

Emneansvarlige på UiB og pedagogiske redskaper i endringstid
- Barrierer og endringsrom

Maria Grönen

Masteroppgave i pedagogikk våren 2017

PED 395



Universitetet i Bergen
Det psykologiske fakultet
Institutt for pedagogikk

Forord

Gjennom studieløpet har jeg stadig kommet tilbake til det Roth foreslår som omdreiningspunkt for pedagogikkens vitenskaps- og forskningsfelt, nemlig ”menneskets forandringmulighet” (Dale, 2005). Menneskets evne til å få kontakt med egen og andres forandringmulighet er for meg en kilde til uendelig inspirasjon. Parallelt med det å skrive denne masteroppgaven som i mangt og mye handler om forandringmulighet og barrierer for endring, har jeg hatt mitt eget endringsløp, og som nok er tilfellet for de fleste masterstudenter, en del barrierer å overkomme. Masterstudent Maria anno mai–17 måtte samskrive denne oppgaven med masterstudent Maria anno oktober -16, selv om jeg i det siste ofte har tenkt at det jammen er beklagelig at man ikke kan skrive en masteroppgave like godt *før* man begynner, som det man kan etter fullført løp. Et ferdig produkt ble det til slutt, men ikke uten betydelig støtte og hjelp fra andre mennesker.

For det første skal min kjære ektemann Daniel, mor og far nevnes. Takk for at dere har tatt dere av barn og hus, og de siste månedene stort sett vært eneansvarlig for at det kom mat på bordet. Takk til søster Silje og Rebecca for korrekturlesning, og til barndomsvenninne Iris og medstudent Madeleine som har delt av sin metodekunnskap og SPSS ekspertise. Takk til institutt for pedagogikk og alle som har vært involvert i å støtte meg og medstudenter gjennom hele studieløpet. Takk til min veileder Yael Harlap, som stødig har guidet meg gjennom denne lange prosessen, minnet meg om sakens kjerne og støttet meg i troen på at dette skal ende godt. Ikke medansvarlig for ferdig produkt, men Yael’s refleksjoner og tilbakemeldinger har vært til stor hjelp og inspirasjon i mitt arbeid. Biveileder Anne-Grete Danielsen skal ha takk for gode råd til løsninger på metodiske problemstillinger, og takk til Robert Gray for godt samarbeid i konstruksjon og gjennomføring av spørreundersøkelsen. Ved veis ende kjenner jeg også på takknemlighet for å bo i et land som gir meg muligheten til å fullføre en mastergrad helt gratis. Med tilhørende frustrasjon og mestring, takk for muligheten til å igjen kjenne på egen forandringmulighet.

Sammendrag

Høyere utdanning i Norge har nylig gjennomgått store endringer og er fortsatt i endringstid når det gjelder bruk av nye undervisnings- og vurderingsformer. Endringer er inspirert av læringsteori og forskning, men er hovedsakelig drevet av utdanningspolitiske reformer som de siste to tiårene har lagt tydelige føringer også på faglig pedagogisk innhold. I lys av dette er målet til denne masteroppgaven å undersøke hvordan emneansvarlige på Universitetet i Bergen (UiB) forholder seg til pedagogiske redskaper som læringsutbyttebeskrivelser og nye vurderingsformer. Med innføringen av Nasjonalt kvalifikasjonsrammeverk for livslang læring (NKR) og læringsutbyttebeskrivelser, sies det at høyere utdanning nå har tatt steget inn i et læringsparadigme, og med det gjennomgått et perspektivskifte fra utdanning til læring. Tidligere forskning peker på at både vurderingsform, og hvorvidt denne sammen med andre undervisningselementer utgjør en meningsskapende helhet, kan spille en avgjørende rolle for hva og hvordan studenter lærer. Valg av vurderingsform og en helhetlig tankegang blir med det sentralt i et læringsparadigme. Samtlige emneansvarlige ble invitert til å delta i en spørreundersøkelse med spørsmål relatert til bruk av læringsutbyttebeskrivelser og vurderingsformer for om mulig belyse hvorvidt dette perspektivskiftet fra utdanning til læring er reflektert i praksis. Resultatene antyder at læringsutbyttebeskrivelser er i bruk utover pålagt krav om å formulere dem og at selv om summative vurderingsformer står sterkt på UiB, er ni av ti også i berøring med formative vurderingsmetoder. Mange respondenter svarer at deres egen vurderingspraksis i stor grad er lik tenkt ideell praksis, men de fleste svarer i noen grad. Respondentenes svar har gjennom statistiske analyser gitt innsikt i sammenhenger mellom demografiske variabler (alder, kjønn og antall års undervisningserfaring) og bruk av pedagogiske redskaper. Resultatene indikerer også samvariasjon mellom lavere grad av enighet i ”endringsskeptiske” utsagn og bruk av flere formative vurderingsmetoder. Resultatene gir samlet sett et inntrykk av en populasjon endringsvillige emneansvarlige, men endringer i høyere utdanning tar tid og barrierer skal overkommes. Undersøkelsen antyder at emneansvarlige flest opplever barrierer når man ønsker å innføre nye vurderingsmetoder, og at disse for noen kan være en effektiv stopper for ønsket endring.

Abstract

Higher education in Norway has the last two decades been through major changes, and are still going through changes related to teaching and assessment practice. These changes are inspired by learning theory and research, but mainly driven by recent educational policy reforms with implications also for educational content. In light of this, the purpose of this study is to examine how course responsible teaching staff at the University of Bergen (UiB) relate to pedagogical tools such as learning outcome statements and new assessment forms. It is said that by introducing learning outcome statements through Norwegian qualification framework (NQF), higher education in Norway has entered a “learning paradigm” that comes with a change in perspective whereby higher education now produce learning instead of merely providing instruction. Research literature suggest that assessment forms, and whether the assessment form is constructively aligned with other educational components, are both influencing how and what students learn. Consequently, choice of assessment form and constructive alignment becomes important in a learning paradigm. A survey was sent out to all course responsible educators with questions related to their use of learning outcome statements and assessment methods that might shed some light on whether their practice assimilated this change in perspective from instruction to learning. The findings suggest that learning outcome statements are used beyond the demand to formulate them and that even though summative assessment methods have a strong position at UiB, nine out of ten teachers have formative assessment methods in their courses. Many respondents express they feel their own assessment practices are largely comparable to an ideal practice, with most responding that their assessments are at least somewhat similar. Findings also suggest that there is a relationship between demographic variables (age, gender, years teaching) and the use of pedagogical tools. The results show covariation between the use of more formative assessment methods and disagreement with statements that are “change skeptical.” All in all, the findings give the impression of a population of course responsible educators that are willing to make changes, though change in higher education tends to be slow and barriers will have to be overcome. This study suggests that most course responsible educators will experience barriers when trying to implement new assessment methods, and for some, these barriers might effectively hinder desired change.

INNHOLD

1. Introduksjon	7
1.1 Pedagogiske redskaper i endringstid	8
1.1.1 Læringsutbyttebeskrivelser.	9
1.1.2 Vurderingskriterier.	11
1.1.3 Vurderingsmetoder.	12
1.1.3.1 Summativ vurdering.	12
1.1.3.2 Formativ vurdering.	13
1.1.3.3 Sameksistens av summativ og formative vurderingsmetoder.	16
1.2 Meningsskapende helhet	17
1.3 Summative og formative vurderingsmetoder på UiB.	19
1.4 Barrierer for endring	20
1.5 Problemstilling og Forskningsspørsmål	24
2. Metode	25
2.1 En kvantitativ pragmatisk tilnærming	25
2.2 Spørreundersøkelse som metode	26
2.2.1 Et kritisk blikk på spørreundersøkelse som metode.	26
2.2.2 Forskningsetiske vurderinger.	28
2.3 Utvalg	29
2.3.1 Kort beskrivelse av respondentene.	29
2.3.2 Frafallsanalyse.	30
2.4 Konstruksjon av spørreskjema	31
2.5 Måleinstrument - Oversikt over variabler	31
2.5.1 Vurderingspraksisvariabler.	32
2.5.2 Demografiske variabler.	35
2.5.3 Endrings skeptiske og endringspositive utsagnsvariabler.	36
2.6 Prosedyre	37
2.6.1 Pilottesting.	37
2.6.2 Administrasjon.	37
2.7 Bearbeiding av data	38
2.7.1 Behandling av Missing data.	39
2.7.2 Signifikansnivå .05.	39
2.8 Statistiske analyser	40
2.8.1 Deskriptiv univariat og bivariat analyse.	40
2.8.2 Bivariat korrelasjonsanalyse.	40
2.8.3 Multivariate analyser	41
2.8.3.1 Faktoranalyse av utsagnsvariabler.	41
2.8.3.2 Stegvis Lineær regresjonsanalyse.	42
2.9. Validitet og Reliabilitet.	43
2.9.1 Reliabilitet.	43
2.9.2 Validitet.	44

3. Resultater	45
3.1 Deskriptiv univariat og bivariat analyse	45
3.1.1 Kjønn, alder, og antall års undervisningserfaring.	45
3.1.2 Nåværende versus ideell praksis.	46
3.1.3 Vurderingsmetoder.	46
3.1.4 Læringsutbyttebeskrivelser og vurderingskriterier.	48
3.1.5 Utsagn knyttet til undervisnings- og vurderingspraksis.	49
3.1.6 Barrierer.	51
3.1.7 Å ønske kursing eller ikke.	52
3.2 Bivariate korrelasjonsanalyser	52
3.2.1 Korrelasjon mellom utsagnsvariabler.	57
3.2.2 Samvariasjon mellom utsagnsvariabler og demografiske variabler.	57
3.2.3. Bruk av Formative vurderingsmetoder.	58
3.2.4 Bruk av summative vurderingsmetoder.	58
3.2.5 Bruk av læringsutbyttebeskrivelser.	59
3.2.6 Å definere og å formidle vurderingskriterier.	59
3.2.7 Barrierer for nye vurderingsmetoder.	59
3.2.8 Å ønske kursing eller ikke.	60
3.3 Multivariate analyser	61
3.3.1 Variasjon i antall formative vurderingsmetoder.	61
3.3.2 Variasjon i antall summative vurderingsformer.	62
4. Diskusjon	64
4.1 Begrensninger med studien	64
4.1.1 Utvalgets representativitet og studiens generaliserbarhet.	64
4.1.2 Variabler og statistisk behandling.	65
4.1.3 Andre begrensninger.	66
4.2 Diskusjon av Resultat	67
4.2.1. Summativ og formativ vurdering i bruk på UiB.	67
4.2.1.1 Kjønn, undervisningserfaring, endringsskepsis og vurderingsformer.	69
4.2.1.2 Vurderingsform, innhold og hensikt.	70
4.2.1.3 Vurderingsmetoder og helhet.	71
4.2.2. Læringsutbyttebeskrivelser, vurderingskriterier og meningsskapende helhet.	72
4.2.3. Å si seg enig eller uenig i endringsskeptiske utsagn.	74
4.2.4. Barrierer - en effektiv stopper for endring?	76
4.2.5 Endringsrom.	78
4.2.6 Videre forskning.	80
4.3 Oppsummering	81
4.4 Konklusjon	82
Litteraturliste:	84
Vedlegg 1 – Godkjenning fra NSD	89
Vedlegg 2 – Invitasjon til spørreundersøkelsen	91

1. Introduksjon

Vurderingsformer i formell høyere utdanning har nærmest vært uendret frem til dette århundre. Utdanning har i stor grad blitt forstått som en serie forelesninger der kunnskap blir overlevert studenten, og deretter testet gjennom kvalifiserende slutteksamen. Høyere utdanning var de dyktige og de bemedlede arena, og læringsarbeidet mellom gitt og anvendt kunnskap var utdanningsinstitusjonen i liten eller ingen grad involvert i. I den klassiske utdanningsmodellen var det forbeholdt de få å lykkes, men i dag skal det store flertallet ha tilgang til kunnskapskapital gitt gjennom høyere utdanning, og utdanningsinstitusjonen skal i større grad bistå mangfoldet av studenter til å lære og å lykkes i sitt studieløp. Høyere utdanningsinstitusjoner eksisterer ikke lenger kun for å produsere kunnskap og å overlevere denne, men også for å produsere læring. Utdanningen skal fortsatt kvalifisere, men ansvaret forlenges fra det å legge kunnskapen på bordet, til å sammen med studenten jobbe for at kunnskapen fører til læring.

Tidligere forskning antyder at uavhengig studenters individuelle ulikheter, tilpasser studenter sine læringsstrategier og studieatferd i tråd med undervisnings- og læringsmiljø (Xiaoyan Wang, Su, Cheung, Wong, & Kwong, 2013). Dette gjelder også i møte med vurderingsformer, som forskning påpeker kan være styrende for studenters læringsprosesser (Graham Gibbs & Simpson, 2004; Ramsden, 1992). En gjennomtenkt bruk av ulike undervisnings- og vurderingsformer er ikke bare viktig sett hver for seg. Constructive alignment (CA) legger et teoretisk fundament for en helhetstankegang fra formulering av læringsutbyttebeskrivelser, gjennom undervisning og til vurdering av oppnådd læringsutbytte. Forskningsresultat indikerer en sammenheng mellom helhetlig samsvar i emnet og større grad av dyp læring, men også mindre grad av overflatisk læring (Xiaoyan Wang, et al., 2013). Mens mer overflatisk læring er knyttet til tilegnelse, lagring, og reproduksjon av kunnskap, er dypere læring knyttet til meningskonstruksjon og forståelse av sammenhenger, kognitive prosesser og overføring til andre eventuelt mer utfordrende oppgaver (Hattie & Timperley, 2007).

Barr og Tagg (1995) skrev for over 20 år siden at amerikansk høyere utdanning var i ferd med å gjennomgå det de kalte et nødvendig og ønsket paradigmatisk perspektivskifte fra undervisning til læring. Dette vil blant annet si at man går bort fra en praksis knyttet til en monologisk overføring av kunnskap gjennom forelesninger, og vender fokuset til studenters læring og hva som kan støtte denne. I dette

læringsparadigmet kan utdanningsinstitusjonen også se seg selv som en lærende, som gradvis lærer hvordan man best kan produsere ønsket læringsutbytte (Barr & Tagg, 1995). Skal man skape en meningsskapende helhet i dette nye læringsparadigmet, må man stadig tenke nytt i alle ledd, også i design av kvalitetssystemer (Gynnild, 2014). Perspektivskiftet fra utdanning til læring har også influert hvordan vi snakker om høyere utdanning på vårt kontinent. Her møter læringsteori siste års internasjonalt, nasjonalt og lokalt reformarbeid i utdanningssektoren (Stensaker, 2008). Nasjonale utdanningspolitiske reformer har i dette århundre et utvidet fokus. Det er ikke bare snakk om styring, organisering og effektivitet, men også innholdet i utdanning. En internasjonal diskurs preger utdanningspolitiske reformer i Norge (Roos & Trippestad, 2015), og man har de siste 10-15 år sett den største omveltningen i måten man tenker og snakker om utdanning, og hvordan den skal gjennomføres (Raaheim, 2016). Den samlede europeiske Bologna-prosessen førte til Kvalitetsreformen i Norge (Kirke-, utdannings- og forskningsdepartementet [KUF], 2001), der endring av vurderingsformer var den mest uttalte pedagogiske konsekvensen (Dysthe, Raaheim, Lima, & Bygstad, 2006; Karseth, Lycke, & Solbrekke, 2008). Arbeidet med et samlet europeisk kvalifikasjonsrammeverk ledet til det nasjonale kvalifikasjonsrammeverket (NKR) med formulering av læringsutbyttebeskrivelser som krav i samtlige studieprogram og emner (Kunnskapsdepartementet [KD], 2011). Med innføringen av NKR hevdes det ofte at høyere utdanning i Norge har tatt steget inn i et læringsparadigme (Haakstad, 2011).

1.1 Pedagogiske redskaper i endringstid

Som allerede nevnt har det vært en økende tendens til at utdanningspolitiske myndigheter gir tydeligere føringer for faglig innhold og hvilke pedagogiske redskaper som skal tas i bruk (Witteck, 2008; Aamodt, Hovdhaugen, & Prøitz, 2014). I det følgende vil begrepet pedagogiske redskaper bli brukt som et samlebegrep for læringsutbyttebeskrivelser, vurderingskriterier og vurderingsformer. Nye undervisningsansatte på et universitet i Norge beskriver tiden de arbeider i som en ”endringstid” (Hanssen, Husebø, & Moen, 2017) og denne overordnede beskrivelsen av ansatte i høyere utdanningsinstitusjoner sin kontekst, vil tas med videre i tekst og diskusjon.

1.1.1 Læringsutbyttebeskrivelser. Læringsutbytte blir i nasjonalt kvalifikasjonsrammeverk for livslang læring (NKR) definert som: ”Det en person vet, kan og er i stand til å gjøre som et resultat av en læringsprosess” (KD, 2011, s. 37). Dette skal uttrykkes i kategoriene kunnskap, ferdigheter og generell kompetanse. Det er bestemt at høyere utdanning skal benytte seg av læringsutbyttebeskrivelser, men hvordan dette skal gjøres i praksis er uklart.

For det første er selve begrepet utydelig. Det eksisterer ingen klar definisjon eller universell enighet om hva læringsutbytte er (Gynnild, 2011; Aamodt, Prøitz, Hovdhaugen, & Stensaker, 2007). Slik begrepet brukes kan det bety både læringsmål og læringsresultat, intendert og oppnådd læringsutbytte (Gynnild, 2011; Haakstad, 2011; Aamodt, et al., 2007). Det engelske begrepet ”learning outcomes” dekker både læringsmål og læringsresultat som oppnås (Aamodt, et al., 2007), og begrepets uklarhet på norsk var noe av kritikken som ble fremsatt fra høyere utdanningsinstitusjoner i prosessen frem mot fastsetting av NKR (Skodvin, Handal, Lycke, & Solbrekke, 2012). Ved gjennomgang av ulike dokument og notat fra Universitetet i Bergen (UiB), som er fokus for denne masteroppgaven, kan man se en viss uklarhet i forholdet mellom læringsmål og læringsutbytte. I ”Handbok for kvalitetssikring av universitetsstudia” vises en figur som beskriver sentrale element i læringsprosessen. Her blir begrepet læringsmål brukt som startpunkt og begrepet læringsutbytte er satt i andre enden av figuren i det som må være betydningen læringsresultat (Universitetet i Bergen [UiB], 2013, s. 2). Dette er i og for seg tydelig nok, men figuren gir lite støtte til dem som måtte stille seg undrende til hva læringsutbyttebeskrivelsene er ment å uttrykke. Med innføringen av læringsutbyttebeskrivelser er nettopp studenters læringsutbytte mål for utdanningen, og læringsutbyttebeskrivelser skal med det uttrykke hva man ønsker at studentene skal kunne (Raaheim, 2016). En hovedutfordring er kanskje likevel inkonsekvent bruk av begrepet læringsutbytte, og Gynnild hevdet i 2011 at begrepene læringsmål og læringsutbytte sidestilles slik de brukes på UiB (Gynnild, 2011). Et eksempel på det kan finnes i et notat fra studieadministrativ avdeling på UiB. Her blir et avsnitt fra ”Kvalitet i utdanning” (Universitetet i Bergen [UiB], 2012), som utelukkende inneholder begrepet læringsmål, referert til og oppsummert slik: ”Særlig blir forholdet mellom læringsutbytte og vurdering vektlagt” (Studieadministrativ avdeling UiB, 2012, s. 2). I kommunikasjon med en vitenskapelig ansatt i Komite for oppretting av nye studieprogram, blir det antydning at man i dag på UiB arbeider aktivt for å ikke sidestille

begrepene læringsmål og læringsutbytte, og at det er begrepet læringsutbytte som skal benyttes (Y. Harlap, personlig kommunikasjon, 18.04.2017).

For det andre er det en utfordring at læringsutbytte skal skrives etter en bestemt mal som ikke skiller mellom ulike program og fagtradisjoner. Å skrive læringsutbytte for et fag som pedagogikk eller filosofi, kan vanskelig skrives på samme måte som for eksempel matematikk eller fysikk. De ulike kunnskapstradisjonene har ulike epistemologisk utgangspunkt og ikke all kunnskap kan enkelt eller ønskelig oversettes til en punktvis liste (Nicol & Macfarlane-Dick, 2006; Raaheim, 2016). Reduksjon i kompleksitet fra NKR sin side, som var nødvendig da rammeverket skulle omfavne samtlige høyere utdanningsinstitusjoner og fagtradisjoner, må muligens gjenopprettes i det lokale arbeidet om læringsutbyttebeskrivelser skal ha epistemologisk gyldighet innenfor ulike fag (Skodvin, et al., 2012). Ved UiB er dette blant annet uttrykt slik: ”Overenstemmelse med kvalifikasjonsrammeverket må i hovedsak vurderes kvalitativt for hvert enkelt studieprogram”, og videre at ”Læringsutbyttebeskrivelser på emnenivå ikke må spesifisere alle elementer i kvalifikasjonsrammeverket (...)” da det ikke er ”(...) hensiktsmessig (eller mulig) å vurdere studentene på alle kriterier innen rammene av hvert enkelt emne” (Studieadministrativ avdeling UIB, 2012, s. 13).

Det er nødvendigvis utfordringer knyttet til operasjonalisering av et begrep som er tvetydig og/eller blir brukt inkonsekvent. Overkommer man begrepsforvirringen, og klarer man å gjøre nødvendig justeringer for å sikre epistemologisk gyldighet i eget fag, kan læringsutbyttebeskrivelser egne seg som pedagogisk redskap for å formulere læringsmål, og som grunnlag for å planlegge undervisning og vurdering av læringsresultat (Skodvin, et al., 2012). I Aamodt et al. (2014) sin studie ble faglig ansatte spurt i hvilken grad læringsutbyttebeskrivelsene spiller en rolle for tilrettelegging av pensum og undervisningsplaner, for egen undervisning og veiledning og i vurdering av studenters prestasjoner. Flertallet svarer i stor eller noen grad under alle svaralternativ, men det er færrest respondenter fra gamle universiteter som UiB som svarer i stor grad, spesielt på svaralternativet tilknyttet vurdering (s. 46). Kvinner mener i større grad enn menn at læringsutbyttebeskrivelsene spiller en rolle, også kontrollert for stilling og alder (s.48).

I st. meld. nr. 16 ”Kultur for kvalitet i høyere utdanning”, kan vi lese at en undersøkelse gjort av NOKUT i 2015 konkluderte med at kun halvparten av læringsutbyttebeskrivelser var i tråd med NKR og samtidig hadde en tydelig faglig profil (Kunnskapsdepartementet [KD], 2017). Med tanke på utfordringer knyttet til

begrepet over, er det kanskje som forventet at læringsutbyttebeskrivelsene ikke har nådd sitt fulle potensial som pedagogisk redskap og at læringsutbyttebeskrivelser til nå ikke har hatt "(...) den strukturerende effekten de er ment å ha" (KD, 2017, s. 49). En tydelig formulert forventning i St.Meld.nr 16 er derfor at læringsutbyttebeskrivelser skal videreutvikles eller revideres og brukes som et "strategisk og faglig verktøy" (KD, 2017, s. 49). En omfattende gjennomgang av samtlige studieprogram, inkludert oppdatering av læringsutbyttebeskrivelser, er i skrivende stund i gang ved UiB (UU-sak 40-15, 18/16). I den anledning tilbyr Program for universitetspedagogikk (UPED) og Læringsstøttegruppen ved studieadministrativ avdeling (SA) denne våren blant annet tre miniseminarer med læringsutbyttebeskrivelser som tema. Her adresseres spørsmål tilknyttet helhetlig tankegang og bruk av læringsutbyttebeskrivelser og vurderingskriterier (Universitetet i Bergen [UiB], 2017a).

1.1.2 Vurderingskriterier. Mens læringsutbyttebeskrivelser skal uttrykke det alle studenter skal kunne etter fullført studieløp, med andre ord alle som ikke stryker, skal vurderingskriterier bestemme grad av måloppnåelse uttrykt gjennom en karakterskala A til F eller bestått / ikke bestått. Høyere utdanning skal siden begynnelsen av 2000-tallet benytte seg av kriteriebasert vurdering. UHR har utarbeidet generelle beskrivelser for prestasjoner til grunn for det enkelte karaktertrinn (Universitets- og høyskolerådet [UHR], 2004). I retningslinjer for bruk av det nasjonale karaktersystemet kan vi lese at "Den verbale beskrivelsen må forstås med referanse til anerkjente faglige standarder for det aktuelle studiet og de kunnskapsmål som er satt for det enkelte program eller emne" (Utdannings- og forskningsdepartementet, 2004). Oversatt til dagens kontekst og i lys av avsnittet over kan vi si at vurderingskriterier skal knyttes til læringsutbyttebeskrivelser, i den grad de er sikret epistemologisk gyldighet. Kriteriebasert vurdering betyr at man vurderer grad av måloppnåelse opp mot på forhånd fastsatte kriterier, i motsetning til en normbasert vurdering der studenters karakter på eget arbeid blir influert av hvordan andre studenter har prestert på samme oppgave (Maki, 2012). Kriteriebasert og normbasert vurdering er to helt ulike prinsipper for vurdering. Likevel sies det at selv kriteriebasert vurdering over tid og for et stort antall skal være i samsvar med en normalfordeling (Gynnild, 2013). Hvordan kriteriebasert vurdering skal gjøres i praksis er fortsatt uklart, og Gynnild (2013) viser i sin artikkel "Kriteriebasert vurdering" – Hva innebærer det i praksis? hvordan vurderer

og fagmiljø sin ulik operasjonalisering og tolkning av forskriften gir seg utslag i ulike slutt karakterer.

På samme måte som med læringsutbyttebeskrivelser er det altså utfordringer knyttet til begrepet i seg selv og i bruk. Tross utfordringer, litteratur på området argumenterer sterkt for at vurderingsprosessen skal være transparent. Dette betyr at kriteriene er eksplisitt definert og tilgjengelig for alle involvert, noe som direkte kan støtte studenters læring (Brown, 2004-05). Når vi i avsnittene under skal se på vurderingsformer og ulike teorier knyttet til dem, er hvordan man forholder seg til vurderingskriterier et viktig element. Brown utfordrer leseren med en guide for hvordan man kan bruke vurdering til å forhindre læring. På toppen av listen er: "Keep the student in the dark about the rules of the game" (Brown, 2004-05, s. 85).

1.1.3 Vurderingsmetoder. Med tanke på høyere utdanningsinstitusjoner sin oppgave i å både produsere læring og være kvalifiserende, kan vi med Kvale i intervju med Smith si at "Eksamener er et knutepunkt for motsigelser i utdanningssystemet" (Smith, 2007a). Krysningpunktet mellom vurdering og læring har blitt gitt stor betydning for kvaliteten på utdanning (Baird, Hopfenbeck, Newton, Gordon, & Steen-Utheim, 2014), men utfordringer kan oppstå i det vurderingen skal oppfylle to ulike hensikter. Når fokuset er på studenters læring blir vurderingsmetoder viktig da tidligere forskning har funnet at vurdering kan være styrende for læringsprosesser (Graham Gibbs & Simpson, 2004; Ramsden, 1992). Dette gjelder læringsprosessen både før, etter og under vurderingen (Dochy, Segers, Gijbels, & Struyven, 2007). Det eksisterer ikke enighet om en felles generell definisjon av hva vurdering er (Evans, 2013), men begrepet brukes i dag om all form for prøving av studenter (Raaheim, 2016). Det er vanlig å kategorisere ulike typer vurdering i hovedkategoriene summativ og formativ (Raaheim, 2016; Slemmen, 2010). Som allerede nevnt har det i høyere utdanning tradisjonelt vært fokus på sluttvurdering, summativ vurdering og vurdering *av* læring (Dysthe, et al., 2006). Disse omtales som oftest i motsetning til vurdering *for* læring, underveisvurdering og formativ vurdering. Når man snakker om innføring av alternative eller nye vurderingsmetoder er det nettopp fordi den avsluttende og kvalifiserende eksamen til nå har vært ensomt regjerende (Raaheim, 2016).

1.1.3.1 Summativ vurdering. Hensikten med summativ vurdering er som ordet i seg selv peker på, en oppsummering og kontroll av hva studentene kan. Om rammen er

4 timer i en klassisk skoleeksamen, en 7 dagers hjemmeeksamen eller en 30 min muntlig eksamen, er felles for summativ vurdering at den foregår *etter* at undervisningen er avsluttet (Raaheim, 2016). Tidligere forskning har funnet at avsluttende og meritterende eksamen kan opprettholde et miljø som kun er i favør for studenter med visse læringsforutsetninger. Dette læringsmiljøet er gjerne dominert av overføring av kunnskap og høyt strukturerte aktiviteter (W Harlen & Deakin, 2002). Tidligere studier antyder at en puggekultur og overflatisk læring fortsatt dominerer i norsk høyere utdanning, og at puggehefter og tidligere eksamensoppgaver brukes flittig før avsluttende eksamen (Thingnes, Stalsberg, & Sitter, 2015). Når det er sagt kan en vurderingsform fylles med både summative og formative element, og sistnevnte skal redegjøres for i neste avsnitt. Eksempler på det er semesteroppgaver og prosjektarbeid som har et formativt element da studenten mottar tilbakemeldinger fra underviser og/eller medstudenter og kan utbedre oppgaven før den leveres til vurdering og karaktersetting (Raaheim, 2016).

1.1.3.2 Formativ vurdering. I formativ vurdering er hensikten å støtte studenters læring, også gjennom vurdering. Det finnes et stort utvalg av litteratur som tar for seg vurdering med hovedfokus på læring og læringsprosesser. Formativ vurdering, underveisvurdering, vurdering for læring (AfL) er alle ulike navn på vurdering som har læring som mål og utgangspunkt. I tråd med perspektivskiftet fra utdanning til læring, har det i de siste tiårene vært en dreining mot vurdering som fremmer læring (Dysthe, et al., 2006), og med det vurdering som støtter studenten i sin selvregulerte læring (Nicol & Macfarlane-Dick, 2006). Der formålet med eksamen er å vurdere hvorvidt studentene har tilegnet seg spesifiserte læringsutbytter, søker en ved formativ vurdering å være læringsfremmende, ved å “trene inn” læringsutbytter gjennom de aktiviteter som skjer underveis (Raaheim, 2016). Hovedfokuset er ikke på kontroll, men å bistå studenten i egen læringsprosess med mål om dypere læring. Dette kan være utfordrende, selv med pedagogisk kompetanse i bagasjen (Smith, 2007b). Viktige moment i formativ vurdering er tilbakemeldinger, selvurdering og medstudentvurdering. Tilbakemelding er en grunnleggende del av læringsfremmende vurdering (Baird, et al., 2014; Nicol & Macfarlane-Dick, 2006; Raaheim, 2016; Slemmen, 2010), og Black & William sin metastudie fra 1998 konkluderer med at tilbakemelding er essensielt for læring og økt læringsutbytte (Black, 2015; Black & Wiliam, 1998). Ifølge Hattie og Timperley (2007) er det muligens også den enkeltfaktoren som betyr mest for læring. Å gi tilbakemelding

er en ferdighet som tar tid å øve opp. Et første steg er uansett å bruke tilbakemeldinger, og Nicol & Macfarlane-Dick (2006) hevder at man kan måle positiv effekt av tilbakemelding selv når den ikke kjennetegnes av det som ofte blir beskrevet som gode tilbakemeldingsprinsipper. Dette forklares med at studenter allerede genererer sin egen tilbakemelding (Nicol & Macfarlane-Dick, 2006), noe som også er en viktig læringsstrategi. Gjennom å få tilbakemelding på eget arbeid kan studentene internalisere en akademisk standard og blir i stand til å gi seg selv tilbakemelding (Raaheim, 2016). Slik selvvurdering er en essensiell komponent i vurdering som har fokus på læring (Black & Wiliam, 1998) og i teorier om selvregulert læring (Baird, et al., 2014; de Bruijn-Smolders, Timmers, Gawke, Schoonman, & Born, 2016; Therese Nerheim Hopfenbeck, 2011). Selvregulert læring består av ulike selvreguleringsprosesser som for eksempel metakognitive strategier. Metakognisjon er studenters refleksjon og læring om egen læring, læringsprosess og læringsstrategier (Black, 2015; Brown, 2004-05; de Bruijn-Smolders, et al., 2016). Forskning har funnet at når det legges til rette for selvvurdering, kan dette ha en signifikant betydning for læring og læringsutbytte idet det effektivt oppmuntrer til dypere læring. Man kan bruke quiz eller lignende for å tilrettelegge for selvvurdering, men også gjennom medstudentvurdering kan studenter utvikle evnen til å gjøre objektive vurderinger mot på forhånd angitte kriterier (Brown, 2004-05; Nicol & Macfarlane-Dick, 2006) (s. 208). Dette vil igjen hjelpe studenten når man skal tilbake til eget arbeid (Black, 2015, s. 169). I tråd med avsnitt om vurderingskriterier over blir det derfor viktig med en transparent vurderingspraksis der kriterier blir eksplisitt definert og gjort tilgjengelig for studentene. Studenter kan kun effektivt vurdere seg selv og andre om de har et tydelig nok bilde av målet for læringen, og hvilke kriterier som beskriver måloppnåelse (Black & Wiliam, 1998; Brown, 2004-05; Slemmen, 2010).

Vurderingskriterier og tilbakemeldinger som støtter selvvurdering henger sammen, og det ene kan vanskelig være læringsstøttende uten det andre. Når det for eksempel i karakterskalabeskrivelsen til en masteroppgave i pedagogikk på UiB står at ”kunnskapen benyttes til selvstendig drøfting og analyse” for karakteren B og videre for karakteren A ”(...) på høyt faglig nivå”, gir dette meg, om ikke veldig detaljert, en viss informasjon om hva som forventes. Hvorvidt en student kan nå dette nivået vil likevel være knyttet til hvorvidt man gjennom emnet og studieløp for øvrig har fått trening i akademisk skriving. Videre at tilbakemeldingene på arbeidet er knyttet opp mot de aktuelle vurderingskriterier. Skriftlige arbeider med tilbakemelding og

vurderingskriterier kan da av studenten samlet sett brukes til å justere egen forståelse av hvordan for eksempel ”selvstendig drøfting og analyse på høyt faglig nivå” ser ut i praksis.

Det har vært en betydelig økning i bruk av tilbakemeldinger i norsk høyere utdanning etter Kvalitetsreformen, men det er lite forskning gjort på omfang og type tilbakemeldingspraksis (Dysthe, et al., 2006). Selv om det eksisterer få dokumenterte sammenhenger mellom obligatorisk skriftlige arbeider med tilbakemeldinger og økt læring (Baird, et al., 2014; Dysthe, et al., 2006), gir Dysthe et al. sitt studie klare indikasjoner på at skriftlige arbeider med tilbakemeldinger har betydning for eksamensresultater, og at studenter blir flinkere til å skrive (Dysthe, et al., 2006).

Undervisningsformer er ikke direkte tema i denne masteroppgaven, men skal kommenteres kort da en formativ vurderingsform i større grad enn den summative overlapper med undervisning, eller kanskje heller utfordrer et skille mellom undervisning og vurdering. Forelesninger som undervisningsform, spesielt monologforelesninger, har blitt kritisert. Blant annet antyder en studie fra 2007 at studenter en time etter forelesninger husket under 10 % av lærestoffet som ble presentert på en tradisjonell forelesning (Hrepic, Zollman, & Rebello, 2007). Begrepet ”medlesning” er blant annet introdusert, som påpeker viktigheten av en aktiv inkludering av studenten (Raaheim, 2016). Hvorvidt forelesninger er passé i det nye læringsparadigmet skal ikke diskuteres her, men lar Thingnes et al. konkludere med at ”Forelesningen er og blir en del av undervisningstilbudet i UH-sektoren, selv om form og verdi diskuteres” (Thingnes, et al., 2015). Å tilsidesette forelesning som undervisningsform er heller ikke et uttalt mål på UiB. Det som vektlegges er refleksjon rundt valg av undervisningsform, og at man skal velge den formen som er mest tjenlig, og ikke for eksempel velger forelesninger av sedvane (UiB, 2012).

Begrepsparet summativ og formativ vurderingsformer eller metoder er ofte brukt, men er ikke alltid like beskrivende eller meningsfulle. Black (2014) hevder at det er en fundamental feil å se formativ vurdering adskilt fra summativ vurdering, og hevder at en slik forståelse er bunnet i en underkjennelse av vurderingens kompleksitet (Black, 2015). Harlen (2005) skriver at skillet mellom de to utelukkende bunner i ulik hensikt med vurderingen. Videre at informasjon om læring, samlet på samme måte, kunne blitt kalt formativ dersom den ble brukt for å støtte læring og undervisning, eller summativ, dersom informasjonen kun ble brukt for karaktersetting og rapportering (s. 208). Når begrepene likevel blir brukt videre i teksten er det med en forståelse av at

vurderingsformene kan fylles med ulikt *innhold*. I neste avsnitt blir vurderingsformens innhold også sett i sammenheng med underviseres implisitte teori, og gitt en rolle i hvorvidt man kan oppfylle *hensikten* om å støtte studenters læringsprosess.

1.1.3.3 Sameksistens av summative og formative vurderingsmetoder.

Når Kvale omtaler eksamen som et knutepunkt for motsigelser i utdanningssystemet (Smith, 2007a), kan det nettopp være fordi fundamentalt ulike hensikter møtes i vurderingspunktet. Kvale sier videre i intervju med Smith (2007) at i den grad læringsfremmede vurderingsmetoder har vært vanskelig å få til i utdanningssystemet, er det kanskje fordi de kan anses å utfordre andre viktige funksjoner som seleksjon og kontroll (Smith, 2007a). Men et fokus på mer formativ vurdering i høyere utdanning trenger ikke å bety en avvisning av den summative vurderingens plass. Knight (2007) anerkjenner den summative vurderingens rolle i kvalifikasjonsøymed og Barnett (2007) konkluderer med at summativ vurdering ikke er et problem i seg selv så lenge den ivaretar den aktive, lærende student. Black (2015) beskriver dette som formativ bruk av summative vurderingsmetoder, og konkluderer med at summative vurderingsmetoder skal være, og må anses å være, en positiv del av læringsprosessen (Black, 2015).

Med fare for å bli arrestert av Black kan det likevel sies å være en utfordring at typiske summative og formative vurderingsformer springer ut fra helt ulike grunntanker. En tradisjonell og summativ vurderingsform er gjerne knyttet til en klassisk utdanningsmodell med en objektivistisk forståelse av verden og en dualisme mellom ”den som skal kunne” og ”det som skal kunnes” (Biggs, 1996). En slik dualisme blir avvist i konstruktivisme, og grunntanken er at studenten er aktivt lærende og konstruerer kunnskap gjennom både individuell og sosial aktivitet. Biggs (1996) skriver videre at underviserens valg baseres på en eksplisitt, men oftest implisitt teori om undervisning og læring. Ligger en objektivistisk teori til bunn for underviserens praksis, vil det gjøre seg utslag i en undervisningshverdag preget av overføring av kunnskap og kan oppmuntre til en puggekultur. Studentenes jobb blir i hovedsak å motta kunnskap, lagre den og hente den frem til eksamen. (s. 347). Når Knight (2007), Barnett (2007) og Black (2015) gir et håp om harmonisk sameksistens mellom summativ og formativ vurderingspraksis, er det kanskje fordi fokuset er på vurderingsformens innhold og hensikt. Med et bevisst reflektert forhold til egen implisitt teori, kan også en vurderingsform knyttet til objektivisme fylles med formativt innhold i tråd med en

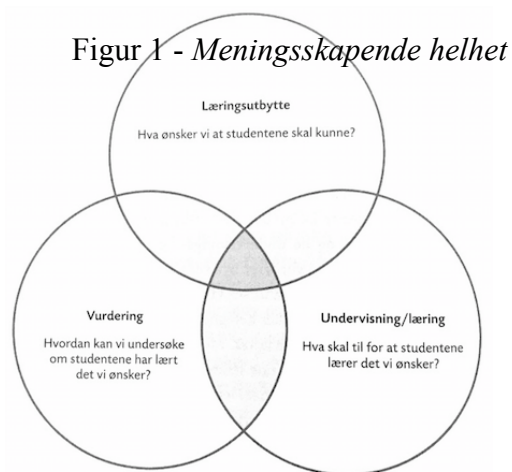
konstruktivistisk tankegang. Oppsummert kan man si at en vurderingsform i seg selv av sedvane kan være knyttet til ulike teoretiske utgangspunkt, men vurderingsformen kan til en viss grad fylles med *innhold* som enten støtter eller ikke støtter opp om vurderingens *hensikt*. I vår utdanningspolitiske kontekst er hensikten å støtte studenters dype og varige læring, men fortsatt også å kvalifisere gjennom kontroll og oppsummering av oppnådd læringsutbytte.

Biggs (1996) hevder at konstruktivisme har en lang historie og er i ferd med å bli den dominerende teorien som ligger til grunn for utdanning, men at det gjenstår å se den som en vanlig anvendt teori i undervisningshverdagen (s.348). Det er nå 20 år siden Biggs skrev sin artikkel, og en konstruktivistisk tankegang er gjenspeilet i utdanningspolitiske dokument på overnasjonalt, nasjonalt og lokalt plan. Det er trolig at den klassiske utdanningsmodellen fortsatt preger praksis i høyere utdanning, (Thingnes, et al., 2015) selv om vi på papiret lever i et læringsparadigme. I læringsparadigmet blir det viktig at teorien som styrer underviserens praksis er lik teorier læringsparadigmet bygger på. Mer spesifikt, blant annet at innholdet i vurderingsformen(e) brukt har formative element, og støtter hensikten om studenters dype læring. Å prioritere vurderingens hensikt med å fremme læring først, vil også tjene andre hensikter med vurdering, for eksempel kvalifikasjon (Smith, 2007b). Følgelig kan det tenkes at en forutsetning for en god sameksistens mellom summative og formative vurderingsmetoder, er at grunntanken bak det formative element må ha forrang og samlet styre praksis. En helhetstankegang ved bruk av både klassiske summative og nyere formative vurderingsmetoder blir dermed også viktig.

1.2 Meningsskapende helhet

Raaheim (2016) foreslår i sin bok ”Eksamensrevolusjonen” å oversette Constructive alignment (CA) til “meningsskapende samsvar” eller “meningsskapende helhet” på norsk. Figur 1 til høyre er lånt med Arild Raaheim sin tillatelse fra samme bok og illustrerer hovedtanken i CA (Raaheim, 2016, s. 86).

Som det engelske navnet tilsier ligger konstruktivisme til grunn, noe som blant annet gjør



sitt utslag i et fokus på studentaktive undervisnings- og vurderingsformer. "Alignment" viser til tanken om harmoni mellom de ulike delene av undervisnings- og vurderingsprosesser (Biggs, 1996; Raaheim, 2016). Et meningsskapende samsvar i vår nasjonale og lokale kontekst er at det som er uttrykt gjennom læringsutbyttebeskrivelsene skal reflekteres i undervisningsform, vurderingsform og vurderingskriterier. I en slik tankegang er det ikke delene hver for seg, men helheten som er en forutsetning for at ønsket læringsutbytte skal kunne oppnås (Biggs, 1996). Som allerede referert til i første avsnitt ble det i en studie gjennomført av Wang et al. (2013) påvist en sammenheng mellom mer dypere læring og mindre overflatisk læring i emner som var helhetlig planlagt og gjennomført.

I høyere utdanning skal læringsutbyttebeskrivelser og vurderingskriterier nettopp uttrykke kunnskaper og evner knyttet til dypere læring (KD, 2011), og ikke bare "gjengivelse av en pugget power point slide" (UiB, 2012, s. 33). Men å vurdere slike komplekse læringsutbyttebeskrivelser krever for det første en reell investering, og det i et klima preget av minskende ressurser, for det andre at "assessors need evidence of achievement that comes from a variety of modules, just as students need in order to construct robust achievements" (Knight, 2007, s. 82). Dette er et sterkt argument for bruk av varierte vurderingsmetoder, som også er blitt en realitet på mange studiesteder etter Kvalitetsreformen (Michelsen & Aamodt, 2007). Selv om nye vurderingsmetoder er tatt i bruk, er den tradisjonelle slutteksamen i stor grad beholdt, resultatet er med det en volumøkning i vurdering av studenters prestasjoner (Michelsen & Aamodt, 2007). Med det må man også være på vakt for eventuell overvurdering og for stort arbeidspress. En helhetlig tankegang, der ulike vurderingsformer drar i samme retning, kan muligens motvirke dette.¹ I st.meld.nr 16 kan vi lese at "Tradisjonell skriftlig eksamen kan med fordel suppleres med andre vurderingsformer (...)" (KD, 2017, s. 56), men også at det bør sikres en sammenheng mellom alle delene i læringsprosessen (KD, 2017, s. 24). Ved UiB er bruk av varierte og tilpassede vurderingsformer i en meningsskapende helhet løftet frem både på emne- og programnivå. Kvalitetssikringsarbeidet på UiB er vist gjennom "Handbok for kvalitetssikring av universitetsstudia".² I denne kan man lese at UiB har en løpende evaluering av

¹ Ved bruk av flere vurderingsmetoder i samme emne eller i emner som går parallelt, er det fare for eventuell "overevaluering". Lydhørhet i studentevalueringer er viktig med tanke på å unngå dette (Dysthe, et al., 2006)

² Kvalitetssikringssystemet på UiB fikk sitt godkjentstempel i 2014 av Nasjonalt Organ for Kvalitet i Utdanning (Nasjonalt organ for kvalitet i utdanning [NOKUT], 2014).

studieprogram og emner. En tredjedel av alle emner det blir undervist i skal evalueres hvert år, og dette har emneansvarlig ansvar for. I evalueringen skal man blant annet komme med ”forslag til tiltak for vurdering av sammenhengen mellom fastsett læringsutbytte og undervisnings- og vurderingsformene i emnet” (UiB, 2013, s. 8). I ”Kvalitet for utdanning” blir det også poengtert at det skal være innbyrdes forhold og sammenheng mellom alle deler fra utvikling av læringsmål, læringsprosess med undervisning og vurdering, til læringsutbytte (UiB, 2012, s. 30). En antagelse i ”Quality assurance” (QA) og til grunn for kvalitetssikringssystem, er at systemet i seg selv aktivt og indirekte kan føre til endring i undervisning og læring. Få studier kan bevise at en slik sammenheng eksisterer (Mårtensson, Roxå, & Stensaker, 2014). Gynnild (2014) påpeker også at kvalitetssikringssystem muligens ikke er tilpasset et paradigmeskifte fra undervisning til læring og at kvalitetssikringssystemet kan oppmuntre til en rutinemessig gjennomgang basert på pålagte krav, uten dokumenterte forbedringer men med betydelig ressursforbruk (Gynnild, 2014). At dette til en viss grad kan være beskrivende for UiB antydes i ”Kvalitet i utdanning” (UiB, 2012). Her skrives det blant annet at evalueringene må ha et reelt innhold og at formålet ikke må være å ”tilfredsstille NOKUT” (s.36). At varierte vurderingsformer og en meningsskapende helhet er adressert i kvalitetssystemet på UiB, trenger med andre ord ikke bety at kvalitetssystemet er designet for å sikre det, eller at man har grunnlag for å tro at det i realiteten fører til ønsket praksis.

1.3 Summative og formative vurderingsmetoder på UiB.

Selv om vurdering som fremmer læring har hatt fokus både nasjonalt og internasjonalt i både ett og to tiår, er det en del som tyder på at overgangen til en slik vurderingspraksis kan ta tid (Therese N. Hopfenbeck, Flórez Petour, & Tolo, 2015; Nicol & Macfarlane-Dick, 2006). Når det er sagt påpeker Dysthe et al. at visse elementer som skriftlige arbeider med tilbakemeldinger og mappevurdering etter Kvalitetsreformen har blitt implementert raskt, muligens også for raskt. En mangel på forståelse av hva mappevurdering innebar, resulterte blant annet i at en del undervisere på grunn av ressursknapphet reduserte innholdet i formen til det ugjenkjennbare, eller avviste mappevurdering som ugjennomførbar i høyere utdanning (Dysthe, et al., 2006). Gjennom notatet fra Studieadministrativ avdeling ved UiB som hadde til formål å beskrive vurderingsformer ved UiB i lys av kvalifikasjonsrammeverket (2012), får vi et

inntrykk av bruk av summative og formative vurderingsmetoder i perioden 2008 til 2012. Her blir det slått fast at summative vurderingsformer jevnt over har høyest frekvens, og at det også var en *økning* i antall emner som kun bruker skriftlig eksamen i samme periode, både på lavere og høyere studienivå. Omfanget av mappevurdering og skriftlige karaktersettende oppgaver reduseres derimot. I lys av Dysthe sitt funn over, er dette interessant, og kan tyde på at noe av det samme har funnet sted ved UiB. Når det er sagt økte obligatoriske arbeidskrav som ikke skal karaktersettes i perioden (UIB, 2012). I notatet konkluderes det med at ”Karaktergivende vurdering er altså fremdeles i hovedsak summativ, og vurdering som gir studentene tilbakemeldinger i løpet av studiet, reduseres noe” (s.4). Undersøkelsen som ble gjort skilte ikke på type krav, så dette kan være skriftlige arbeider, med eller uten tilbakemeldinger, men også krav som ”oppmøte”. Det blir et problem, med tanke på ønsket endring, hvis det er en trend at bruk av skriftlige arbeider med tilbakemeldinger reduseres, men det trenger ikke i seg selv være et problem at de skriftlige arbeidene ikke er karaktergivende. Gibbs konkluderer med at hovedsaken er at det blir gitt (gode) tilbakemeldinger, fra underviser og/eller medelever, ikke at arbeidet karaktersettes (G. Gibbs, 2010). Dysthe et al. fant at undervisere og studenter opplevde at skriftlige arbeider med tilbakemelding førte til bedre skriving og økt læringsutbytte, men at studenter var kritiske til at man ikke fikk tilstrekkelig uttelling på karakteren for den tid og innsats som var lagt inn i arbeidet underveis i semesteret (Dysthe, et al., 2006, s. 46). Karaktersatt eller ei, i tråd med redegjørelsen av sameksistens mellom summative og formative vurderingsmetoder, og i lys av målet om meningssskapende samsvar, er det likevel et problem om de skriftlige arbeidene underveis ikke har sammenheng med den summative karaktergivende eksamen.

1.4 Barrierer for endring

Utdanningssystem med lange og sterke tradisjoner i endring, og det er i økende grad eksterne føringer på faglig pedagogisk innhold. Hvorvidt og i hvilken grad det har vært en reell ekstern innflytelse på innholdet er fortsatt usikkert (Aamodt, et al., 2014). Det er også slik at selv om læringsteori beskriver et paradigmatisk perspektivskifte fra utdanning til læring, og utdanningspolitiske dokument reflekterer det samme, har ikke nødvendigvis teorien som styrer undervisnings- og vurderingspraksis gjennomgått det samme perspektivskiftet. I avsnittene under beskrives potensielle barrierer på system-

og individnivå som kan være med på å belyse hvorfor endringsarbeidet i høyere utdanningsinstitusjoner tar tid. De to første punktene handler om skepsis til, eller konfliktlinjer i endringsprosjektet i seg selv, den tredje om kunnskap og implisitt teori og det fjerde og siste om systembarrierer.

For det første kan det tenkes at endringstregghet henger sammen med skepsis til et ”neoliberalistisk prosjekt”, her med tanke på innføring av kvalifikasjonsrammeverk og læringsutbyttebeskrivelser spesielt. Med det sagt hevder Skodvin, Handal, Lycke og Solbrekke (2012) at universitetspedagoger på UiO sine innsigelser mot endringene hovedsakelig var pedagogiske, og ikke rettet mot Bologna-prosessens ideologiske overbygning. Dette betyr likevel ikke at universitetspedagogene var uten ideer på overordnede nivåer, ”(...) men at slike begrunnelser sjelden ble gjort eksplisitte” (s.26). Det kan tenkes at det faktum at endringene i høyere utdanning i Norge er påvirket av globale utviklingstrekk, kan styrke en skepsis til nye pedagogiske redskaper i tillegg til eventuelt pedagogisk begrunnet skepsis.

For det andre kan det at endringene kommer ”ovenfra” og ”utenfra” være et moment som kan forklare endringstregghet. Når det gjelder endringene som Kvalitetsreformen førte med seg (mer skriving, tilbakemelding og nye vurderingsformer), konkluderer Dysthe et al. at disse endringene møtte et behov ansatte i høyere utdanningsinstitusjoner selv opplevde, og at det derfor var lite prinsipiell motstand mot de pedagogiske sidene ved reformen. Men da endringene ble oppfattet styrt gjennom en ovenfra-ned prosess var det ”ikke stor og utbredt entusiasme for å gå i gang med endringene” (Dysthe, et al., 2006, s. 54). Utdanningspolitiske føringer kommer i dag fra nye aktører og organisasjoner. Hva som er god kvalitet, og hvilke pedagogiske redskaper som skal brukes, blir i stor grad definert av et ytre kvalitetssystem. Utdanningsinstitusjonenes egne kvalitetssikringssystem reflekterer gjerne det samme, men dette kvalitetssynet harmonerer ikke nødvendigvis med andre syn på hva kvalitet er innad i organisasjonen. Dette fører til ulike konfliktlinjer som alle påvirker pedagogiske endringsarbeid (Kårhus, 2011; Stensaker, 2008). Aamodt et al. (2014) spurte faglig ansatte blant annet om hvorvidt kvalitetssystemet bidrar til bedre kvalitet i studieprogrammet. Relativt mange svarte verken eller og omtrent like mange er uenige som enige. Aamodt et al. skriver at dette tyder på at kvalitetssikringssystemene ikke er spesielt godt forankret i faglig virksomhet (s. 43) og videre at holdningen til kvalitetssystemet må betegnes som lunken, spesielt ved gamle universitet (s. 49).

For det tredje omhandler muligens den viktigste barrieren konvensjoner, innarbeidede rutiner og tillærte tankemønstre (Raaheim, 2016), men også hvilke pedagogisk kunnskap undervisere har. Et moment som kun kort skal nevnes her er at innføring av nye formative vurderingsformer fører med seg mindre grad av kontroll over hvordan studenter bruker eller misbruker ressurser for å løse en oppgave. Bekymring for fusk kan med det bli styrket hvis man fokuserer på tapet av kontroll som en tradisjonell summativ vurderingsform som skriftlig skoleeksamen i større grad sikrer. Bekymring for fusk er en reell utfordring basert på det forskning antyder er et reelt problem (Raaheim, 2016) og man må tenke nytt om man skal håndtere aktuelle problemstillinger knyttet til dette.³ Forskning indikerer at skal man få til en endring i undervisning og læring, må det først bli en endring i tanker og ideer rundt det samme (Watkins, Dahlin, & Ekholm, 2005, s. 283). Dette gjelder for eksempel en økt bevissthet rundt hvordan vurdering påvirker studenters læringsprosess. Hanssen, Husebø og Moen (2017) nevner implisitte teorier og forforståelse av rollen som sentral når nye ansatte skal tiltre en stilling som underviser i høyere utdanning. Denne forforståelsen er gjerne basert på det klassiske bildet av foreleseren ved kateteret, med eller uten Power Point. Man kan bli ”sjokket” ut av en slik type forforståelse i møte med nye perspektiv, bilder og begreper (Hansen, Husebø og Moen, 2017). Et kurs i universitetspedagogikk kan utgjøre en slik ”vekker”. I en studie fra Storbritannia (2013) blant universitetsansatte som nettopp hadde fullført, eller var i gang med et slikt kurs, ble det vist at deltakerne gjennom kurset hadde endret syn på vurdering og uttrykte at egen vurderingspraksis kunne bli bedre. Institusjonstilhørighet, fagdisiplin, kjønn, undervisningserfaring og hvorvidt undervisere hadde fullført kurs i universitetspedagogikk var indikatorer for eller korrelerte statistisk signifikant med ønsket vurderingspraksis og opplevde barrierer (Norton, et al., 2013).

Ulike typer systembarrierer og mangel på insentiv og støtte er fjerde punkt. Over 50 % av deltakerne i Norton et al. (2013) sin undersøkelse svarte at det var praktiske restriksjoner eller barrierer for å utvikle eller endre vurderingspraksis. Studien viste også en oppfatning blant undervisere om at det var lite insentiv til innovasjon i undervisning og vurdering, selv om man altså selv opplevde at det var rom for bedring i egen vurderingspraksis (Norton, et al., 2013). Albers (2009) skriver at det ikke kun er mangel på insentiv, men den institusjonelle konteksten kan også effektivt hindre

³ Viser til Dysthe et al. (2006) og Raaheim (2016) for en utvidet diskusjon rundt dette.

underviseres forsøk på å være innovativ i undervisnings- og vurderingspraksis. Institusjonen som helhet kan med andre ord bidra til å fastholde både underviser og student i gamle mønster (Albers, 2009). Incentiv kan blant annet være merittering av undervisningsarbeid, noe som UiB tok med i sitt innspill til Kunnskapsdepartementet i arbeidet mot den nye stortingsmeldingen "Kultur for kvalitet i høyere utdanning" (KD, 2017), og opplevde å bli hørt (Universitetet i Bergen [UiB], 2017b). Meritteringsordninger kan være en viktig del av en kvalitetskultur som løfter frem god undervisnings- og vurderingspraksis. Wang og Hurley (2012) og Knight & Trowler (2000) hevder at i takt med endringene er det av avgjørende betydning å skape en kultur for vurdering i institusjonen med kollektiv innsats mot ønsket endring. I notatet "Kvalitet i utdanning" (UiB, 2012) nevnes det i sammenheng med utvikling av en kvalitetskultur blant annet tillit til undervisere, deres prestasjoner og yrkesstolthet samt at forslag til kvalitetsfremmende tiltak skal bli ønsket velkommen (s. 37). Men selv med incentiv som merittering og videreutvikling av en kultur som støtter underviseren i endringsarbeid er det mulig at den støtten må forlenges i form av tilførte ressurser. Spørsmål blir ofte reist om hvorvidt forskning og undervisning er aktiviteter som er konkurrerende med tanke på tidsbruk og oppmerksomhet (Aamodt, et al., 2014). Med tanke på endring mot vurderingspraksis var mangel på ressurser barrieren flest respondenter i Norton et al. sitt studie påpekte (2013). Dette gjaldt blant annet mangel på tid, store studentgrupper og kostnader. Skepsis til eller konfliktlinjer i endringsprosjektet, implisitt teori som ikke harmonerer med innføring av nye metoder og til slutt systembarrierer, har tidligere forskning vist eksisterer i høyere utdanningsinstitusjoner. Det er rimelig å anta at disse barrierene også er å finne på UiB.

Oppsummert kan vi si at de løse trådene i læringsteori og forskning ikke hovedsakelig er knyttet til *hvorfor* det er viktig eller riktig å endre undervisning - og vurderingspraksis hvis fokuset på studenters læring er overordnet. De løse trådene er blant annet knyttet til *hvorvidt* og *i hvilken grad* "nye" pedagogiske redskaper blir tatt i bruk, *hvordan* man skal leve ut ønsket om endring i praksis, *kvaliteten* på de endringene som blir gjort, og begrunnelser og barrierer for nåværende eller ønsket praksis. Nordkvelle skriver i 2013 at Kunnskapsdepartementet for få år siden dokumentert at høyere utdanning er en sektor med lavest innsats på forskning om egen virksomhet (Nordkvelle, 2013). Denne masteroppgaven kan i så måte bidra med å rette søkelyset mot noen av disse spørsmålene, og i neste avsnitt vil problemstilling og

forskningsspørsmål bli presentert. Spørsmål om innhold i de pedagogiske redskapene i bruk og kvalitet på eventuelle endringer, vil ikke kunne besvares gjennom denne undersøkelsen. At kvalitet og innhold likevel er adressert i introduksjonen, er gjort fordi spørsmål knyttet til dette er en viktig del av konteksten når denne undersøkelsens forskningsspørsmål kan få sitt svar eller blir belyst.

1.5 Problemstilling og Forskningsspørsmål

Problemstillingen er *”Hvordan forholder emneansvarlige på UiB seg til nye pedagogiske redskaper som læringsutbyttebeskrivelser og formative vurderingsmetoder?* Problemstillingens hovedverb “forholder” er valgt med tanke på at det inkluderer både emneansvarliges bruk av redskapene i tillegg til grad av enighet i utsagn knyttet til redskapene og undervisnings- og vurderingspraksis for øvrig.

For å belyse problemstillingen ble følgende forskningsspørsmål formulert:

1. *Hvorvidt og/eller i hvilken grad er emneansvarlige i berøring med ulike vurderingsmetoder, læringsutbyttebeskrivelser og eksplisitte vurderingskriterier?*
2. *I hvilken grad opplever emneansvarlige barrierer som hindrer implementering av nye vurderingsformer, og hvilke barrierer oppleves?*
3. *Hvordan samvarierer bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderingspraksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring)?*
4. *I hvilken grad kan bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderingspraksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring) forklare variansen i bruk av formative og summative vurderingsmetoder?*

2. Metode

Dette kapitlet er en redegjørelse av metodisk tilnærming og forskningsdesign. Her vil jeg begrunne valg av metode og gjøre forskningsetiske vurderinger. Videre vil jeg vise hvordan datamaterialet er bearbeidet statistisk og hvilke grep som er tatt med tanke på reliabilitet og validitet i undersøkelsen.

2.1 En kvantitativ pragmatisk tilnærming

Valget mellom kvantitativ og kvalitativ metode kan bli sett på som et pragmatisk valg eller som et valg av vitenskapsfilosofisk standpunkt (Ringdal, 2013, s. 104). Kvantitativ metode er ofte blitt knyttet til positivisme som fremholder et 1=1 forhold mellom verden og vår persepsjon av den. Få kvantitative forskere vil kjenne seg igjen i et rent positivistisk verdenssyn (Silverman, 2001, s. 38), og få som bruker kvantitativ metode i utdanningssammenheng ser sosiale hendelser som kun objektive realiteter (Danielsen, 2013, s. 139). Som for de fleste kvantitative undersøkelser har heller ikke denne et mål om å produsere en “science of laws (like physics)”, målet er rett og slett å generalisere basert på kritisk bearbeidelse av data (Silverman, 2001, s. 38). Mitt verdenssyn er tettere knyttet opp mot sosio-konstruktivisme som ofte er til grunn for kvalitative undersøkelser. Dette betyr en anerkjennelse av at menneskelig erfaring og persepsjon er mediert gjennom historie, kultur og språk (Willig, 2008), og med det en ydmyk tilnærming til ”sannhet”. Jeg vil derfor ikke påstå at min tolkning av innsamlet empiri representerer den éne sanne virkelighet. Når jeg har valgt en kvantitativ tilnærming er det først og fremst et pragmatisk valg, og forskningsspørsmål er satt over metode og verdenssynet metoden bygger på (Tashakkori & Teddlie, 1998, s. 21). Undersøkelsens mål er å få oversikt og større innsikt i deler av emneansvarlige på UiB sin vurderingspraksis. For å få oversikt er det hensiktsmessig å spørre mange, om relativt mye (Befring, 1998). På bakgrunn av det overnevnte har jeg ut fra et konstruktivistisk og pragmatisk verdenssyn utarbeidet et deskriptivt-analytisk kvantitativt design med spørreundersøkelse som metode. Undersøkelsen er utforskende, og selv om kvantitative metoder er mer stringente av natur er det også her rom for en viss fleksibilitet og tilpasning (Creswell, 2014). Min forståelse av metoden er at det er et redskap som kan støtte en hensikten med undersøkelsen, nemlig å få svar på eller belyse forskningsspørsmålene. Metoden har likevel en ramme som jeg må operere innenfor og

med den følger både muligheter og begrensninger. Disse skal adresseres i påfølgende avsnitt.

2.2 Spørreundersøkelse som metode

Spørreundersøkelser er en vanlig metode i samfunnsvitenskapene (Svartdal, 2009) og i utdanningskontekst (Danielsen, 2013). Informasjon man kan få fra spørreundersøkelser er demografiske data og atferd, men også holdninger og meninger (Fekjær, 2016; Svartdal, 2009). I tillegg til oversikt, som gjør spørreundersøkelse til en hensiktsmessig metode, er målet også å få større innsikt i emneansvarlige sin vurderingspraksis på UiB. I denne sammenhengen forstår jeg innsikt som økt forståelse av sammenhenger mellom for eksempel ønsket vurderingspraksis og barrierer for endring, eller berøring med ulike vurderingsmetoder og grad av enighet i utsagn knyttet til pedagogiske redskaper og praksis. Kvalitetstegnet og validiteten i en spørreundersøkelse er det samme som i forskningsprosessen generelt; om den gjør det den er ment å gjøre og om den belyser forskningsspørsmålet på best mulig måte (Oppenheim, 1992). Spørsmålene ble utformet på bakgrunn av relevant teori og forskning, men bygger også på den erfaring min veileder Y. Harlap samt medforfatter av spørreundersøkelsen R. Gray har gjennom sine møter med undervisere på UiB. Spørsmålsutvikling har med andre ord vært både teoristyrte og mottakerstyrte ut fra et visst råmateriale (Haraldsen, (1999, s. 97).

2.2.1 Et kritisk blikk på spørreundersøkelse som metode. En fallgrube ved en kvantitativ metode som spørreundersøkelser er en tendens til empirisme, som kan bety at man tenker at “data are left to speak for themselves” (Denscombe, 2003, s. 28). Med et konstruktivistisk og pragmatisk utgangspunkt, og med et kritisk blikk samt refleksjon rundt dette, har jeg forsøkt å balansere mellom ydmykhet og tydelighet i tolkning av datamaterialet. Spørreundersøkelsen som helhet inkluderer både åpne og lukkede spørsmål, men til grunn for denne masteroppgaven har jeg trukket ut en samling lukkede spørsmål som er relevant for å belyse problemstillingen. At det er lukkede spørsmål, letter analysearbeidet (Oppenheim, 1992). Lukkede spørsmål er også enkle for respondenter å svare på, men bare i den grad man kjenner seg igjen i svaralternativene og svarkategoriene dekker aktuelle alternative (Ringdal, 2013). Dette er samtidig også en av utfordringene med lukkede spørsmål. Denscombe skriver at

“Pre-coded questions can bias the findings towards the researcher’s rather than the respondent way of seeing things” (Denscombe, 2003, s. 160). Selv om våre spørsmål både er teoristyrte og mottakerstyrte, har jeg ikke oversikt over respondentenes virkelighet. Både spørsmålsformulering og faste svaralternativer, speiler *min* persepsjon av respondentenes virkelighet. For å bøte på dette ble svaralternativet ”Annet” i de aller fleste spørsmål inkludert, og respondenten fikk også mulighet til å gi kommentarer i et åpent felt merket ”ytterligere kommentarer” etter de fleste spørsmål. Planen var i utgangspunktet analysere disse kommentarene induktivt og la de stå som mer fyldige eksempler på det tallmaterialet viser, eller som representanter for en alternativ tolkning. Disse kommentarene kunne vært en berikelse i tolkning og diskusjon, men det viste det seg at planen ikke lot seg gjennomføre innenfor rammen av denne masteroppgaven. Svar oppgitt under ”Annet” er derimot inkludert, hvordan blir redegjort for lenger nede.

I spørreundersøkelser er mangel på detalj og dybde også en begrensning (Denscombe, 2003). Når spørreundersøkelsen i tillegg er selvadministrert, er ikke forskeren tilstede og mister dermed muligheten til oppfølgingsspørsmål ved interessante spørsmål eller om respondent har misforstått spørsmålet. Som ved alle innsamlingsmetoder som baserer seg på opplysninger gitt av forskningsobjektet selv, er det også en mulighet for unøyaktighet eller uærlighet i svarene. Dette gjelder særlig i situasjoner der sensitive tema tas opp. Spørreundersøkelsens tema er ikke spesielt sensitivt, men det kan være en fare for at respondenten, spesielt i situasjoner der man ikke har tenkt over spørsmålet før, eller er usikker på svaret, dreier svaret i den retning man tenker er sosialt akseptert (Haraldsen, 1999).

Spørreundersøkelsen er del av et større prosjekt som administreres av Gray, og flere kvalitative intervjuer skal gjennomføres. 50 respondenter har gjennom spørreundersøkelsen gitt oss sin e-post adresse for å bli kontaktet for et eventuelt oppfølgingsintervju. Basert på et utkast fra Gray som viste hvilke spørsmål han ønsket å inkludere, fikk jeg frihet til å legge til eller fjerne spørsmål, og konstruere spørreskjemaet og invitasjon til undersøkelsen i SurveyXact etter mine ønsker. Vi var begge med på ferdigstillingen av den engelske versjonen, og står med det begge som forfattere. Basert på den engelske versjonen oversatte jeg deretter spørreundersøkelsen til norsk. I oversettelsesarbeid er det mange valg å ta med tanke på setningsoppbygging og vektning og plassering av ord. Jeg valgte å legge mest vekt på å sikre konsistens mellom den engelske og norske versjonen, i noen tilfeller til dels på bekostning av språklig flyt i den norske versjonen. Spørsmålene i undersøkelsen berører temaene

vurderingsformer, utsagn knyttet til undervisning og vurdering som respondentene ble bedt om å ta stilling til, hvilke typer kurs undervisere kunne tenke seg og for hvilken grunn, samt den nye digitale læringsplattformen Mitt UiB. Ikke alle spørsmål er relevant for å belyse problemstilling og forskningsspørsmål formulert for denne masteroppgaven. Spørsmålene jeg har valgt ut blir presentert under overskriften måleinstrumenter i et senere avsnitt.

2.2.2 Forskningsetiske vurderinger. I en forskningsprosess er det mange valg som skal tas, og noen av dem har allerede blitt redegjort for. Etter hvert som arbeidet med undersøkelsen skred frem, ble det også tydelig at det var flere valg å ta stilling til enn de jeg hadde klart å se for meg i starten av prosjektet. I visse tilfeller var jeg ikke en gang bevisst at det var et valg som ble tatt før gjennom refleksjon post faktum. Et eksempel er knyttet til oversettingsarbeidet nevnt over, valg av statistiske analyser som ble gjort, og hvilke funn jeg forfulgte. Da problemstillingen var vid, forskningsspørsmålene relativt åpne og datamaterialet relativt stort, hadde jeg mulighet til å forfølge flere ulike spor. De veiene jeg så og valgte, er influert av min personlige interesse og forforståelse i møte teori og empiri. Med det anerkjenner jeg at mitt epistemologiske og personlige ståsted har influert og formet forskningsprosessen (Willig, 2008, s. 17). En annen forsker kunne valgt andre veier og belyst andre deler av datamaterialet. Forskningsetikk handler også om en anerkjennelse av egen feilbarlighet, om hvordan man går frem for å produsere resultater og sannferdige presentasjoner av dem (Ilstad, 1989). Blant annet å inkludere resultater som ikke støtter ens egne hypoteser, noe som også er knyttet til respekt for respondentene og deres svar. Det handler om å balansere mellom å komme så nær sine egne data at man blir kjent med dem, men også ha såpass avstand at man kontinuerlig stiller kritiske spørsmål til egen prosess og gjør aktiv feilsøking. Forskningsetikk handler ikke bare om gode intensjoner, men er også knyttet til kvaliteten og validiteten i arbeidet utført. Man må med det aktivt jobbe for å redusere antall utilsiktede feil, blant annet gjennom trening og opplæring i metode og analysearbeid (Befring, 1998). For egen del la metodekurset i masterløpet et godt grunnlag, men et selvregissert metodekurs der jeg har dratt på styrker til medstudenter, veiledere og andre, opplevdes nødvendig i møte med denne oppgaven. Da spørreundersøkelsen henvender seg til voksne og temaet er ikke sensitivt, er ikke vanlige forskningsetiske vurderinger rundt det å beskytte respondenter relevant utenom

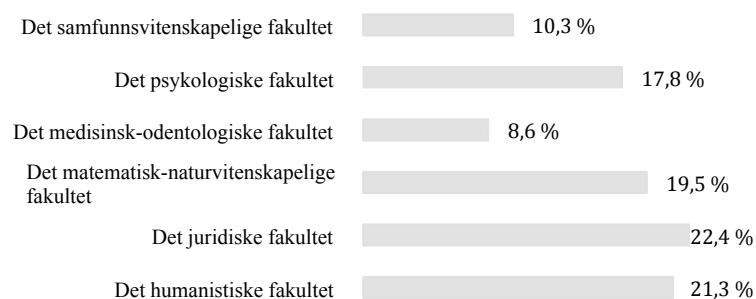
det å sikre anonymitet og konfidensialitet. Informasjon gitt i invitasjon til undersøkelsen samt godkjenning fra NSD kan finnes i vedlegg 1 og 2.

2.3 Utvalg

Spørreundersøkelsen ble sendt til samtlige vitenskapelig ansatte ved UiB, med forbehold om at e-postlisten vi fikk av en administrativt ansatt, er korrekt. Vi gjorde med andre ord ikke en tilfeldig trekning av utvalget men henvendte oss til hele populasjonen. Av emneansvarlige, som er min undersøkelsesenheter, var det 176 som svarte. Basert på informasjon gitt av samme administrativt ansatt, er det rundt 1300 emneansvarlige ved UiB.

2.3.1 Kort beskrivelse av respondentene. Utvalget består av 66 (39 %) kvinner og 103 (61 %) menn. 8 oppga ikke kjønn. Alder ble oppgitt under svaralternativene ”40 eller yngre”, ”41-50”, ”51-60”, 61 ”eller eldre” og aldersgruppen 41-50 har høyest frekvens (39 %). Respondentene har alt fra 0-2 års undervisningserfaring til over 20 år, og de fleste har over 10 års undervisningserfaring (68 %). Respondentenes stillingsbeskrivelser og fakultetstilhørighet blir kort presentert her, men forskjeller i svar basert på fakultet eller stilling er ikke inkludert som forskningsspørsmål i denne undersøkelsen. Respondentene er ansatt ved UiB som Professor (49,5 %), Førsteamanuensis (35,5 %), Postdoktor / Stipendiat / PhD kandidat (9 %), Universitetslektor (4 %) eller Forsker (2 %). Vi har respondenter fra alle utenom tre institutt og som figur 1 under viser, er samtlige fakultet representert.⁴

Figur 1 – *Fakultetstilhørighet*



⁴ Det nye fakultet for kunst, musikk og design som ble etablert i Januar 2017 er ikke representert da det viste seg at e-post listen vi fikk av en administrativt ansatt ikke inkluderte ansatte ved dette fakultetet.

2.3.2 Frafallsanalyse. I en selvadministrert spørreundersøkelse som denne, er frafall en vanlig utfordring (Ringdal, 2013). Om mulig, bør man finne ut hvem som ikke svarer, og i hvilken grad frafallet er systematisk (Denscombe, 2003; Oppenheim, 1992). Den vanligste årsaken til non-response er utilgjengelighet i innsamlingsperioden (Haraldsen, 1999) og vi fikk et relativt stort antall automatiske e-post svar om ferie eller permisjon. Andre grunner til frafall er at teknikken sviktet. Flere respondenter ga oss beskjed om at linken til spørreundersøkelsen ikke virket, eller at de opplevde tekniske problemer i gjennomføring av spørreundersøkelsen. Vi kan anta at dette da gjaldt betydelig flere enn de som sa ifra. Siden vi ikke hadde bakgrunnsinformasjon om språkpreferanser tilgjengelig, inneholdt e-posten med invitasjon til spørreundersøkelse informasjon på både norsk og engelsk. Vi valgte å la linken til spørreundersøkelsen gå til en engelsk versjon av spørreundersøkelsen, men opplyste om at man kunne bytte til norsk både i e-posten og på første side i spørreundersøkelsen. På bakgrunn av tilbakemeldinger vi fikk, vet vi at det var en del som likevel ikke fikk med seg denne informasjonen. Det kan tenkes at noen har unnlatt å svare på spørreundersøkelsen om de ikke visste at en norsk versjon var tilgjengelig. Selv om vurderingsformer er del av samtlige emneansvarlige sitt ansvarsområde, kan det kan tenkes at de som er spesielt interessert i vurderingsformer i større grad har svart. Basert på svar og kommentarer vi fikk ser det likevel ut som om vi har fått svar fra ”hele spekteret”. Da jeg har manglende kunnskap om fordelinger i populasjonen emneansvarlige på UiB, er det heller ikke mulig å sammenligne utvalget med populasjonen for å trekke en sikker konklusjon i forhold til systematisk frafall. Jeg kan derimot sammenligne med datamateriale til grunn for Aamodt et al. (2014) sitt studie, som også hadde en lignende skjevfordeling for ”gamle universitet”. Tilstandsrapport for høyere utdanning 2016 viser også at det generelt for sektoren høyere utdanning i Norge eksisterer en skjevfordeling lik den vi ser i denne undersøkelsen (Kunnskapsdepartementet [KD], 2016). På bakgrunn av dette er det da rimelig å anta at en slik skjevfordeling også eksisterer i populasjonen emneansvarlige på UiB. Frafallsanalysen gir få indikasjoner på at frafallet er systematisk, men respons-bias kan likevel ikke utelukkes. Med det menes hvilken påvirkning på resultatet de som ikke svarer, ville hatt om de svarte (Creswell, 2014). I tolkning av resultat og i generalisering til populasjon spesielt, må man derfor ta høyde for eventuelle representasjonsfeil.

2.4 Konstruksjon av spørreskjema

Hvordan spørreskjemaet konstrueres, hvilke element som blir inkludert og hvordan delene blir satt sammen er viktig med tanke på å redusere non-response. Å skape et spørreskjema med minst mulig friksjon er viktig for både villighet til å delta og å gjennomføre spørreundersøkelsen (Haraldsen, 1999). Med friksjon menes her ulike typer feil i formulering av spørsmål, svaralternativer og design. Da alle spørsmål utenom de demografiske i vår undersøkelse var obligatorisk, risikerte vi frustrasjon hos respondent ved uklarhet, noe som kan føre til at spørreundersøkelsen ble avbrutt.

Begrepsbruk er kanskje det viktigste punktet med tanke på utforming av spørsmål, men også det som bærer med seg flest utfordringer. Det er viktig fordi dette går direkte på spørreundersøkelsens validitet og reliabilitet, noe som skal adresseres lenger nede, og utfordrende fordi mange av våre begreper ikke har en entydig teoretisk betydning. Selv begrep som har en entydig teoretisk betydning, kan i praksis være mangetydige (Haraldsen, 1999, s. 155). Begrepet ”vurdering”, som er hovedbegrepet i spørreundersøkelsen, er heftet med mange ulike betydninger og forståelser og er knyttet til fundamental ulik praksis. Det som har vært viktig her, er å gjøre vårt beste for at spørsmålene er entydige selv om begrepene ikke er det (Danielsen, 2013, s. 145). Her har vi blant annet fulgt et generelt råd om å tilpasse spørsmålsformulering etter målgruppen, og ikke overvurdere gruppens kunnskapsnivå. Vi har med en kompetent og kunnskapsrik populasjon å gjøre, men ville unngå å bruke ord som det kunne tenkes ikke er en del av gruppens dagligtale (Ringdal, 2013, s. 204). Vi valgte blant annet å utelate begrepsparet summativ og formativ vurdering, selv om det å utelate det kunne skape frustrasjon for de som kjenner begrepene godt. Valget ble tatt da det kan tenkes at emneansvarlige utenfor vårt eget fagmiljø ikke så lett kan fylle disse begrepene med samme meningsinnhold.

2.5 Måleinstrument - Oversikt over variabler

I en deskriptiv-analytisk undersøkelse som vår, er målet å kartlegge basert på innsamlet data fra populasjonen som skal undersøkes, med bakgrunn i vide problemstillinger, på mange variabler (Befring, 1998, s. 67). Spørsmål ble stilt for å få oversikt og innsikt i hvilken grad emneansvarlige bruker eller er i berøring med ulike vurderingsmetoder, læringsutbyttebeskrivelser, vurderingskriterier og opplevde barrierer (i det følgende kalt ”vurderingspraksisvariabler”). Videre ble respondentene

bedt om å oppgi kjønn, alder og antall års undervisningserfaring (demografiske variabler). Det jeg har kalt ”utsagnsvariabler” er en serie utsagn emneansvarlige ble spurt om å gi tilsagn til. Spørreundersøkelsen hadde til sammen 37 spørsmål, og jeg har valgt ut 23 av disse som er relevant for å belyse mine forskningsspørsmål. Det kan være verdt å merke seg at målgruppen til medforfatter var samtlige undervisere, ikke bare emneansvarlige. Spørsmålene er derfor formulert med tanke på den større populasjonen.

Visse spørsmål kan brukes direkte som analysevariabler mens andre må bearbeides før de kan benyttes i statistiske analyser (Ringdal, 2013). Her følger en presentasjon av variablene brukt i undersøkelsen, og hvordan svaralternativ på spørsmålene er kodet og omkodet for å kunne brukes som måleinstrument i bivariate og multivariate analyser.

2.5.1 Vurderingspraksisvariabler. Det første spørsmålet faller under typen atferdsspørsmål. Alle spørsmål vi har inkludert om underviseres ”adferd” er relativt tett knyttet til underviserens vurderingspraksis gjennom hele semesteret. Selv om utfordringer knyttet til langtidshukommelse kan være en utfordring i atferdsspørsmål (Haraldsen, 1999, s.167), er dette mest sannsynlig ikke relevant her.

1. Antall formative og 2. Antall Summative. Disse variablene er knyttet til spørsmålet: *Hvilke av følgende strategier brukes for å la undervisere og/eller studenter måle eller vurdere læring i emnene du er involvert i? (Sett ett eller flere kryss).* Samtlige svaralternativ er vist i figur 2 (s.47). Da dette er et flerkrysspørsmål blir hvert av svaralternativene automatisk kodet 0 = ikke avkrysset / 1= avkrysset ved overføring av datasettet fra SurveyXact til SPSS, og hvert svaralternativ blir med det en egen variabel. For å få innsikt i hvor stor grad ulike emneansvarlige er i berøring med formative strategier for vurdering, ble svaralternativ som omhandlet typiske formative vurderingskategorier samlet i en ny variabel: ”1 Antall formative”. Dette er en vanlig og skånsom behandling variabler da det kun er snakk om å summere og ikke manipulere datapunktet. Det er for eksempel ikke snakk om å transformere variabler for å møte forutsetninger for parametriske tester, noe som kan være mer kontroversielt (Pallant, 2010). Dette ble gjort i flere steg. Først slo jeg sammen svaralternativ som var variasjoner av samme type formativ metode. De tre svaralternativene som omhandler Quiz til bruk for ulike formål ble slått sammen da bruk av 1 quiz kan utføre alle tre formål: med tanke på studenters selvevaluering, en måte å sjekke om studenter er forberedt til undervisning og med formål om å tilpasse videre undervisning. Når det

gjelder de tre svaralternativene som omhandler skriftlige arbeider eller øvelser med tilbakemelding enten fra lærer, medstudent eller begge, kan dette sies å være tre ulike praksiser, men teori redegjort for i introduksjonen skiller ikke så sterkt på disse. Etter å ha veid tapet av informasjonen opp mot muligheten av at respondenten har krysset av på flere svaralternativ på bakgrunn av en og samme praksis, valgte jeg å kollapse også disse svaralternativene. Når det gjelder svaralternativene ”mappevurdering” og ”refleksjonsnotater” er dette et eksempel som speiler egen læringskurve, og der det manglet en på forhånd lagt tilstrekkelig detaljert plan for hvordan svarene skulle tolkes. I utgangspunktet var det et forsøk på å si noe kvalitativt om innholdet i vurderingsformen ”mappevurdering”, men dette er ikke mulig basert på disse svarene. Det kan tenkes at de som opererer med refleksjonsnotat i en mappevurdering, lar være å krysse av på svaralternativet ”refleksjonsnotater” da man oppfatter det som en del av svaralternativet ”mappevurdering”. Hele 32 av 57 som svarte mappevurdering har ikke krysset av på refleksjonsnotater, og 23 har kun krysset av for refleksjonsnotater. Da usikkerheten knyttet til hva respondenter har tenkt når de krysset av eller ikke på refleksjonsnotater er stor, valgte jeg å ikke slå sammen disse to variablene, men å beholde dem som to separate skårer selv om det i noen tilfeller kan tenkes at de viser til en og samme vurderingsstrategi. Andre og siste steg var å summere summen av skåren på de samlede tilbakemelding og quiz kategoriene, samt skåre på mappevurdering og refleksjonsnotat i den nye variabelen ”1. Antall formative”. Høyest mulig skår er 4 der 0 = ingen, 1= 1 og så videre. ”1. Antall formative” er ment som et mål på i hvilken grad emneansvarlige er i berøring med ulike formative vurderingsmetoder. Svaralternativene skriftlige skoleeksamener, muntlige skoleeksamener og andre skriftlige skoleeksamener ble samlet i variabelen ”2. Antall summative”. Denne er kodet 0 -3 på samme måte som 1. Antall formative.

7. Ønsket bruk av formative. Denne variabelen er basert på svar under spørsmålet: *Ønsker du å ta i bruk noen av følgende vurderingsformer som du ikke bruker pr. dags dato?* Svaralternativene i dette spørsmålet er de samme som i spørsmål 1 over. Dette kan gi oss en pekepinn på emneansvarlige sitt ønske om å ta i bruk andre eller flere formative vurderingsformer enn de gjør per dags dato. Dette spørsmålet ble automatisk kodet på samme måte som spørsmål 1. Svaralternativene ble også kollapset på samme måte som i variabelen ”1. Antall formative”.

3. Antall tidspunkt LUB. Denne variabelen er knyttet til spørsmålet: *På hvilke tidspunkt tar du med læringsutbyttebeskrivelsene i arbeidet med vurdering? (Sett ett*

eller flere kryss). Svaralternativene her var: (1) Når emnet planlegges og bygges opp, (2) Når studentene blir gitt tilbakemeldinger, (3) Når studentene får instruksjoner om hvordan de skal gi tilbakemelding til medstudenter, (4) Når forelesninger planlegges, (5) Ved karaktersetting, (6) Ved samarbeid med ekstern sensor, (7) Aldri, (8) Annet. De ulike svarkategoriene er kodet 0 = ikke avkrysset, 1=avkrysset. Den nye variabelen "3. Antall tidspunkt LUB" er basert på skåre for svaralternativ (1), (4), (5), (6), (8). Dette for å få et mål på i hvilken grad læringsutbyttebeskrivelsene (LUB) blir brukt i vurderingsrelatert arbeid. Svaralternativ 2 og 3 ble utelatt siden dette er knyttet til spesifikk vurderingspraksis (tilbakemelding) som ikke alle benytter. Variabelen er kodet 0 til 4.

4. Definert vurderingskriterium. Nei/Ja. Denne variabelen er en dikotom kategorivariabel skapt på bakgrunn av svaralternativ på spørsmålet: *Har du definert spesifikke kriterier for vurdering (for eks. vurderingsrubrikker) basert på læringsutbyttebeskrivelsene? (Sett ett kryss)* Dette spørsmålet har 3 svaralternativer. (1) Ja, (2) Nei, men jeg bruker andre på forhånd definerte kriterier, og (3) Nei, jeg bruker ikke på forhånd definerte kriterier. Gjennom dette spørsmålet får vi svar på to spørsmål i ett. Hvorvidt emneansvarlige bruker på forhånd definerte vurderingskriterier, og om disse basert på læringsutbyttebeskrivelsene. Kategoriene ble kollapset til den dikotome kategorivariabelen "4. Definert vurderingskriterium. Nei/Ja". (1 = nei, 2=ja) som mål på hvorvidt emneansvarlige bruker på forhånd definerte kriterier eller ikke. For denne og andre dikotome kategorivariabler har jeg valgt å inkludere Nei/Ja i variabelnavnet for å vise at det nettopp er en dikotom kategorivariabel og for å støtte leseren i tolkning av retningen på resultatet av korrelasjonsanalysene.

5. Formidler vurderingskriterium. Nei/Ja. Dette variabelnavnet er en forkortet versjon av spørsmålsformuleringen: *Formidler du aktuelle vurderingskriterier til studentene?* Svaralternativene her er "Ja" eller "Nei". I utgangspunktet var disse kodet 1=Ja, 2 = Nei, men disse ble snudd for å sikre konsistens og unngå sjanse for feiltolkning av resultat. Med andre ord er 1 = Nei, 2=Ja.

6. Vurderingspraksis vs. Ideell. Dette variabelnavnet er som over bare en forkortelse av spørsmålsformuleringen: *Er din nåværende vurderingspraksis lik det du tenker er en ideell måte å vurdere studenters læring på?* Svaralternativene her var (1) I stor grad, (2) I noen grad, og (3) i liten eller ingen grad. Holdningsspørsmål angis vanligvis i vurderingsskalaer fra negative til positive ytterpunkter (Haraldsen, 1999, s.

163, Ringdal s. 202). For å lette analysearbeidet og fortolkning av resultat valgte jeg derfor å omkode også denne variabelen der 1 ble 3 og 3 ble 1.

8a Antall barrierer og 8b. Barrierer Nei/Ja. Disse to variablene er knyttet til spørsmålet: *Opplever du at det eksisterer barrierer som hindrer deg i å implementere nye vurderingsformer? (Sett ett eller flere kryss).* Full oversikt over svaralternativer kan sees i tabell 5 (s.51). Som med de andre flerkrysspørsmål ble hver av kategoriene kodet 0/1. For å få mål på totalt antall barrierer ble samtlige kategorier utenom ”Nei” tatt med i variabelen ”8a. Antall barrierer” som er kodet 0 – 7 fra lav til høy skåre. Den dikotome variabelen ”8b. Barrierer Nei/Ja” ble også skapt på bakgrunn av spørsmål 10 der 0 = Nei, 1 = Ja. 2 respondenter hadde krysset av for ”Nei”, men samtidig krysset av for flere barrierer. De ble da likevel kodet som ”1”.

9. Ønske Kursing Nei/Ja. Denne dikotome kategorivariabelen er skapt på bakgrunn av spørsmålet: *28. Ønsker du kursing i hvordan du kan planlegge eller gjennomføre vurdering? Hvis ja, hvilke type kursing ønsker du? (Sett ett eller flere kryss).* Svaralternativene var Halvdagskurs, Heldagskurs, 3-dagers kurs, Instruksjonsvideoer på nett, Artikler med instruksjoner på nett, Nei, trenger ikke kurs, Annet. For å kunne meningsfullt sammenligne ønsket om kurs eller ikke med andre variabler ble dette kategoriene omkodet til variabel ”9. Ønske kursing Nei/Ja” der 1 = ønsker ikke kursing og 2= ønsker kursing.

I utgangspunktet var også spørsmål knyttet til Mitt UiB en del problemstilling og forskningsspørsmål. Her har jeg avveket fra original plan da svar og kommentarer tydet på at det er for tidlig å stille spørsmål om Mitt UiB. Mange av respondentene har nylig har tatt den i bruk, og har dermed hatt lite erfaring med den, samt tid til å reflektere rundt bruken av den. Vi fikk likevel gode innspill til potensielle bedringer i plattformen som skal sendes videre til riktig instans.

2.5.2 Demografiske variabler.

K/M. Kjønn. Under spørsmål om kjønn var svaralternativene (1) Kvinne, (2) Mann, og (3) Transkjønnet/annet, ikke-binær. Da ingen krysset av for svaralternativ (3) ble denne gjort om til en dikotom kategorivariabel ”K/M -Kjønn” der 1 = Kvinne og 2 = Mann. 7 respondenter har valgt å ikke svare på dette spørsmålet.

A. Alder. Svaralternativene her var (1) Under 30, (2) 31 – 40, (3) 41-50, (4) 51-60, (5) 61-70 og (6) Over 70. Da kun 2 hadde krysset av på (1) og 1 på (60) kollapset jeg disse kategoriene sammen med nærmeste kategori og satt da igjen med 4 kategorier.

(1) 40 og yngre, (2) 41-50, (3) 51-60 og (4) 61 og eldre. Syv respondenter unnlot å svare på dette spørsmålet.

ERF. Ant. år undervisningserfaring. Denne variabelen har kun fått nytt navn og er ikke blitt omkodet. I SPSS er den kodet (1) 0-2 år, (2) 3-5 år, (3) 6-10 år, (4) 11-20 år, (5) Over 20 år.

2.5.3 Endrings skeptiske og endrings positive utsagnsvariabler.

Respondentene ble bedt om å ta stilling til serie en rekke utsagn gjennom en Likert-skala (tilslutningsskala) fra Helt Uenig – Uenig – Nøytral - Enig – Helt Enig).

Utsagnene er variasjoner av utsagn som Gray har hørt i møte med undervisere i intervju, kurs og andre sammenhenger og er kategorisert som endringspositive eller endrings skeptiske. Ved å kategorisere dem som endringspositive og endrings skeptiske foregriper jeg til dels begivenhetenes gang, men dette er gjort for å lette lesing og tolkning av resultat i neste kapittel. Betegnelsene er nødvendigvis overforenklet og i siste kapittel vil dette diskuteres. Utsagnene har blitt brukt direkte som analysevariabler i statistiske analyser med visse unntak. I frekvensfordelingen er noen av kategoriene kollapset slik at vi står igjen med Uenig, Nøytral og Enig. Å kollapse like svar er akseptabelt når man vil se et mer tydelig mønster i funnene (Babbie, 2010). Da de ulike utsagnsvariablene viste stor grad av signifikant samvariasjon gjennom bivariat analyse, ble også en faktoranalyse utført med tanke på å eventuelt samle utsagnene i en variabel. I den sammenheng ble 1 variabel reversert. Dette blir redegjort for i avsnitt 2.8 under som omhandler statistiske analyser. Utsagnene er kodet 1 = Helt Uenig, 2 = Uenig , 3 = Nøytral , 4 = Enig og 5 = Helt Enig. Utsagnene er listet under med forkortet variabelnavn foran.

U1. Nye metoder – forbigående trender - Nye måter å undervise og lære på, som for eksempel undervisvurdering og aktive læringsstrategier, er bare forbigående trender.

U2. Skriftlige arbeider forbedrer læring - Skriftlige arbeider forbedrer studenters læring.

U3. Forelesning er best - Forelesninger er den beste måten å undervise på i høyere utdanning.

U4. Tilbakem. krever for mye tid -Tiden som kreves for å gi tilbakemelding på skriftlige arbeider (med eller uten karakter), gjør det vanskelig å tilby undervisvurdering i et emne.

U5. Tilbakem. bortkastet tid - Å gi tilbakemeldinger er bortkastet arbeid da studentene mest sannsynlig ikke vil bruke de.

U6. Forbedre praksis – ikke verdsatt - *Jeg ser ingen grunn til å forbedre min undervisningspraksis siden det i min profesjonelle utvikling og karriere ikke er vurdert like høyt som forskning.*

U7. Studenter motvillige til nye met. - *Studenter vil være motvillige til å gjøre nye aktiviteter på nett (for eks. diskusjoner, oppgaver, quiz, videoer osv.).*

U8. Vurdering bør harmoneres med LUB - *Det er viktig at studenters læring er vurdert på en måte som harmonerer med læringsutbyttebeskrivelsene.*

U9. Forelesninger – ineffektivt for læring - *Forelesninger er en ineffektiv måte for studenter å lære på.*

U10. Flere karakterer bedre - *Å gi flere karakterer gjennom semesteret er en bedre måte å vurdere studenters læring på, enn en avsluttende vurdering i siste del av semesteret.*

2.6 Prosedyre

2.6.1 Pilottesting. Det var lite respons når jeg henvendte meg samlet til grupper av potensielle pilottestere, men fikk ved direkte forespørsel ja fra to emneansvarlige ved institutt for pedagogikk. Tross iherdige forsøk på å være “feilfri” avdekket pilottestere både designfeil og innholdsfeil. I ett tilfelle hadde respondenten blitt bedt om å vurdere følgende utsagn på en skala fra Helt uenig til Helt enig: “Jeg ønsker å bruke mer skriftlige arbeider i mine emner”. Hva skal den som allerede bruker mye skriftlige arbeider svare her? Svarskalaen passet med andre ord ikke overens med utsagnet og innsamlede data ville vært meningsløse. Vi endret derfor til utsagnet presentert over: “Skriftlige arbeider forbedrer studenters læring”. Her er det rom for alle til å gi sitt tilsagn uavhengig av variasjoner i nåværende praksis.

2.6.2 Administrasjon. Gray forfattet og sendte inn meldeskjema til Norsk senter for forskningsdata (NSD) for det større prosjektet denne spørreundersøkelsen er en del av. Etter at prosjektet ble godkjent av NSD ble utsendelse av den selvadministrerte spørreundersøkelsen gjort via SurveyXact Mandag 28. Februar 2017. Etter første dag hadde vi ca. 70 besvarelser og etter første arbeidsuke omtrent 140. Status før påminnelse var 153 fullførte besvarelser. Påminnelse ble sendt mot slutten av påfølgende arbeidsuke den 9. Februar. Vi lukket undersøkelsen for deltakelse den 26. Februar og hadde da totalt samlet inn 238 fullførte besvarelser. Se vedlegg 2 for kopi

av invitasjon og informasjon gitt når respondentene ble invitert til å delta. Som tidligere nevnt har jeg av de 238 besvarelsene trukket ut 176 som var svar fra emneansvarlige. Av disse hadde 166 svart på samtlige spørsmål, og 10 respondenter hadde unnlatt å svare på noen av de frivillige demografiske spørsmålene.

2.6.3 SurveyXact og SPSS. UiB har rettigheter til bruk av SurveyXact og dette er et program som oppfylte alt jeg trengte i forbindelse med administrering av spørreundersøkelsen. I analysearbeidet var også SurveyXact forenkende på den måten at man kunne generere deskriptive rapporter basert på valgte analysepunkter. Man kan også enkelt eksportere datasettet til IBM SPSS Statistics. Bearbeiding av data og statistiske analyser ble gjort i IBM SPSS Statistics version 24.

2.7 Bearbeiding av data

Før man setter i gang med statistiske analyser for å se grad av sammenheng mellom variabler, eller undersøke signifikante forskjeller mellom grupper, må man sjekke om datasettet oppfylder visse underliggende forutsetninger som de aller fleste analyser har (Tabachnick & Fidell, 2013, s. 63). For det første må man kontrollere om alt har gått riktig for seg ved import av datasettet i SPSS. Da SurveyXact genererte en fil med datasettet som kunne importeres direkte i SPSS, unngikk jeg å måtte manuelt kode samtlige variabler. Som vist over har jeg likevel manuelt omkodet visse variabler, men har kontrollert disse flere ganger for feil. Det andre man må gjøre er å håndtere manglende data. Dette blir redegjort for i avsnittet under. For det tredje må man sjekke hvorvidt og i hvilken grad man møter forutsetningene som de ulike statistiske testene bygger på. (Tabachnick & Fidell, 2013, s. 94). I hvilken grad, fordi man samfunnsvitenskapene stort sett aldri har gleden av ”perfekte” variabler som møter alle forutsetningene (Pallant, 2010, s. 111). Det viktige er hvilke valg man tar videre, og hvorvidt man kan underbygge disse. Dersom ens variabler ikke møter forutsetningene har man tre valg (Field, 2009; Pallant, 2010; Tabachnick & Fidell, 2013). Man kan velge å transformere variablene ved å endre data eller ekskludere outliers. Dette er mulig, men mer kontroversielt (Pallant, 2010). Et annet valg er å bruke ikke-parametriske analyseteknikker i stedet, da disse har mindre strenge forutsetninger. Ikke-parametriske tester er ikke like sensitive og oppdager kanskje ikke forhold eller forskjeller mellom grupper som en parametriske test ville vist (Field, 2009). Field (2009)

hevder at dette alternativet likevel er bedre om variablene ikke møter forutsetningene for parametriske tester i tilstrekkelig grad. Et tredje alternativ er å likevel bruke parametriske tester kjent som såkalte ”robuste” tester, som tåler en grad av brudd med forutsetningene (Field, 2009; Pallant, 2010; Tabachnick & Fidell, 2013). I den grad jeg oppdaget outliers eller skjevfordeling ved visuell inspeksjon, viste preliminaire statistiske analyser ingen alvorlige brudd på forutsetningene bortsett fra en variabel. I tilfellet der variabelen ikke i akseptabel grad møtte forutsetningene, valgte jeg å ekskludere variabelen fra videre analyse. Dette blir vist i sammenheng med resultat fra normalitetstesting rapportert i tabell 6 (s.54) i neste kapittel. Korrelasjonsanalyser kan være sensitive for outliers, og disse kan på ulik måte påvirke resultatet og vise sterkere eller svakere sammenhenger enn det som egentlig er tilfelle (Tabachnick & Fidell, 2013). Da outliers som jeg oppdaget ikke kunne sies være resultat av målefeil eller kodingsfeil, men var reelle datapunkt, valgte jeg likevel å la dem stå urørt da man skal ha gode grunner for å ekskludere dem (Tabachnick & Fidell, 2013).

2.7.1 Behandling av Missing data. Eventuelle mønster i missing data (manglende data) er viktigere enn hvor mye missing det er i et datasett da dette påvirker generaliserbarheten til resultatene (Tabachnick & Fidell, 2013). For hver person som unnlater å svare på et spørsmål i et spørreskjema blir det et hull i datamatriksen. Som en standardinnstilling i SPSS blir personer som har manglende informasjon på minst 1 variabel utelukket (stepwise) (Ringdal, 2013). I denne undersøkelsen har vi missing data kun på de demografiske variablene. Det er verken snakk om mye missing data da det kun er 2 til 7 respondenter som har unnlatt å svare på de enkelte demografiske spørsmålene, eller et spesielt mønster i missing data. Missing data er derfor ikke et stort problem og det ble løst ved bruk av pairwise deletion i bivariate og multivariate analyser. Dette vil si at en respondent som har manglende skåre på en variabel, kun ble utelatt i kalkulasjoner som involverer variabelen med manglende data, og ikke som i listwise at respondenten blir utelatt fra hele analysen (Pallant, 2010).

2.7.2 Signifikansnivå .05. En vanlig standard i forskning er å sette signifikansnivå til .05 (Ringdal, 2013), og denne undersøkelsen er ingen unntak. At resultatet er statistisk signifikant på .05 nivå vil si at det er 95 % sjans for at resultatet ikke er basert på tilfeldigheter (Clark-Carter, 2010), eller at man har gjort Type 1 eller Type 2 feil. Type 1 feil vil si at man basert på resultat fra statistisk analyse konkluderer

med at en nullhypotese om populasjonen er falsk, men at den egentlig er sann. Type 2 feil betyr at man godtar nullhypotesen som sann, mens den egentlig er falsk (Befring, 1998). Selv om jeg i denne undersøkelsen ikke har formulert spesifikke hypoteser, ligger det likevel hypoteser til grunn for samtlige tester jeg utfører. Eksempelvis en hypotese om sammenheng mellom bruk av ulike vurderingsmetoder og endrings skeptiske utsagn.

2.8 Statistiske analyser

I henhold til forskningsspørsmålet er formålet med studien både å få oversikt og mer innsikt i emneansvarlige på UiB sin vurderingspraksis. Dette vil gjøres gjennom statistikk for å oppsummere data, organisere og fortolke observasjoner og til å forstå og presentere data i håndterlige størrelser (Lund & Christophersen, 1999). Som for de fleste spørreundersøkelser har analysen gått gjennom flere stadier fra univariat til bivariat og til slutt multivariat analyse (Oppenheim, 1992). Under følger en redegjørelse over de statistiske testene som er brukt.

2.8.1 Deskriptiv univariat og bivariat analyse. Det innsamlede datamaterialet ble bearbeidet i SPSS gjennom frekvensanalyser og andre deskriptive analyser for å bedre oversikt over hvordan undersøkelsesenheten emneansvarlige svarte på ulike spørsmål. Gjennom frekvensfordelinger kan vi enkelt komme frem til sentraltendens (middelverdi), spredning (hvor konsentrert fordelingen er på den/de mest hyppige verdiene), og form på skårefordelingen (Lund & Christophersen, 1999). Basert på resultatene fra univariat analyse, ble det i noen tilfeller også gjennomført krysstabulering. Der det er hensiktsmessig, og for at leseren skal bli kjent med datamaterialet, vil variablene bli presentert i grafiske fremstillinger som tabulering og histogram i første del av resultatkapittelet. Gjennom de univariate analysene blir også 1. og 2. forskningsspørsmål forsøkt besvart: *Hvorvidt og/eller i hvilken grad er emneansvarlige i berøring med ulike vurderingsmetoder, læringsutbyttebeskrivelser og eksplisitte vurderingskriterier? og: I hvilken grad opplever emneansvarlige barrierer som hindrer implementering av nye vurderingsformer, og hvilke barrierer opplever?*

2.8.2 Bivariat korrelasjonsanalyse. I spørreundersøkelser er man ofte interessert i å utforske eventuelle sammenhenger (Danielsen, 2013, s. 142), og denne

undersøkelsen er intet unntak. Pearson product-moment korrelasjon blitt brukt for å belyse forskningsspørsmål 3: *Hvordan samvarierer bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderingspraksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring)?* Pearsons korrelasjonskoeffisient er det mest brukte målet på relasjon mellom variabler (Tabachnick & Fidell, 2013) og denne antyder også retning på samvariasjonen (Pallant, 2010). Point-biseral korrelasjon er brukt når den ene variabelen er en dikotom kategorivariabel (Field, 2009). Samtlige korrelasjoner blir presentert samlet i korrelasjonsmatrise (tabell 7, s.56). Basert på gjennomgangen av resultat fra univariat og bivariat analyse ble det også i visse tilfeller gjennomført delkorrelasjoner som kontrollerer for en tredje variabel som kan tenkes justerer resultatet (Pallant, 2010).

2.8.3 Multivariate analyser

2.8.3.1 Faktoranalyse av utsagnsvariabler. Som nevnt over ble det på bakgrunn av på høy grad av signifikant interkorrelasjon mellom flere av utsagnsvariablene også gjennomført en faktoranalyse. Faktoranalysen er en datareduksjonsteknikk (Pallant, 2010) og viste at 5 av utsagnene kunne med samles i 1 ny variabel jeg i det videre refererer til som "U20. Endrings skepsis". Variabel U7. *Studenter vil være motvillige til å gjøre nye aktiviteter på nett (for eks. diskusjoner, oppgaver, quiz, videoer osv.)* ble i utgangspunktet inkludert men hadde koeffisient verdi under .3 og ble derfor tatt ut (Pallant, 2010). De 5 variablene samlet i U20. Endrings skepsis er: U1. *Nye måter å undervise og lære på, som for eksempel undervisvurdering og aktive læringsstrategier, er bare forbigående trender.* U3. *Forelesninger er den beste måten å undervise på i høyere utdanning* U5. *Å gi tilbakemeldinger er bortkastet arbeid da studentene mest sannsynlig ikke vil bruke de.* U6. *Jeg ser ingen grunn til å forbedre min undervisningspraksis siden det i min profesjonelle utvikling og karriere ikke er vurdert like høyt som forskning.* U8. *Det er viktig at studenters læring er vurdert på en måte som harmonerer med læringsutbyttebeskrivelsene.* Det siste utsagnet er det eneste som kan sies å være ladet endringspositivt. Variabelen ble derfor reversert før faktoranalysen slik at Helt Uenig ble kodet 5, Uenig, 4, Nøytral 3, Enig 2 og Helt Enig 1 (Field, 2009). Disse utsagnsvariablene ble undersøkt og funnet passende for faktoranalyse gjennom koeffisienter på .3 eller over. Kaiser Meyer – Olkin verdien var .696 som er over anbefalt verdi på .6. Bartlett's test of sphericity var statistisk signifikant noe som støtter

faktorisering av korrelasjonsmatrisen (Pallant, 2010). ”U20. Endrings skepsis” hadde god reliabilitet testet med Cronbach’s Alpha (α , .603). Gjennomsnitt, standardavvik og faktorladning blir presentert i tabell 1 under.

Tabell 1 - *M, SD og faktorladning for utsagn i U20. Endrings skepsis.*

Utsagn	M	(SD)	Faktorladning
U1	2,24	,888	,753
U3	2,88	,952	,581
U5	1,94	,889	,545
U6	2,05	,970	,571
U8	1,97	,851	,657

Note: M = gjennomsnitt, SD = standardavvik.

For U20. Endrings skepsis er M, 2,2 og SD ,566

2.8.3.2 Stegvis Lineær regresjonsanalyse. Regresjonsanalyser ble gjort for å belyse siste forskningsspørsmål: *I hvilken grad kan bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderings praksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring) forklare variansen i bruk av formative og summative vurderingsmetoder?* Regresjon kan brukes for se i hvilken grad en rekke uavhengige variabler samlet kan forklare variansen i en avhengig variabel, men også hvor stort unikt bidrag hver av variablene bidrar med til totalvariansen (Tabachnick & Fidell, 2013). I en regresjonsanalyse blir variablene kontrollert for hverandre, og kontrollert for multikollinearitet vil variablene forklare ulike deler av totalvariansen (Pallant, 2010). Resultatet fra en regresjonsanalyse vil dermed gi et mer utfyllende bilde enn en serie bivariante korrelasjonsanalyser. Det er en rekke ulike måter å gjøre en regresjonsanalyse på. Her er stegvis regresjonsanalyse valgt da denne oftest brukes når målet er eksplorativt og ikke direkte baserer seg på tidligere forskning eller teori (Field, 2009). Stegvis regresjonsanalyse vil si at man mater SPSS med en rekke variabler som kan tenkes forklare variansen og deretter lar SPSS matematisk regne ut hvilken variabel som passer best i modellen (Field, 2009). Som Field (2009) skriver, SPSS er et smart men ikke intelligent verktøy og jeg har selv vurdert hvilke variabler som skal inkluderes i potensielle modeller (s. 225).

To regresjonsanalyser ble gjort med 1 Antall formative og 2 Antall summative som avhengig variabel. Testene gjort for å sjekke at forutsetningene regresjonsanalyse bygger på ble møtt, blir presentert sammen med resultatet i neste kapittel.

2.9. Validitet og Reliabilitet.

All statistikk kan brytes ned i en enkel modell: Resultat = (analysemodell) + målefeil (Field, 2009). Validitet og reliabilitet handler om ulike målefeil i forskningsprosessen.

2.9.1 Reliabilitet. Mål som har høy reliabilitet har små tilfeldige målefeil. Man kan beskytte seg mot målefeil i en undersøkelse, men de vil alltid være til stede (Ringdal, 2013). Empirien består av respondenters egenrapportering, og selv om spørsmål omhandler ”faktiske” forhold, kan respondenten huske feil eller krysse av på et annet svaralternativ enn de mente. Grep jeg har tatt for å beskytte reliabiliteten er å utforme spørsmålene så entydig som mulig, og å ha et nøyaktig arbeid i registrering og bearbeidelse av data. Da begrepene vi berører i seg selv ikke er entydig, kan det tenkes at respondentene likevel har forstått spørsmål og svaralternativer ulikt. Der vi har forsøkt å unngå begrepet, kan det diskuteres hvorvidt det er til hjelp eller ikke for respondenten. I oversettelsen til norsk kan det tenkes at en kyndig oversetter ville tørre å ta valg som sikret enda bedre flyt på norsk. I en perfekt verden ville ulike svar på spørsmål i undersøkelsen bety at det er en faktisk forskjell i respondentene (Fowler, 2009, s. 89), men det kan som redegjort for over også bety at respondentene tolket spørsmål eller svaralternativer ulikt. Som eneste variabel som består av flere items i undersøkelsen ble Cronbach`s alpha brukt som reliabilitetsmål for U 20.

Endringsskepsis. Når det gjelder reliabilitet for resultatene av statistisk analyse kan vi i forhold til svarprosent (13,5 %) finne at feilmarginen er på 6,9 % (95 % sikkerhetsintervall). Det har vist seg vanskelig å finne sikre kilder på hvor høy feilmargin som er ”akseptabelt” da søk i metodelitteratur viser at man unngår å si noe konkret om dette. Clark – Carter (2010) skriver eksempelvis og generelt at målet er å ha minst mulig feilmargin, og om mulig utvide utvalget hvis man har høyere feilmargin enn man er villig til å akseptere. Mer uformelle kilder som www.surveystar.com påstår at det er vanlig å operere med en feilmargin mellom 4 og 8 % på 95 % konfidensnivå i survey forskning. Ved hjelp av programmet G*power, som Field (2009) anbefaler, har jeg kalkulert power nivå for korrelasjon og regresjonsanalyser basert på 176

respondenter, .05 signifikansnivå og 0,2 nivå på korrelasjonen. For begge oppnås anbefalt nivå på 0.8 eller over (Field, 2009). Det vil si en 80 % sjans for å avdekke en effekt hvis den faktisk eksisterer i populasjonen, med andre ord unngå Type 2 feil (Field, 2009).

2.9.2 Validitet. Som kort beskrevet over, blir reliabiliteten påvirket av tilfeldige målefeil, mens systematiske målefeil går på dataenes validitet. Høy reliabilitet er en forutsetning for høy validitet (Ringdal, 2013). Validitet handler om hvorvidt jeg har målt det jeg mente å måle (Field, 2009) ergo hvorvidt spørreundersøkelsen gjør det den er ment å gjøre. Kilder til feil kan være at vi ha stilt spørsmålet ”feil”, eller respondenten har svart ”feil”. På bakgrunn av teori og erfaring har vi i spørsmålskonstruksjon forsøkt å få spørsmål som er gode mål på det vi ønsker å måle. Når det gjelder respondentens del i det å oppgi sannferdige og ”rette” svar basert på våre spørsmål, ble respondentene informert om at vi har sikret anonymitet gjennom SurveyXact, og at alle demografiske spørsmål er frivillige. Om respondenten opplever at de kan svare reelt anonymt, kan det bidra i rett retning (Ringdal, 2013).

3. Resultater

I det følgende skal resultat fra de statistiske analysene gjort for å belyse problemstilling og forskningsspørsmål bli presentert. Resultatkapittelet er tredelt. Først skal datamaterialet beskrives gjennom univariate og bivariat deskriptiv statistikk, deretter blir resultat fra de bivariate korrelasjonsanalysene presentert og til slutt resultat fra regresjonsanalyser.

3.1 Deskriptiv univariat og bivariat analyse

3.1.1 Kjønn, alder, og antall års undervisningserfaring. Som vist i tabell 1 på neste side består respondentene av 66 kvinner (39 %) og 103 menn (61 %). Med 66 respondenter i alderskategorien 41-50 (39 %) har denne høyest frekvens. Når det gjelder antall års undervisningserfaring har både kategorien ”11-20” og ”Over 20” 58 respondenter (34 %). Kategoriene som angir lengst undervisningserfaring har med andre ord høyest frekvens og utgjør til sammen 68 % av respondentene.

Tabell 1 – Frekvensfordeling for demografiske variabler i antall og prosent

Variabler

K/M.	Kvinne	Mann				
	66	103				
	37,5 %	58,5 %				
A.	< 41	41-50	51-60	>60		
	31	66	46	26		
	18,5 %	39 %	27 %	15,5 %		
ERF.	0-2	3-5	6-10	11-20	Over 20	
	12	18	25	58	58	
	7 %	10,5 %	14,5 %	34 %	34 %	

Note: ERF = antall års undervisningserfaring (n = 171), A=Alder (n=169), K/M = Kjønn (n=169)

Krysstabulering av de demografiske variablene viser at samtlige utenom én respondent i høyeste alderskategori har mer enn 20 års undervisningserfaring. Selv om

lang undervisningserfaring tilsier høyere alder, har også et flertall i laveste alderskategori over 5 års undervisningserfaring. I alderskategori 41-50 har 63% over 10 års undervisningserfaring. Når det gjelder kjønn og alder viste krystabulering at det kun er i laveste alderskategori at kvinner (K) og menn (M) er tilnærmet likt representert. En skjevfordeling øker suksessivt med høyere alder: Under 41 (K=48,5 %, M=51,5 %) 41-50 (K=43 %, M= 57 %), 51-60 (K=33 %, M=66 %) 61 og over (K=27 %, M=73 %).

3.1.2 Nåværende versus ideell praksis. I tabell 2 på neste side ser vi emneansvarlige i utvalget sitt svar på spørsmålet: *Er din nåværende vurderingspraksis lik det du tenker er en ideell måte å vurdere studentenes læring på?* 115 respondenter krysset av for "I noen grad" og dette svaralternativet har med det høyest frekvens (65%).

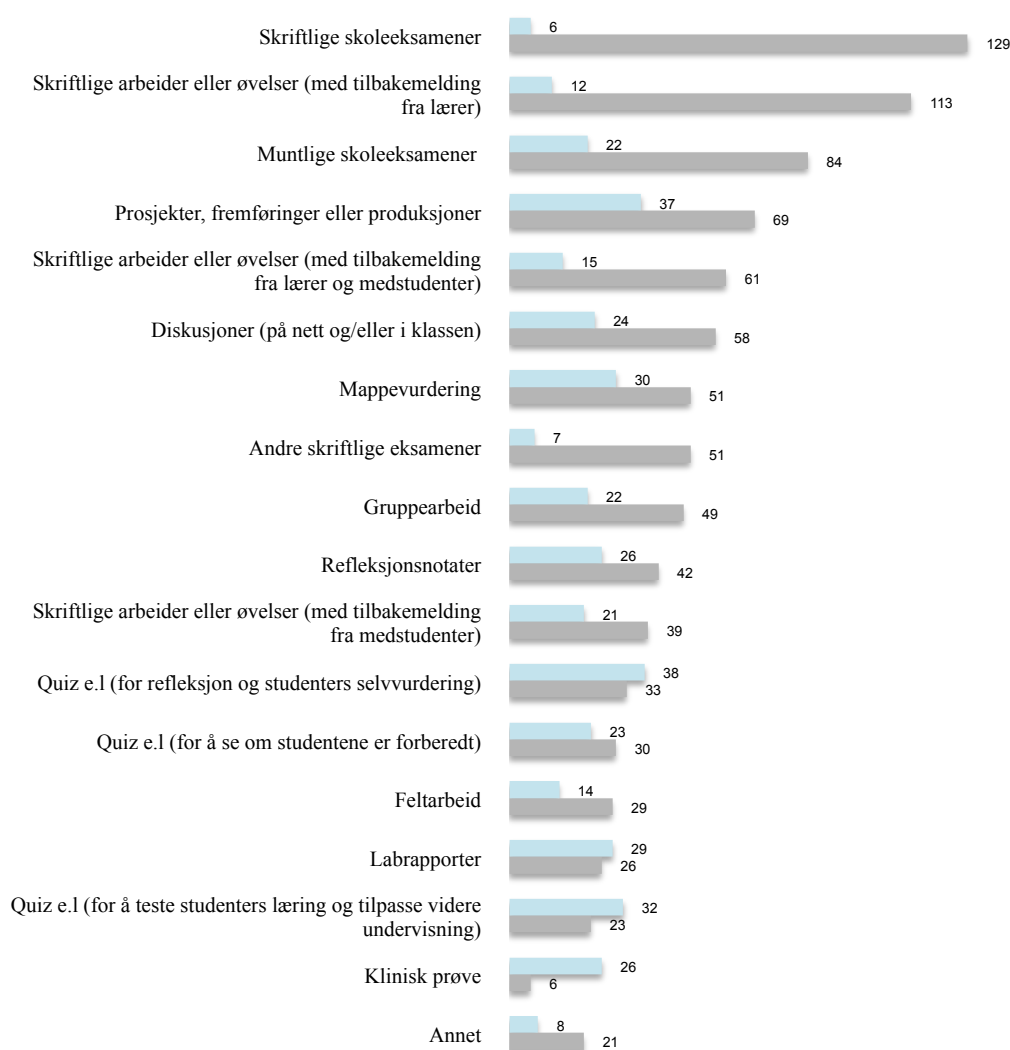
Tabell 2 - *Nåværende vurderingspraksis vs. ideell praksis*

Svaralternativ	Frekvens	Prosent
I liten eller ingen grad	3	1,7 %
I noen grad	115	65,3 %
I stor grad	58	33 %
Totalt	176	100 %

3.1.3 Vurderingsmetoder. Figur 2 på neste side viser frekvensfordeling av emneansvarliges svar på spørsmålet: *Hvilke av følgende strategier brukes for å la undervisere og/eller studenter måle eller vurdere læring i emnene du er involvert i?* og: *Ønsker du å ta i bruk noen av følgende vurderingsformer som du ikke bruker pr. dags dato?* På spørsmål om ønsket bruk var det 26 (15 %) som krysset av for "ingen av de overnevnte", og 22 (12,5 %) om svarte "Vet ikke". Resultatene er rangert fra høyest til lavest frekvens med utgangspunkt i svar gitt på spørsmålet om nåværende bruk av vurderingsformer. Det er ikke alle svaralternativer som inkluderes i videre statistiske analyser og her kommenteres kun de som er samlet i nye variabler. Når det gjelder svaralternativ til grunn for variabelen 1. Antall formative kan vi av figuren også se at 113 (64 %) svarer at de bruker skriftlige arbeider eller øvelser med tilbakemelding fra lærer i sine emner, 61 (35 %) oppgir at de bruker skriftlige arbeider eller øvelser med

tilbakemelding fra lærer og medstudenter, og at 39 respondenter krysser av for at det er medstudenter som gir tilbakemelding (22 %). 51 respondenter (29 %) svarer at mappevurdering er i bruk, og 42 (24 %) krysser av for refleksjonsnotater. Når det gjelder Quiz tas dette i bruk av 38 respondenter (22 %) for studenters refleksjon og selvvrdering, 30(17 %) svarer at Quiz blir brukt for å sjekke om studenter er forberedt, og 23 (13 %) krysser av for Quiz med formål om å teste læring for å tilpasse videre undervisning. Når det gjelder svaralternativ som er inkludert i variabelen 2. Antall Summative finner vi skriftlige skoleeksamener øverst med høyest frekvens (129, 73 %). Videre var det 84 (47 %) som krysset av for muntlige skoleeksamener, og 51 (29 %) for andre skriftlige eksamener. I figuren er vurderingsformer i bruk merket med grå farge, og vurderingsformer ønsket brukt med lyseblå.

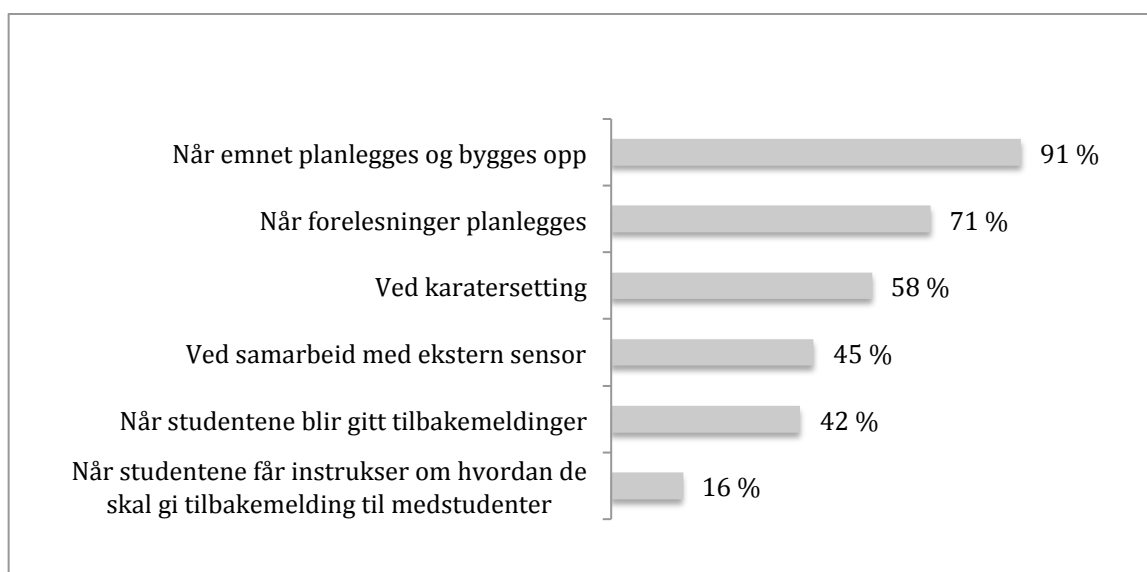
Figur 2 *Vurderingsformer i bruk og ønsket brukt*



Deskriptiv univariat analyse av variabel 1. Antall formative og 2. Antall Summative viser at alle utenom 1 respondent er i berøring med vurderingsmetoder som her er klassifisert som summative eller formative, og de aller fleste er i berøring med både summativ og formativ vurdering. 157 respondenter (89 %) har krysset av for 1 eller flere summative vurderingsmetoder. Resultatene overlapper ikke fullt ut, men det er også 157 respondentene som har krysset av for 1 av de formative. 143 respondenter (81 %) har krysset av for 1 eller flere svaralternativ som omhandler tilbakemelding.

3.1.4 Læringsutbyttebeskrivelser og vurderingskriterier. Figur 3 under viser frekvensfordeling for svar på spørsmålet *På hvilke tidspunkt tar du med læringsutbyttebeskrivelsene i arbeidet med vurdering?* Ikke vist i figuren er at 8 respondenter (4,5 %) krysset av for "Annet". De fleste påpekte her at de også bruker læringsutbyttebeskrivelsene når de lager eksamensoppgaver.

Figur 3 - Tidspunkt læringsutbyttebeskrivelser blir tatt i bruk



Note: n = 176

Tabell 3 under viser den nye variabelen 3. Antall tidspunkt LUB som angir antall tidspunkt respondentene tar i bruk læringsutbyttebeskrivelser. Som redegjort for i metoddelen har jeg i den nye variabelen utelatt tidspunkt knyttet til bruk av spesifikke metoder som tilbakemelding og medstudentvurdering.

Tabell 3 - Antall tidspunkt læringsutbyttebeskrivelsene blir tatt i bruk

Antall tidspunkt	Frekvens	Prosent
0	8	4,5 %
1	26	14,8 %
2	38	21,6 %
3	51	29 %
4	53	30,1 %
Totalt	176	100 %

På spørsmålet: *Har du definert spesifikke kriterier for vurdering (for eks. vurderingsrubrikker) basert på læringsutbyttebeskrivelsene?* har 70 respondenter (40 %) svart “Ja”, 56 (32 %) har svart “Nei, men jeg bruker andre på forhånd definerte kriterier” og 50 (28 %) har svart “Nei, jeg bruker ikke på forhånd definerte kriterier. Basert på dette spørsmålet ble variabel “4. Definert vurderingskrit. Nei/Ja” skapt. Totalt 126 (71,5 %) oppgir at de bruker på forhånd definerte kriterier (basert på læringsutbyttebeskrivelser eller ikke), og 50 (28,5 %) oppgir at de ikke bruker på forhånd definerte kriterier.

Når det gjelder variabel 5. Formidler vurderings.krit Nei/Ja har 136 respondenter (77 %) svart “Ja” på spørsmålet “ Formidler du aktuelle vurderingskriterier til studentene?”, og 40 (23 %) har svart “Nei”. Av dem som har svart at de formidler kriteriene, har 22 respondenter også svart at de ikke definerer vurderingskriterier.

3.1.5 Utsagn knyttet til undervisnings- og vurderingspraksis. Tabell 4 under viser en samlet oversikt over hvordan respondentene i ulik grad har gitt sitt tilsagn til en serie utsagn. I utgangspunktet var svaralternativene en 5-punkts Likert-skala, men i denne tabellen er kategoriene Helt Uenig og Uenig kollapset til kategorien “Uenig”, og Helt Enig og Enig er kollapset til “Enig”. Dette er gjort med tanke på at det tydeliggjør resultatet (Babbie, 2010). I tabell 6 (side 54) vil disse variablenes gjennomsnitt, standardavvik, skjevhet og kurtose også bli presentert.

Tabell 4

Samlet oversikt over fordeling av tilsagn til utsagnsvariablene vist i antall og prosent (%)

Variabler	Uenig	Nøytral	Enig
U1. Nye metoder – forbigående trender	109 (62 %)	56 (31,8 %)	11 (6,2 %)
U2. Skriftlige arbeider forbedrer læring	2 (1,1 %)	13 (7,4 %)	161 (91,5 %)
U3. Forelesning er best	61 (34,7 %)	75 (42,6 %)	40 (22,7 %)
U4. Tilbakemeldinger krever for mye tid	24 (13,6 %)	39 (22,2 %)	113 (64,2 %)
U5. Tilbakemeldinger bortkastet tid	138 (78,4 %)	31 (17,6 %)	7 (4 %)
U6. Ingen grunn til å forbedre praksis – ikke verdsatt	133 (75,6 %)	27 (15,3 %)	16 (9,1 %)
U7. Studenter motvillige til nye metoder	114 (64,7 %)	50 (28,4 %)	12 (6,9 %)
U8. Vurdering bør harmonere med LUB	10 (5,6 %)	5 (14,2 %)	141 (80,1 %)
U9. Forelesninger – ineffektivt for læring	103 (58,5 %)	41 (23,3 %)	32 (18,1 %)
U10. Flere karakterer bedre	30 (17,1 %)	53 (30,1 %)	93 (52,9 %)

Note: n for alle variabler er 176. Variabelbeskrivelsene er her forkortet. Utsagnsformuleringen er i sin helhet presentert på side 36

3.1.6 Barrierer. Frekvensfordelingene under dette punktet gir oss svar på forskningsspørsmål 2: *I hvilken grad opplever emneansvarlige barrierer som hindrer implementering av nye vurderingsformer, og hvilke barrierer oppleves?*

Under spørsmålet ”*Opplever du at det eksisterer barrierer som hindrer deg i å implementere nye vurderingsformer?*” var det 33 respondenter (18,7 %) som krysset av for svaralternativet ”Nei”. Under ”Annet” nevner flere manglende administrativ og teknisk support når endringer skal implementeres, spesifikt tekniske løsninger til mappe og UiB sin bruk av flere IT plattformer som ikke samfungerer. Frekvensfordeling av de øvrige svaralternativ vises i tabell 5 under. Av tabellen kan vi se at det er ulike typer systembarrierer som har høyest frekvens. På toppen av listen er det barrierer som har med ressursknapphet å gjøre, deretter barrierer knyttet til policy. Videre kommer støtte fra fagmiljø, begrenset kunnskap og usikkerhet rundt fremtidig undervisningsansvar og til slutt usikkerhet om studentene vil like det.

Tabell 5 - *Barrierer*

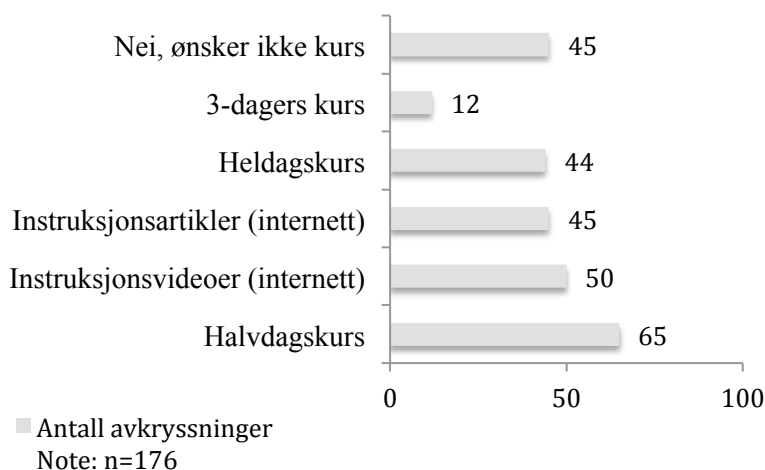
Svaralternativ	Frekvens	Prosent
Begrenset tid	95	54 %
For mye administrativt arbeid involvert	71	40 %
Kostnader (for eks. ekstern sensor)	47	27 %
Studenters rett til å klage på karakterer	34	19 %
Universitetspolicy	31	18 %
Lite støtte for å gjennomføre endringer i fagmiljø	32	18 %
Begrenset kunnskap om alternative vurderingsformer	29	16 %
Begrenset kunnskap om vurderings- og læringsteori	17	10 %
Usikker på fremtidig undervisningsansvar	15	9 %
Annet	13	7 %
Usikker på om studentene vil like det	10	6 %

Note N = 176

Deskriptiv analyse av den nye variabelen 8a. Antall barrierer for endring viser at 145 respondenter (82 %) oppgir 1 eller flere barrierer.

3.1.7 Å ønske kursing eller ikke. Figur 4 under gir en oversikt over respondentenes avkryssninger på ulike svaralternativer tilknyttet spørsmålet: *Ønsker du kursing i hvordan du kan planlegge eller gjennomføre vurdering? Hvis ja, hvilke type kursing ønsker du? (Sett et eller flere kryss)*

Figur 4 – *Ønsker du kurs?*



10 respondenter svarer "Annet" og de fleste skriver variasjoner av at "Kurs er bra, men har ikke tid". 1 respondent formulerer et ønske om en faglig ansatt med "drop-in" tider man kan benytte seg av når man trenger veiledning. Totalt 131 (74,5 %) svarer at de ønsker en eller annen form for støtte til å planlegge eller gjennomføre vurdering.

Krysstabulering ble også gjort for å undersøke forholdet mellom det å svare ja på barrierene som omhandlet mangel på kunnskap, og det å ønske kurs. Samtlige som krysset av for barrieren "Begrenset kunnskap om vurderings- og læringsteori" svarer at de ønsker kursing, og 27 av 29 som krysset av for barrieren "Begrenset kunnskap om vurderingsformer" ønsket ulike former for kursing.

3.2 Bivariate korrelasjonsanalyser

I avsnittene under presenteres resultat av korrelasjonsanalyser gjort for å belyse forskningsspørsmål 3: *Hvordan samvarierer bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderingspraksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring)?* Først skal resultat fra preliminære undersøkelser gjort for å teste normalitet da dette er en forutsetning for parametriske analyse som Pearson product-moment korrelasjonsanalyse og regresjonsanalyse. Verdier på skjevhet (symmetri) og kurtose (spissitet) fra -1 til +1 blir vurdert som god

fordeling, men -2 til +2 kan også aksepteres (Field, 2009). Skjevhet og kurtose på 0 vil si at fordelingen er normal (som en symmetrisk bjelle), men det er uvanlig at man får gleden av slike perfekte skårer i samfunnsvitenskapelig forskning (Pallant, 2011). Som vi så av tabell 4 (s. 50) over, er respondentene stort sett enig eller helt enig i utsagn U2 om at skriftlige arbeider forbedrer studentenes læring. Dette tilsier at variabelen vil være spisset, og fordelingen vil være samlet ved ene halen i normalfordelingen. Som vises i tabell 6 under bekrefter mål på kurtose at den er utenfor det som bør aksepteres (4,6) og symmetrien (skjevhet) er akseptabel men ikke god (1,2). På bakgrunn av dette har variabelen blitt ekskludert fra videre analyser.

Tabell 6 - Variablene på skalanivå sitt gjennomsnitt (*M*), standardavvik (*SD*), skjevhet og kurtose.

Variabler	M	SD	Skjevhet	Kurtose
1 Antall formative	1,64	1,004	,323	-,484
2. Antall summative	1,50	,835	,030	-,546
3. Antall tidspunkt LUB	2,65	1,185	-,507	-,737
6. Egen vurderingspraksis vs. ideell	2,31	,500	,395	-,886
7. Ønsket bruk av formative	,852	,875	,965	-,878
8a. Antall barrierer for endring	2,24	1,663	.474	-,201
A. Alder	2,40	,959	,197	-,882
ERF. Antall år undervisningserfaring	3,77	1,223	-,844	-,253
U1. Nye metoder – forbigående trender	2,24	,888	.354	-,090
U2. Skriftlige arbeider forbedrer læring	4,20	,668	-1,183	4,64
U3. Forelesning er best	2,88	,952	,242	-,135
U4. Tilbakemeldinger krever for mye tid	3,67	1,000	-,653	,011
U5. Tilbakemeldinger bortkastet tid	1,94	,889	1,061	1,606
U6. Ingen grunn til å forbedre praksis – ikke verdsatt	2,05	,970	,962	,717
U7. Studenter motvillige til nye metoder	2,19	,891	,358	-,358
U8. Vurdering bør harmonere med LUB	4,03	,851	-,954	1,201
U9. Forelesninger – ineffektivt for læring	2,48	1,008	,504	-,364
U10. Flere karakterer bedre	3,45	,955	-,370	-,326
U20. Endrings skepsis	2,22	,566	,352	,298

Note: A = Alder (N = 169), ERF – antall års undervisningserfaring (n = 172) N for øvrige variabler er 176. For variabel U20. Endrings skepsis er Cronbach's α .603.

Tabell 7 på neste side viser samtlige Pearson product – moment og point-biseral korrelasjoner mellom vurderingspraksisvariabler, demografiske variabler og utsagnsvariabler. I de påfølgende avsnittene vil også resultatene presenteres i tekst og korrelasjonskoeffisient og signifikansnivå vil gjentas til støtte for leseren.

Korrelasjonskoeffisienten har verdi fra -1 til +1, der 0 uttrykker at det ikke er noen sammenheng mellom variablene. 0,1 – 0,3 antyder liten sammenheng, 0,3 – 0,5 medium og 0,5 – 1,0 antyder sterk sammenheng (Pallant, 2010). Først skal sammenhengen mellom grad av tilsagn til de ulike utsagnene presenteres, deretter hvordan utsagnene korrelerer med kjønn, alder og undervisningserfaring. Til slutt presenteres korrelasjoner med utgangspunkt i vurderingspraksisvariabler og hvordan disse korrelerer med utsagnsvariabler og demografiske variabler.

Tabell 7 - Bivariat korrelasjonsmatrise av vurderingspraksisvariabler, demografiske variabler og utsagnsvariabler.

Variabler	1	2	3	4	5	6	7	8a	8b	9	K/M	A	ERF	U1	U3	U4	U5	U6	U7	U8	U9	U10
1. Antall Formative	-																					
2. Antall Summative	,119																					
3. Antall tidspunkt LUB	,034	-,020																				
4. Definert vurderingskrit. Nei/Ja	,077	,061	,316**																			
5. Formidler vurderingskrit. Nei/Ja	,090	-,016	,231**	,320**																		
6. Vurderingspraksis vs. ideell	-,060	-,075	,097	-,085	,095																	
7. Ønsket bruk av formative	-	,070	,099	,038	-,076	-,233**																
8a. Antall barrierer	,069	,082	,022	-,016	-,077	-,166*	,240**															
8b. Barrierer Nei/Ja	-,058	,206**	-,009	,039	-,073	-,219**	,178*	-														
9. Ønske Kursing Nei/Ja	,051	-,055	,115	,064	-,007	-,103	,348**	,265**	,310**													
K/M. Kjønn	-,188*	,018	-,148	-,147	-,064	,007	,021	,057	-,043	-,106												
A. Alder	,031	,102	-,050	,019	,071	,015	-,147	-,116	-,048	-,141	,150											
ERF. Ant. år undervisningserfaring	,191*	,178*	-,030	-,004	,101	,001	-,165*	-,131	-,059	-,252**	,168*	-										
U1. Nye metoder–forbigående trender	-,282**	,046	-,127	-,101	-,084	,076	-,219**	-,147	-,145	-,283**	,076	,270***	,121									
U3. Forelesning er best	-,272**	-,018	-,072	-,026	-,140	,103	-,104	-,209**	-,200**	-,252**	,165*	,069	-,084	,331**								
U4.Tilbakem. krever for mye tid	-,101	,096	,043	,032	-,043	-,147	,166*	,261**	,281**	,081	,074	,010	,064	,018	,000							
U5.Tilbakem. bortkastet tid	-,115*	,073	-,075	-,087	-,191*	,006	-,085	,029	,001	-,100	,136	,048	-,013	,301**	,106	,086						
U6. Forbedre praksis – ikke verdsatt	-,163*	,011	-,163*	-,149*	-,027	,038	-,002	,070	,024	-,063	,055	-,036	-,023	,231**	,168*	,076	,222**					
U7. Studenter motvillige til nye met.	-,052	-,027	-,187*	-,108	-,114	-,004	-,089	-,092	-,070	-,184*	,133	,004	,102	,182*	,121	-,065	,217**	,108				
U8.Vurdering bør harmonere med LUB	,152*	,004	,180*	,095	,130	,073	-,010	,015	-,037	,158*	-,154*	,063	,031	-,357**	,236**	,031	-,164*	-,237**	-,188**			
U9.Forelesninger–ineffektivt for læring	,189*	-,078	-,003	,077	,126	-,108	,042	,176*	,133	,139	,022	,018	,008	-,174*	-,547**	-,023	-,049	-,025	-,019	,244**		
U10.Flere karakterer bedre	-,049	,082	,012	,204**	,099	-,116	,128	,184*	,202**	,208**	-,025	,135	,012	-,127	-,117	,066	-,034	-,006	-,086	,139	,189*	
U20.Enddringskepsis	-,318**	,034	-,198**	-,147	-,182*	,051	-,127	-,088	-,093	-,274**	,186*	,093	-,011	,706**	,602**	,050	,570**	,613**	,259**	-,624**	-,336**	-,134

Note: ***= Korrelasjon er signifikant på .005-nivå (2-halet) ** = Korrelasjon er signifikant på .01-nivå (2-halet). * =Korrelasjon er signifikant på 0.05-nivå (2-halet).

- =Resultatet er ikke oppført da variablene er variasjoner av samme variabel

3.2.1 Korrelasjon mellom utsagnsvariabler. Respondentene ga sitt tilsagn til en serie utsagn gjennom en skala fra Helt Uenig, Uenig, Nøytral, Enig og Helt Enig. Korrelasjonsmatrisen over viser at sterkeste korrelasjon er mellom utsagnene U3 *Forelesninger er den beste måten å undervise på i høyere utdanning* og U9 *Forelesninger er en ineffektiv måte for studenter å lære på* ($r = -.547, p < .01$). At korrelasjonen er negativ vil si at høyere grad av enighet uttrykt på den ene variabelen assosieres med lavere grad av enighet på den andre variabelen.

Med utgangspunkt i tabell 7 og første utsagnsvariabel U1 kan vi se at tanken om at nye metoder er forbigående trender, har en signifikant positiv assosiasjon med U3 at forelesning er best i høyere utdanning ($r = .331, p < .01$), U5 at det å gi studenter tilbakemeldinger er bortkastet da de likevel ikke bruker de ($r = .301, p < .01$), U6 at det er ingen grunn til å forbedre praksis da det ikke er verdsatt ($r = .231, p < .001$), og U7 at studenter er motvillige til nye metoder ($r = .182, p < .05$). I tillegg ser vi en negativ signifikant assosiasjon med følgende utsagn: U8 om at vurdering bør harmonere med læringsutbyttebeskrivelser ($r = -.357, p < .01$) og U9 at forelesninger er ineffektivt med tanke på læring ($r = -.174, p < .05$). Leser man korrelasjonsmatrisen med utgangspunkt i flere av de andre utsagnsvariablene vil man også se at det er høy grad av interkorrelasjon. Som redegjort for i metodekapittelet ble det på bakgrunn av dette utført en faktoranalyse. Basert på denne ble den nye samlevariabelen U20. Endringsskepsis skapt av gjennomsnittsverdier fra U1, U3, U5, U6 og U8 reversert. Lavere grad av enighet i U20 vil bety at man har uttrykt lavere grad av enighet i endringsskeptiske utsagn, og omvendt.

3.2.2 Samvariasjon mellom utsagnsvariabler og demografiske variabler.

Når det gjelder kjønn kan vi av korrelasjonstabell 7 for det første se at det å være kvinne er signifikant assosiert med lengre undervisningserfaring ($r = .168, p < .05$). Korrelasjonsmatrisen over viser også en signifikant positiv korrelasjon mellom kjønn og U20. Endringsskepsis ($r = .186, p < .05$) samt utsagn U3 om at forelesninger er best ($r = .165, p < .05$). Korrelasjonen viser derimot en negativ assosiasjon mellom kjønn og U8 om at vurdering bør harmonere med læringsutbyttebeskrivelser ($r = -.154, p < .05$). Slik variabelen kjønn er kodet vil det si at resultatet antyder at det å være kvinne er assosiert med å ha lavere grad av enighet i endringsskeptiske utsagn, eller omvendt, at menn er assosiert med høyere grad av enighet i de samme.

Da det som rapportert over viste seg at kvinner var assosiert med lengre undervisningserfaring, og krysstabulering viste at menn var forholdsvis overrepresentert i høyere alderskategorier, ble også delkorrelasjoner utført. Resultatet av delkorrelasjonen viser at sammenhengen mellom kjønn og utsagnsvariabler fortsatt er signifikant på .05 nivå kontrollert for Alder, (U20: $r = .174$, U3: $r = .156$, U8: $r = -.165$) og antall års undervisningserfaring (U20: $r = .190$, U3: $r = .182$, U8: $r = -.161$).

Videre viser korrelasjonsmatrisen (tabell 7) at den eneste utsagnsvariabelen som korrelerer signifikant med alder er U1 ($r = .270$, $p < .005$). Å være eldre er med andre ord assosiert med enighet i at nye undervisningsformer er forbigående trender. På samme måte som over ble en delkorrelasjon utført for å kontrollere resultatet for kjønn og undervisningserfaring. Resultatet viser at kjønn ($r = .263$, $p < .001$) og antall års undervisningserfaring ($r = .255$, $p < .005$) i liten grad justerte korrelasjonskoeffisienten og styrken på signifikans.

Til slutt kan vi av korrelasjonsmatrisen (tabell 7) se at antall års undervisningserfaring ikke korrelerer signifikant med noen av utsagnsvariablene, kun med kjønn omtalt over.

3.2.3. Bruk av Formative vurderingsmetoder. Variabelen 1. Antall formative samvarierer ikke signifikant med noen av de andre vurderingspraksisvariablene, men bruk av flere formative vurderingsmetoder er assosiert med det å være kvinne ($r = -.188$, $p < .05$) og lengre undervisningserfaring ($r = .191$, $p < .05$). Det er også signifikante korrelasjoner med flere av utsagnsvariablene. Det å være i berøring med et høyere antall formative vurderingsmetoder er signifikant assosiert med lavere grad av enighet i samlevariabelen U20. Endringsskepsis ($r = -.318$, $p < .01$). Ser vi på utsagnsvariablene hver for seg er berøring med et høyere antall formative vurderingsmetoder signifikant assosiert med lavere grad av enighet i utsagnene som uttrykker at U1 nye metoder er forbigående trender ($r = -.282$, $p < .01$), U3 at forelesning er beste måten å undervise på i høyere utdanning ($r = -.272$, $p < .01$) og U6 at man ikke ser grunn til å forbedre praksis da det ikke er verdsatt ($r = -.163$, $p < .05$). Derimot antyder resultatet en høyere grad av enighet i U8 at vurdering bør harmonere med læringsutbyttebeskrivelsene ($r = .165$, $p < .05$) og U9 at forelesning er en ineffektiv måte for studenter å lære på ($r = .162$, $p < .05$).

3.2.4 Bruk av summative vurderingsmetoder. Å være i berøring med flere summative vurderingsmetoder er assosiert med å oppleve barrierer som hindrer

implementering av nye vurderingsmetoder ($r = .206, p < .01$) og lengre undervisningserfaring ($r = .178, p < .05$). Bruk av summative vurderingsmetoder er ikke signifikant assosiert med noen av utsagnsvariablene.

3.2.5 Bruk av læringsutbyttebeskrivelser. Når det gjelder korrelasjon med andre vurderingspraksisvariabler er det å ta i bruk læringsutbyttebeskrivelser målt på antall tidspunkt i vurderingsprosessen signifikant assosiert med å definere eksplisitte vurderingskriterier ($r = -.316, p < .01$) og å formidle dem til studentene ($r = .231, p < .01$). Det er ingen signifikant korrelasjon med demografiske variabler, men det å ta i bruk læringsutbyttebeskrivelser i høyere grad er assosiert med en lavere grad av enighet i endringsskeptiske utsagn, blant annet U20. Endringsskepsis ($r = -.198, p < .01$). Når det gjelder de ulike utsagnsvariablene sett hver for seg, er det å ta i bruk læringsutbyttebeskrivelsene på flere tidspunkt korrelert med lavere grad av enighet i at man ikke ser grunn til å forbedre undervisningspraksis da det ikke er verdsatt like høyt som forskning ($r = -.163, p < .05$), at studenter er motvillige til nye metoder ($r = -.187, p < .05$), og til slutt at vurdering bør harmonere med læringsutbyttebeskrivelsene ($r = .180, p < .05$).

3.2.6 Å definere og å formidle vurderingskriterier.

Det å definere spesifikke vurderingskriterier og å formidle aktuelle vurderingskriterier til studentene er signifikant korrelert ($r = .320, p < .01$). Av korrelasjonsmatrisen (tabell 7) kan man også se at det å eksplisitt definere og formidle vurderingskriterier er signifikant assosiert med lavere grad av enighet i U6 at man ikke ser noen grunn til å forbedre praksis da det ikke er verdsatt ($r = -.149, p < .05$) og U10 at flere karakterer er bedre ($r = .204, p < .01$). Å definere vurderingskriterier er som i bruk av Læringsutbyttebeskrivelser ikke signifikant korrelert med noen av de demografiske variablene.

3.2.7 Barrierer for nye vurderingsmetoder. Det er to variabler inkludert i korrelasjonsmatrisen (tabell 7) som sier noe om barrierer for det å ta i bruk nye vurderingsmetoder. Variabel 8a angir antall barrierer, og variabel 8b er kodet som en dikotom kategorivariabel som angir hvorvidt respondenten opplever barrierer eller ikke. Disse to variablene er skapt på bakgrunn av samme variabel og korrelerer derfor i all

hovedsak signifikant med samme variabler, men med noe ulik styrke på sammenhengen.

Med utgangspunkt i Variabel 8a som måler grad av opplevde barrierer kan vi av korrelasjonsmatrisen i tabell 7 for det første se at det å oppleve et høyere antall barrierer er signifikant assosiert med det å være mindre fornøyd med egen vurderingspraksis målt opp mot en tenkt "ideell" praksis ($r = -.166, p < .05$), et ønske om å ta i bruk et høyere antall formative vurderingsmetoder ($r = .240, p < .01$), og et ønske om å lære mer om hvordan man kan planlegge eller gjennomføre vurdering ($r = .265, p < .01$). For det andre er det å oppleve flere barrierer assosiert med en høyere grad av enighet i utsagn om at tilbakemeldinger krever for mye tid ($r = .261, p < .01$), U9 at forelesninger er ineffektivt for læring og U10 at det å gi flere karakterer gjennom semesteret er en bedre måte å vurdere studentenes læring på, enn en avsluttende vurdering i siste del av semesteret ($r = .202, p < .01$). Å oppleve større grad av barrierer er også assosiert med en lavere grad av enighet i U3 at forelesninger er den beste måten å undervise på i høyere utdanning ($r = -.209, p < .01$). Den dikotome kategorivariabelen 10b som angir hvorvidt respondenten har oppgitt at han/hun opplever barrierer, korrelerer også signifikant med antall summative vurderingsformer ($r = .206, p < .01$). Slik variabelene er kodet betyr dette at det er en signifikant assosiasjon mellom det å oppleve barrierer og å være i berøring med flere summative vurderingsmetoder.

3.2.8 Å ønske kursing eller ikke. Variabel 9. Ønske kursing Nei/Ja er en dikotom kategorivariabel som uttrykker hvorvidt den emneansvarlige ønsker kursing i hvordan man kan planlegge eller gjennomføre vurdering. Resultatene rapportert i tabell 7 antyder at et ønske om å lære mer gjennom ulike typer av kursing er assosiert med et ønske om å innføre flere formative vurderingsmetoder ($r = .348, p < .01$) og færre antall års undervisningserfaring ($r = -.252, p < .01$). Som vist i avsnittet over er det å ønske kurs også assosiert med å oppleve barrierer ($r = .310, p < .01$) og høyere grad av barrierer ($r = .265, p < .01$) i implementering av nye vurderingsmetoder. Det å ønske kurs er også assosiert med lavere grad av enighet i U20. Endrings skepsis ($r = -.274, p < .01$). Ser vi på utsagnsvariablene separat er variabelen signifikant assosiert med uenighet i U1 at nye metoder er forbigående trender ($r = -.283, p < .01$), U3 at forelesning er den beste måten å undervise på ($r = -.252, p < .01$), U7 at studenter vil være motvillige til nye metoder ($r = -.184, p < .05$). I tillegg er det å ønske kursing korrelert med en høyere grad av enighet i U8 at vurdering bør harmonere med

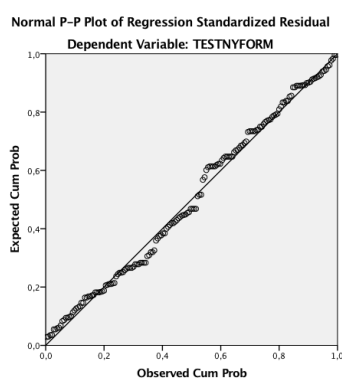
læringsutbyttebeskrivelser ($r = .158$, $p < .05$), og til slutt U10 at flere karakterer er bedre ($r = .208$, $p < .01$).

3.3 Multivariate analyser

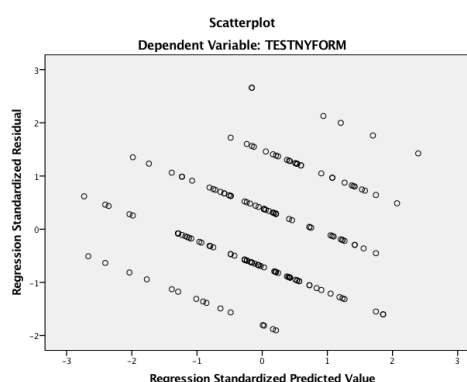
Regresjonsanalyser ble gjennomført for å belyse forskningsspørsmål 4: *I hvilken grad kan bruk av pedagogiske redskaper, tilsagn til utsagn om undervisnings- og vurderingspraksis, barrierer og demografiske variabler (kjønn, alder og undervisningserfaring) forklare variansen i bruk av formative og summative vurderingsmetoder?* Over har vi sett hvordan berøring med antall formative og summative vurderingsmetoder korrelerer med andre variabler. Gjennom resultatene av de stegvise regresjonsanalysene under får vi svar på hvilke variabler som kan være med å forklare variansen i variablene 1. Antall formative og 2. Antall summative når de ulike variablene er kontrollert for hverandre.

3.3.1 Variasjon i antall formative vurderingsmetoder. Forutsetningen om linearitet, normalitet og homosedasitet kan sies møtt basert på visuell inspeksjon av P – plot av forventet mot observert standardisert regresjons residual (Figur 5 under) og spredningsplot (Figur 6 under) av standardisert regresjons residual mot standardisert regresjon predikert verdi.

Figur 5 -P plot



Figur 6 - Spredningsplot



Residualene var også uavhengige i tilstrekkelig grad basert på Durbin-Watson verdi 1,128 (Field, 2009). Det var heller ingen antydninger til multikollinearitet basert på undersøkelse av delkorrelasjonsmatriser og toleranse verdier større enn 0.1, og VIF mindre enn 10 (Pallant, s.158). I et normalt distribuert utvalg kan man forvente mindre

enn 1 % som faller utenfor kritisk verdi på $-.3$ og $.3$ (Pallant, s. 160). Det var ingen som var utenfor $-.3$ og $.3$ men 3 cases (1,7 %) med standard residualer utenfor $-.2$ og $.2$. Dette er innenfor kritisk grense for 5 % (Field, 2009, s. 247). Basert på Mahalanobis distance og verdi på 18.47 (df = 4) er alle innenfor kritisk grense, og samtlige hadde Cook's distance under 1, noe som også antyder at det ikke er store problem i datasettet (Pallant, s. 160). Forutsetningene som regresjonsanalysen bygger på kan sies møtt i tilstrekkelig grad basert på gjennomgangen over. Regresjonsmodellen er statistisk signifikant på $<.0005$ nivå og indikerer at de uavhengige variablene i modellen samlet kan forklare 19,3 % av variansen i 1. Antall formative ($R_2 = .193$). Justert R_2 er $.173$ hvilket gir en bedre indikasjon med tanke på generalisering til populasjonen (Pallant, s.160, Field, 2009, s. 235). Utsagnsvariabel U1 har det høyeste unike signifikante bidraget til modellen ($\beta = -.272$). Tabell 9 under viser de øvrige variablenes bidrag. Oppsummert kan man si at resultatet antyder at det å være uenig i at nye metoder er forbigående trender, enig i at forelesninger er ineffektivt for læring, å være kvinne og ha lengre undervisningserfaring sammen kan være med å forklare bruk av mer formative vurderingsmetoder. Regresjonskoeffisienter og standardfeil vises i Tabell 8 under.

Tabell 8 - Oppsummering av regresjonsanalysen

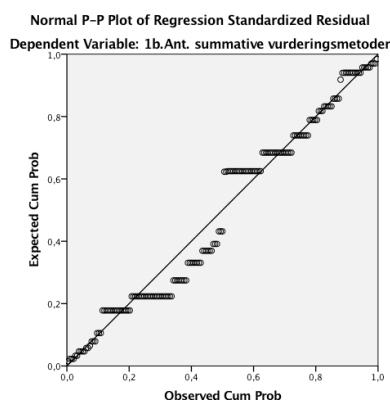
Variabel	<i>B</i>	SE _B	β
1. Antall formative	1.879	.388	-
U1. Nye metoder -forbigående trender	-.307	.082	-.272***
ERF. Undervisningserfaring	.208	.059	.259***
K/M. Kjønn	-.439	.147	-.214**
U9. Forelesninger -ineffektivt	.143	.071	.144*

Note: * $p < .05$, ** $p < .01$, *** $p < .0005$, *B* = ustandardisert regresjonskoeffisient, SE_B = standardfeil av koeffisient, β = standardisert koeffisient

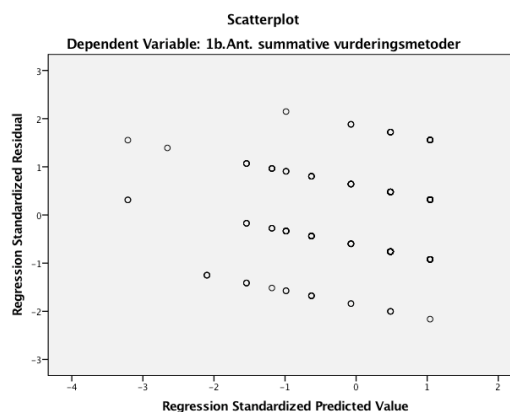
3.3.2 Variasjon i antall summative vurderingsformer.

På samme måte som i regresjonsanalysen over ble analyser gjort for å sjekke om forutsetninger ble møtt i tilstrekkelig grad. Forutsetningen om linearitet, normalitet og homosedasitet ble inspisert gjennom visuell inspeksjon av P - plot av forventet mot observert standardisert regresjons residual (Figur 7 under) og spredningsplot (Figur 8 under) av Standardisert residual mot standardisert predikert verdi.

Figur 7 - P plot



Figur 8 – Spredningsplot



Residualene er uavhengige basert på Durbin-Watson verdi 2,176 og det var ingen bevis på multikollinearitet basert på undersøkelse av delkorrelasjonsmatrisen og toleranse verdier større enn 0.1, og VIF mindre enn 10. 5 cases (2,8 %) har standard residualer utenfor -2 og 2 . Mahalanobis distance sin maksimum verdi er under kritisk verdi på 13.82 (df = 2). Cook's distance er under 1. Regresjonsmodellen er statistisk signifikant på $<.001$ nivå og antyder at de uavhengige variablene samlet kan forklare 8 % av variansen i 2. Antall summative ($R^2 = .079$, justert $R^2 = .068$). Slik variablene er kodet antyder resultatet at det å oppleve barrierer og ha lengre undervisningserfaring predikerer berøring med flere typer summativ vurderingsformer. Regresjonskoeffisienter og standardfeil vises i Tabell 9 under.

Tabell 9 - Oppsummering av regresjonsanalysen

Variabel	B	SE_B	β
2. Antall Summative	.617	.247	-
ERF. Undervisningserfaring	.475	.162	.217**
8b. Barrierer Nei/Ja	.130	.051	.191*

Note: * $p < .05$, ** $p < .01$, B = ustandardisert regresjonskoeffisient, SE_B = standardfeil av koeffisient, β = standardisert koeffisient.

4. Diskusjon

Målet for studien var å finne ut hvordan emneansvarlige på UiB forholder seg til pedagogiske redskaper som læringsutbyttebeskrivelser og nye vurderingsmetoder. Spørreundersøkelsen har generert data som gir oversikt over hvordan 176 emneansvarlige svarte på spørsmål knyttet til nåværende og ønsket bruk av pedagogiske redskaper, grad av tilsagn til utsagn om undervisnings- og vurderingspraksis og barrierer knyttet til implementering av nye vurderingsformer. I resultatkapittelet over har vi gjennom statistisk bearbeiding av data sett hvordan visse demografiske variabler og holdningsvariabler samvarierer og/eller forklarer variansen i emneansvarliges vurderingspraksis på statistisk signifikansnivå .05 eller .01. I teorien skal dette bety at det er 95 % - 99 % sjanse for at resultatet ikke er tilfeldig, og at vi basert på utvalget kan si at variablene samvarierer slik i populasjonen. I tolkning av resultatene må man alltid legge til grunn en viss grad av usikkerhet. Som Pallant skriver: "Statistics are used to make rational decisions under conditions of uncertainty" (2010, s. 134). I diskusjonen videre vil jeg også ta med meg at selv om noe er statistisk signifikant, betyr det for eksempel ikke at resultatet er viktig i praksis, at forskningsdesignet er godt, at reliabilitet og validitet på målene nødvendigvis er bra, eller om disse er blitt bearbeidet i tråd med ulike statistiske metoders forutsetninger. Resultatene må derfor vurderes i lys av flere momenter enn statistisk signifikans (Tabachnick & Fidell, 2013, s. 65) og dette skal diskuteres i det følgende.

4.1 Begrensninger med studien

4.1.1 Utvalgets representativitet og studiens generaliserbarhet. For det første er svarprosenten i denne spørreundersøkelsen 13,5 %. Lav responsrate er vanlig i selvadministrerte spørreundersøkelser, men bidrar til at utvalgets representativitet har høyere grad av usikkerhet (Tabachnick & Fidell, 2013, s. 69). Feilmarginen er på 6,9 % (95 % konfidensintervall) og power nivå for de statistisk analysene gjennomført her over anbefalt nivå på 0.8 (Field, 2009). Selv om frafallsanalysen ikke tyder på det, kan det ikke utelukkes at respondenter og ikke-respondenter er systematisk forskjellige fra hverandre. Det er et flertall av menn, undervisere med lengre undervisningserfaring og aldersgruppen 41-50 har høyest frekvens, men som redegjort for i metodekapittelet er det på bakgrunn av statistikk fra andre studier som involverer vitenskapelig ansatte sannsynlig at også populasjonen har skjevheter i disse demografiske variablene.

4.1 2 Variabler og statistisk behandling. Hvordan variablene er omkodet og bearbejdet før statistiske analyser er redegjort for i metoddelen. Variablene i denne undersøkelsen er ikke uten lyte, som få er i samfunnsvitenskapelig forskning (Pallant, 2010). Forutsetningene for de statistiske testene ble møtt i akseptabel grad, men selv om analysene er robuste, kan det tenkes at resultatene viser statistisk signifikant grad av sammenheng der det ikke er det, eller omvendt (Type 1 og Type 2 feil). Gjennomfører man 100 statistiske tester er det med 95 % konfidensintervall sannsynlig at 5 av dem viser feil resultat. Basert på antallet tester gjort i denne undersøkelsen er det en reell sjans for at slike feil kan ha forekommet. Da visse variabler måtte manuelt omkodes for bruk i statistiske analyser er det også en økt sjans for feil. Selv om dette arbeidet ble gjort nøyaktig og sjekket flere ganger kan man ikke helgardere seg mot feil gjort med inntasting eller lignende (Ringdal, 2013). Et annet usikkerhetsmoment er knyttet til outliers. Valget om å ikke transformere variabler eller fjerne outliers er begrunnet i metodekapittelet. Det kan tenkes at resultatet i noen tilfeller viser sterkere eller svakere assosiasjoner enn sammenhengene som eventuelt eksisterer i populasjonen (Tabachnick & Fidell, 2013).

I analyser er utsagnsvariablene med Likert-skala som i utgangspunktet er ordinalnivå, behandlet som om de var på intervallnivå, noe som er vanlig praksis (Tabachnick & Fidell, 2013). Field (2009) problematiserer og fraråder dette, selv om han også anerkjenner at det som oftest blir gjort på denne måten. I valget jeg har tatt har jeg støttet meg til Ringdal (2013) som skriver at tendensen i dag heller er å se på hvor robuste analyseteknikkene er, enn å la målenivå styre. ”Søking etter nye og interessante mønstre i data bør ikke avgrenses ved at målenivåene tolkes for bokstavelig” (s. 92).

Når det gjelder de statistiske analysene utført skal noen begrensninger ved regresjonsanalysen nevnes. Ved bruk av stegvis regresjon lot jeg SPSS basert på rene matematiske utregninger plassere variabler i modellen. Dette er vanlig når testen er eksplorerende og variablene ikke er valgt ut med støtte i tidligere forskning eller teori (Pallant, 2010). En annen forsker kunne inkludert andre også for meg ukjente variabler i modellen som kunne forklart variansen i den avhengige variabelen bedre. Det ”tredje variabel problemet” er muligheten for at en ukjent variabel kan ha en effekt på forholdet mellom to andre variabler (Field, 2009). Kontrollert for en ukjent variabel er det mulig at også resultatene fra korrelasjonsanalysene hadde blitt justert.

Noen ord skal også skrives om kausalitet. Å påvise årsakssammenhenger er ikke vanlig, og som oftest ikke et oppnåelig mål i en eksplorerende deskriptiv-analytisk

undersøkelse som denne. Selv om korrelasjoner angir retning på forholdt, er slike analyser i seg selv ikke tilstrekkelig for å påvise årsakssammenheng (Danielsen, 2013, s. 143; Pallant, 2010). Regresjonsanalyser forutsetter en kausal sammenheng mellom variablene, men i tolkningen av resultatet må man ta med seg at vi ikke kan si noe om årsak - virkning, selv om det skulle være dekning for å anta det basert på for eksempel tidsrekkefølgen i variablene (Befring, 1998).

4.1.3 Andre begrensninger. Det at spørreundersøkelsen også er utformet med tanke på et forskningsprosjekt utover rammen av denne masteroppgaven, har gitt seg utslag i at visse spørsmål har blitt formulert med tanke på å inkludere samtlige vitenskapelige ansatte, ikke kun emneansvarlige. Noen spørsmålsformuleringer hadde vært tydeligere om populasjonen vi ønsket å nå var den samme. Blant annet ble det i spørsmål 1 spurt om hvilke vurderingsmetoder som ble brukt i emnene de var ”involvert i”. Hadde målgruppen vært kun emneansvarlige, hadde formuleringen vært emnene de har ”ansvar for”.

To spørsmål ble utelatt på grunn av min forforståelse av at samtlige undervisere hadde fullført et kurs i universitetspedagogikk, og at det kun er et fåtall som delte ansvaret eller hadde emneansvar for flere emner. Et spørsmål om hvorvidt den emneansvarlige har fullført kurs hadde vært en interessant variabel, men påvirker ikke direkte de resultat jeg har rapportert over. Spørsmål om hvilke emner respondentene refererte til ble av hensyn til anonymitet utelatt, men det hadde vært rimelig å spørre om antall emner de var emneansvarlig for. Dette er en større begrensning. En emneansvarlig som krysser av på flere antall vurderingsformer fordi de har flere emner, kan ikke direkte sammenlignes med en emneansvarlig som kun har ett emne. Dette skaper et ekstra usikkerhetsmoment knyttet til tolkning av resultatene, og må tas høyde for. Det kan tenkes at det er systematiske forskjeller i hvilke respondenter som eventuelt har ansvar for flere emner, for eksempel lengre undervisningserfaring. Kontrollert for antall emner kunne resultat knyttet til denne variabelen bli rapportert og tolket med større grad av sikkerhet. Når det er sagt er likevel hovedpoenget å få oversikt over hvilke vurderingsformer *emneansvarlige* er i berøring med. Dette er interessant fordi den emneansvarlige bærer en stor del av ansvaret for hvilke vurderingsformer som blir tatt i bruk. Ingen av resultatene blir tolket som mål på antall vurderingsformer i ett emne, eller som mål på antall vurderingsmetoder i bruk på UiB.

Mangel på dybde og detalj er en annen begrensning. Da prosjektplanen for masteroppgaven ble godkjent, var det med kommentar om at planen om å både bruke spørreundersøkelse og kvalitativt intervju var for omfangsrik. Hvorfor har blitt tydeligere for undertegnede nå, men med en kvalitativ tilnærming gjennom intervjuer kunne man fått dypere innsikt i hvorfor emneansvarlige har svart som de har gjort. Planen om å følge opp med intervjuer består, men utenfor rammen av denne masteroppgaven. Når resultatene i det videre skal tolkes og diskuteres, blir det gjort i lys av de overnevnte begrensninger.

4.2 Diskusjon av Resultat

Oppsummert tyder svar gitt fra emneansvarlige i spørreundersøkelsen på at de aller fleste emneansvarlige på UiB er i berøring med både summative og formative vurderingsmetoder. Å være i berøring med flere formative vurderingsmetoder er blant annet assosiert med lavere grad av enighet i endrings skeptiske utsagn. De aller fleste tar i bruk læringsutbyttebeskrivelser, og de fleste både definerer og formidler vurderingskriterier. De aller fleste opplever også barrierer. Denne korte oppsummeringen av resultat kan tyde på at vi har med en gruppe progressive emneansvarlige som enten har gjort nylige endringer i praksis eller ønsker å gjøre det, og at vurderingspraksis på UiB speiler definisjonen av kvalitet på dette området vist gjennom kvalitetssikringssystemet. Bildet er mer komplekst enn så, og introduksjonen viste forhåpentligvis at det gjelder både tema og kontekst. Resultatene skal i det videre tolkes i lys av dette, og av det følger det at diskusjonen vil føre til langt flere nye spørsmål enn svar.

4.2.1. Summativ og formativ vurdering i bruk på UiB. Med få unntak oppgir respondentene at ulike typer skriftlig eller muntlig avsluttende eksamen blir tatt i bruk i sine emner. Dette resultatet kan med det sies å være i tråd med den tidligere omtalte studien ved UiB (2012) som slo fast at slike summative vurderingsmetoder sin posisjon over tid står sterkt. Som redegjort for over, tar vår undersøkelse utgangspunkt i den emneansvarlige og ikke i emner. Resultatene kan derfor ikke sammenlignes direkte, men de peker mot samme trend. Dysthe et al. (2006) viser til Gipps (1994) og Shepard (2001) og skriver at etter at teori som tradisjonelle vurderingsformer bygger på for lengst er forlatt, består gjerne vurderingsformene (s. 49). Selv om ”nyere” forskning og teori legger et solid grunnlag for bruk av mer formative vurderingsmetoder, kan man

likevel spørre seg om dette betyr at mer tradisjonelle vurderingsformer er passé. Med tanke på at høyere utdanningsinstitusjoner både skal støtte hensikten om kvalifikasjon og læring, kan både bruk av vurdering med fokus på kontroll av oppnådd læringsutbytte (summative) og vurdering med fokus på å støtte læringsarbeidet mot ønsket læringsutbytte (formative) forsvares. Det mer interessante spørsmålet i lys av læringsparadigmet man nå skal operere i, er kanskje heller hvorvidt hensikten om å støtte studenters dype læring kan ivaretas ved valgt vurderingsform, her eksemplifisert gjennom skriftlig skoleeksamen som var den vurderingsmetoden flest respondenter oppga.

Med Black (2015) kan man si at mer formativ bruk av summative vurderingsmetoder er mulig, og at summativ vurdering også må sees som en positiv del av læringsprosessen. Men en summativ vurderingsform som skriftlig skoleeksamen har visse rammer som det er vanskelig å komme utenom. For det første er denne tradisjonelle vurderingsformen gjennom teoretisk grunnlag, av sedvane eller begge deler, ofte knyttet til en undervisningshverdag preget av overføring av kunnskap, og studenters bruk av mer overflatiske læringsstrategier. Som Ramsden skriver, fra studenters ståsted er vurderingen den faktiske læreplanen (curriculum) (1992, s. 182). For det andre utelukker det faktum at en skoleeksamen gjennomføres etter endt undervisning, en læringsprosess der studenter får trening i å skrive, får tilbakemeldinger på arbeidet og deretter får muligheten til å bruke disse mot ferdig produkt som leveres til karaktersetting. En praksis som forskning har funnet kan være knyttet til metakognisjon og dypere læring. På bakgrunn av dette kan man si at bruk av avsluttende eksamen som eneste vurderingspunkt tvilsomt er egnet til å støtte samt vurdere studenters oppnådde læringsutbytte når dypere læring er målet. Dersom begge hensikter skal ivaretas er bruk av vurderingsmetoder som bedre kan ivareta både det summative og formative element, som for eksempel mappevurdering, en mulig løsning. En annen løsning kan være å supplere med en formativ vurderingsmetode som kan støtte det formative element. I Michelsen og Aamodt sin sluttrapport etter Kvalitetsreformen blir nettopp dette omtalt som en vanlig strategi (Michelsen & Aamodt, 2007).

Studien fra 2012 (UiB) kunne ikke si noe sikkert med tanke på hvorvidt formative vurderingsmetoder i økende grad ble brukt på UiB. Mappevurdering ble redusert i perioden 2008 – 2012, mens arbeidskrav uten karakterer økte. Igjen, resultatene fra vår spørreundersøkelse kan ikke sammenlignes direkte, men ca. 30 % av

respondentene svarer at de tar i bruk mappevurdering, ca. 80 % svarer at de bruker tilbakemeldinger fra underviser, medstudenter eller begge på skriftlige arbeid eller øvelser, og samlet sett er det ca. 90 % av emneansvarlige som er i berøring med det som i denne undersøkelsen er blitt definert som typiske formativ vurderingsmetoder.

4.2.1.1 Kjønn, undervisningserfaring, endringskepsis og vurderingsformer.

Bivariat korrelasjonsanalyse indikerer at det å bruke flere ulike typer summative vurderingsmetoder er signifikant korrelert med lengre undervisningserfaring og det å oppleve barrierer. Regresjonsanalysen antyder at samme variabler kan være med å forklare variansen i antall summative vurderingsmetoder i bruk. At lenger undervisningserfaring er forklarende, kan tolkes som at visse emneansvarlige kan ha satt seg fast i gamle tanke- og praksismønstre. At korrelasjonsanalysen også viste at lengre undervisningserfaring er assosiert med å ikke ønske kursing i hvordan man kan planlegge eller gjennomføre vurdering, kan tyde på det samme. Men å ikke ønske kursing kan også forklares med at man tidligere har deltatt på kurs. Et annet funn som utfordrer en tolkning om at undervisere med lengre erfaring har satt seg fast i gamle mønstre er at lengre undervisningserfaring også var med å forklare deler av variansen i berøring med flere antall formative vurderingsmetoder. En alternativ tolkning om at undervisningserfaringen gjør en tryggere i å prøve ut nye metoder er med det vel så sannsynlig.⁵ I tillegg kan det at opplevelsen av barrierer også var forklarende for berøring med flere summative vurderingsmetoder, også støtte en tolkning om at undervisere ønsker endring, men blir stoppet av barrierer. Hvordan man kan tolke korrelasjonen med barrierer skal diskuteres nærmere i et senere avsnitt. I motsetning til variabel 2. Antall summative kan vi av korrelasjonsmatrisen (Tabell 7 s. 56) se at variabel 1. Antall formative også slår ut signifikant med en rekke utsagnsvariabler. Dette kan forstås i lys av at de fleste emneansvarlige har relativt lik praksis når det kommer til summativ vurdering, og at forskjellene hovedsakelig omhandler berøring med flere typer formativ vurderingsmetoder. I regresjonsmodellen med 1. Antall formative som avhengig variabel var, i tillegg til lengre undervisningserfaring, uenighet i utsagnsvariabelen ”Jeg ser ingen grunn til å forbedre min undervisningspraksis siden det i min profesjonelle utvikling og karriere ikke er vurdert like høyt som forskning” og enighet i at ”Forelesninger er en ineffektiv måte for studenter å lære på” sammen med

⁵ Som omtalt i begrensninger med studien kan det også tenkes at resultatet hadde blitt justert kontrollert for antall emner, da jeg ikke kan utelukke at lengre undervisningserfaring har sammenheng med emneansvar i flere emner.

det å være kvinne forklarende for bruk av flere formative vurderingsmetoder. Regresjonsanalysen antyder med andre ord at disse variablene samlet sett er viktig når nye formative vurderingsmetoder skal implementeres. En tolkning er at lengre undervisningserfaring gjøre en tryggere i rollen som underviser, man har sett behov for endring og ønsker å forbedre praksis tross mangel på insentiv, er klar for å tenke nytt i forhold til undervisningsformer og innføre formative vurderingsmetoder som i større grad også overlapper med undervisning. Det at kjønn skal være forklarende er mer utfordrende å tolke. Da kjønn slår ut signifikant i flere korrelasjoner, skal dette bli diskutert videre i sammenheng med disse funnene i senere avsnitt. I denne omgang skal det kun nevnes at bivariat korrelasjonsanalyse også viste at det å være kvinne hadde en signifikant, om enn liten, assosiasjon med å være i berøring med flere formative vurderingsmetoder.

Resultatene rapportert og diskutert over antyder at den summative vurderingen ikke lenger er ensomt regjerende, og at formative vurderingsmetoder også i stor grad er i bruk. Varierte vurderingsformer er noe av det kvalitetssystemet på UiB er ment å sikre, og resultatene over kan tyde på en grad av suksess. Men, det er flere aspekt som må tas med i betraktningen før man kan konkludere om hvorvidt hensikten om studenters dype læring er ivaretatt. For det første er det ikke sikkert at det som oppgis som bruk av formativ vurdering faktisk er det med tanke på innhold. I de tilfeller der den formative vurderingsmetoden alene kan sies å støtte hensikten om dypere læring, kan man spørre seg hvorvidt bruk av både en tradisjonell summativ vurderingsform og formative vurderingsmetoder i samme emne sammen effektivt støtter denne hensikten.

4.2.1.2 Vurderingsform, innhold og hensikt. At formative vurderingsmetoder i stor grad er i bruk, og i stor grad ønskes brukt, kan være en indikasjon på at konstruktivistiske ideer med fokus på studenters aktive og dype læring er reflektert i praksis. Men en slik konklusjon forutsetter også at innholdet i vurderingsformen gjenspeiler teorien den er bygget på. Det å ta i bruk en typisk formativ vurderingsform sier i seg selv ikke nok om innholdet støtter en formativ hensikt. Dysthe et al. (2006) skriver at i implementering av nye vurderingsformer etter kvalitetsreformen i noen tilfeller ble gjort uten forståelse for hva dette faktisk innebærer. Dette kan være med å forklare stor variasjon i hvordan for eksempel mappevurdering er praktisert (Dysthe, et al., 2006; Michelsen & Aamodt, 2007). Hvorvidt paradigmeskiftet fra utdanning til læring (Barr & Tagg, 1995), eller fra en objektivistisk til en konstruktivistisk (Biggs,

1996) tankegang har skjedd også i praksis, kan med det hverken tall fra denne undersøkelsen eller UiBs egen studie fra 2012 svare på.

Dersom det er slik at undervisere må arbeide ut fra et konstruktivistisk teoretisk ståsted for å best sikre den formative hensikten, kan endringen noen undervisere må gå gjennom sies å være av fundamental karakter. Men, det kan også tenkes at dette er et overkomplisert bilde, og at det som Dysthe et al. (2006) skriver, er snakk om trinnvise heller enn fundamentale endringer. Dette begrunnes med at det stort sett handler om at undervisere må gjøre mer av noe som er kjent fra egen tid som student eller tidligere undervisningspraksis; ”mer skriving, mer aktivisering, mer tilbakemelding og mer oppfølging” (s. 52). Det kan også tenkes at dette er en overforenkling. Ressonnementet forutsetter for det første at underviseren faktisk har erfaring med skriving, aktivisering, tilbakemelding og oppfølging. For det andre at det er et ønske om å gjøre mer av disse aktivitetene, som kan tenkes er knyttet til en forståelse av hvorfor det er viktig. Følgelig kan det likevel være avgjørende at emneansvarlige har en konstruktivistisk forforståelse eller har kjennskap til aktuell læringsteori og forskningsresultat. Watkins et al. (2015) skriver blant annet at en effektiv måte å forandre måten studenter lærer på er å forandre vurderingspraksis, men at underviserens forståelse av vurderingens rolle i forhold til studenters læring også kan være avgjørende (s. 283).

4.2.1.3 Vurderingsmetoder og helhet. Hovedfunnet når det gjelder vurderingsformer er muligheten at de aller fleste emneansvarlige virker å være i berøring med både tradisjonelle summative og nyere formative vurderingsmetoder. At det er vanlig å ta i bruk formative vurderingsmetoder i tillegg til tradisjonell slutteksamen, er som nevnt i tråd med Michelsen og Aamodt (2007) sine funn i evaluering av Kvalitetsreformen. Over har jeg diskutert hvorfor det ikke i utgangspunktet trenger å være snakk om enten eller, men argumentert for at bruk av summative vurderingsmetoder som avsluttende eksamen, kanskje må støttes av formative elementer utenfor rammen av nevnt vurderingsform for å kunne støtte dypere læring mot ønsket læringsutbytte. Men en slik praksis gir også grunnlag for å spørre hvorvidt det er et meningsskapende samsvar i bruk av både summative og formative vurderingsmetoder, at de ikke drar i ulike retninger og om de sammen faktisk sammen støtter opp om hensikten om dyp læring. Basert på tidligere forskning om hva som styrer studenters innsats, og ulike fallgruver knyttet til dette (Dysthe, et al., 2006; G. Gibbs, 2010; Ramsden, 1992) kan man se for seg følgende scenario: Der et emne inneholder både en

klassisk vurderingsform som skriftlig skoleeksamen som alene angir slutt karakter, i tillegg til obligatorisk skriftlige oppgave(r) underveis i semesteret, gir dette visse signaler. Med mindre studenten vet at det er en direkte sammenheng mellom de obligatoriske oppgavene og avsluttende eksamen, altså at arbeidsinnsatsen med oppgavene vil gi uttelling, vil studenter gjerne legge inn minimal innsats tilsvarende det å få bestått på oppgaven(e). Tiden og energien vil brukes til å lese og pugge til eksamen som er karaktergivende, gjerne basert på puggehefter eller tidligere eksamensoppgaver. Som Michelsen og Aamodt skriver i sin sluttrapport i Evaluering av Kvalitetsreformen (2007), ”Studentenes måte å studere på er preget av at de gjør det de må gjøre (...)” (s. 9). På den måten kan den skriftlige skoleeksamen overskygge ønsket effekt av de formative element. Det å supplere med en ny vurderingsmetode vil ikke automatisk øke kvalitet eller støtte den formative hensikt. Raaheim (2016) skriver at slik han ser det, er ikke grad av endring etter Kvalitetsreformen det viktigste, men hvordan praksis er forankret og begrunnet (s.82). Overordnet det å implementere en ny vurderingsform er i lys av teori om meningsskapende helhet (CA) hvorvidt den passer inn i helheten, og hvorvidt den sammen med de andre delene støtter det man ønsker studenter skal kunne, uttrykt gjennom læringsutbyttebeskrivelsene.

4.2.2. Læringsutbyttebeskrivelser, vurderingskriterier og meningsskapende

helhet. Som redegjort for i introduksjonen er det en forventning om at læringsutbyttebeskrivelser skal kunne være et pedagogisk redskap i undervisning og vurdering. Overkommer man begrepsforvirring og lykkes man i å gjøre nødvendige justeringer for å sikre epistemologisk gyldighet, kan læringsutbyttebeskrivelser være en generator for helhetlig tankegang i undervisning og vurderingsprosesser.

Emneansvarlige fikk gjennom spørreundersøkelsen spørsmål om tidspunkt læringsutbyttebeskrivelsene blir tatt med i arbeidet med vurdering. Kun 4 % svarer at de aldri tar i bruk læringsutbyttebeskrivelser. De aller fleste tar i bruk læringsutbyttebeskrivelser også utover planlegging og oppbygging av emnet. Resultatene antyder med andre ord at læringsutbyttebeskrivelser av de fleste er i bruk også utover pålagte krav om å formulere dem. De aller fleste er også enig i utsagnet om at det er viktig at vurdering harmonerer med læringsutbyttebeskrivelser, og enighet i dette utsagn korrelerer også signifikant med bruk av læringsutbyttebeskrivelser på flere tidspunkt. 60 % av respondentene inkluderer dem i arbeid med karaktersetning og 45 % svarer at de tas i bruk ved samarbeid med ekstern sensor. At ikke flere bruker

læringsutbyttebeskrivelsene i sammenheng med vurdering kan tyde på en manglende bevissthet rundt et helhetlig samsvar mellom vurderingsform og læringsutbyttebeskrivelser. Disse resultatene kan sies å være i tråd med Aamodt et al. sin studie fra 2014, der færrest hadde svart i stor grad når det gjaldt læringsutbyttebeskrivelser sin rolle i vurdering av studenters prestasjoner. En alternativ tolkning, og kanskje mer trolig tolkning med tanke på utfordringer knyttet til begrepet, er at man enda ikke har funnet løsningen på hvordan læringsutbyttebeskrivelser kan defineres slik at de egner seg som pedagogisk redskap i en helhetstankegang. Aamodt et al. (2014) skriver at med tanke på at læringsutbyttebeskrivelser er relativt nytt fenomen som man er pålagt å bruke og ikke nødvendigvis er godt forankret, er resultatet overraskende positivt.

At det å være kvinne i større grad er assosiert med enighet i utsagnet om at vurderingsformen bør harmonere med læringsutbyttebeskrivelser, er også et funn som er i tråd med resultatet fra Aamodt et al. (2014) sin studie. Her var det flest kvinner som svarte at læringsutbyttebeskrivelser i stor grad spilte en rolle (s.50). I tolkningen av dette resultatet må man for det første ta med seg at korrelasjonskoeffisienten var liten, for det andre at relativt få var uenig i påstanden. For det tredje var det ingen signifikant assosiasjon mellom kjønn og læringsutbyttebeskrivelser i bruk. Men, av de som var mindre enig i at vurdering bør gjenspeile læringsutbyttebeskrivelsene, var altså flest menn.

Ifølge UHR sine generelle beskrivelser skal vurderingskriterier knyttes til ”kunnskapsmål” i faget og i dagens kontekst kan man si at læringsutbyttebeskrivelser skal uttrykke disse. I lys av dette er det et interessant funn at det å bruke læringsutbyttebeskrivelser på flere tidspunkt er assosiert med å definere vurderingskriterier, og av de 72 % som definerer vurderingskriterier svarer ca. halvparten at de definerer dem basert på læringsutbyttebeskrivelsene. Å definere vurderingskriterier, basert på læringsutbyttebeskrivelser eller ikke, er også assosiert med å formidle vurderingskriteriene til studentene, som nærmere 80 % av respondentene oppgir at de gjør. At 22 respondenter (12,5 %) svarer at de ikke definerer kriterier, men formidler dem, er et funn som er utfordrende å tolke. Først og fremst vet vi ikke med sikkerhet hva respondentene legger i det å definere og formidle aktuelle vurderingskriterier. Flere studier har funnet at vurderingskriterier av ulike årsaker er vanskelig å formulere i skriftlig form (Nicol & Macfarlane-Dick, 2006), og det kan tenkes at det er en forvirring rundt hva vurderingskriterier er ment å være. Det at noen

mener man kan formidle vurderingskriterier uten å først definere dem, kan tyde på en oppfatning av at man ”kjenner en god oppgave når jeg ser den” og at spesifikke eksplisitte vurderingskriterier med det blir overflødig. Lang undervisningserfaring er ikke signifikant korrelert med å definere eller formidle vurderingskriterier, men det kan tenkes at faglig ekspertise vil kunne gjøre en i stand til å nettopp kjenne igjen en god oppgave uten å konsultere spesifikke vurderingskriterier. I tillegg til at det å ikke definere vurderingskriterier kan tyde på en normbasert vurderingspraksis, er det et problem dersom dette vanskeliggjør studenters arbeid med å internalisere ønsket akademisk standard i den grad studenter ikke får spesifikke kriterier å måle egen læringsprosess opp mot.

Selv om det å utforme læringsutbyttebeskrivelser og vurderingskriterier er en relativt liten del av den emneansvarliges ansvar, kan arbeid med disse sies å ha avgjørende betydning for et meningskapende samsvar og gjennomsliktighet i vurderingsprosessen. Som redegjort for innledningsvis støtter både helhetlig samsvar i undervisningen i seg selv (Xiaoyan Wang, et al., 2013), og transparens i vurdering studenters dype læringsarbeid (Black & Wiliam, 1998; Brown, 2004-05). Samlet sett antyder disse resultatene knyttet til læringsutbyttebeskrivelser og vurderingskriterier at UiB har gode muligheter for forbedring om man skal etterleve krav om kriteriebasert vurdering, at disse skal knyttes til læringsutbyttebeskrivelser og oppfordringer til en transparent og helhetlig vurderingspraksis.

4.2.3. Å si seg enig eller uenig i endrings skeptiske utsagn. Resultatene antyder at bruk av pedagogiske redskaper som vurderingsformer, læringsutbyttebeskrivelser og vurderingskriterier samvarierer med grad av enighet i endrings skeptiske utsagn. Som redegjort for i metoddelen, er utsagnene emneansvarlige ble bedt om å gi sitt tilsvaret til, variasjoner av det universitetspedagoger ved UiB har hørt i sitt virke. Når utsagnene har blitt omtalt som ”endrings skeptiske” og ”endrings positive” er dette en karikert betegnelse. Et eksempel der betegnelsen er grovt overforenklet er utsagnet ”*Forelesninger er den beste måten å undervise på i høyere utdanning*”. Som med vurderingsform, er det viktige med undervisningsform hvilket innhold man legger i formen, og hvorvidt det støtter studenten i sitt læringsarbeid mot ønsket læringsutbytte. Av alle utsagn er det også flest som har svart ”nøytral” på dette spørsmålet (ca. 40 %). Hvilket innhold de som svarer ”Helt enig” på dette spørsmålet fyller i denne undervisningsformen kan ikke sies noe om basert på denne undersøkelsen.

Som redegjort for tidligere ble dette utsagnet sammen med fire andre utsagn basert på høy grad av interkorrelasjon samlet i ett mål kalt ”Endrings skepsis”. Det store flertallet angir lavere grad av enighet i endrings skeptiske utsagn (se tabell 4 på side 50). Lav grad av enighet i endrings skeptiske utsagn er assosiert med bruk av flere formative vurderingsformer, å bruke læringsutbyttebeskrivelser i utstrakt grad, å formidle vurderingskriterier til studenter, og et ønske om å lære mer om hvordan man skal planlegge og gjennomføre vurdering gjennom kursing. Hva man tenker om pedagogiske redskaper er med andre ord til en viss grad reflektert i praksis, og forskning indikerer også at skal man få til en endring i undervisning og læring, må det først bli en endring i tanker og ideer rundt det samme (Watkins, et al., 2005, s. 283). I tolkning av disse resultatene må man likevel huske på at slike korrelasjoner ikke angir kausalitet. Det er vanlig å tenke at praksis springer ut fra tanken, men det kan tenkes at også det man gjør, kanskje også som et resultat av ytre påvirkning, i sin tur er med på å forme det man tenker. Et eksempel kan være fokuset på skriftlige arbeider som blant annet O. Dysthe har vært en forkjemper for, og kvalitetsreformen har ført med seg. Hvordan emneansvarlige ville stilt seg til utsagnet ”*Skriftlige arbeider forbedrer studenters læring*” før dette ble vanlig praksis på universitetet, vites ikke. Men det kan tenkes at det å prøve dette ut i praksis og få positiv erfaring med det, gjorde at nærmest alle i vår undersøkelse nå sa seg enig eller helt enig i dette utsagnet. Dysthe et al. (2006) rapporterer også at undervisere var positive til skriftlige arbeider og at grunnen oppgitt var at de erfarte at det virket.

Et siste punkt når det gjelder grad av endrings skeptiske utsagn gjelder kjønn og alder. Som allerede nevnt var det å være kvinne assosiert med enighet i utsagnet om at vurderingsformer bør harmonere med læringsutbyttebeskrivelsene. Kjønn var også signifikant korrelert med andre utsagn på den måten at kvinner ser ut til å være minst endrings skeptisk. De aller fleste er nøytral eller uenig i endrings skeptiske utsagn, men av mindretallet som er endrings skeptisk, er flest menn. Hvorvidt dette betyr at menn har tendenser til å være mer endrings skeptisk enn kvinner, kan ikke konkluderes basert på resultater fra denne undersøkelsen. En alternativ tolkning kan også handle om sosial ønskebarhet. Det kan tenkes at det oppfattes mer sosialt akseptert å si seg uenig i endrings skeptiske utsagn, og det kan tenkes at kvinner i større grad enn menn svarer endrings positivt, men at man likevel ikke er mer endrings positive enn menn. Når det er sagt, har kjønn vist seg signifikant i en rekke studier når det gjelder variabler knyttet til undervisningspraksis, med samme retning. Aamodt et al. (2014) er allerede nevnt, men

også i Norton et al. (2013) sitt studie var kjønn signifikant. Norton et al. viser til samme funn i Demetriou, Wilsow & Winterbottom (2009) og Zukov (2012) sine studier (Dysthe, et al., 2006). Felles for dem alle er at funn knyttet til kjønn ikke blir tolket og diskutert særlig dypt eller grundig. Dette kanskje fordi at den praktiske relevansen av betydningen av kjønn diskuteres, og i den grad man skal trekke konklusjoner trengs mer forskning som tar for seg akkurat dette.

Når det gjelder alder var det å være eldre assosiert med en større grad av enighet i at nye metoder er forbigående trender. En tolkning er at høyere alder tilsier større tendens til å være låst i innarbeidede rutiner og tillærte tankemønstre, at man rett og slett er endrings skeptisk. En annen og kanskje mer empirisk forankret forklaring er at høyere alder kan tilsie at man har sett nye metoder komme og gå, og erfaringen tilsier med det at nye metoder nettopp er forbigående trender.

4.2.4. Barrierer - en effektiv stopper for endring? Basert på den utstrakte bruken av ulike formative vurderingsformer på UiB kan det tyde på at selv om barrierer finnes, er de ikke effektive stoppere for endring. Andre funn kan nyansere dette bildet. Regresjonsanalysene viste at det å oppleve barrierer er med på å forklare variansen i antall summative vurderingsmetoder. Det kan tolkes slik at visse har forsøkt å innføre nye formative vurderingsmetoder, men at barrierene har stoppet dem. I tolkningen må man også ta med seg at spørsmålet om barrierer ikke skiller mellom dem som ikke en gang har forsøkt å endre praksis basert på en opplevelse av mange barrierer, de som eventuelt prøver, opplever barrierer og gir opp og de som overkommer disse barrierene. Det skiller heller ikke på implementering av nye formative eller summative metoder. Man kan med andre ord oppleve barrierer når man vil innføre en annen summativ vurderingsmetode, for eksempel muntlig eksamen eller hjemmeeksamen. Gjennom resultat fra korrelasjonsanalyser så vi i tillegg at det å oppleve barrierer, også i større grad, korrelerer med et høyere antall ønskede formative vurderingsmetoder. Det vil si at de som ikke tar i bruk disse vurderingsformene per i dag, men ønsker det, i stor grad også opplever barrierer. At barrierer både kan forklare det å ta i bruk flere summative vurderingsmetoder i regresjonsanalysen og samtidig samvarierer med ønsket bruk av formative vurderingsmetoder, kan derfor tolkes i retning av at en opplevelse av barrierer for visse per i dag er en effektiv stopper for en type vurdering som bedre støtter studenters dypere læring.

Resultatene som omhandler ulike typer barrierer er hovedsakelig i tråd med tidligere forskning. Et av resultatene som skiller seg fra tidligere forskning er usikkerhet knyttet til om studenter vil være positiv til innføring av nye metoder. Albers (2009) påpeker at studenter var motvillige til endring. Mål på hvorvidt emneansvarlige var usikker på det samme ved UiB i dag, var grad av enighet i utsagnet ”studenter vil være motvillige til nye metoder” der få sa seg enig (12), og barrieren: ”usikker på om studenter vil like det”, der kun 10 krysset av. En tolkning av dette funnet er at det nå er 8 år siden Albers sin studie, og undervisere vet at studenter de siste årene har blitt mer vant til variasjon i metode for undervisning og vurdering både i videregående opplæring og høyere utdanning.

At systembarrierer som ressursknapphet toppet listen av opplevde barrierer, er i tråd med tidligere forskning (Norton, et al., 2013, s. 248). Dette gjelder begrenset tid, for mye administrativt arbeid og kostnader. Dysthe et al. (2006) skriver om ”tidsklemmen” som ansatte opplever og arbeidsbelastningen summen av samtidige endringer gir (s.52). Det er nok få forunt å oppleve overflod av ressurser. I Kvalitet for utdanning (2012) påpekes det at forholdet mellom tilførte ressurser og kvalitet ikke nødvendigvis er lineært, men det finnes samtidig en anerkjennelse av at en rekke tiltak er avhengige av at det blir tilført ekstra ressurser (s.39). Ambisjonsnivået er høyt og forventning om endringsarbeid tydelig. Når helhetlig tankegang, bruk av læringsutbyttebeskrivelser som virkemiddel til å sikre helhet, varierte og tilpassede undervisnings- og vurderingsformer med fokus på studenters dype og varige læring er definert som kvalitet i utdanning også på UiB, kan det tenkes at tiltak for å sikre dette krever tilføring av ressurser. Under svaralternativet ”Annet” påpekte også flere manglende administrativ og teknisk support når endringer skal implementeres, spesifikt tekniske løsninger til mappe og UiB sin bruk av flere IT plattformer som ikke samfungerer. At teknikken og systemet emneansvarlige skal operere i skal fungere, er rimelig å forvente. Når det ikke gjør det, kan dette være en effektiv barriere for endring.

I tillegg til ressursmangel svarer ca. 20 % at lite støtte fra fagmiljø til å gjennomføre endringer er en barriere som oppleves. Kollegial støtte (Hanssen, et al., 2017) og kultur for vurdering på fakultetsnivå har vist seg viktig i endringsarbeid knyttet til innføring av undervisnings- og vurderingspraksis som støtter dypere læring (Knight & Trowler, 2000; Xueli Wang & Hurley, 2012). En tolkning er at lav grad av støtte til å gjennomføre endringer kan ha med mangel på kunnskap eller bevissthet rundt vurderingsformens avgjørende rolle i en helhetlig tankegang, eller det kan bunne i ulikt

syn på kvalitet i utdanning. Det som er viktig for å skape en kultur for kvalitet er kontinuerlig dialog rundt hva som er kvalitet, og å innta en aktiv rolle i endringsarbeidet (Stensaker, 2008) Det er ikke avgjørende at alle involverte har samme syn på hva kvalitet i utdanning er, i denne konfliktlinjen er det iboende mulighet til fortsatt utvikling. I Kvalitet i utdanning (UiB, 2012) er et av punktene for ønsket kultur for kvalitet at forslag til kvalitetsfremmende forslag blir ønsket velkommen (s.37).

Husebø og Moen (2017) forteller om kollegaveiledning for alle nyansatte i vitenskapelige stillinger på UiS og foreslår å opprette *undervisningsgrupper* på lik linje med *forskergrupper* samt å etablere og benytte begrepet universitetslærerarbeid. Slike initiativ og endringer kan være med å oppmuntre undervisere til å bruke knappe ressurser også på utvikling av undervisnings- og vurderingsarbeid. Meritteringsordninger blir også tillagt vekt i en kvalitetskultur for undervisning. Mangel på insentiv ble i denne undersøkelsen målt gjennom utsagnet ”*Jeg ser ingen grunn til å forbedre min undervisningspraksis siden det i min profesjonelle utvikling og karriere ikke er vurdert like høyt som forskning*”. Av alle utsagn var det dette som signifikant korrelerte med fleste vurderingspraksisvariabler. Høyere grad av enighet i dette utsagnet var assosiert med å bruke formative vurdering i mindre grad, å bruke læringsutbyttebeskrivelser på færre tidspunkt og å ikke spesifisere vurderingskriterier. Meritteringsordninger sender et signal om at kvalitetsarbeid i undervisning og vurdering blir verdsatt. Om det blir insentiv også for andre enn de som allerede er ”de beste” undervisere vil tiden vise.

Mangel på kunnskap er også for noen en opplevd barriere for endring. 16 % krysser av for begrenset kunnskap om alternative vurderingsformer og 10 % for begrenset kunnskap om vurderings – og læringsteori. Det er et interessant at av de som selv oppgir at mangel på kunnskap oppleves som en barriere, ønsker også nærmest alle å lære mer gjennom kursing. Dette kan tyde på at det finnes villighet til utvikling og endring, i alle fall hos de som selv oppgir at de opplever begrenset kunnskap som en barriere.

4.2.5 Endringsrom. At 2/3 av respondentene sier de ønsker kursing og at det store flertallet sier seg nøytral eller uenig i endrings skeptiske utsagn, kan vise at det eksisterer et endringsrom hos de aller fleste emneansvarlige. At svaralternativet i ”noen grad” har høyest frekvens (65 %) på spørsmål om de opplever at egen vurderingspraksis er lik det de tenker er ideell praksis, kan også tyde på det samme. Hva emneansvarlige

ser for seg av ”ideell” praksis, og hvorvidt denne oppfatningen er basert på en bevissthet rundt vurderingens viktige rolle i en helhetlig tankegang med fokus på læring vites ikke. Resultatene sier heller ikke noe om at en anerkjennelse av at man kun i ”noen grad” har vurderingspraksis lik en ideell praksis innebærer et ønske om å bevege seg mot en ideell praksis. Dette funnet kan tolkes i sammenheng med at 74 % av de som svarer ”i noen grad” ønsker kursing. Også 69 % av de som svarer at de ”i stor grad” synes egen vurderingspraksis er lik ideell, og de 3 respondentene som svarte i ”liten eller ingen grad” ønsker kursing. Det å ønske kursing er også signifikant assosiert med et ønske om å ta i bruk flere formative vurderingsmetoder, og uenighet i en rekke endrings skeptiske utsagn knyttet til nye pedagogiske redskaper (se tabell 7, s. 56). Dette gjelder utsagnet om at nye metoder er forbigående trender, at forelesning er den beste måten å undervise på og at studenter vil være motvillig til nye metoder. Et ønske om kursing er også assosiert med enighet i utsagnet om at vurdering bør harmonere med læringsutbyttebeskrivelser og at flere karakterer er bedre. Mer forskning trengs for å kunne si noe med større grad av sikkerhet, men en mulig tolkning av resultatene over samlet sett, er at det foreligger et ønske om å bevege seg mot ideell praksis hos de fleste i det man ønsker kurs, og at ideell praksis for mange betyr mer formativ vurdering og helhetlig tankegang.

Selv om et ønske om ulike typer kursing foreligger, er ikke dette det samme som å delta på kurs. Som rapportert over var nesten samtlige kommentarer under ”Annet” at man synes at kurs var bra, men at man ikke kjenner man har tid til å delta. De fleste har også krysset av for svaralternativer knyttet til kursing som er mindre tidkrevende (figur 4, s.52). Det kan peke på at selv om emneansvarlige har et endringsrom, er ressursknapphet som barriere en effektiv stopper for en del i å følge opp et ønske om utvikling. Følgelig kan eventuelt svakt oppmøte på kurs ikke begrunnes med uvilje, men mer trolig med ulike systembarrierer. Biggs (2001) skriver om ”Quality enhancement” (QE) som et tillegg til QA (kvalitetsikringsarbeid) og påpeker at ”There is an enormous body of scholarship of teaching that individual teachers and administrators cannot be expected to know and apply” (s. 236). Gynning (2014) konkluderer også med at i kvalitetssikring og endringsarbeid i undervisning og vurdering er tilgang til riktig kompetanse i ulike faser av arbeidet vesentlig. I tillegg til teknisk og administrativ support som nevnt i avsnitt om barrierer over, kan det være verdt å utforske muligheten av å imøtekomme ønsket uttrykt av en av våre respondenter om en faglig ansatt med ”drop-in” tider man kan benytte seg av når man trenger veiledning.

4.2.6 Videre forskning. At undersøkelsesenheten var emneansvarlige på UiB, og målet var å generalisere til populasjonen emneansvarlige på UiB, betyr ikke at resultatene er uten interessere utover UiB. Det kan også tenkes at visse resultat er overførbare, og forslag til videre forskning presentert her, er aktuelt også for andre universitet i Norge.

Hvorvidt et perspektivskifte fra utdanning til læring er reflektert også i praksis må det mer forskning til for å svare på. Selv om et av hovedfunnene i denne undersøkelsen antyder at de fleste emneansvarlige på UiB tar i bruk nye pedagogiske redskaper, sier dette likevel lite eller ingenting om kvaliteten eller innholdet av disse. Mer forskning trengs for å undersøke hvorvidt læringsutbyttebeskrivelser er i tråd med NKR, og kanskje viktigere, om nødvendige justeringer er blitt gjort for å sikre epistemologisk gyldighet, hvorvidt læringsutbyttebeskrivelsene faktisk uttrykker det man ønsker studenter skal lære, og hvorvidt de brukes som et pedagogisk redskap i en helhetstankegang. Videre bør det forskes mer på hvorvidt vurderingsformens innhold avspeiler ønsket læringsutbytte og hvorvidt spesifikke vurderingskriterier er definert som kan angi grad av måloppnåelse. I den grad emneansvarlige tar i bruk både en summativ og formativ vurderingsform i samme emne, må mer forskning til for å si hvorvidt den summative vurderingsformen eventuelt overskygger ønsket effekt av den formative. Mer forskning trengs også for å si noe om hva emneansvarlige forstår som vurderingskriterier og om disse i tilstrekkelig grad blir gjort kjent for studenten. Når det gjelder tilbakemeldinger på skriftlig arbeid eller øvelser, trengs det mer forskning for å si noe om kvaliteten på disse.

Resultat fra denne undersøkelsen antyder videre at emneansvarlige flest ved UiB opplever barrierer. Videre forskning kan si mer om hvorvidt disse barrierene er effektive stoppere for endring, eller hvorvidt de vanskeliggjøre varig endring med tanke på for eksempel ressursknapphet. Det er store krav til emneansvarlige og fagmiljø med tanke på nøyaktig og epistemologisk gyldig formulering av læringsutbyttebeskrivelser, bruk av varierende undervisnings- og vurderingsformer, samt å sikre en helhet. Med tanke på dette er et interessant forskningsspørsmål hvilke type støtte emneansvarlige og andre vitenskapelig ansatte selv uttrykker er nødvendig i prosessen med å endre eller ta i bruk nye pedagogiske redskaper.

Kvalitetssikringssystemet er ment å være utdanningsinstitusjonens redskap for å skaffe seg nødvendig kunnskap for å kunne vurdere kvaliteten i egne utdanningstilbud

(NOKUT, 2014), men mekanismer for å vurdere kvalitet i utdanning kan stå i konflikt med en kvalitetskultur (UiB, 2012). Et forskningsspørsmål er hvilke kunnskap kvalitetssikringssystemet på UiB faktisk genererer per i dag, samt hvem og hvordan man skal vurdere kvaliteten på endringer foretatt. Herunder forholdet mellom kvalitetssystemets tiltak og læringsutbytte spesifikt, og hvorvidt systemet slik det er designet per i dag kan støtte og lede til ønsket praksis (Gynnild, 2014, 2017).

4.3 Oppsummering

Høyere utdanning er i endringstid, også UiB. Nasjonale reformer og interne UiB dokument legger tydelige føringer på et helhetlig innholdet i utdanning, og definerer på lang vei hva kvalitet i utdanning er. Forskningsspørsmålene i denne undersøkelsen ble stilt for å få oversikt og innsikt i deler av emneansvarlige sin vurderingspraksis relatert til bruk av nye pedagogiske redskaper som man er oppmuntret til eller pålagt å bruke. Med tanke på forskning som antyder at vurdering influerer både hva og hvordan studenter lærer, blir valg av vurderingsform viktig. Det foreligger også resultat fra forskning og argumenter basert på læringsteori som tyder på at en bevisst helhetlig tilnærming til alle deler av undervisnings- og vurderingsprosessen kan føre til bruk av dypere læringsstrategier. Det gjelder ikke kun valg av vurderingsform, men også bruk av tydelige læringsutbyttebeskrivelser som egner seg som pedagogisk redskap, formulering av vurderingskriterier og å gjøre disse kjent for studenten.

Det første forskningsspørsmålet var i hvilken grad emneansvarlige på UiB er i berøring med eller tar i bruk ulike vurderingsformer, læringsutbyttebeskrivelser og vurderingskriterier. Resultatene peker i retning av at emneansvarlige på UiB flest er i berøring med både summative og formative vurderingsformer, men tallene sier lite eller ingenting om innholdet i vurderingsformene hver for seg, eller samlet, ivaretar hensikten om studenters dype læring. Videre antyder resultatene at læringsutbyttebeskrivelser blir brukt utover pålagt krav om å formulere dem, men også at bruken for de fleste ikke reflekterer en helhetlig tankegang.

Det andre forskningsspørsmålet omhandlet hvilke og i hvilken grad emneansvarlige opplever barrierer i implementering av nye vurderingsformer. Basert på resultatene er det sannsynlig å anta at emneansvarlige på UiB flest opplever barrierer og at barrierene er en effektiv stopper for noen i implementering av nye vurderingsformer. Barrierene som omhandlet ressursknapphet som tid og kostnader toppet listen, og disse

kan muligens vanskeliggjøre det å prioritere knappe ressurser til kursing, selv om man anerkjenner eget behov og ønsker å delta. Mangel på helhetstankegang i de mange IT-systemene ansatte på UiB skal bruke, er også en barriere som effektivt kan hindre ønsket endring.

Et tredje forskningsspørsmål ble stilt for å belyse eventuelle sammenhenger mellom bruk av pedagogiske redskaper, barrierer, demografiske variabler og grad av enighet i utsagnsvariabler. Det fjerde i hvilken grad samme variabler være med forklare variansen i bruk av formative og summative vurderingsmetoder. Resultatene antyder for det første at tilsagn til endrings skeptiske og endringspositive utsagn samvarierer med hvorvidt og i hvor stor grad man bruker pedagogiske redskaper. Enighet i utsagnet om at man ikke så grunn for å endre praksis da det ikke var verdsatt, viste seg assosiert med mindre bruk av formative vurderingsmetoder, læringsutbyttebeskrivelser og det å ikke definerer vurderingskriterier. I regresjonsmodellen er lavere grad av enighet i endrings skeptiske utsagn også med å forklare variansen i berøring med formative vurderingsformer. Av de relativt få som er endrings skeptiske, er flest menn, og høyere alder er assosiert med enighet i at nye metoder er forbigående trender. Lengre undervisningserfaring var forklarende for bruk av flere både formative og summative vurderingsmetoder i regresjonsanalysene, og det å være kvinne var med å forklare berøring med flere formative vurderingsmetoder. At regresjonsmodellene totalt forklarte henholdsvis 19 % (formative) og 8 % (summative) av variansen kan peke på kompleksiteten i både vurderingspraksisen og mennesket bak. Det å forstå emneansvarliges vurderingspraksis kan ikke gjøres gjennom en statistisk analyser alene. Andre ukjente variabler, også variabler som kanskje ikke så enkelt lar seg måle i tall, kan muligens gi bedre forklaringsmodeller.

4.4 Konklusjon

De fleste som underviser har nok tanker om hva det er viktig at studenter skal lære. Det store spørsmålet som ikke kan besvares ved hjelp av tall og statistiske analyser alene, er hvorvidt undervisnings- og vurderingsformer støtter opp om studenters læring av ønsket læringsutbytte. Med andre ord hvorvidt et perspektivskifte fra utdanning til læring også er reflektert i praksis. Listen over nye forskningsspørsmål er lengre enn de spørsmålene undersøkelsen har vært med på å belyse, men det er et viktig funn i seg selv at spørsmål om kvalitet og innhold ikke kan besvares med tall alene.

Samlet sett peker resultatene i retning av at de fleste emneansvarlige er endringspositive og har endringsrom. Dette viser seg også i bruk av nye pedagogiske redskaper, og et ønske om å lære mer gjennom kursing. Dette er et godt utgangspunkt for en videreutvikling av en helhetlig undervisnings- og vurderingspraksis på UiB, og for å samlet arbeide for at et perspektivskiftet fra utdanning til læring skal gjennomsyre hele utdanningssystemet. Når systembarrierer topper listen av barrierer, og lite støtte fra fagmiljø også oppleves av noen, kan dette bety at i den grad høye ambisjoner skal innfris, handler det om faktorer andre enn den emneansvarlige må finne løsninger på. Administrativ, teknisk og faglig støtte til emneansvarlige i denne endringstiden kan være avgjørende, og det er trolig at både mer pedagogisk kunnskap og ressurser i form av tid og midler må tilføres om UIB som en lærende organisasjon skal lære mer om hva som kan produsere økt læringsutbytte, og hvordan man kan leve ut dette i praksis.

Litteraturliste:

- Albers, C. (2009). Teaching: From Disappointment to Ecstasy. *Teaching Sociology*, 37(3), 269-282. doi:10.1177/0092055X0903700305
- Babbie, E. R. (2010). *The practice of social research* (12th ed. ed.). Belmont, Calif: Wadsworth Cengage Learning.
- Baird, J.-A., Hopfenbeck, T. N., Newton, P., Gordon, S., & Steen-Utheim, A. T. (2014). *Assessment and learning : state of the field review*. Oxford University Centre for Educational Assessment Report, Vol. 14/2. Hentet fra <http://www.forskningsradet.no/servlet/Satellite?c=Rapport&cid=1253996755700&lang=en&pagename=kunnskapscenter%2FHovedsidemal>
- Barr, R. B., & Tagg, J. (1995). From Teaching to Learning: A New Paradigm for Undergraduate Education. *Change*, 27(6), 12-25.
- Befring, E. (1998). *Forskningsmetode og statistikk* (3. utg. ed.). Oslo: Samlaget.
- Biggs, J. (1996). Enhancing Teaching through Constructive Alignment. *Higher Education*, 32(3), 347-364.
- Biggs, J. (2001). The reflective institution: Assuring and enhancing the quality of teaching and learning. *The International Journal of Higher Education and Educational Planning*, 41(3), 221-238. doi:10.1023/A:1004181331049
- Black, P. (2015). Formative assessment – an optimistic but incomplete vision. *Assessment in Education: Principles, Policy & Practice*, 22(1), 161-177. doi:10.1080/0969594X.2014.999643
- Black, P., & Wiliam, D. (1998). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappan*, 80(2), 139-144.
- Brown, S. (2004-05). Assessment for learning. *Learning and Teaching in Higher Education*(1), 81-89.
- Clark-Carter, D. (2010). *Quantitative psychological research : the complete student's companion* (3rd ed. ed.). Hove: Psychology Press.
- Creswell, J. W. (2014). *Research design : qualitative, quantitative, and mixed methods approaches* (4th ed. ed.). Los Angeles, Calif: SAGE.
- Dale, E. L. (2005). *Kunnskapsregimer i pedagogikk og utdanningsvitenskap*. Oslo: Abstrakt forl.
- Danielsen, A. G. (2013). Kunnskapsbygging i skolen via kvantitative verktøy : statistikk og spørreskjema (pp. 138-154). Oslo: Universitetsforl., cop. 2013.
- de Bruijn-Smolanders, M., Timmers, C. F., Gawke, J. C. L., Schoonman, W., & Born, M. P. (2016). Effective Self-Regulatory Processes in Higher Education: Research Findings and Future Directions. A Systematic Review. *Studies in Higher Education*, 41(1), 139-158. doi:10.1080/03075079.2014.915302
- Denscombe, M. (2003). *The good research guide : for small-scale social research projects* (2nd ed. ed.). Maidenhead: Open University Press.
- Dochy, F., Segers, M., Gijbels, D., & Struyven, K. (2007). Assessment engineering: breaking down barriers between teaching and learning, and assessment. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education : learning for the longer term* (pp. 87-101). London: Routledge.
- Dysthe, O., Raaheim, A., Lima, I., & Bygstad, A. (2006). Evalueringen av Kvalitetsreformen. Delrapport 7. Undervisnings- og vurderingsformer: Pedagogiske konsekvenser av Kvalitetsreformen.
- Evans, C. (2013). Making Sense of Assessment Feedback in Higher Education. *Review of Educational Research*, 83(1), 70-120. doi:10.3102/0034654312474350
- Fekjær, S. B. (2016). *Statistikk i praksis*. Oslo: Gyldendal akademisk.

- Field, A. (2009). *Discovering statistics using SPSS : (and sex and drugs and rock 'n' roll)* (3rd ed. ed.). Los Angeles: SAGE.
- Fowler, F. J. (2009). *Survey research methods* (4th ed. ed. Vol. 1). Los Angeles: Sage.
- Gibbs, G. (2010). *Using assessment to support student learning*: Leeds Met Press.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education*(1), 3-31.
- Gynnild, V. (2011). Kvalifikasjonsrammeverket: Begreper, modeller og teoriarbeid. *Uniped*, 34(02), 18-32.
- Gynnild, V. (2013). «Kriteriebasert vurdering» - hva innebærer det i praksis? *Uniped*, 36(01), 26-40.
- Gynnild, V. (2014). Kvalitetssystemet i praksis: Ressurser på avveie? *Uniped*, 37(03), 4-22.
- Gynnild, V. (2017). Kvalitetssystemer til stryk: En empirisk, konstruktivstudie. *Uniped*, 40(01), 38-53.
- Hanssen, B., Husebø, D., & Moen, V. (2017). Universitetslærerarbeidet – utfordring av den naturlige innstilling? *Uniped*, 40(01), 7-17.
- Haraldsen, G. (1999). *Spørreskjemametodikk : etter kokebokmetoden*. Oslo: Ad Notam Gyldendal.
- Harlen, W. (2005). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16(2), 207-223.
doi:10.1080/09585170500136093
- Harlen, W., & Deakin, C. R. (2002). *A systematic review of the impact of summative assessment and tests on students' motivation for learning*. (EPPI-Centre Review, version 1.1*). Research Evidence in Education Library. Hentet fra <http://www.forskningsradet.no/servlet/Satellite?c=Rapport&cid=1250822859733&lang=en&pagename=kunnskapscenter%2FHovedsidemal>
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81-112. doi:10.3102/003465430298487
- Hopfenbeck, T. N. (2011). Vurdering og selvregulert læring. *Bedre skole*(4), 26-30.
- Hopfenbeck, T. N., Flórez Petour, M. T., & Tolo, A. (2015). Balancing Tensions in Educational Policy Reforms: Large-Scale Implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy & Practice*, 22(1), 44-60. doi:10.1080/0969594X.2014.996524
- Hrepic, Z., Zollman, D. A., & Rebello, N. S. (2007). Comparing Students' and Experts' Understanding of the Content of a Lecture. *Journal of Science Education and Technology*, 16(3), 213-224.
- Haakstad, J. (2011). Læringsutbytte: Begrepets anvendelighet i kvalitetsvurdering av høyere utdanning. *Uniped*, 4(04), 72-82.
- Ilstad, S. (1989). *Survey-metoden : en veiledning i utvalgsundersøkelser* (Ny utg. ed.). Trondheim: Tapir.
- Karseth, B., Lycke, K. H., & Solbrekke, T. D. (2008). Høyere utdanning i endring – pedagogiske og kunnskapsmessige utfordringer. *Norsk pedagogisk tidsskrift*, 92(06), 413-416.
- Kirke- utdannings- og forskningsdepartementet [KUF]. (2001). *Gjør din plikt - Krev din rett*. (St. meld. nr. 27 (2000-2001)). Hentet fra <https://www.regjeringen.no/no/dokumenter/stmeld-nr-27-2000-2001-/id194247/>
- Knight, P. (2007). Grading, classifying and future learning. In D. Boud & N. Falchikov (Eds.), *Rethinking assessment in higher education : learning for the longer term* (Vol. Routledge, pp. 72-86). London.

- Knight, P., & Trowler, P. (2000). Department-level Cultures and the Improvement of Learning and Teaching. *Studies in Higher Education*, 25(1), 69-83.
doi:10.1080/030750700116028
- Kunnskapsdepartementet [KD]. (2011). Nasjonalt kvalifikasjonsrammeverk for livslang læring (NKR). Hentet fra
http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Norsk_utdanning/NKR/250414_Nasjonalt_kvalifikasjonsrammeverk_for_livslang_l%C3%A6ring_NKR.pdf
- Kunnskapsdepartementet [KD]. (2016). *Tilstandsrapport for høyere utdanning 2016*. (Rapport). Hentet fra
<https://www.regjeringen.no/no/dokumenter/tilstandsrapport-for-hoyere-utdanning-2016/id2498657/>
- Kunnskapsdepartementet [KD]. (2017). *Kultur for kvalitet i høyere utdanning*. (St.meld. nr. 16, 2016-2017). Hentet fra
<https://www.regjeringen.no/no/dokumenter/meld.-st.-16-20162017/id2536007/>
- Kårhus, S. (2011). Utdanningsmarkedet og rekontekstualisering av utdanningskvaliteter. *Uniped*, 34(02), 43-54.
- Lund, T., & Christophersen, K.-A. (1999). *Innføring i statistikk*. Oslo: Universitetsforlaget.
- Maki, P. L. (2012). *Assessing for Learning: Building a Sustainable Commitment Across the Institution* (2nd ed ed.). Sterling: Sterling : Stylus Publishing.
- Michelsen, S., & Aamodt, P. O. (2007). Evaluering av Kvalitetsreformen. Sluttrapport.
- Mårtensson, K., Roxå, T., & Stensaker, B. (2014). From quality assurance to quality practices: an investigation of strong microcultures in teaching and learning. *Studies in Higher Education*, 39(4), 534.
- Nasjonalt organ for kvalitet i utdanning [NOKUT]. (2014). *Evaluering av system for kvalitetssikring av utdanningen ved Universitetet i Bergen*. Hentet fra
http://www.nokut.no/Documents/NOKUT/Artikkelbibliotek/Kunnskapsbasen/Rapporter/Kvalitetssikringsrapporter/2014/UiB_system_kvalitetssikring_2014.pdf
- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative Assessment and Self-Regulated Learning: A Model and Seven Principles of Good Feedback Practice. *Studies in Higher Education*, 31(2), 199-218.
- Nordkvelle, Y. T. (2013). Forskningsmelding med budskap om undervisning. *Uniped*, 36(01), 1-3.
- Norton, L., Norton, B., & Shannon, L. (2013). Revitalising assessment design: what is holding new lecturers back? [journal article]. *Higher Education*, 66(2), 233-251.
doi:10.1007/s10734-012-9601-9
- Oppenheim, A. N. (1992). *Questionnaire design, interviewing, and attitude measurement* (New ed. ed.). London: Pinter.
- Pallant, J. (2010). *SPSS survival manual : a step by step guide to data analysis using SPSS* (4th ed. ed.). Maidenhead: McGraw-Hill Open University Press.
- Ramsden, P. (1992). *Learning to teach in higher education*. London: Routledge.
- Ringdal, K. (2013). *Enhet og mangfold : samfunnsvitenskapelig forskning og kvantitativ metode* (3. utg. ed.). Bergen: Fagbokforl.
- Roos, M., & Trippestad, T. A. (2015). Reformer eller deformer? *Norsk pedagogisk tidsskrift*, 99(06), 397-402.
- Raaheim, A. (2016). *Eksamensrevolusjonen : råd og tips om eksamen og alternative vurderingsformer*. Oslo: Gyldendal akademisk.
- Silverman, D. (2001). *Interpreting qualitative data : methods for analysing talk, text and interaction* (2nd ed. ed.). London: Sage.

- Skodvin, A., Handal, G., Lycke, K. H., & Solbrekke, T. D. (2012). Føyelige eller ubøvelige? Universitetspedagogers rolle i implementering av nasjonalt kvalifikasjonsrammeverk. *Uniped*, 35(02), 18-31.
- Slemmen, T. (2010). *Vurdering for læring i klasserommet* (2. utg. ed.). Oslo: Gyldendal akademisk.
- Smith, K. (2007a). Om eksamensforskning – intervju med Steinar Kvale. *Norsk pedagogisk tidsskrift*, 91(02), 94-99.
- Smith, K. (2007b). Vurdering som et motivasjonsfremmende redskap for læring. *Norsk pedagogisk tidsskrift*, 91(02), 100-106.
- Stensaker, B. (2008). Endringsarbeid i høyere utdanning: nye konfliktlinjer og nye muligheter. *Norsk pedagogisk tidsskrift*, 92(06), 417-427.
- Studieadministrativ avdeling UIB. (2012). *Vurderingsformer, kvalifikasjonsrammeverk og læringsutbyttebeskrivelser*. (UU-sak 63-12-2012). Hentet fra https://kvalitetsbasen.app.uib.no/rapport.php?rapport_id=3965
- Svartdal, F. (2009). *Psykologiens forskningsmetoder* (3. utg. ed.). Bergen: Fagbokforl.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using Multivariate Statistics: Pearson New International Edition* (International edition of sixth edition. ed.): United Kingdom: Pearson Education M.U.A.
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology : combining qualitative and quantitative approaches* (Vol. v. 46). Thousand Oaks, Calif: Sage.
- Thingnes, E. R., Stalsberg, R., & Sitter, B. (2015). Er forelesninga effektiv, interessant og meningsfull? - Oppfatninger om og betydningen av forelesninger som undervisningsform. *Uniped*(04), 390-397.
- Universitetet i Bergen [UiB]. (2012). Kvalitet i Utdanning. Hentet fra http://www.uib.no/filearchive/kvalitet_i_uttanning_.pdf
- Universitetet i Bergen [UiB]. (2013). Handbok for kvalitetssikring av universitetsstudia. Hentet fra <http://www.uib.no/sa/49342/kvalitetshandboka>
- Universitetet i Bergen [UiB]. (2017a). "Fra prat til praksis" - om læringsutbyttebeskrivelser som ressurs i undervisningen og i Mitt UiB. Hentet fra <http://www.uib.no/sa/105104/fra-prat-til-praksis-om-l%C3%A6ringsutbyttebeskrivelser-som-ressurs-i-undervisningen-og-i-mitt>
- Universitetet i Bergen [UiB]. (2017b). UiB ble hørt i regjeringens stortingsmelding om høyere utdanning.
- Universitets- og høgskolerådet [UHR]. (2004). Karaktersystemet - generelle, kvalitative beskrivelser. Hentet fra http://www.uhr.no/documents/karaktersystemet_generelle_kvalitative_beskrivelser.pdf
- Utdannings- og forskningsdepartementet. (2004). *Retningslinjer for bruk av det nasjonale karaktersystemet*. Hentet fra [https://www.regjeringen.no/no/dokumenter/retningslinjer-for-bruk-at-det-nasjonale/id91189/?q=Retningslinjer for bruk av det nasjonale karaktersystemet](https://www.regjeringen.no/no/dokumenter/retningslinjer-for-bruk-at-det-nasjonale/id91189/?q=Retningslinjer+for+bruk+av+det+nasjonale+karaktersystemet).
- Wang, X., & Hurley, S. (2012). Assessment as a Scholarly Activity?: Faculty Perceptions of and Willingness to Engage in Student Learning Assessment. *Journal of General Education*, 61(1), 1-15. doi:10.1353/jge.2012.0005
- Wang, X., Su, Y., Cheung, S., Wong, E., & Kwong, T. (2013). An exploration of Biggs' constructive alignment in course design and its impact on students' learning approaches. *Assessment & Evaluation in Higher Education*, 38(4), 477-491. doi:10.1080/02602938.2012.658018
- Watkins, D., Dahlin, B., & Ekholm, M. (2005). Awareness of the backwash effect of assessment: A phenomenographic study of the views of Hong Kong and

- Swedish lecturers. *An International Journal of Learning and Cognition*, 33(4), 283-309. doi:10.1007/s11251-005-3002-8
- Willig, C. (2008). *Introducing qualitative research in psychology : adventures in theory and method* (2nd ed. ed.). Maidenhead: Open University Press.
- Wittek, L. (2008). Pedagogiske redskaper og studenters læring. *Norsk pedagogisk tidsskrift*, 92(06), 443-455.
- Aamodt, P. O., Hovdhaugen, E., & Prøitz, T. S. (2014). Utdanningskvalitet i høyere utdanning: noen empiriske eksempler. Resultater fra en undersøkelse blant faglig ansatte våren 2013.
- Aamodt, P. O., Prøitz, T. S., Hovdhaugen, E., & Stensaker, B. (2007). *Læringsutbytte i høyere utdanning*. (Rapport 40/2007). Hentet fra http://www.hivolda.no/neted/upload/attachment/site/group36/NIFU_STEP_Rapport_40_2007%5B1%5D.pdf

Vedlegg 1 – Godkjenning fra NSD

Robert Gray Jr
Institutt for pedagogikk Universitetet i Bergen
Christies gate 13
5020 BERGEN

Vår dato: 23.01.2017

Vår ref: 51624 / 3 / BGH

Deres dato:

Deres ref:

TILBAKEMELDING PÅ MELDING OM BEHANDLING AV PERSONOPPLYSNINGER

Vi viser til melding om behandling av personopplysninger, mottatt 16.12.2016. Meldingen gjelder prosjektet:

<i>51624</i>	<i>Teaching and Learning in the Digital Age: Online Tools and Assessment Practices</i>
<i>Behandlingsansvarlig</i>	<i>Universitetet i Bergen, ved institusjonens øverste leder</i>
<i>Daglig ansvarlig</i>	<i>Robert Gray Jr</i>
<i>Student</i>	<i>Maria Grønner</i>

Personvernombudet har vurdert prosjektet og finner at behandlingen av personopplysninger er meldepliktig i henhold til personopplysningsloven § 31. Behandlingen tilfredsstillers kravene i personopplysningsloven.

Personvernombudets vurdering forutsetter at prosjektet gjennomføres i tråd med opplysningene gitt i meldeskjemaet, korrespondanse med ombudet, ombudets kommentarer samt personopplysningsloven og helseregisterloven med forskrifter. Behandlingen av personopplysninger kan settes i gang.

Det gjøres oppmerksom på at det skal gis ny melding dersom behandlingen endres i forhold til de opplysninger som ligger til grunn for personvernombudets vurdering. Endringsmeldinger gis via et eget skjema, <http://www.nsd.uib.no/personvern/meldeplikt/skjema.html>. Det skal også gis melding etter tre år dersom prosjektet fortsatt pågår. Meldinger skal skje skriftlig til ombudet.

Personvernombudet har lagt ut opplysninger om prosjektet i en offentlig database, <http://pvo.nsd.no/prosjekt>.

Personvernombudet vil ved prosjektets avslutning, 19.06.2020, rette en henvendelse angående status for behandlingen av personopplysninger.

Vennlig hilsen

Kjersti Haugstvedt

Belinda Gloppen Helle

Kontaktperson: Belinda Gloppen Helle tlf: 55 58 28 74

Dokumentet er elektronisk produsert og godkjent ved NSDs rutiner for elektronisk godkjenning.

Vedlegg: Prosjektvurdering
Kopi: Maria Grønner mgr045@student.uib.no

Personvernombudet for forskning



Prosjektvurdering - Kommentar

Prosjektnr: 51624

INFORMATION AND CONSENT

The sample will receive written information about the project, and give their consent to participate. The letter of information is well formulated. However the participants that will answer the online survey have to get information equivalent to this in order for their consent to be valid.

DATA PROCESSOR AGREEMENT

According to the notification from personal information will be gathered by using an online questionnaire. The Data Protection Official presumes that you will be using a data processor. Therefore a data processor agreement should be made. For advice on what the data processor agreement should contain, please see:
<http://www.datatilsynet.no/English/Publications/Data-processor-agreements/>.

INFORMATION SECURITY

The Data Protection Official presupposes that the researcher follows internal routines of Universitetet i Bergen regarding data security. If personal data is to be sent by email, the information should be adequately encrypted.

PUBLICATION

It is stated that personally identifiable information will be published. The Data Protection Official presupposes that the participants give their explicit consent to this. Further, we recommend that participants are given the opportunity to read through their own information and give their approval before publication.

THE END OF THE PROJECT

Estimated end date of the project is 19.06.2020. According to the notification form all collected data will be made anonymous by this date.

Making the data anonymous entails processing it in such a way that no individuals can be recognised. This is done by:

- deleting all direct personal data (such as names/lists of reference numbers)
- deleting/rewriting indirectly identifiable data (i.e. an identifying combination of background variables, such as residence/work place, age and gender)
- deleting digital audio and video files

Vedlegg 2 – Invitasjon til spørreundersøkelsen



UNIVERSITETET I BERGEN

Spørreundersøkelse "Vurderingspraksis på UiB"

Klikk på linken for å delta: [http://www.survey-xact.no/answer?](http://www.survey-xact.no/answer?key=2NV2JSGQWLCX)

[key=2NV2JSGQWLCX](http://www.survey-xact.no/answer?key=2NV2JSGQWLCX) Gjennomføring av spørreundersøkelsen blir ansett som samtykke til å delta. Du kan endre språk til norsk på spørreundersøkelsens første side.

Informasjon om spørreundersøkelsen

Målet med denne frivillige spørreundersøkelsen er å få innsikt i hvordan undervisere ved UiB vurderer studenters læring i dag, hvordan du ønsker å gjøre vurdering og hvorvidt tilgjengelige verktøy til vurderingsformål blir tatt i bruk. Studien gjennomføres av dr. Robert Gray og masterstudent Maria Grönen ved Institutt for Pedagogikk. Undersøkelsen er anonym og samtlige undervisere ved UiB er forespurt om å delta i undersøkelsen.

Alle personopplysninger vil bli behandlet konfidensielt. E-post adresser / koblingsnøkkel er lagret adskilt fra øvrige data og er anonymisert gjennom SurveyXact. Deltakerne vil ikke kunne gjenkjennes i publikasjon. Prosjektet skal etter planen avsluttes 19.06.2020.

Spørsmål? Ta gjerne kontakt med dr. Robert Gray, robert.gray@uib.no eller masterstudent Maria Grönen, tlf. 92224481, mgr045@student.uib.no

Studien er meldt til Personvernombudet for forskning, NSD - Norsk senter for forskningsdata AS.

Survey "Assessment Practices at UiB"

Link to participate: [http://www.survey-xact.no/answer?](http://www.survey-xact.no/answer?key=2NV2JSGQWLCX)

[key=2NV2JSGQWLCX](http://www.survey-xact.no/answer?key=2NV2JSGQWLCX) Filling out the survey will be regarded as an active consent. You can take the survey in Norwegian or English by setting the language on the first page in the survey.

Information about the survey

This voluntary survey is intended to discover how teaching staff at UiB currently assess student learning, how you would like to do assessment, and if tools available for assessment purposes are being used. Dr. Robert Gray and master student Maria Grönen (Institutt for Pedagogikk) will be conducting this survey. The survey is anonymous and all teaching staff at UiB has been asked to participate.

Your e-mail address and respondent key are stored separate from all other data. The information you provide will remain confidential and will be de-identified for all analyses and publication. Planned end date for this research project is June 19, 2020.

Any questions? Feel free to contact dr. Robert Gray, robert.gray@uib.no or master student Maria Grönen, tlf. 92224481, mgr045@student.uib.no

NSD, Norwegian centre for research data, is notified of this study.