Gro Tellsgård

Identitet og karakter

En kvantitativ studie av sammenhengen mellom matematisk identitet og prestasjoner i matematikk på videregående skole

Masteroppgave i matematikkdidaktikk
Veileder: Eivind Kaspersen
Mai 2020

Masteroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for lærerutdanning

NTNU
Norwegian University of
Science and Technology

Gro Tellsgård

Identitet og karakter

En kvantitativ studie av sammenhengen mellom matematisk identitet og prestasjoner i matematikk på videregående skole

Masteroppgave i matematikkdidaktikk
Veileder: Eivind Kaspersen
Mai 2020

Norges teknisk-naturvitenskapelige universitet
Fakultet for samfunns- og utdanningsvitenskap
Institutt for lærerutdanning

NTNU
Norwegian University of
Science and Technology

Sammendrag

Formålet med denne oppgaven har vært å se på sammenhengen mellom matematisk identitet (MI) og prestasjoner i matematikk på en videregående skole. Bakgrunn for oppgaven er at det er vist at mange sider ved matematikkrelatert affekt kan ha sammenheng med matematikkprestasjoner (Ma, 1999; OECD, 2015). Sammenhengen mellom elevers identitet som utøvere av matematikk (elevers MI) og deres prestasjoner i matematikk er imidlertid ikke undersøkt i like stor grad.

Studien er kvantitativ, og datagrunnlaget var elevers responser på et spørreskjema der de på 20 utsagn skulle svare i hvilken grad de var enige. Utsagnene var laget med bakgrunn i karakteristikker av personer som tenker og jobber som matematikere, hentet fra teori om dybdelæring i matematikk og positiv affekt for matematikk (Burton, 1998; Hiebert, 1986; Skemp, 1976). Svarene på spørreskjemaet ble analysert med Rasch-modellen som gir mål på intervallnivå, og derfor kan brukes videre i statistiske analyser. For å undersøke sammenhengen mellom MI og henholdsvis matematikkarakterer gitt av faglærer og resultater på kartleggingsprøven *Kartleggeren* (Fagbokforlaget.no, 2020), ble det utført korrelasjonsanalyser (Pearsons *r*). Til slutt ble det også undersøkt om de to korrelasjonene som ble funnet var signifikant ulike, med bakgrunn i teori om at lærere som kjenner elevene de vurderer kan legge andre ting til grunn for vurderingen enn oppnådd kompetanse, som for eksempel progresjon eller holdninger (Prøitz & Borgen, 2010). Kartleggeren er en digital prøve som rettes automatisk og anonymt, og vil ikke kunne legge vekt på annet enn det elevene presterer faglig.

Instrumentet (spørreskjemaet) som ble brukt i undersøkelsen var ikke tidligere validert for elever i videregående skole. En betydelig del av oppgaven handler derfor om å sikre at utsagn, og responser på disse, passer i Rasch-modellen etter Thurstones prinsipper for måling (additivitet, endimensjonalitet og invarians) (Andrich, 1989). Valideringen er gjort etter et rammeverk av Wolf og Smith (2007). Spørreskjemaet ble utviklet av Kaspersen (2018) og utsagnene valgt ut på bakgrunn av at de la seg i samme dimensjon i Rasch-analysen. Dimensjonen av MI som blir målt omfatter positiv affekt for matematikk og dypt arbeid med faget, men det er ikke utelukket at MI også kan ha andre dimensjoner.

I studien fant jeg moderat korrelasjon mellom MI og karakterer (r = 0,32) og mellom MI og resultater på Kartleggeren (r = 0,33). Det ble ikke funnet signifikant forskjell mellom de to korrelasjonene (p = 0,89).

Funnene i studien underbygger at matematikkrelatert affekt generelt, og MI spesielt, henger sammen med prestasjoner i matematikk. Funnene støtter ikke tanker om at lærere legger affektive aspekter, som MI, til grunn når de vurderer elevene.

Abstract

The purpose of this study has been to examine the relationship between mathematical identity (MI) and achievements in mathematics at a senior high school. The study is based on the knowledge that many aspects of mathematics related affect may be related to achievements in mathematics (Ma, 1999; OECD, 2015). However, the relationship between the students' identity as mathematics performers (students' MI) and their achievements in mathematics has not been investigated to the same degree.

The study is quantitative. The basic data has been gathered from the students' replies to a questionnaire in which they were asked to define their degree of agreement with 20 given statements. These statements were designed based on characteristics of individuals with a "mathematical" character, drawn from theory on deep approach to mathematics and positive affect for mathematics (Burton, 1998; Hiebert, 1986; Skemp, 1976). The answers to the questionnaire were analysed using the Rasch model which gives results at an interval scale, and thus can be used in subsequent statistical analyses. Correlational analyses (Pearson's *r*) were used to investigate the relationship between MI and (1) teacher assessed achievement in mathematics, and (2) results of assessment tests using "Kartleggeren" (Fagbokforlaget.no, 2020). Finally, the two correlations which were obtained were examined for significant differences, based on the theory that teachers who know the students they are assessing may use other criteria in their assessments than attained ability, for example progress or attitude (Prøitz & Borgen, 2010). "Kartleggeren" is a digital test with anonymous, automatic assessment which can only use the students' academic ability as an assessment criterium.

The instrument (the questionnaire) used in this study has not previously been validated for students at a senior high school. Therefore, a considerable part of the study was devoted to checking that the statements, and the students' response to them, were suitable for the Rasch model, in accordance with Thurstone's principles of measurement (additivity, uni-dimensionality and invariance) (Andrich, 1989). The validation was performed using a framework defined by Wolf and Smith (2007). The questionnaire was developed by Kaspersen (2018) and the statements were chosen on the background of their position in the same dimension in the Rasch analysis. The dimension of MI which was measured encompassed positive affect for mathematics and a deep approach to the subject, but the possibility that MI can include other dimensions cannot be excluded.

My study showed a moderate correlation between MI and examination results (r = 0,32) and between MI and the results of the "Kartleggeren" test (r = 0,33). No significant difference between the two correlations was found (p = 0,89).

The results of this study confirm that mathematical affect in general, and specifically MI, are related to achievements in mathematics. The results do not support the idea that teachers base their assessment of students on affective aspects such as MI.

Forord

Etter to år med én fot i Tønsberg og én i Trondheim, er det med glede og vemod jeg nå avslutter mitt masterstudium i matematikkdidaktikk ved NTNU. Jeg vil benytte anledningen til å takke alle som har bidratt, både praktisk og emosjonelt, til at jeg har fått i havn dette dokumentet som markerer avslutningen av det hele.

Først vil jeg få takke min veileder Eivind Kaspersen som tidlig og sent har delt av sin kunnskap om matematisk identitet, Rasch-analyse og referanseføring. Uten din vennlige kritikk, ditt engasjement og din humor hadde jeg ikke klart å gjennomføre.

Takk til skolen som deltok i undersøkelsen. Takk til elevene som svarte på spørreskjemaet, og til lærerne som slapp meg inn i klasserommene sine og som bidro med informasjon til studien.

Takk til min arbeidsgiver som har gitt meg mulighet til videreutdanning gjennom «Kompetanse for kvalitet». Jeg håper jeg skal kunne gi tilbake i form av god og forskningsbasert undervisning for elevene.

Takk til Maria som har hatt den vanskelige jobben å språkvaske oppgaven, og takk til Donald som har oversatt sammendraget til engelsk.

Takk til studievenner som har bidratt med gode innspill og godt humør underveis i oppgaveskrivingen, og for hyggelig selskap på samlinger i Trondheim.

Takk til familie, venner og kollegaer som har heiet meg fram.

Sist men ikke minst takk til min egen lille familie. Takk til Ingrid og Frida som udelt har støttet mitt prosjekt, selv om det har medført eksamensstress i førjulstid, og mange uker de siste to åra uten at jeg har vært hjemme. Og takk Magnus. Du har vært unik. Uten din forståelse, tålmodighet og støtte hadde jeg ikke kunnet gjøre dette.

Gro Tellsgård

Tønsberg, mai 2020

Innhold

Figurer

Tabeller

Formler

Forkortelser/symboler

DIF	Differential item functioning
ICC	Item characteristic curve
Infit MNSQ	Infit mean square
IRT	Item response theory
MI	Matematisk identitet
Outfit MNSQ	Outfit mean square
PCA	Principal component analysis
PM Corr.	Point measure correlation

1 Innledning

Formålet med denne studien har vært å se på sammenhengen mellom matematisk identitet (MI) og prestasjoner i matematikk. Det er en økende anerkjennelse av at affektive faktorer spiller en avgjørende rolle i læring og undervisning av matematikk (Ma, 1999; McLeod, 1992). Affektive faktorer innebærer for eksempel følelser, holdninger, oppfatninger, motivasjon og identitet (Goldin et al., 2016; McLeod, 1992). McLeod (1992) beskriver hvor viktig de affektive faktorene er i matematikklasserommet på denne måten: «Når lærere snakker om matematikklassene sine, nevner de like gjerne elevenes entusiasme eller fiendtlighet til matematikk, som de nevner elevenes kognitive prestasjoner. På samme måte vil elever like gjerne gi affektive som kognitive responser» (min oversettelse) (McLeod, 1992).

Elevene har holdninger, følelser og oppfatninger om matematikkfaget, og de har en matematisk identitet. MI kan være en følelse av å høre til og bidra i et matematikkfellesskap (Solomon, 2007; Wenger, 2010), det kan være historiene du forteller om deg selv, eller som blir fortalt om deg, som utøver av matematikk (Sfard & Prusak, 2005), eller det kan være egne forestillinger om hvem du er i matematiske sammenhenger eller aktiviteter (Bishop, 2012). Felles for mange definisjoner av MI er at den uttrykker hvordan personer ser seg selv relativt til matematikk som sosial struktur eller aktivitet (f.eks; Deaux, 1993; Kaspersen, 2018).

Siden 50-tallet har det vært mye forskning på sammenhengen mellom affektive faktorer og elevers prestasjoner i matematikk. PISA-undersøkelsen (OECD, 2015) viste at det finnes en sammenheng mellom «Self-Efficacy» (mestringstro) og prestasjoner i matematikk. Mestringstro handler om elevenes tro på at de, gjennom sine handlinger, kan oppnå ønsket resultat (OECD, 2015). Rapporten viser at elever med høy grad av mestringstro i gjennomsnitt ligger faglig ett år foran elever med lav mestringstro. Studier om matematikkangst (Ma, 1999) viser at det også finnes en sammenheng mellom elevers grad av matematikkangst og deres prestasjoner i matematikk, der høy grad av angst henger sammen med lavere prestasjoner. Jeg har ikke funnet tidligere kvantitative undersøkelser av sammenhengen mellom MI og prestasjoner i matematikk, så denne undersøkelsen kan være et bidrag til å underbygge og rettferdiggjøre et syn på matematikkrelatert affekt som et viktig aspekt i læring og undervisning i matematikk.

Affekt har generelt blitt sett på som noe annet enn matematisk tenkning, heller enn som en del av det (Zan, Brown, Evans & Hannula, 2006), men flere teorier bygger på anerkjennelsen av sammenhengen mellom affekt og læring i matematikk. Kilpatrick, Swafford og Bradford (2001) beskriver fem komponenter som må være til stede for at elever skal kunne utvikle gode matematikkferdigheter. Disse komponentene er, som oversatt av Matematikksenteret (matematikksenteret.no, 27.1.20): forståelse, beregning, anvendelse, resonnering og engasjement. Den siste komponenten, «engasjement», blir forklart som det å være motivert for å lære matematikk, å se på matematikk som nyttig og verdifullt, og å tro at egen innsats bidrar til økt læring i matematikk (mestringstro). Tråden «forståelse» i Kilpatricks (2001) rammeverk viser til en konseptuell forståelse av matematikk. Konseptuell forståelse forklares som et integrert og funksjonelt grep om matematiske idéer, i motsetning til instrumentell innlæring av isolerte fakta og metoder (Kilpatrick et al., 2001; Skemp, 1976).

Konseptuell forståelse omtales som viktig fordi den bidrar til å knytte ny og gammel kunnskap sammen, og gjør at det er lettere å huske og velge riktige metoder. Konseptuell forståelse kan også omtales som dybdelæring, og er framtredende i fagfornyelsen av læreplanen (Utdanningsdirektoratet, 2020) som gradvis trer i kraft fra høsten 2020. Dybdelæring defineres i fagfornyelsen som «gradvis utvikling av kunnskap og varig forståelse av begreper, metoder og sammenhenger i fag og mellom fagområder» (Utdanningsdirektoratet, 2019a). Elever som er engasjerte, som har mestringstro, og som er motiverte for dybdelæring, kan vise dette ved å like å diskutere matematikk, søke etter sammenhenger og bruke ulike strategier i oppgaveløsning. Disse karakteristikkene finner vi igjen i instrumentet som ble brukt til å måle MI i denne undersøkelsen. Dybdelæring og positiv affekt er altså to sider av MI slik begrepet er definert her.

Kilpatricks komponenter er avhengige av hverandre (Matematikksenteret, 2020). Det kan på denne måten se ut til at Kilpatrick et al. (2001) mener at motivasjon og mestringstro er en forutsetning for å utvikle gode matematikkferdigheter samtidig som de sier at gode mestringsopplevelser fører til bedre engasjement og motivasjon. Tilsvarende gjensidige sammenhenger kan også gjelde for annen matematikkrelatert affekt, som mellom MI og prestasjoner i matematikk. Denne sammenhengen ønsker jeg å belyse i denne oppgaven.

De fleste studier på MI er kvalitative (Graven & Heyd-Metzuyanim, 2019), men Kaspersen (2015) har utviklet et instrument som måler MI. Måling av psykologiske faktorer, som MI, kalles psykometri og gjør det mulig å måle ulike sider ved personligheten vår (Wright & Stone, 1979). I psykologi kan psykometri for eksempel brukes til å måle grad av lykke (Lyubomirsky & Lepper, 1999) eller personlighet, som i den kjente Big Five-analysen (John & Srivastava, 1999). I utdanningsforskning kan det brukes til å måle kompetanse, som i nasjonale prøver (Utdanningsdirektoratet, 2018), personlighetstrekk som motivasjon og mestringstro (Fennema & Sherman, 1976), eller som vist i studier gjort av Kaspersen og Ytterhaug, til å måle matematisk identitet (Kaspersen, 2018; Kaspersen, Pepin & Sikko, 2017; Ytterhaug, 2019). Den psykometriske metoden som brukes i denne studien er Rasch-modellen (Linacre, 2006).

De nevnte studiene (Kaspersen, 2018; Kaspersen et al., 2017; Ytterhaug, 2019) tar for seg måling av MI hos lærerstudenter, ingeniørstudenter og ungdomsskoleelever. Jeg har brukt samme instrument og metode til å måle matematiske identitet hos elever på en videregående skole. Siden dette ikke har vært gjort før, blir undersøkelsen også en validering av instrumentet for elever på videregående skole, og derfor en undersøkelse av om MI vil la seg måle i denne gruppen.

Prestasjoner hos elever i ungdomsskolen og videregående skole blir målt jevnlig gjennom blant annet fagsamtaler, prøver og oppgaver laget og rettet av faglærere, og gjennom standardiserte prøver som kartleggingsprøver, nasjonale prøver, PISA-undersøkelsen og eksamen. Slik skolesystemet er lagt opp nå, vil elevenes karakterer ved endt skolegang representere kompetansen eleven har i hvert fag. Grunnlaget for vurdering er kompetansemålene i læreplanene (Lovdata, 2006), og standpunktkarakteren settes vanligvis av en faglærer som kjenner eleven.

Man kan undres om en faglærer legger mer i karakteren enn den faglige kompetansen som skal vurderes. Kan det hende at elevens måte å jobbe på, eller affekt til matematikkfaget vektlegges av læreren i vurderingen? De normerte prøvene vurderes vanligvis anonymt. Studier av sammenhengen mellom matematikkangst og prestasjoner i matematikk har vist at det er større korrelasjon (negativ) mellom elevens grad av angst og deres prestasjoner når det er en lærer som kjenner eleven som setter karakterer, enn det er når elevenes kompetanse blir vurdert ut fra anonyme prøver (Ma, 1999). Dette kan tyde på at lærere kan la seg påvirke av elevenes affekt for faget når de vurderer dem. Gjelder denne sammenhengen også når vi ser på sammenhengen mellom MI og karakterer som blir gitt av lærere og resultater på en anonym kartleggingsprøve?

1.1 Problemformulering

I denne kvantitative studien ønsket jeg å undersøke sammenhengene mellom elevers matematiske identitet og deres prestasjoner i matematikk. Jeg jobbet ut fra følgende tre forskningsspørsmål:

1. Er det mulig å måle MI hos elever på videregående skole?
2. Hva er korrelasjonen mellom elevenes mål på MI og deres matematikkprestasjoner?
3. Hvordan samsvarer korrelasjonen mellom MI og karakterer gitt av faglærer og korrelasjonen mellom MI og resultater på en anonym kartleggingsprøve?

Et positivt svar på det første forskningsspørsmålet var en forutsetning for å kunne svare på det andre og det tredje forskningsspørsmålet. Data fra undersøkelsen kan gi grunnlag for å se nærmere på årsaker og virkninger gjennom kvalitative studier på matematikkrelatert affekt, dybdelæring og MI. Instrumentet som brukes kan også standardiseres ved å validere for større grupper. Da kan det brukes som et verktøy i forskning på tvers av kontekster. Slik jeg har brukt instrumentet, er det kun gyldig for gruppen jeg selv undersøker.

1.2 Begrepsavklaringer

MI er i denne oppgaven definert som «den relative posisjonen mellom personer og den sosiale strukturen av å være matematisk i den aktiviteten man deltar i» (Kaspersen, 2018). MI er derfor definert både som et forhold mellom person og sosial struktur, og som hva den «sosiale strukturen å være matematisk» innebærer. I denne undersøkelsen er den sosiale strukturen operasjonalisert som et sett med karakteristikker og rekkefølgen disse danner i analysen. Karakteristikkene beskriver egenskaper som for eksempel positiv affekt for og dyp tilnærming til matematikk.

Dybdelæring eller dybdekunnskap er i oppgaven brukt som samlebegrep for dyp, konseptuell og relasjonell kunnskap definert av Skemp (1976), Hiebert (1986) eller Utdanningsdirektoratet (2019a). Se nærmere forklaring i kapittel 2.2.4.

Prestasjoner brukes i oppgaven som en samlebetegnelse på prestasjoner, ferdigheter og kunnskap. Hva som bør eller kan ligge til grunn for vurdering av prestasjoner, utdypes i kapittel 2.3.

1.3 Oppgavens oppbygning

Teorikapittelet vil inneholde teori om måling generelt, og psykometri og Rasch-modellen spesielt, et utvalg definisjoner av MI, teori om hva det vil si å være «matematisk», definisjon av MI for denne studien, og forskning på hva som ligger til grunn for læreres vurdering av elever i matematikk.

Metodekapittelet vil blant annet gi en begrunnelse for valg av kvantitativ metode, beskrive utvalg og instrument. En betydelig del av kapittelet er viet beskrivelse av Wolf og Smiths (2007) rammeverk for validering og analyse av datamaterialet som er samlet inn slik at det tilpasses Rasch-modellen. Metoder som er brukt for undersøkelse av sammenheng mellom MI og prestasjoner er Pearsons *r* og en metode for sammenligning av korrelasjoner, og kapittelet gir en forklaring for disse metodene. Til slutt kommer noen etiske betraktninger.

Resultatkapittelet beskriver resultater fra valideringen av instrument og elevresponser, resultater fra korrelasjonsanalyser og sammenligning av korrelasjoner, og kapittelet vil gi svar på forskningsspørsmålene.

I diskusjonene diskuterer jeg om instrumentet er egnet til å måle MI, og hvilke begrensninger jeg har oppdaget ved instrumentet underveis i prosessen. Sammenhengen mellom MI og matematikkprestasjoner diskuteres, og her vektlegges en diskusjon om bruk av karakterer i studien, som kan ha lav reliabilitet, validitet og som ikke nødvendigvis gir mål på intervallnivå. Kapittelet vil videre inneholde en diskusjon om studiens aktualitet i forbindelse med fagfornyelsen, resultater sett i lys av tidligere forskning, samt pedagogiske implikasjoner og forslag til videre forskning.

2 Teori

For å kunne svare på forskningsspørsmålet mitt om hva som er sammenhengen mellom MI og prestasjoner i matematikk, må jeg først få klarhet i hva som menes med MI, og hvordan det er mulig å måle denne slik at jeg får mål som kan brukes i videre statistiske analyser.

MI er ikke entydig definert blant forskere som bruker begrepet (Darragh, 2016; Graven & Heyd-Metzuyanim, 2019; Radovic, Black, Williams & Salas, 2018). Definisjonen av holdninger til matematikk vil sjeldent være passende for alle situasjoner, og selv om man ble enige om én definisjon, vil den sannsynligvis være for generell til å være nyttig (Kulm (1980) referert i Zan & Di Martino, 2007). Definisjonene vi lager av affektive aspekter tilpasses undersøkelsen man skal gjøre. I mitt tilfelle er den psykometriske metoden en viktig faktor i definisjonen av det affektive aspektet MI. Jeg vil derfor i dette kapittelet først skrive om måling av psykologiske faktorer generelt, og om Rasch-modellen spesielt, før jeg går inn på definisjonen av MI. Siden metoden er en del av definisjonen av MI, vil dette kapittelet inneholde teori om måling.

I tillegg til det metodologiske aspektet ved definisjonen av MI i denne undersøkelsen vil begrepet også defineres av et sett med karakteristikker av «matematiske» personer, skrevet om til utsagn i et spørreskjema. Det å være matematisk settes blant annet i sammenheng med å ha positive følelser for matematikk og å jobbe dypt med matematikk, og teorier om dette beskrives i kapittel 2.2.4.

Målene på MI som dannes i Rasch-analysen blir satt i sammenheng med prestasjoner i matematikk, og jeg vil til slutt i teorikapittelet forklare hva jeg legger i begrepet prestasjoner, forskjellene på en vurdering gitt av lærer og en vurdering som er gjort automatisk og anonym, og forskning på hva lærere legger vekt på når de vurderer elever i matematikk.

2.1 Måling og psykometri

Hva måling er, og spesielt hva som er målbart, har vært diskutert gjennom historien. Vi kan være enige om at fysisk lengde lar seg måle, men hva med psykologiske fenomener? Fysikken har fundamental direkte (fysisk sammensatt) måling og derivert (indirekte registrert eller konstruert) måling til å dekke objekters målbare fysiske egenskaper (Bond & Fox, 2015). Lengde kan måles direkte mot den skalaen som brukes, i hvert fall dersom vi måler relativt små lengder. Vekt kan også direkte måles mot kjente størrelser. I tillegg kan vi gjøre måling på grunnlag av hvordan noe virker på andre ting. Et eksempel på dette er temperatur. Vi kan ikke måle temperatur direkte, men vi kan måle hvordan ulik temperatur virker på andre stoffer og fenomener. Et mål av temperatur med et kvikksølvtermometer vil være et mål på hvor mye kvikksølv utvider seg og våre erfaringer og teorier knyttet til dette. Til slutt måler vi i prinsippet lengden på kvikksølvsøylen som danner seg, men oversetter det til den temperaturen det tilsvarer.

Å måle psykologiske faktorer som kompetanse og personlighet kan sammenliknes med å måle temperatur. Vi kan ikke ta og føle på personlighet, men er avhengige av å se på hvordan personlighet eller andre psykologiske faktorer kommer til uttrykk. Dette betyr at de psykologiske faktorene må operasjonaliseres før vi kan måle dem.

Måling av mentale egenskaper og psykologiske fenomener kalles psykometri. Psykometri ble definert av Galton som «kunsten å gjøre målinger på, og å tillegge tall på tankens handlinger» (min oversettelse) (Galton, 1879). Hvordan måling av noe så vagt som «tankens handlinger» kunne gjøres riktig, har vært gjenstand for mye diskusjon siden tidlig i forrige århundre, og det har resultert i utvikling av flere metoder for måling av mentale egenskaper og psykologiske fenomener. Metoden som brukes i denne studien er Rasch-modellen, som ble utviklet av Georg Rasch på 1950- og 60-tallet (Wright & Stone, 1979).

Felles for måling av fysiske og psykologiske faktorer er at målene er relasjonelle. Dette betyr at man måler mot en kjent skala eller bakgrunn. Skalaen for lengde er gitt av for eksempel en meterstokk, mens skalaene i psykologiske målinger er litt mindre åpenbare. I Rasch-modellen dannes skalaen av utsagnene eller oppgavene som er brukt i spørreskjemaet respondentene har svart på. Jeg vil forklare hvordan dette skjer i kapittelet om Rasch-modellen (2.1.3) og i kapittelet om definisjon av MI (2.2).

Videre vil jeg først si litt om debatten i psykometrifeltet på 1900-tallet og litt om noen teorier og prinsipper for måling, før jeg argumenterer for at Rasch-modellen er en tilfredsstillende modell for den type måling som gjøres i denne undersøkelsen.

2.1.1 Stevens skalaer

Etter at en komité av fysikere og psykologer, nedsatt av «British Association for the Advancement of Science», hadde jobbet i åtte år med spørsmålet om det i det hele tatt var mulig å måle menneskelig sansing, konkluderte de i 1940 i «Final Report» med at dette ikke var mulig (Michell, 1997). Begrunnelsen var at en måling måtte være en fysisk operasjon, som å legge stenger etter hverandre for lengdemåling, eller stable murstein for å måle vekt (Linacre, 2005).

Psykologen Stanley S. Stevens var uenig og svarte i 1946 med en ny definisjon av måling: «måling, i videste forstand, er definert som kobling av tall til objekter eller hendelser etter regler» (min oversettelse) (Stevens, 1946). Stevens skrev i sin artikkel at reglene som definisjonen henviser til ville resultere i måling på ulike nivåer og med ulike skalaer: nominell, ordinal, intervall og ratio. Jeg vil nå prøve å utdype hvordan Stevens (1946) definerte de ulike skalaene.

Nominelle skalaer klassifiserer objekter eller hendelser etter kvalitative karakteristikker. Objekter eller hendelser gis navn eller nummer som viser hvilken klasse de tilhører, for eksempel klassifisering etter farge, kjønn, form eller art. Statistisk analyse av data fra nominelle skalaer begrenser seg til å telle opp og finne typetall, og eventuelt sammenligne typetall i ulike grupper (Stevens, 1946).

Ordinale skalaer brukes til å rangere data. Ordinale skalaer kan være av typen intelligenstester, der du får vite om du er bedre enn andre, men hvor betydningen av 10 poeng høyere intelligens ikke nødvendigvis betyr det samme midt på skalaen som det gjør i endene av skalaen (Eysenck, 1998). Intervallene er ikke nøyaktig definert, og en konsekvens er at man ikke kan bruke resultatene til å gjøre visse statistiske analyser, som gjennomsnitt. Gyldig statistisk mål i ordinale skalaer er median og tilhørende persentiler (Stevens, 1946).

Intervallskalaer har kjente intervaller og kan brukes til sammenlikninger og til å finne forskjeller (Stevens, 1946). De fleste statistiske analyser kan gjøres her, med unntak av de som krever at det finnes et absolutt nullpunkt. Ofte er nullpunktet på en intervallskala satt et sted der det ble funnet hensiktsmessig. Et eksempel er temperatur. Celsius-skalaen har sitt nullpunkt der vann fryser/smelter (ved havoverflaten). Det finnes mange ulike temperaturskalaer, men de er alle intervallskalaer, og det er mulig å gå fra den ene til den andre ved hjelp av en koeffisient, og/eller ved justering av nullpunktet. Problemet oppstår hvis man ønsker å si at noe er dobbelt så mye som noe annet. Ti grader Celsius er ikke dobbelt så mye temperatur som fem grader, for eksempel. Det ville vært det samme som at 50 grader Fahrenheit er det dobbelte av 41 som jo ikke gir mening. Intervallene er like, men hva vi kaller dem og hvor nullpunktet er, er ulikt. Kelvin-skalaen har derimot et absolutt nullpunkt og kan defineres som neste skala-type, nemlig ratio-skala.

Ratio-skalaer er skalaer som oppfyller alle kravene i de tre foregående skalaene i tillegg til at det finnes relasjonelle sammenhenger (Stevens, 1946). En dobling av verdi vil bety en dobling av det som er målt. Skalaen for lengde er et eksempel på en slik skala. Den har nominelle egenskaper i at de ulike lengdemålene har ulike navn, lengdemålene er ordinale da det er tydelig når noe er lengre enn noe annet, skalaen har like intervaller, enten det er centimeter eller tommer, og den har et absolutt nullpunkt. Det absolutte nullpunktet, som er et kriterium for denne typen skala, gjør det mulig å gå fra én enhet til en annen kun ved hjelp av en koeffisient.

Selv om Stevens definerte egenskapene til de ulike skalaene og målene de gir, er det flere som hevder at hans definisjon på måling har skapt mer forvirring enn oppklaring. Linacre (2012) er kritisk til Stevens «vide» definisjon på måling og hevder at det innen sosialvitenskap har vært vanlig å kalle det for mål uansett hva slags tall de har fått, nettopp på grunn av formuleringen i Stevens definisjon, og at dette har ført til forvirring i det sosialvitenskapelige forskningsfeltet. Analyser med Rasch-modellen gir mål på intervallnivå, og målene kan derfor brukes til sammenlignende statistiske analyser, som gjennomsnitt og korrelasjon som er utført i denne studien.

2.1.2 Thurstone

Thurstone definerte noen kriterier som må være til stede for at man skal kunne gjøre en statistisk analyse av sosiale/psykologiske faktorer. Hans tre hovedkriterier var endimensjonalitet, additivitet og invarians (Andrich, 1989).

Endimensjonalitet handler om å måle det man skal måle, og ikke noe annet i tillegg (Andrich, 1989). I fysikk kan det være enkelt å se at en meterstokk bare måler lengde og ikke for eksempel litt masse og temperatur i tillegg. Dette kan være vanskeligere å skille i psykologi. Hva måler man når man måler intelligens? Kun intelligens? Eller spiller leseferdighet og tallforståelse også inn? Egenskapen man måler må kunne legge seg et sted på en linje, et kontinuum, der et mål av egenskapen som er større, ligger på den ene siden, og et mål som er mindre, vil ligge på den andre (Andrich, 1989).

Additivitet handler om at enhetene på skalaen du måler etter må ha samme intervaller, slik at dersom du legger til én enhet, så legger du til like mye uavhengig av hva du hadde fra før (Andrich, 1989). Nominelle og ordinale skalaer har som nevnt ikke denne egenskapen, mens intervall- og ratioskala tilfredsstiller kriteriet om additivitet (Stevens, 1946).

Invarians handler om at selve måleinstrumentet ikke må endre seg dersom det brukes på ulike grupper (Andrich, 1989). Det skal også være mulig å måle noen med færre oppgaver eller utsagn uten at målet endrer seg, og resultatet av en måling skal ikke være avhengig av de som lagde instrumentet.

Disse tre kriteriene som Thurstone formulerte for måling av sosiale variabler, var nettopp *kriterier*, og ikke *antagelser* (Andrich, 1989). For å kunne gjøre målinger på datamaterialet måtte man først sjekke om datamaterialet tilfredsstilte kravene til måling. Data måtte passe til analysemodellen. Motsatt tilfelle var ifølge Andrich (1989) utbredt blant forskere på psykologiske variabler. De hadde et sett med data, og søkte etter en analysemodell som passet datasettet.

2.1.3 Rasch-modellen

Georg Rasch lyktes i å vise hvordan de strenge kriteriene i fysikken kunne brukes i sosialvitenskap gjennom sine modeller, som han selv kalte for «Modeller for måling» (Models for Measurement) (Linacre, 2012). Rasch bygde på Thurstones kriterier for måling, og fant en måte der man kunne gi personer mål på samme skala som utsagnene/oppgavene de ble målt etter (Wright & Stone, 1979). Datagrunnlaget i Rasch-analysen er typisk responser på et spørreskjema med spørsmål man skal svare på (ja/nei) eller utsagn man skal angi i hvilken grad man er enige i.

Skalaen dannes av rekkefølgen utsagnene får i analysen, og både rekkefølgen og intervallene bestemmes av hvor mange som svarer at de er enige i hvert utsagn. Resultatet blir en intervallskala. Et utsagn som mange er enige i, vil legge seg langt nede på skalaen, mens et utsagn som få er enige i, vil legge seg høyere på skalaen. Målet personen får, vil være det samme som vanskelighetsgraden til utsagnet der det er like sannsynlig at personen vil være enig som uenig (Linacre, 2012). En skala som ikke endrer struktur når utsagn testes på ulike grupper, vil være invariant for disse gruppene.

Dersom det finnes et utsagn der det er like sannsynlig at respondentene er enige, enten de har høyt eller lavt mål, vil dette vises i analysen og kan tolkes som at utsagnet tilhører en annen dimensjon, altså at det måler noe annet. Dette utsagnet kan etter en statistisk og kvalitativ vurdering tas ut av analysen slik at målingen blir endimensjonal.

2.1.4 Sammendrag

Rasch-modellen, som brukes i denne studien, støtter seg til Thurstones teorier om additivitet, endimensjonalitet og invarians (Andrich, 1989). Rasch-målene er på intervallnivå og kan være grunnlag for videre statistiske analyser (Stevens, 1946). I kapittel 2.2.3 vil jeg knytte målingsteori til begrepet MI. I tillegg til den metodologiske definisjonen av MI vil det videre i kapittelet følge en definisjon basert på teori, med utgangspunkt i utsagnene i spørreskjemaet som er brukt i denne studien.

2.2 Matematisk identitet

MI har vært mye omtalt i matematikkdidaktiske studier de siste tiårene, men begrepet har blitt omtalt som blant annet vagt, inkonsistent og førparadigmatisk (Darragh, 2016; Graven & Heyd-Metzuyanim, 2019; Radovic et al., 2018). Det at det ikke er noen overordnet enighet om hvordan begrepet skal defineres trenger ikke være begrensende for forskning innen feltet, men kan gi mulighet til å gjøre pragmatiske valg av definisjon som er tilpasset den aktuelle studien. I denne studien er det tatt utgangspunkt i at matematikk er en sosial aktivitet, og at å være matematisk er en struktur i denne aktiviteten.

2.2.1 Identitetsbegrepet

De aller fleste har en formening om hva identitet er. På skolen lærer vi at identitet er de kvalitetene og egenskapene som gjør deg unik og skiller deg fra andre (Aksnes, 2019).

Identitet kan være hvordan du definerer deg selv (Deaux, 1993), at du er en bestemt type person når du samhandler og handler i en gitt kontekst (Gee, 2000), eller hvem man er i et gitt samfunn (Bishop, 2012).

Forskere skiller mellom personlig identitet og sosial identitet. Personlig identitet er det som skiller deg fra andre mennesker som kjønn og utseende. Sosial identitet kan beskrives som følelsen av å tilhøre bestemte sosiale fellesskap som interessegrupper, politisk ståsted og kulturelle fellesskap, men at det er ulikt hva man personlig legger i en slik tilhørighet (Deaux, 1993). Identitet er et sentralt tema innen psykologi og sosiologi, der psykologene gjerne snakker om den personlige identiteten, og sosiologene snakker om den sosiale identiteten. Felles for mange definisjoner av identitet er at det handler om en kombinasjon av det personlige og det sosiale. Deaux (1993) skriver at det ikke er enighet om å skille mellom sosial og personlig identitet, og mener at de fundamentalt henger sammen.

2.2.2 Noen teorier om MI

Vi kan ha mange identiteter samtidig (Deaux, 1993; Gee, 2000). Jeg kan blant annet identifisere meg (i ulik grad) som kvinne, mor, lærer, musiker og matematiker. MI er altså én av mange identiteter man kan ha. Hvordan definerer vi MI? Hvordan kommer MI til uttrykk? Når MI skal måles, må vi definere og operasjonalisere begrepet slik at vi kan finne noe målbart.

Radovic et al. (2018) gikk igjennom forskningsartikler som omhandler MI de siste 20 årene, og prøvde å danne seg et bilde av både hvordan forfatterne definerer MI, og hvordan de operasjonaliserer begrepet. Bakgrunnen for artikkelen er at litteraturen om MI gir inkonsise definisjoner av begrepet. Definisjonene ble delt inn i tre hoveddimensjoner: (1) sosial/subjektiv, (2) handlende/representativ, og (3) endring/stabilitet (min oversettelse). Operasjonaliseringene som var gjort i litteraturen ble delt inn i fem hovedkategorier: (1) Identitet som individuell attributt, (2) identitet som narrativ, (3) identitet som forhold til bestemte praksiser, (4) identitet som måter å handle på, og (5) identitet som gitt og begrenset av lokale praksiser (min oversettelse) (Radovic et al., 2018).

For å få et innblikk i hvordan rammeverket til Radovic (2018) kan brukes til å forstå ulike definisjoner av identitet, skal jeg komme med noen eksempler fra matematikkdidaktisk forskning. Eksemplene viser at det er mange måter å tilnærme seg begrepet identitet, og spesielt MI. Definisjonene er mer ulike tenkemåter om identitet, og ikke nødvendigvis motsetninger av hverandre. Jeg vil også beskrive hvordan Kaspersen (2018), som har utviklet instrumentet som brukes i denne studien, definerer MI. Definisjonen som er brukt i denne studien støtter seg på Kaspersens (2018) definisjon, og har, i tillegg til definisjon ut fra målingsteori, elementer fra flere teorier om MI. Teorier om MI operasjonalisert som narrativer, som forhold til spesielle praksiser, og som måter å handle på (Radovic et al., 2018) kan alle til en viss grad knyttes til utsagn i instrumentet som er brukt i denne studien.

Sfard og Prusak (2005) ser på identitet som narrativer og definerer identitet som et sett av konkretiserende, signifikante og bekreftbare historier om en person. Sfard og Prusak (2005) både operasjonaliserer og definerer identitet til å være historiene som blir fortalt om oss. Dette gjelder både historiene andre forteller om oss til oss eller til andre, eller historiene vi selv forteller om oss selv til oss selv eller til andre. Historiene *er* identiteten, ikke et uttrykk for, eller en representasjon av den. Identiteten er et produkt av kollektiv historiefortelling, og læring kan sees på som å lukke gapet mellom aktuell identitet og ønsket identitet (Sfard & Prusak, 2005). Noen av studiene som omtaler identitet som narrativer la vekt på det subjektive: selvforståelse, selvrefleksjon og personens rolle i konstruksjon av personlige historier. Andre studier så på narrativene mer som selvposisjonering i diskursive rom, og vektla sosiale forhold og strukturelle begrensninger heller enn individuelle måter å handle på (Radovic et al., 2018).

Solomon (2007) definerer MI som en følelse av inkludering eller ekskludering i et matematisk fellesskap, en definisjon som Radovic (2018) kategoriserer som «identitet som forhold til bestemte praksiser». Solomon intervjuet matematikkstudenter om hvorvidt de følte at de var del av et matematikkfellesskap der de selv var aktive bidragsytere, inspirert av Wengers (2010) teori om identitet som tilhørighet til praksisfellesskap. Mange av studentene følte at matematikk var noe som ble «gjort med dem», ikke «gjort av dem». De var ekskludert fra et viktig identitetsaspekt: deltakelse i forhandling om mening (Solomon, 2007). Wenger (2010) hevder at du ikke bare definerer din identitet ut fra hva som er kjent, og fra hvor du deltar og er med på å forhandle om mening, men også ut fra hva du ikke er, hva som er ukjent, noe som fører til «ikke-deltakelse» (non-participation). De som deltar i utvikling av ideer og mening, og derfor har eierskap til innholdet, har høy grad av tilhørighet til praksis (Solomon, 2007). Radovic (2018) hevder at forfattere som omtaler identitet som «forhold til bestemte praksiser», ser på praksisene der det forhandles om mening som lokale, for eksempel som praksis i et klasserom, og ikke som forhandling om mening i den sosiale diskursen.

Bishop (2012) definerer identitet som et dynamisk syn på seg selv, forhandlet i en spesifikk sosial kontekst og påvirket av tidligere historie, hendelser, personlige narrativer, erfaringer, rutiner og måter å delta på. Identitet er hvem man er i et gitt samfunn, og er derfor både individuelt og kollektivt definert (Bishop, 2012). Hun mener at identitet inkluderer affektive aspekter som følelser, holdninger og forestillinger. Videre konkretiserer hun MI til ideene man har om hvem man er med tanke på matematikk og aktivitetene som hører til. Hun mener at MI er avhengig av hva matematisk aktivitet betyr i et gitt samfunn, klasserom eller liten gruppe, og derfor at MI er avhengig av konteksten. Radovic et al. (2018) klassifiserer dette synet på identitet til å være identitet som måter å handle på i spesifikke sosiale kontekster. Felles for flere artikkelforfattere med dette synet på identitet er at de ser på hvordan elever posisjonerer seg selv i forhold til andre.

Kaspersen (2018), som utviklet instrumentet som brukes i denne studien, definerte MI til å være «den relative posisjonen mellom personer og den sosiale strukturen av å være matematisk i den aktiviteten man deltar i». Den sosiale strukturen til MI, også kalt den sosiale identiteten, er operasjonalisert som et sett med karakteristikker på det å være matematisk, og deres interne struktur. Den personlige identiteten blir hvordan personer posisjonerer seg selv relativt til den sosiale strukturen. På samme måte som Deaux (1993) setter Kaspersen den personlige identiteten i sammenheng med den sosiale identiteten og mener at identitet er relasjonell av natur (Kaspersen, 2018). Der andre forskere kan se på MI som noe flerdimensjonalt, har Kaspersen (2018) valgt å definere MI ut fra Thurstones (Andrich, 1989) prinsipper for måling (additivitet, endimensjonalitet og invarians), noe som gjøre at hans definisjon begrenser seg til én dimensjon av MI, samt en nødvendig tilpassing av data til Rasch-modellen.

2.2.3 MI og prinsipper for måling

Måling av identitet gjør identitet til noe *relasjonelt*, fordi måling i seg selv er relasjonell. Måling av lengde er mulig fordi vi har en referanse å måle etter. Måling av psykologiske variabler må derfor også være mål i forhold til en standard, målestokk eller kjent struktur. For MI vil denne målestokken være rekkefølgen og strukturen som utsagnene danner under analyse med Rasch-modellen, og personmålet vil være hvor man plasseres relativt til strukturen.

Additivitet er det første kriteriet Thurstone satte for måling av psykologiske faktorer (Andrich, 1989). Rasch-modellen gir mål på intervallskala (Bond & Fox, 2015, s. 1), en skala som tilfredsstiller kravet om additivitet. Man kan ikke si at MI på 2 er dobbelt så mye som MI på 1, men vi kan si at økning av 1 enhet MI betyr det samme uansett hvor på skalaen økningen finner sted. Fordelen med data på intervallmål er at de kan være utgangspunkt for flere statistiske analyser som sammenligning av gjennomsnitt i ulike grupper (f.eks. *t*-test) og korrelasjonsanalyser med Pearsons *r*.

Endimensjonalitet handler ifølge Thurstone (Andrich, 1989) om at måling bare kan skje i én dimensjon av gangen. Dimensjonen av MI som jeg har brukt i min studie, er knyttet til positive følelser og dypt arbeid med matematikk, men det er ikke utelukket at det også finnes andre dimensjoner av MI. Perfekt endimensjonal måling er vanskelig å oppnå. Hvis meningen er å måle MI, kan det hende at litt av det som måles, også er for eksempel leseforståelse. Rasch-modellen søker etter en måling som er endimensjonal nok til det formålet testen er utviklet for (Bond & Fox, 2015). Utsagnene som definerer den sosiale strukturen «å være matematisk» i denne studien, ble av Kaspersen (2018) valgt ut på grunnlag av hvilke utsagn som la seg i samme dimensjon i Rasch-modellen. Vi vet derfor at utsagnene representerer noe målbart. Observasjon av noe målbart var utgangspunkt for teorien om at dette «noe» kunne være MI (Kaspersen, 2018)

Invarians handler om at MI kan defineres ut fra karakteristikker som er relativt stabile når de analyseres i undergrupper av datamaterialet, for eksempel oppdelt i kjønn eller i ulike matematikkfag. Utsagn som legger seg på signifikant ulike steder på skalaen når de testes på ulike undergrupper, bør undersøkes nøyere før man kan konkludere med at et instrument er invariant. Dersom ulik plassering av utsagn ikke har signifikant betydning for personmål, kan man likevel vurdere å godta instrumentet som invariant. Et instrument som skal brukes på større deler av befolkningen og i ulike kontekster, må testes for invarians i flere representative kontekster og undergrupper.

2.2.4 Hva vil det si å være matematisk?

MI handler i denne oppgaven om hva som karakteriserer en person som er «matematisk» i en gitt kontekst (Kaspersen, 2018). Det å være matematisk kan sees på som å tenke og jobbe som en matematiker, å jobbe konseptuelt med matematikk eller å jobbe dypt med matematikk (Entwistle, McCune & Tait, 2013; Kaspersen, 2018; Skemp, 1976). Når jeg senere i oppgaven bruker begrepet dypt arbeid med matematikk eller dybdelæring, vil dette begrepet også innebære konseptuelt og relasjonelt arbeid med matematikk. Å være matematisk vil i denne undersøkelsen i tillegg handle om positive følelser for matematikk (Kaspersen, 2018), som glede over faget og indre motivasjon for å lære. Jeg skal nå komme med eksempler på teori om hva det vil si å være matematisk slik det er definert i denne undersøkelsen.

Matematisk tankegang knyttes til å jobbe konseptuelt eller relasjonelt med matematikk (Skemp, 1976). Skemp (1976) sammenlikner det å lære matematikk med det å bli kjent i en ny by. Du kan lære deg en veibeskrivelse fra A til B, huske på hvor du skal ta til venstre og ved hvilket bygg du skal ta til høyre for å komme fram. Et annet alternativ er å bli kjent i byen på egenhånd slik at du finner fram uten å huske den nøyaktige veibeskrivelsen. Du vil da risikere å ta omveier, men du vil kanskje lære av det, og dersom du går feil, vil du kunne rette opp og finne fram likevel. På samme måte kan elever lære matematikk ved å følge oppskrifter, med den risikoen at de ved den minste feil ikke vil kunne hente seg inn igjen for å komme til riktig løsning, noe Skemp (1976) beskriver som en instrumentell tilnærming. Alternativet er en relasjonell tilnærming der elevene bygger opp en begrepsmessig struktur som hjelper dem å finne løsningsstrategier uavhengig av hvor i prosessen de er.

Hiebert (1986) har definert konseptuell kunnskap og prosedyrekunnskap litt på samme måte som Skemp (1976) har definert relasjonell og instrumentell forståelse. Hiebert (1986) beskriver konseptuell kunnskap som «rik på sammenhenger» og «et sammenhengende nett av kunnskap, et nettverk hvor koblingene er like framtredende som de enkeltstående bitene av informasjon» (min oversettelse). Han ser ikke på de to typene kunnskap som motsetninger, men som viktige sammen. Konseptuell kunnskap oppnås ved å koble sammen biter av informasjon. Dette kan skje enten ved å koble sammen biter av informasjon som man allerede har tilegnet seg, eller ved å koble ny informasjon til tidligere opparbeidet kunnskap (Hiebert, 1986).

Burton (1998) intervjuet matematikere om hvordan de gjennom oppvekst og karriere hadde lært seg matematikk. Metaforen om et kart gikk igjen også her. Matematikerne beskrev et kart med mange veier, et kart med huller der de måtte bygge broer og en reise mot en fjelltopp der de hele tiden hadde målet i syne, men at veien dit var uklar og uforutsigbar. En annen metafor de brukte, var et puslespill der flere og flere brikker falt på plass, og de beskrev gleden ved å legge en brikke som var viktig for å komme videre. Felles for tankegangen er at matematikk er mer enn endimensjonale «veier» eller oppskrifter for å nå målet. Matematikerne i Burtons (1998) studie pekte også på det å se sammenhengen mellom ny kunnskap og gammel kunnskap som avgjørende for å lære matematikk.

2.2.5 Definisjon av MI for denne studien

I denne studien har jeg valgt å bruke samme formulering for definisjonen av MI som Kaspersen (2018). MI defineres altså som «den relative posisjonen mellom personer og den sosiale strukturen av å være matematisk i den aktiviteten man deltar i». Den sosiale MI er et sett med karakteristikker på det «å være matematisk» og hvordan de strukturerer seg i Rasch-Modellen. Den personlige MI er hvordan personer responderer på disse karakteristikkene, og målet de får i Rasch-modellen. Definisjonen tar hensyn til Thurstones (Andrich, 1989) prinsipper for måling. Dimensjonen av MI som måles i denne studien er knyttet til dypt arbeid med matematikk, positive følelser for faget.

En Rasch-analyse gjør det mulig å studere den sosiale MI og den personlige MI samtidig. Siden jeg skal sammenligne MI med prestasjoner i matematikk er det den personlige MI og måltallene fra analysen jeg kommer til å konsentrere meg om. Det er likevel et poeng i å se på hvordan den sosiale MI strukturerer seg. Hvilke utsagn legger seg øverst på skalaen? Disse utsagnene vil være utsagn som kjennetegner personer med høy MI, og kan brukes som et kompass for bedre undervisning dersom styrking av MI er et mål.

I oppgaven vil jeg veksle mellom å snakke om «sosial struktur» og «skalaen som dannes av utsagnene i Rasch-modellen» når jeg omtaler den sosiale identiteten av å være matematisk. Når jeg senere i oppgaven snakker om personmål, Rasch-mål eller MI vil dette handle om posisjonen elevene i studien tar relativt til den sosiale identiteten «å være matematisk». MI blir derfor et mål som er relativt til konteksten som undersøkelsen er gjort i.

2.3 Prestasjonsbegrepet

Jeg omtaler både elevenes terminkarakterer i matematikk og deres resultater på kartleggingsprøven som prestasjoner i denne studien. Andre begreper som kunne vært brukt, og som kan ligge under samlebetegnelsen *prestasjoner*, er *ferdigheter* og *kompetanse*. Siden det er hva elevene faktisk klarer å vise at de kan som skal være grunnlag for å sette en karakter i matematikk, mener jeg at ordet prestasjoner er mest dekkende. Jeg skal i det følgende forklare hva jeg mener med begrepet når jeg bruker det, samt belyse hva lærere legger til grunn når de vurderer elever.

Terminkarakter i matematikk settes av faglærer halvveis i skoleåret. Opplæringsloven sier at «Grunnlaget for vurdering i fag er kompetansemåla i læreplanane for fag slik dei er fastsette i læreplanverket», jf. § 1-1 eller § 1-3 (Lovdata, 2006). Kompetansemålene er en beskrivelse av hvilken spesifikk faglig kompetanse eleven skal ha etter endt skoleår, og inneholder ikke punkter som handler om affekt. Selv om det i kapittelet «Føremål» står at matematikken skal skape positive holdninger, er det ingen kompetansemål som sier noe om dette (Utdanningsdirektoratet, 2013). Elevene skal derfor ikke vurderes på holdninger eller progresjon i matematikk.

En terminkarakter er noe annet enn en standpunktkarakter. Eleven er halvveis i året når terminkarakteren settes, og eleven har helt fram til sluttvurderingen mulighet til å vise sin kompetanse, også for det som ble vurdert til første termin. Terminkarakteren har derfor ikke noen direkte sammenheng med standpunktkarakteren, men vil i de fleste tilfeller gi en pekepinn på hvordan elevens kompetanse er på slutten av skoleåret.

Det er faglærer som setter terminkarakterer og standpunktkarakterer, og det er gjort undersøkelser på hva lærere faktisk legger til grunn for karakterene de setter. Prøitz og Borgen (2010, s. 61) undersøkte hva et utvalg lærere ved seks norske skoler mener at en standpunktkarakter skal uttrykke, og kom fram til fire ulike tilnærminger: (1) oppnådd kompetanse til slutt i opplæringen, (2) oppnådd kompetanse til slutt og underveis, (3) oppnådd kompetanse, holdninger, innsats og andre sider ved eleven og (4) helhetlig kompetanse. Utvalget av matematikklærere sier at innsats og progresjon ofte vektlegges, og spesielt når det dreier seg om svake elever (Prøitz & Borgen, 2010). Dette kan tyde på at faglig svake elever kan få bedre karakter enn kompetansen deres skulle tilsi.

Kartleggeren måler elevenes faglige nivå i matematikk (Fagbokforlaget.no, 2020). Den er ment som et kartleggingsverktøy for å lage tilpasset undervisning, og tester elevene i følgende områder:

- De fire regneartene: Eleven skal løse blandede oppgaver med de fire regneartene. Det er også noen oppgaver der eleven skal bestemme hvilken regneart som må benyttes for å løse en oppgave.
- Tallsystemet: Eleven skal gjennom forskjellige oppgaver vise at de forstår 10-tallssystemet.
- Hverdagsliv: Oppgaver der eleven skal regne med tid, omregning av enheter, valuta og hastighet.

- Brøk og prosent: Beregne og forkorte ulike brøker, gjøre om tall fra grunnform til prosent og utføre ulike beregninger ut fra prosentsats.
- Geometri: Eleven skal navngi figurer, beregne areal, omkrets og volum.
- Statistikk: Elevene skal lese av ulike diagram, tolke størrelsessammenhenger, lese av grafer og angi gjennomsnittsverdier.
- Ligninger (fra 8. trinn og opp): Eleven skal løse ulike ligninger med en ukjent.

(Fagbokforlaget.no, 2020)

Kartleggingsprøven rettes automatisk og anonymt. Forhold som holdninger, progresjon og MI vil derfor ikke kunne vektlegges når prøven vurderes. Det samme gjelder skriftlig eksamen i matematikk, som også vurderes anonymt.

Statistikk fra SSB viser at eksamenskarakterer, som rettes anonymt, ligger litt lavere enn standpunktkarakterer for elever på ungdomsskole og videregående skole.

Tabell 1. *Gjennomsnittlige standpunkts- og eksamenskarakterer for matematikk etter endt grunnskole i Norge, 2012-2019 (ssb.no, 2019).*

	2012	2013	2014	2015	2016	2017	2018	2019
Standpunkt	3,5	3,5	3,5	3,5	3,5	3,6	3,7	3,7
Eksamen	3,1	3,1	3,0	2,9	3,3	3,4	3,6	3,6

Ulikhetene i standpunktkarakterer og eksamenskarakterer kan for eksempel tyde på at faglærere som kjenner elevene vektlegger andre faktorer enn elevenes ferdigheter når de setter karakterer, at eksamensresultatet er mer sårbart for at elevene har en dårlig dag, eller at elever sliter med angst for prøvesituasjoner.

2.4 Oppsummering

Matematisk identitet defineres i dette kapittelet som «den relative posisjonen mellom personer og den sosiale strukturen av å være matematisk i den aktiviteten man deltar i». Utgangspunktet er karakteristikker av personer som er «matematiske» og strukturen karakteristikkene lager i Rasch-modellen. Rasch-modellen er en psykometrisk modell som gjør det mulig å måle psykologiske faktorer, som MI, etter Thurstones prinsipper om additivitet, endimensjonalitet og invarians. Strukturen danner den sosiale MI som fungerer som en målestokk for den personlige MI. Den personlige MI er målene elevene får i Rasch-analysen. Å være «matematisk» er definert teoretisk som å jobbe dypt med matematikk og å ha positive følelser for faget. Prestasjoner er brukt som samlebegrep for prestasjoner, ferdighet og kompetanse.

3 Metode

I denne studien har jeg undersøkt korrelasjonen mellom videregående elevers MI og resultater de har fått på kartleggingsprøven Kartleggeren og terminkarakterene de fikk første termin skoleåret 2019/2020.

Instrumentet jeg brukte til å måle MI, var ikke ferdig validert til bruk på tvers av kontekster, og jeg har derfor selv validert det for min kontekst: elever på videregående skole. Det var resultatene fra denne valideringen som gjorde det mulig å svare på det første forskningsspørsmålet. Dataprogrammer som er brukt i analyse av data, er Winsteps (Linacre, 2006) og IBM SPSS Statistics 25.

Korrelasjonsanalyser var mitt grunnlag for å svare på det andre og tredje forskningsspørsmålet. Jeg har kun sett på korrelasjon mellom variabler i denne undersøkelsen, ikke på årsakssammenheng.

Metodekapittelet inneholder beskrivelser av metoder for innsamling av data, både om praktisk gjennomføring og om instrumentet/spørreskjemaet som ble brukt til innsamlingen av data fra elevene. Videre følger en del om metode for analyse av data som viser hvordan Rasch-modellen kan gi mål på MI, validering av instrumentet og statistiske metoder som er brukt for å se på sammenhenger. Til slutt kommer et underkapittel om etiske betraktninger rundt informasjon til elevene, samtykke, og behandling av forskningsdata.

3.1 Begrunnelse for valg av kvantitativ metode

Et tilgjengelig og utprøvd instrument for måling av MI var avgjørende for mitt valg av kvantitativ metode, men det finnes også andre grunner til å ta et slikt valg. Kvantitativ metode gir mulighet for å se «det store bildet», siden store mengder data kan analyseres samtidig (Cohen, Manion & Morrison, 2018). Resultatene fra kvantitative analyser kan skape grunnlag for utvelgelse av tilfeller for videre kvalitativ forskning (Cohen et al., 2018, s. 847). Kvantitativ analyse kan også brukes til å undersøke et fenomen som er oppdaget kvalitativt, i stor skala (Cohen et al., 2018, s. 850).

Tidlig forskning på matematikkrelatert affekt var hovedsakelig kvantitativ forskning der respondentene svarte på spørreskjemaer som «Mathematics Anxiety Rating Scale» (Zan et al., 2006), eller «Mathematics Attitude Scales» (Fennema & Sherman, 1976). Testene målte affektive dimensjoner som verdier, forestillinger, selvtillit i matematikklæring, matematikkangst, og motivasjon i problemløsning.

McLeod (1994) kritiserte, i en gjennomgang av forskning på matematikkrelatert affekt, den utstrakte bruken av kvantitativ metode, som han så preget forskningsfeltet. Han mente at kvantitativ metode hadde bidratt med nyttig informasjon, men at feltet måtte komplementeres med den nye tilgjengelige innsikten om hva forskning på affekt kan være. Han mente også at forskere som brukte psykometriske metoder, la mer vekt på metodens reliabilitet enn på metodens validitet (McLeod, 1994).

Forskning på MI ble ikke nevnt i McLeods (1994) gjennomgang. Vi vet derimot at forskning på MI begynte, og gradvis økte, i tiden etterpå og fram til nå, og at de fleste studiene på MI er gjort kvalitativt (Darragh, 2016; Graven & Heyd-Metzuyanim, 2019;

Radovic et al., 2018). Det har altså skjedd et skifte i metode, noe som gjør at denne studien kan bidra til å komplementere de kvalitative studiene av MI.

3.2 Metode for innsamling av data

3.2.1 Utvalg

Skolen som var med i undersøkelsen, var en videregående skole med yrkesfag som bygg og anlegg, teknologi og industriell produksjon, elektro, og design og håndverk, samt studieforberedende for kunst, design og arkitektur, og teknologiske og allmenne fag, i tillegg til påbygg for generell studiekompetanse. Alle elever som hadde et matematikkfag dette skoleåret ble spurt om å delta, og de var fordelt på vg1, vg2 og vg3. Noen elever hadde to ulike matematikkfag, og da rapporterte de det faget de hadde første gang de ble spurt om å delta i undersøkelsen. Se fordeling av fag i Tabell 2. 15 % av elevene ved skolen hadde minoritetsspråklig bakgrunn.

Tabell 2. *Antall respondenter for ulike matematikkfag*

Fagkode	Beskrivelse	Antall elever	Prosent
1PY	Praktisk matematikk for VG1 på yrkesfag	143	40 %
2PY	Praktisk matematikk for påbyggingsfag	128	36 %
1P	Praktisk matematikk for VG1, studieforberedende	16	5 %
2P	Praktisk matematikk for VG2, studieforberedende	24	7 %
1T	Teoretisk matematikk for VG1, studiespesialiserende	24	7 %
R1 og R2	Teoretisk matematikk for VG2/3, studiespesialiserende	20	6 %
Sum		355	100 %

Jeg fikk inn 355 spørreskjemaer med samtykke om å bruke opplysningene fra skjemaet, men 28 av disse samtykket ikke til at jeg kunne samle inn resultater fra Kartleggeren og terminkarakterer. I tillegg var det 18 elever som ikke fikk karakter i faget fordi de hadde sluttet, byttet fag, eller ikke hadde oppfylt kravet om 90 % tilstedeværelse aktuell termin. Av de som hadde samtykket, manglet 31 elever resultater på Kartleggeren, enten fordi de ikke var til stede da den ble tatt, fordi de har tatt den på en annen skole tidligere, eller fordi de ikke hadde fullført og levert testen da de tok den. Se oversikt i Tabell 3.

Tabell 3. *Antall respondenter for parameterne MI, terminkarakterer og kartleggingsresultat*

Parameter	Antall elever
Mål på MI	355
Terminkarakterer	309
Kartleggingsresultater	296
Alle tre data	284

3.2.2 Instrumentet

Instrumentet som ble brukt til måling av MI i denne undersøkelsen, var et sett med 20 utsagn som elevene skulle svare i hvilken grad de var enige i. Jeg vil i teksten veksle mellom å omtale instrumentet som spørreskjema og instrument.

Instrumentet ble i utgangspunktet laget for å måle hvorvidt elever jobbet konseptuelt med matematikk, eller hvor dypt de jobbet med matematikk (Kaspersen, 2018), og hadde i utgangspunktet 40 utsagn. Inspirasjon til de 40 utsagnene ble hentet fra (a) liknende instrumenter, spesielt instrumentet «Approaches and Study Skills Inventory for Students (ASSIST) (Entwistle et al., 2013), som er et instrument ment for å måle studenters tilnærminger til læring i tre hoveddimensjoner: dybdelæring, overflatelæring og strategisk læring, (b) eksisterende litteratur om forståelse i matematikk (f.eks., Hiebert, 1986; Skemp, 1976) og (c) medlemmer av matematiske fellesskap (f.eks. doktorgradsstudenter i ingeniørfag) (Kaspersen, 2018). Av de 40 opprinnelige utsagnene ble 20 valgt ut til å danne instrumentet. Dette fordi de la seg i samme dimensjon i Rasch-modellen.

Instrumentet var allerede validert for to andre grupper, ingeniørstudenter og lærerstudenter, i en tidligere studie (Kaspersen, 2018). Jeg valgte å bruke det samme instrumentet uten å gjøre andre endringer enn utseende og informasjonstekst til elevene (se vedlegg 2-5). Grunnen til at jeg ikke endret på utsagnene, var at jeg visste at instrumentet var under validering for bruk på tvers av kontekster i to andre masterprosjekt, og jeg tenkte at jeg kunne bidra med informasjon om hvordan instrumentet fungerer for konteksten videregående skole med yrkesfag. Utsagnene kommer her i nummerert rekkefølge:

1. Jeg tar initiativ til å lære mer om et matematisk emne enn skole/jobb legger opp til.
2. Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode.
3. Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker.
4. Jeg har problemer med å legge fra meg matematiske oppgaver.
5. Dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv.
6. Jeg blir engasjert når noen starter en matematisk diskusjon.
7. Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.
8. Matematiske ideer jeg leser eller hører om, setter meg på sporet av egne tankerekker.
9. Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre.
10. Hvis jeg prøver på en metode som ikke fører frem, bruker jeg tid på å finne ut hvorfor denne ikke virker.
11. Når jeg lærer en ny metode/algoritme, prøver jeg å finne ut hvorfor den virker.
12. Når jeg kommer over et matematisk bevis/forklaring, studerer jeg det til det gir mening.
13. Når jeg møter et matematisk problem, tenker jeg jeg over om det finnes flere måter å løse oppgaven på.
14. Når jeg jobber med et matematisk problem, hopper jeg mellom ulike strategier.
15. Når jeg lærer noe nytt, fører det til at det er flere ting jeg ønsker å finne ut av.
16. Når jeg jobber med en oppgave, stopper jeg opp underveis og reflekterer over hva jeg gjør.
17. Hvis jeg står fast, prøver jeg å visualisere problemet.

18. Jeg kan forklare hvorfor løsningen min er rett.
19. Jeg prøver å koble det jeg lærer opp mot det jeg vet fra før.
20. Jeg fortsetter å prøve meg fram selv om jeg ikke får det til med en gang.

Utsagnene i instrumentet har svaralternativer på en Likert-skala. Dette er en skala som gir respondentene graderte svaralternativer. I mitt tilfelle var de fem alternativene: «aldri/nesten aldri», «noen ganger», «ofte», «alltid/nesten alltid», og «vet ikke». Hvert alternativ har en verdi knyttet til seg, i dette tilfellet verdiene fra 1-4 på de ladede utsagnene og X på «vet ikke». Observasjoner av responser for alternativet «vet ikke» ble registrert som «missing data».

Utfordringen med en Likert-skala er at det kan være vanskelig å lage svaralternativer med like intervaller (Cohen et al., 2018). Er man dobbelt så mye enig om man svarer «ofte» som når man svarer «aldri/nesten aldri»? Rasch-modellen tar hensyn til at intervallene mellom svaralternativene varierer, og bestemmer intervallene basert på hvordan respondentene har brukt svaralternativene i den aktuelle undersøkelsen (Bond & Fox, 2015, s. 117). Dersom det er litt lengre avstand mellom «aldri/nesten aldri» og «noen ganger» enn det er mellom «noen ganger» og «ofte», går det altså fint så lenge det samme mønsteret gjelder for alle utsagnene.

Svaralternativet «Vet ikke» er et av de fem mulige i denne undersøkelsen, et svaralternativ som i visse undersøkelser kan være uheldig (Cohen et al., 2018, s. 481). Jeg oppfordret elevene til å velge dette alternativet dersom de ikke forstod innholdet i utsagnet. Dette gjorde jeg for at respondentene ikke skulle svare vilkårlig dersom det var noe de ikke forstod. Rasch-modellen er robust for «missing data» (Bond & Fox, 2015, s. 171), og et «vet ikke»-alternativ er i så måte et bedre alternativ enn tilfeldige svar som ville gitt høyere misfit-verdier (se kapittel 3.4 om validering av instrumentet) for elever og utsagn. «Vet ikke» kan være et problematisk svaralternativ dersom spørsmålene er sensitive og man kan risikere å svare noe som er sosialt lite akseptert (Cohen et al., 2018, s. 481). Jeg vurderer utsagnene i den aktuelle undersøkelsen som lite sensitive og anser derfor en inkludering av svaralternativet «vet ikke» som bedre enn å ikke ha denne muligheten.

For at elevene ikke skulle kopiere eleven ved siden av mens de fylte ut skjemaet, lagde jeg fire ulike versjoner av samme skjema. Spørsmålene var i utgangspunktet nummerert fra 1-20, og jeg brukte «tilfeldig»-funksjonen i Excel for å lage fire ulike rekkefølger av de 20 utsagnene. Skjemaene ble merket med bokstavene a, b, c og d. Elevene fikk informasjon om at skjemaene var ulike, slik at de ikke skulle tro at naboen hadde likt skjema, for så å bli fristet til å kopiere det den andre svarte.

3.2.3 Pilot

Jeg gjennomførte en pilot tidlig i prosessen. Den ble utført i én 1PY-klasse på en annen skole. Hensikten med piloten var for lettere å kunne gjennomføre datainnsamlingen i de 23 klassene som jeg hadde planlagt. Jeg fikk gjennom piloten informasjon om hvor lang tid undersøkelsen ville ta, hvilken informasjon jeg skulle gi lærere og elever, og på hvilken måte jeg skulle gi informasjonen. Fra jeg begynte med informasjon i klassen og til siste elev hadde levert inn spørreskjemaet, tok det i overkant av 15 minutter. På grunnlag av dette planla jeg innsamling i maksimalt to klasser for hver økt på 45 minutter. Det ga en plan over innsamling som strakk seg over to uker høsten 2019.

Data fra spørreskjemaene i piloten ble testet i analyseprogrammet Winsteps (Linacre, 2006) og fungerte som en øvelse i Rasch-analyse. Siden det bare var 11 elever som svarte på pilotundersøkelsen, ble det for lite data til at jeg kunne bruke resultatene til å gjøre justeringer av instrumentet før den faktiske datainnsamlingen. Jeg brukte derfor det samme skjemaet i datainnsamlingen som jeg gjorde i piloten.

3.2.4 Innhenting av informasjon om prestasjoner

Jeg ønsket å bruke informasjon både om elevenes terminkarakter i matematikk og om deres resultater på kartleggingsprøven. For å skaffe informasjon om elevenes resultater på Kartleggeren, måtte en ansatt ved skolen hente resultater for én og én elev fra en digital database på Kartleggeren.no. Kartleggeren var gjennomført i alle klasser i løpet av de første ukene i skoleåret, og resultatene lå derfor klare tidlig i prosessen. Resultatene ble plottet inn i et regneark sammen med elevenes mål på MI. Terminkarakterer ble hentet inn ved at jeg sendte lister til faglærerne om hvilke av deres elever som hadde samtykket til at jeg kunne hente informasjon om terminkarakter. Da lærerne hadde satt terminkarakter, fikk jeg listene tilbake ferdig utfylt. Dette var mot slutten av januar 2020. Terminkarakterene ble også koblet mot informasjonen jeg hadde om elevene fra før, og ført inn i samme regneark.

3.3 Rasch-modellen

Jeg valgte Rasch-modellen for analyse av dataene fra spørreskjemaet. Rasch-modellen er en sannsynlighetsmodell, og er en av flere metoder for psykometri. Modellen er først og fremst valgt fordi den tidligere har vært brukt for måling av MI med samme instrument som jeg har brukt, men også fordi den gir mål på intervallnivå.

Rasch-modellen gir mulighet for å måle utsagn og personer på den samme skalaen, noe som gir mulighet for å studere både strukturen utsagnene danner og målene elevene får. Et utsagn som mange er enige i, vil få lav vanskelighetsgrad, og et utsagn som få er enige i, vil få høy vanskelighetsgrad. For personmål vil en person som er enig i få utsagn få et lavt mål, mens en person som er enig i mange utsagn vil få et høyt mål. Rasch-modellen vil kunne gi personmål selv om det mangler respons på noen av oppgavene fordi modellen er robust mot manglede data (Bond & Fox, 2015).

Utgangspunktet til Rasch var dikotome tester som målte skoleferdigheter eller intelligens, og hans hovedspørsmål var: «Når en person på dette nivået (antall riktige) prøver på en oppgave av denne vanskelighetsgraden (antall personer som har klart den), hva er sannsynligheten for at denne personen klarer oppgaven? Svar: Sannsynligheten for suksess avhenger av differansen mellom nivået på personen og vanskelighetsgraden til oppgaven» (min oversettelse) (Bond & Fox, 2015, s. 11).

Andrich (1978) videreutviklet modellen til også å gjelde for tester med flere graderte svaralternativer (polytome tester); modellen som brukes i denne studien. Jeg vil likevel vise hvordan utregningen foregår for den dikotome modellen, da de polytome modellene tar utgangspunkt i denne.

Vi ønsker å finne sannsynligheten for at en person svarer «rett» (eller ja) på en oppgave. Dersom oppgaven er dikotom, vil det typisk bety at rett svar, eller ja, tilsvarer 1 poeng, og galt svar, eller nei, tilsvarer 0 poeng. For å regne ut sannsynligheten for at en elev får rett (1) på en oppgave, må vi vite målet til personen og vanskelighetsgraden til oppgaven.

$$log_e\left(\frac{P_{ni}}{1-P_{ni}}\right) = B_n - D_i$$

Formel 1. *Rasch-modellen for dikotome tester (Linacre, 2012)*

P_{ni} = *sannsynligheten for å svare rett (1) på oppgave i for person n*

B_n = *målet til person n*

D_i = *vanskelighetsgraden til oppgave i*

Dersom vanskelighetsgraden til oppgaven er lik målet til personen, vil dette gi en sannsynlighet for at personen skårer rett på oppgaven på 0,5. Er vanskelighetsgraden større enn personmålet, vil sannsynligheten bli mindre enn 0,5, og hvis vanskelighetsgraden til oppgaven er mindre enn personmålet, vil sannsynligheten bli større enn 0,5. Den oppgaven som det er 50 % sannsynlig at en person klarer, vil altså ha samme vanskelighetsgrad som målet til denne personen. Målene for personer og utsagn (sannsynlighetene for suksess) konverteres til naturlige logaritmer (log-odds) som gir måleenheten logit (log-odds unit). Denne konverteringen gjøres for å få mål på intervallnivå, for å unngå at skalaen komprimeres i endene, og for å gjøre skalaen uavhengig av antall oppgaver/utsagn (Bond & Fox, 2015, s. 30).

Rasch-modellen er en variant av *item response theory* (IRT). Forskjellen mellom Rasch-modellen og andre IRT-modeller er at når man bruker Rasch-modellen, må man sikre seg at dataene passer med modellen, i motsetning til andre modeller der det er mulig å tilpasse modellen til dataene man har (Bond & Fox, 2015, s. 265). Fit-analyser av datamaterialet er derfor en viktig del av datavalideringen. Fordelen med Rasch-modellen og andre IRT-modeller er at målene blir på intervallnivå, og at personmålene er uavhengige av vanskelighetsgraden til spørsmål/utsagn. Rasch-modellen krever derimot at det måles i kun én dimensjon, noe som ikke er en forutsetning for andre IRT-modeller.

En alternativ behandling av data fra spørreskjemaer er å telle opp råskår fra et spørreskjema og bruke poengsummen som et slags mål. Problemet med denne type databehandling er at poengsummen blir påvirket av vanskelighetsgraden til oppgavene. Dersom en overvekt av oppgavene er lette, vil mange få høyt mål, og dersom en overvekt av oppgavene er vanskelige, vil mange få lave mål. En skala som i en psykometrisk undersøkelse baserer seg på opptelling av råskår vil heller ikke bli på intervallnivå.

I Rasch-modellen brukes ordet *item* om oppgaven eller spørsmålet som respondenten skal svare på, eller utsagnet som respondenten skal si seg mer eller mindre enig i. Siden det ikke finnes noe godt ord for *item* på norsk bruker jeg ordet utsagn, siden det er *det* testen jeg har brukt inneholder.

3.4 Validering av instrumentet

Wolfe og Smith (2007) har utviklet et rammeverk for validering av psykometriske instrumenter. Rammeverket fungerer som en veiledning, og det beskriver ulike metoder for validering innen åtte aspekter.

De fire første aspektene blir behandlet i analysekapittelet og består av innholdsaspektet, det substansielle aspektet, det strukturelle aspektet og generaliserbarhetsaspektet (min oversettelse). De fire siste aspektene blir behandlet i ulik grad i diskusjonskapittelet og består av det eksterne aspektet, konsekvensaspektet, responsivitetsaspektet og tolkningsaspektet (min oversettelse). Videre kommer en forklaring på hva som ligger i disse aspektene, og hvilke metoder jeg har valgt for å sikre noen av de ulike aspektene for validitet.

3.4.1 Innholdsaspektet

Validering av innholdsaspektet handler om å sikre at utsagnene er basert på relevant og representativt innhold, samt å sikre teknisk kvalitet på utsagnene (Wolfe & Smith, 2007). Det er også en kontroll av at respondentene forstår, eller om de har lik forståelse av innholdet i utsagnene. Elever som ikke forstår utsagnet, vil kunne svare noe helt annet enn det som er forventet ut fra målet eleven har på MI.

Teorien som ligger bak utsagnene i instrumentet er beskrevet i teorikapittelet, og ble validert ved kollegakontroll da instrumentet ble utviklet (Kaspersen, 2018). Jeg vil derfor konsentrere meg om den tekniske valideringen av utsagnene i gruppen jeg skal måle.

En metode som kan brukes, er å se på utsagnenes «point-measure correlation» (PM corr.), og ifit og outfit «mean-squared fit» (MNSQ). PM corr. viser korrelasjonen mellom personers mål på MI og deres respons på et utsagn. Infit MNSQ og outfit MNSQ viser om det er mange uventede responser på et utsagn.

PM corr. måler Pearsons korrelasjon (se kapittel 3.5.1) mellom skårene på ett bestemt utsagn og målene til respondentene som har gitt disse skårene (Wolfe & Smith, 2007). Negativ korrelasjon kan tyde på at et utsagn er kodet feil, altså at personer med lav MI vil være enige i utsagnet, mens de med høy MI ikke er enige. Lav korrelasjon (enten positiv eller negativ) kan tyde på at det er tilfeldig om en med høy eller lav MI er enige i utsagnet, og kan for eksempel tyde på at utsagnet måler noe som er i en annen dimensjon. En korrelasjon på over 0,4 er tilfredsstillende for en undersøkelse med polytom responsskala (Wolfe & Smith, 2007).

Det er en forutsetning at innsamlet data passer modellen tilfredsstillende for å oppnå invariante målinger på intervall-nivå (Bond & Fox, 2015, s. 266). Fit-analysene infit MNSQ og outfit MNSQ indikerer om det er uoverensstemmelser mellom dataene vi har samlet og Rasch-modellens antagelser (Bond & Fox, 2015, s. 266). Dersom det for et utsagn er mange uventede responser (som at en respondent med lav MI likevel er enig i et utsagn med høy vanskelighetsgrad), vil dette gi høye «fit»-verdier. Dersom både infit MNSQ og outfit MNSQ er høye, kan det tyde på et systematisk avvik, og at det er noe ved utsagnet som gjør at det ikke passer inn i modellen. Dersom outfit MNSQ er høy, mens infit MNSQ er innenfor kritiske grenser, kan det tyde på at det er tilfeldige feil som slår ut, altså at det for eksempel kun gjelder én respondent, men at det likevel gir stort utslag.

En infit MNSQ eller outfit MNSQ på 1 betyr at det er svært god overenstemmelse mellom hva respondentene har svart og hva modellen har forutsagt. En fit-verdi på 1,3 betyr at det er 30 % mer variasjon enn det modellen har forutsagt, og en verdi på 0,7 betyr at det er 30 % mindre variasjon enn modellen har forutsagt (Bond & Fox, 2015, s. 269). Jeg har i denne undersøkelsen satt grensene for infit MNSQ og outfit MNSQ til 30 %, som er samme grenser som Kaspersen (2018) brukte i sin studie, noe som er et litt strengere kriterium enn Winsteps' manual oppgir (0,5-1,5) (Linacre, 2006).

Beregning av outfit MNSQ tar utgangspunkt i at Rasch-modellen regner ut et sannsynlig responsmønster for hver respondent basert på målet til respondenten. Videre sjekkes respondentens faktiske responser på utsagnene opp mot det som var forventet. For hvert utsagn kan man derfor kontrollere hvor mye avvik det er mellom forventet respons og faktisk respons hos hver enkelt respondent.

Residualet (*R_{ni}*) er differansen mellom forventet respons (*E_{ni}*) for person *n* på utsagn *i,* og observert respons (*P_{ni}*). Standardisert residual (*Z_{ni}*) fås ved å dele residualet på standardavviket ($\sqrt{W_{ni}}$) til de observerte responsene:

$$Z_{ni} = \frac{R_{ni}}{\sqrt{W_{ni}}}$$

Formel 2. *Standardisert residual (Linacre, 2012)*

Videre beregnes outfit MNSQ for utsagn *i* (*U_i*) ved å finne gjennomsnittet av de kvadrerte standardiserte residualene, når *N* er antall responser på utsagn *i*:

$$U_i = \frac{\sum_{n=1}^{N} Z_{ni}^2}{N}$$

Formel 3. *Outfit MNSQ (Linacre, 2012)*

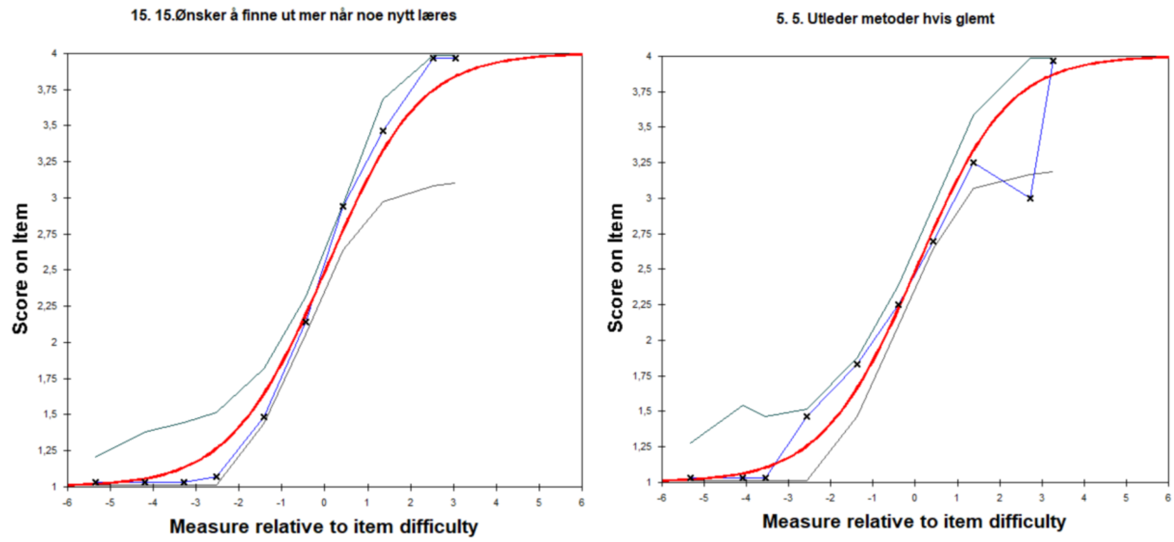
Et stort avvik vil gi relativt større utslag enn et lite avvik ved utregning av outfit MNSQ. Dette gjør outfit MNSQ god til å kontrollere for tilfeldige feil. Få respondenter med store avvik vil gi større utslag enn mange respondenter med små avvik. Dette justeres i utregningen av infit MNSQ, da de små avvikene vektes. Infit-verdien er derfor god til å avdekke systematiske avvik (Bond & Fox, 2015).

$$U_i = \frac{\sum_{n=1}^{N} Z_{ni}^2 W_{ni}^2}{\sum_{N=1}^{N} W_{ni}^2}$$

Formel 4. *Infit MNSQ (Linacre, 2012)*

Forventet og observert responsmønster for hvert utsagn, relativt til utsagnets vanskelighetsgrad, kan også fremstilles grafisk slik det er vist i Figur 1 av to «item characteristic curves» (ICC).

Den røde kurven (jevn S-form) viser hva Rasch-modellen har beregnet som den mest sannsynlige responsen (y-aksen) for personer med mål relativt til vanskelighetsgraden til utsagnet (x-aksen). En person som har samme Rasch-mål som vanskelighetsgraden til utsagnet blir beregnet til 0 (residual=0) på x-aksen, og vil mest sannsynlig svare 2,5 på utsagnet (i praksis like gjerne 3 som 2). De observerte personresponsene er gruppert i intervaller på 1 logit, og det er gjennomsnittet for gruppen som vises i den blå grafen (kurve med kryss). Store residualer for høye og lave personmål kan skyldes at gjennomsnittet beregnes av færre respondenter.



Figur 1. *Item Characteristic Curves (ICC) for utsagn 15 og 5.*

Rød graf (jevn S-form) angir forventet respons, blå graf (med kryss) viser gjennomsnittlig observert respons med intervaller på 1 logit, og svarte linjer (ytterkant) angir 95 % konfidensintervall.

3.4.2 Det substansielle aspektet

Respondenten uttrykker sin respons etter først å ha tolket utsagnet, og deretter å ha tolket svaralternativene. Det substansielle aspektet ved valideringen handler om å kontrollere om responsene opptrer konsistent med intensjonene til de som utformet svaralternativene i spørreskjemaet (Wolfe & Smith, 2007). Det substansielle aspektet sikres ved å analysere svaralternativene, og ved å se på *person fit.*

Etter å poengtere at det språklige i hva vi velger å kalle de ulike *svaralternativene* er viktig for at de skal kunne forståes slik vi ønsker, beskriver Linacre (2002) åtte retningslinjer for hvordan man kan lage gode spørreskjemaer med Likert-skala. De fire første retningslinjene er essensielle (Wolfe & Smith, 2007). Jeg beskriver her seks av de åtte retningslinjene:

1) Det må være minst 10 observasjoner for hvert svaralternativ (Linacre, 2002). Dette er for å få presise analyser av hvordan svaralternativene forholder seg til hverandre. En årsak til at et svaralternativ ikke er representert, kan være at det er svært vanskelig å velge. Hadde jeg i min undersøkelse byttet ut alternativet «alltid/nesten alltid» med bare «alltid», ville det kanskje vært enda vanskeligere å velge dette alternativet. Ved få observasjoner av et alternativ bør man derfor kontrollere svaralternativene kvalitativt. Dersom et svaralternativ har under 10 observasjoner, kan det tas bort da det ikke har noen verdi for analysen.
2) Det bør være jevn distribusjon av observasjoner mellom svaralternativene (Linacre, 2002). Dette er også en kontroll på om det er svaralternativer som er vanskelige å være enige i. I noen undersøkelser vet man derimot at man kommer til å få lav respons på de ekstreme kategoriene, som for eksempel hvis man har et spørreskjema om ungdomskriminalitet (Linacre, 2002).
3) Gjennomsnittlig mål på deltakerne som svarer for hvert svaralternativ bør øke med verdiene til svaralternativene for å støtte at de rangerte svaralternativene brukes konsistent på tvers av utsagnene (Linacre, 2002). Avvik kan tyde på at verdiene til svaralternativene ikke er satt rett.

4) Outfit MNSQ for svaralternativene må være under 2,0 (Linacre, 2002) for å sikre at ikke ett av alternativene har tendens til å velges tilfeldig eller uventet blant respondentene, og derfor ikke er i samsvar med Rasch-modellen.
5) Alle svaralternativene må være mest sannsynlige for et intervall av personmål relative til utsagnenes vanskelighetsgrad (Wolfe & Smith, 2007).
6) Tersklene der et svaralternativ går over til å bli mer sannsynlig enn det tilstøtende, bør ha en avstand på mer enn 1,1 logit for en undersøkelse med fire svaralternativer (Wolfe & Smith, 2007). Respondentene bør altså ha i gjennomsnitt mer enn 1,1 logit høyere mål for å svare en kategori høyere på utsagnene.

Analyse av *person fit* viser om det finnes respondenter som har tendens til å ha mange uventede responser ved å se på personenes infit og outfit MNSQ. Respondenter som har over 2,0 i infit MNSQ eller outfit MNSQ, bør kontrolleres (Wolfe & Smith, 2007), og kan vurderes utelatt fra analysen. Det uvanlige svarmønsteret kan skyldes at respondenten har spesielle interesser som ikke samsvarer med den vanlige strukturen, at det er utsagn de ikke forstår, eller at de krysser av mer eller mindre tilfeldig (Linacre, 2002).

3.4.3 Det strukturelle aspektet

Validering av det strukturelle aspektet kan være å sikre at modellen kun måler i én dimensjon. Data kan aldri passe Rasch-modellen helt perfekt, men det er et mål at instrumentet er endimensjonalt nok til den type måling instrumentet er laget for (Bond & Fox, 2015). Utsagn som legger seg i en annen dimensjon bør kontrolleres både kvantitativt og kvalitativt før man avgjør om det skal være med i instrumentet eller ikke. I min studie undersøkte jeg dimensjonalitet med en «principal component analysis» (PCA).

3.4.4 Generaliserbarhetsaspektet

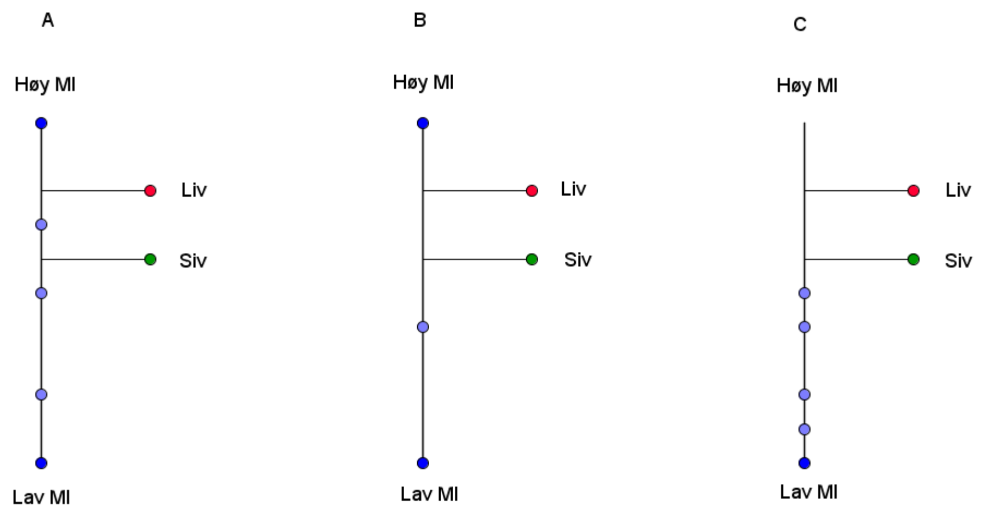
Generaliserbarhetsaspektet handler om i hvilken grad målene er meningsfylte på tvers av kontekster (Wolfe & Smith, 2007). Kan man sammenligne målene til respondenter som tilhører ulike grupper og kontekster? Vil strukturen av utsagnene bli lik for ulike undergrupper? For å sikre generaliserbarhetsaspektet ble det i denne undersøkelsen kontrollert for «differential item functioning» (DIF) mellom ulike grupper/kontekster innad i hovedkonteksten (den aktuelle videregående skolen). DIF ble kontrollert mellom elever på yrkesfag og elever som tar studieforberedende, mellom de fire ulike skjemaene som ble brukt, og mellom kjønn (20 elever var ikke definert for kjønn og ble tatt ut av analysen). DIF-testen lager Rasch-modeller for undergrupper i datamaterialet, for eksempel én for jentene og én for guttene. Disse modellene vil bli litt ulike med tanke på hvor vanskelighetsgraden for hvert utsagn ligger. Så gjøres det en signifikanstest for å se om forskjellen er av betydning.

Utsagnene kontrolleres for å se at differansen er under 0,64 logit (Boone, Staver & Yale, 2014, s. 282) og at p-verdien er over 0,05. Der begge er utenfor disse grensene, bør det kontrolleres for om differansen vil ha signifikant betydning for målene elevene får.

Jo flere DIF-analyser man ser på, jo høyere er sannsynligheten for å finne en DIF som er signifikant. Sannsynligheten for at en DIF som viser seg å være signifikant, faktisk ikke er det, ligger i kriteriets definisjon på 5 %, noe som kalles for Type 1-feil. En metode som kan brukes til å korrigere for dette, ble beskrevet av Bonferroni, og går ut på at hver gjennomført signifikanstest skal bruke et signifikanskriterium på (vanligvis) 0,05 delt på antall tester som er utført. Dette er effektivt, men kan igjen bli litt for strengt når det er veldig mange tester som utføres (Field, 2013).

Reliabilitet blir ofte sett på som noe annet enn validitet, men regnes av Wolfe og Smith (2007) som en del av generaliseringsaspektet til instrumentets validitet. Reliabilitet er et mål på om resultatene på testen ville blitt de samme dersom en tilsvarende test ble utført på en tilsvarende gruppe mennesker. Reliabilitet deles opp i «multiple form»-reliabilitet og «internal consistency»-reliabilitet (intern reliabilitet) (Wolfe & Smith, 2007). Den første tester for reliabilitet mellom tester med ulikt innhold, men som skal måle det samme, mens den andre tester et instrument internt uten at det må testes mot andre instrumenter. Jeg har testet den interne reliabiliteten til Rasch-modellen både for utsagn (item reliability index) og for personer (person reliability index) (Bond & Fox, 2015). Det er ikke noen fastsatt regel for hva som er grensen for god reliabilitet, men den bør ligge så nært 1 som mulig. Over 0,9 er svært godt, og over 0,8 er godt (Cohen et al., 2018, s. 774).

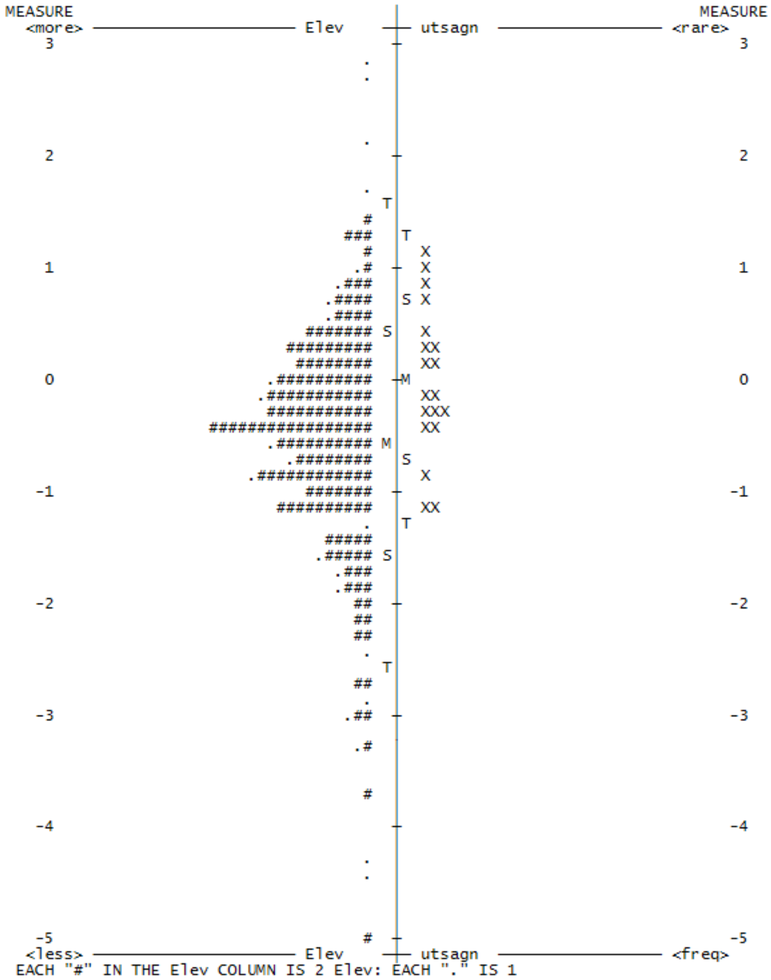
Reliabilitet kan påvirkes av antall utsagn i instrumentet, og spredningen av vanskelighetsgraden til utsagnene (Cohen et al., 2018, s. 282; Linacre, 2006) (se Figur 2). Det beste er utsagn som sprer seg jevnt utover hele skalaen som personmålene ligger på, men i praksis blir aldri spredningen perfekt. Det kan hende at utsagnenes vanskelighetsgrad er gjennomgående lav, slik at mange er enige i alle. Hvilket mål har personene da? De ligger over høyeste utsagn, men hvor høyt? På samme måte kan det være store mellomrom mellom vanskelighetsgraden til grupper av utsagn. Dersom det havner personer i områder på skalaen hvor det ikke finnes noen utsagn som kan fungere som målestreker, hva er da målet på disse personene? Dårlig reliabilitet vil stort sett bli bedre ved å legge til flere utsagn i måleområder som ikke er dekket av eksisterende utsagn (Cohen et al., 2018, s. 282; Linacre, 2006).



Figur 2. *Tre skalaer med ulik reliabilitet*

Liv og Siv (Figur 2) har i realiteten ulik MI, men det er ikke sikkert dette vil fanges opp av instrument B og C. Instrument A har god spredning av utsagn i hele måleområdet slik at Liv og Siv separeres og får ulike mål. I instrument B og C mangler det målepunkter i området der Liv og Siv er, og det vil derfor bli vanskeligere å skille mellom dem. Dette betyr at Liv og Siv i ulike undesøkelser risikerer å få samme mål. Test B og C har derfor dårlig reliabilitet.

Et Wright map (Figur 3) viser fordeling av utsagn og personmål i en Rasch-modell. Høyre side viser utsagnene plassert etter vanskelighetsgrad, og venstre side viser personmål på samme skala. Her kan man ta en visuell vurdering av om flere utsagn, eller utsagn med annen vanskelighetsgrad, ville kunne økt reliabiliteten.



Figur 3. *Wright map*

Personmål (venstre side) og utsagnenes vanskelighetsgrad (høyre side) legger seg på samme skala i Rasch-modellen.

3.4.5 Det eksterne aspektet

Det eksterne aspektet ved validitet er muligens det viktigste, og det som ligger nærmest det som tradisjonelt har blitt referert til som konstruktvaliditet (Wolfe & Smith, 2007). Hvordan henger resultatene sammen med andre mål av samme konstrukt? Da det ikke finnes mange andre kvantitative metoder for måling av MI blir dette vanskelig å få til i denne undersøkelsen.

3.4.6 Responsivitetsaspektet

Dette aspektet handler om muligheten instrumentet har for å måle endring. I medisin omtales dette aspektet som sensitivitet og handler om hvor lite endring som skal til før instrumentet fanger det opp (Wolfe & Smith, 2007). Et instrument med god reliabilitet, og derfor lite målefeil, vil lettere kunne oppdage små endringer.

3.4.7 Konsekvensaspektet

Dette aspektet ser på verdien av resultatene som grunnlag for handling (Wolfe & Smith, 2007). For hvem er resultatene viktige, og hvorfor? Hvilke konkrete tiltak kan gjøres på grunnlag av resultatene i studien? Disse spørsmålene prøver jeg å belyse i diskusjonskapittelet.

3.4.8 Tolkningsaspektet

Tolkningsaspektet handler om i hvilken grad betydningen av målene kommuniseres til de som skal tolke dem (Wolfe & Smith, 2007). Dette gjelder spesielt dersom instrumentet man validerer, skal brukes av andre enn de som har utviklet instrumentet, slik tilfellet er med kommersielle psykometriske tester. Da bør det kobles kvalitativ mening til de kvantitative målene som testen gir. Jeg skal ikke utvikle instrumentet for kommersiell bruk, men har likevel skrevet litt om hvilken kvalitativ mening et høyt mål på MI kan ha i diskusjonskapittelet.

3.5 Sammenheng mellom MI og prestasjoner

Det andre og tredje forskningsspørsmålet mitt var å undersøke om det var noen sammenheng mellom elevenes MI og deres prestasjoner i matematikk, og om det er noen forskjell i sammenhengen avhengig av om det er terminkarakter eller om det er karakter fra standardisert prøve.

Jeg har valgt å gjøre korrelasjonsanalyser mellom elevenes MI og henholdsvis karakterer og kartleggingsresultater. Jeg valgte å bruke Pearsons product-moment correlation coefficient (Pearsons *r*) siden dette er en test som passer for kontinuerlig data på intervallnivå (Cohen et al., 2018, s. 766).

MI er målt med Rasch-modellen, som gir mål på intervallnivå, men det kan diskuteres om karakterer som gis av lærere også er på intervallnivå, da de ikke er mål fra en metode som gir dette. Praksis er uansett som om karakterer er på intervallnivå da det brukes et gjennomsnitt av karakterer som grunnlag for inntak til videre utdanning. Resultater på Kartleggeren oppgis i prosent og er ikke nødvendigvis på intervallnivå, da det er uklart hvordan prosenten regnes ut, men jeg gjør en antagelse av at intervallene er gode nok for denne analysemetoden.

3.5.1 Korrelasjon

Et punktdiagram viser punkter grafisk for hver elev og deres plassering i forhold til to variabler, for eksempel MI på x-aksen og karakter på y-aksen. En titt her kan gi en indikasjon på hvordan sammenhengen er mellom de to variablene. Det er ikke alltid en sammenheng vil synes godt, og en korrelasjonsanalyse vil være hensiktsmessig for å kunne argumentere for en eventuell sammenheng.

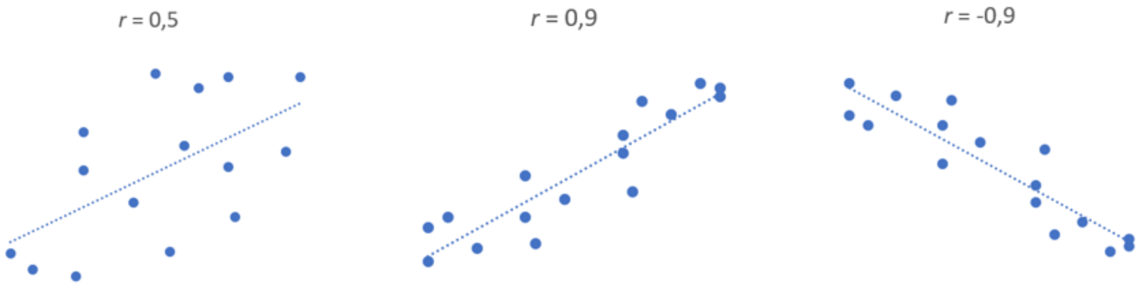
Korrelasjonsanalysen gir informasjon om hvor sterkt punktene i punktdiagrammet samler seg rundt en eventuell lineær regresjonslinje (Lysø, 2010), i hvilken retning linjen går, og om denne sammenhengen er signifikant. Pearsons Product Moment Correlation er en korrelasjonsanalyse som passer for parametrisk datamateriale på intervall-nivå (Cohen et al., 2018, s. 766).

Utregning av korrelasjonskoeffisienten (r) tar utgangspunkt i de to variablenes (x og y) kovarians (*cov*), og variablenes (x og y) egen varians, uttrykt ved standardavviket (s).

$$r = \frac{cov_{xy}}{s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(N-1)s_x s_y}$$

Formel 5. *Pearson Product Moment Correlatin (Pearsons r) (Field, 2013, s. 266)*

Korrelasjonskoeffisienten (*r*) uttrykker to ting. Tallet som oppgis, er mellom -1 og 1. Fortegnet uttrykker i hvilken retning sammenhengen går. Dersom fortegnet er positivt, betyr dette at det er positiv korrelasjon, og at for eksempel høy MI vil være knyttet til høye karakterer. Negativt fortegn betyr negativ korrelasjon, og viser til at når den ene variabelen øker, så vil den andre minke. Tallet som oppgis, uttrykker hvor godt de observerte punktene legger seg rundt den linja som er funnet ved regresjon (Lysø, 2010).



Figur 4. *Ulik grad av korrelasjon mellom to variabler rundt deres regresjonslinje*

Perfekt korrelasjon (1 eller -1) finnes bare der det er en absolutt sammenheng, som den teoretiske sammenhengen mellom diameter og omkrets av en sirkel. Korrelasjon i sosial forskning er sjelden bedre enn +/- 0,50 (Cohen et al., 2018). Tabell 4 viser en veiledning til hvordan vi kan tolke Pearsons *r* (Cohen et al., 2018, s. 746):

Tabell 4. *Tolkning av Pearsons r*

$Pearsons\ r$	Effekt	R^2	Forklaring
>0,10	liten	0,01	Forklarer 1 % av total varians
>0,30	medium	0,09	Forklarer 9 % av total varians
>0,50	stor	0,25	Forklarer 25 % av total varians

En korrelasjonsanalyse vil også teste om korrelasjonen er signifikant, altså sannsynligheten for at den estimerte korrelasjonen vi fikk kunne skjedd tilfeldig, og uten at det egentlig var noen korrelasjon der. Det er vanligvis ønskelig at denne sannsynligheten ligger under 0,05 (Cohen et al., 2018, s. 744).

3.5.2 Sammenligning av korrelasjoner

For å se om en korrelasjon er signifikant ulik en annen, kan man gjøre en *t*-test for å sammenligne korrelasjonene. Korrelasjonskoeffisientene (*r*) må da konverteres til *z*-verdier (z_r) for å gjøre prøvedistribusjonen normalfordelt, og derfor også standardavviket kjent (Field, 2013, s. 268).

$$z_r = \frac{1}{2} log_e \left(\frac{1+r}{1-r}\right)$$

Formel 6. *Konvertering av Pearsons r til z-verdi (z_r)*

Videre regner vi ut *z*-verdien til differansen (*N*=utvalget) (Field, 2013, s. 286):

$$z_{differanse} = \frac{z_{r1} - z_{r2}}{\sqrt{\frac{1}{N_1 - 3} + \frac{1}{N_2 - 3}}}$$

Formel 7. *Sammenligning av to korrelasjoner: z-verdien til differansen mellom z-verdiene til de to korrelasjonene*

Vi kan lete opp *z*-verdien vi nå får i en tabell, og lese av *p*-verdien. Denne angir så om differansen mellom korrelasjonene er signifikant.

3.6 Etikk og personvern

Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH) har et rammeverk med etiske retningslinjer for forskning i humaniora som gir grunnlag for gode refleksjoner over valg som tas underveis i en studie (De nasjonale forskningsetiske komiteene, 2016). Retningslinjene om «hensyn til personer» er av spesiell interesse i denne studien, da jeg har valgt å innhente respondentenes navn sammen med svar på spørreskjemaet.

Punkt 8 i rammeverket handler om samtykke og informasjonsplikt, og legger vekt på at et samtykke må være fritt, informert og uttrykkelig. Jeg skal nå gå igjennom hvordan jeg har prøvd å ivareta disse tre aspektene.

Fritt samtykke betyr at det er avgitt uten ytre press eller begrensninger av personlig handlefrihet (De nasjonale forskningsetiske komiteene, 2016). Det at forskeren selv er til stede når samtykke gis, kan føre til at deltakerne føler press til å signere. Jeg valgte likevel å selv være til stede, informere elevene, og organisere gjennomføringen av spørreundersøkelsen fordi jeg var redd for at elevene ville la seg påvirke av at faglærer skulle se besvarelsen deres da de fylte ut undersøkelsen. Jeg passet på å informere tydelig om at det var helt frivillig å delta, og var påpasselig med å ikke gi negativ tilbakemelding til de som valgte å avstå fra å svare.

Informert samtykke betyr at elevene får god informasjon om hva det vil innebære å delta i prosjektet. Informasjon ble gitt muntlig før spørreskjemaet ble utdelt, og faglærer fikk en digital versjon (se vedlegg 7) som ble lagt ut på elevenes digitale læringsplattformer. Da var det lett for elevene å gå tilbake og lese informasjonen på nytt, og eventuelt å finne informasjon om hvor man måtte henvende seg om man skulle trekke seg fra undersøkelsen.

Uttrykkelig samtykke betyr at deltakerne klart og tydelig gir uttrykk for at de er innforstått med hva det faktisk innebærer å delta i forskningsprosjektet (De nasjonale forskningsetiske komiteene, 2016). Elevene må aktivt krysse av for at de ønsker å delta, og for at vi kan innhente og bruke informasjon om dem. Ønsker vi samtykke til behandling av flere typer data om dem, bør det være egne avkryssingsbokser for hver type. Jeg hadde én avkryssingsboks for samtykke til å bruke svarene elevene hadde gitt på spørreskjemaet og én avkryssingsboks for samtykke til at jeg kunne innhente informasjon om karakterer og resultater på kartleggingsprøven. Av alle elevene som ga samtykke til at jeg fikk bruke svarene på spørreskjemaet, var det 28 stykker som ikke samtykket til ag jeg kunne hente informasjon om prestasjoner. Samtykkeerklæringen finnes som vedlegg (se vedlegg 6).

Jeg søkte Norsk senter for forskningsdata (NSD) om å få innhente personopplysninger i forbindelse med studien og fikk søknaden godkjent. Jeg ønsket at elevene skulle levere spørreskjemaet med navn av praktiske årsaker. Svarene elevene ga på undersøkelsen, skulle settes i sammenheng med deres terminkarakterer og kartleggingsresultater. For at ikke faglærer skulle bli unødvendig involvert i koding av elever og kobling av data, valgte jeg å be om elevenes navn slik at jeg kunne gjøre denne jobben selv. En annen grunn var som nevnt over at jeg ikke ville at elevene skulle la seg påvirke av at læreren skulle se hva de hadde svart på skjemaet.

Jeg valgte å fortelle hensikten med prosjektet til respondentene. De visste altså at deres svar på spørreskjemaet skulle settes i sammenheng med deres prestasjoner i matematikk. Dette var for å forklare hvorfor jeg trengte navn og samtykke til innhenting av terminkarakterer og kartleggingsresultater. At de visste hensikten med undersøkelsen, kan ha påvirket hvordan de svarte på spørreskjemaet.

3.7 Forskningsdata

NTNU har egne retningslinjer for forvaltning og håndtering av forskningsdata, med mål om å være så åpne som mulig (Research Data @NTNU, 2020). Alle prosjekter skal ha en datahåndteringsplan med beskrivelse av hvordan data i et forskningsprosjekt skal håndteres helt fra oppstart av prosjektet, gjennom hele forskningsprosessen og i tiden etter avsluttet prosjekt. Datahåndteringsplan for denne studien finnes som vedlegg (1). NTNU oppfordrer alle til å dele forskningsdata i NTNUs eget åpne arkiv for forskningsdata (Research Data @NTNU, 2020), noe jeg planlegger å gjøre ved prosjektets slutt.

3.8 Sammendrag

Jeg har i dette kapittelet forklart hvordan data har blitt samlet inn, validert og analysert i dette prosjektet. Det har blitt lagt vekt på Wolfe og Smith (2007) sitt rammeverk for validering av data for Rasch-modellen, da god validering av data er en forutsetning for å kunne få gode intervallmål på elevenes MI. Resultatene fra valideringen, og justeringene som følge av disse, vil jeg beskrive i resultatkapittelet. Jeg har hatt mindre kontroll på hvordan karakterene og kartleggingsresultatene har oppstått, men bruker dem i videre analyser på samme måte som målene jeg har på MI. I neste kapittel vil jeg også vise hvilke resultater jeg har fått på korrelasjonsanalysene, og analyse for sammenligning av korrelasjoner som er beskrevet i dette kapittelet.

4 Resultater

I denne delen av oppgaven viser jeg resultatene på analysene som jeg har gjort for å svare på mine forskningsspørsmål.

Mitt første forskningsspørsmål var: «Er det mulig å måle MI hos elever på videregående skole?» For å svare på dette måtte jeg validere måleinstrumentet som er brukt i innsamling av datamaterialet. Jeg vil i dette kapittelet vise hvordan jeg ved hjelp av Wolfe og Smith (2007) sitt valideringsverktøy sikret at utsagnene som ble brukt i instrumentet og responsene fra elevene passet Rasch-modellen.

Videre forskningsspørsmål dreier seg om hvorvidt det er sammenheng mellom elevenes MI og deres prestasjoner i matematikk; «Hva er korrelasjonen mellom elevenes mål på MI og deres matematikkprestasjoner?» For å svare på dette spørsmålet har jeg gjort to korrelasjonsanalyser av datamaterialet, den første av elevenes MI og deres terminkarakterer i matematikk, og den andre av elevenes MI og deres resultater på Kartleggeren*.*

Mitt siste forskningsspørsmål var: «Hvordan samsvarer korrelasjonen mellom MI og karakterer gitt av faglærer, og korrelasjonen mellom MI og resultater på en anonym kartleggingsprøve?» For å svare på dette måtte jeg undersøke om det var noen signifikant forskjell på korrelasjonskoeffisientene jeg fant mellom MI og henholdsvis karakterer og resultater på kartleggingsprøven.

Kapittelet er delt i to hvor første del handler om validering av innsamlet data og instrumentet som er brukt, og andre del handler om sammenligning av data.

4.1 Er det mulig å måle MI hos elever på vidergående skole?

Før jeg kunne bruke elevmålene som jeg fikk da jeg analyserte elevresponsene med Rasch-modellen, måtte jeg være sikker på at både utsagn, elevresponser og svarkategorier oppførte seg i samsvar med hensikten til instrument og analysemodell. Før jeg begynte på analysene, gjorde jeg derfor en Rasch-analyse av alt datamaterialet. Tabell 5 viser resultater for utsagnenes vanskelighetsgrad, infit MNSQ, outfit MNSQ og PM corr. i den første analysen.

Kolonnen helt til venstre viser i hvilken rekkefølge utsagnene har lagt seg, med det utsagnet som færrest elever var enige i øverst (høyest vanskelighetsgrad) og det som flest var enige i (lavest vanskelighetsgrad) nederst. Måltallet for utsagnene, eller vanskelighetsgraden, ser vi i andre kolonne. Rasch-analysen har nå også gitt alle elevene et mål på samme skala som utsagnene. Før jeg visste om jeg kunne bruke disse målene direkte, måtte jeg validere dataene. De ulike trinnene i valideringen blir gjennomgått i de neste avsnittene.

Tabell 5. *Utsagnenes vanskelighetsgrad, infit MNSQ, outfit MNSQ og PM Corr etter første analyse*

Utsagn	Vanskelighetsgrad	Infit MNSQ	Outfit MNSQ	PM Corr.
4	1,01	1,22	1,26	0,53
3	0,90	1,18	1,08	0,51
1	0,80	1,03	0,94	0,60
2	0,64	1,05	1,11	0,51
9	0,60	1,94	3,30	-0,01
6	0,37	1,02	0,97	0,59
7	0,30	0,95	0,99	0,58
8	0,30	0,84	0,89	0,61
16	0,12	0,97	0,98	0,48
10	0,07	0,99	0,95	0,57
14	-0,15	0,93	0,93	0,47
15	-0,17	0,75	0,72	0,70
11	-0,24	1,08	1,09	0,56
13	-0,24	0,88	0,86	0,59
5	-0,36	0,82	0,86	0,56
12	-0,40	0,94	0,91	0,63
17	-0,46	0,94	0,93	0,55
18	-0,77	0,95	1,01	0,51
20	-1,15	0,85	0,86	0,57
19	-1,16	0,82	0,85	0,63

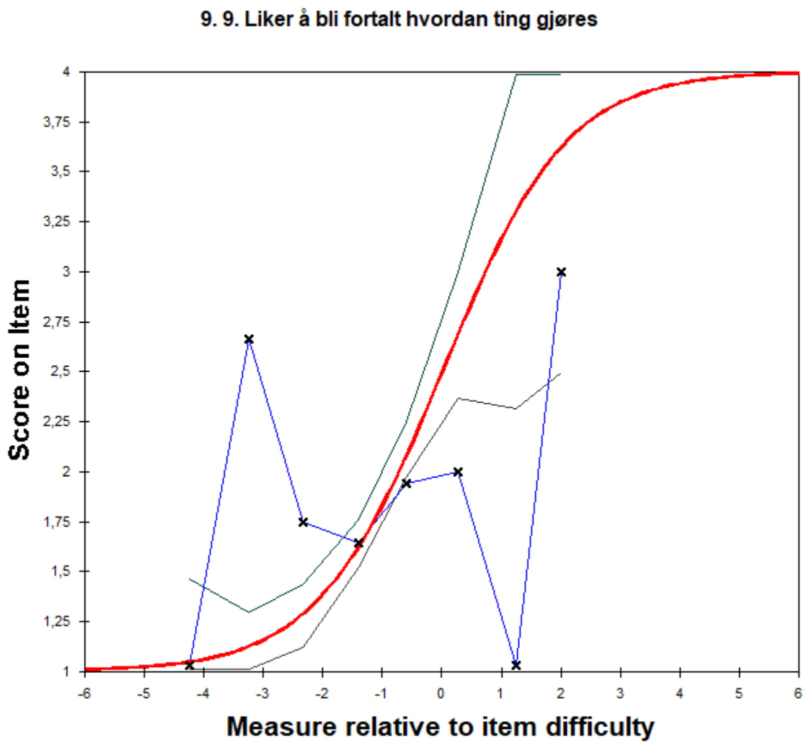
4.1.1 Innholdsaspektet

Validering av innholdsaspektet skal som nevnt i metodekapittelet kontrollere at utsagnene er basert på relevant og representativt innhold for at man skal kunne si at instrumentet faktisk måler det det er ment å måle (Wolfe & Smith, 2007). Innholdsaspektet inkluderer også at det gjøres en teknisk undersøkelse av kvaliteten til utsagnene for å sikre at respondentene forstår og har relativt lik forståelse av utsagnene. Elever som ikke forstår utsagnet på samme måte som de fleste andre, vil kunne svare noe helt annet enn det som er forventet ut fra målet eleven har på MI. Det samme gjelder dersom utsagnet ikke handler om det samme som de andre utsagnene og derfor ikke bidrar til å måle det samme, altså at utsagnet er i en annen dimensjon.

Kaspersen (2018) har tidligere validert instrumentet teoretisk på grunnlag av litteratur om dyp (Entwistle et al., 2013), konseptuell (Hiebert, 1986) og relasjonell (Skemp, 1976) forståelse i matematikk. Jeg gjør ikke noen egen analyse her, men vil i diskusjonskapittelet knytte innhold i utsagnene opp mot nevnt teori i tillegg til ulike definisjoner av MI.

For teknisk kontroll av om respondentene var enige om hva utsagnene betydde, sjekket jeg infit MNSQ og outfit MNSQ for utsagnene, samt Point Measure Correlations (PM corr.) (se Tabell 5).

Kritisk intervall for infit MNSQ og outfit MNSQ er mellom 1,3 og 0,7 (Bond & Fox, 2015, s. 269). Utsagn 9 hadde høy infit MNSQ og outfit MNSQ (Tabell 5), noe som kunne tyde på at det var mange uventede responser på utsagnet. PM corr. bør være over 0,4 (Wolfe & Smith, 2007). For utsagn 9 var PM corr. −0,01 mellom elevenes mål på MI og hva de svarte på dette utsagnet. ICC-kurven til utsagnet (Figur 5) viser svarmønsteret grafisk. På x-aksen ser vi elevenes mål på MI relativt til målet på utsagnet, og på y-aksen ser vi hva elevene svarte på utsagnet. Den blå kurven viser dermed sammenhengen mellom elevmål og hva de svarte på utsagnet.

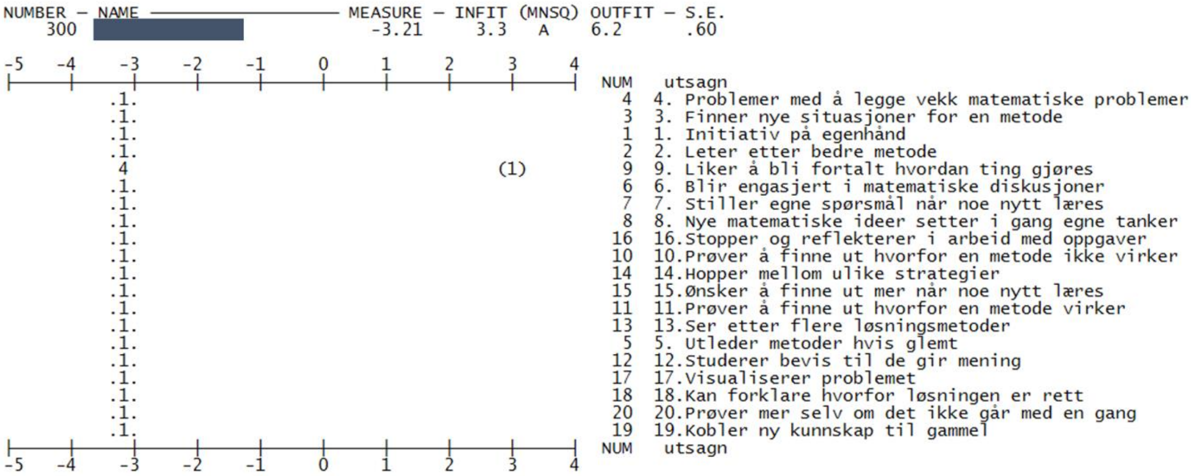


Figur 5. *ICC-kurve for utsagn 9*

Rød graf (jevn S-form) viser forventet respons, blå graf (med kryss) viser observert respons og sorte linjer (i ytterkant) markerer 95 % konfidensintervall.

ICC-kurven bekrefter mistanken om at det ikke er noen klar sammenheng mellom hva elevene svarer på dette utsagnet og hvilket mål de har.

Hva gjorde utsagn 9 så spesielt? Utsagnet lyder slik: «Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre». Utsagnet var det eneste som var kodet omvendt, slik at svaralternativet «alltid/nesten alltid» var det alternativet som ga lavest skår. Alle de andre utsagnene hadde motsatt kodede svaralternativer. Noen elever med lav MI krysset kanskje av på «aldri» på nesten alle utsagn slik at det var lett å svare «aldri» også på utsagn 9. Nærmere analyse av responsmønsteret til elever med uventet respons på utsagn 9 var typisk slik. Figur 6 viser en elev (elev nummer 300) som har svart «aldri» på alle utsagn, inkludert utsagn 9. På elevens spørreskjema var alle kryss satt i ruta lengst til venstre.



Figur 6. *Svarmønster for elev 300*

Observert respons vises med punktum foran og bak, observert (men svært uventet) respons i parentes, og forventet respons uten parentes og uten punktum.

Både infit MNSQ og outfit MNSQ for utsagn 9 var over kritisk grense. Høy infit MNSQ tydet på at det var en systematisk feil ved utsagnet. Dette ble bekreftet av lav PM corr. og kvalitativ kontroll av utsagnet som bekreftet at utsagnet skilte seg ut ved at det hadde motsatt kodede svaralternativer. På bakgrunn av disse analysene valgte jeg å fjerne utsagn 9 før videre validering av instrumentet.

PM corr., infit MNSQ og outfit MNSQ ble på nytt kontrollert etter en analyse uten utsagn 9 (Tabell 6).

Tabell 6. *Utsagnenes vanskelighetsgrad, infit MNSQ, outfit MNSQ og PM Corr. uten utsagn 9*

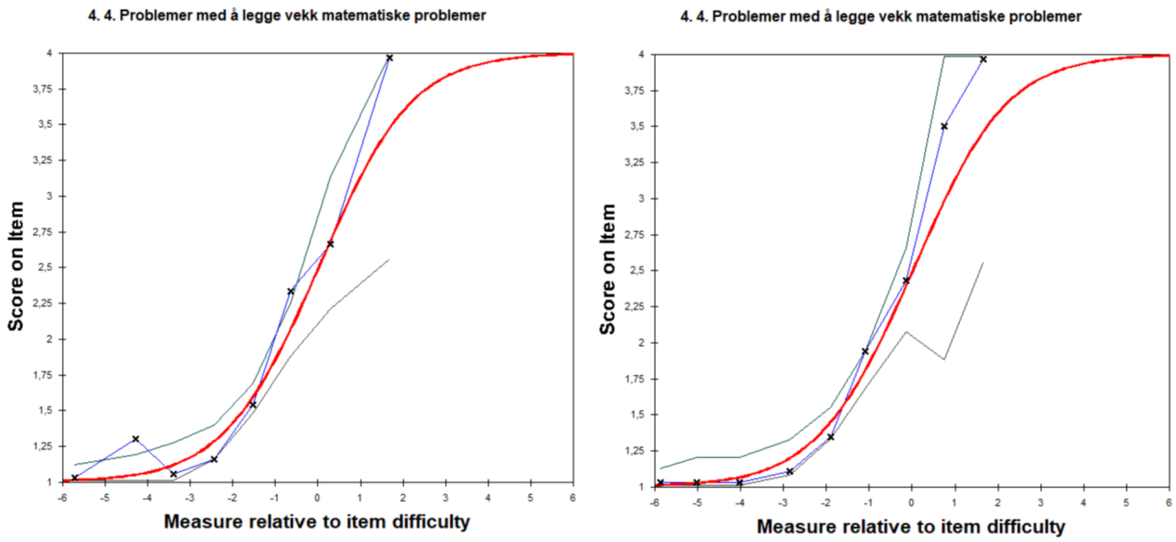
Utsagn	Vanskelighetsgrad	Infit MNSQ	Outfit MNSQ	PM Corr.
4	1,10	1,29	1,48	0,50
3	0,99	1,24	1,14	0,49
1	0,89	1,07	0,96	0,58
2	0,71	1,11	1,22	0,49
6	0,43	1,08	1,02	0,57
7	0,35	1,02	1,20	0,56
8	0,34	0,88	0,93	0,59
16	0,16	1,04	1,07	0,47
10	0,11	1,05	1,00	0,56
14	-0,12	1,01	1,02	0,48
15	-0,15	0,77	0,74	0,68
11	-0,22	1,14	1,21	0,56
13	-0,22	0,92	0,90	0,58
5	-0,35	0,86	0,92	0,56
12	-0,40	0,98	0,94	0,62
17	-0,45	0,99	0,97	0,56
18	-0,79	1,02	1,13	0,51
20	-1,18	0,90	0,92	0,58
19	-1,20	0,85	0,90	0,64

Utsagn 4 fikk nå infit MNSQ innenfor kritisk intervall på 0,7-1,3. Outfit MNSQ var derimot over kritisk grense, noe som kunne tyde på at det var tilfeldige feil knyttet til utsagnet. ICC-kurven (Figur 7) viste også at det var uregelmessigheter i området med lave mål på MI, noe som tyder på at elever med lav MI var enige i utsagnet selv om Rasch-modellen hadde forutsagt at de ville være uenige.

Tabell 7. *Mest uventede responser*

Elev	Utsagn	Observert respons	Forventet respons
323	4. Problemer med å legge vekk matematiske problemer	4	1,06
302	7. Stiller egne spørsmål når noe nytt læres	4	1,10
298	11. Prøver å finne ut hvorfor en metode ikke virker	4	1,16
340	2. Leter etter bedre metode	3	1,09
331	18. Kan forklare hvorfor løsningen er rett	2	3,89

Jeg fant at elev 323 hadde svart svært uventet på utsagn 4 (Tabell 7). Der Rasch-modellen hadde forutsatt at eleven skulle svare 1,06, var observert respons 4. Jeg analyserte datamaterialet uten at elev 323 var med i analysen. Figur 7 viser ICC-kurven for utsagnet med og uten elev 323.



Figur 7. *ICC-kurve for utsagn 4 med (venstre) og uten (høyre) elev (323) med uventet respons på utsagnet.*

Alle utsagn, inkludert utsagn 4, fikk nå PM corr., infit MNSQ og outfit MNSQ innenfor kritiske intervaller.

Jeg valgte å tilbakeføre elev 323 fordi man ikke vet hvorfor denne eleven svarte slik vedkommende gjorde. Vi kan ikke anta at eleven *ikke* mente at den synes det er vanskelig å legge fra seg matematiske oppgaver. For å finne ut om det ville ha noen betydning at elev 323 var inkludert, analyserte jeg datamaterialet med og uten eleven inkludert, for å kunne sammenligne elevmålene i de to analysene, og se om det var signifikant forskjell på målene. Dette er beskrevet i neste underkapittel, da jeg måtte ta ut flere elever fra datamaterialet på grunn av dårlig «fit». Foreløpig antagelse var at en enkeltelevs respons sannsynligvis ikke ville påvirke instrumentet i så stor grad at det ville endret elevmålene i analysen.

4.1.2 Det substansielle aspektet

For å undersøke det substansielle aspektet har jeg analysert svaralternativene, og jeg har undersøkt *person fit*. Begge analysene gjøres for å kontrollere om svaralternativene er tydelige, og at elevene legger den samme betydningen i av hver av dem (Wolfe & Smith, 2007).

Tabell 8 gir informasjon som bidrar til å svare på noen av kontrollspørsmålene fra Linacres (2002) retningslinjer for utforming av spørreskjema med Likert-skala. Videre følger en gjennomgang av hvilke resultater jeg fikk da jeg analyserte svaralternativene etter seks av Linacres (2002) retningslinjer.

Tabell 8. *Sammendrag av struktur for svaralternativene*

Svaralternativ	Antall responser	Responser (%)	Gj.snitt mål på respondent	infit MNSQ	outfit MNSQ	Andrich threshold	Kategorimål
1	1761	27	-1,64	0,94	0,96	-	(-2,59)
2	2277	35	-0,56	0,97	0,91	-1,33	-0,75
3	1659	26	0,12	0,97	0,99	0,06	0,78
4	728	11	0,73	1,13	1,30	1,27	(2,55)
Missing	320	5					

1. Er det minst ti observasjoner av hver svarkategori?

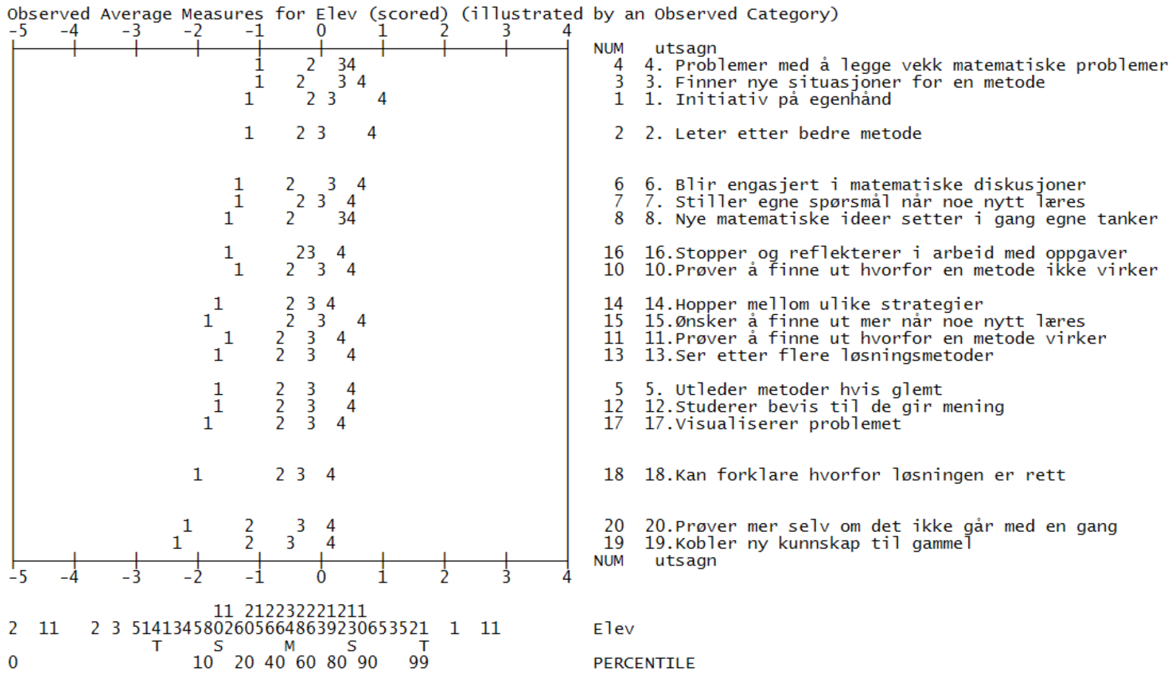
Andre kolonne i Tabell 8 viser at det er godt over ti observasjoner for hvert svaralternativ.

2. Er det jevn distribusjon av observasjoner mellom kategoriene?

Andre kolonne i Tabell 8 viser at det er ganske jevn fordeling av observasjoner for hver svarkategori. Kategori 4 (alltid/nesten alltid) skiller seg ut med færre observasjoner enn for de andre. Dette kan skyldes at det oppfattes som et litt mer ekstremt alternativ enn de andre. Det viktigste å kontrollere på dette spørsmålet er at det ikke veksler mellom mange og få observasjoner flere ganger for tilstøtende alternativ, et såkalt berg-og-dalbane-mønster (Linacre, 2002), noe som ikke er tilfelle for svaralternativene i denne undersøkelsen.

3. Øker gjennomsnittsmålet til deltakerne med verdiene til svaralternativene?

Fjerde kolonne i Tabell 8 viser gjennomsnittsmålene til respondentene som svarte i hver kategori, og at dette øker med verdien av svaralternativet. Figur 8 viser gjennomsnittsmålet til respondentene i hver svarkategori for hvert utsagn. Her ser vi at målet til respondenten øker med verdiene til svarkategoriene for hvert utsagn, og at dette gjelder for alle utsagn i denne undersøkelsen.



Figur 8. *Gjennomsnittsmål (x-akse) til elever som svarer de ulike svaralternativene for hvert av utsagnene (y-akse).*

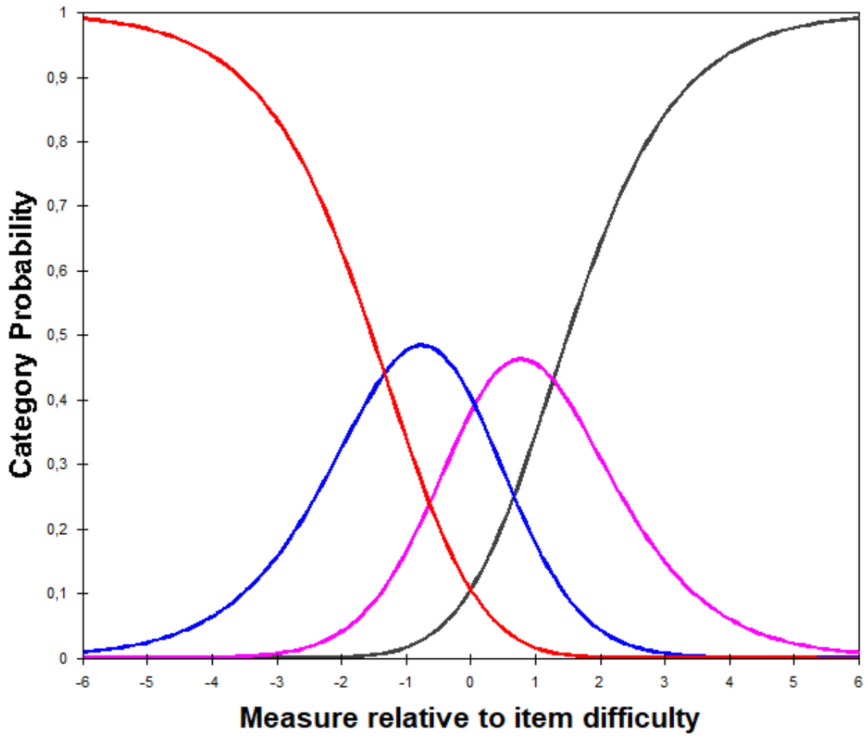
På Figur 8 kan vi ane at det er lengre avstand mellom svarkategori 1 og 2 enn det er mellom 2 og 3. Dette ser ut til å gjelde for nesten alle utsagn.

4. Er det noen svaralternativer som har mye uventet respons?

Jeg ser på outfit MNSQ i Tabell 8 for hvert svaralternativ og kontrollerer at denne er under 2 (Linacre, 2002). Alle svaralternativene har outfit MNSQ som ligger godt under 2.

5. Er alle svaralternativene mest sannsynlige for et intervall av personmål relativt til utsagnenes vanskelighetsgrad?

På Figur 9 ser vi elevenes mål relativt til utsagnets vanskelighetsgrad på x-aksen, mens sannsynligheten for å svare det aktuelle svaralternativet er på y-aksen. Vi ser at alle svaralternativene har en egen topp der det er mest sannsynlig at dette blir valgt.



Figur 9. *Category Probability Curve (CPC)*

Sannsynligheten for hvert svaralternativ (y-aksen) for elever med mål relativt til utsagnets vanskelighetsgrad (x-aksen). Rød graf (lengst til venstre) for svaralternativ 1, blå graf (nummer to fra venstre) for svaralternativ 2, lilla graf (nummer tre fra venstre) for svaralternativ 3 og sort graf (nummer fire fra venstre) for svaralternativ 4.

6. Er det mer enn 1,1 logit mellom tersklene der det går over fra å være mest sannsynlig å svare ett alternativ til det går over til å være mest sannsynlig å svare det neste alternativet?

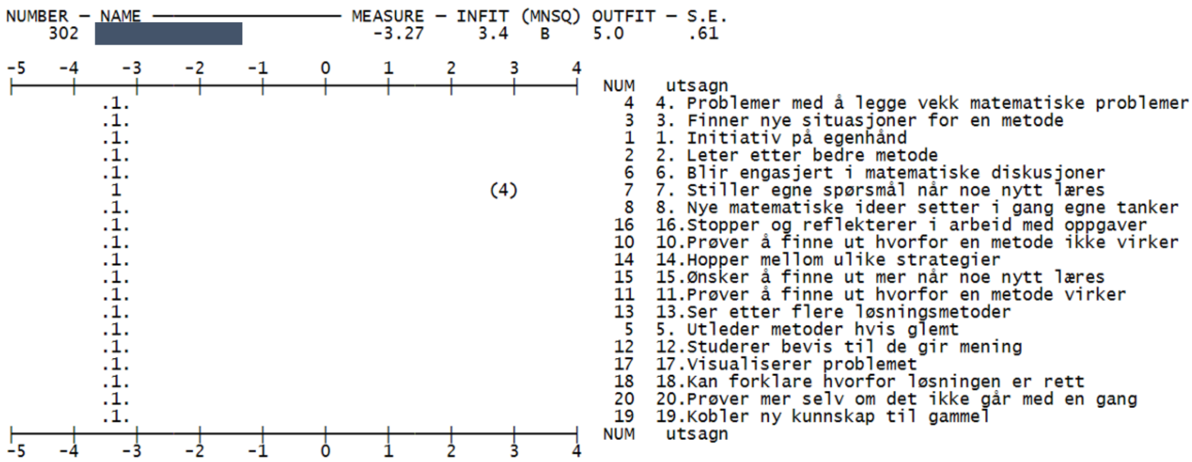
Overgangene der grafene krysser hverandre på Figur 9, er tersklene hvor det går fra å være mest sannsynlig å svare det ene alternativet til å bli mer sannsynlig å svare det andre alternativet. Avstanden mellom disse overgangene bør være over 1,1 for en undersøkelse med fire svaralternativer (Wolfe & Smith, 2007). I syvende kolonne (Andrich threshold) i Tabell 8 ser vi nøyaktige mål på hvor disse tersklene befinner seg. Jeg regnet ut avstanden mellom disse og fant at det var 1,39 logit mellom tersklene til alternativ 1 og 2, og 2 og 3, mens det var 1,21 logit mellom tersklene til alternativ 2 og 3 og 3 og 4. Det er god nok avstand mellom tersklene til svarkategoriene.

En annen analyse som kan gjøres for å kontrollere for det substansielle aspektet, er å se på *person fit*. Høy infit MNSQ eller outfit MNSQ (over 2,0) for en elev kan tyde på at denne har et uvanlig svarmønster (Wolfe & Smith, 2007). Det var 17 elever som hadde infit MNSQ eller outfit MNSQ over 2,0.

Tabell 9. *Elever sortert etter misfit*

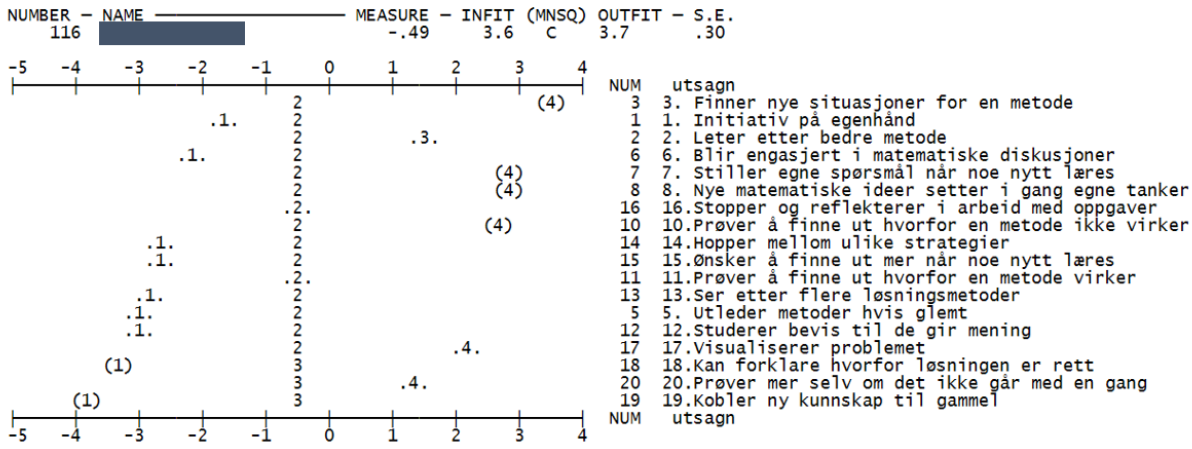
Elevnummer	Rasch-mål	infit MNSQ	outfit MNSQ
323	-2,95	2,88	7,56
302	-3,27	3,36	4,98
116	-0,49	3,60	3,73
298	-3,27	3,21	2,96
170	0,06	2,38	2,53
340	-2,95	1,51	2,50
40	-0,39	2,44	2,43
240	-0,72	2,36	2,37
331	2,65	1,40	2,27
229	0,48	2,20	2,25
355	-0,23	2,14	2,08
319	-1,39	2,04	2,12
276	0,31	2,11	2,09
133	-2,95	1,69	2,10
163	0,14	2,08	2,06
37	-0,72	2,02	2,07
253	-1,18	1,96	2,03

Høy person-misfit (høy infit MNSQ og/eller outfit MNSQ) kan skyldes at eleven ikke har forstått utsagnet, eller at eleven har en uvanlig kombinasjon av egenskaper (Linacre, 2002). For å få en idé om hva høy misfit kan skyldes, så jeg nærmere på noen elever med høy misfit (Tabell 9). Elev 323 var kontrollert tidligere, og jeg fant at denne eleven hadde svart svært uventet på utsagn 4. Elev 302 og 116 ga innblikk i andre mulige årsaker (Figur 10 og Figur 11).



Figur 10. *Svarmønster for elev 302*

Observert respons vises med punktum foran og bak, observert (men svært uventet) respons i parentes, og forventet respons uten parentes og uten punktum.

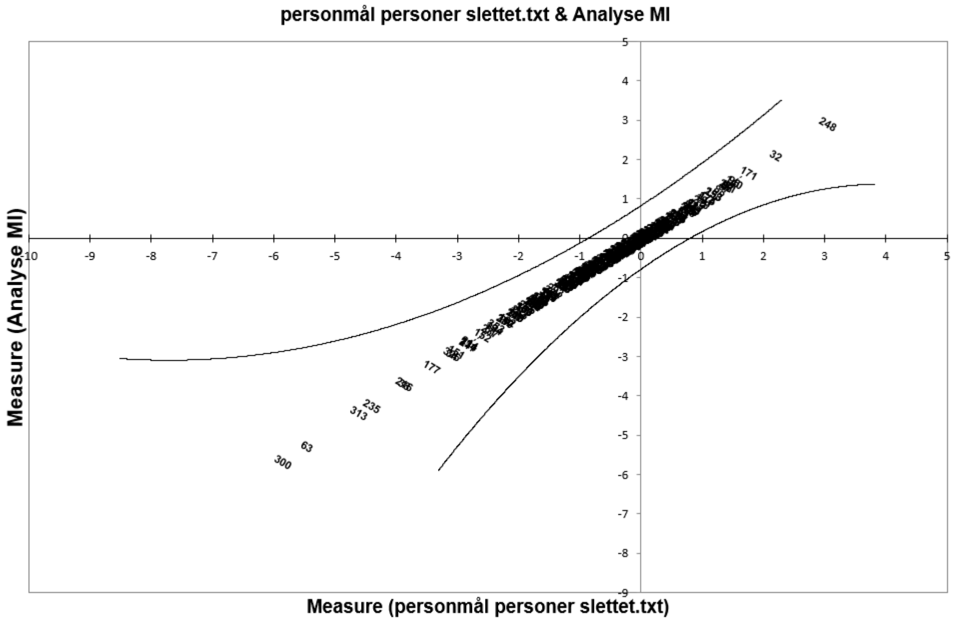


Figur 11. *Svarmønster for elev 116*

Observert respons vises med punktum foran og bak, observert (men svært uventet) respons i parentes, og forventet respons uten parentes og uten punktum.

Elev 302 hadde et tilsvarende mønster som elev 323 og svarte 1 på alle utsagn bortsett fra ett, som i dette tilfellet var utsagn 7; «Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.». Elev 116 hadde et litt annet mønster. Eleven hadde en tendens til å svare i de ekstreme kategoriene, «aldri/nesten aldri» eller «alltid/nesten alltid» og med nesten like mange i hver ende. Dette førte til et mål på MI som lå litt midt imellom. Responsene eleven ga ble derfor nesten alltid et stykke unna det Rasch-modellen forutså, og eleven fikk høy misfit.

Jeg ønsket å teste om det at elever med høy misfit var med i analysen, ville påvirke de andre elevenes mål på MI. Jeg gjorde derfor en Rasch-analyse der de 17 elevene med infit MNSQ og outfit MNSQ over 2,0 ikke var med. Elevenes mål på MI fra denne analysen ble så sammenlignet med elevenes mål på MI der de 17 elevene var med. Plottet (Figur 12) viser om elevmålene ble signifikant forskjellige i analysene med og uten elever med høy person-misfit inkludert. Mål fra analysen uten elever med høy misfit finnes på x-aksen, og mål fra analyse med alle elever inkludert er på y-aksen. De to buede linjene i plottet viser 95 % konfidensintervall, og alle elevene legger seg innenfor her. Dette betyr at forskjellen ikke er signifikant (p < 0,05 og r = 1,00), og jeg velger å beholde elevene i analysen videre.

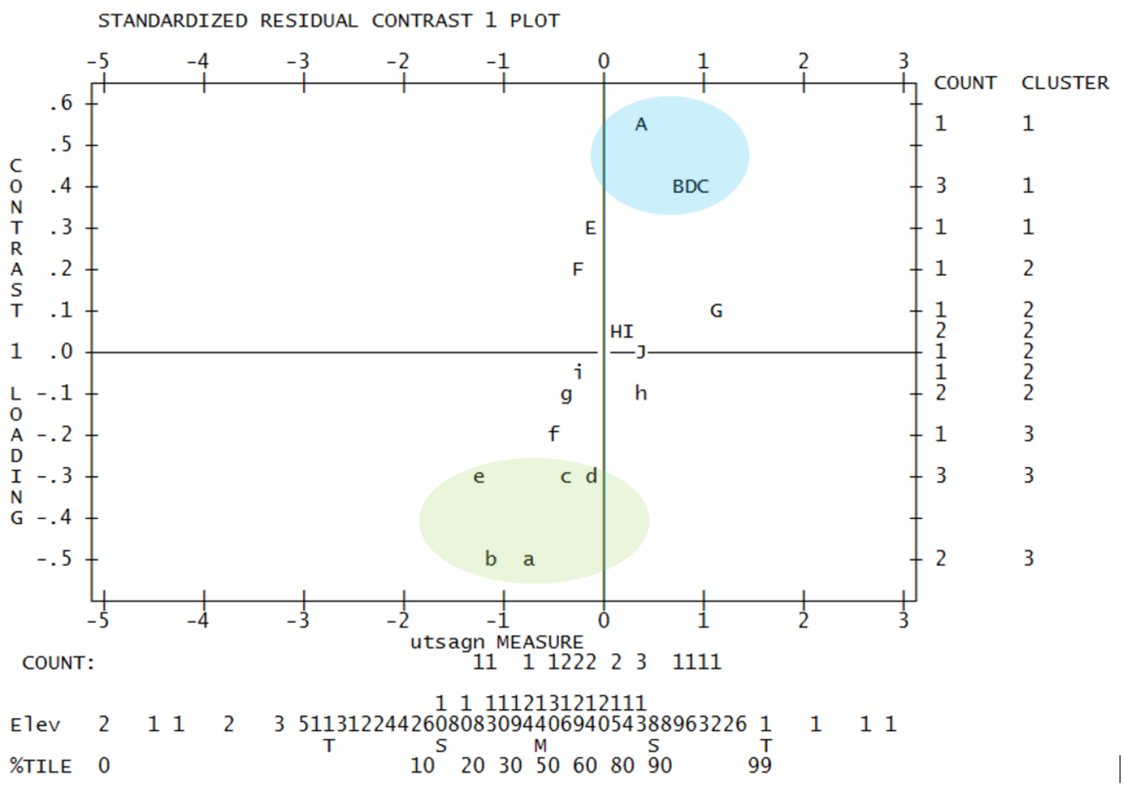


Figur 12. *Personmål med og uten elever med høy infit MNSQ og outfit MNSQ*

Elev 323 som ble omtalt i kapittel 4.1.1 var én av elevene som ble fjernet i analysen over. Som ventet hadde inkludering av eleven i datamaterialet ingen signifikant påvirkning av elevmålene selv om eleven hadde uventet respons på utsagn 4.

4.1.3 Det struktureelle aspektet

En PCA viste at instrumentet hadde 1,7988 *eigenvalue* for «Unexpalined variance in 1st contrast». Når *eigenvalue* er under 2,0 ansees instrumentet som endimensjonalt (Linacre, 2012), noe som var tilfelle for dette instrumentet. Selv om modellen ansees som endimensjonal, kan utsagnene likevel ha underdimensjoner som eventuelt kan undersøkes kvalitativt. Figur 13 viser målene til utsagnene på x-aksen og hvilken ende av dimensjonen utsagnene tilhører på y-aksen.



Figur 13. *Standardized Residual Contrast 1 Plot*

Utsagnenes ladning i dimensjonen (y-aksen) mot utsagnets vanskelighetgrad (x-aksen). «Klaser» av utsagn som legger seg omtrent på samme sted i dimensjonens ytterkanter er markert.

Det kan se ut som utsagn som måler elever i den høye enden av skalaen ligger i den ene enden av dimensjonen, mens utsagn som måler i den nedre del av skalaen tilhører en annen ende av dimensjonen. Tabell 10 viser hvilke utsagn som la seg øverst i dimensjonen, og

Tabell 11 viser hvilke utsagn som la seg nederst i dimensjonen.

Tabell 10. *Utsagn sortert etter størst ladning i 1. dimensjon (PCA)*

Ladning	Vanskelighets-grad	Navn	Utsagn
0,55	0,35	A	7: Stiller egne spørsmål når nytt læres
0,41	0,71	B	2: Leter etter bedre metode
0,41	0,99	C	3: Finner nye situasjoner for metode
0,40	0,89	D	1: Tar initiativ til å lære på egenhånd
0,29	-0,15	E	15: Ønsker å finne ut mer når noe nytt læres

Tabell 11. *Utsagn sortert etter minste ladning i 1. dimensjon (PCA)*

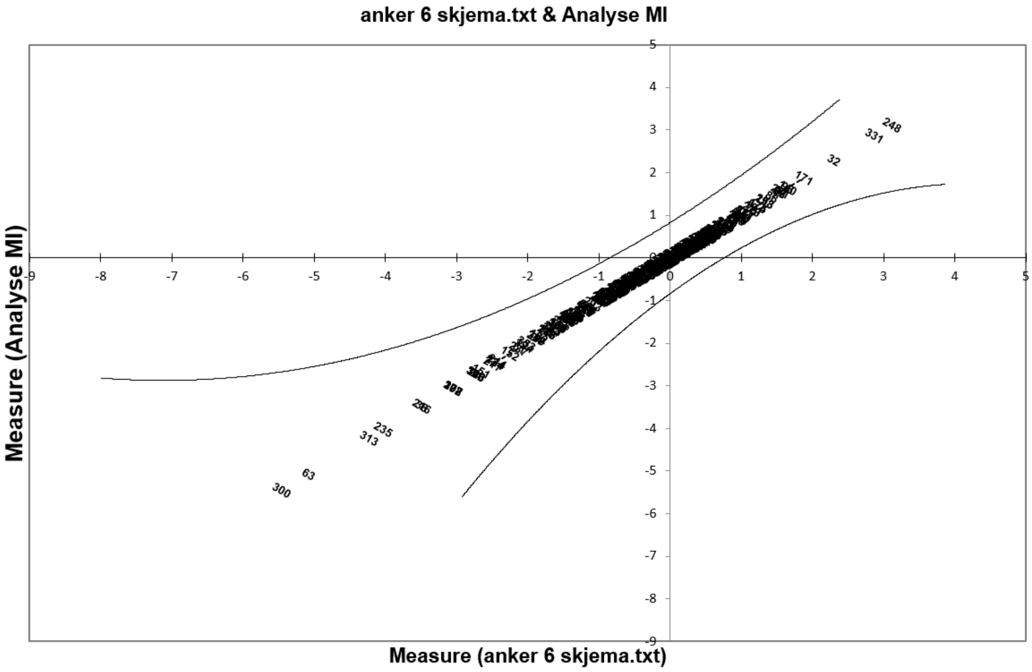
Ladning	Vanskelighets-grad	Navn	Utsagn
-0,51	-0,79	a	18: Kan forklare at løsningen er rett
-0,50	-1,18	b	20: Prøver mer selv om det ikke går med en gang
-0,32	-0,35	c	5: Utleder metode selv hvis jeg har glemt den
-0,30	-0,12	d	14: Hopper mellom ulike strategier
-0,29	-1,20	e	19: Kobler ny kunnskap til gammel

Utsagnene som la seg øverst i dimensjonen, og som viste seg å også være kjennetegnende for elever med høy MI, var utsagn 7: «Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med», utsagn 2: «Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode», og utsagn 3: «Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker». Dette er utsagn som kanskje uttrykker overskudd hos eleven; at elevene gjør en ekstra innsats etter at de har lært og forstått det de skal, og kan vitne om indre motivasjon for å lære. I den andre enden av dimensjonen finner vi utsagn som uttrykker forhold som til en viss grad kreves av alle elever. De to utsagnene som lå nederst i dimensjonen, var utsagn 18: «Jeg kan forklare hvorfor løsningen min er rett» og utsagn 20: «Jeg fortsetter å prøve meg fram selv om jeg ikke får det til med en gang».

4.1.4 Generaliserbarhetsaspektet

DIF-analyser mellom undergrupper i datamaterialet ble brukt for å kontrollere om instrumentet var invariant i konteksten det ble brukt. Datamaterialet ble delt opp på tre ulike måter: (1) Skjema A/B/C/D, (2) gutt/jente og (3) yrkesfag/studieforberedende. På denne måten ble det gjort 8 ulike DIF-analyser for alle 19 utsagn, noe som tilsvarer 152 signifikanstester for om utsagnene legger seg på ulike steder på skalaen i de ulike gruppene. Jeg korrigerer ikke signifikansnivået med Bonferronis korreksjon, og beholder kriteriene på $p < 0,05$ og DIF > 0,64.

Av de 152 DIF-analysene fant jeg kun én DIF som var signifikant og større enn 0,64 logit. Utsagn 6 hadde i instrument A en vanskelighetsgrad på 0,06 logit, mens det i instrument B hadde en vanskelighetsgrad på 0,72 logit. Differansen var 0,66 logit ($p = 0,0025$). For å finne ut om denne differansen hadde signifikant betydning for elevmålene, gjorde jeg to Rasch-analyser av alle elevene der jeg først låste utsagn 6 til 0,06, og deretter til 0,72. Et punktdiagram (Figur 14) viser punkter for hver elev relativt til de to målingene med 95 % konfidensintervall markert som buede linjer.



Figur 14. *Elevmål fra analyser med utsagn 6 «ankret» to ulike steder på skalaen. 95 % konfidensintervall er markert.*

Alle målene lå innenfor 95 % konfidensintervall (p < 0,05 og r = 1,00). Jeg konkluderte med at jeg kunne godta instrumentet slik det var, selv om det ble litt ulikt hvor utsagn 6 la seg ettersom hvilket skjema som var brukt. Jeg fant ingen kvalitativ forklaring på hvorfor utsagn 6 fikk så ulik vanskelighetsgrad i skjema A og B. Utsagnet lå omtrent på midten i begge skjemaer.

Reliabiliteten for elever og utsagn var henholdsvis 0,86 og 0,99. Dette er definert som god reliabilitet (Cohen et al., 2018), og kan tyde på at instrumentet har god spredning og jevn fordeling av utsagnene som danner skalaen elevene måles ette.

Wright map (Figur 15) gir en visuell kontroll av hvordan utsagnene fordelte seg langs skalaen. Utsagn der få elever sa seg enige legger seg høyt på skalaen, mens utsagn som mange var enige i legger seg lavt på skalaen. Elevenes mål er det samme som vanskelighetsgraden til utsagnet som det er 50 % sannsynlighet for at de vil være enige i.

```
MEASURE      Elev - MAP - utsagn
               <more>|<rare>
    3                +
                  .  |
                  .  |
                     |
                     |
                     |
                  .  |
    2                +
                     |
                  .  |
                    T|
                  .  |
                 ##  |T
                  .  |  4. Problemer med
    1             #  +  3. Finner nye sit
                .##  |  1. Initiativ på e
                ###  |S 2. Leter etter be
                ###  |
              .#### S|  6. Blir engasjert
             ######  |  7. Stiller egne s  8. Nye matematisk
             .#####  |  10.Prøver å finne  16.Stopper og ref
    0       #######  +M
           .#######  |  14.Hopper mellom   15.Ønsker å finne
           .#######  |  11.Prøver å finne  13.Ser etter fler  5. Utleder metode
       .###########  |  12.Studerer bevis  17.Visualiserer p
            ####### M|
             .#####  |S
          .########  |  18.Kan forklare h
   -1         .####  +
            .######  |  19.Kobler ny kunn  20.Prøver mer sel
                  .  |T
               .###  |
               .### S|
                .##  |
                .##  |
   -2            .#  +
                 .#  |
                 .#  |
                  .  |
                    T|
                 .#  |
                  .  |
   -3            .#  +
                     |
                  #  |
                     |
                     |
                  .  |
                     |
   -4                +
                     |
                  .  |
                  .  |
                     |
                     |
                     |
   -5             .  +
               <less>|<freq>
 EACH "#" IS 3: EACH "." IS 1 TO 2
```

Figur 15. *Wright map*

Utsagn (høyre side) og elevmål (venstre side) på samme skala. «Vanskelige» utsagn ligger høyt på skalaen og «lette» utsagn ligger lavt på skalaen. Hver «#» representerer tre elever, og hver «.» representerer én til to elever.

Det er mange elevmål som ligger under det nederste utsagnet på skalaen. Ellers ser det ut til at utsagnene sprer seg jevnt utover i det området der de fleste elevmålene ligger. Noen flere utsagn i det nedre området kunne bedret målenøyaktigheten for lave elevmål, og derfor økt reliabiliteten til instrumentet (Linacre, 2006).

4.1.5 Sammendrag

Etter validering av data og justering av analyseinstrument ser det ut til at instrumentet egner seg for måling i gruppen det er validert for. Jeg har valgt å bruke elevmålene jeg fikk fra Rasch-analysen til å gjøre de videre analysene som kreves for å svare på det andre og det tredje forskningsspørsmålet mitt.

4.2 Hva er sammenhengen mellom MI og prestasjoner?

I dette avsnittet vil jeg beskrive resultater på analysene jeg har gjort for å besvare det andre og det tredje forskningsspørsmålet mitt som begge handler om å undersøke korrelasjonene mellom MI og prestasjoner. For å besvare det andre forskningsspørsmålet, «Hva er korrelasjonen mellom elevenes mål på MI og deres matematikkprestasjoner?», har jeg gjort to korrelasjonsanalyser. For å svare på det tredje forskningsspørsmålet, «Hvordan samsvarer korrelasjonen mellom MI og karakterer gitt av faglærer og korrelasjonen mellom MI og resultater på en anonym kartleggingsprøve?», har jeg testet om de to korrelasjonene jeg fikk var signifikant forskjellige. Før jeg beskriver resultatene fra korrelasjonsanalysene, vil jeg kort referere noen deskriptive funn fra datamaterialet.

4.2.1 Noen deskriptive funn fra datamaterialet

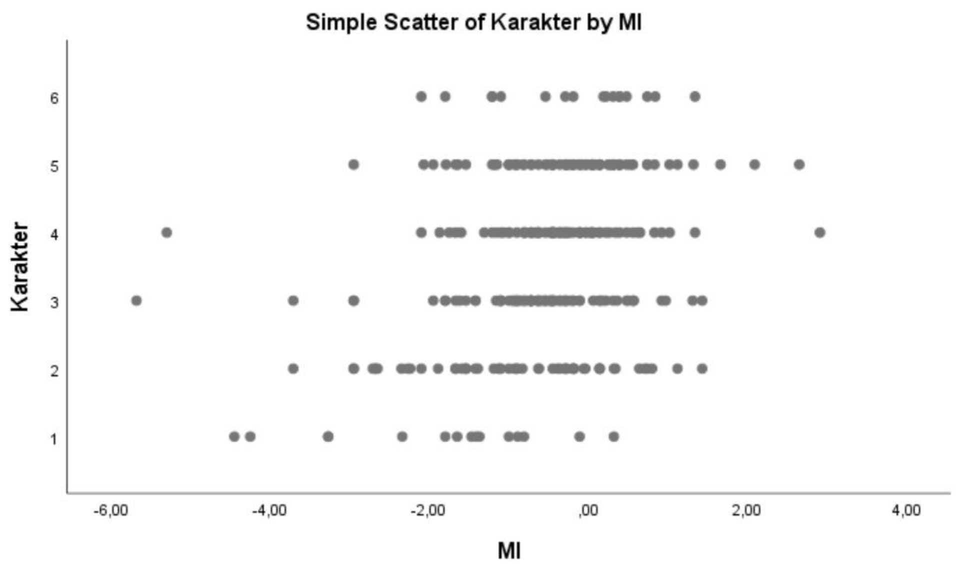
Elevene som var med i undersøkelsen fikk karakterer i ulike matematikkfag. Det høyeste faglige nivået har fagene R1 og R2. Av de 50 elevene med høyest MI, var det 10 stykker (20 %) som tok et av de teoretiske matematikkfagene 1T, R1 eller R2 (disse elevene utgjør 12 % av det totale utvalget). De to elevene med høyest MI tok begge faget R1. Blant de 50 elevene med lavest MI var det ingen elever som tok et teoretisk matematikkfag. Ingen av de 17 som fikk karakteren 6 fikk dette i et teoretisk matematikkfag, noe som kan tyde på at det er relativt vanskeligere å få en god karakter i et teoretisk matematikkfag enn i et praktisk matematikkfag. Kartleggeren tester derimot elevene på den samme prøven uavhengig av hvilket matematikkfag de tar. Blant de 50 elevene med høyest kartleggingsresultater var det 22 som tok et teoretisk matematikkfag (44 %).

Tabell 12. *Gjennomsnitt for MI, karakterer og resultater fra Kartleggeren, fordelt på fag*

Matematikkfag	Antall elever	Gjennomsnitt MI	Gjennomsnitt karakter	Gjennomsnitt Kartleggeren
1PY	143	-0,96	3,1	92
2PY	128	-0,40	3,9	95
1P	16	-0,04	4,0	110
2P	24	-0,61	4,0	102
1T	24	0,00	3,9	122
R1	13	0,23	3,6	131
R2	7	-0,09	3,9	144
Hele datamaterialet	355	-0,56	3,6	100

4.2.2 Korrelasjonsanalysene

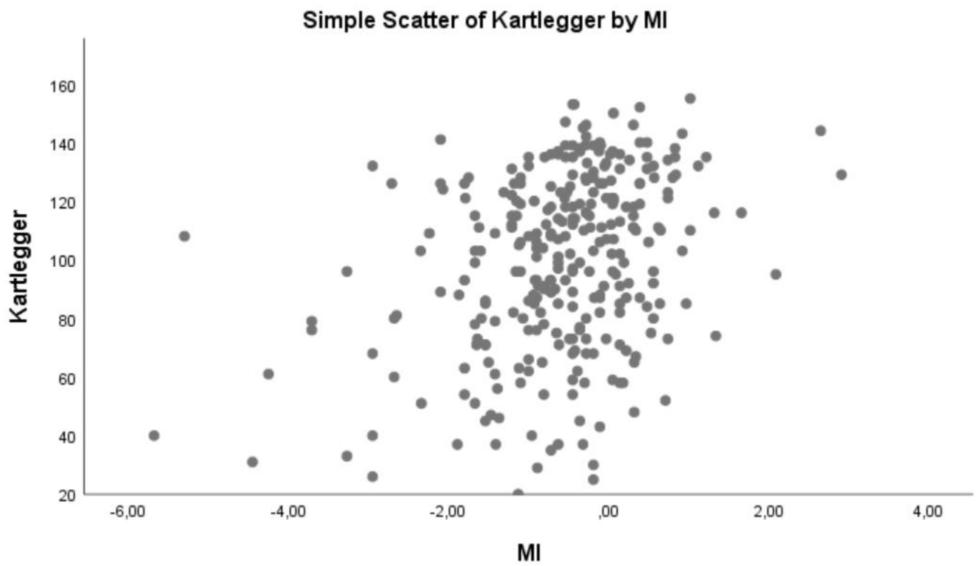
Punktdiagrammer ga et visuelt inntrykk av om det var sammenheng mellom variablene. Jeg valgte å ha MI på x-aksen både i analysen av sammenhengen mellom MI og karakterer og i analysen av sammenhengen mellom MI og kartleggingsresultater, selv om det ikke er gitt at MI er den uavhengige variabelen.



Figur 16. *Punktdiagram for sammenhengen mellom MI og karakterer. MI på x-aksen og karakter på y-aksen.*

Punktdiagrammet i Figur 16 viser sammenhengen mellom MI og karakterer. Det er flere punkter i det høye området for karakterer for elever som har høy MI enn det er for elever som har lav MI. Motsatt ser vi også at det er flere punkter i det lave området for karakterer for elever med lav MI enn for elever med høy MI. Dette kan tyde på at elever med høy MI i gjennomsnitt har høyere karakterer enn elever med lav MI.

Den samme tendensen kan vi se i Figur 17, som viser sammenhengen mellom MI og resultater på Kartleggeren.



Figur 17. *Punktdiagram for sammenhengen mellom MI og kartleggingsresultater. MI på x-aksen og kartleggingsresultater på y-aksen.*

Blant elever med middels MI ser det ut som resultatene på Kartleggeren varierer fra laveste til høyeste resultat. For elever med mer ekstreme verdier for MI ser det ut som det finnes en tendens der de med veldig lav MI har dårligere resultater på Kartleggeren enn de med veldig høy MI.

Jeg gjorde korrelasjonsanalyser i begge tilfeller for å se om det fantes signifikante sammenhenger. Korrelasjonen var signifikant ($p < 0,01$) i begge tilfeller, slik det vises i Tabell 13 og Tabell 14 fra analyseprogrammet SPSS.

Tabell 13. *Korrelasjon mellom MI og karakterer*

Correlations

		MI	Karakter
MI	Pearson Correlation	1	,322**
	Sig. (2-tailed)		,000
	N	355	309
Karakter	Pearson Correlation	,322**	1
	Sig. (2-tailed)	,000	
	N	309	309

**. Correlation is significant at the 0.01 level (2-tailed).

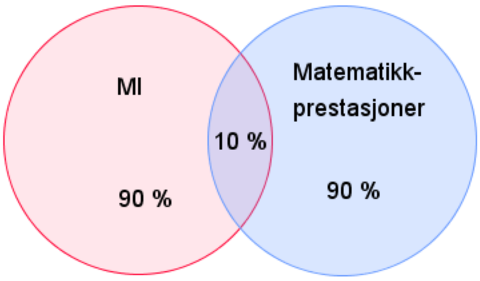
Tabell 14. *Korrelasjon mellom MI og kartleggingsresultater*

Correlations

		MI	Kartlegger
MI	Pearson Correlation	1	,333**
	Sig. (2-tailed)		,000
	N	355	296
Kartlegger	Pearson Correlation	,333**	1
	Sig. (2-tailed)	,000	
	N	296	296

**. Correlation is significant at the 0.01 level (2-tailed).

Korrelasjonene mellom MI og karakter og MI og kartlegger hadde korrelasjonskoeffisienter på henholdsvis $r = 0,32$ og $r = 0,33$. Hvor mye av variansen denne effekten forklarer, kan beregnes ved å kvadrere r. Effekten forklarer i disse tilfellene 10,2 % og 10,9 % av variansen, noe som betegnes som medium effekt (Field, 2013).



Figur 18. *MI og prestasjoner deler ca 10 % av variansen.*

4.2.3 Sammenligning av korrelasjoner

For å svare på det tredje og siste forskningsspørsmålet måtte jeg finne ut om det var signifikant forskjell på korrelasjonen mellom MI og karakter og MI og kartleggingsresultater. Jeg sammenlignet *z*-skårene til begge koeffisienter for å se hvor stor avstand det var mellom dem, for så å sjekke om denne avstanden var signifikant (Field, 2013, s. 286). Det var ingen signifikant forskjell på korrelasjonene (p = 0,89).

4.2.4 Sammendrag

Jeg fant signifikant korrelasjon med medium effekt både mellom MI og karakterer (r = 0,32) og mellom MI og kartleggingsresultater (r = 0,33). Svaret på mitt andre forskningsspørsmål blir derfor at det finnes en moderat sammenheng mellom MI og prestasjoner i matematikk.

Jeg fant ikke signifikant forskjell mellom de to korrelasjonskoeffisientene til henholdsvis MI og karakterer og MI og kartleggingsresultater (p = 0,89). Svaret på det tredje forskningsspørsmålet mitt er derfor at det i denne undersøkelsen ikke var noen forskjell i hvor stor korrelasjonen var mellom MI og karakterer som lærere satte, og MI og resultater fra en anonym prøve.

5 Diskusjon

Formålet med denne studien har vært å se om det er sammenheng mellom elevers matematiske identitet og deres prestasjoner i matematikk. Forskningsspørsmålene jeg har jobbet med er:

1. Er det mulig å måle MI hos elever på videregående skole?
2. Hva er korrelasjonen mellom elevenes mål på MI og deres matematikkprestasjoner?
3. Hvordan samsvarer korrelasjonen mellom MI og karakterer gitt av faglærer og korrelasjonen mellom MI og resultater på en anonym kartleggingsprøve?

Det første spørsmålet handler om kvaliteten til instrumentet og om Rasch-modellen er egnet for gruppen jeg undersøkte. Jeg har konkludert med at MI lar seg måle, og jeg vil diskutere metoden og instrumentet i kapittel 5.1.

Det andre spørsmålet handler om hvorvidt det er sammenheng mellom elevenes mål på MI og deres prestasjoner. Jeg fant korrelasjoner på r = 0,32 og r = 0,33. Begge korrelasjonene var signifikante (p < 0,05). Disse resultatene diskuterer jeg nærmere i kapittel 5.3. I kapittel 5.3. vil jeg også se på hva tidligere forskning sier om sammenheng mellom prestasjoner og variabler som er sammenlignbare med MI.

Det tredje og siste spørsmålet har jeg svart på ved å sammenligne korrelasjonene fra de to analysene for å se om de er signifikant forskjellige. Det var ingen signifikant forskjell på de to korrelasjonskoeffisientene. Jeg vil i kapittel 5.4 se på resultatet i lys av tidligere forskning.

Kapittelet inneholder til slutt en diskusjon om instrumentets og undersøkelsens aktualitet i forbindelse med ny læreplan, implikasjoner for undervisning, og forslag til videre forskning.

5.1 Er instrumentet egnet til måling av MI?

Resultatene fra denne undersøkelsen har vist at det er mulig å måle MI hos elever i videregående skole. Siden det ikke er noen felles forståelse av hvordan vi definerer MI, er det viktig å ha med en forklaring av begrepet slik det er definert for instrumentet som er brukt. Dette er i tråd med den åttende retningslinjen for validering av data, *tolkningsaspektet* (Wolfe & Smith, 2007), som påpeker at det må kobles kvalitativ mening til de kvantitative målene testen gir.

MI var i denne studien definert som «den relative posisjonen mellom personer og den sosiale strukturen av å være matematisk i den aktiviteten man deltar i» (Kaspersen, 2018). Den sosiale strukturen dannes av karakteristikker som beskriver det å «være matematisk». Karakteristikkene var igjen skrevet om til utsagn som man kunne være mer eller mindre enige i. Det er derfor innholdet i utsagnene som i prinsippet operasjonaliserer hva det vil si å «være matematisk» (ha høy MI). Utsagnene bygger på teori om dybdelæring, og om relasjonell og konseptuell læring, samt om hvordan noen personer i matematiske miljøer beskriver hva de mener kjennetegner mennesker som er «matematiske» (Kaspersen, 2018).

De fleste utsagnene handler om måter å jobbe eller tenke på når du jobber med matematiske oppgaver, som «Jeg fortsetter å prøve meg fram selv om jeg ikke får det til med en gang», eller «Når jeg jobber med et matematisk problem, hopper jeg mellom ulike strategier». Noen utsagn hadde også en positiv affektiv ladning, som «Jeg blir engasjert når noen starter en matematisk diskusjon», eller «Jeg har problemer med å legge fra meg matematiske oppgaver». Målet elevene fikk på MI, sa derfor noe om deres egenrapporterte arbeidsmetoder og tenkemåter, som videre kunne tolkes som ulik grad av indre motivert og dyp tilnærming til matematikk. Målet sa også noe om deres grad av positiv affekt for faget.

Utsagnene i instrumentet kan knyttes til ulike definisjoner og operasjonaliseringer av MI. For eksempel kan utsagnene tolkes som narrativer slik Sfard og Prusak (2005) definerer identitet. Elevenes responser er deres egen fortelling om seg selv, fortalt til seg selv eller forskeren, om hvem de er som utøvere av matematikk. Når en elev er enig i utsagn som «Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med», er dette elevens egen fortelling om seg selv. Undersøkelsen kan ikke kontrollere om eleven faktisk gjør dette. Solomon (2007) definerer MI som en følelse av inkludering eller ekskludering i et matematikkfellesskap. Hun mener at elever som føler de deltar i utvikling av idéer og forhandling om mening vil få eierskap til innholdet, og derav få høy grad av tilhørighet til praksis. Utsagn som «Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker» og «Matematiske idéer jeg leser eller hører om setter meg på sporet av egne tankerekker» er begge eksempler på utsagn som viser at eleven føler den har noe å bidra med og derfor føler hun er med i forhandling om mening. Bishop (2012) hevder at MI er idéene man har om hvem man er med tanke på matematikk og aktivitetene som hører til. Utsagn som «Jeg blir engasjert når noen starter en matematisk diskusjon» uttrykker hvem eleven føler hun er i en situasjon der det diskuteres matematikk. De fleste utsagn i instrumentet er fint kompatible med de nevnte definisjonene av MI, og det kan ikke utelukkes at flere definisjoner også vil passe.

Elever med sterk ytre motivasjon vil også kunne si seg enige i mange utsagn dersom de har lært effektive metoder for læring. Jeg tror likevel at selv om arbeidsmetoder og tenkemåter som støtter dybdelæring og relasjonell læring til en viss grad kan læres, er de kanskje først og fremst et resultat av positiv affekt for matematikkfaget. Dette kan vi se hos matematikere som selv forteller om hvordan de tenker og jobber med matematikk (Burton, 1998). Gode tenkemåter og arbeidsmetoder vil derfor kunne være et uttrykk for positiv affekt til faget, og som et uttrykk for at man har en sterk matematisk identitet.

5.2 Instrumentets begrensninger

Instrumentet viste seg egnet for måling av MI på videregående skole. Jeg ser likevel forbedringsmuligheter knyttet til språket som er brukt i utsagnene, og til den visuelle utformingen av spørreskjemaet.

5.2.1 Språket i utsagnene

Instrumentet var opprinnelig laget for studenter. Studenter er eldre enn mine respondenter, og de har valgt en akademisk utdanning. Elevene i min studie bestod av 40 % yrkesfagelever, som i utgangspunktet har valgt en ikke-akademisk karrierevei. Matematikknivået til studentene og respondentene i denne undersøkelsen var også stort sett på ulikt nivå. I tillegg hadde 15 % av elevene i undersøkelsen minoritetsspråklig bakgrunn, mange med kort botid i Norge. Det er ikke usannsynlig at det var ord eller formuleringer i utsagnene i instrumentet som elevene i denne undersøkelsen ikke forstod slik de var ment, at de la noe annet i dem enn en student ville gjort, eller som de ikke forstod i det hele tatt. Et eksempel er utsagn 5: *dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv*. I studien av studenter la dette utsagnet seg på 5. øverste plass, mens det i min undersøkelse la seg på 14. plass. En av grunnene til at utsagnet var relativt lettere å si seg enig i for elevene på videregående skole enn for studenter, kan være at elevene ikke forstod hva ordet *utlede* betyr. Utledning av formler og matematiske bevis blir kanskje lite vektlagt i de praktiske matematikkfagene. Elevene kan for eksempel ha tenkt at utsagnet handlet om å prøve å gjøre oppgaven selv om han/hun har glemt formelen eller metoden. En annen grunn kan være at elevene forstår ordet, men at matematiske formler i 1PY er relativt enkle, og at de kanskje tenker at de ville klart å utlede dem selv. Det er vanskeligere å utlede formler selv i arbeid med matematikk på høyere nivåer.

5.2.2 Instrumentets utforming

Instrumentet var utformet med ruter for avkrysning av svaralternativer som lå ved siden av hverandre, og ikke som separate avkrysningsbokser (vedlegg 2-5). Mange av elevene i undersøkelsen valgte å sette et kryss midt mellom to alternativer der de var usikre på hvilket av dem de ville svare. Alle responser som lå mellom to svaralternativer ble registrert som manglede data, og gjorde at datamaterialet ble mindre enn det kunne vært. Dersom avkrysningsboksene hadde ligget mer separat ville muligens fristelsen til å svare midt imellom vært mindre. Det at elever hadde behov for å krysse av mellom to alternativer kan indikere det hadde vært behov for en mer nyansert responsskala med for eksempel 5 alternativer.

5.3 Hva er sammenhengen mellom MI og matematikkprestasjoner?

I denne undersøkelsen har jeg funnet korrelasjon mellom MI og prestasjoner i matematikk (r = 0,32 og r = 0,33, p < 0,05). Korrelasjonen ansees å ha middels effekt (Field, 2013). Selv om effekten regnes som middels, forklarer den bare 10 % av variasjonene. Det er altså omtrent 90 % av variasjonen i MI og prestasjoner som må forklares på annen måte enn som en effekt av den andre variabelen. Før jeg ser på tidligere funn og diskuterer implikasjoner for undervisning og videre forskning, vil jeg argumentere for at korrelasjonen *kan* være større enn det som vises i analysen, samt diskutere hvorfor karakterer som mål på prestasjoner kan være lite valide og reliable.

5.3.1 Betydningen av lav reliabilitet i korrelasjonsanalysene

Korrelasjonskoeffisientene som er funnet i denne studien for sammenhengen mellom MI og henholdsvis matematikkarakterer og kartleggingsresultater, bygger på tre ulike målinger. Det er kun målet på MI som har kjent reliabilitet. Det er grunn til å tro at reliabiliteten til karakterene ikke er veldig god. Karakterene gis på en skala med seks målepunkter. En skala med så få målepunkter vil nesten alltid gi lav reliabilitet (Linacre, 2006). Det er ikke slik at to elever med karakteren 4 er på nøyaktig samme faglige nivå, og i overgangene mellom to karakterer vil kanskje elever som er på samme faglige nivå få ulike karakterer.

En korrelasjonsanalyse mellom to sett mål med lav reliabilitet vil kunne gi en falsk lav korrelasjon grunnet målefeilene i datasettene (Muchinsky, 1996). Spearman (1904) utviklet en metode for å korrigere for målefeilene til datasettene for å få en mer riktig korrelasjonskoeffisient, «correction for attenuation» (korrigering for demping).

$$P_{xy} = \frac{r_{xy}}{\sqrt{r_{xx}}\sqrt{r_{yy}}}$$

Formel 8. *«Correction for attenuation»*

P_{xy}*= korrigert korrelasjonskoeffisient,* r_{xy}*= opprinnelig korrelasjonskoeffisient,* r_{xx}*= reliabiliteten til første variabel,* r_{yy}*= reliabiliteten til andre variabel* (Muchinsky, 1996).

Vi vet at reliabiliteten til mål på MI i denne undersøkelsen er 0,86. Reliabiliteten for karakterer vet vi ikke, men jeg lar reliabiliteten være 0,7 for å kunne vise et eksempel på hvordan korrelasjonskoeffisienten endrer seg når man korrigerer for dempingen som lav reliabilitet forårsaker. Korrigert korrelasjonskoeffisient blir da 0,41. Dersom reliabiliteten til karakterene er enda lavere, vil korrigert korrelasjon bli enda bedre.

Reliabiliteten til resultatene i Kartleggeren er muligens høyere enn for karakterer. Kartleggeren opererer med prosent som skala, og har derfor mange målepunkter. Det er likevel usannsynlig at reliabiliteten er perfekt, så vi kan regne med at en korrigering for demping også ville gitt høyere korrelasjon mellom MI og resultater fra Kartleggeren.

Cohen (2018, s. 573-574) argumenterer for at også læreres ulike praksis når det kommer til karaktersetting kan påvirke reliabiliteten til karakterer som mål. Dette gjelder både mellom ulike lærere, men også det at lærere ikke alltid gjør likt selv fra gang til gang når de setter karakterer. Ulike matematikkfag kan også ha ulik vanskelighetsgrad slik at det er vanskeligere å få god karakter i et fag med høyere faglig nivå enn i et fag med lavere faglig nivå. Noe av dette diskuterer jeg i neste underkapittel, og det kan bidra til å underbygge min antagelse om at MI og karakterer har høyere korrelasjon enn det som kommer fram i denne studien.

5.3.2 Usikkerhet knyttet til karakterer som mål

Det er mulig å sette spørsmålstegn ved validiteten til karakterene. Hva er det som blir målt? Måler de det samme? Som nevnt i teorikapittelet kan lærere som setter terminkarakterer, legge ulike forhold til grunn når de setter karakterer (Prøitz & Borgen, 2010), selv om det i opplæringsloven tydelig er beskrevet at det bare skal være elevenes oppnådde kompetanse, i henhold til kompetansemål i læreplanen, som skal være grunnlag for standpunktkarakter (Lovdata, 2006). Noen lærere mener et profesjonelt faglig skjønn må ligge til grunn for en valid og rettferdig standpunktvurdering (Hovdhaugen, Prøitz & Seland, 2018), og de kan legge både innsats og progresjon til grunn for vurderingen (Prøitz & Borgen, 2010). Selv om man bare legger kompetansemål fra læreplanen til grunn ved karaktersetting, er det igjen mange måter å tolke disse på. Lærere vil derfor både legge ulike forhold til grunn for sin vurdering, óg ha ulik tolkning av læreplanen. Dette vil til slutt kunne påvirke validiteten til karakterer som settes av faglærere. I min studie er det syv ulike matematikklærere som har satt elevenes karakterer, noe som åpner for syv ulike tolkninger av kompetansemål og grunnlag for karaktersetting.

Validiteten til karakterer som mål i denne studien kan være påvirket av at elevene tok ulike matematikkfag. En elev som fikk fire i et teoretisk matematikkfag på høyt faglig nivå, som R1 eller R2, ville kanskje fått fem eller seks i et fag med lavere faglig nivå. Elevene har i praksis blitt målt på forskjellige ting. Elevene som tok et teoretisk fag hadde i gjennomsnitt høyere MI enn elevene som tok et praktisk matematikkfag, men ble vurdert strengere på prestasjoner. Disse funnene støtter teorien om at det kan være høyere korrelasjon mellom MI og karakterer enn det som kommer fram i analysene.

Det er mulig å tenke seg at en lærer mener det skal mer til for å gå fra en femmer til en sekser, enn hun mener det skal til for å gå fra en treer til en firer i et fag. Dette gir i praksis en karakterskala som ikke er på intervallnivå. En sensorveiledning for matematikkeksamen viser noe av det samme. Det regnes ut en poengsum etter hvor mange oppgaver som er riktige, og veiledningen gir kriterier for hvor mange poeng som skal til for å oppnå de ulike karakterene (Tabell 15). Poengintervallet for å få en toer er mye større enn poengintervallet for å få en sekser (Utdanningsdirektoratet, 2019b). Riktignok vektes oppgaver med høy vanskelighetsgrad til en viss grad, og veilederen inneholder en spesifisering av at poenggrensene er veiledende.

Tabell 15. *Poenggrenser fra sensorveilederen til eksamen i matematikk 2PY vår 2019. Høyest mulige poengsum er 60 poeng.*

Karakter	1	2	3	4	5	6
Poeng		12	24	35	45	56

(Utdanningsdirektoratet, 2019b)

Intervallene for karakterene er altså ikke like, og skalaen for karakterer kan ikke sies å være på intervallnivå. Et gjennomsnitt av karakterer brukes likevel som grunnlag for opptak til videre studier, selv om gjennomsnitt ikke er en «lovlig» statistisk metode for ordinale data (Stevens, 1946). Karakterer brukes i praksis som intervallmål til langt mer avgjørende formål enn jeg har gjort i denne studien. Jeg har derfor valgt å behandle karakterer som intervallmål i mine analyser, selv om jeg ikke kan argumentere for at de faktisk er det.

5.3.3 Korrelasjon og årsakssammenheng

Det at det er funnet korrelasjon mellom variablene MI og henholdsvis karakterer og kartleggingsresultater, sier bare at det er sammenheng mellom dem. Korrelasjon sier ikke noe om årsakssammenhenger, og vi kan derfor ikke si at elever automatisk vil få bedre karakterer i matematikk dersom de øker MI. Like gjerne kan det hende at de vil få høyere MI dersom de gjør det bedre på skolen.

Det er ikke godt å si om det er høy MI som fører til gode arbeidsmåter og gode prestasjoner, eller om de gode prestasjonene kommer som følge av gode arbeidsmetoder og tenkemåter, som igjen kan føre til høyt mål på MI. Det vil likevel være de som også er enige i utsagn som har en positiv affektiv ladning som får de høyeste målene på MI. De fleste av utsagnene med positiv affektiv ladning ligger i øvre del av skalaen. Men her kan det igjen diskuteres om det er de gode prestasjonene som fører til positiv affekt, eller om det er positiv affekt som fører til gode prestasjoner. Det kan i tillegg være eksterne faktorer som påvirker begge parametere, som for eksempel støtte hjemmefra, relasjon til lærer, læringsmiljø eller undervisningsmetoder.

5.3.4 Sammenheng mellom MI og prestasjoner i tidligere forskning

Det er i denne studien funnet en sammenheng mellom MI og prestasjoner i matematikk hos elever på en videregående skole. Siden det ikke er konsensus blant forskere om hvordan MI defineres, vil det også være vanskelig å sammenligne resultater fra andre studier på MI. Det er heller ikke gjort kvantitative studier tidligere som ser på sammenhengen mellom MI og prestasjoner. Derimot er det gjort en rekke studier med funn av sammenheng, både kvantitativt og kvalitativt, mellom blant annet skoleprestasjoner og affekt (Ahmed, Van der Werf, Kuyper & Minnaert, 2013; Ma, 1999; OECD, 2015; Singh, Granville & Dika, 2002), og prestasjoner og dybdelæring (Ahmed et al., 2013; Purdie & Hattie, 1999). Tidligere funn støtter funn som er gjort i denne studien. Ahmed et al. (2013) fant i en kvantitativ studie at det var sammenheng mellom prestasjoner og graden av variabler som angst (negativ korrelasjon), glede over faget og dybdelæring. Av disse variablene var det dybdelæring som hadde sterkest korrelasjon med prestasjoner (Ahmed et al., 2013). PISA-undersøkelsen har funnet sammenheng mellom elevers mestringstro og prestasjoner (OECD, 2015). Singh et al. (2002) fant i en kvantitativ studie positiv korrelasjon mellom i hvilken grad elever gleder seg til matematikktimene og deres karakterer i matematikk. Ulike holdnings- og motivasjonsfaktorer ble testet for korrelasjon, direkte og indirekte, med matematikkprestasjoner, og konklusjonen var at disse faktorene spiller en stor rolle i forklaringen av variasjonen i prestasjoner (Singh et al., 2002).

5.4 Legger lærere vekt på MI når de vurderer elevene?

Det var ingen signifikant forskjell på korrelasjonen mellom MI og karakterer som var satt av lærerne og resultatene fra den normerte og anonyme kartleggingsprøven. Dette kan tyde på at lærerne som var med i denne studien, ikke la vekt på arbeidsmåtene eller det affektive hos elevene da de satte terminkarakterer i matematikk. Dette er i tråd med opplæringsloven som sier at det kun skal være oppnådd kompetanse i henhold til kompetansemålene i læreplanen som skal legges til grunn for karakterer i fag (Lovdata, 2006).

Tidligere forskning har vist at lærere kan legge elevers affekt til grunn når de setter karakterer (Ma, 1999; Prøitz & Borgen, 2010). Ma (1999) fant i tillegg at det var høyere korrelasjon mellom lærersatte karakterer og elevers matematikkangst enn det var mellom resultater på normerte og anonyme prøver og matematikkangst. Resultatene i min studie støtter ikke disse resultatene. Det må likevel tas et lite forbehold om at korrelasjonen mellom MI og karakterer kan være høyere enn jeg har beregnet grunnet dårlig reliabilitet på karakterene. Det kunne gitt større forskjell på korrelasjonene.

5.5 Studiens aktualitet i forbindelse med fagfornyelsen

Vi skal fra høsten 2020 og de kommende år gradvis gå over til ny læreplan, *fagfornyelsen*, for grunnskolen og videregående skole i Norge. Nåværende *generell del* i LK06 (Utdanningsdirektoratet, 2015) erstattes i fagfornyelsen med *overordna del* (Utdanningsdirektoratet, 2020), og kompetansemål for alle fag er revidert.

En av de nye større endringene som er gjort i fagfornyelsen, er vektlegging av dybdelæring. «Skolen skal gi rom for dybdelæring slik at elevene utvikler forståelse av sentrale elementer og sammenhenger innenfor et fag, og slik at de lærer å bruke faglige kunnskaper og ferdigheter i kjente og ukjente sammenhenger» (Utdanningsdirektoratet, 2020). Dybdelæring skal ifølge fagfornyelsen skje ved at elevene utvikler evne til å stille spørsmål, utforske og eksperimentere, og elevene skal få rike muligheter til å utvikle engasjement og utforskertrang (Utdanningsdirektoratet, 2020). Engasjement kan sees på som en type positiv affekt. Engasjement sammen med evne til, eller ønske om, tenkemåter og arbeidsmetoder som fremmer dybdelæring, er egenskaper vi finner igjen hos elever med høy MI.

5.6 Pedagogiske implikasjoner

5.6.1 Bruk av instrumentet i klasserommet

Som følge av at fagfornyelsen vektlegger engasjement og dybdelæring, vil måling av MI kunne fungere som en monitorering av dette hos elevene. Lærere som ønsker å bedre sin undervisning med tanke på dybdelæring, konseptuell/relasjonell forståelse, engasjement og eventuell annen positiv affekt hos elevene har, med et validert instrument lik det jeg har brukt i denne studien, mulighet til å måle elevenes MI før og etter pedagogiske tiltak som er ment å fremme egenskapene. Siden målene instrumentet gir er på intervallnivå, vil det være mulig å bruke dem i statistiske undersøkelser og for eksempel gjøre en signifikanstest for å se om det pedagogiske tiltaket har ført til endring hos elevene. Før- og etter-testing vil kunne gi indikasjoner på instrumentets *responsivitetsaspekt* (Wolfe & Smith, 2007).

Ettersom grunnlaget for vurdering kun skal være basert på kompetansemålene i læreplanen, bør måling av MI ikke brukes som en del av vurderingen av elevene. Forhold fra *generell del*, eller den kommende *overordnet del*, av læreplanen skal ikke være grunnlag for vurdering. Lærere kan derfor ikke vurdere elever etter om de viser engasjement, positiv affekt eller om de har gode arbeidsmetoder eller tenkemåter i faget. Det er derimot ikke usannsynlig at disse forholdene vil bidra indirekte til at elevene oppnår høyere kompetanse.

Bruk av tester som del av undervisning skal man generelt være svært forsiktig med. International Test Commission (ITC) har utviklet retningslinjer (International Test Commission, 2001) for hvordan tester skal brukes, og retningslinjene brukes blant andre av Norsk psykologiforening. Hensikten med retningslinjene er å sikre kompetente testbrukere som vil «bruke tester hensiktsmessig, med profesjonalitet, og på en etisk forsvarlig måte, og som tar i betraktning behov og rettigheter til de involverte i testprosessen, grunnene for testing, formålet med testing, og den bredere konteksten som testingen foregår i» (International Test Commission, 2001).

I tillegg til å understreke viktigheten av at testbrukerne må ha nok teknisk og teoretisk kunnskap om testenes validitet og reliabilitet, legger retningslinjene vekt på at resultatene fra en test må tolkes riktig, at de ikke skal kunne misforstås, og at de ikke må misbrukes. Man må unngå situasjoner der den som administrerer testen kan ha egeninteresse i vurderingens resultat, eller der vurderingen kan skade forholdet til testtakeren (International Test Commission, 2001). Dersom en lærer skal utføre testen på elevene sine, er det fare for at elevene ikke vil svare ærlig dersom de vet at det er faglæreren som også skal vurdere resultatene av testen. Dersom en elev som er uenig med de fleste utsagnene i testen er ærlig, risikerer han/hun at læreren vil legge vekt på dette når hun/han skal sette karakterer, da noen lærere legger mer i karakterene enn oppnådd kompetanse (Prøitz & Borgen, 2010). Dersom eleven ikke er ærlig, vil testen ikke ha noen verdi. Testen bør derfor administreres av kvalifiserte personer som ikke er karaktersettende faglærere for elevene som blir testet.

Andre situasjoner der testen kan komme til nytte i skolehverdagen, er som en bevisstgjøring av lærere og elever. Utsagnene kan brukes som utgangspunkt for diskusjon mellom matematikklærere, eller i klasserommet, med kunnskap om at de beskriver karakteristikker hos elever som har høy MI, som har gode arbeidsmetoder i matematikk, og som har gode prestasjoner i faget. Ved bruk av instrumentet på denne måten er det ikke nødvendig å se på testresultater, da det er innholdet i utsagnene som er interessant.

5.6.2 Pedagogiske tiltak for å øke MI og matematikkprestasjoner

Kunnskapen om at høy MI henger sammen med prestasjoner i matematikk, kan på flere måter utnyttes pedagogisk. Vi vet ikke årsakssammenhengen, men kan anta at det er en gjensidig påvirkning der høy MI vil føre til bedre prestasjoner, og gode prestasjoner vil føre til høy MI. Som faglærer kan man derfor hjelpe elever ved å gi oppgaver de vil mestre for å bygge opp MI, samt hjelpe dem med å utvikle gode arbeidsmåter og tenkemåter som fremmer dybdelæring.

En læringsteori som støtter tankegangen om at affekt, arbeidsmåter og prestasjoner påvirker hverandre gjensidig, er Kilpatricks rammeverk om «The Five Strands of Mathematical Proficiency» (De fem trådene i matematisk kompetanse) (Kilpatrick et al., 2001). Rammeverket legger vekt på at det er like viktig å oppleve matematikk som meningsfylt og engasjerende, som det er å ha gode regneferdigheter og arbeidsmetoder. Wæge og Nostrati (2015) foreslår ut fra Kilpatricks (2001) rammeverk «ambisiøs undervisning» som en type undervisning som vil fremme den matematiske kompetansen. «Ambisiøs undervisning støtter matematisk meningsskaping, identitetsforming, og likeverdige muligheter til å lære. Det krever en undervisningspraksis som tillater at lærere engasjerer seg dypt i elevenes tenkning, som gir robuste muligheter for at alle barn kan lære, som støtter meningsfull matematikdeltakelse for alle barn, og som rokker ved tidligere antagelser om hvem som kan og ikke kan matematikk» (min oversettelse) (Kazemi et al., 2012 referert i Wæge & Nosrati, 2015). Wæge og Nosrati (2015) foreslår undersøkende undervisningskontekster som lar elevene selv utforske problemer for å finne mønstre og systemer, og der diskusjon av løsningsstrategier elevene imellom er en viktig del av prosessen, som et godt pedagogisk tiltak for å øke elevenes matematiske kompetanse i henhold til Kilpatricks (2001) rammeverk.

5.7 Videre forskning

Resultater fra kvantitative studier kan utløse behov for kvalitative studier innenfor samme tema for å belyse tendenser som dukker opp i det kvantitative materialet (Cohen et al., 2018). I tillegg kan instrumentet i seg selv brukes i videre forskning.

5.7.1 Bruk av instrumentet

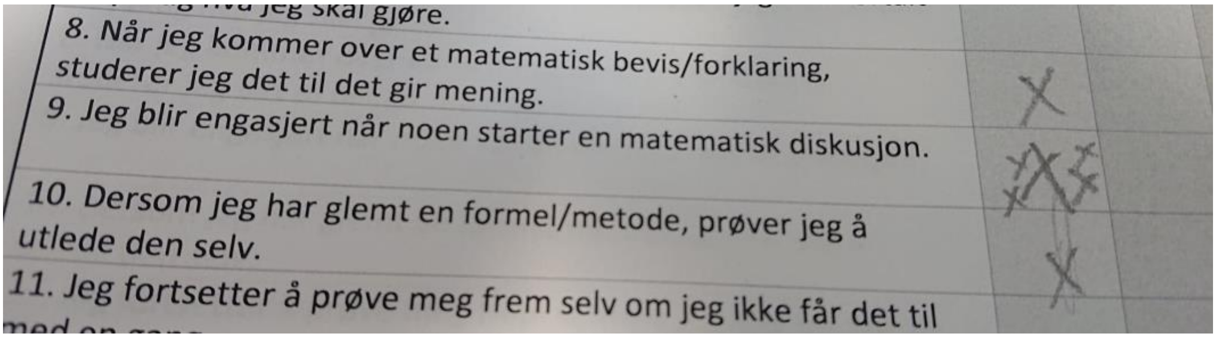
Dersom instrumentet skal brukes til videre forskning, må det først valideres for gruppen det skal forskes på. Inntil videre finnes det ikke validering av instrumentet for nok undergrupper til å kunne si at det er gyldig på tvers av kontekster (Kaspersen, 2018). Videre forskning kan derfor innebære en validering og sammenligning av instrumentet på tvers av kontekster slik at det kan standardiseres og brukes uten ny validering for hver gruppe som skal undersøkes.

Dersom instrumentet skal brukes på nye grupper, kan det være en fordel å tilpasse språket i utsagnene til gruppen som skal undersøkes. Dette er i tråd med retningslinjene til ITC som sier at en test skal være tilpasset gruppen den utføres på (International Test Commission, 2001).

Formålet med å bruke instrumentet i forskning kan være av samme type som for undervisning. Instrumentet kan brukes for å se på utvikling av elevers MI over tid, eller før og etter pedagogiske tiltak, som for eksempel «ambisiøs undervisning» som ble nevnt i kapittel 5.6.2.

5.7.2 Forskning med utgangspunkt i resultater fra denne studien

Det var flere elever som prøvde å formidle noe mer på spørreskjemaet enn det ble spurt om. Figur 19 viser respons fra en elev som hadde behov for å uttrykke ekstra tydelig at han ikke var enige i utsagnet «Jeg blir engasjert når noen starter en matematisk diskusjon». Det var mange slike forsøk på kommunikasjon fra elevene, og dette viser at det kan komme annen informasjon i en kvalitativ undersøkelse, for eksempel et intervju, enn det kan komme fra en kvantitativ spørreundersøkelse.



Figur 19. *Eksempel på elevrespons med alternativ avkrysning*

Et forslag for videre kvalitativ analyse vil være å intervjue elever med ulike mål på MI, og høre hvordan de selv identifiserer seg som utøvere av matematikk. Dersom elevene her får snakke mer fritt, og ikke er bundet til utsagnene i instrumentet, kan resultater fra en slik undersøkelse være med på å validere eller utvikle instrumentet ytterligere.

Dersom instrumentet brukes til å måle MI før og etter pedagogiske tiltak, kan supplerende intervjuer med elever som viste stor eller liten endring bidra med nyttig informasjon om hva som fungerte eller ikke fungerte i tiltaket.

Med elevenes kompetanse målt med Rasch-modellen, ville vi fått et mål på kompetanse som var på intervallnivå. En eventuell studie som søker å bekrefte eller avkrefte funn fra denne studien vil få en mer presis korrelasjon med kompetanse på intervallmål.

Det at ingen av de 50 elevene med lavest MI var elever som selv hadde valgt å ta et teoretisk matematikkfag, viser at det kan være en sammenheng mellom høy MI og valg av matematikk som studieretning senere. Det kunne vært interessant å finne ut om elever med høy MI faktisk oftere velger matematikk i videre studier, og om det er forskjell i prestasjoner til studenter med høy MI og elever med høye karakterer fra videregående skole. Kan det hende at MI er viktigere enn høye karakterer i høyere studier som innebærer matematikk?

Tidligere forskning har vist at lærere kan legge elevers affekt til grunn når de setter karakterer i matematikk (Ma, 1999; Prøitz & Borgen, 2010). Denne studien støttet ikke opp under dette. Jeg mener likevel at det kan være interessant å se litt nærmere på denne problemstillingen. Studien kunne vært gjort på en gruppe som tok samme matematikkfag, slik at karakterene som ble brukt i målingen var mer sammenlignbare. Studien kunne også vært gjort med flere lærere, slik at det ikke bare var praksis på én skole som ble undersøkt.

5.8 Avslutning

I denne studien har jeg sett på sammenhengen mellom elevers matematiske identitet og deres prestasjoner i matematikk, og funnet medium effekt (r = 0,32 og r = 0,33) som forklarer om lag 10 % av total varians (Field, 2013, s. 82). Jeg fant ingen signifikant forskjell på korrelasjonen mellom MI og karakterene som var gitt av faglærere og MI og kartleggingsresultater (p = 0,89), noe som tyder på at lærerne i denne studien ikke la elevenes MI til grunn da de vurderte dem.

Bildet av hva som gjør at elever presterer bra i matematikk, er komplekst. Både affekt, læringsmiljø, faginnhold og undervisningsmetoder kan spille inn, og ingen av aspektene fungerer optimalt uten at de andre er på plass. Elever formes av hva som skjer inne i klasserommet, og det kan gi grunnlag for identiteten de utvikler. Dybdelæring kan være noe elever med sterk MI selv streber etter, men det kan også være noe lærere legger opp til slik intensjonen er i fagfornyelsen (Utdanningsdirektoratet, 2020). Boaler (2002, s. 132) poengterer at «ulik pedagogikk er ikke bare verktøy for mer eller mindre kunnskap, den former kunnskapens natur og definerer identiteten elevene utvikler som utøvere av matematikk» (min oversettelse). Identiteten til de lærende er, slik hun beskriver det, et resultat av hvordan kunnskap presenteres. Matematikkunnskap som oppfattes som oppnåelig, og hvor man selv føler man kan være med å forhandle om innhold og mening, vil kanskje kunne skape en sterk matematisk identitet, samtidig som det kan føre til bedre resultater.

Referanser

Ahmed, W., Van der Werf, G., Kuyper, H. & Minnaert, A. (2013). Emotions, self-regulated learning, and achievement in mathematics: A growth curve analysis. *Journal of Educational Psychology, 105*(1), 150. https://doi.org/10.1037/a0030160

Aksnes, M. (2019). Hva er identitet? Hentet fra https://ndla.no/subjects/subject:18/topic:1:185340/topic:1:71188/resource:1:65716

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-573.

Andrich, D. (1989). Distinctions between assumptions and requirements in measurement in the social sciences. *Mathematical and Theoretical Systems, 4*, 7-16.

Bishop, J. P. (2012). "She's always been the smart one. I've always been the dumb one": Identities in the mathematics classroom. *Journal for Reasearch in Mathematics Education, 43*(1), 34-74. Hentet fra https://www.jstor.org/stable/10.5951/jresematheduc.43.1.0034?seq=1

Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning* (2. utg.). Mahwah, N.J.: Lawrence Erlbaum Associates.

Bond, T. & Fox, C. M. (2015). *Applying the Rasch model: Fundamental Measurement in the Human Sciences* (bd. 3). New York: Routledge.

Boone, W. J., Staver, J. R. & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-6857-4

Burton, L. (1998). The practices of mathematicians: What do they tell us about coming to know mathematics? *Educational Studies in Mathematics, 37*(2), 121-143.

Cohen, L., Manion, L. & Morrison, K. (2018). *Research methods in education* (8. utg.). London: Routledge.

Darragh, L. (2016). Identity research in mathematics education. *Educational Studies in Mathematics, 93*(1), 19-33.

De nasjonale forskningsetiske komiteene. (2016, 27. april). Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi. Hentet fra https://www.etikkom.no/forskningsetiske-retningslinjer/Samfunnsvitenskap-jus-og-humaniora/

Deaux, K. (1993). Reconstructing social identity. *Personality and Social Psychology Bulletin, 19*(1), 4-12.

Entwistle, N., McCune, V. & Tait, H. (2013). Approaches and study skills inventory for students (ASSIST): Report of the development and use of the inventories. Hentet fra https://www.researchgate.net/publication/50390092_Approaches_to_learning_and_studying_inventory_ASSIST_3rd_edition

Eysenck, H. J. (1998). *Intelligence: A new look*. New Brunswick: Transaction Publishers.

Fagbokforlaget.no. (2020). Kartleggeren. Hentet 01.02 2020 fra https://kartleggeren.no/

Fennema, E. & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education, 7*(5), 324-326.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4. utg.). London: SAGE Publications Ltd.

Galton, F. (1879). Psychometric experiments. *Brain, 2*(2), 149-162.

Gee, J. P. (2000). Identity as an analytic lens for research in education. *Review of Research in Education, 25*, 99-125.

Goldin, G. A., Hannula, M. S., Di Martino, P., Pantziara, M., Zhang, Q., Morselli, F., … Jansen, A. (2016). Attitudes, beliefs, motivation, and identity in mathematics education: An overview over the field and future directions. *ICME-13* (s. 1-35): Springer. https://doi.org/10.1007/978-3-319-32811-9

Graven, M. & Heyd-Metzuyanim, E. (2019). Mathematics identity research: The state of the art and future directions. *ZDM, 51*, 1-17.

Hiebert, J. (1986). *Conceptual and procedural knowledge : the case of mathematics*. Hillsdale, NJ: Erlbaum. Hentet fra https://ebookcentral.proquest.com

Hovdhaugen, E., Prøitz, T. S. & Seland, I. (2018). Eksamens-og standpunktkarakterer–to sider av samme sak? *Acta didactica Norge, 4*(18). Hentet fra https://utdanningsforskning.no/artikler/eksamens--og-standpunktkarakterer--to-sider-av-samme-sak/

International Test Commission. (2001). International guidelines for test use. *International Journal of Testing, 1*(2), 93-114.

John, O. P. & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research, 2*(1999), 102-138.

Kaspersen, E. (2018). *On measuring and theorising mathematical identity* University of Agder, Faculty of Engineering and Science, Kristiansand.

Kaspersen, E., Pepin, B. & Sikko, S. A. (2017). Measuring STEM students' mathematical identities. *Educational Studies in Mathematics, 95*(2), 163. https://doi.org/10.1007/s10649-016-9742-3

Kilpatrick, J., Swafford, J. & Bradford, F. (2001). *Adding it up: Helping children learn mathematics*. Washington DC: National Academy Press. https://doi.org/10.17226/9822

Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement, 3*(1), 85-106.

Linacre, J. M. (2005). Measurement, meaning and morality. *Pacific Rim Objective Measurement Symposium and International Symposium on Measurement and Evaluation. Kuala Lumpur*.

Linacre, J. M. (2006). WINSTEPS Rasch measurement. Chicago, IL: Winsteps.com.

Linacre, J. M. (2012). Rasch-Winsteps-Facets online Rasch tutorial PDFs Hentet fra https://www.winsteps.com/tutorials.htm

Lovdata. (2006). Forskrift til opplæringslova. Hentet 01.02 2020 fra https://lovdata.no/dokument/SF/forskrift/2006-06-23-724/KAPITTEL_4#KAPITTEL_4

Lysø, K. O. (2010). *Sannsynlighetsregning og statistisk metodelære* (3. utg.). Bergen: Caspar forlag.

Lyubomirsky, S. & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46*(2), 137-155.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30*(5), 520-540.

Matematikksenteret. (2020). Hentet 27.01. 2020 fra https://www.matematikksenteret.no/videreg%C3%A5ende/vurdering-og-kartlegging/digitale-verkt%C3%B8y-til-eksamen-i-matematikk

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. I D. A. Grouws (Red.), *Handbook of Research on Mathematics Teaching and Learning* (bd. 1, s. 575-596). New York: Macmillan Publishing Co, Inc.

McLeod, D. B. (1994). Research on affect and mathematics learning in the JRME: 1970 to the present. *Journal for Research in Mathematics Education, 25*(6), 637-647.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355-383.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*(1), 63-75.

OECD. (2015). How confident are students in their ability to solve mathematics problems? *PISA in Focus, 56*. https://doi.org/10.1787/5jrs3cfzg836-en

Prøitz, T. S. & Borgen, J. S. (2010). *Rettferdig standpunktvurdering–det (u) muliges kunst? Læreres setting av standpunktkarakter i fem fag i grunnopplæringen* (8272186837). Oslo: NIFU STEP.

Purdie, N. & Hattie, J. (1999). The relationship between study skills and learning outcomes: A meta-analysis. *Australian Journal of Education, 43*(1), 72-86.

Radovic, D., Black, L., Williams, J. & Salas, C. (2018). Towards conceptual coherence in the research on mathematics learner identity: a systematic review of the literature. *Educational Studies in Mathematics, 99*(1), 21-42. https://doi.org/10.1007/s10649-018-9819-2

Research Data @NTNU. (2020). Hentet 4/5 2020 fra https://innsida.ntnu.no/researchdata

Sfard, A. & Prusak, A. (2005). Telling identities: In search of an analytic tool for investigating learning as a culturally shaped activity. *Educational Researcher, 34*(4), 14-22.

Singh, K., Granville, M. & Dika, S. (2002). Mathematics and science achievement: Effects of motivation, interest, and academic engagement. *The Journal of Educational Research, 95*(6), 323-332.

Skemp, R. R. (1976). Relational understanding and instrumental understanding. *Mathematics Teaching, 77*(1), 20-26.

Solomon, Y. (2007). Not belonging? What makes a functional learner identity in undergraduate mathematics? *Studies in Higher Education, 32*(1), 79-96.

Spearman, C. (1904). The proof and measurement of association between two things. . *The American Journal of Psychology, 15*(1), 72-101. https://doi.org/10.2307/1412159

ssb.no. (2019). Karakterer ved avsluttet grunnskole. Hentet fra https://www.ssb.no/statbank/table/07501/tableViewLayout1/

Stevens, S. S. (1946). On the theory of scales of measurement. *Science, New Series, 103*(2684), 677-680. Hentet fra http://www.jstor.org/stable/1671815

Utdanningsdirektoratet. (2013). Læreplan i matematikk fellesfag. Hentet 24.02 2020 fra https://www.udir.no/kl06/MAT1-04

Utdanningsdirektoratet. (2015). *Generell del av læreplanen*. Hentet fra https://www.udir.no/laring-og-trivsel/lareplanverket/generell-del-av-lareplanen/

Utdanningsdirektoratet. (2018). *Metodegrunnlag for nasjonale prøver*. Hentet fra https://www.udir.no/globalassets/filer/vurdering/nasjonaleprover/metodegrunnlag-for-nasjonale-prover-august-2018.pdf

Utdanningsdirektoratet. (2019a). *Dybdelæring*. Hentet fra https://www.udir.no/laring-og-trivsel/dybdelaring/

Utdanningsdirektoratet. (2019b). *Sensorveiledning MAT1005 Matematikk 2P-Y 20. mai 2019*. Udir.

Utdanningsdirektoratet. (2020). Overordnet del- verdier og prisnipper for grunnopplæringen. Hentet 07.04 2020 fra https://www.udir.no/lk20/overordnet-del/

Wenger, E. (2010). Conceptual tools for CoPs as social learning systems: Boundaries, identity, trajectories and participation. I C. Blackmore (Red.), *Social Learning Systems and Communities of Practice* (s. 125-143). London: Springer London.

Wolfe, E. & Smith, J. E. (2007). Instrument development tools and activities for measure validation using Rasch models: part II--validation activities. *Journal of Applied Measurement, 8*(2), 204-234.

Wright, B. D. & Stone, M. H. (1979). *Best test design*. Chicago, IL: MESA Press.

Wæge, K. & Nosrati, M. (2015). Sentrale kjennetegn på god undervisning i matematikk. Hentet fra https://www.matematikksenteret.no/sites/default/files/attachments/product/Oppdatert%20september%202019%20Sentrale%20kjennetegn%20på%20god%20læring%20og%20undervisning%20i%20matematikk.pdf

Ytterhaug, B. O. (2019). *Matematisk identitet i ungdomsskolen. En kvantitativ studie av elevers matematiske identitet.* NTNU, Trondheim.

Zan, R., Brown, L., Evans, J. & Hannula, M. S. (2006). Affect in mathematics education: An introduction. *Educational Studies in Mathematics, 63*(2), 113-121. Hentet fra https://www.jstor.org/stable/25472116

Zan, R. & Di Martino, P. (2007). Attitude toward mathematics: Overcoming the positive/negative dichotomy. *The Montana Mathematics Enthusiast,* (3), 157-168.

Vedlegg

Vedlegg 1: Datahåndteringsplan

Vedlegg 2: Spørreskjema A

Vedlegg 3: Spørreskjema B

Vedlegg 4: Spørreskjema C

Vedlegg 5: Spørreskjema D

Vedlegg 6: Samtykkeerklæring

Vedlegg 7: Informasjonsskriv til elever

Vedlegg 1: Datahåndteringsplan

Datahåndteringsplan for masterprosjektet

«Matematisk identitet og prestasjoner i matematikk»

Datainnsamling og metoder

Innsamlet data vil bestå av
- Spørreskjemaer som matematikkelever har fylt ut.
- Elevenes terminkarakterer i matematikk og resultater fra kartleggingsprøver.
- Elevenes navn og klasse.

Det er jeg, prosjektleder, som deler ut og samler inn spørreskjemaer.
Faglærere rapporterer terminkarakterer i matematikk og kartleggingsresultater direkte til prosjektleder.

Beskrivelse av data, formater, organisering og metadata

All innsamlet data er i papirform.

Lagring, arkivering, backup og deling

Papirskjemaer blir oppbevart i låst skap. Data kodes og anonymiseres før de lagres på en kryptert minnepinne. Minnepinnen lagres i låst skap. All data som inneholder identifiserbart materiale vil makuleres når prosjektperioden er over. Anonymisert data vil deles via NTNU etter at prosjektet er over.

Rettigheter, lisenser, personvern og etikk

Alle som deltar skriver under på en samtykkeerklæring der de gir samtykke til at jeg kan bruke svarene de har gitt på spørreskjemaet, samt samle inn navn og informasjon om resultater fra kartleggingsprøven og terminkarakterer i matematikk fra dem.

Kostnader og ansvar

Ingen kostnader knyttet til innsamling, lagring og deling av data.

Vedlegg 2: Spørreskjema A

Hvordan jobber du, og hva tenker du, når du jobber med matematikk?

Les påstandene nøye og svar på hvor enig du er i hver påstand. Læreren din skal ikke lese svarene dine.

Aldri/nesten aldri (1), Noen ganger (2), Ofte (3), Alltid/nesten alltid (4), Vet ikke

	1	2	3	4	Vet ikke
1. Jeg har problemer med å legge fra meg matematiske oppgaver.					
2. Dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv.					
3. Når jeg lærer en ny metode/algoritme, prøver jeg å finne ut hvorfor den virker.					
4. Hvis jeg prøver på en metode som ikke fører frem, bruker jeg tid på å finne ut hvorfor denne ikke virker.					
5. Hvis jeg står fast, prøver jeg å visualisere problemet.					
6. Jeg prøver å koble det jeg lærer opp mot det jeg vet fra før.					
7. Jeg fortsetter å prøve meg frem selv om jeg ikke får det til med en gang.					
8. Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.					
9. Jeg tar initiativ til å lære mer om et matematisk emne enn skole/jobb legger opp til.					
10. Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker.					
11. Når jeg jobber med et matematisk problem hopper jeg mellom ulike strategier.					
12. Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre.					
13. Jeg blir engasjert når noen starter en matematisk diskusjon.					
14. Når jeg jobber med en oppgave, stopper jeg opp underveis og reflekterer over hva jeg gjør.					
15. Når jeg kommer over et matematisk bevis/forklaring, studerer jeg det til det gir mening.					
16. Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode.					
17. Når jeg lærer noe nytt, fører det til at det er flere ting jeg ønsker å finne ut.					
18. Når jeg møter et matematisk problem, tenker jeg over om det finnes flere måter å løse oppgaven på.					
19. Jeg kan forklare hvorfor løsningen min er rett.					
20. Matematiske ideer jeg leser eller hører om setter meg på sporet av egne tankerekker.					

Navn (tydelig): ____ Klasse: _______________

Vedlegg 3: Spørreskjema B

Hvordan jobber du, og hva tenker du, når du jobber med matematikk?

Les påstandene nøye og svar på hvor enig du er i hver påstand. Læreren din skal ikke lese svarene dine.

Aldri/nesten aldri (1), Noen ganger (2), Ofte (3), Alltid/nesten alltid (4), Vet ikke

	1	2	3	4	Vet ikke
1.Når jeg møter et matematisk problem, tenker jeg over om det finnes flere måter å løse oppgaven på.					
2. Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode.					
3. Hvis jeg prøver på en metode som ikke fører frem, bruker jeg tid på å finne ut hvorfor denne ikke virker.					
4. Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.					
5. Når jeg lærer en ny metode/algoritme, prøver jeg å finne ut hvorfor den virker.					
6. Når jeg jobber med et matematisk problem hopper jeg mellom ulike strategier.					
7. Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre.					
8. Når jeg kommer over et matematisk bevis/forklaring, studerer jeg det til det gir mening.					
9. Jeg blir engasjert når noen starter en matematisk diskusjon.					
10. Dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv.					
11. Jeg fortsetter å prøve meg frem selv om jeg ikke får det til med en gang.					
12. Matematiske ideer jeg leser eller hører om setter meg på sporet av egne tankerekker.					
13. Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker.					
14. Jeg har problemer med å legge fra meg matematiske oppgaver.					
15. Jeg prøver å koble det jeg lærer opp mot det jeg vet fra før.					
16. Jeg kan forklare hvorfor løsningen min er rett.					
17. Hvis jeg står fast, prøver jeg å visualisere problemet.					
18. Jeg tar initiativ til å lære mer om et matematisk emne enn skole/jobb legger opp til.					
19. Når jeg jobber med en oppgave, stopper jeg opp underveis og reflekterer over hva jeg gjør.					
20. Når jeg lærer noe nytt, fører det til at det er flere ting jeg ønsker å finne ut.					

Navn (tydelig): ____ Klasse: _______________

Vedlegg 4: Spørreskjema C

Hvordan jobber du, og hva tenker du, når du jobber med matematikk?

Les påstandene nøye og svar på hvor enig du er i hver påstand. Læreren din skal ikke lese svarene dine.

Aldri/nesten aldri (1), Noen ganger (2), Ofte (3), Alltid/nesten alltid (4), Vet ikke

	1	2	3	4	Vet ikke
1. Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre.					
2. Når jeg jobber med en oppgave, stopper jeg opp underveis og reflekterer over hva jeg gjør.					
3. Når jeg jobber med et matematisk problem, hopper jeg mellom ulike strategier.					
4. Når jeg lærer en ny metode/algoritme, prøver jeg å finne ut hvorfor den virker.					
5. Når jeg lærer noe nytt, fører det til at det er flere ting jeg ønsker å finne ut.					
6. Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker.					
7. Jeg har problemer med å legge fra meg matematiske oppgaver.					
8. Jeg tar initiativ til å lære mer om et matematisk emne enn skole/jobb legger opp til.					
9. Når jeg møter et matematisk problem, tenker jeg over om det finnes flere måter å løse oppgaven på.					
10. Jeg prøver å koble det jeg lærer opp mot det jeg vet fra før.					
11. Jeg kan forklare hvorfor løsningen min er rett.					
12. Matematiske ideer jeg leser eller hører om setter meg på sporet av egne tankerekker.					
13. Dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv.					
14. Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode.					
15. Jeg fortsetter å prøve meg fram selv om jeg ikke får det til med en gang.					
16. Jeg blir engasjert når noen starter en matematisk diskusjon.					
17. Når jeg kommer over et matematisk bevis/forklaring, studerer jeg det til det gir mening.					
18. Hvis jeg prøver på en metode som ikke fører fram, bruker jeg tid på å finne ut hvorfor denne ikke virker.					
19. Hvis jeg sitter fast, prøver jeg å visualisere problemet.					
20. Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.					

Navn (tydelig): ____ Klasse: _______________

Vedlegg 5: Spørreskjema D

Hvordan jobber du, og hva tenker du, når du jobber med matematikk?

Les påstandene nøye og svar på hvor enig du er i hver påstand. Læreren din skal ikke lese svarene dine.

Aldri/nesten aldri (1), Noen ganger (2), Ofte (3), Alltid/nesten alltid (4), Vet ikke

	1	2	3	4	Vet ikke
1. Når jeg lærer noe nytt, stiller jeg meg selv egne spørsmål som jeg jobber med.					
2. Jeg blir engasjert når noen starter en matematisk diskusjon.					
3. Når jeg jobber med et matematisk problem hopper jeg mellom ulike strategier.					
4. Når jeg lærer en ny metode, bruker jeg tid på å se om jeg kan finne en bedre metode.					
5. Hvis jeg prøver på en metode som ikke fører frem, bruker jeg tid på å finne ut hvorfor denne ikke virker.					
6. Matematiske ideer jeg leser eller hører om setter meg på sporet av egne tankerekker.					
7. Når jeg lærer en ny matematisk metode, liker jeg å bli fortalt nøyaktig hva jeg skal gjøre.					
8. Jeg kan forklare hvorfor løsningen min er rett.					
9. Når jeg møter et matematisk problem, tenker jeg over om det finnes flere måter å løse oppgaven på.					
10. Når jeg lærer en ny metode, prøver jeg å finne situasjoner hvor denne ikke virker.					
11. Jeg prøver å koble det jeg lærer opp mot det jeg vet fra før.					
12. Jeg fortsetter å prøve meg frem selv om jeg ikke får det til med en gang.					
13. Når jeg lærer en ny metode/algoritme, prøver jeg å finne ut hvorfor den virker.					
14. Når jeg lærer noe nytt, fører det til at det er flere ting jeg ønsker å finne ut.					
15. Jeg har problemer med å legge fra meg matematiske oppgaver. .					
16. Dersom jeg har glemt en formel/metode, prøver jeg å utlede den selv.					
17. Når jeg jobber med en oppgave, stopper jeg opp underveis og reflekterer over hva jeg gjør.					
18. Jeg tar initiativ til å lære mer om et matematisk emne enn skole/jobb legger opp til.					
19. Hvis jeg står fast, prøver jeg å visualisere problemet. .					
20. Når jeg kommer over et matematisk bevis/forklaring, studerer jeg det til det gir mening.					

Navn (tydelig): ____ Klasse: _______________

Vedlegg 6: Samtykkeerklæring

Samtykkeerklæring

Jeg har mottatt og forstått informasjon om prosjektet *Matematisk identitet og prestasjoner i matematikk,* og har fått anledning til å stille spørsmål. Jeg samtykker til:

- ☐ at svarene i spørreskjemaet brukes i forskning.
- ☐ at lærer kan gi opplysninger om mine karakterer i matematikk og resultater på kartleggingsprøver i matematikk.

Jeg samtykker til at mine opplysninger behandles frem til prosjektet er avsluttet, ca. juni 2020.

(Signert av prosjektdeltaker, dato)

Vedlegg 7: Informasjonsskriv til elever

Informasjon om masterprosjektet «Matematisk identitet og prestasjoner i matematikk».

Prosjektet har som formål å se om det er en sammenheng mellom matematisk identitet og prestasjoner i matematikk. Det betyr at vi lurer på om måten man tenker om matematikk og om hvordan man jobber med matematikk har noe å si for hva man får vist at man får til. Alle elever som har matematikk på Færder videregående skole vil få spørsmål om å delta.

Masterprosjektet er knyttet til NTNU, institutt for lærerutdanning, og det er bare jeg (student Gro Tellsgård) og min veileder Eivind Kaspersen som får tilgang til informasjonen om deg. Informasjonen vil lagres digitalt og anonymiseres slik at den ikke kan kobles til deg hverken i tekst som publiseres, eller digitalt når prosjektet er ferdig juni 2020.

Alt du trenger å gjøre er å svare på et spørreskjema med 20 spørsmål som skal besvares med avkrysning. Læreren din skal ikke se hva du svarer. Jeg ønsker også å samle inn informasjon om dine resultater på Kartleggeren og din terminkarakter første halvår dette skoleåret (19/20).

Det er frivillig å delta i prosjektet. Hvis du velger å delta, kan du når som helst trekke samtykket tilbake uten å oppgi noen grunn. Alle opplysninger om deg vil da bli anonymisert. Det vil ikke ha noen negative konsekvenser for deg hvis du ikke vil delta eller senere velger å trekke deg.

Vi vil bare bruke opplysningene om deg til formålene vi har fortalt om i dette skrivet. Vi behandler opplysningene konfidensielt og i samsvar med personvernregelverket.

Dine rettigheter
Så lenge du kan identifiseres i datamaterialet, har du rett til:
- innsyn i hvilke personopplysninger som er registrert om deg,
- å få rettet personopplysninger om deg,
- få slettet personopplysninger om deg,
- få utlevert en kopi av dine personopplysninger (dataportabilitet), og
- å sende klage til personvernombudet eller Datatilsynet om behandlingen av dine personopplysninger.

Hva gir oss rett til å behandle personopplysninger om deg?
Vi behandler opplysninger om deg basert på ditt samtykke.

Hvis du har spørsmål til studien, eller ønsker å benytte deg av dine rettigheter, ta kontakt med NTNU, institutt for lærerutdanning, veileder Eivind Kaspersen (eivind.kaspersen@ntnu.no) eller student Gro Tellsgård (grote@stud.ntnu.no). Du kan også kontakte NSD-Norsk senter for forskningsdata AS på epost (personverntjenester@nsd.no) eller telefon: 55582117

Med vennlig hilsen

Gro Tellsgård

Student ved institutt for lærerutdanning, NTNU.

NTNU

Norwegian University of
Science and Technology