, 150-155) viser i sin studie av 20.000 norske skolebarn at det er en sterk korrelasjon mellom kontaktlæreres vurdering av elevenes sosiale ferdigheter, og hvordan de vurderer de faglige ferdighetene deres. Videre rangerer norske lærere jenters sosiale ferdigheter over guttenes (Nordahl og Gustavsen, 2021,150-151; Gustavsen, 2018, 58). Muligens begunstiges altså jentene av å vise «korrekt» oppførsel i større grad enn guttene.

Flere av forskerne undersøker hvordan lærere kan påvirke kjønnsvurderingsgapet ved å se på fenomenets omfang etter variabler som kjønn og erfaring. Her finner de ikke et konsistent

bilde. Tallene «shows a large variation by teachers' characteristics and a pattern that varies from subject to subject» (Lavy, 2008, 2001). Eksempelvis viser Lavy til at variabelen «kvinnelig lærer» dobler kjønnsvurderingsgapet i fysikk kontra mannlige lærere, mens effekten er motsatt i matematikk (Lavy, 2008, 2101) – enda disse fagene er beslektet og undervisningen typisk er ganske lik. Videre finner både Falch & Naper (2013, 24), og Terrier (2020, 15) at uerfarne lærere øker kjønnsvurderingsgapet i matematikk, mens Lavy finner det motsatte (2008, 2101). Selv om det ikke er et ensartet mønster å finne, peker fenomenets variasjon etter læreregenskaper på at noe av forklaringen kan ligge i hvordan lærere oppfører seg, og i interaksjonen mellom lærer og elev. Lavy underbygger dette videre med å se på fordypningsoppgavene til israelske elever, som vurderes av eksterne men *ikke* anonyme sensorer. Kandidaten leverer oppgaven og må deretter delta på en muntlig høring som avgjør karakteren. Han sammenligner resultatene på fordypningsoppgaven med eksamenskarakterer i samme fag. Jentene øker forspranget til guttene, fra den anonyme eksamenen til de ikke-anonyme fordypningsoppgavene (Lavy, 2008, 2098-99). Det tyder altså på at guttenes relative fordel på eksterne vurderinger minker når de ikke vurderes anonymt.

Til sist peker også en del av forskningen på at læreres skjeve vurderingspraksis får utslag utover i skoleløpet. Jentene får muligens et faglig utbytte på tidspunkt B av å bli vurdert fordelaktig på tidspunkt A:

« for two classes where the achievement gap between boys and girls would be identical in grade 6, randomly assigning a teacher who is one standard deviation more biased against boys to one of the classes would decrease boys' relative progress in that class by 0.123 SD in math and by 0.1106 SD in french»(Terrier, 2020, 8)

Det kan med andre ord være snakk om at kjønnsvurderingsgapet ikke bare er uheldig for guttene der og da, men at fenomenet også bidrar til å gi jentene en bedre faglig utvikling over tid.

Forskningen på området er altså tydelig på *at* fenomenet er til stede. Hvorfor guttene gjør det relativt bedre på eksamen er mer i luften. På den ene siden virker det å være liten empirisk støtte for at forklaringen(e) på kjønnsvurderingsgapet ligger på selve eksamen, gjennom at gutter skjerper seg eller jenter stresser. På den andre siden peker litteraturen ellers bare til at forklaringen(e) da ligger i klasserommet og i interaksjonen mellom lærer og elev, uten å klare å være mer spesifikk. Det er med andre ord ingen bestemt forklaring som har vunnet fram.

Som nevnt innledningsvis jeg valgte en problemstilling som går på fenomenets årsaker for å borte på denne mangelen. Videre har jeg valgt en kvalitativ tilnærming av samme grunn; når foreliggende forskning peker på at kjønnsvurderingsgapets årsaker er i klasserommet, er det de som står i klasserommet vi må snakke med for å forstå fenomenet bedre. Dette kommer jeg tilbake til i metodekapitlet. I analysen kommer jeg til å bruke den eksisterende forskningen, og da især den kvantitative empirien, til å belyse og vurdere plausibiliteten til forklaringene på kjønnsvurderingsgapet som kommer fram i intervjuene mine. På samme vis vil jeg også trekke på noen teoretiske analyseverktøy, som er det jeg skal komme nærmere inn på i det følgende kapitlet.

3. Teori

I oppgaven min forsøker jeg altså å forklare at gutter gjør det relativt bedre enn jenter på eksamen enn sammenlignet med standpunkt. Spørsmålet er *hvorfor* og *hvordan*, og det er her teoriaspektet i oppgaven kommer inn. Teorier er, tross alt, «[...] nets cast to catch what we call ‘the world’; to rationalize, to explain, and to master it.» (Popper, 1934, 59, referert til i Joas & Knöbl, 2009, 5). Hvilke begreper kan belyse og forklare hvorfor og hvordan kjønnsvurderingsgapet finner sted? Hvordan passer de med empirien som tidligere forskning og informantene mine kommer med? I de følgende delkapitlene skal jeg gjøre rede for både den vitenskapsteoretiske plasseringen av prosjektet mitt - som har innvirkning på hvordan jeg samler og analyserer data-, og de teoretiske analyseredskapene jeg benytter meg i oppgaven.

3.1 Forskningslogikk og paradigme

«Explicitly or implicitly, researchers usually work within the context of [...] theoretical ideas and ontological and epistemological assumptions.» (Blaikie & Priest, 2019: 9). Det er ikke mulig å forske på den sosiale virkeligheten uten noen grunnleggende antakelser om hva denne består av, eller hvordan man kan innhente kunnskap om den. Her gir forskjellige paradigmer ulike innfallsvinkler for forskeren, noe det er viktig å være bevisst på – ingen kan se på noe uten en synsvinkel. «Every observation made in everyday life, and every statement about it, is already permeated by theory. The same also applies to scientific observations and statements.» (Joas & Knöbl, 2009, 9), som Joas og Knöbl mindre banalt uttrykker denne postempiriske innsikten.

Utgangspunktet for denne oppgaven er det fortolkende paradigmet. Det ontologiske grunnlaget her er at den sosiale virkeligheten skapes gjennom kontinuerlig konstruering og reforhandling aktørene imellom. Dermed er den sosiale virkeligheten allerede fortolket idet forskeren entrer banen (Blaike & Priest, 2017, 103). Dette legger igjen epistemologiske føringer for hvordan forskeren må gå fram: skal forskeren få kunnskap om et felt, er det dermed aktørenes subjektive forståelse eller motivasjon for handlinger etc. som er det primære interessefeltet. Framgangsmåten i det fortolkende paradigmet kan minne om den postempiriske innsikten jeg skrev om i forrige avsnitt, men det er viktig å holde første- og annenhåndshermeneutikk adskilt her: det at forskeren selv må være klar over at observasjonene hennes alltid fortolkes av et teoretisk utgangspunkt, er noe annet enn at hun

som forsker er interessert fortolkningen til *aktørene* i feltet hun ser på. Den fortolkende forskeren -slik som jeg- må ha et reflektert forhold til begge disse aspektene.

Paradigmer forplanter seg gjerne nedover i forskningsprosjekt fordi de ofte legger føringer på andre metodiske valg; eksempelvis er de ofte tettere koblet til noen typer forskningslogikk enn andre (Blaike & Priest, 2017, 13-15). Det fortolkende paradigmet forbindes gjerne med induktiv eller abduktiv forskningslogikk, siden den fortolkende forskeren ofte jobber «nedenfra og opp». En tar utgangspunkt i aktørenes subjektive syn og håndtering av verden; «their construction of reality; their way of conceptualizing and giving meaning to their social world; and their tacit knowledge» (Blaike, 2019, 9). Forskeren abstraherer datamaterialet for å øke forståelsen av studiefeltet, og bruker denne innsikten til å generere sosiologiske forklaringer på sosiale fenomener. Dermed har fremgangsmåten min, hvis analyse tar utgangspunktet i intervjuer med lærere, en del til felles med databasert teoriutvikling – såkalt grounded theory (Glaser & Strauss 1999). Den abduktive forskningslogikken jeg benytter meg av, trekker dog også mye på teori «ovenfra» for å finne de mest plausible forklaringene på kjønnsvurderingsgapet, så fullstendig teoribasert og «grounded» er oppgaven ikke.

Abduksjon kan defineres som «[...] the process of forming an explanation» (Pierce 1934, 171, referert til i Swedberg, 2014,101). Ta en observert regelmessighet, slik som at gutter gjennomgående gjør det relativt bedre på eksamen. Den abduktive forskeren må foreslå noe som kan forklare denne regelmessigheten, ut fra studiefeltet hun observerer; kanskje vet hun at det er flere kvinnelige lærere i skolen enn mannlige, for eksempel, og tenker seg at disse er strengere når de vurderer gutter? «Adopting a hypothesis as being suggested by the facts, is what i call abduction» (Pierce, 1958, 121-122, referert til i Merton, 1987, 3), som Pierce sier. Disse mulige forklaringene blir kontinuerlig moderert og supplert av andre etter hvert som forskningsprosjektet utfolder seg (Swedberg, 2014, 101-110). De holdes med andre ord opp mot og belyses av både teori og empiri, for å finne fram til hvilke av forklaringene som er mest treffende. Den abduktive forskningen beveger seg i så måte hele tiden «[...] mellom teori og empiri. Det teoretiske utgangspunktet justeres etter hvert som empirien samles og datainnsamlingen endres etter hvert som det utvikles nye teorier» (Busch, 2013, 51-52).

Abduksjon er særlig nyttig i eksplorerende oppgaver som min. Selv om jeg gikk inn i prosjektet med noen tanker om både teoretiske perspektiver og mulige forklaringer, er kjønnsvurderingsgapet ganske utrådt terreng. Jeg ønsket derfor å være åpen for nye idéer,

synspunkt og forklaringer underveis i datainnsamlingen, og til slutt stå igjen med noen plausible svar på forskningsspørsmålene mine.

3.2 Begreper som brukes i analysen

Jeg skal i det følgende gjøre rede for de teoretiske begrepene jeg bruker for å belyse resultatene mine i analysekapitlet. For mitt vedkommende er jeg på utkikk etter noe som kan forklare hvorfor kjønnsvurderingsgapet finner sted, over tid og i ulike sammenhenger. Derfor benytter jeg meg primært av begreper som betegner såkalte kausale mekanismer, som Elster definerer som «frequently occurring and easily recognizable causal patterns» (Elster, 2015, 26). Disse kausale mønstrene er tett sammenkoblet med samhandlingsprosesser. Samhandling slik som den i klasserommet styres av meningen vi gir den; derfor er det viktig å koble på begreper som sier noe om hvordan lærere tenker og vurderer, og (derigjennom) hvordan elever og lærere interagerer. Nærmere bestemt blir jeg å belyse resultatene mine med matteuseffekten, selvoppfyllende profetier, klasseromskapital, glorieffekten og ankereffekten.

Matteuseffekten

Som tidligere nevnt er det en kjensgjerning at jenter gjør det bedre på skolen enn gutter. Dette vet også lærere, og det kan da være snakk om at læreres forventninger former vurderingene deres. Derfor benytter jeg meg av *matteuseffekten*, som allment kan forstås som det at de privilegerte i en kontekst gjerne øker forspranget sitt over de ikke-privilegerte (Merton, 1968; Merton, 1988). Eller, for å sitere mannen effekten er oppkalt etter: «For den som har, han skal få, og det i overflod. Men den som ikke har, skal bli fratatt selv det han har» (Matteus 13,12, referert til i Schiefloe, 2019, 249).

«This principle represents a special application of the self-fulfilling prophecy somewhat as follows: (*known scientists*) see fit to report this in print and so it is apt to be important (since, with some consistency, they have made important contributions in the past); since it is probably important, it should be read with special care; and the more attention one gives it, the more one is apt to get out of it. This becomes a self-fulfilling process.» (Merton, 1968, 61-62, min klamme i kursiv)

I sitatet over ser Merton på hvordan kjente forskere får uforholdsmessig mye oppmerksomhet for arbeidet sitt kontra relativt ukjente kolleger, som øker forspranget mellom de store forskerne og resten. Merton (1988) kaller det en kumulativ fordel; de allerede (angivelig) dyktige på et område får arbeidet sitt gjennomgått på et fordelaktig vis og oppnår dermed lettere gode resultater. De gode resultatene vil i sin tur gi videre grobunn for uforholdsmessig

positive tolkninger av dem ved senere anledninger. «(The mathew effect) violates the norm of universalism» (Merton, 1968, 62, min klamme). Effekten er blitt brukt på flere områder, som i økonomien hvor effekten belyser hvordan det er langt lettere å tjene penger hvis du allerede er bemidlet, enn hvis du ikke er det. Det er tenkelig at samme tankegang kan brukes i forbindelse med kjønnsvurderingsgapet. Utdanningssystemet i Norge er meritokratisk, og krever dermed presise og nøyaktige vurderinger av elevers kompetanse – kanskje rokker matteeffekten med universalismen også her? Får de elevene som oppfattes som dyktige av lærerne, en fordelaktig vurdering av at arbeidet deres tolkes med en slik forventning i mente?

Selvoppfyllende profetier

En annen forventningseffekt jeg bruker i oppgaven er *selvoppfyllende profetier*. Forventer vi et godt eller dårlig resultat av noen vi interagerer med, blir det mer sannsynlig at utfallet faktisk blir henholdsvis godt eller dårlig. Her er det i prinsippet snakk om at en gal eller upresis situasjonsdefinisjon preger oppførselen til den som forventer, til den grad at det påvirker utfallet: «If men define situations as real, they are real in their consequences» (Thomas & Thomas, 1928, 572), som det kjente thomasteoremet lyder. Det klassiske eksemplet er den velfungerende banken som går konkurs fordi kundene *tror* at den skal gå konkurs; kundene skynder seg å ta ut pengene sine før banken angivelig kollapser, og forårsaker dermed det de frykter selv (Rosenthal & Jacobsen, 1968, 3-4).

Rosenthal & Jakobsen (1968) viste i sin banebrytende bok på området hvordan selvoppfyllende profetier gjør seg gjeldende på flere andre områder – i spørreundersøkelser, i eksperimentell psykologi, i arbeid med dyr, i helsevesenet og i utdanningssystemet. Essensen i den selvoppfyllende profetien er at dersom den som skal arbeide med eller vurdere noen, går inn i arbeidet med en positiv eller negativ forventning om utfall, kommuniseres denne forventningen til nevnte noen bevisst eller ubevisst. Særlig sistnevnte kommunisering av forventning er vanskelig å sette fingeren på hvordan fungerer:

We know that the process whereby the experimenter communicates his expectancy to his subject is a subtle one. We know that it is subtle because for five years we have tried to find in sound films the unintended cues the experimenter gives the subject-and for five years we have failed, at least partly (Rosenthal & Jacobsen, 1968, 28)

Forventningen som blir kommunisert, preger deretter subjektet slik at forventningen blir en selvoppfyllende profeti. De vi tror blir å gjøre det godt, arbeide hardt, bli frisk etc, gjør oftere

det – og motsatt. Det som er viktig å presisere for tydelig å skille mellom matteuseffekten og selvoppfyllende profetier, er at forventningen gitt på tidspunkt A preger utfallet på tidspunkt B *uavhengig* av hvem som måler det. Det er ikke snakk om en uforholdsmessig positiv vurdering som øker forspranget mellom de begavede og ikke (slik matteuseffekten belyser), men snarere snakk om en reell fordel i det at noen uttrykker positive forventninger til subjektet. Eksempelvis kan en se til forsøk hvor elever, hvis lærere ble fortalt at var særlig begavet på starten av året, hadde større utvikling på IQ-tester enn jevnaldrende kontrollgrupper i samme klasse – og testen ble ikke målt av lærerne selv (Rosenthal, 1968, 82-85). Et av forskningsspørsmålene mine går på konsekvensene av kjønnsvurderingsgapet. Dersom elevenes oppnådde karakterer ses på som en forventning om kompetanse («han er en sekserelev, hun er en toer»), kan de fungere som selvoppfyllende profetier som preger elevens faglige utvikling senere.

Klasseromskapital

Videre benytter jeg meg også av noe av Bourdieus kapitaltypologi. I denne oppgaven er den feltspesifikke kapitalen som er relevant. Med feltspesifikk kapital sikter Bourdieu til ferdigheter, handlingsmønstre og kunnskap som gir anerkjennelse og belønning et gitt område i samfunnet – altså felt (Bourdieu, 1993, 6-9); Aakvaag, 2008: 155). I min sammenheng er det tenkelig at måten undervisning og vurderingspraksis drives på, legger føringer for nettopp hvilke ferdigheter og handlingsmønstre som begunstiger elevene i klasserommet. Jeg bruker derfor *klasseromskapital* som en feltspesifikk betegnelse på de ferdighetene, verdiene, handlingsmønstrene o.l. som gir anerkjennelse og belønning i undervisning og vurdering. Her sikter jeg altså til en kroppsliggjort kapital (Bourdieu 1986, 17-18) – en «know how» om hvordan vi bør oppføre oss. Tanken er at hvis jenter i større grad innehar denne feltspesifikke klasseromskapitalen, kan det potensielt forklare kjønnsvurderingsgapet, fordi jentene da vil ha lettere for å gjøre det godt når elevene vurderes.

Glorieeffekten

Til sist benytter jeg også Kahnemanns (2012) arbeid om hvordan avgjørelser og dømmekraft ofte foregår relativt ubevisst gjennom såkalte biaser (den nærmeste norske oversettelsen blir mentale tommelfingerregler/ tilbøyeligheter). Især benytter jeg meg av glorieeffekten og ankereffekten. Begge disse er i grunn psykologiske mekanismer, men sosiologien bør og må forholde seg til andre disipliner også, slik som psykologi eller biologi, all den tid disse gir

innsikt i forhold som er med på å prege hvordan vi mennesker samhandler med hverandre. I dette tilfellet bruker jeg begrepene fordi klasserommet er en samhandlingsarena. Som nevnt innledningsvis, styres samhandling av den meningen aktørene gir den.

Klasseromsundervisning ville aldri fungert uten at lærere og elever først kom til en felles situasjonsdefinisjon. Læreren kan eksempelvis ikke gi en elev ordet når hen rekker opp hånden, dersom partene ikke tillegger en hevet hånd meningen «jeg vil si noe». Siden jeg skriver om vurdering og interaksjon mellom lærere og elever i klasserommet, er det derfor relevant komme på innsiden av hvordan lærerne tenker, slik at jeg bedre kan forstå hvordan de eksempelvis tolker elevers oppførsel og arbeid, eller setter karakterer.

Glorieeffekten går på at vårt inntrykk av en person på ett bemerkelsesverdig eller viktig område, påvirker hvordan vi bedømmer dem på andre. Sinnet har en tendens til å ønske koherens, og dermed har vi lettere for å overføre et godt eller dårlig inntrykk fra en situasjon til en annen; «good people do only good things, and bad people are all bad» (Kahnemann, 2012, 199-200). Kahnemann viser blant annet til at attraktive forelesere ofte vurderes som faglig dyktige (2012, 4) eller at vi sliter med å tro at Hitler likte hunder og barn (2012, 200). I forbindelse med kjønnsvurderingsgapet, er standpunkt karakterer basert på et (eller flere) års samhandling mellom lærer og elev. Elevene og lærerne tilbringer svært mange timer i lag. Kanskje kan glorieeffekten brukes til å forklare at lærere til en grad blåser opp noen elevers karakterer, gjennom at lærerne ubevisst lar atferd og oppførsel farge vurderingene deres?

Ankereffekten

Ankereffekten, på sin side, handler om at vi blir preget av eventuelle utgangspunkt vi blir gitt når skal estimere en ukjent kvantitet: «If you are asked whether ghandi was more than 114 years old when he died you will end up with a much higher estimate of his age of death than you would if the anchoring question referred to death at 35» (Kahnemann, 2012, 119-120). Et annet eksempel kan være at den som første som foreslår pris under pruting, gjerne legger utgangspunktet for resten av forhandlingen. Det er også mulig å anvende effekten i vurderingsarbeid i skolen:

[...] The first question i scored had a disproportionate effect on the overall grade. The mechanism was simple: if i had given a high score on the first essay, i gave the student the benefit of the doubt whenever i encountered a vague or ambiguous statement later on. Surely a student who had done so well [...] would not make a foolish mistake!» (Kahnemann, 2012, 85).

For skolen er tanken altså at elevens «vanlige» karakter kan fungere som et anker, som senere vurderinger trekkes mot. Kanskje sitter det eksempelvis veldig inne å gi en elev som pleier å få treere, en sekser, fordi spriket blir så stort?

4. Metode

Siden kjønnsvurderingsgapet er relativt nytt som forskningsfelt, har litteraturen til nå fokusert på å vise at fenomenet er reelt. I og for seg har nok dette vært rett tilnærming: «Take care to establish a phenomonon (or a historical event) before proceeding to interpret or explain it» (Merton, 1989, 5), som Merton så kjent poengterte. Dog betyr dette at forskningsfeltet har vært preget av kvantitativ forskning til nå, og mangler den forstående innsikt kvalitativ forskning ofte kan supplere kvantitativ med; *hvorfor* er det slik at guttene gjør det relativt bedre på eksamen enn jentene? Hvilke tanker har aktørene i feltet om dette? Hvilke mekanismer kan de tenke seg? Jeg har derfor gjennomført ustrukturerte dybdeintervjuer med lærere i skolen, og hovedtyngden i oppgaven er derfor kvalitativ. Samtidig var den tilgjengelige kvantitative dataen fra Norge gammel. Forut for den kvalitative datainnsamlingen min - i tråd med Mertons læresetning- hentet jeg derfor også inn karakterstatistikk selv for å etablere at kjønnsvurderingsgapet *fortsatt* finnes i Norge, og for å se om det har økt eller minket i omfang over tid. I det følgende skal jeg derfor først gå gjennom hvordan jeg samlet karakterstatistikk, før jeg forklarer valgene mine rundt de kvalitative intervjuene.

Før den tid, dog, skal jeg minne om det overordnede perspektivet i hvordan jeg har gått fram for å samle og analysere data i denne oppgaven. Som tidligere nevnt i teoridelen, baserer denne eksplorerende oppgaven seg på det fortolkende paradigmet og abduksjon. Jeg leter etter mulige forklaringer på et komplekst fenomen, og i stedet for induksjonens slutning *fra* empiri, eller deduksjonens testing av hypoteser *på* empiri, innebærer en abduktiv tilnærming (blant annet) en kombinasjon av disse innfallsvinklene – såkalt metodisk triangulering. Dette er viktig for meg å presisere, for det har vært rådende for både datainnsamlingen og analysen min. På den ene siden gikk jeg inn i intervjuene mine preget av tidligere forskning på området. Denne har gjort at jeg hadde noen antakelser på forhånd som jeg ønsket å holde opp mot det informantene kom med. På den andre siden ønsket jeg også å bruke intervjuene til å finne forklaringer jeg eller tidligere forskning ikke hadde tenkt på. Oppgaven bærer derfor preg av en kontinuerlig pendling mellom empiri og teori (og mellom induksjon og deduksjon) hvor jeg kobler disse sammen for til slutt å kunne holde opp noen plausible svar på forskningsspørsmålene mine.

4.1 Kvantitativ del

Jeg samlet karakterstatistikk for å kunne svare på det første forskningsspørsmålet mitt, og skal i det følgende gjennomgå hvordan jeg gikk fram. Falch & Naper (2013), som er de eneste som eksplisitt har skrevet om kjønnsvurderingsgapet i Norge, brukte karakterstatistikk fra årene 2002-2005 i sin forskning. Det er litt for mange år siden til at jeg kunne ta utgangspunkt i dataene deres, siden mye kan ha skjedd siden den gang. Videre var karakterene de brukte i sin forskning fra ungdomsskolen, og fra før kunnskapsløftet 06 ble innført. Dette ble avgjørende for hvordan jeg la opp datainnsamlingen min. Jeg ønsket å belyse flere deler av utdanningsløpet til norske elever, så i den forstand passet det godt å sette søkelyset på videregående opplæring i stedet for ungdomsskolen. En kan selvfølgelig innvende at det hadde vært nyttig å samle data fra ungdomsskolen for å se etter en utvikling fra da Falch & Naper forsket på dette i Norge, men i årene 2002-2005 var det læreplanen L97 som la føringer for skolegangen. Skal jeg kunne si noe om utviklingen kjønnsvurderingsgapet har hatt, er det etter min mening mer presist å sammenligne karaktersetning innenfor samme læreplan. Hvis ikke kan det som ser ut som en endring over tid skyldes at læreplanene er ulike, og ikke at eventuelle forklaringer på fenomenet har økt eller sunket i innvirkning. Videre valgte jeg videregående opplæring ikke bare fordi det i så måte var et ubelyst område; jeg mener også at dette var riktig valg siden det er ganske få eksamener på ungdomsskolen. Hver elev har bare én skriftlig og én muntlig. Datagrunnlaget for å belyse det første forskningsspørsmålet mitt er dermed langt større i videregående. Et større antall eksamener er også en av to grunner til at studiespesialiserende linje er valgt over yrkesfaglige linjer innad i videregående opplæring. Den andre grunnen er at yrkesfagene, som Grøgaard og Arnesen (2016) peker på, ofte har ulik vurderingspraksis på de forskjellige linjene. Dermed er yrkesfag utfordrende å arbeide med som en helhet, og studiespesialiserende virket som det beste valget.

Karakterene hentet jeg fra utdanningsdirektoratets nettsider, som samler karakterstatistikk på individnivå direkte fra fylkeskommunenes inntakssystem VIGO (Utdanningsdirektoratet, 2023). Alle elever som får karakter i et fag, telles dermed med. Statistikken viser gjennomsnittskarakterer for hvert fag for både standpunkt og eksamen(er), og gir mulighet til å lese statistikken etter variabler som fylkestilhørighet, kjønn og eierskap (offentlig, privat). Jeg kommer tilbake til hvordan jeg bearbeidet denne statistikken.

Jeg samlet inn data fra to treårsperioder: karakterer fra de skoleårene 2016/17, 2017/18 og 2018/19, og årene 2009/10, 2010/11 og 2011/12. Læreplanen KL06 ble brukt alle de seks årene. To treårsperioder ble valgt slik at et avvikende år ikke skulle bli for utslagsgivende. Fordi noe tidligere forskning pekte på at kjønnsvurderingsgapet minker idet elevene ikke lengre er anonyme for de eksterne sensorene, samlet jeg data for både skriftlig og muntlig eksamen (siden sistnevnte ikke vurderes anonymt). Jeg samlet data fra de samme fagene alle årene for sammenligningen sin del – 30 skriftlige og 26 muntlige. Fagutvalget er basert på størrelse med utgangspunkt i de tre siste årene kronologisk sett: alle fag som hadde over 1000 elever hvorav minst 400 gikk opp til eksamen *alle* disse tre årene, ble tatt med. Grunnen til at jeg satte en grense for hvor mange elever som måtte ta faget og ha eksamen, er at en oftere finner ekstreme resultater i små utvalg (Kahnemann, 2012, 109-118). Et mindre ekskluderende utvalg kunne dermed ha påvirket hvor robuste resultatene mine ble.

Årene som er valgt må også kommenteres. Ideelt sett ønsket jeg å bruke de tre første årene etter at KL06 ble innført, og de tre siste før vi gikk over til den helt nye læreplanen LK20. Dette fordi jeg ønsket å se om kjønnsvurderingsgapet har økt eller minket i omfang over tid. KL06 ble innført i skolen høsten 2006, men var ferdig implementert i videregående opplæring først skoleåret 2008/09, fordi elever får fullføre opplæring med den læreplanen de begynte på skolenivået med. Det var altså først høsten 2008 at alle trinnene på studiespesialiserende linje hadde KL06. Denne læreplanen ble gradvis utskiftet fra skoleåret 2020/21. Dog var skoleåret før, 2019/2020, preget av en avslutning med korona, hjemmeundervisning og avlyst eksamen. Dermed ble tre siste årene med KL06 hvor det ble avholdt eksamen, 2016/17, 2017/18 og 2018/19.

De kronologisk første årene ble tatt fra skoleåret 2009/10 og ikke fra 2008/09, på grunn av at jeg ønsket å samle statistikk fra de samme fagene alle årene. Flere av fagene jeg ønsket å se på var det av flere årsaker ikke mulig å finne statistikk for i 2008/09. Noen fag hadde rett og slett for få elever dette året, bare tall for et av kjønnene eller så var det ikke var tall å oppdrive grunnet omlegging i fagkoding. Dermed skjøv jeg starten for treårsperioden fram et år til 2009/10. Jeg vurderte dette som mer hensiktsmessig enn å tviholde på at jeg ville samle karakterer fra skoleåret 2008/09 – ett års mindre gap mellom treårsperiodene var et mindre tap enn å minke utvalget og derigjennom sammenligningsgrunnlaget.

Hva framgangsmåte gjelder, har jeg brukt den samme difference in differences – metoden som tidligere forskning har, og sammenlignet internt vurderte standpunktkarakterer med eksternt vurderte eksamener. Scoren for kjønnsvurderingsgapet beregnes av å regne ut forskjellen på hvert kjønns gjennomsnittskarakter i et fag på standpunkt og eksamen, og deretter se på diskrepansen mellom kjønnene. For eksempel vil et fag x hvor jentene går fra karakter 5 til 4 fra standpunkt til eksamen, og guttene fra 4.5 til 4, gi en score på 0.5 for kjønnsvurderingsgapet. Beregningen blir $(5-4) - (4,5-4) = 1 - 0.5 = 0.5$. Positiv score vil si at guttene tar innpå på eksamen. Negativ score vil si at jentene øker forspanget. Dette gjorde jeg for alle fagene som passet kriteriene jeg gikk gjennom ovenfor. Jeg regnet i flere omganger, slik at jeg fikk tall for det gjennomsnittlige kjønnsvurderingsgapet for hvert av årene, hver av treårsperiodene, og alle de seks årene regnet sammen. Til arbeidet benyttet jeg Microsoft Excel.

4.2 Kvalitativ del

Komplekse problemstillinger og forskningsspørsmål, slik som mine, trekker ofte oppgaver mot intensive forskningsdesign med lengre intervjuer og få informanter (Busch, 2013, 53). Hovedtyngden i oppgaven ligger derfor på kvalitative dybdeintervjuer med lærere i skolen. Et mer ekstensivt design ville fungert dårlig, all den tid korte intervjuer eller spørreskjemaer ville krevd at jeg i større grad kunne formulere en del forklaringer på kjønnsvurderingsgapet på forhånd, som informantene deretter kunne ta stilling til. Siden jeg verken har (nok av) forklaringene klar på forhånd, og snarere *leter* etter disse i prosjektet mitt, ble et intensivt intervjudesign å foretrekke. De komplekse forskningsspørsmålene mine var også grunnen til at intervju som metode ble valgt ovenfor andre kvalitative alternativer. For eksempel handler ikke oppgaven min om hvordan elever oppfører seg så mye som den handler om hvordan lærerne *opplever* at elevene oppfører seg. Det førstnevnte kunne jeg ha brukt observasjon til, men det samme gjelder ikke lærernes opplevelse og gjengivelse av flere års erfaring i klasserommet. Dermed ble intervjuet det naturlige valget.

Uvalg

Det å bearbeide kvalitative intervjuer kan være veldig tidkrevende. Når jeg i tillegg visste at intervjuene mine kom til å bli ganske lange, valgte jeg å gå for et lite antall informanter (6). Som Brinkmann og Tanggaard sier, er det «[...] bedre å gjennomføre relativt få intervjuer og gjennomanalysere disse» (2012, 21) enn å drukne i et for stort datamateriale. Videre har noe

tidligere forskning pekt på at lærerens kjønn har hatt et visst utslag på kjønnvurderingsgapet (Falch & Naper, 2013; Lavy, 2008). Jeg landet derfor på å ha tre kvinnelige og tre mannlige informanter, for å se om det var en forskjell å spore i hva de tenker og sier om fenomenet.

Utvalget ble gjort strategisk i den forstand at jeg spurte lærere som visste hvem jeg var på skoler jeg har en kjennskap til, om å være informanter. Dette gjorde jeg fordi lærere er en særdeles knapp ressurs for forskningsprosjekter; mange fagdisipliner intervjuer gjerne lærere. Det kunne tatt lang tid å få tak i informanter dersom jeg hadde tatt en mer indirekte vinkling – å be skoleadministrasjon henge opp informasjonsskrivet mitt på lærerværelset, for eksempel, eller ta det opp på møter. Dessuten er det en fare for at de som hadde meldt seg ved en mer indirekte informantsøken hadde hatt en særlig interesse i feltet jeg forsker på. Kvalitative data er ikke generaliserbare, men her kunne en stilt spørsmål ved hvorvidt refleksjonene fra et slikt utvalg er *representative* for den allmenne lærerstanden. Jeg leter jo etter noen stabile mønstre som finnes hos de fleste lærere og/eller elever, som kan forklare hvorfor guttene år etter år gjør det relativt bedre på eksamen. En kan jo også vri på det og si at lærere med særlig interesse for kjønnsforskning i skolen kunne kommet med veldig interessante tanker om kjønnsvurderingsgapet. Likevel vektla jeg ingen andre egenskaper enn kjønn hos informantene mine, fordi jeg ikke ønsket å legge noen føringer for hvilke «typer» lærere jeg snakket med – for eksempel ut fra erfaring eller fagkrets. Derfor valgte jeg å sende ut e-post til litt over 20 lærere på fire skoler, og valgte de tre første av hvert kjønn som svarte. Utvalget var dermed ikke helt uvilkarlig utover kjønn; de lærerne jeg kom på å sende e-post til kan for eksempel ha vært lærere jeg liker godt eller har sett mye i gangene på skolene. Men siden jeg sendte ut til mange flere enn jeg trengte, og konsekvent valgte de første seks som svarte, mener jeg at jeg fant en fin linje mellom det å få tak i lærere raskt, og ikke å legge føringer for hvilke (typer) lærere utvalget besto av.

Dette ble tydeliggjort av at utvalget ble relativt variert gjennom metoden min: tre av lærerne jeg intervjuet er språk - og samfunnsfagslærere, tre er realister, og aldersspennet blant dem er omtrent 40 år. Jeg ble også utover i intervjuene sikrere på at det å ha seks informanter var et godt valg. Selv om lærerne var forskjellige i både fagkrets og alder, merket jeg at metningspunktet nærmet seg når jeg var inne i de siste intervjuene. Lærerne tok på det tidspunktet opp mange av de samme tingene som de foregående informantene hadde. Samtidig var det en fordel for meg å ha intervjuet nok lærere til at jeg kunne se hvilke poeng

og forklaringer som oftere ble tatt opp. Dermed virker det for meg som at utvalget på seks passet ganske godt til oppgavens mål og (tids)omfang.

De potensielle informantene fikk en mail hvor jeg forklarte hva jeg forsket på, og et informasjonsskriv (se vedlegg 1) med videre detaljer og informasjon om hva deltakelsen deres betød. Jeg søkte parallelt med dette om tillatelse til å samle dataene hos NSD, og gikk i gang med intervjuene først etter at de godkjente prosjektet mitt. NSD gav grønt lys i midten av desember (se vedlegg 2), og intervjuene foregikk derfor i januar 2023. I forbindelse med intervjutidspunktet ser jeg to metodiske svakheter jeg ønsker adressere før jeg går videre.

For det første mener jeg at det ideelle tidspunktet å snakke med lærerne på ville vært i slutten av skoleåret, fordi lærerne da både har hatt et år på å bli kjent med elevene sine, akkurat har satt standpunkt og gjennomført eksamener. En kan selvfølgelig innvende at lærerne i intervjuene snakket ut fra flere års observasjoner og erfaring, og at intervjutidspunktet slikt sett ikke har så mye å si. Likevel mener jeg at det ville vært gunstigere å snakke med lærere i juni, all den tid vi gjerne husker den siste tiden best. Flere av lærerne i intervjuene måtte eksempelvis tenke en stund før de kom på hvordan eksamenstrekk og lignende fungerte fordi pandemien har gjort at det ikke har vært eksamen på noen år. Det gikk av praktiske hensyn dessverre ikke an i min oppgave, siden masteren leveres i mai. For det andre er det en svakhet at jeg ikke har fått snakket med lærere på flere ulike tidspunkt; forskningen min er på det området synkron. Den kvantitative datainnsamlingen min er derimot diakron, og måler utviklingen av kjønnsvurderingsgapet over tid. Det hadde vært interessant å se om lærernes vurderingspraksis og oppfatning av jenter og gutter, for eksempel, har endret seg over tid. Hvordan ville de kvalitative og kvantitative dataene passet sammen i så tilfelle? Dette hadde jeg ikke anledning til å i denne oppgaven, og er noe eventuelle andre forskere kan ha i mente dersom de ønsker å bygge videre på arbeidet mitt.

Formål og intervjuguide

Hovedpoenget med intervjuene var å få refleksjoner rundt kjønnsvurderingsgapet – hva kunne være årsakene til at guttene minker jentenes forsprang når de vurderes på eksamen?

Intervjuene var delvis strukturerte. Jeg lagde en intervjuguide (se vedlegg 3) som fungerte som en felles ramme slik at informantene og jeg var innom omtrent de samme hovedpunktene i de forskjellige intervjuene. Vi snakket om hvordan gutter og jenter oppfører seg ulikt i klasserommet, hva som pleier skje når eksamen nærmer seg, og om hvordan lærerne vanligvis

går fram for å sette standpunktkarakterer. I tillegg spurte jeg konkret om lærerne kunne tenke seg noen forklaringer på kjønnsvurderingsgapet, og om hva lærere kan gjøre for å være mest mulig presis i vurderingsarbeidet sitt og derigjennom minke fenomenets omfang. Spørsmålene i intervjuguiden delte jeg i fem deler: «forklaringer på fenomenet», «oppførsel og vurdering», «om eksamen», «fenomenet i skolen» og «presise vurderinger». Dette ble gjort fordi jeg vurderte det som mer oversiktlig å gjennomføre den senere analysen av intervjuene ut fra fem deltemaer heller enn å måtte analysere og kode hvert enkelt spørsmål.

Samtidig forholdt jeg meg ikke altfor rigid til intervjuguiden, fordi jeg som nevnt la opp til en abduktiv forskningsprosess. For det første ville jeg derfor ikke bare diskutere det jeg på forhånd ønsket å ta opp, og lot derfor samtalene få leve sitt eget liv med digresjoner og diskusjoner informantene og jeg kom på underveis. For det andre var intervjuene kumulative i den forstand at jeg kontinuerlig tok med meg det informantene sa videre; eksempelvis kunne én informant ta opp en interessant vinkling, som jeg deretter dro inn i de neste intervjuene for å se om dette var noe flere lærere så. Det jeg på forhånd tenkte ta opp med informantene, eller hvordan jeg formulerte spørsmål, endret seg altså noe underveis.

Gjennomføring

Intervjuene ble gjennomført på informantenes respektive skoler. Jeg ba informantene om å booke oss et gruppe - eller møterom, slik at vi kunne prate i ro og fred. Som regel hadde informantene ordnet kaffe til oss på forhånd, mens jeg hadde med noe vi kunne tygge på underveis. Dette var bevisst foreslått fra min side, slik at intervjuet skulle bli så lite formelt og stivt som mulig. Siden jeg kjente til de fleste informantene fra før av, snakket vi en del løst og fast om hvordan det gikk med hverandre og slikt før jeg bevegde samtalen inn på selve intervjuet. På forhånd hadde jeg varslet informantene om at jeg kom til å ta opp intervjuet. Dette gjorde jeg via Nettskjema.no. Alle UiTs studenter har tilgang til nettskjemaet via FEIDE-innlogging, som er den anbefalte måten å samle og lagre data til masterprosjekter. Nettskjema har en tilknyttet diktafonapplikasjon som du kan laste ned til smarttelefonen din, som sikkert sender opptakene til nettskjemaet ditt. Før jeg startet opptakene, spurte jeg alltid informantene om det var noe de lurte på om prosjektet og om deltakelsen deres. Etter opptaksstart innledet jeg som regel intervjuet med å la informantene få fortelle litt om seg selv som lærer – hvilke fag de underviser i, hvor lenge de hadde vært lærere og slikt. Selv om det gav undertegnede en del å transkribere som ikke var direkte relevant for oppgaven, opplevde

jeg at det fungerte som en god start på en *samtale* heller enn et formelt intervju med høye skuldre. Deretter rekapitulerte jeg kjønnsvurderingsgapet kort, og spurte om informantenes umiddelbare tanker til det jeg fortalte dem. Utover det diskuterte informantene og jeg fenomenet i den retningen samtalen tok oss. Intervjuguiden ble brukt som en rettesnor, men som tidligere nevnt var jeg ikke særlig rigid her, og lot i stor grad intervjuet leve sitt eget liv så lenge vi var innom punktene i guiden på et eller annet tidspunkt. Det korteste intervjuet varte i 50 minutter, og det lengste i halvannen time.

Noe som bør problematiseres i intervjusammenheng er at informantene mine kjente til meg på forhånd. Forskningen min er tross alt inne på noe som kan være ømt å prate om for lærere. Intervjuer som omhandler sensitive temaer blir av den grunn noen ganger gjennomført digitalt og/eller anonymt (Brinkmann & Tanggaard, 2012, 24). Kanskje ville lærerne turt være mer åpne med noen de ikke kjente eller visste de ikke kom til å møte flere ganger, eller hvis jeg gjennomførte intervjuene med en skjerm mellom oss? Samtidig vil jeg argumentere for at det tjente prosjektet at jeg var en aspirerende lærer som informantene kjente til. Det å posisjonere seg som en likesinnet kan jo ofte bidra til at intervjuer blir mer åpne og avslappet (Aase & Fossåskaret, 2018, 124). Jeg brukte dette aktivt i samtale, og fikk gjennomgående et inntrykk av at informantene opplevde at de og jeg diskuterte kjønnsvurderingsgapet *sammen som lærere*. Dermed gav de meg den tilliten som kreves for å diskutere profesjonsutfordringer – som det å sette karakterer, eller å snakke om forskjeller mellom elevgrupper i klassene deres. Det at jeg allerede var kjent med feltet jeg forsket i, gjorde også at de slapp å bruke tid på å sette meg inn i lærerdiskursen, eller forklare hva de mente med utsagn om nye læreplaner eller lignende. Som IP6 utbrøt da vi diskuterte hvor vanskelig det kan være å sette karakterer: «Så her *sitter vi som lærere* og lider frivillig hæ?» (IP6, min kursivering).

Bearbeiding av data

Bearbeidingen av intervjuene ble gjort i flere trinn. Aller først transkriberte jeg alle seks intervjuene. Siden det tar tid å transkribere, valgte jeg en relativt enkel transkripsjonsmetode inspirert av Brinkmann og Tanggaard (2012, 35): Jeg brukte initialene IPX for intervjuperson x og F for forsker, og byttet linje hver gang en ny person begynte å prate. De gangene informantene og jeg pratet nesten samtidig, benyttet jeg meg av ... for å vise at setningene gled over i hverandre. Videre brukte jeg BLOKKBOKSTAVER for å markere roping eller

særlig trykk på ord. I tillegg benyttet jeg parenteser for å vise til ulike forhold som ikke kommer fram i skriftlig form, slik som smiling, latter, pauser, spesielle stemninger etc.

Etter dette hørte jeg gjennom alle intervjuene mens jeg leste over transkripsjonene for å luke ut feil. Selve analysen begynte jeg med å lese gjennom transkripsjonene i papirform og kondensere meningsinnholdet for meg selv. Deretter gikk jeg mer systematisk til verks og analyserte hvert intervju skjematisk. Som tidligere nevnt var spørsmålene i intervjuguiden var delt i fem deler. Det informantene snakket om utenfor guiden lot seg også plassere inn i en eller flere av disse undertemaene, og de fem delene ble dermed utgangspunkt for den formelle analysen. For hver del trakk jeg fram særlig relevante sitater, deretter meningskondenserte jeg, før jeg også forkortet og kategoriserte meningskondenseringen til nøkkelsetninger. Det sistnevnte ble gjort slik at jeg lettere kunne sammenligne hva de ulike informantene fortalte. Nedenfor følger et utdrag fra analysen av IP2 sitt intervju:

Utdrag 1 – eksempel på analyse

IP2, mann, samfunnsfag			
Spørsmål/tema	Transkripsjon/særlig gode sitat	Meningskondensering	Kategorisering
Forklaringer på fenomenet	<p>«Det jeg kanskje umiddelbart tenkte er at guttene har kanskje i sterkere grad enn jenter den her skippertakmentaliteten. Når det VIRKELIG gjelder, og i en kort periode, så hiver de seg til og jobber kolossalt» (s.3)</p> <p>«Sånn generelt vil jeg si at jentene jobber vel mer jevnt enn gutter [...] Og det kan jo også hende, at vi, ubevisst, bevisst, ubevisst kanskje, belønner, SER at de gjør mer innsats i det de gjør, at vi på en måte dermed bedømmer dem litt bedre? Kanskje?» (s.4)</p>	<p>Først og fremst attribueres anonymitetsfenomenet til at guttene hiver seg voldsomt rundt. De tar igjen jentene litt på eksamen, fordi de girer sånn opp.</p> <p>I tillegg snakker IP2 om at jentene jobber jevnt og trutt, og at lærere kanskje ubevisst eller bevisst belønner dette.</p>	<p>Gutter skjærer seg på eksamen</p> <p>Jenter <i>begunstiges</i> av rett oppførsel i klasserommet (arbeidsinnsats)</p>
Oppførsel og vurdering	<p><i>Oppførsel</i></p> <p>«I løpet av året virker det som at jenter har rutiner for hvordan de jobber og sånn at, at de har, kanskje, er bedre å planlegge ukene sine og er jevnt mer MED i det som foregår.» s.4</p>	<p><i>Oppførsel</i></p> <p>Jentene er mer med og mer flittig. De gjør seg også mer klar for enkeltvurderinger underveis i året. Guttene slapper mer av.</p> <p>Sekserelever er jevnt over elever som deltar</p>	<p><i>Oppførsel</i></p> <p>Jentene mer flittig og er mer serios. Gutter mer avslappet.</p> <p>Sekserelever er elever som deltar.</p> <p><i>Karaktersetting</i></p>

Slik gikk jeg gjennom hvert intervju. Deretter satte jeg opp en tabell for å sammenligne det informantene hadde sagt. Leseren kan se et glimt av sammenligningstabellen nedenfor i utdrag 2. Den observante leser legger merke til at det som står under «kategorisering» hos IP2 i utdrag 1, er å finne igjen i rutene for Del 1 – IP2 og Del 2 – IP2. Tabellen er med andre ord en samling av den forenklete kategoriseringen av det informantene sa fra hver intervjudel.

Utdrag 2 – sammenligning av kategorisering

Analyse av intervjuer						
Spørsmålene er delt i fem deler. De er fargekodet. Spørsmål eller temaer som handler om disse vil være kodet. Øverst ligger en forenklet oversikt over hva alle informantene sier. Lengre ned er en egen tabell for hver informant med en inndeling i sitat – meningskondensering – forenkling.						
<ul style="list-style-type: none"> - Del 1: Forklaringer på fenomenet - Del 2: Oppførsel og vurdering - Del 3: om eksamen - Del 4: Fenomenet som problem i skolen - Del 5: Presise vurderinger 						
	IP1	IP2	IP3	IP4	IP5	IP6
Del 1	Gutter skjærer seg på eksamen Jenter begunstiges av tradisjonell vurderingspraksis i klasserommet (Måte skriftlig) Vippekarakterer skaper diskrepansen	Gutter skjærer seg på eksamen Jenter begunstiges av rett oppførsel i klasserommet (arbeidsinnsats)	Jenter begunstiges av tradisjonell vurderingspraksis i klasserommet. (i KL06, KL KL20 må)	Jenter begunstiges av lærernes pedagogiske vurderingspraksis i større grad enn guttene? Jenter begunstiges av at de mestrer skolens kode (oppførsel)	Jenter begunstiges av tradisjonell vurderingspraksis (kaver seg oppfor vurderinger) Jenter begunstiges av rett oppførsel (pliktoppfyllende) Dårlege lærere gir rom for upresise karakterer	Jenter begunstiges fordi de knækker skolens kode (forstår og passer rammene) Jenter begunstiges av tradisjonell vurderingspraksis Diskrepansen skapes av vippekarakterer. Jentene stresser mer på- og rundt eksamen.
Del 2	Oppførsel Jentene mer flink og stresset. Karakterpress. Gutter har ansvar og kontroll. Mer praktisk av seg. Karaktersetting Kombinasjon av formelle vurderinger og innsats i klasserommet. Å sette karakterer er et stressende ansvar for læreren.	Oppførsel Jentene mer flittig og er mer seriøs. Gutter mer avslappet. Sekserelever er elever som deltar. Karaktersetting Kombinasjon av formelle vurderinger og deltakelse i klasserommet.	Oppførsel Jenter mer pliktoppfyllende og stresset. Elever som er ambisiøse får mer oppmerksomhet av læreren Karaktersetting Prosess i fokus. Endings- og revisjonskompetanse, innsats, belønnes, ikke rene svar.	Oppførsel Gutter passer ikke inn i dagens skole. Gutter og jenter behandles ulikt. Jentene får mer slakk enn guttene. Det er mer krevende å irettesette jenter (sensitiv). Lærer må tilpasse seg forventet motreaksjon... Gutter stoler mer på seg selv. Mer rolig. Jenter stoler mindre på seg selv, stresset.	Oppførsel Jentene er mer moden og stresset. Gutter blir oftere irettesatt av lærer enn jentene Sekserelever assosieres ofte til å være jenter, mens toere er gutter. Karaktersetting Kombinasjon av formelle vurderinger og innsats og bidrag i timene.	Oppførsel Jentene er mer pliktoppfyllende og stresset. Gutter og jenter behandles ulikt. Strengere med gutter, ettergivende med jenter. Karaktersetting Kombinasjon av innsats i timene og formelle vurderinger. Formelle vurderinger ofte urettferdig mål av

Dette var til stor hjelp for å få oversikt over datamaterialet, fordi nøkkelsetningene lot meg trekke linjer mellom intervjuene. Særlig var det nyttig å se hvilke poeng som ble tatt opp oftest. Når flere informanter uavhengig av hverandre forteller det samme -muligens med andre ord-, veier dette naturligvis tyngre enn det eksempelvis bare én informant mener.

4.3 Forskningsetiske hensyn

I denne oppgaven har det vært flere etiske utfordringer å hankses med. Først og fremst har disse vært relatert til anonymisering, samtykke og tolkning. Jeg skal i det følgende kort redegjøre for disse, og vise hvordan jeg har forsøkt imøtegå dem.

Det er nærliggende å tro at de som leser masteroppgaven til en student ved Universitetet i Tromsø, antar at datainnsamlingen er Troms og Finnmark. Det var derfor ekstra viktig for meg å ivareta anonymiteten til informantene, noe som videre blir understreket av at intervjuene omhandlet et potensielt sårt tema. Av den grunn transkriberte jeg intervjuene til bokmål slik at ingen informanter kunne bli identifisert av en særskilt dialekt. Videre fikk informantene kodenavn av typen intervjuperson x (IPX) uavhengig av den kronologiske rekkefølgen på intervjuene (Thagaard, 2018, 115). Dette snakket jeg også om med informantene på forhånd, slik at de kunne være sikre på at anonymiteten deres ble ivaretatt. I tillegg har jeg vært forsiktig med sitater og andre gjengivelser av det informantene sa – det er alltid en fare for at andre egenskaper eller trekk ved det informantene sier, utover (steds)navn, kan gjenkjennes av personer i samme miljø (Aase & Fossåskaret, 2014, 213). Kanskje kjenner en eventuell sensor til en lærer i området som er veldig glad i Tromsø IL, for eksempel. Dersom informanten kontinuerlig bruker sportsmetaforer i intervjuene, og jeg ukritisk bruker disse uten omskriving, kan faktisk noe såpass banalt være med på å bryte konfidensialiteten mellom informanten og meg. Videre forsøkte jeg være helt tydelig i informasjonsskrivet mitt - og i den innledende korrespondansen på E-post - på hva jeg forsker på og hvorfor informantene ble spurt (Thagaard, 2018, 86). Deltakerne fikk også vite at de kunne trekke seg - og da også de dataene de har bidratt med - fra forskningen når som helst gjennom prosessen fram til levering. Samtlige informanter ble også tilbudt å bli tilsendt lydfil og transskripsjon fra intervjuet sitt, for å se om det var noe de ønsket å fjerne eller kommentere. Ingen tok i på dette. Det kan selvfølgelig være symptomatisk for at lærere er travle mennesker, men det kan også tolkes dithen at informantene følte seg trygge på at jeg forvaltet dataene deres godt.

En utfordring i forhold til informert samtykke var at informantene i utgangspunktet bare samtykket til å bidra med data i intervjuer, og ikke nødvendigvis til den tolkningen jeg har gjort i oppgaven (Aase & Fossåskaret, 2014: 217-219). Dette spennet mellom «gjenfortellingen» og «diagnosen», som Aase & Fossåskaret treffende beskriver det, kan være sår for deltakerne. Jeg har muligens sett noen linjer som de ikke selv ser, eller analysert det de

sier i en retning de ikke er enig i. Det har derfor vært viktig for meg å argumentere godt for tolkningene mine; jeg som forsker skal ikke «[...] tillegge mennesker irrasjonelle eller lite aktverdige motiver uten å kunne gi overbevisende dokumentasjon og begrunnelse» (NESH, 2016). Dette er noe av grunnen til at jeg ønsket delvis strukturerte intervjuer, slik at jeg kunne følge opp deltakerne så – «Forstår jeg deg rett når [...]» er eksempelvis et hyppig spørsmål i intervjuene. På den måten klarte jeg kanskje bedre å forstå hva informantene mente, samtidig som jeg også gav informantene et innsyn i hvordan jeg fortløpende tolket data, og derigjennom en sjanse til å korrigere meg der det trengtes.

4.4 Metodens troverdighet

Avslutningsvis i dette kapitlet skal jeg skal skrive om hvordan jeg har forsøkt å styrke troverdigheten til metoden min. I tillegg skal jeg kort kommentere i hvilken grad metoden gjør oppgavens funn overførbare.

I kvalitative studier er redegjørelsen for fremgangsmåter og metodisk refleksjon et viktig grunnlag for reliabilitet, og begrunnelser for tolkninger et viktig grunnlag for validitet (Thagaard, 2018, 188-193). Med unntak av den korte og relativt ukompliserte karakterstatistikken jeg har samlet inn, er det meste av datainnsamlingen min kvalitativ. I så måte er det kanskje mer treffende å snakke om prosjektets troverdighet i stedet for reliabilitet, siden sistnevnte i utgangspunktet viser til en generell repliserbarhet som det ikke er mulig å ha i en kvalitativ forskning som min (Thagaard, 2018, 188; Ringdal, 2018, 247). Først og fremst har jeg forsøkt å styrke prosjektets troverdighet ved å skrive en gjennomiktig methodedel. Jeg har etterstrebet å gi grundig argumentasjon for både utvalg, hvordan jeg la opp og gjennomførte intervjuene, og deretter bearbeidet dem. I tillegg har jeg lagt stor vekt på at det er informantene som skal «bli hørt» i resultatkapitlet – primærdataen bør (først) bli presentert i fred fra forskerens tolkninger (Thagaard, 2018, 188). Tankegangen her er at den kritiske leser kan se på resultatene mine, og deretter vurdere om hun synes at tolkningene av dem i analysedelen er treffende. Videre har jeg også fortalt at jeg hadde kjennskap til informantene og feltet på forhånd, og forsøkt å tydeliggjøre hvordan det kan ha påvirket datainnsamlingen.

Validiteten i oppgaven har jeg forsøkt å ivareta gjennom å være tydelig på grunnlaget for tolkningene mine. Analysedelen av denne abduktivt-eksplorerende oppgaven preges av at jeg siler de ulike forklaringene på kjønnsvurderingsgapet som kommer fram i intervjuene, gjennom både empiri og teori, for å finne de som virker å ha størst forklaringsevne. Empirien

fra informantene holdes altså opp mot og tolkes ved hjelp tidligere forskning og teori. Utgangspunktet for denne tolkningen har jeg forsøkt å presentere tydelig i kapittel 2 og 3, hvor jeg både redegjorde for tidligere forskning på området og presenterte det teoretiske utgangspunktet mitt, samt begrepene jeg bruker i analysen. Videre etterstreber jeg i analysekapitlet kontinuerlig å vise leseren *hvordan* jeg analyserer og tolker, gjerne i form av å tydeliggjøre hvorfor «[...] alternative tolkninger er mindre relevante» (Thagaard, 2018, 189).

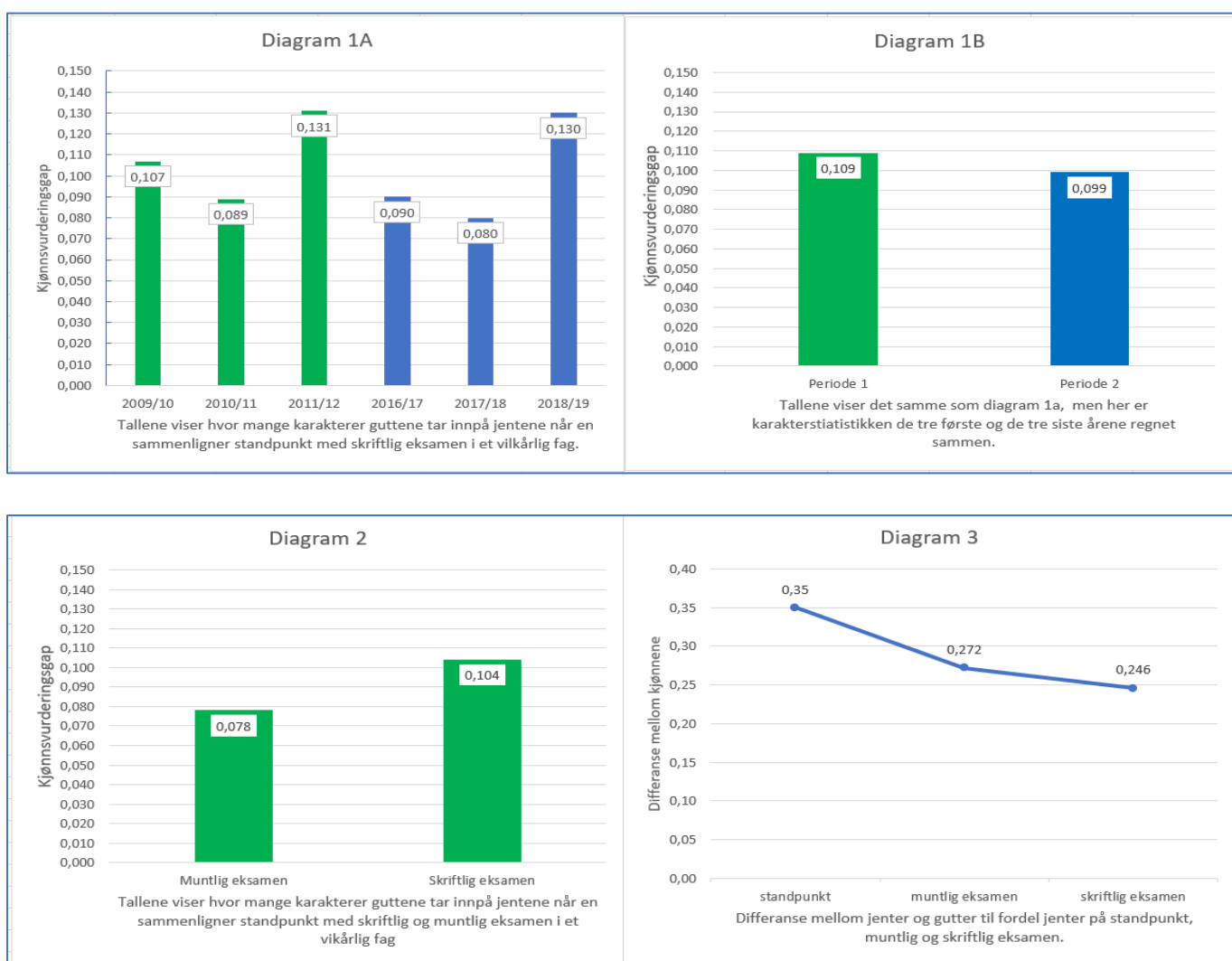
Hva overførbarhet gjelder, har jeg (bare) snakket med seks lærere. Som nevnt tidligere, var dette ganske forskjellige lærertyper, og intervjuene nærmet seg et metningspunkt i de to siste intervjuene. Samtidig, som jeg kommer tilbake til avslutningsvis i oppgaven, vokter jeg meg for å presentere forskningen min som representativt for hva lærere tenker om dette fenomenet siden utvalget er beskjedent. Jeg vil heller si at forskningen min er et utgangspunkt å bygge videre på. Seks lærere er tross alt bare seks lærere, og jeg ønsker å understreke behovet for å snakke med flere lærere i ytterligere kvalitativ forskning på kjønnsvurderingsgapet, slik at en kan belyse årsakene til fenomenet enda bedre.

5. Resultater

I de følgende to delkapitlene blir jeg å presentere resultatene fra datainnsamlingen min. Jeg viser de kvantitative resultatene først, fordi de danner et viktig bakteppe for de kvalitative. Dette er også i tråd med kronologien i forskningsspørsmålene mine; jeg ønsker først å vise kjønnsvurderingsgapets omfang i skolen, før jeg går videre til mulige forklaringer på det.

5.1 Presentasjon av kvantitative data

Som tidligere nevnt, er kjønnsvurderingsgapet et uttrykk for hvor mye guttene tar inn jentenes forsprang i karakter, når en sammenligner standpunkt med eksamen. Nedenfor har jeg presentert karakterstatistikken min i fire diagram.



Det første sentrale funnet å vise til i karakterstatistikken er at kjønnsvurderingsgapet på skriftlig eksamen er omtrent 0.1 karakter. Diagram 1A og 1B viser hvor stort

kjønnsvurderingsgapet var i årene jeg samlet karakterstatistikk for. Det er ingen større forskjell å snakke om i utvikling; treårsperioden rett etter at KL06 ble innført har et kjønnsvurderingsgap på 0.11, mens tilsvarende tall for treårsperioden før pandemien er 0.10. Dermed tyder det på at årsaken(e) til fenomenet har vært til stede i skolen over noe tid. Regnes alle årene sammen, er kjønnsvurderingsgapet på skriftlig eksamen 0.104 karakterer.

Det andre sentrale funnet er at kjønnsvurderingsgapet er større på skriftlig eksamen enn på muntlig, slik vist i diagram 2. Gapet er stabilt også på muntlig eksamen, og ligger på 0.078 karakterer på muntlig hvis en regner alle de seks årene samlet. Diagram 3 viser hvor stor differansen er mellom gutter og jenters snittkarakter på henholdsvis standpunkt, muntlig eksamen og skriftlig eksamen.

Til slutt er det også viktig å presisere at karakterstatistikken spriker veldig. For tallene for skriftlig eksamen er det slik at halvparten av årene jeg har samlet data for, har standardavvik som er større enn gjennomsnittet for kjønnsvurderingsgapet. De fleste årene har for eksempel gjerne noen fag hvor kjønnsvurderingsgapet er 0,3 eller 0,4 karakterer, samtidig som noen få andre fag kan ha 0 eller til og med negativ score (altså at jentene gjør det relativt bedre kontra gutter på eksamen, sammenlignet med standpunkt). Det er disse sprikene som skaper et så stort standardavvik. Dette er ekstremer, dog; typetallet for kjønnsvurderingsgapet er 0.2 for fire av årene jeg samlet statistikk for, og 0,1 på de to andre. Og som diagram 1a og 1b viser, holder gjennomsnittlig kjønnsvurderingsgap seg relativt stabilt. Dermed er det en klar trend å se i tallene; guttene gjør det (fortsatt) jevnt over relativt bedre enn jentene på eksamen sammenlignet med standpunkt.

5.2 Presentasjon av kvalitative data

I intervjuguiden min delte jeg spørsmålene inn i fem deler. Jeg blir derfor å presentere funnene etter disse, i hvert sitt delkapittel, og oppsummerer det viktigste å trekke ut fra hver intervjudel før jeg går videre til den neste. Delene presenteres ikke i samme rekkefølge som i intervjuguiden av narrative hensyn. Først skriver jeg om hva lærerne sa om oppførsel, vurdering og eksamen, før jeg etterpå går over informantenes forslag til hva som kan være årsakene til fenomenet, og hva skolen og lærere kan gjøre å minke fenomenets omfang. Underveis forsøker jeg å være tydelig på hvor mange av informantene som tok opp de ulike poengene, for ikke å være misvisende.

5.2.1 Oppførsel og vurdering

I denne delen av intervjuene fortalte informantene om hvordan elever oppfører seg i klasserommet og på vurderinger, med vekt på om de så noen forskjeller mellom kjønnene. Jevnt over har jeg sortert dataene i to underkategorier: hvordan elevene oppfører seg, og hvordan lærerne vurderer og setter karakter.

Oppførsel

Sitatet «Jentene stresser mer, de kaver mer, de skal være flinke og de skal ha karakter» (IP3), er på mange vis representativt for det informantene sier om jentenes oppførsel. I det store og hele oppfattes jentene som mer skjerpet og aktive. «I løpet av året virker det som at jentene har rutiner for hvordan de jobber [...] og er jevnt mer MED i det som foregår.», som IP2 formulerer det. De kommer seg fortere i gang og er mer lydhøre for beskjeder. Alle informantene er også tydelige på at jentene stresser mer enn guttene. De har i større grad kontroll på når det er vurderinger, og forbereder seg mer til disse enn guttene. IP5 forteller om et typisk eksempel: «Jeg så jo i dag når jeg satt meg med en gutt [...] og så sier jeg husk prøve på torsdag og da sier han ER DET PRØVE PÅ TORSDAG? (Humrer)». Samtidig er lærerne klare på at det ikke bare er positive sider ved at jentene tar fagene mer seriøst. De er ansente, blir lei seg når vurderingene ikke går slik de ønsker, og er veldig opptatte av å prestere:

«Jeg synes at trenden er guttene er litt sånn du nå har jeg fem ting på én gang jeg må få utsettelse på den her (humrer), altså at de kanskje (pause) har lettere for å prioritere bort noe. Mens jentene (sukker), jentene, de vil så gjerne få til alt samtidig, og derfor så, det er liksom de godtar ikke at de skal ikke gjøre noe» (IP6)

Guttene beskrives derimot som litt mer tilbaketrukket underveis i året. Informantene utbroderer dette på litt forskjellige vis, dog. IP1, IP2 og IP3 forteller at guttene er mer avslappet. IP3 prater en del om at det ikke nødvendigvis gjør at guttene får med seg mindre av *essensen* i faget selv om de ikke kaver veldig med pugging; «Tradisjonelle prøver er kunnskap man egentlig ikke klarer å ta med seg videre. Det er noe man har der og da [...] Det du husker best er det du GJØR i klasserommet gjennom året» (IP3).

IP4, på sin side, forteller at guttene er mer sikre på seg selv. De bruker mindre tid på å se over arbeidet sitt. Jentene bruker derimot mye tid på finpussing: «Jeg synes VELDIG ofte at jentene må liksom hele tiden ha den her bekræftelsen på altså, utenifra ser det ut som at de er litt mindre selvstendige fordi de vil på en måte ha fasit.» (IP4).

Et siste interessant moment om oppførsel er at både IP4 og IP6 tar opp og problematiserer at gutter og jenter møter ulike forventninger og dermed behandles ulikt i skolen.

«Men han (mannlig elev på annen skole som lærer diskuterte dette med) har liksom snakket om [...] at jentene kan sitte å bråke og de blir ikke snakket til, men hvis at to gutter snakker i lag i timen så får vi masse kjeft mens jentene de kan sitte å skråle og flire og nærmest HYLE og de blir ikke snakket til. Så han har liksom i mange år kjent på en sånn urettferdighet. Og da lurer jeg jo på [...] om liksom om det er sånn at jentene liksom at de tilpasser seg fortere at de er sånn her må jeg oppføre meg, sant inne i den her firkanten her er mitt handlingsrom (mimer på bordet), og så blir det sett av læreren tidlig, sant, og så blir de oppfattet som at de kan holde seg inni dette handlingsrommet så da blir det større slakk» (IP4)

IP4 forklarer altså at enkelte typer oppførsel kan belønnes med at du får friere tøylar. Videre poengterer hun at det er langt enklere å irettesette gutter enn jenter, fordi guttene ikke legger mer i det enn at de må skjerpe seg. Jentene på sin side er mer sensitive og leser mer i det hun sier. Hun må dermed tilpasse seg en forventet motreaksjon når jentene skal få tilbakemelding, for å bevare relasjonen mellom henne og elevene:

«Hvis jeg sier at skjerp dere NÅ er det nok så tåler de (guttene) det da er det ok sorry *IP4s navn*. Jentene de tåler det de også men hvis du blir oppfattet feil, jeg kan ikke nødvendigvis ordlegge meg på samme måte [...] jeg husker på VG1 da var jeg søkkfortvilt over de her jentene for det var så mye mæmæmæmæ (mimer prating) hele tiden. Og bare (sukker), sant, den irettesettingen må du bruke jævlig mye energi på fordi de er jenter» (IP4)

IP6 er inne på det samme, men knytter det også til at gutter og jenter som tuller i timene gjør det på ulike måter. Guttene vil oftere gjøre fysiske ting, mens jentene gjerne fniser og tisker for seg selv og forstyrrer klasserommet i mindre grad. Hun presiserer at ingen av måtene å ikke følge med på er greit, men at det nok er slik at lærere raskere hanker inn guttene. Hun reflekterer videre:

«er man litt sånn ekstra streng med de som man ja på en måte vet kan være bråkete og stjeler mye oppmerksomhet? Men så føler jeg jo ikke selv at jeg er det, i hvert fall, det er jo ikke sånn at jeg tenker at nå skal jeg være EKSTRA streng med dem. Men man blir jo liksom litt, er det vanskeligere for dem å skille seg ut på positivt vis enn det er for jentene?»

Hun forteller at hun har diskutert dette med elevene sine:

«Og så var det noen som trakk fram den serien Klar til strid [...]den handler om førstegangstjenesten, og de hadde lagt merke til hvor forskjellig guttene og jentene ble behandlet av befalet[...] og de mente det var liksom et bilde av sånn det var ellers i samfunnet. At så fort jenter har et problem for eksempel vil ha utsettelse på innlevering eller uansett så er det sånn (overbærende) ja selvfølgelig, mens med

gutter så er det sånn nei du burde ikke ha sittet og gamet. At det er ikke noen FORSTÅELSE der, og da altså (flau stemme) hæ nei herregud er det sånn?» (Ip6)

Elevene til IP6 forteller altså om det samme som IP4 gjør; lærere er kanskje strengere med gutter enn med jenter.

Vurdering

Med vurdering mener jeg hvordan lærere kartlegger og gir tilbakemelding om faglig kompetanse. (Vurderingssituasjoner blir derav en sosial kontekst hvor læreren får et inntrykk av elevens kompetanse, noen ganger formelt igangsatt gjennom prøver eller lignende). Her er alle lærerne gjennomgående opptatt av det store bildet når de setter karakterer; «[...] det er jo en helhetsvurdering av alle vurderingssituasjonene du har hatt.» (IP2). Disse vurderingssituasjonene er ikke bare de formelle prøvene eller framleggene, for også det elevene viser i timene ellers teller mye. «Hvilket inntrykk har jeg av ham egentlig? Hvor mye har han vært med og bidratt i det faglige og i det muntlige i timene?» som IP5 sier det. I tillegg er lærerne bevisste på at det å vurdere atferd i timene også må vektes mot elevens forutsetninger:

«Altså jeg bruker tid på å forklare dem det herre jeg ser på ALT dere gjør, ikke bare det dere skriver eller det dere forteller når vi har en formell vurdering [...] Men å prøve å ha en sånn tanke hele tiden om hvordan jobber de, hva viser de i undervisningen og hva viser de av progresjon ER det en progresjon (pause), og så må jeg jo selvfølgelig tenke over for jeg har jo elever som ALDRI tar ordet i plenum når det er full klasse[...] og da kan ikke jeg bare si å nei de tar ikke ordet i plenum nei da (krysser på et imaginært ark) lav måloppnåelse. For min jobb er jo å forsøke å legge til rette for at de kan vise meg de de KAN, og da må jeg gjøre det på andre måter» (IP4)

IP1 forteller om det samme: «jeg har en sånn greie at, veldig mange er ikke så muntlige av seg, enkelte er veldig sjenert, og da har jeg tenkt at de ikke skal bli skadelidende for det.» (IP1). Lærerne forsøker altså finne de beste måtene for elevene å vise kompetansen sin på.

Oppsummering for oppførsel og vurdering:

- Jenter oppfattes som mer flittig og pliktoppfyllende. Alle lærerne viser til at jentene stresser mer med å prestere, mens guttene er mer avslappet og mindre seriøse.
- To av lærerne snakker mye om og problematiserer at gutter og jenter har ulike forventninger til seg og behandles ulikt i klasserommet.

- Lærerne snakker om en samlet skjønnsvurdering når de setter standpunktkarakter. Denne er en kombinasjon av formelle vurderinger, elevens progresjon og evne til revisjon, samt hva elevene viser av innsats og arbeid i klasserommet – alt sammen balansert opp mot forutsetningene eleven har.

5.2.2 Om eksamen

I denne delen av intervjuene fokuserte informantene og jeg på to ting. Først pratet vi om den allmenne forskjellen mellom standpunkt – eksamenskarakterer uavhengig av kjønn. Tanken var å undersøke vurderingspraksisene på interne og eksterne vurderinger, samt forskjellen mellom skriftlig og muntlig eksamen. Deretter fortalte informantene om hvordan gutter og jenter oppfører seg før, under og rundt eksamen eller store prøver.

Skriftlig eksamen

På skriftlig eksamen på studiespesialiserende linje er det slik at 47% av elevene går ned en karakter sammenlignet med standpunkt i faget, mens 40% står på samme karakter (utdanningsdirektoratet, 2020). Det er altså vanligst å gå ned, og på informantene mine virker det som at dette er en kjensgjerning i skolen. De gir flere forklaringer på hvorfor det er slik.

Først og fremst snakker lærerne om at er en forskjell på hva som måles og hvilke hensyn som tas på standpunkt og eksamen. På skriftlig eksamen er det kun et produkt uten ansikt som vurderes; «[...] det du skriver på PC'en eller på arket, det som står der når du leverer inn [...] det er jo det sensorene skal vurdere» (IP2). På standpunkt, derimot, er det langt flere komponenter med i bildet: «Jeg tror litt fordi vi ser hele eleven ikke sant, over hele året eller til og med flere, men at vi ser, altså at faglæreren i større grad kan si noe om deres helhetlige kompetanse enn på en eksamen» (IP4). IP5 legger seg også i de baner:

«For vi har jo samtaler med elevene og treffer elevene og hvis de da har fått fire på en tentamen men kanskje har vist mer kompetanse både i timene og på tidligere prøver så kan vi gi fem. Men så kan du jo dette ned på eksamen hvis du er dårlig eller uheldig i den situasjonen» (IP5)

En standpunktvurdering tar altså ifølge informantene mer hensyn til forskjellen mellom dårlig kompetanse og en dårlig dag, og samler et bredere grunnlag for vurderingen enn det ene dokumentet du leverer fra deg på eksamensdagen. Tilknyttet denne forklaringen er også formen på eksamen. Både IP4 og IP3 presiserer at mange sliter med å formulere seg skriftlig, og at mediet du skal prestere gjennom dermed blir en slags «trakt» som bare lar deler av fagkompetansen din komme gjennom. Mens du ellers ville fått alternative måter å vise

kompetansen din på -som nevnt i forrige delkapittel-, er ikke dette et alternativ på skriftlig eksamen.

Videre snakker flere av lærerne om fraværet av pedagogiske hensyn på skriftlig eksamen. I klasserommet, når faglærer skal sette en karakter er det nemlig ikke så enkelt som at en bare kan vurdere faglig kompetanse. Nedenfor er et ganske illustrerende utdrag fra intervjuet med IP6:

«IP6: Jeg tror det er mye lettere å være streng når man ikke kjenner personen bak teksten, og det merker man jo altså jeg tror jo at når du sitter og retter dine egne elever sine tekster så VIL du jo gjerne at de skal føle, for vi vet jo hvor viktig mestringsfølelse er. Og så vet vi jo kanskje altså (pause) at det er stor forskjell på om du setter 2+ eller 3- liksom. Også VIL man kanskje (sukker) gi elevene dine et lite puff og så setter du 3- når du egentlig burde ha satt 2+ fordi du vet at det har ikke noe å si egentlig.

F: ja at det her kanskje er det beste for eleven sin utvikling?

IP6: ja og så typ når du er ekstern sensor så ser du ikke det mennesket og du bryr deg ikke, så jeg tror helt sikkert at det er noe av det. Men så er det jo selvfølgelig sånn at man kanskje drar det litt for langt da og blir FOR snill med sine egne»

IP4 er inne på det samme, og forteller at tankegangen er at elevene trenger å føle at de får til for å jobbe videre. Karakterer blir en litt selvoppfyllende profeti, sier hun, i den forstand at en elev som egentlig ikke burde fått fire på tidspunkt A faktisk ofte kommer seg opp på den måloppnåelsen på tidspunkt B med motivasjonen det gav. Det skjer kanskje ikke med en elev som får en faglig «korrekt» toer. Da kan eleven fort tenke «[...] nei nå du det her faget er bare tapt for meg liksom.», forteller IP4. Det virker altså som at en del lærere er mer opptatt av elevens utvikling enn det å være helt «korrekt» i karaktersetningen dersom disse to hensynene kolliderer på interne vurderinger.

Muntlig eksamen

Hva muntlig eksamen gjelder, er trenden motsatt fra skriftlig. For det siste året før koronaen inntraff -2018/19- var snittet i muntlige fag 4.37 på standpunkt og 4.5 på de tilhørende muntlige eksamenene på studiespesialiserende linje (Utdanningsdirektoratet, 2023). Dette var den delen av intervjuet hvor lærerne var mest unisone i svarene sine. De forteller om to grunner til at det er enklere å gjøre det godt på muntlig enn skriftlig eksamen.

For det første er det slik at «[...] du får HJELP altså, selv om det er eleven som skal prestere så er det jo en kommunikasjon som foregår mellom sensorene og eleven» (IP2). Hjelpen er

ofte i form av «[...] den her lille, liksom, kan du gjenta jeg forsto ikke helt det du sa, kan du si det på nytt? Så du får bort alle disse tullefeilene» (IP1), eller at lærerne aktivt leter etter det eleven kan hvis hen står fast en plass. For det andre er det slik at lærerne oppgir at elevene i større grad er mer forberedt på det som kommer på muntlig:

«[...]Når du går inn til en muntlig eksamen så har du som oftest, da vet du jo hva temaet eller problemstillingen er før du kommer inn der. Og på skriftlig så kommer jo det som en overraskelse når du setter deg. Og så har du liksom fått litt tid til å forberede deg og jeg tror kanskje fordi du har den her veldig korte tidsrammen både på forberedelsestid og på selve eksamen så er det kanskje litt mer OVERKOMMELIG for mange» (IP6)

Lærerne sier med andre ord at elevene vet hva de går til, og får drahjelp av sensorer som aktivt ønsker å hjelpe dem med å vise kompetanse. Dermed er det enklere for de fleste å gjøre det godt.

Generelt om eksamen og oppsummering

Det er ytterligere to poeng jeg ønsker å dra fram fra samtale om eksamen, som gjelder uavhengig av eksamensform. For det første sier flere av lærerne at det ikke alltid er en sammenheng mellom eksamen og det de jobber med i klasserommet. «Jeg synes det er veldig rart at vi jobber prosess i klasserommet, vi jobber karakterdempet og så bang (slipper mobilen i bordet) kommer eksamen. Det bryter med alt [...] den er kunstig», sier IP3. IP4 forteller det samme, og poengterer at læreplanene «Legger opp til DYBDELÆRING og så legger udir i eksamen opp til at det ikke er dybde som måles da er det bredde plutselig [...] Det er en kjempestor mismatch». Det jobbes med andre ord på et annet vis i klasserommet enn det legges opp til på eksamen. Informantene sier at det dermed er ulik kompetanse som måles på standpunkt og eksamen. IP6 forteller at læreplaner og hva elevene skal lære i skolen endrer seg over tid, men spør seg samtidig om «[...] hvor mye innen vurderingspraksis har man egentlig endret siden 1900-tallet du skjønner?» (IP6). Eksamen presenteres altså som en fastlåst brikke som ikke forandrer seg i takt med skolen.

For det andre snakker IP1, IP2 og IP3 om at guttene girer opp før eksamen: «Når det VIRKELIG gjelder, og i en kort periode, så hiver de seg til og jobber kolossalt» (IP2). Flere av lærerne, slik nevnt tidligere i resultatene om oppførsel, snakker om at guttene er mindre stresset og i mindre grad forbereder seg til vurderinger underveis. Angivelig får de likevel med seg essensen i faget. IP1 beskriver prosessen slik:

«Men så kommer de jo til eksamen, og da da kommer, da tenker jeg guttene og jentene de leser like mye [...] Guttene, som har underprestert hele året, de presterer normalt kanskje, på en eksamen. Mens jentene som har prestert godt hele tiden, de presterer òg godt på eksamen, men RELATIVT sett så går guttene, ja.» (IP1)

IP3 er inne på det samme, og forteller at:

«Jentene stresser mer, de kaver mer, de skal være flink og de skal ha karakterer. Og hvert fall som er min erfaring også, så er det i en del klasser, sant, det er grupper som ønsker å være flink og det er ofte grupper med elever som er litt mere sånne middelhavsfarere som absolutt kunne ha trykket på men som ikke gjør det [...] de her gruppene som vil være flink det er overvekt av jenter.» (IP3)

Guttene slår seg altså i større grad til ro med å gjøre det greit nok underveis i semesteret, mens jentene gjerne vil prestere godt på det som kommer. Men noen av informantene virker altså å foreslå at alle elevene ser alvorret når eksamen kommer, og at guttene da relativt sett girer opp mer enn jentene siden sistnevnte har jobbet jevnt og trutt hele året.

Oppsummering for *om eksamen*:

- Alle lærerne forteller at standpunkt har et langt større grunnlag for å sette vurdering enn eksamen, som er en stikkprøve som kan gå alle veier. Eksamen skiller for eksempel ikke mellom dårlig kompetanse og en dårlig dag.
- På skriftlig eksamen vurderer sensor et produkt, og informantene sier at det da er lettere å være streng. Pedagogiske hensyn er borte. IP3 og IP4 poengterer at mange sliter med å utrykke seg skriftlig.
- På muntlig eksamen mener informantene at elevene får drahjelp fra sensorer som aktivt forsøker å finne ut hva de kan. Elevene vet også problemstillingen på forhånd på muntlig.
- Noen av lærerne påpeker at eksamen og standpunkt ikke nødvendigvis måler samme type kompetanse.
- Halvparten av lærerne tar opp at gutter girer opp arbeidsinnsatsen rett før eksamen.

5.2.3 Kjønnsvurderingsgapet i skolen

Jeg spurte alle informantene om kjønnsvurderingsgapet er noe som diskuteres i skolen. Bli det satt søkelys på at gutter konsekvent tar innpå jentene på eksamen? I tillegg snakket vi også litt om hvorvidt kjønnsvurderingsgapet er et problem, sett i forhold til andre ting skolen må håndtere.

For det første er *samtlig*e informanter tydelige på at dette ikke er noe som blir diskutert verken med arbeidskolleger, eller på møter i regi av skoleledelsen. Fire av seks informanter visste ikke om kjønnsvurderingsgapet i det hele tatt, og de to siste var nylig blitt klar over det via en brennpunktdokumentar. Det at guttene systematisk gjør det relativt bedre på eksamen enn jentene, går tilsynelatende helt under radaren i skolen.

For det andre poengterer de fleste informantene at årsaken til kjønnsvurderingsgapet definerer om det er et problem. Særlig virker det som at lærerne synes det er et problem om diskrepansen mellom standpunkt karakterene og eksamens karakterene skapes i klasserommet gjennom at jentene får overbetalt eller guttene underbetalt:

«Men når jeg tenker på det så er det jo kanskje (pause) jeg vet ikke om det er urettferdig heller. Kanskje guttene takler bedre det å ha eksamen? For det, hvis det ikke er sånn at vi lærere faktisk diskriminerer gutter når vi vurderer da, DA er det et problem» (IP5)

«Der tenker jeg at når man får vite ÅRSAKEN, så kan man si om det er et problem. For hvis årsaken er det som vi snakket om i sted, det er sånn at lærere på en måte ubevisst vil gi bedre karakter til en elev som er snill og stille, og gjør det han skal, liksom at det har ikke nødvendigvis noe å si med hva eleven kan, men står en elev og vipper litt, så ja du er så snill og grei og du prøver, så derfor får du opp en karakter, mens bråker du og ikke følger med og fjaser og maser så går du ned en karakter. Hvis DET er årsaken, så tenker jeg at det er et problem. Men (pause) hvis problemet er på en måte i kjønnsforskjellen, at [...] guttene gjør liksom det som må til, minste motstands vei før eksamen[...]» (IP1)

Det IP5 og IP1 sier er på mange vis representativt for det resten av informantene forteller. Lærere ønsker å være rettferdige, og synes det er ubehagelig å tenke på at årsaken kan ligge i klasserommet. «Jeg blir jo egentlig litt skremt av det[...]», som IP6 sier når vi innledningsvis i intervjuet pratet om fenomenet. Samtidig virker det som at kjønnsvurderingsgapet oppfattes som et mindre problem dersom guttene rett og slett skjerper seg inn mot eksamen, eller hvis de takler den bedre enn jentene.

Oppsummering for *kjønnsvurderingsgapet i skolen*:

- Ingen av informantene er blitt informert om kjønnsvurderingsgapet fra skolehold. Det diskuteres heller ikke i kollegiet.
- Lærerne synes kjønnsvurderingsgapet er et større problem dersom årsaken ligger i klasserommet, enn hvis årsaken er at gutter skjerper seg inn mot eller håndterer eksamen bedre enn jenter.

5.2.4 Forklaringer på kjønnsvurderingsgapet

I denne delen av intervjuene spurte jeg informantene om de kunne tenke seg noen årsaker til kjønnsvurderingsgapet. Noe kondensert kom de med fire eksplisitte forslag, som jeg skal gjengi i det følgende. Det er viktig å presisere at de fleste informantene kommer med flere alternative forklaringer. Det fremstår med andre ord som at lærerne oppfatter kjønnsvurderingsgapet som sammensatt med flere årsaker.

Gutter skjerper seg på eksamen

IP1, IP2 og IP3 foreslår, i tråd med det de nevnte i intervjudelen *om eksamen*, at kjønnsvurderingsgapet kan komme av at guttene rett og slett går på en slags cruisekontroll gjennom skoleåret, før de hanker seg inn like før eksamen. Innsatsen gjennom semesteret forklares slik av IP1:

«Hvis du ser gjennom hele skoleåret liksom, så er jentene, liksom, så flink hele tiden. De gjør alt de skal hele tiden. Mens guttene, der er det mer skippertakmentalitet, at de gjør (pause) de forbereder seg ikke nødvendigvis til timen, følger kanskje ikke så mye med, og dermed så kan det jo hende at det har litt negativt utslag på standpunktkarakteren.» (IP1)

Jentene, derimot, har foten på gasspedalen hele tiden. Når det er like før formelle vurderinger, forteller IP2 at «[...] jenter tar det kanskje ETT hakk videre[...] på en måte forbereder seg mer intenst og legger større vekt på den prøven, på betydningen av den prøven enn guttene gjør». Det betyr likevel ikke at guttene ikke er i stand til å kave seg opp før vurderinger; «Når det VIRKELIG gjelder, og i en kort periode, så hiver de (guttene) seg til og jobber kolossalt» (IP2). Det informantene sier er at guttene rett og slett ikke gjør dette underveis i semesteret. De tilhører oftere gruppen av elever som «[...] absolutt kunne ha trykket på men som ikke gjør det» (IP3). IP3 presiserer - som tidligere nevnt - at guttene likevel får med seg essensen i faget selv om de ikke pugger like mye som jentene før vurderinger. IP2 og IP1 sier det samme:

«Men så kommer de jo til eksamen, og da da kommer, da tenker jeg guttene og jentene de leser like mye [...] Guttene, som har underprestert hele året, de presterer normalt kanskje, på en eksamen. Mens jentene som har prestert godt hele tiden, de presterer òg godt på eksamen, men RELATIVT sett så går guttene, ja.» (IP1)

Jenter begunstiges av tradisjonell vurderingspraksis

En annen forklaring som foreslås er at tradisjonell vurderingspraksis har noen aspekter ved seg som kan begunstige jenter. Det er særlig IP1, IP3 og IP5 som peker på at dette kan

påvirke kjønnsvurderingsgapet. Tanken de kommer med er at tradisjonell vurderingspraksis passer jentene bedre enn guttene. Førstnevnte får derfor vist mer av kompetansen sin enn guttene, selv om begge kjønn har tatt essensen i faget. Dermed får jentene en uforholdsmessig høyere standpunkt karakter enn guttene, og denne distansen minker deretter på eksamen som bare er én vurdering. IP3 sier det slik:

«[...] Noen lærere vurderer på de her tradisjonelle punktene, hvor du må sette deg ned og du må pugge dette kapitlet. Og de elevene som orker å gjøre det, de får en bedre karakter enn de som tenker at ehm err nei, det er godt nok, og jeg ser en tendens til at det er mer gutter som tenker slik enn jenter.»(IP3)

Hun sier altså at i klasser hvor det er mange formelle vurderinger som krever pugging, vil jentene skinne i større grad enn guttene all den tid de sistnevnte ikke tar skolen like seriøst. IP5 forteller om det samme, og sier at [...] Jenter ØVER mer til prøver for at de må, mens guttene ikke tar det så hardt.». Det informantene foreslår er at tradisjonell vurderingspraksis lar jentene akkumulere gode karakterer gjennom året fordi de kaver seg opp før vurderinger, som deretter har innvirkning på slutt karakteren deres. Guttene kan ha tatt essensen i faget like godt, men de kaver seg i mindre grad opp før for eksempel prøver underveis i året.

Denne forklaringen underbygger IP3 ved å vise til endringer i karakterene til elevene sine etter overgangen til LK20 (ny læreplan). Den nye læreplanen vektlegger nemlig prosessfokus og revisjonskompetanse over rene læremål av typen «eleven skal kunne x og y»;

«[...]i stedet for å få vurderinger ja nå er du på en toer en treer, så snakker vi heller om at her er to eller tre punkter, dette må du jobbe med nå skal vi jobbe med dette hos deg, mens en annen elev får andre punkter. Og så får de jobbe sånn skreddersydd hele veien uten å jage etter karakterer. De kommer til slutt.» (IP3)

Når læreren sammenligner hvordan de flinke jentene gjorde det med tradisjonell vurderingspraksis, hvor de fikk mange karakterer underveis og hvordan de gjør det nå når karakteren kommer til slutt, sier hun videre:

«Du har en strebergjeng i hver klasse [...] og vi har tatt fra dem toppene og det har jeg faktisk sett også nå [...] de elevene fikk ofte fem på prøven og såne her ting. De ligger på fire nå. [...] Vi har senket skuldrene og tatt vekk puggekunnskapen. [...] de har noen svakheter når det kommer til å anvende kompetanse som IKKE kommer fram om du har en prøve med dem, men som kommer fram når de står i en prosess og utvikler produkter og får veiledning og jobber videre[...] Du kan ikke pugge deg til det her.»(IP3)

De flinke jentene lærer altså ifølge IP3 ikke egentlig så mye mer enn guttene i løpet av året, men at det ofte kan virke sånn hvis man har mange puggeprøver underveis fordi jentene orker å forberede seg mens guttene ikke gjør det. Og, som hun sier,

Det du sitter og nipugger på før et vurderingspunkt er ikke nødvendigvis det du husker best når du kommer til slutten av året. Det du kanskje husker best, det er det du GJØR i klasserommet så jeg er ikke overrasket over at de er ganske lik på eksamenen[...]for jeg tror de jobber ganske likt i klasserommet

IP1 forteller om det samme som IP3, men tar også opp at alle fag domineres av skriftlige vurderinger, og at skriftlige vurderinger passer jenter bedre enn gutter:

«Jeg ser og, for eksempel i fysikk som er mitt fag og et muntlig-praktisk fag, men så ender det veldig ofte opp med at vurderingene som man gir dem i stor grad er skriftlige. Men så egentlig burde det vært en muntlig-praktisk, det er DER karakteren skal ligge. [...] Så om det kan være noe, at det er i de fagene som er muntlig eller praktisk, så er tradisjonelt sett lærerne sånn at vi vil ha det på papir, hva eleven presterer, og kanskje da jentene får skinne litt der, mens guttene som er veldig praktisk av seg, ikke får betaling for det.» (IP1)

Jenter begunstiges av «rett» oppførsel i klasserommet

En tredje forklaring som tas opp er at jentene oppfører seg på en måte som kan gjøre at læreren vurderer arbeidet deres med et fordelaktig blikk. Som IP5 spør seg selv: «Blir man blendet av at de er pliktoppfyllende?». Alle informanter bortsett fra IP3 tar opp dette når de blir spurt om årsaker til kjønnsvurderingsgapet. IP2 forteller det slik:

«Sånn generelt vil jeg si at jentene jobber vel mer jevnt enn gutter.[...] Og det kan jo også hende, at vi, ubevisst, bevisst, ubevisst kanskje, belønner, SER at de gjør mer innsats i det de gjør, at vi på en måte dermed bedømmer dem litt bedre? [...] Vi setter jo pris på de som er med og bidrar i timene, og jeg vet ikke, kanskje vi ubevisst ikke klarer å være 100% nøytrale, at vi LAR det påvirke oss? Ikke nødvendigvis [...] jeg tenker ikke faglig, jeg tenker, ubevisst, at de er med, de viser innsats, at bare DET gjør at vi bedømmer dem annerledes?» (IP2)

IP1 forteller det samme, men kobler også på dikotomien mellom den snille og stille eleven på den ene siden, og den forstyrrende bråkmakeren på den andre:

Og så (humrer litt) har du jo den andre, som, ja, er litt sånn, fyfy å snakke om. [...] om det er sånn at lærere vil favorisere en stille, snill elev. Hvis du har en bråkete elev, så, står den eleven og vipper mellom to karakterer, 4 eller 3, så får han tre, for han der han er en bråkemaker, han forstyrrer timen. Mens en jente som er flink og prøver, og ligg mellom en treer og firer, og prøver og prøver og prøver, så er det liksom (overgivende) ja okei hun har jo prøvd så hun får en firer. Det kan være en årsak.(IP1)

Lærerne er altså åpne for at det å oppføre seg godt kan være utslagsgivende for karakteren som gis, eller at det motsatte kan telle negativt – selv om den snille eleven egentlig ikke viser mer kompetanse enn en tilsvarende flink men bråkete elev. Videre sier IP5 at det kanskje ikke trenger handle om den flittige og flinke eleven heller; det kan rett og slett bare være snakk om en god relasjon mellom lærer og eleven:

«man PRØVER jo å være objektiv og, men det er jo noen elever man har bedre kontakt med enn andre i et klasserom og kanskje man vurderer dem litt ubevisst bedre enn det de egentlig er. [...] Kanskje det er noen som irriterer deg som får dårligere enn de egentlig er? Det skal jo ikke være sånn men jeg prøver å tenke, bare, kan det være det? (IP5)

Dette kobler lærerne på til kjønnsvurderingsgapet, og foreslår altså at når elevene vurderes på eksamen, fjernes den betydningen oppførsel og relasjon eventuelt har. Dermed tar guttene innpå jentenes forsprang.

IP4 og IP6 forsøker i denne sammenhengen å forklare hvorfor jenter oppfører seg på en måte som gir dem fordelaktig blick, og drar en linje til skolens rammer. De snakker begge mye om at jentene knekker skolekoden i større grad: «Systemet i seg selv er kanskje ikke til fordel for guttene, for jentene lærer seg jo tidlig at det er sånn her jeg skal oppføre meg HER» (IP4). IP6 forteller om det samme, og spør seg i intervjuet sitt om:

«er man litt sånn ekstra streng med de som man ja på en måte vet kan være bråkete og stjeler mye oppmerksomhet? Men så føler jeg jo ikke selv at jeg er det, i hvert fall, det er jo ikke sånn at jeg tenker at nå skal jeg være EKSTRA streng med dem. Men man blir jo liksom litt, er det vanskeligere for dem å skille seg ut på positivt vis enn det er for jentene?»

Poenget deres er at de elevene som i større grad mestrer å sitte i ro, følge med og lignende, vil fremstå som flinkere fordi det som oftest er slik klasseromsundervisningen er lagt opp. Og disse elevene er gjerne jenter.

Jenter stresser mer på og rundt eksamen

Den fjerde forklaringen informantene trekker fram, er at jenter blir såpass stresset rundt og på eksamen at de underpresterer. Det er IP2, IP5 og IP6 som tar opp dette:

[...] Men også det her med (pause) for det må jo ha noe med den eksamen å gjøre? Hvorfor gjør ikke jentene det like bra som de vanligvis gjør da? Hvorfor har de plutselig blitt, ja, det gir liksom ikke mening, er det noe med DET kanskje? At de lar seg stresse, jentene?» (IP6)

Men sånn generelt så kan det jo virke som (pause) som, hvis det er en årsak til at karakteren nærmer seg, at på eksamen så nærmer gutter og jenter seg, at det, guttene er mer SKJERPET uten å få angst altså. At jentene lar FRYKTEN for det som skal skje dominere. (IP2)

Det er viktig å presisere at informantene forteller at store vurderinger er stressende for de aller fleste elever, men de tror altså at jentene stresser *mer*:

IP6: Og så er det så KUNSTIG med prøvesituasjoner. VET du hvor mange elever jeg har som går og gruer seg i ukesvis før de skal ha en prøve enten muntlig eller skriftlig, spesielt de store prøvene som de vet er karakterbaserte da og som kanskje VIL gjøre det mye dårligere enn de egentlig har vist...

F: Ja for du opplever at det er vanlig, at det er det vanligste?

IP6: ja (trekker på det), jeg synes på en måte det. Og ikke nødvendigvis at de gjør det typ en karakter dårligere men de kjenner på det stresset og det skaper en ordentlig uro. [...]

F: Men de her som er stresset, opplever du at det er ganske allment? Eller er guttene mer stresset enn jentene, jentene mer stresset enn guttene?

IP6: Jeg synes i alle fall at jentene gir mer uttrykk for at de lar det stresser seg.» (IP6)

IP2, IP5 og IP6 lufter altså tanken om at jentene vippes litt av pinnen av stresset, og guttene derfor gjør det relativt bedre på eksamen.

Oppsummering for *forklaringer på fenomenet*:

- Halvparten av informantene sier at guttene skjærer seg inn mot og på eksamen, og foreslår at kjønnsvurderingsgapet kan forårsakes av det.
- Fem av seks informanter foreslår at jentenes oppførsel i klasserommet begunstiger dem når lærere skal sette standpunkt, og at kjønnsvurderingsgapet kan komme av det fordi jentenes derav ervervede forsprang forsvinner på eksamen.
- Halvparten av informantene foreslår at jentene begunstiges av tradisjonell vurderingspraksis. Mange puggevurderinger og mye skriftlige vurderinger passer jentene bedre enn guttene, som i mindre grad pugger eller klarer å sitte i ro og skrive selv om de forstår essensen i faget omtrent like godt som jentene. Dermed blir jentenes standpunktkarakter uforholdsmessig mye bedre enn guttenes, fordi jentene får akkumulert gode karakterer gjennom året som ikke (bare) reflekterer fagkompetanse. Eksamen avslører den reelle distansen mellom kjønnene.
- Halvparten av lærerne foreslår at jentene stresser mer enn guttene på eksamen, og dermed underpresterer.

5.2.5 Presise vurderinger

Avslutningsvis i alle intervjuene mine snakket jeg med informantene om hva skolen og lærere kan gjøre for å være mest mulig presis i vurderingsarbeidet sitt – og derigjennom muligens minke kjønnsvurderingsgapet. De kom med fire forslag, som jeg skal gjengi nå.

Mer samarbeid om vurderingspraksis

IP1 og IP6 foreslår at mer samarbeid både innad og mellom skoler om vurderingspraksis trengs for at lærere skal kunne være så presise som mulig når de setter standpunkt karakterer. Dette av flere grunner. For det første tar de opp at skillet mellom ulike karakterer kan være vagt:

«altså skillet mellom mange karakterer er jo ganske diffus, altså sånn ja hvis man sitter og diskuterer hva skiller en treer fra en firer for eksempel? Og så er alle enige om at det her er middels måloppnåelse, kan kikke i vurderingskriteriene ok det her skal til, men hvis det nå viser seg at det er lettere å sette guttene på treer og jentene på firer har vi jo et problem.» (IP6)

«[...] Det å sette den her standpunkt karakteren, man snakker jo ofte om å kaste den her terningen, at det blir litt tilfeldig liksom. Okkult ritual, hva er det egentlig som gjør at du får tre og du får fire? Hva er det du har gjort? [...] enkelte ganger, tenker jeg, blir det der bare på magefølelsen til og med kanskje» (IP1)

De forteller altså at det ofte er så små nyanser innad i måloppnåelsene at det virkelig kan være vanskelig å treffe rett. Dermed kan andre forhold enn de faglige bli utslagsgivende – som de mener kan bidra til et økt kjønnsvurderingsgap.

For det andre forteller disse to informantene om alvoret de kjenner på når de skal sette karakterer: «Du har så utrolig mye ansvar for potensielt framtiden til hvor elevene dine kommer inn, og det er klart det er viktig at man setter de riktige karakterene [...]» (IP6). Dette gjør i sin tur at når elever vipper på en standpunkt karakter, så vipper de ofte opp all den tid karakterer er viktig for elevens framtid. Som IP1 sier: «[...] tvil tenker jeg skal komme tiltalte til gode».

IP6 forteller videre om at det er lite opplæring i karaktersetting i lærerutdanningen, og at man er helt prisgitt samarbeidsvillige kolleger for å få diskutert vurdering med noen. Det er derfor hun og IP1 tenker at der er en idé i større grad å samkjøre vurderingspraksis med andre lærere både på egen og andre skoler; mer organisert samarbeid på den fronten gir klarere retningslinjer for vurdering, som deretter gir mindre rom for at andre forhold enn de faglige preger karakteren.

Mer informasjon om kjønnsvurderingsgapet i skolen

IP5 foreslår i sitt intervju at det kan være en idé å gi lærere mer informasjon om kjønnsvurderingsgapet. De trenger ifølge ham å vite at det faktisk *er* slik at gutter relativt sett gjør det bedre på eksamen:

For det første så må vi lærere være oppmerksomme på at det faktisk ER sånn. For jeg visste ikke det [...] Så da må vi være oppmerksom på at ser ut til å være en forskjell der og det KAN være vi som er årsaken til det. [...] Det må du ta med deg i praksisen din når du utøver vurdering. [...] (IP5)

IP5 forteller videre at hvis lærerne blir opplyst om kjønnsvurderingsgapet, blir de nok mer reflekterte og bevisst egen praksis, som i sin tur gjør at de kanskje blir mer faglig presise når de setter karakterene sine. Derfor foreslår han altså å informere lærerne om at fenomenet *er* der: «[...] vi må bli oppmerksomme på det, for hvis ikke så kan vi jo ikke gjøre noe med det» (IP5)

Ha anonyme vurderinger internt for å sjekke egen praksis

Å vokte seg for forutinntatthet tas opp av IP3 og IP6 i denne delen av intervjuene. De foreslår begge at anonyme vurderinger er noe lærere kan være tjent med å ha nå og da, for å overvåke egen praksis. «Det er noe med også som lærer å være litt bevisst på at jeg ikke tenker at, den her eleven hun er jo en femmer, da er det bare å legge seg tilbake og høre» (IP3). De forteller at det er veldig lett å la de første vurderingene eller månedene i klasserommet farge alt som skjer etterpå. IP6 forklarer det slik:

Jeg synes jo at anonyme vurderinger burde vært en mye større greie[...] Ikke nødvendigvis for at man skal gjøre det hver gang men for å ha et slags sjekkpunkt. Det trenger jo ikke være bare kjønn men PERSONEN også, trynefaktor for altså, hvor hvor rettferdig er det egentlig? [...] For du har jo når man etter hvert kjenner elevene sine-uavhengig av kjønn- så vet man jo har en formening om at hun bruker å gjøre det bra han bruker å gjøre det bra, og da når du får inn den prøven eller oppgaven så har du en slags forventning om hvordan det skal ha gått. Så spørsmålet er jo egentlig hvor farget er man av den da, altså hvor viktig er de første vurderingene elevene har for liksom, standpunkt etter hvert da.

IP6 forteller videre at noen læreverktøy som skoler bruker lar deg markere for at du eksempelvis vil lese og vurdere tekstene anonymt, mens andre ikke gjør det. Hun presiserer at det må være *enkelt* å vurdere anonymt, slik at lærere kan gjøre det sporadisk gjennom året for å sjekke egen praksis. Informantene mener altså at dette kan bidra til å minke kjønnsvurderingsgapet; gjennom slike anonyme vurderinger kan læreren se om hen lar seg prege av for eksempel forutinntatthet, og deretter justere vurderingspraksisen sin. Dermed vil

man kanskje minke forskjellen mellom eksamen og standpunkt, fordi de interne vurderingene blir mer faglig presise.

Skolene må ha engasjerte, relasjonelle og aktive lærere som leter etter hva elevene kan

Fire av lærerne (IP2, IP3, IP4 og IP5) tar opp at presise vurderinger krever en dyktig og aktiv lærer. «Du må forsøke å se hele eleven, slik at du i størst mulig grad klarer å gi en helhetsvurdering», som IP2 sier. Lærerne nevner flere måter å jobbe mot dette på.

For det første forteller de at en god faglærer gir elevene mange og varierte måter å vise kompetanse på. Du må i så stor grad som mulig variere hvordan du tester elevene, slik at ingen taper på at det hele tiden vurderes gjennom den samme formen (som eleven ikke mestrer). I den sammenhengen understrekes det at læreren bør være villig til å gi alternative måter å vise kompetanse på til de som av ulike årsaker ikke klarer prestere på de konvensjonelle vurderingsformene: «Hvis du for eksempel har gjort det dårlig på en skriftlig prøve så kan du jo, for noen er jo dårlig skriftlig da kan man få gjøre noe muntlig i stedet. [...] Du må lete etter hva elevene KAN» (IP5).

For det andre sier informantene at må du være tett på elevene for å sette presise karakterer.

IP3 forteller at hun følger arbeidsprosessen til elevene nøye, og:

Det har gjort at når jeg, når Per eller Kari detter av lasset, så ser jeg det fordi jeg går over deres arbeid hver uke sånn hva er det de har gjort nå. Jeg hadde en elevsamtale nå før du kom inn her, med en elev og en forelder for jeg spurte den eleven forrige uke hva det er som gjør at det har gått så ned her? Hva er det som foregår i livet ditt? Og så begynte eleven å skrike og fortalte om veldig mange vanskelige ting. Og da sier jeg du dette må vi ta tak i (*derav dagens møte*). Han hadde ALDRI (banker i bordet med pekefingeren) sagt ifra til meg og jeg tror ikke jeg hadde oppdaget det heller. Og nå tenker jeg jo at nå vet jeg hvorfor det har svinget. [...] da er det på en måte greit. (IP3, min klamme i kursiv))

IP4 er inne på akkurat det samme, og presiserer at å jobbe tett på elevene i større grad lar deg skille mellom dårlig kompetanse og en dårlig periode:

Altså får du en dårlig karakter på den her vurderingen så suger jo det, men hva var det som skjedde med deg i denne perioden? Jo da døde bestemoren min, så da vet jeg at der hadde du en dårlig periode. Da er det ikke nødvendigvis representativt for det du KAN (IP4)

Informantene understreker at å (få lov til) å være tett på eleven krever høy relasjonell kompetanse. Dette er også mer tidkrevende enn det er med mer tradisjonell undervisning og prøver. I den forbindelse problematiserer IP5 at det ikke er noen inngangintervjuer for å bli

lærer. «[...] ansettelsesprosessen burde vært mye bedre. For at noen sånne typer slipper gjennom et intervju (rister på hodet)» (IP5). Samme bekymring uttrykker han også overfor hvor lett det er å komme inn på lærerstudiet. Det er ifølge ham svært mange uegnede lærere i skolen, og det resulterer ofte i upresise vurderinger. «Det har jeg jo sett med mine karer når de gikk på skolen. Særlig i spansken, han fikk fem i standpunkt og så vidt han klarte å stå på eksamen [...] Da har du jo holdt eleven for narr» (IP5).

Essensen i det disse fire informantene forteller er at for å være presis i karaktersettingen, må læreren «[...] være våken på jobben (humrer). Hele tiden» (IP4). De forteller at læreren bør legge opp til varierte vurderingsformer, gi alternativer til de som trenger det og ha forutsetninger i bakhodet. Samtidig må du også kjenne eleven såpass godt at du kan skille mellom dårlig kompetanse og en dårlig dag. Alt dette krever mye av læreren, og derfor blir også ansettelsesprosessen problematisert, all den tid det er elevene som blir skadelidende dersom de har en lærer som ikke er dyktig nok. Dyktige lærere er en forutsetning for å være presis i vurderingsarbeidet, og det å jobbe for at det skal være flest mulig slike i skolen foreslås altså som en måte å minke kjønnsvurderingsgapet på.

Oppsummering for *Presise vurderinger*:

- To av informantene etterlyser mer samarbeid om karaktersetting. De mener at klarere retningslinjer gir mindre rom for at andre forhold enn de faglige blir utslagsgivende når læreren setter karakter.
- IP5 poengterer at lærere må få vite om at kjønnsvurderingsgapet finnes. Det mener han kan gjøre at flere blir mer kritisk og reflektert til egen vurderingspraksis.
- IP3 og IP6 foreslår at man bør ha flere anonyme vurderinger gjennom året, for å sjekke egen praksis for forutinntatthet. Skoler bør ha læremidler som enkelt lar lærerne gjøre dette.
- Fire av lærerne tar opp behovet for at dyktige lærere trengs for å ha presise vurderinger. Disse lærerne må ha stor relasjonell kompetanse.

6. Analyse av data

Jeg vektla et eksplorerende forskningsdesign i denne oppgaven. Kjønnsvurderingsgapet er relativt utrådt lende, og jeg ønsket derfor å snakke med de som står i klasserommet hver dag for å se om de kunne sette meg på sporet av noen plausible forklaringer. Lærerne jeg snakket med gav meg grundige beskrivelser av hvordan de opplever at jenter og gutter oppfører seg, hvordan lærere vurderer, hva som skiller standpunkt fra eksamen – og til sist diskuterte vi også hva som kan forklare kjønnsvurderingsgapet. Lærernes intervjuer var også bakgrunnen for valget mitt av teori. De kausale mekanismene jeg gikk gjennom i kapittel 3 ble valgt ut på bakgrunn av det som ble foreslått -både eksplisitt og implisitt, som jeg skal komme tilbake til i intervjuene. Samtidig er den tidligere forskningen på området kvantitativ, og gir i så måte et godt faktisk grunnlag å vurdere forklaringer og mekanismer opp mot. Som Pierce sa det, «Adopting a hypothesis as being suggested by the facts, is what i call abduction» (Pierce, 1958, 121-122, referert til i Merton, 1987, 3) – og det motsatte er selvfølgelig også tilfelle. Abduksjon er i bunn og grunn prosessen hvor en *leter* etter forklaringer på et observert fenomen. Da må de forslagene som ikke foreslås av faktagrunnlaget skrotes til fordel for andre som passer bedre. I dette analysekapitlet er det nettopp det jeg skal gjøre; jeg skal holde forklaringene lærerne eksplisitt foreslo, og også to ytterligere jeg selv kondenserte fra intervjuene -mer om dette i kapittel 6.2-, opp mot teori og tidligere forskning for å se hvilke av dem som virker som de(n) mest plausible.

Den kommende analysen er strukturert etter forskningsspørsmålene mine. Derfor skal jeg først skrive om kjønnsvurderingsgapets omfang. Deretter analyserer jeg forklaringene på fenomenet, før jeg diskuterer konsekvenser av kjønnsvurderingsgapet. I tråd med problemstillingen brukes jeg mest plass på det andre forskningsspørsmålet.

Forskningsspørsmålene mine er gjengitt nedenfor.

1. Er det kjønnsforskjeller i prestasjonsavvik mellom standpunkt og eksamen i Norge for studiespesialiserende linje?
2. Hva er årsakene til kjønnsvurderingsgapet?
3. Hva er konsekvensene av kjønnsvurderingsgapet for gutters posisjon i den meritokratiske skolen?

6.1 Er det kjønnsforskjeller i prestasjonsavvik mellom standpunkt og eksamen i Norge for studiespesialiserende linje?

Tidligere forskning på kjønnsvurderingsgapet i Norge brukte karakterstatistikk samlet fra forrige læreplan på ungdomsskolen. Jeg samlet derfor data selv. Slik det framgikk i kapittel 5.1 er det et kjønnsvurderingsgap mellom standpunkt og eksamen for studiespesialiserende linje, som har holdt seg relativt stabilt gjennom en tiårsperiode fra 2009/10 og til 2018/19. Jeg viste til at kjønnsvurderingsgapet er størst i skriftlige fag. Regnes alle seks årene under ett, tar gutter innpå 0.104 karakterer fra standpunkt til eksamen i skriftlige fag, og 0.078 i muntlige fag. Det at fenomenet er mindre på muntlig eksamen, er i overenstemmelse med en del av den tidligere litteraturen på området, som foreslår at gutters relative fortrinn på eksamen eller andre eksterne vurderinger minker når anonymiteten blir borte (Lavy, 2008; Falch & Naper, 2013).

6.2 Hva er årsakene til kjønnsvurderingsgapet?

Som jeg har vært tydelig på gjennom teksten, har intervjuene vært utgangspunktet for forklaringene jeg nå skal analysere, både eksplisitt og implisitt. Med det førstnevnte sikter jeg til at informantene kom med eksplisitte forslag til forklaringer på kjønnsvurderingsgapet. Samtidig fortalte informantene mye om både elevers oppførsel og undervisnings- og vurderingspraksis i intervjuene, som for meg utkrystalliserte ytterligere to mulige forklaringer. Jeg skriver om informantenes forslag først, og deretter om de to jeg selv kondenserte fra intervjuene. For å utforske forklaringene trekker jeg på både den tidligere forskningen på området og begrepene som ble presentert i henholdsvis kapittel 2. og 3. i oppgaven, og forsøker å dra fram hvilke som virker mest plausible.

Begunstiger tradisjonell vurderingspraksis jenter?

Denne forklaringen adresseres ikke i stor grad i den eksisterende, kvantitative forskningen på kjønnsvurderingsgapet. Som nevnt i teorikapitlet, er det mulig å tenke seg klasserommet som et sosialt felt, hvor enkelte typer oppførsel, ferdigheter og handlingsmønstre blir belønnet mer enn andre. Tanken er at hvis jenter i større grad innehar denne feltspesifikke klasseromskapitalen, kan det potensielt forklare kjønnsvurderingsgapet, fordi jentene da vil få bedre betalt enn guttene når elevene vurderes. Jeg kommer til å bruke klasseromskapital på to vis. For det første har lærere klare oppfatninger om hvordan elever bør oppføre seg. Dersom jenter i større grad ter seg i tråd med lærerens forventninger om hva som er god oppførsel, kan

det gjøre at de lettere fremstår som flinke i sistnevntes øyne, eller belønnes for den gode oppførselen sin når lærerne setter karakter. Dette aspektet av klasseromskapital kommer jeg tilbake til litt senere i analysen. For det andre kan jentene kanskje ha lettere for å få gode resultater dersom undervisningen er formet på et vis som passer dem – hvis de innehar ferdigheter eller atferdsmønstre som i større grad overlapper med hvordan lærerne legger opp eksempelvis prøver eller innleveringer. Det er dette aspektet som er mest relevant for å belyse hvordan jenter kan bli begunstiget av tradisjonell vurderingspraksis.

Lærernes intervjuer tyder på at hvordan en vurderer spiller en rolle for hvordan ulike elever både forbereder seg og presterer. I delkapittel 5.2.4 viste jeg at lærerne særlig problematiserte to aspekter ved tradisjonell vurderingspraksis som begunstiger jenter. For det første var det slik at:

«[...] Noen lærere vurderer på de her tradisjonelle punktene, hvor du må sette deg ned og du må pugge dette kapitlet. Og de elevene som orker å gjøre det, de får en bedre karakter enn de som tenker at ehm err nei, det er godt nok, og jeg ser en tendens til at det er mer gutter som tenker slik enn jenter.»(IP3)

Guttene har ifølge informantene skjönt essensen i temaet klassen går gjennom omtrent like godt som jentene, men gidder rett og slett ikke å kave seg opp før prøver som tester puggekunnskap. Det gjør derimot jentene. Spør læreren etter puggesvar, vil dermed jentene virke flinkere i faget enn guttene, selv om dette, som IP3 fortalte, ikke er kunnskap du egentlig klarer ta med deg videre eller som reflekterer faktisk fagkompetanse. Samtidig viser intervjuene også at lærerne ønsker å lete etter hva elevene *kan* når de setter karakter, som blant annet gjøres ved at de ser hele året under ett: «For vi har jo samtaler med elevene og treffer elevene og hvis de da har fått fire på en tentamen men kanskje har vist mer kompetanse både i timene og på tidligere prøver så kan vi gi fem.» (IP5) I så forstand kan det at jentene akkumulerer gode karakterer gjennom året i klasserom med mange puggeprøver, bli utslagsgivende for standpunkt karakteren deres. Guttene, som ikke kaver like mye på puggeprøver, har ikke de samme karakterene å vise til på slutten av året. Kanskje får jentene dermed en uforholdsmessig høy standpunkt karakter i forhold til deres reelle kompetanse – eller motsatt for guttene. Deretter synliggjøres dette på eksamen, som oftere spør etter et produkt eleven skal lage – som et essay, et framlegg eller annet hvor en må anvende kunnskap. IP3 fortalte eksempelvis at når hun begynte å vurdere arbeidsprosess og sluttet med puggeprøver, så gikk streberjentene hennes tilbake i karakter: «[...]de har noen svakheter

når det kommer til å anvende kompetanse som IKKE kommer fram om du har en prøve med dem» (IP3).

For det andre var noen av informantene tydelige på at majoriteten av formelle vurderinger er skriftlige:

«Jeg ser og, for eksempel i fysikk som er mitt fag og et muntlig-praktisk fag, men så ender det veldig ofte opp med at vurderingene som man gir dem i stor grad er skriftlige. Men så egentlig burde det vært en muntlig-praktisk, det er DER karakteren skal ligge. [...] Så om det kan være noe, at det er i de fagene som er muntlig eller praktisk, så er tradisjonelt sett lærerne sånn at vi vil ha det på papir, hva eleven presterer, og kanskje da jentene får skinne litt der, mens guttene som er veldig praktisk av seg, ikke får betaling for det.» (IP1)

Dersom det stemmer at majoriteten av alle vurderinger er skriftlige, kan det hende at guttene ikke får vist kompetansen sin like godt som jentene, all den tid guttene behersker det å sitte i ro og få ned det de kan i en tekst, dårligere enn førstnevnte (NOU, 2019:3; Krumsvik, 2021). Dette kan også fungere andre veien: IP1 forteller at «når vi skal ha praktisk arbeid, så er guttene veldig påkoblet», i kontrast til hvordan det er i tradisjonell undervisning. Guttene, som i sistnevnte undervisning ikke er så engasjert og med, blir oppfattet som mer innsatsgivende og engasjerte når de rammene endrer seg i mer fysisk og praktisk retning. Det at skriftlige vurderinger dominerer, underbygges av at IP5 forteller at muntlige vurderinger av tidshensyn ofte blir gjort i grupper. Det er rett og slett ikke tid til at alle i klassen skal ha en presentasjon *hver*, for eksempel. Det har implikasjoner for vurderingspraksisen:

Og det sier jeg til elevene, at gruppearbeid de teller mindre enn de karakterene du skaffer alene. For noen kan jo komme med noen som er veldig dårlig, og det trekker ned arbeidet. Og da kan du ikke straffe de andre elevene for DET (IP5).

Intervjuene kan altså tyde på at det ikke bare er flest skriftlige vurderinger, men at de muntlige teller mindre fordi de ofte er gruppearbeid. Samtidig utfordres dette aspektet av forklaringen om tradisjonell vurderingspraksis, av at diskrepansen mellom jenter og gutter på standpunkt er omtrent identisk i skriftlige og muntlige fag. Jenter har i snitt 0.317 høyere standpunktkarakter enn gutter i skriftlige fag i datamaterialet mitt (Utdanningsdirektoratet, 2023). Tilsvarende tall for muntlige fag er 0.315 – altså er kjønnsforskjellen nesten den samme. Dersom guttene får ut mer av kompetansen når det i mindre grad er skriftlige vurderinger, ville en forvente at diskrepansen mellom kjønnene på standpunkt var mindre i muntlige fag. Men, som IP1 sier, domineres *også muntlig-praktiske fag* av skriftlige

vurderinger. Det kan være at det egentlig ikke er en større forskjell i klasserommet mellom muntlige og skriftlige fag når det gjelder vurderingspraksis (forskjellen blir dermed bare hvilken eksamenstype elevene går opp til). Likevel er det trolig at det er muntlige vurderinger i *høyere grad* i muntlige fag, og det burde vært reflektert i standpunktkarakterene.

Det første aspektet holder derimot mer vann. Legger en opp til mange puggeprøver, vil kanskje jentene få uforholdsmessig godt betalt i forhold til reell fagkompetanse. Jentene kan i så måte akkumulere gode karakterer gjennom året fordi de orker å kave seg opp til å gjengi faktakunnskap. Deretter preger disse puggekarakterene standpunktkarakteren deres (som ofte kan bli et produkt av tidligere karakterer og innsats i klasserommet) i positiv retning. Dette kan ikke guttene vise til i like stor grad, fordi de ikke orker pugge seg opp selv om de tar essensen i faget noenlunde like greit. Dermed blir standpunktkarakteren deres lavere enn den faktiske fagkompetansen deres – som blir synliggjort på eksamen, hvor de minker jentenes forsprang fra standpunkt.

Ulik konkurranseintensitet på standpunkt og eksamen – gutter skjerper seg, jenter stresser?

Flere av informantene snakket om at guttene tar skippertak like før eksamen, og at jentene muligens stresser mer enn guttene før tilsvarende. Begge disse forklaringene går på det tidligere forskning kaller ulik konkurranseintensitet mellom standpunkt og eksamen, og følgelig samler jeg disse to forklaringene under samme overskrift.

I utgangspunktet er det en plausibel forklaring at guttene skjerper seg på eksamen, all den tid det er noe flere informanter uavhengig av hverandre sier at de opplever, og disse er tross alt tilstede «på bakken» gjennom året. Guttene har fått med seg essensen i faget, og «Når det VIRKELIG gjelder, og i en kort periode, så hiver de seg til og jobber kolossalt» (IP2).

«Men så kommer de jo til eksamen, og da da kommer, da tenker jeg guttene og jentene de leser like mye [...] Guttene, som har underprestert hele året, de presterer normalt kanskje, på en eksamen. Mens jentene som har prestert godt hele tiden, de presterer òg godt på eksamen, men RELATIVT sett så går guttene, ja.» (IP1)

Samtidig utfordres forklaringen av resultater på nasjonale prøver gjennom grunnskolen. Slike prøver er i større grad «lavterskelprøver» enn eksamen er. Det er dermed lite trolig at guttene kaver seg veldig opp før disse på grunn av situasjonens alvor. Likevel er kjønnsvurderingsgapet så absolutt tilstede på nasjonale prøver (Falch & Naper, 2013, 19-22; Lindahl, 2007, 17-18). Dog er det viktig å huske på at nasjonale prøver ikke bare

gjennomføres på et annet tidspunkt enn eksamener og standpunktkarakterer settes på -som dermed gjør at en bør være varsom med å sammenligne ukritisk-, de tester heller ikke helt det samme som måles av lærere og sensorer for hele skolefag (NOU, 2019:3, 42). Dette anerkjenner Protivínský & München (2018) i sin forskning på kjønnsvurderingsgapet i Tsjekkia. De sammenlignet derfor kjønnsvurderingsgapet på eksamen for alle elever, med kjønnsvurderingsgapet for elever som i spørreundersøkelser oppga at de i liten grad stresses på høyterskelprøver, og fant ingen forskjell i fenomenets omfang (Protivínský & München, 2018, 146-147).

Hva stress for jentene gjelder, er informantene tydelig på at jentene stresser og kaver mer enn guttene, og foreslår at dette kanskje setter førstnevnte mer ut av balanse når store prøver slik som eksamen nærmer seg. Likevel er det mye som taler mot denne forklaringen.

Argumentasjonen over viste at kjønnsvurderingsgapet eksisterer i like stor grad blant elever som oppgir at de *ikke* er stresset, og på lavterskelprøver som det ikke knyttes særlig press til. Lavy (2008, 2103) viser at fenomenet også er tilstede på skoler hvor guttene i det hele og store presterer bedre enn jenter. Videre har israelske ungdomsskoleelever to sluttvurderinger i hvert fag som er svært like (her kan en sammenligne med tentamen og eksamen i Norge), hvor sluttkarakteren blir en fusjon av disse. Den ene vurderes internt av faglærer og den andre av en ekstern sensor, anonymt. Der finner Lavy (2008) at guttene vurderes substansielt bedre på den eksterne prøven. Begge disse vurderingene teller like mye for karakteren til elevene, og det er dermed lite trolig at guttene og jentene henholdsvis skjerper seg eller stresser mer på den ene av de to. Forklaringen som går på ulik konkurranseintensiteten på eksamen kontra standpunkt, virker derfor å ha mindre empirisk støtte enn de andre forklaringene jeg analyserer.

Begunstiges jenter av «korrekt» oppførsel?

Denne forklaringen er en gjenganger i tidligere forskning, og er den med bredest oppslutning blant informantene mine. De forteller at jentene er flittigere, høfligere og flinkere enn guttene, og problematiserer at det kanskje farger over på vurderingene lærere gjør. I Norge har Gustavsen & Nordahl (2021) vist at det er en markant korrelasjon mellom hvordan lærere vurderer de sosiale og faglige ferdighetene til elevene sine. Her er det viktig å huske på at eksisterende skoleforskning er tydelig på at sosiale ferdigheter er viktige for å få et godt læringsutbytte i skolen (Krumsvik, 2021; NOU, 2019; Gustavsen, 2018). Det kan dermed

være at slike korrelasjoner faktisk bare reflekterer at jenter er flinkere enn gutter i skolen, blant annet *fordi* de har overlegne sosiale ferdigheter. Likevel foreslår informantene og de som har forsket på kjønnsvurderingsgapet tidligere at jentenes gode oppførsel og beherskelse av sosiale ferdigheter ikke bare gir et faglig utbytte, men også en uforholdsmessig god vurdering fra lærernes side.

Spørsmålet er hvordan dette i så fall foregår. I tråd med den tidligere gjennomgangen av klasseromskapitalen, er det noen måter å oppføre seg på som oppfattes mer positivt enn andre av lærere. «[...] Vi setter jo pris på de som er med og bidrar i timene, og jeg vet ikke, kanskje vi ubevisst ikke klarer å være 100% nøytrale» (IP2). Kanskje ender lærerne da -bevisst eller ikke- med å [...] sanction/reward these gender non-neutral skills by giving to students lower/higher marks and this would result in a difference between blind and non-blind scores in favour of girls (Casula et al., 2022, 484).

Glorieeffekten kan være med på å belyse hvordan dette foregår. Som nevnt i teorikapitlet, er det gjennomgående slik at sinnet ønsker koherens; også lærere har lettere for å vurdere noen fordelaktig på område B dersom de allerede har vurdert dem tilsvarende på område A. Altså, hvis jentene oppfører mer i tråd med lærernes ønsker enn guttene, kan det være at lærerne tar de med seg det gode inntrykket av jentenes oppførsel, og ubevisst lar det påvirke vurderingene deres. Rent konkret kan en tenke seg at Martin og Hanne akkurat har hatt en prøve. I ukene før prøven fulgte Martin i liten grad med, pratet mye uten lov, vandret i gangene og fungerte ganske dårlig sosialt i gruppene som forberedte seg til prøven. Hanne, derimot, noterte det læreren sa i timene, møtte presis, stilte spørsmål og var raus og inkluderende i gruppearbeidet. Dette har læreren selvfølgelig sett, og inntrykket fra elevenes oppførsel kan godt tenkes å prege hvordan læreren leser og vurderer prøvene. Eksempelvis kan vi se for oss at begge elevene får fire – en normalt god karakter. I konteksten gitt ovenfor, vil nok de fleste av oss ganske intuitivt tenke at Martin var heldig, og Hanne uheldig.

Det lærerne forteller er i stor grad i tråd med den ovenfornevnte tankegangen. Jentene er mer flittige og oppfører seg oftere i tråd med det som anses som hensiktsmessig i skolen. De er mer forberedt til timene, og «[...] er jevnt mer MED i det som foregår» (IP2). Lærerne er i den sammenheng inne på at skolen passer bedre for jentene. Det er slik at skolen «[...] på en måte har blitt litt skapt for dem. De situasjonene med å sitte å skrive, sitte i ro og liksom konsentrere seg over tid». IP4 forteller det samme, og sier at:

«gutter som [...] kanskje er litt senere i utviklingen enn jenter og liksom YNGRE er jo også mer aktiv fordi de har kanskje ikke kommet ditt enda. [...] Det kan sitte mange gutter rundt omkring som kjenner på det, at det her er jo en tvangstrøye for meg for jeg har jo behov for noen andre ting og så ja vi snakker så fint om tilpasset undervisning og opplæring for alle ikke sant, men skolen er likevel strukturert sånn som den er [...] (IP4)

Det virker altså som at skolens rammer legger føringer for at bestemte typer oppførsel belønnes, som jenter innehar i større grad enn gutter. Det i sin tur kan være med på å farge lærernes faglige vurderinger. Denne forklaringen er også i samsvar med empirien på området, som viser at kjønnsvurderingsgapet er større på skriftlig eksamen enn på muntlig. Jentenes gode oppførsel gir dem uforholdsmessig godt betalt i klasserommet, mens denne ikke kommer til uttrykk i det hele tatt på en anonymt vurdert skriftlig eksamen. På muntlig er kjønnsvurderingsgapet mindre enn på skriftlig. Dette kan i denne konteksten forklares av at den ene av sensorene på muntlig tross alt er elevens faglærer, og vedkommende vil da ha med seg en potensiell glorieeffekt inn på eksamen, som kan påvirke hvordan sensorene i lag setter karakter. I tillegg trenger ikke glorieffekten å utvikle seg over lengre tid. Elever som på muntlig er høflig, smiler, virker godt forberedt eller lignende, kan levne et godt førsteinntrykk på en ekstern sensor, som i sin tur kanskje preger vedkommende sin faglige vurdering av eleven. All den tid lærere jevnt over vurderer jentene til å ha bedre sosiale ferdigheter enn guttene (Gustavsen, 2018; Gustavsen & Nordahl, 2021), er det tenkelig at jentene på muntlig eksamen dermed bevarer noe av den eventuelle fordelene oppførselen deres gir på standpunkt. Det gir derfor mening, ut fra forklaringen om oppførsel, at kjønnsvurderingsgapet er mindre på muntlig eksamen enn skriftlig.

Kumulativ forutinntatthet - preger de første vurderingene lærernes karaktersetning til fordel for jenter?

Denne forklaringen er en jeg selv har kondensert fra datamaterialet. Noen av informantene snakket nemlig om at lærere bør ha anonyme vurderinger i ny og ne, for å sjekke egen vurderingspraksis for forutinntatthet. Det fikk meg til å gruble på betydningen forutinntatthet har i skolen. For å sitere IP6: «[...] spørsmålet er jo egentlig hvor farget er man av den da, altså hvor viktig er de første vurderingene elevene har for liksom, standpunkt etter hvert da».

Slik jeg var inne på når jeg skrev om matteuseffekten i teorikapitlet, er det nok slik at lærere flest *vet* at jenter er litt flinkere på skolen enn gutter. For det første kan det bety at lærere kanskje allerede fra første vurdering har en forventning med seg når de vurderer jenter, som

kan påvirke tolkningen deres. De er kanskje vant til at jenter gjør det godt. For det andre er det - fordi jenter *er* noe dyktigere på skolen - sannsynligvis slik at de elevene som får gode karakterer på de første vurderingene på høsten, oftest er jenter. Dette kan også ha innvirkning på senere tolkning – på samme måte som Merton (1968; 1988) viste at allerede kjente vitenskapsmenn får uforholdsmessig mye oppmerksomhet og ros for arbeidet sitt fordi folk leser det med en positiv forutinntatthet, kan en tenke seg at en også finner en kumulativ fordel eller ulempe for elever som tidligere har gjort det godt eller dårlig. Mekanismen er den samme: en elev har tidligere levert gode resultater. Siden læreren da oppfatter eleven som kompetent, vil hun lese arbeidet til eleven mer inngående, fordi det som regel er noe produktivt å hente ut fra elevens tekster. Og jo mer inngående en leser et arbeid, desto mer sannsynlig er det nok at en får noe ut av det. Lærerne jeg snakket med er veldig tydelige på at de leter etter det eleven *kan*. Dog, for å sette det litt på spissen, er det tenkelig at forutinntattheten om hvor mye eleven angivelig kan, påvirker hvordan og hvor lenge læreren leter.

Denne forklaringen kan videre bygges opp ved å koble den til ankereffekten. Karakteren eleven vanligvis får, kan fungere som et anker fra hvis det er vanskelig å flytte seg som lærer:

[...] The first question i scored had a disproportionate effect on the overall grade. The mechanism was simple: if i had given a high score on the first essay, i gave the student the benefit of the doubt whenever i encountered a vague or ambiguous statement later on. Surely a student who had done so well [...] would not make a foolish mistake!» (Kahnemann, 2012, 85).

Dersom Olav er en strålende elev og bare får femmere og seksere, vil det kanskje sitte lengre inne for en lærer å gi ham en treer. Det blir et såpass stort avvik fra normen at en går noen ekstra runder med produktet som vurderes – og så finner en kanskje noe som rettferdiggjør en karakter som ikke viker så langt fra ankeret. På samme vis kan en se for seg at lærere også vil nøle noe med å gi en svært god karakter dersom elever som vanligvis gjør det dårlig, plutselig presterer godt.

Dermed kan elever som er flinke på de første vurderingene, eller tilhører en gruppe som det oftere er knyttet positive forventninger til, flyte på dette utover i skoleåret. De blir i så måte vurdert som *litt* flinkere enn det de egentlig er. Samtidig vil elevene med mer negativ forutinntatthet mot seg, vurderes uforholdsmessig dårligere enn det de faktisk er. Denne fordelene eller ulempen er kumulativ, fordi den kontinuerlig vil forsterke seg selv;

Uforholdsmessig gode vurderinger bygger opp inntrykket av en flink elev enda mer, som i sin tur gjør at neste vurdering blir enda mer preget av en positiv forutinntatthet – og motsatt. Slikt sett kan matteus- og ankereffekten forklare kjønnsvurderingsgapet – for den kumulative forutinntattheten er jo borte på en anonym eksamen. Jentene vurderes uforholdsmessig positivt på standpunkt, og så korrigeres dette på eksamen. Forklaringen kan også styrkes av at Lavy (2008) finner at guttenes relative fortrinn på eksterne vurderinger minker markant idet anonymiteten forsvinner – det *kan* tyde på at lærere har med seg en generell positiv forutinntatthet om jenter som gruppe. Videre finner Gustavsen & Nordahl (2021) at norske kontaktlærere rangerer jentenes faglige ferdigheter over guttenes i klassene sine. Internasjonal skoleforskning peker også på at det er høyere forventninger til jenter (Wang et al., 2018, 131). Slikt sett passer forklaringen til at kjønnsvurderingsgapet er større på skriftlig eksamen (hvor elevene vurderes anonymt), enn på muntlig (hvor de ikke blir det).

Samtidig er det slik at dersom kumulativ forutinntatthet forklarte hele kjønnsvurderingsgapet, burde fenomenet vært mye mindre på muntlig eksamen enn på skriftlig. På førstnevnte eksamen vurderes elevene av en ekstern sensor som vet at jenter oftere er flinkere enn gutter, og en intern sensor som har med seg en forutinntatthet om den aktuelle elevens tidligere prestasjoner. Her burde med andre ord jentene -ut fra en slik forståelse- holde på forspranget fra standpunkt i større grad. Likevel gjør gutter det relativt bedre også her, og ikke veldig mye mindre enn de gjør på skriftlig (de tar innpå 0.078 karakterer på jentenes forsprang fra standpunkt på muntlig eksamen, mot 0.104 på skriftlig). Kumulativ forutinntatthet kan altså godt ha noe å si for kjønnsvurderingsgapet, men det virker ikke som at det *alene* er nok til å forklare fenomenet.

Får jenter i større grad pedagogiske karakterer enn gutter?

Dette er den andre forklaringen jeg selv kondenserte fra datamaterialet. Underveis i intervjuene fikk jeg nemlig et inntrykk av at lærerne var mer opptatt av elevens utvikling enn å sette 100% korrekte karakterer faglig sett – Slik en kan se i de følgende utdragene:

[...]Jeg vet jo at jeg har jo gitt pedagogiske karakterer, de får kanskje en pluss hvor de egentlig ikke skal få en pluss hvis du skjønner[...] For at du tenker KAN det her motivere kan det her få deg til å kjenne på at du mestrer noe (IP4)

Og så vet vi jo kanskje altså (pause) at det er stor forskjell på om du setter 2+ eller 3- liksom. Også VIL man kanskje (sukker) gi elevene dine et lite puff og så setter du 3- når du egentlig burde ha satt 2+ (IP6)

Å gi en pedagogisk karakter vil si at en med overlegg gir en elev en noe bedre karakter enn hen egentlig burde få ut fra en rent faglig vurdering. Tanken bak er at elevene skal kjenne på en mestring som motiverer dem til å jobbe videre med faget. Som IP4 forklarer det, er det ofte slik at elever som på tidspunkt A ikke fortjener en firer, faktisk jobber seg opp til den kompetansen på tidspunkt B med motivasjonen det gav – karakteren blir en selvoppfylgende profeti. Selvoppfylgende profetier er noe jeg gå mer inn på i neste delkapittel om konsekvensene av kjønnsvurderingsgapet.

I seg selv belyser pedagogiske karakterer aller mest hvorfor det er vanlig å gå ned i karakter på skriftlig eksamen – det medmenneskelige hensynet er borte, og tekstene vurderes strengere. Likevel tror jeg dette kan ha en betydning for kjønnsvurderingsgapet også. I intervjuene fortalte IP4 at det er langt vanskeligere å rette jenter enn gutter; «Jeg kan ikke nødvendigvis ordlegge meg på samme måte til jentegrupper som jeg kan til guttegrupper [...]» (IP4), fordi jentene tar til seg ting og leser mer i det lærerne sier enn gutter gjør. Videre sier flere av informantene at jentene gjerne vil ha bekreftelse på at det de gjør er rett – de er mindre selvstendige i arbeidet sitt en guttene, og har mindre tro på egne ferdigheter. I det store og hele virker det som at jentene er mer sensitive for tilbakemeldinger. Samtidig peker eksisterende skoleforskning på at lærere ofte tilpasser seg en forventet motreaksjon når de skal gi elever tilbakemeldinger (se for eksempel Mechtenberger, 2008). Ut fra det informantene sier kan dermed virke som at lærere i større grad må tilpasse seg en motreaksjon når de skal komme med en negativ tilbakemelding til jenter, enn til gutter:

[...]hvis en jente får en dårlig karakter, så er det jo nesten så tårene kommer frem. Du ser jo hvor lei seg de blir. Mens guttene blir, ja (fnyser og kaster noe over skulderen) det fjaser vi bort, det er, ja, ingen problem. (IP1)

Hvis lærerne vet at jenter tar seg mer nær av et dårlig resultat, *samtidig* som de også er ganske klare på effekten karakterer kan ha for elevens motivasjon, kan det faktisk være at de i større grad gir pedagogiske karakterer til jenter enn til gutter. En kan selvfølgelig innvende at pedagogiske karakterer handler om at elevene skal få motivasjon til å jobbe videre, og det i så måte ikke er hensiktsmessig å benytte seg av en slik karaktersetning på utgangskarakteren til elevene (altså standpunkt), for da er de jo ferdige på skolen. Men samtidig er standpunkt ifølge informantene en helhetsvurdering basert på det eleven har gjort gjennom året – herunder også de karakterene elevene har oppnådd underveis. Dermed kan det at jenter i større grad får pedagogiske karakterer enn gutter underveis i året (hvis så er tilfelle), være

med på å blåse opp standpunktkarakteren til jentene noe i forhold til deres reelle faglige kompetanse.

Samtidig handler pedagogiske karakterer om elevens utvikling, og brukes dermed på interne vurderinger av lærere med et slikt perspektiv i mente. Slik er det ikke på eksamen. Det at en ser en forskjell i kjønnsvurderingsgapets omfang mellom skriftlig og muntlig eksamen, peker på at det må være noe mer enn bare den ovenfornevnte forklaringen i bildet. Et eller annet gjør at jentene taper noe mindre av forspranget sitt på muntlig eksamen, og det lar seg ikke forklare av at de i høyere grad får pedagogiske karakterer underveis i skoleåret.

Diskusjon av forklaringene

Flere av forklaringene jeg har gått gjennom ovenfor har absolutt potensiale til å belyse kjønnsvurderingsgapet. Likevel er det noen som virker mer lovende enn andre. I det følgende diskutere hvilke av forklaringene som er mest plausible, i rekkefølge fra den jeg argumenterer for at har minst forklaringsevne, til den med tilsvarende størst.

For det første får forklaringen som går på ulik konkurranseintensitet mellom standpunkt og eksamen, liten empirisk støtte. I eksisterende forskning er det gjort mange undersøkelser som gir lite hold til at kjønnsvurderingsgapet kan forklares av at guttene skjerper seg såpass mye eller at jentene stresser til den grad at de underpresterer: fenomenet er til stede på lavterskelprøver som det er knyttet lite press til, og det samme er tilfelle på godt sammenlignbare internt og eksternt vurderte prøver hvor begge burde inspirere like stor grad av alvor og stress (Lavy, 2008).

Dernest følger forklaringen om pedagogiske karakterer. De lærerne som snakker om pedagogiske karakterer, er tydelige på at elevens utvikling trumfer helt presis karaktersetning når de hensynene kolliderer. All den tid de fleste informantene forteller at jentene tar seg mer nære av tilbakemeldinger, i langt større grad trenger bekreftelse utenfra og blir lei seg av dårlige karakterer, er det ikke så usannsynlig at slike pedagogiske karakterer i større grad gis til dem. Dersom jentene derigjennom får uforholdsmessig høye karakterer gjennom året, vil de i så måte stille sterkere når standpunktkarakteren settes. Samtidig utfordres dette av at det er en mindre - men markant- forskjell i kjønnsvurderingsgapets omfang avhengig av hvilken eksamensform en ser på; guttene tar i større grad inn på jentene fra standpunkt til skriftlig eksamen, enn de gjør fra standpunkt til muntlig eksamen. Pedagogiske karakterer er langt på vei et fenomen som handler om elevens utvikling, og forekommer dermed på interne

vurderinger; på eksamen er det ikke snakk om utvikling videre. Det at det da er en forskjell i kjønnsvurderingsgapets omfang mellom eksamensformene, tyder på at det må være noe mer inne i bildet enn bare at jentene muligens får pedagogiske karakterer i større grad enn gutter.

Forklaringen om tradisjonell vurderingspraksis er mer tiltalende. Lærerne forteller at jentene i større grad enn guttene kaver seg opp før vurderinger som vektlegger pugging, og får betalt i karakter som ikke nødvendigvis reflekterer faktisk fagkompetanse – de gjengir bare faktakunnskap. Den reelle distansen mellom gutter og jenters fagkompetanse er i så måte noe som avsløres på eksamen. IP3 opplever eksempelvis at de flinke jentene har gått tilbake i karakter med de nye læreplanene, som i mindre grad vektlegger puggevurderinger. Det er mindre empirisk støtte for at mye skriftlige vurderinger gir fordel til jentene, som var det andre aspektet ved tradisjonell vurderingspraksis lærerne tok opp. Men, dersom en bare ser på tradisjonell vurderingspraksis som det å ha mange puggevurderinger, ser forklaringen lovende ut. Dog sliter også denne forklaringen med å belyse hvorfor kjønnsvurderingsgapet er mindre på muntlig eksamen enn på skriftlig, fordi den bare sier noe om den interne karaktersettingen.

Forklaringen om kumulativ forutinntatthet virker også å ha mye for seg. På den ene siden peker skoleforskning på at lærerne har høyere forventinger til jenter ((Wang et al., 2018, 131; Gustavsen & Nordahl, 2021, 155). I tillegg er det nærliggende å anta at jentene, som faktisk *er* flinkest på skolen, også er flinkest tidlig i semesteret. Dermed kan lærere både bli påvirket av en forutinntatthet om jenter som gruppe, men også av de faktiske individene i en klasse – eksempelvis gjennom matteus-og ankereffekten. Likevel er det slik at dersom kumulativ forutinntatthet var den eneste forklaringen på kjønnsvurderingsgapet, skulle en forventet at gapet var langt mindre på muntlig eksamen enn på skriftlig. Elevene vurderes som nevnt da av en intern sensor som kjenner dem fra før av, og en ekstern sensor som vet at jenter oftere gjør det bra. Kjønnsvurderingsgapet *er* litt mindre på muntlig, som er i tråd med at tidligere forskning peker på at guttenes fortrinn på eksterne vurderinger minker idet anonymiteten forsvinner (Lavy, 2008; Falch & Naper, 2013). Men samtidig er kjønnsvurderingsgapet bare *noe* mindre på den ikke-anonyme muntlige eksamenen, og da kan ikke forklaringen om forutinntatthet alene forklare fenomenet. Det må, også her, være noe annet med i tillegg.

Dermed virker det for meg som at den mest plausible forklaringen er at jenter begunstiges av «korrekt» oppførsel i klasseromsinteraksjonen. Denne forklaringen vises til både i eksisterende forskning på kjønnsvurderingsgapet, og det er den forklaringen flest av

informantene mine tar opp. Det som særlig gjør denne forklaringen mer plausibel enn de to foregående, er at den kan brukes til å svare på hvorfor vi ser en forskjell i kjønnsvurderingsgapets omfang på muntlig og skriftlig eksamen. Jentene kan ifølge informantene få et utbytte i vurderingssammenheng av at de viser god oppførsel. Jo mindre av oppførselen sin de får vist, desto mindre burde forspranget deres til guttene dermed være. Dette passer overens med de kvantitative dataene mine. Distansen mellom kjønnene i oppnådd karakter er nemlig størst på standpunkt (0.35 favør jenter), mindre på muntlig eksamen (0.272 favør jenter) og minst på skriftlig (0.246 favør jenter). Disse tre vurderingene har synkende grad av sosial interaksjon knyttet til seg; standpunkt baserer seg på et års samhandling, muntlig på en halvtimes, mens den skriftlige eksamenen vurderes helt anonymt. Det ser altså ut til at guttene tar inn mer av jentenes forsprang når graden av sosial interaksjon minker.

Det betyr ikke at denne forklaringen er den eneste bak kjønnsvurderingsgapet. Som nevnt flere ganger, er dette et komplekst fenomen som det nok ligger flere mekanismer bak. Jeg sikter heller til at den er den mest utslagsgivende *blant* mekanismene jeg gikk gjennom – de andre spiller nok også en rolle, og da særlig forklaringen om kumulativ forutinntatthet og forklaringen om tradisjonell vurderingspraksis. Dette skal jeg komme tilbake til i avslutningen av oppgaven. Der skal jeg også videre poengtere at de forklaringene jeg har gått gjennom, baserer seg på intervjuene mine og i så måte antakeligvis ikke er de eneste mekanismene som kan forklare kjønnsvurderingsgapet; det kan godt være andre forklaringer som lærerne og jeg ikke kom på som vil komme fram gjennom videre forskning.

6.3 Hva er konsekvensene av kjønnsvurderingsgapet for?

Det tredje og siste forskningsspørsmålet mitt gikk på konsekvensene av kjønnsvurderingsgapet. Her er det viktig å presisere at jeg i denne oppgaven hverken har fokus på eller plass til å diskutere konsekvensene av at jenter *i det hele og store* gjøre det bedre på skolen. Det jeg ønsker å adressere er derimot konsekvensene av kjønnsvurderingsgapet. Hvordan en angriper dette er avhengig av hvor årsaken til gapet ligger. I det foregående delkapitlet pekte jeg på at det virker mindre sannsynlig at årsaken ligger på selve eksamen gjennom at gutter skjærper seg eller at jenter stresser. Jeg skal derfor utforske konsekvensene av at kjønnsvurderingsgapet forårsakes av skjeve standpunktvurderinger. Hva jeg legger i konsekvenser må også adresseres, slik at det ikke i

det følgende fremstår som at forfatteren taler gutters sak i skolen uten holdepunkt. Som jeg var inne på innledningsvis i oppgaven, er det viktig å forske på kjønnsvurderingsgapet fordi fenomenet rokker ved den meritokratiske tankegangen i skolen. En av utdanningssystemets manifeste -altså intenderte og tydelig uttalte (Merton, 1967, 37-138) - funksjoner er å sortere innbyggere etter kompetanse; de flinkeste får de beste karakterene, som deretter gir gode muligheter for videre studier og arbeid. Denne funksjonens legitimitet hviler på at alle vurderes etter læreplanene i de ulike fagene – det kreves en universalisme i vurderingen av elevene. Kjønnsvurderingsgapet tyder på at dette ikke er tilfellet. Det er da konsekvensene for gutters posisjon i den meritokratiske skolen jeg utforsker i det følgende.

Dersom kjønnsvurderingsgapet skyldes skjeve vurderinger på standpunkt i favør jenter som deretter «avsløres» på eksamen, vil konsekvensene primært gå på gutters videre utvikling og sjanser i utdanningssystemet. Den skjeve vurderingen blir gjeldene på to måter. For det første peker kjønnsvurderingsgapet på at differansen mellom gutter og jenter egentlig ikke burde vært så stor på standpunkt. I oppgaven har jeg vist at guttene tar innpå 0.104 karakterer av jentenes forsprang på skriftlig eksamen, og 0.078 på muntlig – som en kan se på som et uttrykk for hvor mye mindre differansen burde vært på standpunkt. Dersom guttene i snitt burde hatt 0,104 bedre karakter i skriftlige fag, tilsvarer det at hver tiende gutt skulle hatt én hel bedre karakter. Nå er det selvfølgelig ikke så enkelt. Mange fag på studiespesialiserende har både skriftlige og muntlige eksamener, og det er i så måte ikke enkelt å avgjøre «hvor mye bedre» internt vurdert karakter guttene skulle hatt. Videre er kjønnsvurderingsgapet et sammensatt problem; årsakene er flere og varierte, og en skal derfor vokte seg veldig for bombastisk å si at gutter skulle hatt slik og så mye bedre karakter på standpunkt. Likevel må det adresseres, for uansett om det er vanskelig å fastsette et konkret tall, tyder forklaringene jeg har gått gjennom på at det nok *er* en skjev vurdering i klasserommet. Dermed blir den mest åpenbare konsekvensen av kjønnsvurderingsgapet at gutters karaktersnitt blir dårligere enn deres reelle kompetanse skulle tilsi (eller at jentenes blir bedre enn tilsvarende). Gutter gjør det allment dårligere enn jenter i skolen, har vanskeligere for å få plass på attraktive studier, og det er en overvekt jenter i høyere utdanning. Kjønnsvurderingsgapet kan slikt sett være med på å utvide jentenes fortrinn på disse områdene til et *uforholdsmessig* nivå, som er problematisk ut fra den meritokratiske skolens krav om en universell vurdering etter faglig kompetanse.

Den andre konsekvensen av kjønnsvurderingsgapet, dersom årsakene ligger i upresise vurderinger på standpunkt, er at disse vurderingene kan prege hvordan elevene utvikler seg faglig. Terrier (2020), som så på kjønnsvurderingsgapet i Frankrike, forsket særlig på akkurat dette og fant at:

« for two classes where the achievement gap between boys and girls would be identical in grade 6, randomly assigning a teacher who is one standard deviation more biased against boys to one of the classes would decrease boys' relative progress in that class by 0.123 SD in math and by 0.1106 SD in french»(Terrier, 2020, 8)

Det datamaterialet til Terrier antyder altså at jo skjevare lærere vurderer i favør jenter, desto større blir jentenes relative faglige framgang målt mot guttenes. Det at jentene tilsynelatende får uforholdsmessig gode vurderinger fra lærerne sine er dermed ikke bare uheldig der og da, men kan også utvide den faktiske kompetanseforskjellen mellom gutter og jenter. Denne forskningen kan belyses og potensielt forklares av den selvoppfyllende profetien som mekanisme.

Etter at Rosenthal & Jacobsen (1968) i sin banebrytende bok om selvoppfyllende profetier viste hvordan læreres forventning til elever faktisk gir utslag på faglig prestasjon og IQ-utvikling har dette blitt testet og replisert gjentatte ganger (se f.eks Wang et al., 2018 for en kunnskapsoversikt over forskning på selvoppfyllende profetier i klasserommet de siste 30 årene). Tanken er at læreren får en oppfatning av eleven, som deretter former lærerens atferd mot eleven: «When one "knows" a child is bright, his behavior is evaluated as of higher intellectual quality than is the very same behavior shown by a child "known" to be dull.» (Rosenthal, 1968, 54). For det første vil lærerens atferd i seg selv være med på å prege elevens læring; Lærere gir flere positive og akademiske tilbakemeldinger til elever de har høye forventninger til, og stiller mer krevende spørsmål – noe som har utslag på læringsutbytte (Wang et.al, 2018, 135-138). For det andre er elevene samtidig flinke til å oppfatte lærernes forventninger, som i sin tur preger elevens akademiske motivasjon og innsats i klasserommet. På det eller lignende vis vil altså lærerens «profeti» om elevenes kompetanse, kunne oppfylle seg selv. Og, i samsvar med forskning på kjønnsvurderingsgapet, har lærere oftere høyere forventninger til jenter enn til gutter (Wang et al., 2018, 131; Gustavsen & Nordahl, 2021, 150-155).

I intervjuene mine, som nevnt i forrige delkapittel, sier lærerne at de bevisst kan gi en bedre karakter enn det elevens faktiske kompetanse svarer til, fordi det preger elevens motivasjon og innsats. En kan i så måte se på karakterer som en uttrykt forventning fra lærerens side – «så her god er du». Poenget i forhold til konsekvensene av kjønnsvurderingsapet, er at de upresise vurderingene internt i klasserommet kan bidra til å øke kjønnsforskjellene i skolen. Dersom gutter får en noe dårligere karakterer enn de egentlig burde, vil de da få uttrykt en uforholdsmessig lav forventning til seg, som deretter påvirker den videre utviklingen deres i negativ retning. Det motsatte vil være tilfelle for jenter. Gjennom at de får litt for gode karakterer i forhold til reell fagkompetanse, slik forklaringene jeg har gått gjennom antyder, opplever de i større grad en positiv forventning til seg, som deretter bedrer mulighetene deres til å utvikle seg videre.

7. Avslutning

Temaet for denne mastergradsoppgaven har vært kjønnsforskjeller i skolen. Mer spesifikt forsket jeg på kjønnsvurderingsgapet, som går ut på at gutter - både i Norge og i andre land - gjør det relativt bedre på eksamen enn jenter, sammenlignet med standpunkt. Tidligere forskning på området vektla å etablere dette fenomenet, og har dermed gjennomgående vært kvantitativ med fokus på karakterstatistikk. Dermed manglet feltet innsikten kvalitativ forskning ofte kan supplere kvantitativ med; hvorfor og hvordan gjør gutter det relativt bedre på ekstamen? I tråd med dette ble problemstillingen min *Kjønnsvurderingsgapet i skolen – hvorfor gjør gutter det relativt bedre enn jenter på eksamen?*. Fra den utledet jeg tre forskningsspørsmål: ett om fenomenets omfang, ett om fenomenets årsaker, og ett om konsekvensene av det. For å svare på disse har jeg både hentet inn karakterstatistikk fra utdanningsdirektoratet (2023) som belyser kjønnsvurderingsgapets omfang, og hatt lengre dybdeintervjuer med lærere på studiespesialiserende linje på videregående. Tyngden i oppgaven lå på sistnevnte, og det er intervjuene jeg har brukt som utgangspunkt for å svare på problemstillingen min. I det foregående kapitlet analyserte jeg intervjuene mine ved hjelp av både tidligere forskning på kjønnsvurderingsgapet og teoretiske begreper. Disse begrepene var i all hovedsak kausale mekanismer som kan belyse hvorfor guttene systematisk gjør det bedre enn jentene på eksamen.

Tidligere har Falch & Naper (2013) sett på kjønnsvurderingsgapet i Norge, men de baserte forskningen sin på ungdomsskolen under læreplan L97. Av den grunn valgte jeg å hente min egen karakterstatistikk, og da fra studiespesialiserende linje på videregående; for det første var videregående et ubelyst område, og for det andre har videregående langt flere eksamener enn ungdomsskolen (og dermed et større datagrunnlag). I tillegg så jeg på om kjønnsvurderingsgapets omfang har vært stabilt over tid. Som vist både i resultat - og analysekapitlene, har kjønnsvurderingsgapet holdt seg relativt stabilt i perioden jeg så på (fra 2009 til 2019). Fenomenets omfang varierer eksamenstype. I skriftlige fag tar gutter 0.104 karakterer inn på jentene på eksamen, mens det tilsvarende tallet for muntlige fag er 0.078.

Det andre forskningsspørsmålet mitt var årsaksforklarende. Det at jenter begunstiges av å vise korrekt oppførsel ser ut til å være den mest plausible forklaringen av de jeg har gått gjennom. Gjennom å bruke glorieffekten som en mekanisme hvor en positiv vurdering på område A smitter over når vi skal vurdere noen på område B, har jeg forsøkt vise hvordan jentenes

overlegne sosiale ferdigheter kan bli mer utslagsgivende enn vi tror når lærere skal sette karakterer. Samtidig er nok kjønnsvurderingsgapet et sammensatt og komplekst fenomen som det ligger flere mekanismer bak. Særlig to av de andre forklaringene virket også lovende: kumulativ forutinntatthet og tradisjonell vurderingspraksis. Selv om de hadde sine utfordringer, gir de en god innsikt i andre måter jentene kan begunstiges på når lærerne setter karakterer, utover bare oppførsel. Forklaringen om kumulativ forutinntatthet belyste jeg med matteus- og ankereffekten; de elevene som gjør det godt tidlig i skoleåret, eller tilhører en gruppe som det er positive forventninger til, får uforholdsmessig gode karakterer senere fordi forutinntattheten («Hun her pleier å levere noe bra») preger hvordan lærerne vurderer arbeidet deres. Denne fordelten er kumulativ fordi den kontinuerlig vil forsterke seg selv; uforholdsmessig gode vurderinger bygger opp inntrykket av en flink elev enda mer, som i sin tur gjør at neste vurdering blir enda mer preget av en positiv forutinntatthet – og motsatt, selvfølgelig. Forklaringen om hvordan tradisjonell vurderingspraksis kan begunstige jenter virket også lovende. Hvordan en legger opp undervisning og vurdering vil passe noen elever bedre enn andre. De egenskapene og handlingsmønstrene som gir belønning i klasserommet – klasseromskapitalen – forandrer seg deretter. I klasserom hvor det er mange puggeprøvere vil jentene i større grad få skinne, all den tid informantene mine er tydelige på at de i langt større grad orker kave seg opp til å reprodusere faktakunnskap. Karakterene fra disse puggeprøvene reflekterer egentlig ikke fagkompetanse. Dette blir i så forstand avslørt på eksamen, hvor guttene tar innpå jentenes forsprang.

Det siste forskningsspørsmålet mitt gikk på konsekvensene av kjønnsvurderingsgapet, og da konsekvensene for gutters posisjon i den meritokratiske skolen. Dette valgte jeg å fokusere på fordi en av skolens mest sentrale funksjoner er å sortere innbyggere etter kompetanse. Dette meritokratiets legitimitet hviler på en universell, faglig vurdering av elevene – som det virker som at ikke er tilfelle. For det første tyder kjønnsvurderingsgapet på at guttene får litt dårligere karakterer enn de burde hatt av lærerne sine, sett i forhold til faktisk faglig kompetanse. For det andre argumenterer jeg for at kjønnsvurderingsgapet har innvirkning på elevens faglige utvikling senere. Det å gi elever gode karakterer, ser nemlig ut til å prege den faglige utviklingen deres i positiv retning – og motsatt for elever som får dårlige karakterer. Kjønnsvurderingsgapet kan i så måte bidra til å øke kjønnsforskjellene i skolen, fordi jentene får uforholdsmessig gode karakterer målt mot faktisk faglig kompetanse, og/eller guttene får tilsvarende dårlige. Dermed kan jentene utkonkurrere gutter i kampen om attraktive

studieplasser og jobbmuligheter, i enda større grad enn de ville gjort dersom vurderingene i skolen var presise. Begge konsekvensene jeg tar opp utfordrer universalitetsprinsippet som legitimerer et meritokratisk utdanningssystem.

For å avslutte med å vende tilbake til problemstillingen min, ser det ut fra datamaterialet mitt ut som at gutter gjør det relativt bedre enn jenter på eksamen fordi det foreligger noen mekanismer i klasserommet som gjør at jentene får uforholdsmessig gode standpunktkarakterer. Det er derimot mindre støtte for at årsakene ligger på selve eksamen. Jeg trekker særlig fram at jentes gode sosiale ferdigheter preger hvordan lærerne vurderer elevene på standpunkt, men understreker at dette er den mest utslagsgivende blant flere mekanismer. Den reelle distansen mellom kjønnenes fagkompetanse avsløres deretter på eksamen.

Veien videre

Konsekvensene jeg har utforsket er tankevekkende. De understreker for meg viktigheten av forskningen min, og kan forhåpentligvis motivere andre til å se på det jeg ikke har gjort – nemlig å vurdere hva en kan *gjøre* med kjønnsvurderingsgapet. Jeg skulle gjerne drøftet ulike løsninger, og da især opp mot de potensielle forklaringene jeg har analysert. Originalt var dette planen, og jeg snakket også eksplisitt om løsninger med informantene i intervjuene. Dessuten har funnene mine om forklaringer implikasjoner for eventuelle tiltak mot kjønnsvurderingsgapet, fordi disse selvfølgelig må hvile på en forståelse av årsakene bak. Likevel har jeg latt være å fokusere på løsninger i oppgaven av plasshensyn. Den siste NOU-utredningen om kjønnsforskjeller i skolen gjennomgående klar på at tiltak for å løfte guttene i skolen *ikke* må gå på bekostning av jenter (NOU, 2019:3). Den samme nøye vurderinger bør derfor også gjøres av tiltak for å minke kjønnsvurderingsgapet. Derfor mente jeg at det var bedre å levne tiltakene senere forskning, som kan bygge på funnene mine, enn å forsøke å utforske løsninger på mindre plass enn diskusjonen av disse krever, her i oppgaven min.

Videre er det viktig for meg i å minne om at jeg bare intervjuet seks lærere. Selv om jeg i metodekapitlet var inne på at jeg nærmet meg et metningspunkt når jeg intervjuet de siste lærerne, og at jeg hadde seks svært ulike informanter hva kjønn, alder og fagkrets gjelder, er det et beskjedent datagrunnlag. Intervjuene gav meg en del forklaringer på kjønnsvurderingsgapet -eksplisitt og implisitt-, som virker å ha noe for seg, men poenget mitt er at det er *disse* forklaringene jeg har vurdert. Det kan godt være at det er andre mekanismer

som bidrar til kjønnsvurderingsgapet utover de informantene foreslo eller jeg koblet fra intervjuene. Riktig nok har jeg også forhold meg til den tidligere forskningen på feltet, og ikke funnet andre mekanismer der som jeg ikke har trukket fram i oppgaven. Likevel vil kanskje informanter fra andre deler av landet -eller fra andre land, for den saks skyld- trekke fram andre ting enn mine informanter gjorde. Dessuten kan andre forskere muligens se noe jeg ikke har fått med meg. Det å diskutere kjønnsvurderingsgapet direkte med lærere gjennom kvalitative dybdeintervjuer er det få eller ingen - etter hva jeg kjenner til- andre som har gjort før meg i Norge. Det som derfor slår meg når jeg nå avslutter oppgaven min, er behovet for mer kvalitativ forskning på årsakene til fenomenet. Vil forklaringene jeg dro ut fra mine intervjuer, også være å finne hos andre, og vil de eventuelt suppleres av andre potensielle mekanismer for kjønnsvurderingsgapet? Først når dette er bredere kartlagt, kan en senere se på hvordan en kan minke kjønnsvurderingsgapet på et tilfredsstillende vis.

Referanser

- Berg, P., Palmgren, O., & Tyrefors, B. (2020). Gender grading bias in junior high school mathematics. *Applied economics letters*, 27:11, ss. 915-919.
- Blaikie, N., & Priest, J. (2017). *Social research - paradigms in action*. Cambridge & Malden: Polity press.
- Blaikie, N., & Priest, J. (2019). *Preparing research designs*. Cambridge: Polity press.
- Bourdieu, P. (1984). *Distinction*. London, Melbourne & Henley: Routledge & Kegan Paul.
- Bourdieu, P. (1986). The forms of capital. I J. Richardson, *Handbook of Theory and Research for the Sociology of Education* (ss. 241-258). Westport: Greenwood.
- Brinkmann, S., & Tanggaard, L. (2012). Intervjuet: samtalen som forskningsmetode. I S. Brinkmann, & L. Tanggaard, *Kvalitative metoder - empiri og teoriutvikling* (ss. 17-44). Oslo: Gyldendal.
- Busch, T. (2013). *Akademisk skriving for bachelor - og masterstudenter*. Fagbokforlaget.
- Casula, L., Di Liberto, A., & Pau, S. (2022). Grading practices, gender bias and educational outcomes: evidence from Italy. *Educational Economics*, 30:5, ss. 481-508.
- Cornwell, C., Mustard, D., & Van Parys, J. (2013). Noncognitive Skills and the Gender Disparities in Test Scores and Teacher Assessments: Evidence from Primary School. *The journal of human resources* 48(1), ss. 236-264.
- Elster, J. (2015). *Explaining Social Behavior - more nuts and bolts for the social sciences*. Cambridge University Press.
- Falch, T., & Naper, L. (2013). Educational evaluation schemes and gender gaps in student achievement. *Economics of Education Review* (36), ss. 12-25.
- Glaser, B., & Strauss, A. (1999). *The discovery of grounded theory - strategies for qualitative research*. AldineTransaction.
- Grøgaard, J., & Arnesen, C. (2016, desember 22). Kjønnforskjeller i skoleprestasjon: ulik modning? *Tidsskrift for ungdomsforskning*, 16(2), ss. 42-68.
- Gustavsen, A. M. (2018). *Kjønnforskjeller i sosiale og skolefaglige prestasjoner*. Hentet fra Høgskolen i Innlandet: <https://brage.inn.no/inn-xmlui/handle/11250/2493719>
- Joas, H., & Knöbl, W. (2009). *Social Theory - twenty introductory lectures*. Cambridge: Cambridge university press.
- Kahnemann, D. (2012). *Thinking, fast and slow*. Penguin.
- Krumsvik, R. J. (2021). Det pedagogiske perspektivet på gutters skoleprestasjoner - et blikk på kunnskapsgrunnlaget. I R. J. Krumsvik, *Hva vet vi om skoleprestasjoner? Ulike perspektiver på skolegang og skoleprestasjoner* (ss. 25-66). Fagbokforlaget.
- Lavy, V. (2008). Do gender stereotypes reduce girls' or boys' human capital outcomes? Evidence from a natural experiment. *Journal of public economics* (92), ss. 2083-2105.
- Lindahl, E. (2007). Comparing Teachers' assessments and national test results - evidence from Sweden. *Working paper IFAU-institute for labour market policy evaluation* (24).

- Merton, R. (1967). *On theoretical sociology: five essays, old and new*. The free press.
- Merton, R. (1968). The mathew effect in science. *Science (American Association for the Advancement of Science, 159 (3810))*, ss. 56-63.
- Merton, R. (1988). The mathew effect in science II: Cumulative advantage and the symbolism of intellectual property. *Isis, 79 (4)*, ss. 606-623.
- Merton, R. (1989). Three fragments from a sociologist's notebooks: Establishing the Phenomenon, Specified Ignorance, and Strategic Research Materials. *Annual Review of Sociology (13)*, ss. 1-28.
- NESH. (2016, Desember 16). *Forskningsetiske retningslinjer for samfunnsvitenskap og humaniora*. Hentet fra Forskningsetikk.no: <https://www.forskningsetikk.no/retningslinjer/hum-sam/forskningsetiske-retningslinjer-for-samfunnsvitenskap-og-humaniora/>
- Nordahl, T., & Gustavsen, A. M. (2021). Sammenhengen mellom kjønnsforskjeller i skolefaglige prestasjoner, sosiale ferdigheter og lærerens vurderingspraksis. I R. J. Krumvik, *Hva vet vi om skoleprestasjoner? Ulike perspektiver på skolegang og skoleprestasjoner* (ss. 137-160). Fagbokforlaget.
- Nordahl, T., Løken, G., Knudsmoen, H., Aasen, A. M., & Sunnevåg, A.-K. (2011). *Kjennetegn på skoler med små kjønnsforskjeller - Rapport 14*. Hentet fra Høgskolen i Hedmark: <https://brage.inn.no/inn-xmlui/handle/11250/133966>
- NOU. (2019:3). *Nye sjanser - bedre læring - Kjønnsforskjeller i skoleprestasjoner og utdanningsløp*. Kunnskapsdepartementet.
- Protivínsky, T., & Münich, D. (2018). Gender Bias in teachers' grading: What is in the grade. *Studies in Educational Evaluation (59)*, ss. 141-149.
- Rosenthal, R., & Jacobsen, L. (1968). *Pygmalion in the classroom - teacher expectation and pupils' intellectual development*. Holt, Rinehart and Winston, INC.
- Schifloe, P. M. (2019). *Mennesker og samfunn - innføring i sosiologisk forståelse [3.utg]*. Trondheim: Fagbokforlaget.
- Solli, I. F. (2021). Født for sent? Om aldersforskjeller i klasserommet og skoleresultater. I R. J. Krumsvik, *Hva vet vi om skoleprestasjoner? Ulike perspektiver på skolegang og skoleprestasjoner* (ss. 215-228). Fagbokforlaget.
- Swedberg, R. (2014). *The art of social theory*. Princeton University press.
- Terrier, C. (2020). Boys lag behind: how teachers' gender biases affect student achievement. *Economics of Education review, 77*, ss. 1-20.
- Thagaard, T. (2018). *Systematikk og innlevelse - en innføring i kvalitative metoder*. Fagbokforlaget [5.utg].
- Thomas, W., & Thomas, D. (1928). *The child in america*. Knopf.
- Utdanningsdirektoratet. (2020, September 03). *Karakterforskjeller mellom standpunkt og eksamen i videregående*. Hentet fra Udir.no: <https://www.udir.no/tall-og-forskning/statistikk/statistikk->

videregaende-skole/analyser/karakterforskjeller-mellom-standpunkt-og-eksamen-i-videregaende/

Utdanningsdirektoratet. (2022). *Karaktererstatistikk for videregående skole*. Hentet fra Udir.no: <https://www.udir.no/tall-og-forskning/statistikk/statistikk-videregaende-skole/karakterer-vgs/>

Wang, S., Rubie-Davies, C., & Meissel, K. (2018). A Systematic Review of the Teacher Expectation Literature Over the Past 30 Years. *Educational Research and evaluation*, 24(3-5), ss. 124-179.

Wollscheid, S. (2021). Oppsummert forskning om kjønnsforskjeller i skolen - mangfoldige årsaksfaktorer og sprikende resultater. I R. J. Krumsvik, *Hva vet vi om skoleprestasjoner? Ulike perspektiver på skolegang og skoleprestasjoner* (ss. 113-136). Fagbokforlaget.

Aase, T. H., & Fossekaret, E. (2014). *Skapte virkeligheter - om produksjon og tolkning av kvalitative data*. Universitetsforlaget [2.utg].

Vedlegg 1: Informasjonsskriv

Vil du delta i et forskningsprosjekt om kjønnsforskjeller i prestasjoner på standpunkt og eksamen?

Formål med prosjektet

De fleste er klar over at jenter gjør det bedre på skolen enn gutter – i grunnskolen ligger jenter ganske stabilt en halv karakter over guttene i snitt. Men det er altså slik, både i Norge og andre land, at forskjellen mellom gutter og jenter minker markant på eksamen. I oppgaven min forsøker jeg finne ut av *hvorfor* det er en «forskjell i forskjellene» - hvorfor minker guttene jentenes forsprang på eksamen? På den ene siden kan det være at standpunkt karakterene er upresise på et vis, og at dette korrigeres på eksamen. På den andre kan det være noe med selve eksamen som begunstiger guttene i større grad enn jentene. I oppgaven kaller jeg dette fenomenet for *kjønnsvurderingsgapet*.

Problemstillingen for masteren er derfor følgende:

Kjønnsvurderingsgapet – hvorfor gjør gutter det relativt bedre på eksamen enn jenter?

I tillegg har jeg formulert følgende forskningsspørsmål.

1. Er det kjønnsforskjeller i prestasjonsavvik mellom standpunkt og eksamen?
2. Hva er årsakene til kjønnsvurderingsgapet?
3. Hva er konsekvensene av kjønnsvurderingsgapet?
4. Hva kan lærere og skoler gjøre for å motarbeide fenomenet?

Hvem er ansvarlig for forskningsprosjektet?

Jeg, Herman Hovda Lien, er ansvarlig for prosjektet sammen med veileder Gunnar C. Aakvaag.

Hvorfor får du spørsmål om å delta?

Jeg ønsker å intervju tre mannlige og tre kvinnelige lærere om fenomenet. Utover lærerens kjønn har ingen lærere som har blitt spurt om å delta, blitt kontaktet på bakgrunn av personlige egenskaper, alder, erfaring eller annet. Jeg har kontaktet lærere på skoler jeg har god kjennskap til, heller enn å gå veien om et åpent skriv sendt til administrasjoner på skoler i nærområdet. Dette av rent praktiske hensyn.

Hva innebærer det for deg å delta?

Jeg ønsker å gjennomføre et løst strukturert intervju, hvor jeg har med en del ting jeg ønsker å ta opp, samtidig som samtalen gjerne kan ta en annen retning underveis. Hovedpoenget er refleksjoner rundt fenomenet – hva kan være årsakene til at guttene plutselig minker jentenes forsprang på eksamen?

Særlig er jeg ute etter å høre dine erfaringer med gutter og jenters oppførsel i klasserommet, hvordan de ter seg når eksamen nærmer seg, og litt om vurderingssituasjoner generelt.

Intervjuene vil bli tatt opp, slik at jeg slipper stykke opp samtalen vår ved å gjøre notater av det du sier underveis. Etterpå transkriberes intervjuene. Selve seansen tar omtrent en time, alt ettersom. Jeg kommer til å stille spørsmål om:

1. Dine tanker om fenomenet (visste du om dette fra før av, dine tanker om hva som kan forklare fenomenet, etc)
2. Hvorvidt du ser noen systematiske forskjeller guttene og jentene imellom, når det gjelder oppførsel i klasserommet.
3. Vurderingssituasjoner og eksamen.

Det er frivillig å delta

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle dine personopplysninger vil da bli slettet. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg. Dersom du ønsker det, kan du få tilsendt intervjuet ditt – enten som lydfil eller når det er ferdig transkribert -, for så å gå gjennom om det er noe du ønsker å fjerne eller ta bort.

Ditt personvern – hvordan vi oppbevarer og bruker dine opplysninger

Jeg vil bare bruke opplysningene om deg til formålene jeg har fortalt om i dette skrevet. Jeg behandler opplysningene konfidensielt og i samsvar med personvernregelverket, og det er kun student og veileder i dette prosjektet som kommer til å ha tilgang til det transkriberte intervjuet.

Tiltak for å sikre at ingen kan koble opplysningene du kommer med i intervjuet, til deg, innebærer blant annet:

- Bruk av pseudonymer. Ditt transkriberte intervju vil bli hetende for eksempel «intervju «Pseudonym»». Navnet ditt vil bare stå i dette informasjonsskrivet, som jeg kommer til å oppbevare innelåst. I det forskningen er ferdig og masteren levert - gitt at du ikke har trukket deg fra prosjektet innen den tid - makuleres informasjonsskrivet.
- Svært varsomt bruk av direkte sitat i oppgaven. Dette gjøres slik at jeg ikke har med sitater som gir indikasjoner på hvor lenge du har arbeidet i skolen, alder eller annet som kan være med på å identifisere deg.
- Ingen andre enn meg vil ha tilgang til lydfilen med intervjuet. Denne tas opp via taleopptaksprogram på PC, og lagres deretter på min personlige OneDrive.

- Det transkriberte intervjuet lagres på min personlige OneDrive hos Universitetet i Tromsø. Det kan være aktuelt å skrive ut intervjuene mens jeg skriver oppgaven. Intervjuene vil da låses inn *separat* fra hvor informasjonsskrivet er låst inne hele prosjektet.

Hva skjer med personopplysningene dine når forskningsprosjektet avsluttes?

Prosjektet vil etter planen avsluttes ved slutten av vårsemesteret 2023 (mai/juni). Deretter vil alt datamateriale knyttet til deg – herunder opptaket, det transkriberte intervjuet og informasjonsskrivet – slettes fra OneDrive, og eventuelle fysiske kopier makuleres.

Hva gir oss rett til å behandle personopplysninger om deg?

Vi behandler opplysninger om deg basert på ditt samtykke.

På oppdrag fra prosjektets forfatter og veileder har Personverntjenester hos Norsk Senter for Forskningsdata (NSD) vurdert at behandlingen av personopplysninger i dette prosjektet er i samsvar med personvernregelverket.

Dine rettigheter

Så lenge du kan identifiseres i datamaterialet, har du rett til:

- innsyn i hvilke opplysninger vi behandler om deg, og å få utlevert en kopi av opplysningene
- å få rettet opplysninger om deg som er feil eller misvisende
- å få slettet personopplysninger om deg
- å sende klage til Datatilsynet om behandlingen av dine personopplysninger

Hvis du har spørsmål til studien, eller ønsker å vite mer om eller benytte deg av dine rettigheter, ta kontakt med:

- Herman Hovda Lien ved Universitet i Tromsø (Hli061@uit.no), eller veileder Gunnar C. Aakvaag ved Universitet i Tromsø (Gunnar.C.Aakvaag@uit.no)

Hvis du har spørsmål knyttet til Personverntjenester sin vurdering av prosjektet, kan du ta kontakt med:

- Personverntjenester på epost (personverntjenester@sikt.no) eller på telefon: 53 21 15 00.

Med vennlig hilsen Herman

Prosjektansvarlig

Eventuelt student

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet Er det et kjønnsbias på standpunktverdinger i skolen? og har fått anledning til å stille spørsmål. Jeg samtykker til:


å delta i et løst strukturert kvalitativt intervju

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet

(Signert av prosjektdeltaker, dato)

Vedlegg 2: Vurdering av behandling av personopplysninger fra NSD

12.05.2023, 12:56 Meldeskjema for behandling av personopplysninger

 Sikt

[Meldeskjema](#) / [Er det et kjønnsbias på standpunktvurderinger i skolen?](#) / Vurdering

Vurdering av behandling av personopplysninger

Referansenummer 371965	Vurderingstype Standard	Dato 01.12.2022
----------------------------------	-----------------------------------	---------------------------

Prosjektittel
Er det et kjønnsbias på standpunktvurderinger i skolen?

Behandlingsansvarlig institusjon
UIT Norges Arktiske Universitet / Fakultet for humaniora, samfunnsvitenskap og lærerutdanning / Institutt for samfunnsvitenskap

Prosjektansvarlig
Gunnar C. Aakvaag

Student
Herman Hovda Lien

Prosjektperiode
12.12.2022 - 30.06.2023

Kategorier personopplysninger
Alminnelige

Lovlig grunnlag
Samtykke (Personvernforordningen art. 6 nr. 1 bokstav a)

Behandlingen av personopplysningene er lovlig så fremt den gjennomføres som oppgitt i meldeskjemaet. Det lovlige grunnlaget gjelder til 30.06.2023.

[Meldeskjema](#)

Kommentar
OM VURDERINGEN

Personverntjenester har en avtale med institusjonen du forsker eller studerer ved. Denne avtalen innebærer at vi skal gi deg råd slik at behandlingen av personopplysninger i prosjektet ditt er lovlig etter personvernregelverket.

Personverntjenester har nå vurdert den planlagte behandlingen av personopplysninger. Vår vurdering er at behandlingen er lovlig, hvis den gjennomføres slik den er beskrevet i meldeskjemaet med dialog og vedlegg.

VIKTIG INFORMASJON TIL DEG

Du må lagre, sende og sikre dataene i tråd med retningslinjene til din institusjon. Dette betyr at du må bruke leverandører for spørreskjema, skylagring, videosamtale o.l. som institusjonen din har avtale med. Vi gir generelle råd rundt dette, men det er institusjonens egne retningslinjer for informasjonssikkerhet som gjelder.

TYPE OPPLYSNINGER OG VARIGHET

Prosjektet vil behandle alminnelige kategorier av personopplysninger frem til den datoen som er oppgitt i meldeskjemaet.

LOVLIG GRUNNLAG

Prosjektet vil innhente samtykke fra de registrerte til behandlingen av personopplysninger. Vår vurdering er at prosjektet legger opp til et samtykke i samsvar med kravene i art. 4 og 7, ved at det er en frivillig, spesifikk, informert og utvetydig bekreftelse som kan dokumenteres, og som den registrerte kan trekke tilbake.

Lovlig grunnlag for behandlingen vil dermed være den registrertes samtykke, jf. personvernforordningen art. 6 nr. 1 bokstav a.

PERSONVERNPRINSIPPER

Personverntjenester vurderer at den planlagte behandlingen av personopplysninger vil følge prinsippene i personvernforordningen om:

- lovlighet, rettferdighet og åpenhet (art. 5.1 a), ved at de registrerte får tilfredsstillende informasjon om og samtykker til behandlingen
- formålsbegrensning (art. 5.1 b), ved at personopplysninger samles inn for spesifikke, uttrykkelig angitte og berettigede formål, og ikke behandles til nye, uforenlige formål

<https://meldeskjema.sikt.no/625fb173-0b75-4aa0-9891-4215b6d13c72/vurdering> 1/2

- dataminimering (art. 5.1 c), ved at det kun behandles opplysninger som er adekvate, relevante og nødvendige for formålet med prosjektet
- lagringsbegrensning (art. 5.1 e), ved at personopplysningene ikke lagres lengre enn nødvendig for å oppfylle formålet

DE REGISTRERTES RETTIGHETER

Så lenge de registrerte kan identifiseres i datamaterialet vil de ha følgende rettigheter: innsyn (art. 15), retting (art. 16), sletting (art. 17), begrensning (art. 18), og dataportabilitet (art. 20).

Personverntjenester vurderer at informasjonen om behandlingen som de registrerte vil motta oppfyller lovens krav til form og innhold, jf. art. 12.1 og art. 13.

Vi minner om at hvis en registrert tar kontakt om sine rettigheter, har behandlingsansvarlig institusjon plikt til å svare innen en måned.

FØLG DIN INSTITUSJONS RETNINGSLINJER

Personverntjenester legger til grunn at behandlingen oppfyller kravene i personvernforordningen om riktighet (art. 5.1 d), integritet og konfidensialitet (art. 5.1. f) og sikkerhet (art. 32).

Ved bruk av databehandler (spørreskjemaleverandør, skylagring eller videosamtale) må behandlingen oppfylle kravene til bruk av databehandler, jf. art 28 og 29. Bruk leverandører som din institusjon har avtale med.

For å forsikre dere om at kravene oppfylles, må dere følge interne retningslinjer og/eller rådføre dere med behandlingsansvarlig institusjon.

MELD VESENTLIGE ENDRINGER

Dersom det skjer vesentlige endringer i behandlingen av personopplysninger, kan det være nødvendig å melde dette til oss ved å oppdatere meldeskjemaet. Før du melder inn en endring, oppfordrer vi deg til å lese om hvilke type endringer det er nødvendig å melde: <https://www.nsd.no/personverntjenester/fylle-ut-meldeskjema-for-personopplysninger/melde-endringer-i-meldeskjema>

Du må vente på svar fra oss før endringen gjennomføres.

OPPFØLGING AV PROSJEKTET

Personverntjenester vil følge opp ved planlagt avslutning for å avklare om behandlingen av personopplysningene er avsluttet.

Lykke til med prosjektet!

Vedlegg 3: Intervjuguide

Intervjuguide

Husk å presisere for informanten at det jeg er ute etter, er refleksjoner og tanker fra det som skjer i klasserommet. Det som særlig interesserer meg, er hvordan elever og lærere oppfører seg i klasserommet, hvordan vurderinger formes og hvordan vurderingssituasjoner forløper seg – og da er det veldig fint å snakke med noen som står der ute i klasserommet hver uke. Hvilke erfaringer har informantene her? Det er bare å senke skuldrene, ta seg tid.

START av opptak

1. For å tre oss litt varme; kan ikke du fortelle litt om deg selv som lærer? hvilke fag har du, og hvor lenge du har jobbet i skolen? [få i gang prat]

Spør om informant forstår hva kjønnsvurderingsgapet går ut på får vi går videre. Gutter gjør det relativt bedre enn jentene på eksamen – jentene gjør det fortsatt best, men differansen mellom kjønnene er altså markant mindre på eksamen enn på standpunkt. Vi snakker altså om forskjellen mellom forskjellene her. Har informant noen spørsmål rundt det før vi går videre?

DEL 1 – Forklaringer på fenomenet

1. Hva er dine umiddelbare tanker rundt det jeg fortalte deg nå? Var du klar over at dette fenomenet eksisterte?
2. Hvorfor tror du det er sånn?

DEL 2 – Oppførsel og vurdering

3. Ser du noen forskjeller på hvordan gutter og jenter oppfører seg i klasserommet? (Hvis ja, spør om konkrete eksempler. Spør f.eks om ved vurderingssituasjoner, i diskusjoner, ved selvstendig arbeid).
4. Hvordan oppfører en «sekserelev» seg? (oppføl om motsatt?)
5. Hva er det som former og påvirker vurderingen en lærer gjør av en elevs kompetanse på standpunkt i et fag? (spør gjerne om ulike komponenter)

DEL 3 – Om eksamen

6. a) Det er svært vanlig at årskull markant går ned i karaktersnitt i et fag, fra standpunkt til eksamen. Hvis vi legger gutter og jenter til side – hvorfor tror du, som lærer, at det er så vanlig å gå ned på skriftlig eksamen, generelt sett?

b) – På muntlig eksamen, er det derimot vanlig for årskull at gjennomsnittskarakteren går opp fra standpunkt. Hva er annerledes her?

7. Ser du noen forskjeller på hvilke holdninger gutter og jenter har til eksamen? Oppfører de seg ulikt når eksamen eller store prøver nærmer seg, og på prøvene i seg selv? (kanskje spørre om stress?)

Del 4 – Fenomenet i skolen

8. Er kjønnsvurderingsgapet et problem? I forhold til andre problemer skolen må tenke på og prioritere?
9. Har det kommet noe informasjon om kjønnsvurderingsgapet eller lignende problematikk fra skoleledelsen?
10. Snakkes det om dette eller lignende problematikk i kollegiet?

Del 5 – Presise vurderinger

11. Er det noe den allmenne lærer kan gjøre for å være mest mulig presis i standpunktvurderingene sine?
12. Er det noe skolen kan gjøre for å hjelpe lærere å være mest mulig presis i standpunktvurderingene sine?
13. Nå har vi jo snakket en del om fenomenet... Tidligere spurte jeg deg om hva du tror kan forårsake dette. Har du noe å tilføye om dette nå mot slutten?