MASTEROPPGAVE

Masterstudium i skolerettet utdanningsvitenskap med fordypning i matematikk og matematikkdidaktikk

Mai 2021

Høyt- og lavtpresterende elevers kilder til mestringsforventning i matematikk

Marie Larsen

OsloMet – storbyuniversitetet

Fakultet for lærerutdanning og internasjonale studier

Institutt for grunnskole- og faglærerutdanning

Sammendrag

Siden Bandura (1977) først la frem teorien om mestringsforventning på 70-tallet, har et stadig voksende forskningsfelt kunnet bekrefte at mestringsforventning har en predikativ effekt på akademisk prestasjon. Formålet med denne masteroppgaven er å undersøke elevers kilder til mestringsforventning i matematikk, med utgangspunkt i følgende problemstilling: *Er det forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning i matematikk, og i så fall hvilke?* Jeg ønsker både å undersøke (1) hvordan et instrument for å måle elevers kilder til mestringsforventning kan oversettes og valideres og (2) om det er forskjeller i hvordan høyt- og lavtpresterende elever rapporterer kildene til mestringsforventning, og eventuelt hvilke mønstre som viser seg.

For å undersøke problemstillingen har jeg benyttet en kvantitativ tilnærming. I et samarbeid om forskningsprosjektet SESAM (self-efficacy, its sources and achievement in mathematics) har jeg, to medstudenter og vår felles veileder utviklet og pilotert en undersøkelse, som deretter ble distribuert til 583 elever på 9.trinn i Norge. Elever fra Øst-, Sør- og Vestlandet er representert i utvalget. En matematikkprøve med 12 oppgaver hentet fra TIMMS-undersøkelsen (Trends in International Mathematics and Science Study) dannet grunnlaget for inndelingen av høyt- og lavtpresterende elever. Det validerte instrumentet *Sources of Middle School Mathematics Self Efficacy Scale* (Usher & Pajares, 2009) ble oversatt og benyttet for å undersøke elevers kilder til mestringsforventning. Rasch Rating Scale Model (RSM) ble benyttet som analyseverktøy for å validere det oversatte instrumentet.

Analysene tilsier at den oversatte versjonen av instrumentet er valid og gir reliable målinger. Videre viser analysene at høytpresterende elever rapporterer signifikant større grad av gjenkjennelighet i påstander som antas å påvirke mestringsforventningen i positiv retning, enn det lavtpresterende elever rapporterer. Dette gjelder for alle kilder. I tillegg viser en sammenligning av de ulike kildene i gruppen høytpresterende elever at elevene rapporterer signifikant større grad av gjenkjennelighet i påstandene omhandlende kilden *mestringserfaringer* enn det de rapporterer i påstandene omhandlende de resterende kildene. Tilsvarende sammenligning for lavtpresterende elever viser tvert imot at elevene rapporterer signifikant mindre grad av gjenkjennelighet i påstandene omhandlende kilden *mestringserfaringer* enn det de rapporterer for påstandene omhandlende de resterende kildene. I lys av teori og tidligere forskning diskuteres den praktiske betydningen av disse funnene.

Abstract

Since Bandura (1977) first presented the theory of self-efficacy in the 70s, an ever-growing field of research has been able to confirm that self-efficacy has a predictive effect on academic achievement. The purpose of this master thesis is to investigate the sources of self-efficacy among students in mathematics, based on the following research question: *Are there differences in the sources of self-efficacy among high- and low-achieving students in mathematics, and if so, which ones?* I will examine both (1) how an instrument for measuring the sources of self-efficacy among students can be translated and validated and (2) whether there are differences in how high- and low-achieving students report their sources of self-efficacy, and possibly which patterns appear.

To be able to examine the research question, I have had a quantitative approach. Being part of the collaboration for the research project SESAM (self-efficacy, its sources and achievement in mathematics), I have developed a survey and conducted a pilot study together with two fellow students and our supervisor. The survey was later distributed to 583 students in the 9th grade in Norway. Students from Eastern-, Southern- and Western Norway are represented in our sample. A mathematics test, including 12 tasks from TIMMS (Trends in International Mathematics and Science Study) formed the basis for the division of high- and low-achieving students. *Sources of Middle School Mathematics Self Efficacy Scale* (Usher & Pajares, 2009), which is a validated instrument, was translated and used to examine the sources of self-efficacy among students. The Rasch Rating Scale Model (RSM) was used as an analysis tool to validate the translated instrument.

The analyzes indicate that the translated version of the instrument is valid and provides reliable measurements. Furthermore, the analyzes show that high-achieving students report a significantly greater degree of recognizability regarding statements that are assumed to affect their self-efficacy in a positive direction, compared to what low-achieving students report. This significant difference can be found in all four sources. Additionally, a comparison of the sources within each subgroup shows that the group of high-achieving students report a significantly greater degree of recognizability of the statements concerning *mastery experiences* than they report for the statements concerning the remaining sources. On the contrary, low-achieving students report a significantly lower degree of recognizability of the statements concerning *mastery experiences* than they report for the statements concerning the remaining sources. The practical significance of these findings is discussed in the light of theory and research on the field.

Forord

Fem lærerike år ved OsloMet – storbyuniversitetet nærmer seg nå slutten, og jeg kan se tilbake på en god studietid med inspirerende undervisning. Masterstudiet har gitt meg et innblikk i matematikk- og matematikkdidaktisk forskning og belyst interessante og aktuelle temaer innen matematikkundervisning. Jeg kan med sikkerhet si at utdanningen har formet og inspirert meg, samt gitt meg nyttige refleksjoner og praktiske verktøy jeg vil ta med meg inn i læreryrket.

Min veileder, Anette Hessen Bjerke, fortjener en stor takk for nyttige innspill, konstruktive tilbakemeldinger og god oppfølging med denne masteroppgaven. Takk for din begeistring og tro på prosjektet, og ikke minst tålmodig støtte og hjelp med forståelse og tolkning av Rasch-analyser.

Jeg vil også rette en takk til mine medstudenter Anna og Elisabeth i forskningsgruppa SESAM for et godt samarbeid gjennom hele prosessen. Samarbeidet har resultert i gode, faglige diskusjoner som har bidratt til å gi denne studien høyere kvalitet. Ikke minst har samarbeidet vært hyggelig, støttende og motiverende.

Til slutt vil jeg takke min søster Mirjam som har vært en god støttespiller fra dag én på lærerstudiet. Takk for godt faglig samarbeid og ikke minst en hyggelig studiehverdag.

Marie Larsen

Innhold

Sammendrag i

Abstract ii

Forord iii

Oversikt over tabeller vii

Oversikt over figurer viii

1.0 Introduksjon 1

2.0 Teori og tidligere forskning 3

2.1 Banduras mestringsforventning 3

2.1.1 Mestringserfaringer 4

2.1.2 Vikarierende erfaringer 6

2.1.3 Oppmuntring, støtte og overtalelse fra andre 8

2.1.4 Psykologiske og fysiologiske tilstander 9

2.2 Forskning på mestringsforventning i en skolekontekst 9

2.3 Forskning på mestringsforventning i matematikk 11

2.3.1 Matematikkfagets særtrekk 11

2.3.2 Kildene til mestringsforventning i matematikk 13

2.3.3 Forskning på mestringsforventning i matematikk i norsk kontekst 15

2.4 Høyt- og lavtpresterende elever 17

2.5 Tankesett 19

2.6 Sosiomatematiske normer 20

2.7 Forskningsspørsmål 20

3.0 Metode 22

3.1 Valg av metode 22

3.2 Studiens rammer og utarbeidelse av instrumenter 22

3.2.1 Forskningsgruppa SESAM 22

3.2.2 Instrumentene som måler elevers mestringsforventning (1) og prestasjon (3) i matematikk 23

3.2.3 Instrumentet som måler elevers kilder til mestringsforventning (2) 25

3.3 Rasch-analyse 28
3.3.1 Prinsippene bak Rasch-analyser 28
3.3.2 Endimensjonalitet 31
3.3.3 Tilpasningsstatistikk 32
3.3.4 DTF-analyser 34
3.3.5 Terskelverdier 34
3.3.6 Tolkning av polytome data i det benyttede instrumentet 34
3.4 Pilotering 36
3.5 Datainnsamling 37
3.5.1 Den endelige undersøkelsen 37
3.5.2 Gjennomføring av datainnsamling 37
3.5.3 Utvalg 39
3.6 Klassisk testteori 42
3.6.1 Hypotesetester 42
3.6.2 Uavhengig t-test 43
3.6.3 One-way repeated measure ANOVA 44
3.6.4 Omgjøring til z-scores 45
3.7 Forskningsetikk 45
4.0 Resultater 47
4.1 Instrumentets validitet og målingenes reliabilitet 47
4.1.1 ME-instrument 52
4.1.2 VE-instrument 54
4.1.3 OA-instrument 56
4.1.4 PTF-instrument 58
4.1.5 Et valid instrument 60
4.2 Resultater av studien 61
4.2.1 Sammenligning av høyt- og lavtpresterende elever innenfor hver enkelt kilde 61
4.2.2 Sammenligning av de ulike kildene innenfor hver av elevgruppene 62
4.3.3 Terskelverdier for hvert item 65

5.0 Diskusjon ... 71
5.1 Oversettelse og validering av instrumentet ... 71
5.2 Forskjeller mellom høyt- og lavtpresterende elevers rapportering av kildene ... 72
5.2.1 Signifikant forskjell mellom høyt- og lavtpresterende elever i alle kilder ... 73
5.2.2 Høyt- og lavtpresterende elevers rapportering av mestringserfaringer ... 74
5.2.3 Vikarierende erfaringer for høyt- og lavtpresterende elever ... 77
5.3 Praktiske implikasjoner ... 78
5.3.1 Et stort potensial for å fremme lavtpresterende elevers mestringserfaringer ... 79
5.3.2 Hvordan fremme mestringsforventningene i lys av instrumentenes items? ... 84
5.4 Oppsummering ... 87
6.0 Avslutning ... 88
6.1 Konklusjon ... 88
6.2 Kritiske refleksjoner ... 89
6.3 Innspill til videre forskning ... 91
6.4 Praktisk nytteverdi ... 91
Referanser ... 93
Vedlegg 1: Godkjennelse fra NSD ... 97
Vedlegg 2: Spørreundersøkelse til hoveddatainnsamling ... 99
Vedlegg 3: Retningslinjer for gjennomføringen ... 103
Vedlegg 4: Spørreundersøkelse til pilot ... 104
Vedlegg 5: Sources of Middle School Mathematics Self Efficacy Scale ... 110
Vedlegg 6: Svarkategorienes rekkefølge for delinstrumentene ... 111

Oversikt over tabeller

Tabell 1. Resultater for Rasch-analyse av ME-instrumentet. 52
Tabell 2. Resultater av Rasch-analyse på VE-instrument. 55
Tabell 3. Resultater fra Rasch-analyse på OA-instrumentet. 57
Tabell 4. Resultater av Rasch-analyse på PFT-instrument. 59
Tabell 5. Gjennomsnittet (logits) for hver av de to elevgruppene for hver kilde. 61
Tabell 6. Gjennomsnittlig z-scores for høyt- og lavtpresterende elever for alle kildene 63
Tabell 7. Rekkefølge over hvilke kilder elever rapporterer størst grad av gjenkjennelighet i. 77

Oversikt over figurer

Figur 1. En skjematisk modell av Rasch-målinger. 29
Figur 2. Eksempel på Item Characteristic Curve (ICC) for et item. 50
Figur 3. Eksempel på en DTF-analyse for et instrument med 6 items. 51
Figur 4. ICC for item 2. 53
Figur 5. DTF-analyse av ME-instrumentet. 54
Figur 6. ICC for item 8. 55
Figur 7. DTF-analyse for VE-instrumentet. 56
Figur 8. ICC for item 13. 57
Figur 9. DTF-analyse for OA-instrumentet. 58
Figur 10. ICC for item 22. 59
Figur 11. DTF-analyse for PFT-instrumentet. 60
Figur 12. Signifikansverdier for kildene blant lavtpresterende elever 64
Figur 13. Signifikansverdier for kildene blant høytpresterende elever 64
Figur 14. En visualisering av høyt- og lavtpresterende elevers gjennomsnittlige z-score, samt differansen i z-score for hver kilde. 65
Figur 15. Wright Map for ME-instrumentet (N=583, N_H=110, N_L=113) 67
Figur 16. Wright Map for VE-instrumentet (N=583, N_H=110, N_L=113) 68
Figur 17. Wright Map for OA-instrumentet (N=583, N_H=110, N_L=113) 69
Figur 18. Wright Map for PFT-instrumentet (N=583, N_H=110, N_L=113) 70

1.0 Introduksjon

Gjennom studieprogrammet «master i skolerettet utdanningsvitenskap med fordypning i matematikk og matematikkdidaktikk» har jeg fått innblikk i mye forskning som er gjort på matematikkundervisning. Teoretikere har i lang tid forsøkt å kaste lys over ulike sider ved matematikkundervisning og individers læring, for å skape en bevisstgjøring rundt både læreprosesser og undervisningsmetoder. Banduras teori om «self-efficacy», mestringsforventning, har interessert meg ekstra. Jeg kjenner meg på mange måter igjen i teorien. Som elev erfarte og opplevde jeg effekten av mestringsforventning som stor og verdifull, og dette er også noe jeg ser barn og unge har stor nytte av. Av den grunn alene finner jeg motivasjon for å undersøke nettopp dette konstruktet nærmere.

Min bakenforliggende motivasjon for å begynne på masterstudiet var ønsket om å få innsikt i forskning som er relevant for meg som lærer, og jeg bestemte meg derfor tidlig for at det jeg velger å forske på skal kunne gi meg nyttig og praktisk rettet informasjon, slik at det kan gjøre meg til en bedre og mer reflektert lærer i klasserommet. Teorien om mestringsforventning, og særlig kunnskap om hva som fører til økt mestringsforventning, tror jeg kan gi meg viktige praktiske holdepunkter i mitt fremtidige virke som lærer. Jeg har derfor valgt å sette søkelys på kildene til mestringsforventning i denne oppgaven. Det oppleves som viktig, også på grunn av matematikkfagets særtrekk. Matematikk blir ofte karakterisert som et «rett/galt»-fag, og som et fag mange enten elsker eller hater. Dette finnes det mange praktiske eksempler på, både i og utenfor skolen. Det kan virke som om matematikk, mer enn noe annet fag, er et emne mange mener de enten får til eller ikke får til. Jeg synes det derfor er ekstra interessant å undersøke elevers kilder til mestringsforventning nettopp i matematikkfaget, og tror at den praktiske betydningen av økt mestringsforventning blant elever kan være ekstra viktig i matematikk.

I samarbeid med forskningsgruppa SESAM (self-efficacy, its sources and achievement in mathematics), bestående av meg, to medstudenter og vår felles veileder, har datagrunnlaget for denne kvantitative oppgaven blitt samlet inn og analysert. Et slikt samarbeid muliggjorde et langt større antall deltakere i spørreundersøkelsen. Til tross for et felles datagrunnlag, har min og mine medstudenters masteroppgaver hatt ulike vinklinger og problemstillinger.

I min studie er det kildene til elevers mestringsforventninger i matematikk som står i fokus. Fordi forskere hevder at mestringsforventninger kan forklare en fjerdedel av variansen i akademisk prestasjon (Pajares, 2006), ser jeg dette temaet som svært relevant i forbindelse med

matematikkundervisning. Dette finner jeg støtte til i et stadig voksende forskningsfelt innen temaet. Med utgangspunkt i den forskningen som er publisert, legger Usher og Pajares (2008) frem noen føringer for videre forskning på kildene til mestringsforventning. De vektlegger at målet med forskningen må være å hjelpe elevene med å få det beste ut av kildene til mestringsforventning, slik at de kan bli agenter i egen psykisk helse. Med dette mener de at elevers predisponering for bestemte tankemønstre om egne evner ikke trenger å være til hinder for utfoldelsen. De viser til Banduras (1997) påstand om at mennesker, i alle fall delvis, er arkitekter av sine egne skjebner. Den nye læreplanen (LK20) (Kunnskapsdepartementet, 2017), med folkehelse og livsmestring som ett av tre tverrfaglige temaer, kaster et ekstra lys over aktualiteten av nettopp denne tematikken. LK20 tydeliggjør viktigheten av å både forstå og kunne påvirke faktorer som er av betydning for livsmestring, noe som gjør kildene til mestringsforventninger svært relevant å forske på. Bandura (1997) hevder også at mestringsforventninger er en nøkkelfaktor i menneskelig atferd, og forskning på hvordan man som lærer kan fremme eller hemme elevers mestringsforventning vil derfor være av stor nytte (Usher & Pajares, 2008).

I ethvert matematikklasserom kan en forvente at det faglige spriket er stort (Kunnskapsdepartementet, 2011). Det er derfor grunn til å tro at det er store forskjeller på hvordan ulike elever vil benytte eller ha tilgang på kildene til mestringsforventning. I denne masteroppgaven har jeg valgt å sammenligne høyt- og lavtpresterende elevers kilder til mestringsforventninger med utgangspunkt i følgende problemstilling:

«Er det forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning i matematikk, og i så fall hvilke?»

Masteroppgaven har følgende oppbygning: I kapittel 2.0 vil jeg redegjøre for teori og forskning gjort på mestringsforventning i skolen. Problemstillingen utdypes nærmere i kapittel 2.7. I kapittel 3.0 vil jeg forklare og begrunne metodiske valg for oppgaven. Dette knyttes videre inn i kapittel 4.0, der jeg analyserer den innsamlede dataen. I kapittel 5.0 diskuterer jeg funnene fra analysene med utgangspunkt i oppgavens problemstilling og tilhørende forskningsspørsmål. Helt til slutt vil jeg komme med noen avsluttende kommentarer i kapittel 6.0.

2.0 Teori og tidligere forskning

I denne delen av oppgaven vil jeg redegjøre for hva Bandura legger i begrepet mestringsforventning og dets kilder. Deretter vil jeg gå dypere inn i mestringsforventning i skole- og matematikksammenheng, og vise til ulik forskning som er gjort på feltet. Jeg vil også definere hva jeg legger i kategoriseringen av høyt- og lavt-presterende elever, før jeg med bakgrunn i teorien om mestringsforventning og tidligere forskning presenterer forskningsspørsmålene jeg ønsker å besvare i denne oppgaven.

2.1 Banduras mestringsforventning

Teorien om mestringsforventning ble utviklet av Albert Bandura. Han er en psykolog med et sosial-kognitivt ståsted, og hevder at kognitive prosesser har en viktig rolle i utviklingen av og opprettholdelsen av nye atferdsmønstre (Bandura, 1977). Ifølge Bandura er egne mestringsforventninger et viktig grunnlag for handling. Hvis man ikke tror at egne handlinger kan føre til et ønsket resultat, så har man lite insentiver til handling. Mestringsforventninger kan forstås som en persons tro på egne evner til å utføre de handlingene som kreves for å nå et ønsket resultat (Bandura, 1997). Det kan derfor sies å være et fremtidsrettet konstrukt.

Mestringsforventning er også situasjons – og kontekstavhengig (Bandura, 1997). Det betyr at høy mestringsforventning på ett område, ikke nødvendigvis medfører høy mestringsforventning på andre områder. Bandura påpeker derfor at dersom mestringsforventning skal ha en forklarende og predikativ effekt, så må målinger av mestringsforventning også være kontekstspesifikk. Han forklarer dette med at mestringsforventning er et multidimensjonalt konstrukt, og viser til tre dimensjoner som har viktige implikasjoner for mestringsforventningen. Disse tre er vanskelighetsgrad (*level)*, generalitet (*generality)* og styrke (*strength)*. De tre dimensjonene bidrar til å belyse hvordan mestringsforventninger bygges. Individer vurderer sin oppfattelse av egne evner opp mot oppgavens vanskelighetsgrad. Generalitet spiller også en rolle for hvordan individet vurderer mestringsforventning. Elever kan for eksempel ha ulike mestringsforventninger i matematikk og i samfunnsfag, eller i ulike temaer innenfor samme fag. Samtidig kan det være likhetstrekk mellom ulike oppgaver og fagområder som kan bidra til at mestringsforventningen er mer overførbar. Styrken på mestringsforventning er også av betydning. Svake mestringsforventninger blir lettere berørt av avkreftende erfaringer. Dette vil si at en elev som har svake mestringsforventninger fra før av, ofte kan påvirkes i større grad av for eksempel å mislykkes med en oppgave. Sterke

mestringsforventninger vil derimot være mer bestandige, selv med negative erfaringer (Bandura, 1997).

Videre skiller Bandura (1977) mellom to ulike typer forventning, nemlig de allerede omtalte mestringsforventningene og i tillegg ulike resultatforventninger (outcome expectations). Resultatforventninger er en elevs antagelse om at en bestemt atferd vil føre til et bestemt resultat. Resultat- og mestringsforventninger er adskilt i Banduras modell, fordi en person kan tro at en bestemt handling vil føre til et bestemt resultat, men likevel tvile på sine egne evner til å faktisk utføre handlingen, og dermed velge å ikke utføre handlingene. Resultatforventningene alene vil altså ikke nødvendigvis påvirke atferden. Bandura (1977) gjør det klart at mestringsforventning påvirker både igangsettelse av handlingen, samt personens utholdenhet og innsats i den aktuelle handlingen. I denne oppgaven er det mestringsforventninger jeg vil vie oppmerksomhet, der jeg vil se nærmere på hva som fører til mestringsforventning.

Ifølge Bandura (1997) finnes det fire kilder til mestringsforventning. Disse er mestringserfaringer (enactive mastery experiences), vikarierende erfaringer (vicarious experiences), oppmuntring, støtte og overtalelse fra andre (verbal persuasion) og psykologiske og fysiologiske tilstander (physiological and affective states). De norske oversettelsene er hentet fra Wæge og Nostrati (2018). Nedenfor vil jeg redegjøre for hva Bandura (1977, 1997) legger i hver av kildene.

2.1.1 Mestringserfaringer

Mestringserfaringer brukes som en indikator på egne evner, og er dermed en kilde til mestringsforventning (Bandura, 1997). Denne kilden regnes som den mest innflytelsesrike kilden til mestringsforventning, fordi tidligere erfaringer gir et autentisk bilde av om man faktisk mestrer eller ikke. Gjentatt suksess kan bidra til høyere mestringsforventning, mens gjentatte nederlag vil kunne skape en lavere mestringsforventning. Likevel er det noen faktorer som påvirker hvordan erfaringene tolkes. Bandura (1977) trekker frem timingen, altså når i forløpet nederlaget kommer, og det totale mønsteret av mestringserfaringer som viktige for tolkningen. Hvis man mislykkes på et tidlig tidspunkt, før man har rukket å etablere en solid mestringsforventning, vil den negative erfaringen kunne få større betydning for tolkningen av egne evner. Opplever man derimot nederlag etter å ha utviklet en sterk følelse av mestring gjennom gjentatte mestringsopplevelser, antas det at den negative virkningen vil reduseres. Dessuten vil man etter å ha fått troen på egne evner, lettere holde ut i møte med motgang, og lettere komme tilbake igjen ved et tilbakefall. Bandura (1997) trekker likevel frem at dersom man kun opplever veien til suksess som lett, kan det skape en forventning om å alltid komme

fort frem til suksess (Bandura, 1997). Dette kan medføre at man lett kan bli motløs av å streve eller mislykkes. En stabil følelse av mestringsforventning oppnås gjennom erfaringer i å overvinne utfordringer gjennom vedvarende innsats.

Endringer i oppfattet mestringsforventning avhenger av hvordan individer tolker mestringserfaringene. Erfaringene i seg selv er ikke nødvendigvis avgjørende for mestringsforventningen. Samme grad av oppnådd suksess kan derfor påvirke elevers mestringsforventning på ulik måte (Bandura, 1997). Det er flere faktorer som påvirker hvorvidt mestringserfaringer vil bidra til å øke mestringsforventningen. Personers forhåndsoppfatning av egen mestringsforventning har betydning, samt vanskelighetsgrad på oppgaven, mengden hjelp mottatt, situasjonen rundt prestasjonen, det totale mønsteret av nederlag og suksess og hvordan erfaringene kognitivt er organisert og strukturert i hukommelsen. Av dette blir det tydelig at erfaringen i seg selv ikke nødvendigvis påvirker personens mestringsforventning. Kjennskap til hvordan de ulike faktorene påvirker den kognitive prosesseringen av en erfaring, vil kunne bidra til at man kan få mest mulig ut av sine mestringserfaringer (Bandura, 1997). Av den grunn ønsker jeg å utdype disse faktorene i avsnittene nedenfor.

Eksisterende mestringsforventninger

Allerede eksisterende mestringsforventninger hos individer er av betydning for hvordan nye mestringserfaringer tolkes og påvirker mestringsforventningen (Bandura, 1997). Bandura trekker frem den kognitive konstruksjonen som sentral. Individer strukturerer erfaringer i skjemaer i hodet, og disse skjemaene påvirker igjen hva man ser etter og hvordan man tolker og organiserer informasjon fra omgivelsene. Hvordan nye erfaringer passer inn i allerede eksisterende skjemaer avhenger av arten og styrken til erfaringene. På grunn av dette kan man si at mestringsforventninger både er produkter av og konstruktører av mestringserfaringer. Erfaringer som ikke passer inn med allerede eksisterende skjemaer har en tendens til å bli forminsket, svekket eller glemt. Erfaringer som er i samsvar med eksisterende skjemaer blir derimot lettere husket og blir tillagt større betydning. Allerede eksisterende mestringsforventninger kan være forklarende faktorer for hvorfor gjentatt suksess vil påvirke mestringsforventningene.

Vanskelighetsgrad og kontekstuelle faktorer

Hvordan man tolker en mestringserfaring vil som nevnt avhenge av hvilken oppfatning man har av vanskelighetsgraden på oppgaven (Bandura, 1997). Konteksten oppgaven utføres i vil også påvirke tolkningen av nederlag eller suksess. Har man oppnådd suksess med en oppgave

i en kontekst der man fikk hjelp, vil man mest sannsynlig tillegge skylden for suksessen til de ytre faktorene og ikke til egne evner. Har man utført en oppgave under ugunstige forhold vil et nederlag kanskje få mindre betydning for mestringsforventningen enn dersom forholdene hadde vært optimale. Jo flere faktorer som er involvert, som ikke kan sies å knyttes til en persons evne, desto mindre sannsynlig er det at det vil påvirke personens mestringsforventning. En person som er overbevist om sin svake mestringsevne, vil altså forsøke å finne alle mulige grunner til at en suksessfull erfaring kan tilskrives andre faktorer enn egen evne. I slike tilfeller skal det mange bekreftende opplevelser og suksess under mange ulike omstendigheter til for å kunne skape varige endringer i egen mestringsforventning. En mestringserfaring vil i et slikt tilfelle bli nøye prøvet i andre situasjoner, før individet faktisk tilskriver suksessen til egne evner.

Innsats

Mengden innsats lagt ned i en oppgave, vil påvirke individets oppfattelse av om suksessen skyldtes egne evner (Bandura, 1997). Millers (1984, gjengitt i (Bandura, 1997)) har rapportert at barn ser på høy innsats som en måte å øke sine egne evner på, mens voksne ser det som et tegn på lav evne å måtte legge ned mye innsats for å få til noe. Innsats kan også bli sett på som en kompensasjon for begrensede oppmerksomhets- og minnelagringsprosesser i hjernen.

Prosessen med å utvikle en ferdighet

Veien til å utvikle en ferdighet er en prosess, der man i perioder kan gjøre mye framgang, mens framgangen tilsynelatende er liten i andre perioder (Bandura, 1997). Bandura nevner for eksempel at det kan være sjeldnere å gjøre store fremskritt mot slutten av prosessen. Dette fordi kravene er vesentlig høyere, og fordi store feilkilder allerede er eliminert. Historikken og mønsteret for utviklingen av en ferdighet vil komme til å prege mestringsforventningen. Har man opplevd en periode med mye nederlag, men samtidig ser at fremgang skjer sakte men sikkert over tid, kan likevel mestringsforventningen øke. Andre kan lykkes, men samtidig se at fremgangen avtar i forhold til tidligere utvikling, og det kan bidra til å senke mestringsforventningen. Bandura trekker også frem faren ved å kun se på det nåværende stadiet som grunnlag for oppfattelse av mestringsforventning. Særlig unge, som er mindre utviklet enn voksne, har vanskeligheter med å vurdere egen mestring ut ifra et lengre tidsperspektiv, og vil ha en tendens til å tillegge de siste og nyeste erfaringene stor vekt.

2.1.2 Vikarierende erfaringer

Vikarierende erfaring er en annen kilde som kan bidra til økt mestringsforventning. Wæge og Nosrati (2018) skriver om vikarierende erfaringer at det å se en annen person lykkes med en oppgave, kan føre til at individet selv får en økt overbevisning om å lykkes med en tilsvarende

oppgave. Vurderingen vil skje basert på om individet ser likhetstrekk til personen som gjennomfører oppgaven og individets oppfattelse av personenes evner.

Bandura (1997) tydeliggjør bakgrunnen for denne type stedfortredende erfaring og trekker frem modellering som et verktøy for å fremme mestringsforventning. Jo mer individet kan identifisere seg på modellen, jo mer overbevisende oppleves det at individet selv kan få til samme oppgave. Det å se en annen person mestre en oppgave, kan gi en selv overbevisning om å lykkes med det samme. Det å se en annen person legge ned stor innsats i noe, men likevel ikke lykkes med det, kan derimot senke egne mestringsforventninger.

Modellering ved hjelp av stedfortredende mestringserfaringer kan også være positivt fordi individet som ser på lærer effektive strategier og tenkemåter (Bandura, 1997). På den måten kan modellering fungere for å overføre kunnskap. Vikarierende erfaring kan også ha en motiverende effekt, hvis individet får se at modellen viser utholdenhet og optimisme i møte med utfordringer. Vikarierende erfaring kan også påvirke mestringserfaringen ved at observatøren får en endret oppfatning av oppgavens vanskelighetsgrad. I mange tilfeller er mennesker avhengige av å se til andre og hva andre oppnår, for å kunne vurdere seg selv. Noen aktiviteter har kanskje mer standardiserte normer for hva som regnes som bra. Andre aktiviteter krever at man sammenligner egne resultater med andre for å kunne vurdere om resultatet er bra eller dårlig. På denne måten kan vikarierende erfaring være en indikator på oppgavens vanskelighetsgrad og endre individets oppfatning av den, som igjen kan påvirke mestringserfaringen.

Til tross for at individer stort sett får vikarierende erfaringer ved å se til andre personer de omgås, slik som foreldre og klassekamerater, kan vikarierende erfaringer også skje i form av at individet sammenligner seg selv med sine tidligere erfaringer (Usher & Pajares, 2008). Dette kan skje enten kognitivt, ved at personen kan forestille seg hvordan han/hun handlet ved en tidligere lignende anledning, eller ved hjelp av videoopptak. Til tross for at Bandura (1997) klassifiserte det å se seg selv utføre en oppgave som en vikarierende erfaring, vil slike erfaringer være tett knyttet opp mot mestringserfaringer.

Bandura (1997) trekker generelt mestringserfaringer frem som en sterkere kilde til mestringsforventning enn vikarierende erfaring, da førstnevnte er en direkte erfaring. Likevel påpeker han at det finnes omstendigheter som gjør at vikarierende erfaring kan overgå mestringserfaringer. Har en person mange nederlagserfaringer, kan påvirkningen fra en vellykket vikarierende erfaring medføre at de direkte erfaringene blir tillagt mindre vekt og får

mindre betydning. Innsats kan tilsvarende bli tillagt mer vekt, noe som vil kunne virke positivt inn på videre erfaringer.

2.1.3 Oppmuntring, støtte og overtalelse fra andre

Oppmuntring fra andre er en utbredt metode for å påvirke menneskelig atferd. Mennesker kan bli påvirket av andre til å få tro på å kunne mestre noe som de tidligere ikke har lykkes med (Bandura, 1977). Det at signifikante andre uttrykker tro på at en skal klare noe, kan bidra til økt mestringsforventning. Lærere og foreldre kan for eksempel overbevise et barn om at det kommer til å lykkes med en oppgave (Wæge & Nosrati, 2018). Verbal overtalelse alene er kanskje ikke det beste verktøyet for å skape en langvarig endring i en persons mestringsforventning, men Bandura (1997) ser likevel verdien overtalelse har i form av en slags midlertidig motivasjon. I samspill med andre kilder kan dette bidra til å skape høyere mestringsforventning. Den verbale oppmuntringen kan føre til at individet nedlegger større innsats i oppgaven, som igjen kan føre til en faktisk mestringserfaring. Det er viktig at oppmuntring, støtte og overtalelse fra andre er realistisk (Wæge & Nosrati, 2018). Hvis den ikke er det, kan det resultere i falske forhåpninger for individet. Hvis personen da ikke lykkes med oppgaven, kan det tvert imot bidra til å svekke mestringsforventningen. Av denne grunn vil jeg i avsnittet nedenfor trekke frem noen farer ved bruk av oppmuntring, støtte og overtalelse.

Tilbakemeldinger som fremhever individets evner kan bidra til en «boost» i mestringsforventningen (Bandura, 1997). Tilbakemeldinger som omhandler evner er særlig viktig i de tidlige stadiene av å utvikle en ferdighet (Schunk, 1984, gjengitt i (Bandura, 1997). Tilbakemeldinger som forteller at man har forbedret sine evner gjennom innsats, kan også øke mestringsforventningen. Bruk av tilbakemeldinger som roser innsatsen er temmelig vanlig og kanskje også foretrukket, men Schunk og Rice (1986, gjengitt i (Bandura, 1997)) påpeker likevel at gjentagende ros for innsats kan medføre en oppfattelse av at en slik uendelig stor innsats må bety liten evne. Schunk (1983, gjengitt i (Bandura, 1997)) har også funnet at det å fortelle en person at den har evner og at de har utviklet evnen gjennom innsats medfører lavere mestringsforventning enn det å fortelle personen at fremgangen viser at de har evner, uten å si noe om innsatsen. Bandura trekker også frem en annen fare, nemlig tendensen til å rose personer med begrensede evner indirekte og uklart, eller senke forventningene betraktelig. Slike personer blir ofte rost mye for middelmådige prestasjoner, og får ofte mer skryt enn andre som gjør det like bra. Bandura refererer til forskning som sier at slik respons har en tendens til å senke

oppfatningen av egen evne (Lord, Umezaki & Darley, 1990; Meyer, 1992, gjengitt i (Bandura, 1997).

2.1.4 Psykologiske og fysiologiske tilstander

Den fjerde kilden til mestringsforventning har sitt opphav i at individer ofte tolker ulike psykologiske og fysiologiske reaksjoner på stressende eller krevende situasjoner som et tegn på lav evne (Bandura, 1997). Dette kommer av at kraftige psykologiske og fysiologiske reaksjoner kan svekke prestasjonen, og mennesker vil derfor oftere forvente suksess når de ikke er plaget med slike reaksjoner. Forventningene om at slike reaksjoner dukker opp kan føre til en enda sterkere følelse av stress, og nettopp svekke prestasjonen, slik som fryktet. Den negative påvirkningskraften fra denne fjerde kilden til mestringsforventningen vil kunne endres ved å forsøke å redusere stressnivået og de negative fysiske og følelsesmessige reaksjonene, samt å korrigere feiltolkningen av at slike reaksjoner er et tegn på lav evne.

Bandura (1997) påpeker at det i likhet med de andre kildene er individets tolkning av reaksjonene gjennom kognitive prosesser som er avgjørende for mestringsforventningen. Han nevner flere faktorer som kan virke inn på denne prosessen. Det kan være ulike faktorer knyttet til selve situasjonen som gir individet gode forklaringer på hvorfor de fysiske reaksjonene fant sted. For eksempel kan en som holder en tale tillegge temperaturen og ventilasjonen i rommet skylden for svetting, uten å la en slik reaksjon bli et tegn på lav evne. Hvor sterke de psykologiske og fysiologiske reaksjonene er, kan også ha betydning for fortolkningen av dem. Dette fordi sterke reaksjoner lettere kan forstyrre selve handlingen. Den eksisterende mestringsforventningen til et individ kan også spille en rolle for hvor sensitiv individet er for psykologiske og fysiologiske reaksjoner. I tillegg trekker Bandura (1997) frem humør som en faktor som påvirker mestringsforventningen, fordi humørtilstander ofte følger funksjonaliteten.

2.2 Forskning på mestringsforventning i en skolekontekst

Studier viser at mestringsforventning er en viktig faktor for å lykkes på alle akademiske nivåer (Schunk, 2011, gjengitt i Yurt, 2014). Når jeg nå vil presentere ulik forskning gjort på mestringsforventning i skolen, velger jeg å se bort fra forskning gjort på lærerens mestringsforventning. I denne studien er det elevenes mestringsforventning som er av interesse.

Med bakgrunn i Banduras (1997) teori, viser forskning at elever med høy mestringsforventning vil velge mer utfordrende oppgaver og legge ned større innsats enn det en elev med lav mestringsforventninger vil gjøre (Wæge & Nosrati, 2018), noe som i seg selv understreker viktigheten av å ha høye mestringsforventninger. Mye forskning som er gjort på akademisk

mestringsforventning, bekrefter også Banduras teori (1997) om mestringsforventningens predikative effekt på akademiske prestasjoner (Galyon, Blondin, Yaw, Nalls & Williams, 2012; Salazar & Hayward, 2018; Usher & Pajares, 2008).

Mestringsforventninger er imidlertid et komplekst konstrukt som det finnes mye forskning på innenfor skolekontekst. Zimmerman (1995) trekker frem tre faktorer jeg velger å bruke som en inngang til forskningsfeltet. For det første handler mestringsforventning om å vurdere egen evne til å utføre en gitt oppgave, og ikke generelt vurdere personlige kvaliteter og karaktertrekk. For det andre er mestringsforventning tett bundet til det enkelte fagområdet. En elevs mestringsforventning i matematikk kan være svært forskjellig fra den samme elevens mestringsforventning i andre skolefag. Dette stemmer jo godt overens med Banduras (1997) beskrivelse av mestringsforventning som et kontekst- og situasjonsavhengig konstrukt, som tidligere nevnt. Zimmermans (1995) tredje faktor bygger videre på det samme, nemlig at det kan være andre situasjonsavhengige faktorer som påvirker mestringsforventningen til elever, som medfører at måling av mestringsforventning alltid er kontekstspesifikk. For eksempel kan elevenes mestringsforventning variere ut ifra om de befinner seg i en prøve-kontekst, i et konkurransepreget klasserom, eller i en rolig arbeidsatmosfære. Disse ulike faktorene spiller en rolle for hvordan vi tolker forskningsresultater, og medfører at man ikke ukritisk kan sammenligne resultater fra ulike studier.

Siden Bandura fremla sitt teoretiske rammeverk om mestringsforventning på 70-tallet, har det blitt forsket mye på kildene til konstruktet, både gjennom kvantitative og kvalitative studier (Chen, 2003; Mozahem, Boulad & Ghanem, 2020; Usher & Pajares, 2008; Zimmerman, 1995). Bandura (1997) introduserte mestringserfaringer som den viktigste av de fire kildene. Han initierte imidlertid ikke noe rekkefølge på de tre andre kildene (Usher & Pajares, 2008). Han hevder at det kan være vanskelig å sile ut effekten av en enkelt kilde, da informasjon fra en kilde kan påvirke effekten av de andre kildene. Dette har vi allerede sett eksempler på ved utledningen av de ulike kildene over, der vi så at verbal oppmuntring kan bidra til større innsats, som igjen kan føre til mestringserfaringer. Når Usher og Pajares (2008) i sin review-artikkel går igjennom forskning gjort på kildene til mestringsforventning i skolen fra 1977 til 2008, finner de at mestringserfaringer fremkommer som den viktigste kilden til mestringsforventning uavhengig av fagområde. Funnene er altså overensstemmende med Bandura, som begrunnet det med at mestringserfaringer er det mest autentiske målet på egen evne (Bandura, 1997). Men rekkefølgen og «viktigheten» av de tre andre kildene fremkommer som varierende i ulike studier (Usher & Pajares, 2008). For eksempel har vikarierende erfaringer i noen studier ikke

vist seg å ha noen predikativ effekt på mestringsforventningen (Gainor & Lent, 1998; Lent et al., 1991; Pajares et al., 2007; Usher & Pajares, 2006, gjengitt i (Usher & Pajares, 2008)). Andre studier har imidlertid funnet en korrelasjon mellom vikarierende erfaringer og mestringsforventning i enkelte studentgrupper, for eksempel elever med lærevansker (Hampton, 1998, gjengitt i (Usher & Pajares, 2008)). Lopez et al. (1997) rapporterte at oppmuntring, støtte og overtalelse fra andre hadde en predikativ effekt på mestringsforventning innen avansert algebra, men ikke innen geometri. Studienes variasjon i kildenes interne relasjoner og forutsigbarhet begrunnes av Usher og Pajares (2008) med de kontekstuelle faktorene, samt metodiske begrensninger.

Siden forskningsfeltet på internasjonal basis er stort, og Zimmerman (1995) påpeker at mestringsforventning er tett knyttet til det spesifikke fagfeltet, begrenser jeg meg til den matematikkfaglige konteksten når jeg i neste delkapittel ser nærmere på forskning gjort på kildene til mestringsforventning.

2.3 Forskning på mestringsforventning i matematikk

Slik flere forskere har påpekt, spiller kontekst en rolle for hvordan mestringsforventninger skapes og oppfattes (Bandura, 1997; Haciomeroglu, 2019; Zimmerman, 1995). For å belyse denne studiens kontekst, vil jeg aller først si noe om matematikkfagets særtrekk. Videre vil jeg redegjøre for forskning som er gjort på mestringsforventning i matematikk, og kildene til mestringsforventning i matematikk. Jeg vil også vise til forskning gjort i norsk kontekst på temaet.

2.3.1 Matematikkfagets særtrekk

Matematikkfaget i skolen har gjennom tiden hatt en tendens til å dreie seg om å komme frem til rett svar (Stodolsky, 1985). Skovsmose (1998) tilskriver den matematiske tradisjonen med begrepet *oppgaveparadigmet*, og beskriver en matematikkundervisning sentrert rundt oppgaven. Han påstår at målet i den tradisjonelle matematikkundervisningen har vært å få avklart matematiske forhold, for på den måten å besvare bestemte oppgaver korrekt. Prosessen frem mot det riktige svaret har ofte vært preget av en oppfatning av at læreren er eksperten som besitter kunnskapen om hvordan man kommer dit (Lampert, 1990). Undervisningsformen har typisk vært at læreren først forklarer og at studentene deretter gjør oppgaver. En slik undervisningsform kan føre til en implisitt forventning om at matematikk læres fra en autoritet og deretter kan praktiseres. Elevene kan få en forståelse av at det å gjøre matematikk betyr å følge et sett med regler fastsatt av læreren, og at matematikk handler om å huske reglene. Det

kan også skape en forventing om at matematiske sannheter er styrt av lærerens evaluering av svaret. Vi kan si at slike forventinger skaper et bilde av at det er læreren som har eierskapet til matematikken, og matematikkundervisning derfor blir kunnskap og ferdigheter som læreren porsjonerer ut til elevene i passende mengde (Stigler & Hiebert, 2009). Sett fra et slikt perspektiv, er det forståelig at matematikken i skolen for mange har blitt et fag som dreier som om å komme fortest mulig til rett svar, og skapt et «rett/galt» – fokus.

De ulike måtene å undervise matematikk på har med tiden fått noen merkelapper, som både kan bidra til å forklare og bevisstgjøre. En lærer med «transmissionist»-undervisning har en tendens til å være mer lærersentrert, mens en lærer med «connectionist»-undervisning er mer sentrert på elevenes forståelse eller på å skape sammenheng (Williams et al., 2010). De to undervisningsstilene blir forklart ved å vise til undervisningens motiver og mål. For en «transmissionist» ligger motivet for undervisningen på selve skoleringen, der eksamensforberedelser, prestasjoner og karakterer er målet med undervisningen. Dette er også faktorene læreren motiverer elevene med, og som ligger til grunn for den matematiske aktiviteten i klasserommet. Denne type undervisning kan også relateres til matematikkens eierskap, nevnt i avsnittet over. Elever under denne type undervisning vil ofte se på læreren som eksperten, og tilkalle denne for å få hjelp når de står fast. Undervisningen er også ofte fokusert på rette svar. Williams et al. (2010) trekker frem «connectionist»-undervisning som en motsetning. Slik undervisning er rettet mot å bygge elevenes konseptuelle forståelse, og motivet for undervisningen er å skape forståelse for matematikken. Undervisningen blir da lagt opp slik at ny kunnskap kan kobles på det elevene har lært tidligere. Elevene blir oppfordret til å diskutere sine løsningsforslag i grupper og helklasse. Det å ta feil blir sett på som en mulighet til læring, eller en metode for å vise frem mange ulike løsningsmetoder. Til tross for at begrepene her er skissert opp som to typiske ytterpunkter, kan de ikke tolkes helt unyansert. Williams et al. (2010) påpeker nemlig at man i praksis ser en større kontinuitet mellom disse ytterpunktene, slik at selv den mest typiske «transmissionist» også vil arbeide for at elevene skal forstå matematikken, og at selv den mest typiske «connectionist» også vil være bekymret over å rekke igjennom pensumet før eksamen. Disse merkelappene for matematikkundervisning gir oss imidlertid et nyttig analytisk verktøy, og kan bidra til å bevisstgjøre lærere rundt egne holdninger til matematikken.

Skemp (1976) introduserer begrepene instrumentell og relasjonell forståelse, og forklarer gjennom det at matematikk er mer enn å komme til rett svar. Han belyser grunnen til at mange elever kan sitte med den feilaktige forståelsen av at selve målet i matematikkfaget er å komme

frem til rett svar. Skemp (1976) skiller mellom matematikk som fører til instrumentell forståelse og matematikk som fører til relasjonell forståelse. Instrumentell matematikk består typisk av et sett med regler og metoder som elevene kan benytte seg av for å komme til rett svar. Relasjonell matematikk handler om å bygge opp en konseptuell struktur, eller skjemaer, som eleven selv kan utforme regler fra. Vi kan si at relasjonell matematikk innebærer en underliggende forståelse for sammenhenger i faget, og samsvarer bedre til «connectionist»-undervisningen beskrevet ovenfor. Skemp (1976) påpeker imidlertid at begge formene for forståelse er viktig å fremme i matematikkundervisningen. Likevel tydeliggjør han faren ved at elevene sitter med en forventning om at forståelse kun innebærer den instrumentelle forståelsen, mens læreren egentlig underviser elevene for å fremme relasjonell forståelse, eller motsatt.

Vi ser også av LK20 (Kunnskapsdepartementet, 2019) at det er et ønske å komme bort fra den mer tradisjonelle matematikkundervisningen med et sterkt resultatfokus. Vektleggingen av matematikk som innebærer abstrahering og generalisering gir et solid grunnlag for å undervise for relasjonell forståelse i matematikkfaget. Læreplanens begreper som reflektere, resonnere, argumentere, samt utforsking, problemløsning, dybdelæring, viser at matematikkfaget er i endring. Dette står i motsetning til Skott, Skott, Jess & Hansen (2018) sine beskrivelser av den tradisjonelle skolematematikken med et sterkt resultatfokus. I LK20 (Kunnskapsdepartementet, 2019) beskrives det nå eksplisitt at *«…elevene skal legge mer vekt på strategiene og fremgangsmåtene enn på løsningene»*. Til tross for at vi også i LK06 (Kunnskapsdepartementet, 2006) så at forståelse, problemløsning og utforskende aktiviteter fikk større plass enn tidligere, ble det likevel ikke satt eksplisitt lys på betydningen av den matematiske prosessen. Det er altså tydelig at matematikkfaget er i endring.

Det mer tradisjonelle synet på matematikkfaget vil nok imidlertid fortsatt kunne være en vanlig oppfatning. Siden matematikkfaget for mange elever kan bære et preg av å være et «rett/galt»-fag, kan det tenkes at mestringsforventning er et særlig relevant tema i matematikkfaget. Det kan for eksempel tenkes at elever kan undervurdere sine egne evner i matematikk dersom de kommer frem til feil svar til tross for at fremgangsmåten er korrekt.

2.3.2 Kildene til mestringsforventning i matematikk

Det er de siste tiårene gjort en del forskning på kildene til mestringsforventning i matematikk, og da særlig forskning på de ulike kildenes påvirkning på mestringsforventning (Phan, 2012; Usher, 2009), samt kildenes predikative effekt på matematisk prestasjon (Chen, 2003; Phan, 2012; Yurt, 2014). Forskning omhandlende mestringsforventning i matematikk belyser, i likhet med forskning på mestringsforventninger i skolen generelt, mestringserfaringer som den

viktigste kilden til mestringsforventning (Usher, 2009). Nettopp av denne grunn blir lærere oppfordret til å maksimere mulighetene for autentiske mestringsopplevelser i matematikkfaget, uansett om de er små eller store (Usher, 2009). Usher påpeker at dette krever tålmodighet fra læreren, og at det ikke må gjøres overfladiske forsøk på å skape mestring ved å gi enkle oppgaver som ikke utfordrer.

Nyere forskning på kildene til mestringsforventning i matematikkfaget viser imidlertid samme tendens som nevnt tidligere; kildenes rekkefølge og predikative effekt fremkommer som varierende i ulike studier. Yurt (2014) setter opp følgende rekkefølge på kildene til mestringsforventning med bakgrunn i hans studie: mestringserfaringer, oppmuntring, støtte og verbal overtalelse, fysiologiske og psykologiske tilstander og til slutt vikarierende erfaringer. Mens de tre førstnevnte kildene har en predikativ effekt på den matematiske prestasjonen, har vikarierende erfaringer ikke det. Forfatteren viser til tidligere forskning der det også fremkommer at sammenhengen mellom denne kilden og prestasjon er svak i forhold til de andre kildene (Usher & Pajares, 2006, Lopez & Lent, 1992, Stevens et.al, 2007, gjengitt i (Yurt, 2014)). I studien til Yurt (2014) blir dette forklart med at studiens metode til en viss grad kan ha vært til hinder for gode vikarierende erfaringer, fordi elevene ikke møtte situasjoner der de kunne observere sine rollemodellers matematiske strategier, metoder og teknikker. Dessuten argumenterer forfatteren for at den indirekte erfaringen en elev får ikke har en direkte påvirkning på matematisk prestasjon. Yurt (2014) nevner derimot muligheten for at vikarierende erfaringer indirekte kan ha påvirket matematisk prestasjon gjennom de andre kildene. Dette bekrefter i så fall den tidligere nevnte teorien til Bandura (1997) om at informasjon fra en kilde kan påvirke de andre kildene, og at det dermed er vanskelig å sile ut effekten fra en enkelt kilde.

Det er flere forskere som trekker inn tidligere matematikkprestasjoner som en predikativ faktor for mestringserfaring (Lopez & Lent, 1992; Yildiz & Özdemir, 2019). Lopez og Lent (1992) fant blant annet en korrelasjon mellom mestringsforventning i matematikk og tidligere karakter i matematikk, men understreker at den predikative effekten for mestringsforventning ble betydelig forsterket når de også tok med mestringserfaringer i beregningen.

Usher og Pajares (2008) bemerker at kvantitative studier ennå ikke har funnet positive virkninger av psykologiske og fysiologiske tilstander. Senere forskning vitner imidlertid om delte meninger om dette. En kvalitativ studie påpekte at elever med høy mestringsforventning ble motivert av den fysiske reaksjonen, mens elever med lav mestringsforventning tolket den fysiske reaksjonen som et negativt tegn (Usher, 2009). En annen kvalitativ studie viser til at

elever med høy mestringsforventning tar negative fysiologiske og psykologiske reaksjoner mer på alvor enn elever med lav mestringsforventning (Gao, 2019). Forfatteren peker på muligheten for at dette kan skyldes at elever med høy mestringsforventning opplever disse reaksjonene sjeldnere og ikke er så vant til å kjenne på slike reaksjoner. Studien viser at elever med høy mestringsforventning opplever positive emosjoner (f.eks. stolthet, spenning og tilfredsstillelse) oftere enn negative emosjoner (f.eks. irritasjon og frustrasjon, tretthet) (Gao, 2019). Yildiz og Özdemir (2019) finner i deres kvalitative studie også eksempler emosjoner som virker både positivt og negativt inn på elevenes mestringsforventning i matematikk.

Gao (2019) hevder i sin studie av kinesiske elevers kilder til mestringsforventninger at mange tidligere studier har satt søkelys på om kildene predikerte eller korrelerte med mestringsforventning i matematikk, og peker på at kompleksiteten til mestringsforventning og dets underliggende mekanismer kanskje har blitt noe oversett. Med en mixed method tilnærming finner forfatteren forskjeller i kildenes effekt på elever med høy og lav mestringsforventning. Funnene peker på at elever med høy mestringsforventning har en klarere formening om kildenes innflytelse på dem, mens elever med lav mestringsforventning har svært forskjellige synspunkt på dette. Dette ligger til grunn for Gaos (2019) antakelse om at utvikling av mestringsforventning er svært komplekst. Forfatteren peker på at kontekstuelle og individuelle faktorer bør få større oppmerksomhet fra lærere som vil fremme sine elevers mestringsforventning.

2.3.3 Forskning på mestringsforventning i matematikk i norsk kontekst

Det er gjort relativt lite forskning på elevers mestringsforventninger i matematikkfaget i Norge. Av de få studiene som finnes er det ingen som setter søkelys på kildene til mestringsforventning. Likevel vil jeg i dette delkapittelet presentere forskningen som er gjort på mestringsforventning i matematikk generelt, samt belyse behovet for mer forskning på mestringsforventning i matematikk blant norske elever.

Skaalvik og Skaalvik (2011) har gjennom to longitudinelle studier forsket på sammenhengen mellom selvoppfatning og prestasjoner i matematikk. Forfatterne definerer selvoppfatning til å bestå av selvbilde og mestringsforventning. Studiens datagrunnlag er kvantitative undersøkelser som ble gjort ved tre ulike tidspunkt i løpet av ett år. Skaalvik og Skaalvik (2011) finner i begge studiene at selvoppfatninger er viktige predikatorer for prestasjoner. I den ene studien fanget forskerne overgangen mellom ungdomsskole og videregående skole, og fant at sammenhengen mellom selvoppfatning og prestasjon var ekstra sterk etter at elevene begynte på videregående skole. Forskerne foreslår at dette kan komme av at selvoppfatning spiller en

ekstra viktig rolle når konteksten er ny. Oppløsningen av tidligere statiske roller og posisjoner kan gjøre at elevene kjenner færre forventinger fra medelever og lærere. Elever kan da føle at de får en «ny start» og vil dermed kunne få større tro på seg selv og sin mulighet til å utvikle egne evner. Skaalvik og Skaalvik (2011) foreslår for fremtidige studier å se nærmere på de prosessene som gjør at elevenes selvoppfatning påvirker prestasjonene. De foreslår også at fremtidige studier ser på emosjonelle variabler slik som bekymring og motivasjon, for eksempel gjennom innsats, utholdenhet og selvregulert læring.

Skaalvik, Federici og Klassen (2015) gjorde en kvantitativ studie for å undersøke hvorvidt emosjonell støtte fra lærere og elevers mestringsforventning er predikatorer for motivasjon, samt hvordan dette påvirker forholdet mellom studentenes karakterer og ulike mål på matematisk motivasjon. Studien skulle også kontrollere for om tidligere prestasjoner var det som påvirket motivasjon for skolearbeidet. Gjennom sin forskning fant Skaalvik et al. (2015) at mestringsforventning korrelerte positivt med alle mål på motivasjon, både indre motivasjon, innsats, utholdenhet og hjelp-søkende atferd. Det samme gjorde emosjonell støtte fra læreren. Mestringsforventning var en av de største predikatorene for indre motivasjon og utholdenhet ved arbeid med utfordrende oppgaver, mens emosjonell støtte var den største predikatoren for hjelp-søkende atferd. Forfatterne foreslår at den sterke sammenhengen mellom mestringsforventning og ulike mål på motivasjon muligens kan forklare hvorfor mestringsforventning synes å ha innvirkning på matematisk prestasjon. Forfatterne gjør dessuten en interessant refleksjon rundt sammenhengen mellom mestringsforventning og utholdenhet. De stiller spørsmål ved om denne sammenhengen kan skyldes at mestringsforventning er knyttet til ferdigheter, som igjen kan medføre at elever med gode akademiske ferdigheter ikke på samme måte behøver å holde ut like lenge som elever med dårligere akademiske ferdigheter (Schunk, 1991, gjengitt i (Skaalvik et al., 2015)). Skaalvik et al. (2015) oppfordrer til mer empirisk forskning på denne sammenhengen, samtidig som de påpeker et behov for mer forskning omhandlende prosessene som forårsaker mestringsforventningens påvirkning på prestasjon.

Street, Malmberg og Stylianides (2017) har også gjennomført en studie om mestringsforventning. Studien har et noe mer teoretisk perspektiv, og handler om hvordan ulike sider ved elevers mestringsforventning kan måles. De tar utgangspunkt i Banudras (1997) tre dimensjoner av mestringsforventning som er beskrevet tidligere; vanskelighetsgrad, generalitet og styrke. Med utgangspunkt i disse tre dimensjonene utvikler Street, Malmberg og Sylianides (2017) et instrument; *Self-Efficacy Gradations of Difficulty Questionaire*. Studien er gjort på

norske 5., 8. og 9.klassinger, der de sammenligner resultatene med resultatene fra en nasjonal prøve i regning. Resultatet av studien viser at elevers opplevde vanskelighetsgrad spiller en rolle for hvordan de vurderer egen mestringsforventning og at dette derfor også bør hensyntas i videre studier.

Det er altså ikke mye forskning på mestringsforventning i norsk kontekst. Sett i lys av at mestringsforventning er kontekstavhengig (Bandura, 1997; Zimmerman, 1995), vil jeg argumentere for at det trengs mer forskning på dette på norske elever. Forfatterne av de to førstnevnte norske studiene viser til behovet for forskning som forsøker å forstå prosessene bak mestringsforventningens predikative effekt på matematiske prestasjoner. Jeg vil derfor argumentere for at forskning som går dypere inn i kildene til de norske elevenes mestringsforventning kan være et skritt i riktig retning, og kan være av nyttig for å få bedre forståelse og innsikt i de bakenforliggende prosessene. Med denne masteroppgaven ønsker jeg å bidra til mer forskning på nettopp dette.

2.4 Høyt- og lavtpresterende elever

Siden jeg i denne studien ønsker å benytte elevenes prestasjonsnivå som et komparativt aspekt, vil jeg i dette delkapittelet redegjøre for hva jeg legger i høyt- og lavtpresterende elever i matematikk.

Som en inngang til å forstå begrepet høytpresterende elever, vil jeg referere til rapporten *Mer å hente* (NOU 2016:14). Forfatterne av rapporten trekker frem utfordringer knyttet til begreper omhandlende elevenes evner og prestasjoner. Det nevnes at det er brukt over 100 forskjellige begreper på dette i internasjonal forskning. Nordiske land har tidligere vært noe skeptiske til å sette merkelapper på elever etter faglige evner for å unngå å stigmatisere, samt for å fremme tanken om at alle elever, uavhengig av faglig nivå, skal ha like muligheter til å utvikle seg. Av denne grunn velger forfatterne av NOU-rapporten sine begreper med omhu. De beskriver *elever med stort læringspotensial*, noe som både omfatter elever som presterer på et høyt faglig nivå, men også elever med spesielle evner eller talenter og elever som har potensial til å nå langt faglig. Rentzsch, Shütz og Schröder (2011) nevner kjennetegn som *intelligent, ambisjoner* og *gode karakterer* for høytpresterende elever. Forfatterne av NOU-rapporten referer til en analyse av PISA-resultater i nordiske land med fokus på elever som presterer høyt. På bakgrunn av denne analysen, vises det til typiske kjennetegn for elever som presterer høyt. Blant disse finner vi at disse elevene i større grad enn andre benytter læringsstrategier, blir motivert av konkurransebaserte læringsaktiviteter, er mer selvdrevne og har et bedre forhold til skolen og

lærere. Elevene har også høyere motivasjon, utholdenhet, selvoppfatning og mestringsforventning.

I en av utdanningsdirektoratets publikasjoner beskrives lavtpresterende elever med utgangspunkt i hvordan dette defineres i PISA og TIMSS-undersøkelsene (Utdanningsdirektoratet, 2019). Her fremkommer det at lavtpresterende elever anses å ha prestasjonsnivåer der elevene kun har noe grunnleggende kunnskap i faget, og under et minimumsnivå for å være forberedt på videre skolegang og yrkesliv.

Det er lovpålagt å tilpasse opplæringen for alle (Opplæringslova, 1998, §1-3). Det innebærer at skolen skal sikre at alle elever får best mulig utbytte av undervisningen, uavhengig av faglig nivå og utgangspunkt. Sammenhengen mellom mestringsforventning og prestasjon, beskrevet i delkapitlene ovenfor, medfører at et bevisst arbeid for å tilrettelegge for økt mestringsforventning hos elevene kan være et nyttig verktøy for elevenes utvikling. Denne studiens overordnede mål må derfor være å få en større forståelse for hvordan læreren kan bidra til å øke mestringsforventningen til elevene. Hvis det viser seg å være forskjeller mellom høytpresterende og lavtpresterende elevers kilder til mestringsforventning, kan dette være relevant informasjon for lærere som forsøker å tilpasse undervisningen best mulig til den enkelte elev.

Stuart (2017) argumenterer for at gapet mellom høyt- og lavtpresterende elever i skolen øker på grunn av skolens struktur. Han beskriver skolens testregime, kontinuerlig karaktersetting og sammenligning med medstudenter, som undergravende for lavtpresterende elevers mestringsforventning, innsats og engasjement. Dette er enda en god grunn til å undersøke sammenhengen mellom prestasjonsnivå og kildene til mestringsforventning.

Av annen forskning gjort på kildene til mestringsforventning i matematikk, finner jeg kun en studie som studerer kildene opp mot prestasjonsnivå. Dette er en kvalitativ studie av Yildiz og Özdemir (2019), der de så på en høytpresterende elev og en lavtpresterende elev sine kilder til mestringsforventning. Elevene gikk i 4.klasse. Informantene ble valgt på grunnlag av tidligere karakterer, samt informasjon fra læreren. Gjennom tre semistrukturerte intervjuer innhentet forskerne informasjon om elevenes kilder til mestringsforventning. De finner at det er en tydelig sammenheng mellom elevenes prestasjonsnivå og deres mestringsforventning, som er overensstemmende med annen forskning. De finner også en sammenheng mellom elevenes prestasjonsnivå og deres kilder til mestringsforventning. Yildiz og Özdemir (2019) beskriver at den høytpresterende eleven innhenter positiv informasjon fra kildene til mestringsforventning,

mens den lavtpresterende eleven innhenter negativ informasjon fra kildene. Avslutningsvis etterlyser forfatterne mer forskning og studier med flere respondenter på kildene til mestringsforventning som inkluderer prestasjonsnivå som en faktor.

2.5 Tankesett

Synet på evner er tett knyttet opp mot teorien om mestringsforventning (Bandura, 1993). Den enkelte elev bruker nemlig de fire kildene til mestringsforventning for å vurdere sine egne evner. Samtidig vet vi at elevenes mestringserfaringer også påvirkes av allerede eksisterende mestringsforventning, der nettopp synet på egne evner er sentralt. Dweck (2006, gjengitt i (Boaler, 2013)) definerte to ulike typer tankesett man kan se hos individer, der nettopp synet på evner er det sentrale som skiller de to tankesettene. Hun undersøkte effektene av et *statisk tankesett* og et *dynamisk tankesett*. Et statisk tankesett innebærer at man ser på evner som noe medfødt og uforanderlig, mens man med et dynamisk tankesett vil tro på at evner kan utvikles gjennom innsats (Wæge & Nosrati, 2018). Et kjennetegn for elever med et dynamisk tankesett er at de ikke er så redd for å gjøre feil, og tør å begi seg ut på utfordringer (Boaler, 2013). Disse elevene jobber og lærer mer effektivt. Elever med et statisk tankesett vil være mindre utholdende i møte med utfordringer fordi de ser på sin evne som statisk, og kan ofte bruke unngåelsesstrategier når de møter utfordringer de ikke tror de vil lykkes med.

Boaler (2013) har arbeidet med å fremme et dynamisk tankesett i skolen, og hevder matematikk er det faget i skolen som sterkest kommuniserer et statisk tankesett. Hun har derfor beskrevet hvordan et dynamisk tankesett kan kommuniseres til elevene gjennom klasseromsundervisningen i matematikk. Oppgaver, stilte spørsmål, normsetting, karakterer og tilbakemelding, feil og gruppering av elever er elementer i undervisningen Boaler (2013) mener kan være arenaer for å fremme et dynamisk tankesett. Hun oppfordrer blant annet lærere til å benytte mer åpne oppgaver, og gå bort ifra korte, lukkede oppgaver som enten gir rett eller galt svar. Hun oppfordrer også til å verdsette feil som gjøres i klasserommet, da disse kan være svært utviklende for hjernen.

Siden mestringsforventning dreier seg om å vurdere sine egne evner, er det trolig at denne vurderingen vil påvirkes av det grunnleggende synet man har på evner. Jeg vil derfor påstå at teorien om tankesett kan være et nyttig perspektiv å ha med meg videre i diskusjonen av denne studiens resultater. De to tankesettene kjennetegnes også av ulike holdninger til det å gjøre feil (Boaler, 2013), noe som trolig kan bidra til å nyansere oppfatningen av mestringserfaringer og nederlag innenfor mestringsforventningsteorien.

2.6 Sosiomatematiske normer

Sosiomatematiske normer er normer i klasserommet som er spesifikke for matematikkfaget (Yackel & Cobb, 1996). Slike normer vil være med på å danne en klasseromskultur som kan utvikles i interaksjonen mellom læreren og elevene i klasserommet. Sosiomatematiske normer skiller seg imidlertid fra mer generelle normer fordi de spesifikt omhandler den matematiske aktiviteten. Et eksempel kan være en norm for hva som gjelder for å være en matematisk annerledes forklaring. Elevene er da klar over at det å ordlegge seg annerledes enn tidligere gitte forklaringer ikke nødvendigvis medfører at forklaringen er matematisk annerledes, men at det er den matematiske løsningsmetoden som skal være annerledes. Yackel og Cobb (1996) trekker frem lærerens egne holdninger til matematikk som viktig for normene, da lærerens holdninger vil komme til syne gjennom samhandlingen med elevene.

2.7 Forskningsspørsmål

Med bakgrunn i litteratur og forskning som er presentert i dette delkapittelet, vil jeg nå spisse problemstillingen min (som gitt i introduksjonen) og fremsette denne oppgavens forskningsspørsmål. Forskningsspørsmålene skal bidra til å utdype den overordnede problemstillingen.

Bandura (1997) hevder at individer ikke er spesielt gode på å behandle multidimensjonal informasjon, slik vi kan si at kildene til mestringsforventning er. Dette medfører at man ofte legger for stor vekt på enkelte kilder og ignorerer informasjon fra andre kilder. I denne studien vil jeg undersøke hvorvidt det eksisterer mønstre i hvilke kilder som overvurderes eller undervurderes. Slik som det er beskrevet i ovenstående delkapitler, viser tidligere forskning ikke konsistens i rekkefølgen av kildenes betydning for mestringsforventning, med unntak av at mestringserfaringer tydelig fremkommer som den viktigste kilden. Dette gir meg god grunn til å forsøke å forstå kildenes betydning for elever i matematikk bedre.

I denne studien vil jeg komparativt undersøke hvordan de ulike kildene vektlegges blant høytpresterende og lavtpresterende elever, med utgangspunkt i følgende problemstilling: «*Er det forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning i matematikk, og si fall hvilke?*». Dersom resultatene av studien peker på forskjeller i de to elevgruppenes vurdering av kilder til mestringsforventning, mener jeg det kan ha viktige implikasjoner for praktiseringen av prinsippet om tilpasset opplæring (Opplæringslova, 1998, §1-3). Dette leder meg til følgende forskningsspørsmål:

1. Hvordan kan et instrument som måler elevers kilder til mestringsforventning oversettes og valideres for bruk på norske elever?
2. Er det forskjeller i hvordan høyt- og lavtpresterende elever rapporterer de ulike kildene til mestringsforventning, og eventuelt hvilke mønstre viser seg?

3.0 Metode

I denne delen av oppgaven vil jeg redegjøre for metoden som er benyttet i studien. Jeg vil både si noe om forskningsdesignet, hvordan undersøkelsen er utformet, pilotering og innsamling av datamaterialet, samt komme med noen forskningsetiske betraktninger. I tillegg vil kapittelet inneholde en redegjørelse av Rasch-analyser, som er analyseverktøyet jeg benytter i denne studien.

3.1 Valg av metode

Problemstillingen bør være avgjørende for valg av metode i forskning (Nardi, 2018). Siden jeg i denne studien ønsker å se på forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning, finner jeg det fordelaktig å innhente data fra et større antall elever. Nettopp dette er den kvantitative metodens styrke (Nardi, 2018). Fordelen ved å innhente svar fra mange respondenter er at man kan si noe om hvor utbredt et fenomen er, og også undersøke fenomenet i ulike sosiale grupper (Tufte, 2011).

Forskningsprosessen bør være preget av systematikk, grundighet og åpenhet (Johannessen, Tufte & Christoffersen, 2016). For å etterstrebe dette vil jeg i inneværende kapittel gjøre rede for forskningsprosessen i detalj. I mine rapporteringer forsøker jeg hele tiden å være så transparent som mulig.

3.2 Studiens rammer og utarbeidelse av instrumenter

I dette delkapittelet vil jeg redegjøre for grunnlaget for studien min. Dette innebærer at jeg vil si noe om valg av de overordnede rammene for studien, samt redegjøre for hvordan undersøkelsen benyttet til datainnsamling ble utformet.

3.2.1 Forskningsgruppa SESAM

I tidlig fase av forskningsprosessen dannet vi forskergruppen SESAM: *Self-Efficacy, its Sources and Achivement in Mathematics*. Denne gruppen ble dannet av meg og to andre medstudenter sammen med vår felles veileder, med den hensikt å samarbeide om å samle inn et datamateriale som alle tre kunne benytte som grunnlag for analyser i hver våre masteroppgaver. Til tross for at våre oppgaver skulle besvare ulike problemstillinger med ulike vinklinger, var mestringsforventninger i matematikk det overordnede temaet. Samarbeidet i SESAM har altså primært dreid seg om utforming av forskningsdesign, utarbeidelse av en undersøkelse som bestod av tre instrumenter og gjennomføringen av selve datainnsamlingen. Hvis jeg ser forskningsprosessen i lys de fem fasene *(1) innledning, (2) forskningsdesign, (3)*

datainnsamling, (4) dataanalyse og *(5) konkludering og rapportering* (Thrane, 2018) kan vi si at det særlig er i fase 2 og 3 at samarbeidet i forskningsgruppa har stått sentralt. De resterende fasene har foregått som et individuelt arbeid i samråd med veileder.

Opprettelsen av SESAM var hensiktsmessig av flere grunner. Kanskje den største fordelen var at vi ved datainnsamlingen kunne nå ut til langt flere respondenter enn det en enkeltperson kunne gjort innenfor den samme tidsrammen. Jeg vil også trekke frem verdien av å ha forskningskollegaer å drøfte med. Særlig i startfasen av forskningsprosjektet ble dette samarbeidet betydningsfullt, både for nyanseringen av problemstillingen og utformingen av forskningsdesignet. Når jeg videre beskriver de ulike delene av forskningsprosessen i detalj, vil jeg også påpeke flere fordeler ved å arbeide med studien på denne måten.

Utgangspunktet for utformingen av undersøkelsen benyttet til datainnsamling var våre problemstillinger. Vi i SESAM satt oss altså ned i fellesskap for å se hvilke data vi som en gruppe kom til å trenge. Vi fant ut at vi trengte data om elevenes *mestringsforventninger, kilder til mestringsforventning* og *prestasjonsnivå.* I tillegg ville vi trenge karakteristikker som kjønn, alder, fødselsmåned og fødselsår. Vi kan si at utformingen av undersøkelsen var tredelt. Vi skulle (1) finne en metode for å måle elevers mestringsforventning i matematikk, (2) finne en metode for å undersøke elevenes kilder til mestringsforventning og (3) finne en metode for å vurdere elevenes prestasjonsnivå. Når jeg nedenfor redegjør for utformingen av undersøkelsen, legger jeg størst vekt på (2), da det er denne delen av undersøkelsen som utgjør hovedgrunnlaget for min masteroppgave. Jeg vil i tillegg si noe om (3), da dette instrumentet benyttes til å skille mellom høyt- og lavtpresterende elever i min studie. Til tross for at jeg i denne oppgaven ikke benytter (1), velger jeg likevel å kommentere hovedtrekkene i prosessen, da den henger tett sammen med (3).

3.2.2 Instrumentene som måler elevers mestringsforventning (1) og prestasjon (3) i matematikk

I henhold til Banudas (2006) anbefalinger for målinger av mestringsforventning, utarbeidet vi et instrument som var emne- og oppgavespesifikk. For å måle elevenes mestringsforventning i matematikk, tok vi i bruk en metode som er brukt av flere anerkjente forskere (Chen, 2003; Pajares & Graham, 1999; Stevens, Olivarez Jr., Lan & Tallent-Runnels, 2004). Vi målte elevenes mestringsforventninger ved å vise elevene ulike matematikkoppgaver og stille dette spørsmålet til hver oppgave: *Hvis du skulle gjort denne matematikkoppgaven, hvor trygg er du på at du kunne løst den?* Elevene rangerte svarene sine på en Likert-skala fra 1 (*veldig utrygg*) til 6 (*veldig trygg).* Dette instrumentet utgjorde del 1 av undersøkelsen, heretter kalt instrument 1.

Bandura (2006) anbefaler i sin *Guide for Constructing Self-Efficacy Scales* å benytte en Likert-skala fra 0 – 100 med intervaller på 10 (tilsvarende en 0-10-skala). Han fraråder å bruke skalaer med få kategorier, fordi skalaene er mindre sensitive og reliable. Toland og Usher (2016) foreslår imidlertid at fire kategorier er optimalt for å måle unge voksnes mestringsforventning i matematikk. Dette begrunnes blant annet med at færre kategorier letter den kognitive byrden og gjør det mer sannsynlig at respondentene vil fullføre undersøkelsen, samt at færre kategorier trolig vil skape mindre støy og målefeil i dataene. Wolfe og Smith (2007) påpeker at det vanlige er å tenke at flere svarkategorier vil føre til bedre nyansering og bedre reliabilitet. Forfatterne hevder dette er sant opptil et visst punkt. Som en tommelfingerregel hevder de at voksne sjeldent kan skille mellom mer enn syv nivåer på en skala, og legger til at det for yngre deltakere gjerne kan benyttes et lavere antall. I vår vurdering av antall svarkategorier ble det naturlig å se det i sammenheng med de andre instrumentene vi benyttet i undersøkelsen. For at det skulle være enkelt for elevene å forholde seg til, brukte vi samme antall kategorier både for instrumentet som undersøker mestringsforventningen og instrumentet som undersøker kildene til mestringsforventning. I kap. 3.2.3 utdypes begrunnelsen av vårt valg om å bruke en 6-poengs Likert-skala. Uavhengig av hvor mange svarkategorier man benytter, minner Toland og Usher (2016) om å undersøke konstruktvaliditeten til det instrumentet og den responsskalaen som er brukt i hver enkelt undersøkelse, for å videre kunne legge frem konklusjoner på bakgrunn av undersøkelsen.

Instrumentet for å måle elevers matematiske prestasjonsnivå, heretter kalt instrument 3, ble gjennomført som en typisk matematikkprøve. Elevene skulle da regne de samme oppgavene som ble benyttet i instrument 1. Oppgavene var da i annen rekkefølge, med den hensikt å hindre at elever skulle endre sine svar i instrument 1 etter å ha regnet ut oppgavene i instrument 3.

For å sikre at matematikkoppgavene i undersøkelsen skulle være på riktig nivå i forhold til elevenes klassetrinn, ønsket vi å hente oppgaver fra tidligere nasjonalt gitte prøver. I første omgang tok vi utgangspunkt i en tidligere gitt *nasjonal prøve i regning*, men etter en diskusjon i forskningsgruppa SESAM, valgte vi likevel å gå bort fra disse oppgavene. Dette fordi *nasjonal prøve i regning* er ment for å måle regning som en grunnleggende ferdighet. Dette innebar at det var mange tolkingsoppgaver, og mindre spesifikke matematikkfaglige oppgaver. Vi vurderte det slik at det kan være vanskeligere å raskt avgjøre hvor trygg man føler seg på at man hadde klart en oppgave dersom det kreves mye tolkning og mye tekstlesing. Vi valgte derfor heller å bruke oppgaver fra den internasjonale TIMSS-undersøkelsen (*Trends in International Mathematics and Science Study*) som ble gitt til 8. og 9.trinn på ca. 150 skoler i

Norge i 2015 (Haavik, 2015). TIMSS-undersøkelsen i matematikk er laget for å måle elevenes faglige kunnskap på områdene *tall, geometri, statistikk* og *algebra* på internasjonalt nivå (Utdanningsdirektoratet, 2020). Det er kun halvparten av oppgavene fra TIMSS-undersøkelsen som blir frigitt etter testen. De matematikkoppgavene vi benyttet oss av er derfor et utvalg fra den frigitte halvparten.

I selve utvalgsprosessen gjorde vi det slik at hver av deltakerne i SESAM valgte ut 5 oppgaver fra hvert av de fire matematiske temaene nevnt ovenfor. Deretter samlet vi oss og diskuterte hvilke oppgaver vi mente var passende og gode for vårt formål. Utvelgelsen av matematikkoppgavene ble gjort etter noen kriterier vi hadde satt oss. Ett av disse var å unngå flervalgsoppgaver, da det kan tenkes at elevenes mestringsforventning kan øke noe på slike oppgaven da det uansett vil være en viss sannsynlighet for å tilfeldig treffe riktig svar. Vi har derfor hovedsakelig valgt oppgaver som ikke hadde svaralternativer. Likevel har vi i et fåtall av oppgavene gjort små endringer ved å fjerne svaralternativene. Et annet kriterium vi brukte under utvelgelsen var at det skulle være mulig for elevene å relativt raskt forstå hva oppgaven spurte etter. Dette gjorde at oppgaver som krevde mye tekstforståelse ikke ble tatt med. Vi valgte først ut fem oppgaver fra hver av kategoriene, men kuttet deretter ned til fire oppgaver fra hvert tema, det vil si totalt 16 oppgaver. Denne justeringen gjorde vi for at totaltiden av undersøkelsen ikke skulle overskride en klokketime. Oppgavemengden var ment som et førsteutkast, som senere skulle testes ved en pilotering av undersøkelsen.

3.2.3 Instrumentet som måler elevers kilder til mestringsforventning (2)

For å måle elevers kilder til mestringsforventning benyttet vi oss av et allerede validert instrument, *Sources of Middle School Mathematics Self-Efficacy Scale*, som er utformet av Usher og Pajares (2009). I deres review-artikkel av forskning på kildene til mestringsforventning i skolen som ble utgitt ett år tidligere (Usher & Pajares, 2008), finner de at det mye variasjon i undersøkelser og items som er brukt, noe som resulterer i lite konsistens mellom funnene. I mangel av et dekkende norsk begrep, benytter jeg meg av det engelske begrepet *item/items* for det som kan være en oppgave, en påstand eller et spørsmål. Forfatterne kritiserer flere sider ved tidligere studiers målinger, blant annet at studier bruker objektiv prestasjon som et mål på kilden mestringserfaringer, noe som ikke er i tråd med Banduras (1997) beskrivelse av mestringserfaringer som individets egen oppfattelse av erfaringen. Etter en grundig gjennomgang av tidligere forskning på feltet, publiserer og validerer de et instrument som måler kildene til mestringsforventning (Usher & Pajares, 2009).

Instrumentet er utviklet for å måle kildene til mestringsforventning i matematikkfaget, og er rettet mot elever på 6. - 8.trinn i amerikansk kontekst (Usher & Pajares, 2009). Gjennom en kvantitativ valideringsstudie, viser de i artikkelen at instrumentet pålitelig kan brukes for å måle kildene til den aktuelle elevgruppen. Forfatternes intensjon med studien er blant annet å gi lærere tilgang til et instrument som kan gi større innsikt i hva som bidrar til å øke mestringsforventningen hos deres elever. Usher og Pajares (2009) hevder at denne forståelsen vil være nyttig for alle som ønsker å fremme elevenes kompetanse. Usher og Pajares' (2009) validering av instrumentet er et godt argument for at det er forskningsmessig forsvarlig å ta i bruk instrumentet til datainnsamlingen i denne studien. I påfølgende avsnitt vil jeg redegjøre for forfatternes validering av instrumentet.

Konklusjoner trukket fra empiriske undersøkelser er ikke mer reliable enn itemene i det benyttede instrumentet (Usher & Pajares, 2009). Dette er forfatternes bakgrunnsmotivasjon for å utarbeide og validere et instrument. I deres studie foregår valideringsprosessen i tre faser. I første fase ble et instrument bestående av 84 items besvart av en fokusgruppe som inkluderte både 6.klasse-elever og voksne med noe ulike roller. Her ble det gitt tilbakemeldinger angående ordlyd og klarhet, og førte til en revidert utgave av instrumentet. Noen uker senere besvarte 1111 elever på 6., 7. og 8.trinn instrumentet. Deretter utførte forfatterne en rekke analyser, som viste at en del items var problematiske. I fase 2 ble disse itemene byttet ut med nye items, og en ny undersøkelse med 86 items ble besvart av 824 respondenter på de samme klassetrinnene. Igjen ble det gjort analyser, og forfatterne tok med seg et 39-items-instrument inn i fase 3. I denne fasen gjorde Usher og Pajares (2009) noen justeringer på flere av itemene som ikke overlevde fase 2. I tillegg fikk de tilbakemeldinger på itemene av en gruppe eksperter innenfor den sosialkognitive teorien. Dette for å sørge for at itemene var teoretisk forsvarlige. På grunnlag av ekspertenes tilbakemeldinger tok forfatterne bort noen items, justerte ordlyden på noen items og la til seks nye items. Resultatet var et instrument bestående av 73 items som skulle testes i 3. fase. Med 803 respondenter og en rekke analyser, valgte forfatterne igjen bort problematiske items og valgte til slutt ut 24 items som skulle utgjøre det endelige instrumentet. Deres metodiske fremgangsmåte og analyser brukt i valideringsstudien er detaljert beskrevet i deres artikkel (Usher & Pajares, 2009). Dette vil jeg ikke utdype nærmere her.

Den endelige utgaven av instrumentet består altså av 24 items, der det er 6 items som omhandler hver av de fire kildene (mestringserfaringer, vikarierende erfaringer, oppmuntring, støtte og overtalelse fra andre og psykologiske og fysiologiske tilstander*)*. Itemene omhandlende vikarierende erfaringer er delt inn i undergruppene: Vikarierende erfaringer fra voksne,

vikarierende erfaringer fra jevnaldrende og vikarierende erfaringer fra en selv. Itemene i instrumentet er skrevet som førstepersonspåstander der elevene skal avgjøre hvor godt påstanden stemmer for seg selv.

For å kunne bruke instrumentet i våre undersøkelser, var vi avhengig av å oversette det til norsk, og jeg vil her forklare hvordan utarbeidelsen av den norske versjonen av instrumentet foregikk. Første steg var at samtlige i forskningsgruppa SESAM oversatte instrumentet hver for seg. Deretter diskuterte vi de ulike oversettelsene i fellesskap. På denne måten fikk vi diskutert ulike ordlyder og vår teori om elevenes oppfattelse av ulike ord og begreper. Vi forsøkte situasjonstilpasse språket best mulig; vi hadde målgruppen for øyet når vi forsøkte å finne den mest naturlige måten å si noe på. Det var for eksempel et bevisst valg å bruke ordet «matte» fremfor «matematikk» til tross for at sistnevnte vanligvis ville være det mest korrekte. I dette tilfellet var argumentene i gruppa at elevene i hovedsak kaller faget i skolen for «matte», mens «matematikk» trolig assosieres med et mer altomfattende og overordnet begrep for matematiske regneoperasjoner, og kan oppleves som et mer abstrakt og udefinert begrep. Bruken av plussord var også gjenstand for diskusjonen. Usher og Pajares (2009) beskriver at de har brukt en positiv ordlyd på alle items med unntak av de itemene som omhandler psykologiske og fysiske tilstander og ett av itemene omhandlende mestringserfaringer (henholdsvis item 19-24 og item 3, se vedlegg 5). Det samme gjelder også for den norske oversettelsen. Forfatterne oppfordrer forskere som benytter instrumentet til å være oppmerksom på disse itemene, da resultatet nemlig kan påvirkes av om et item har negativ eller positiv ordlyd (Marsh, 1996).

I likhet med Usher og Pajares (2009) innhentet vi svar fra respondentene ved hjelp av en Likert-skala med 6 kategorier. Elevene ble bedt om å svare på spørsmålet: «*Hvor godt stemmer påstandene for deg?*» og skulle gi svar på en skala fra 1 *(stemmer ikke i det hele tatt)* til 6 *(stemmer veldig godt).* Fordi det validerte instrumentet benyttet 6 Likert-kategorier ble det et naturlig valg å videreføre dette.

Den norske oversettelsen av Usher og Pajares (2009) sitt instrument, utgjorde det andre instrumentet i vår undersøkelse (se del 2, vedlegg 2) og er hovedgrunnlaget for mine analyser i denne studien. Det er dette instrumentet som hovedsakelig omtales videre i oppgaven. For å skille instrumentet fra de to andre instrumentene i inneværende kapittel, refererer jeg til instrumentet som instrument 2. Siden det validerte instrumentet både er oversatt til norsk og benyttes i en helt annet kontekst enn det opprinnelige instrumentet, vil jeg som en del av denne studien validere instrumentet. Usher og Pajares (2009) oppfordrer til nettopp dette; at videre forskning kan benytte de samme itemene og validere bruken av de i ulike kontekster og emner.

Som et ledd i valideringsprosessen vil jeg derfor i kapittel 3.4 gjøre rede for hvordan vi testet ut den norske utgaven av instrumentet gjennom en pilotstudie med tilhørende analyser.

Før jeg går mer i detalj på pilotering og hoveddatainnsamling velger jeg å stoppe opp og gi en redegjørelse for Rasch Rating Scale Model (RSM) og Rasch analyse. Dette gjøres fordi sentrale valg i piloteringsprosessen og utviklingen av undersøkelsen benyttet til hoveddatainnsamlingen ble gjort på grunnlag av Rasch-analyser utført på pilotdataene.

3.3 Rasch-analyse

I denne masteroppgaven benyttes en Rasch-modell som verktøy for å analysere dataene. I dette delkapitlet ønsker jeg å introdusere prinsippene bak Rasch-analyser og begrunne hvorfor jeg i denne studien finner det hensiktsmessig å ta i bruk en Rasch-modell, nærmere bestemt Rasch Rating Scale Model (RSM), som analyseverktøy. Jeg vil også forklare sentrale begreper knyttet til Rasch-analyser. I analysen benyttes WINSTEPS 3.81.0 software (Linacre, 2014) for å teste hvorvidt datamaterialet samsvarer med Rasch-modellen. Siden programvarene og litteratur knyttet til Rasch er engelske, velger jeg for enkelhets skyld å benytte meg av de engelske begrepene, der det ikke er allerede er etablert en passende norsk oversettelse. Begrepene defineres fortløpende i delkapitlene nedenfor.

3.3.1 Prinsippene bak Rasch-analyser

For å forstå de sentrale prinsippene bak Rasch-analyser vil jeg først trekke frem to fundamentale aspekter ved måling. Enhver forsker som ønsker å undersøke et fenomen, må begynne med å definere en variabel som kan måles (Boone, 2016). Eksempelvis kan høyden til ulike blomster måles ved å bruke en linjal. Når det kun er høyden til blomstene forskeren måler, gir det mening å sammenligne de ulike blomstene. Det er altså viktig at instrumentet kun måler en variabel om gangen, for at målingene skal være meningsfulle. En annen fundamental faktor er at skalaen som det måles med er lineær. Det at linjalen i eksempelet er en lineær skala, altså at avstanden mellom målene på skalaen er like store, er helt essensielt for å kunne sammenligne de ulike resultatene av målingen (Boone, 2016).

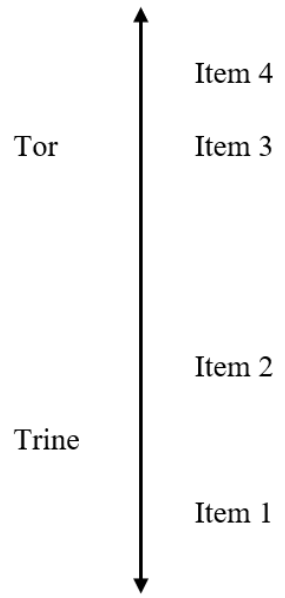
Mange typer variabler vil kunne måles med lineære skalaer, slik som tid eller lengde. Når en forsker samler inn data ved hjelp av ulike spørreundersøkelser fra elever, kan det være fristende å benytte rådataen direkte til å sammenligne elevene, for eksempel i form av score eller poengsum (Boone, 2016). Det problematiske med dette er imidlertid at skalaene ikke vil være lineære. Tar vi utgangspunkt i en matematikkprøve, vil dette komme bedre til syne. Oppgavene på en matematikkprøve kan naturligvis ha forskjellig vanskelighetsgrad. Hvis vi

tenker oss at prøven har 10 oppgaver med varierende vanskelighetsgrad, vil det være fornuftig å si at en elev som har fått til de 8 vanskeligste oppgavene, er på et høyere matematisk nivå enn en elev som har fått til de 8 enkleste oppgavene. Dette til tross for at begge elevene har fått samme poengsum på prøven.

Det samme gjelder for undersøkelser med bruk av Likert-skalaer. Benytter man for eksempel en skala med svarkategoriene helt enig (HE), litt enig (LE), litt uenig (LU) og helt uenig (HU) kan det være fristende å kode datamaterialet slik at HE=4, LE=3, LU=2 og HU=1, og deretter gjøre matematiske analyser ved bruk av tallene (Boone, 2016). Dette er problematisk fordi forskeren da antar at det er like stor avstand mellom alle svaralternativene for alle itemene. Det eneste som faktisk kan sies om svarkategoriene er at HE>LE>LU>HU. På samme måte som de ulike itemene på en matematikkprøve vil ha ulik vanskelighetsgrad, kan det være forskjeller i de ulike itemene i en spørreundersøkelse. For noen av itemene kan det kanskje oppfattes som lettere å svare «helt enig» enn for andre items.

Bond og Fox (2015) hevder at det i humaniora og samfunnsvitenskapene ofte benyttes målinger med ordinale data, slik som data innhentet ved hjelp av en Likert-skala. De kritiserer at målingene i psykologien stort sett har blitt gjort ved å beskrive rådataene, i form av å rapportere hvor mange respondenter som besvarte ulike items korrekt, hvor relatert en besvarelse er til en annen, eller hvilke korrelasjoner som finnes mellom ulike items og totalscoren. Forfatterne mener forskningen lar slike beskrivelser telle for målinger i humaniora. De argumenterer for at vi må bruke mer tid på å undersøke skalaene våre, enn å undersøke med skalaene våre. De hevder data på nominalt og ordinalt målenivå ikke er noen form for måling i seg selv. Stenner (2001) definerer måling som prosessen med å konvertere observasjoner (f.eks. opptelling) til mål (kvantiteter) via en konstruktteori (gjengitt i (Bond & Fox, 2015)). Ved bruk av Rasch-analyse gjøres nettopp en slik konvertering.

I dette avsnittet vil jeg ved hjelp av figur 1 forklare hvordan dataene fra en Likert-skala blir omformet til en kontinuering skala i Rasch-analyser (Bond & Fox, 2015). Modellen viser at items og personer måles ved samme logit-skala. Det er undersøkelsens faktiske data som danner grunnlaget for den kontinuerlige skalaen, der itemenes vanskelighetsgrad og



Figur 1. *En skjematisk modell av Rasch-målinger.*

respondentenes evner er parametere i skalaen. Vi kan si at den vertikale linjen representerer skalaen for det konstruktet som ønskes undersøkt (Boone, 2016). De to respondentenes plassering langs den vertikale linja, markerer deres evne på skalaen. Undersøkelsens items er også plassert langs samme skala, og markerer vanskelighetsgraden til itemene. Plasseringen til hvert item langs linja uttrykker også en sannsynlighet for at en respondent, målt ved et bestemt nivå på skalaen, svarer korrekt på itemet. I dette eksempelet har Tor 50% sannsynlighet for å svare korrekt på item 3, men mindre enn 50% sannsynlighet for å svare korrekt på item 4. Tilsvarende vil Trine ha mer enn 50% sannsynlighet for å svare korrekt på item 1, og mindre enn 50 % sannsynlighet for å svare korrekt på de resterende itemene.

Den kontinuerlige skalaen i Rasch-analyser, kalt *logit-skala*, dannes altså av sannsynligheten for at en person lykkes med et item. Dette kan enklest forstås med utgangspunkt i matematikkprøveeksempelet nevnt tidligere. For hver oppgave vil det være to mulige utfall: eleven kan avgi et riktig svar eller et feil svar. I Rasch-modellen benyttes resultatene til å rangere alle elevene etter evne basert på deres score på prøven. Samtidig rangeres oppgavene etter vanskelighetsgrad, basert på hvor mange av elevene som fikk til hver oppgave. Sannsynligheten beregnes ved hjelp av forskjellen mellom hver oppgaves vanskelighetsgrad og elevens evne. Dataene, som opprinnelig var på ordinalt målenivå, blir nå representert ved en kvantitativ målestruktur, ved hjelp av en logaritmisk transformasjon av sannsynligheten for suksess (Bond & Fox, 2015).

Rasch-modeller er også et godt analyseverktøy for å administrere selve måleinstrumentet og skreddersy instrumentet til den spesifikke konteksten (Cavanagh & Fischer, 2018). Rasch-analysen vil synliggjøre items som ikke bidrar til å skille respondentene fra hverandre. Dette kan være items som alle respondentene får til, items som ingen av respondentene får til, eller flere items som måler det samme. Fordi disse itemene ikke er nyttige diskriminatorer for det som skal undersøkes, bør itemene fjernes eller erstattes. Rasch-analyser vil også synliggjøre respondenter som enten har svart korrekt på alle items eller feil på alle items. Disse respondentene måles ikke godt av instrumentet, fordi vi ikke kan si noe om «hvor mye høyere/lavere» deres evner er. Testen som er utført kan dermed ikke gi oss presis informasjon om slike elevers evner.

De ovenstående avsnittene fungerer som en introduksjon til hva Rasch-analyser er, og forklarer noe av grunnen til at jeg har valgt en Rasch-modell som analyseverktøy i denne studien. For det første muliggjør Rasch-analyser en transformering av ordinale data til kontinuerlig data, og kan på den måten berike datamaterialet (Bond & Fox, 2015). For det andre vekter Rasch-

analyser vanskelighetsgraden på itemene, noe som medfører at resultatene er mer nøyaktige. To elever med like høy score (målt på ordinal skala), vil dermed ikke nødvendigvis klassifiseres som like evnerike (på kontinuerlig skala). For det tredje bidrar Rasch-analyser til å identifisere oppgaver i datasettet som ikke bidrar til å måle det bakenforliggende konstruktet eller blir overflødige i oppgavesettet, og vil dermed være et nyttig verktøy for å sikre god kvalitet i utviklingen av instrumentene som benyttes i undersøkelsen. Det er på bakgrunn av disse argumentene at Bond og Fox (2015) hevder at Rasch-analyser i måling er det nærmeste vi kommer en tilnærming av de fundamentale måleprinsippene for humaniora på nåværende tidspunkt.

Det finnes flere Rasch-modeller, og det er som nevnt Rating Scale Model (RSM) som benyttes i denne masteroppgaven. Jeg vil her forklare hvorfor nettopp denne Rasch-modellen benyttes. Om en undersøkelse har dikotome data vil det kun være to mulige utfall for hvert item, slik som i matematikkprøveeksempelet ovenfor. Da benyttes den dikotome Rasch-modellen. Bruk av Likert-skalaer i undersøkelser gir imidlertid polytome data, som skiller seg fra dikotome fordi det gir mer enn to mulige svarkategorier (Andrich, 2010). Dette krever en polytom Rasch-modell (PRM). Det er to varianter av PRM: Rating scale model (RSM) og Partial credit model (PCM). RSM måler alle items med samme skala, der svarkategoriene er like store for alle items. PCM benyttes dersom svarkategoriene for de ulike itemene varierer i størrelse eller antall. Modellene er likevel identiske hvis man kun ser på en enkeltpersons respons på ett enkelt item og har derfor fått samlebetegnelsen PRM. Vi kan også si at den dikotome Rasch-modellen er et spesialtilfelle av PRM. I denne masteroppgaven benyttes RSM fordi itemene i instrumentene er designet for å måles med samme Likert-skala, og avstanden mellom svarkategoriene er derfor ment for å være lik for alle items (Linacre, 2000).

I avsnittene nedenfor følger en forklaring av sentrale begreper knyttet til Rasch-analyser, og en presentasjon av de ulike analysene jeg vil benytte i kapittel 4.0.

3.3.2 Endimensjonalitet

Rasch-modellen fordrer endimensjonalitet (Bond & Fox, 2015). Det betyr at modellen kun kan måle ett teoretisk konstrukt om gangen, noe som også tidligere nevnt er et velkjent fenomen innenfor vitenskapelig forskning. Hvis en matematikkprøve er ment for å måle matematiske ferdigheter, men i praksis måler både matematiske og språklige ferdigheter, vil ikke resultatene kunne si noe konkret om de matematiske ferdighetene. Bond og Fox (2015) mener derfor det bør være et hovedfokus å identifisere mulige flerdimensjoner. Tilpasningsstatistikk, som

forklares nærmere i understående delkapittel, er et viktig verktøy for å oppdage om et instrument har flere underliggende dimensjoner.

3.3.3 Tilpasningsstatistikk

I Rasch-analyser er *fit* et sentralt begrep, og sier noe om hvor godt dataene passer til modellen (Linacre, u.å.). Jeg bruker oversettelsen *tilpasningsstatistikk* om fit-analsyer. Begrepet er sentralt, fordi det i Rasch-analyser er dataene som må passe til modellen (Bond & Fox, 2015). Dette er i motsetning til *Item Response Theory* (såkalte IRT-analyser), der fit handler om hvor godt modellen passer til dataene. Bond og Fox (2015) gjør det klart at et empirisk datagrunnlag aldri vil passe Rasch-modellens teoretiske ideal perfekt. Det man derimot vurderer ved å se undersøke tilpasningsstatistikk er om dataene passer modellen godt nok til å støtte de slutningene vi ønsker å trekke fra modellen. Tilpasningsstatistikk skal bidra til å oppdage avvik fra Rasch-modellens preferanser og empiriske data. Dette rapporteres i form av *outfit og infit mean-square* (outfit MNSQ og infit MNSQ) og *outfit og infit standardized mean* (outfit og infit *ZSTD*). Før jeg forklarer disse begrepene nærmere, vil jeg aller først definere *residualer*, som er grunnlaget for de ulike målene for tilpasningsstatistikk.

Residualer er differansen mellom den faktiske responsen fra hver person på hvert item og den forventede responsen for hver person på hvert item. Den forventede responsen er sannsynligheten for at en person skal få til en gitt oppgave. Hvis Rasch-modellen estimerer den forventede responsen fra en person på et item til å være 0.95 (altså stor sannsynlighet for at personen klarer oppgaven) og personens faktiske score er 1 (personen klarte oppgaven), blir denne residual-verdien 1 - 0.95 = + 0.05. Jo nærmere residual-verdiene er 0, jo mindre er den og jo nærmere verdien er lik -1 eller 1, jo større er den. Man kan regne ut slike residual-verdier for hver person på hvert item. For å kunne summere de ulike residual-verdiene uten at positive og negative residual-verdier utligner hverandre, er det vanlig å kvadrere de.

MNSQ-verdiene er et mål på «*how much misfit exists in the data?» (*Linacre, *u.å.).* Det sier altså noe om hvor mye tilfeldigheter som finnes i datamaterialet. Outfit MNSQ er *«outlier sensitive fit statistics»* og er gjennomsnittet av de kvadrerte residualene. Denne verdien er mer sensitiv for uventede observasjoner fra personer på items som egentlig burde være veldig vanskelige eller lette for dem. Dette medfører at høye outfit-verdier kan være et resultat av noen få uventede responser. Infit MNSQ er «*inlier sensitive fit statistics»* og er mer sensitiv for uventede observasjoner fra personer på items som er på nivå med personen. Man kan si at infit-verdier er et informasjonsvektet gjennomsnitt (Bond & Fox, 2015). Rasch-analyser genererer både infit og outfit MNSQ-verdier, og man kan spørre seg hvorvidt man burde rapportere og

benytte seg av infit eller outfit-verdiene. Til dette anbefaler Linacre (u.å.) å benytte outfit MNSQ, fordi dette er en chi-kvadrat-statistikk og dermed en velkjent metode for de fleste forskere. Han hevder outfit MNSQ bør rapporteres, så lenge ikke det er gode grunner til å benytte infit MNSQ. Jeg velger derfor å rapportere outfit MNSQ i mine analyser. Man forventer at MNSQ-verdiene ligger rundt 1.0, men de kan variere fra 0 til uendelig. Er MNSQ-verdien over 1.0 indikerer det *underfit,* som betyr at det er mer variasjon i datamaterialet enn forventet av Rasch-modellen. En verdi på 1,4 tilsier at det er 40% mer variasjon i datamaterialet enn modellen forventet. Er verdien under 1.0 indikerer det *overfit,* som betyr at datamaterialet har mindre variasjon enn forventet av Rasch-modellen. En verdi på 0,6 tilsier at det er 40 % mindre variasjon i datamaterialet enn modellen forventet. Bond og Fox (2015) betegner infit og outfit MNSQ-verdier mellom 0,6 og 1,4 som god tilpasning og velfungerende for målinger med instrumenter som benytter Likert-skalaer.

Infit og outfit ZSTD er standardiserte t-tester av hypotesen «*do the data fit the model (perfectly)?"* (Linacre, u.å.). ZSTD rapporterer den statistiske signifikansen for å få de aktuelle MNSQ-verdiene tilfeldig, selv om dataene hadde passet Rasch-modellen perfekt. Den forventede verdien er 0. Er disse verdiene større enn 2 eller mindre enn -2 tolkes det generelt til at dataene har mindre kompatibilitet med modellen enn forventet. Negative ZSTD-verdier indikerer mindre variasjon i datamaterialet enn forventet (for forutsigbar), mens positive ZSTD-verdier indikerer mer variasjon i datamaterialet enn forventet (mangel på forutsigbarhet). Dersom MNSQ-verdiene er tilfredsstillende i henhold til anbefalingene nevnt i forrige avsnitt, er det ikke vanlig å undersøke ZSTD-verdiene nærmere.

Som tidligere nevnt kan tilpasningsstatistikk si noe om hvor mye tilfeldigheter som finnes i datamaterialet, og uttrykkes i form av *underfit* eller *overfit*. Underfit skyldes støy fra uberegnelige items eller personer. Det finnes altså observasjoner i datamaterialet som er for uforutsigbare, mer enn det RSM tar høyde for. Overfit skyldes items eller personer som passer for godt overens med RSM. Slike observasjoner kan helle mot et *Guttman-mønster* (Bond & Fox, 2015). Et slikt mønster er uforventet perfekt og urealistisk, og innebærer at de oppgavene en respondent har fått til står helt i forhold til vanskelighetsgraden på oppgaven. Hvis oppgavene er rangert etter vanskelighetsgrad kan personenes respons se slik ut: 111110000, der 1 er korrekte oppgaver og 0 er ukorrekte oppgaver. Typisk for Guttman-mønsteret er at det aldri vil forekomme en korrekt oppgave (1) til høyre for en ukorrekt oppgave (0) i responsrekken. Overfit kan også skyldes at items mangler lokal uavhengighet, altså at mange items er basert på den samme informasjonen (Bond & Fox, 2015).

3.3.4 DTF-analyser

Ved hjelp av RSM kan man gjøre ulike analyser som viser om testen fungerer på samme måte for ulike undergrupper. En av disse er *Differential Test Functioning* (DTF). Målet er at undersøkelsen skal oppfattes likt blant de undergruppene som undersøkes nærmere. I en DTF-analyse undersøkes dette ved å gjøre separate analyser for hver av undergruppene, og deretter sammenligne itemenes vanskelighetsgrad i de ulike undergruppene (Linacre, 2012c). Det er også mulig å gjøre en analyse av *Differential Item Functioning* (DIF), som undersøker om ett bestemt item oppfattes likt for ulike undergrupper. I denne studien nøyer jeg meg imidlertid med en DTF-analyse for å sjekke testens funksjonalitet i undergruppene som en helhet.

3.3.5 Terskelverdier

Siden dikotome data kun gir to mulige utfall for hvert item, gir den dikotome Rasch-modellen kun èn terskelverdi. Denne verdien vil da angi elevens evnenivå når sannsynligheten for at eleven vil svare riktig eller feil på et item er 50% (Bond & Fox, 2015). Med RSM, trengs det imidlertid flere terskelverdier (minst to), for å skille de ulike svarkategoriene på en Likert-skala. Er det *m* svarkategorier i Likert-skalaen benyttet, vil RSM gi *m-1* terskelverdier. Terskelverdiene angir da elevenes evnenivå når eleven har like stor sannsynlighet for å krysse av i èn kategori som den neste kategorien, altså når det er 50% sannsynlighet for at eleven vil krysse av i to påfølgende kategorier. For at terskelverdiene skal kunne gi nyttig informasjon om respondentene, er det viktig at svarkategoriene opptrer i riktig rekkefølge, slik at $\tau_{m-1} > \ldots > \tau_3 > \tau_2 > \tau_1$ (Andrich, 2010). Dette vil jeg også sjekke i resultatkapittelet mitt.

3.3.6 Tolkning av polytome data i det benyttede instrumentet

Begrepene om *evner* og *vanskelighetsgrad*, benyttet til å forklare grunnleggende prinsipper ved Rasch-analyse, har sitt utspring i den dikotome Rasch-modellen. Når jeg i kapittel 4.0 skal analysere et instrument med 24 items i form av påstander, og respons i form av en 6-point Likert-skala, gir ikke disse begrepene nødvendigvis like mye mening. Jeg vil derfor i avsnittene nedenfor tolke disse begrepene og deres betydning mer konkret for det aktuelle instrumentet som benyttes i denne masteroppgaven.

Jeg vil presisere at eksempelet med matematikkprøven, derav evner og vanskelighetsgrad, bidrar til å forklare hvordan den dikotome Rasch-modellen fungerer uavhengig av hva «evner» og «vanskelighetsgrad» svarer til i ulike instrumenter. Når man undersøker holdninger ved hjelp av en Likert-skala, vil ofte begrepet om *item/oppgave* kunne erstattes med *spørsmål* eller *påstand*, *vanskelighetsgrad* kunne erstattes med *intensitet* eller *affektiv verdi* og *evne* erstattes med *holdningsnivå* (Andrich, 1978). I det benyttede instrumentet er itemene formulert som

påstander. Elevene svarer på hvor godt situasjonene beskrevet i påstandene stemmer for dem. Holdningsnivået kan da forstås som graden av gjenkjennelighet elevene rapporterer de oppgitte situasjonene til å ha. Videre i oppgaven finner jeg det derfor hensiktsmessig å bruke begrepet *gjenkjennelighetsgrad* om elevenes holdningsnivå, da dette i større grad beskriver det aktuelle fenomenet. Siden noen av itemene i instrumentet er reversert, vil jeg i påfølgende avsnitt gjøre en liten avklaring med tanke på begrepet om gjenkjennelighetsnivå.

Usher og Pajares (2009) har utformet itemene med den hensikt å vurdere de fire teoretiserte kildene til mestringsforventning. Siden de fire kildene antas å påvirke elevenes mestringsforventning (Bandura, 1977), blir det naturlig å anta at tilstedeværelsen av de positivt rettede itemene (item 1-2 og item 4-18) vil påvirke elevenes mestringsforventning i positiv retning, i den forstand at de bidrar til å bygge opp elevens mestringsforventning. Tilsvarende kan vi anta at tilstedeværelsen av de negativt rettede itemene (item 3 og item 19-24) vil bidra negativt til elevens mestringsforventning. Gjenkjennelighetsgraden er da definert slik at en høyere grad av gjenkjennelighet antas å påvirke mestringsforventningen i positiv retning og visa versa. Instrumentets Likert-skala fungerer da som en slags gjenkjennelighetsskala og det er hvordan elevene svarer på denne skalaen som legger grunnlaget for den enkelte elevs gjenkjennelighetsgrad. Siden item 3 og item 19-24 er reversert i analysene, vil det imidlertid medføre en motsatt tolkning, altså at lavere grad av gjenkjennelighet rapportert på disse itemene (altså svar i lavere svarkategorier) bidrar til en høyere gjenkjennelighetsgrad hos elevene.

Påstandenes intensitet eller affektive verdi kan for dette instrumentet tolkes som hvordan respondentene drar nytte av de ulike situasjonene som fremlegges i påstandene, og sier noe om graden av påstandenes bidrag til elevenes mestringsforventning. For å ikke skape forvirring rundt tolkningen av instrumentet, velger jeg imidlertid å omtale dette som det elevene faktisk har svart på; i hvilken grad påstandene stemmer eller ikke. Jeg vil også beholde den dikotome modellens uttrykk om itemenes vanskelighetsgrad, der vanskelighetsgraden kan tolkes som hvor lett eller vanskelig det er å kjenne seg igjen i itemene. Jo bedre elevene rapporterer et item til å stemme (altså jo høyere itemet rates), jo lavere blir itemets vanskelighetsgrad, fordi det er lettere å kjenne seg igjen i itemet.

Når jeg nå har redegjort for RSM og sentrale begreper knyttet til Rasch-analyser, vil jeg videre beskrive piloteringen av instrumentet og hvordan hoveddatainnsamlingen ble gjort. Siden jeg i resultatkapittelet benytter meg av analyser fra klassisk testteori i tillegg til Rasch-analyser, vil jeg i kapittel 3.6 redegjøre for dette. Her vil jeg også påpeke hvorfor jeg benytter en kombinasjon av analysemetodene. Jeg vil også kommentere forskningsetiske aspekt ved

studien, før jeg i kapittel 4.0 benytter meg av gjør analyser ved hjelp av RSM og klassisk testteori for å besvare mine forskningsspørsmål.

3.4 Pilotering

Som et ledd i utviklingen av undersøkelsen vår foretok vi en pilotering med 41 respondenter. Dette er over anbefalt nedre grense på N > 30 respondenter (Linacre, 1994). Piloteringen skulle både bli gjenstand for analyser for videre utvikling av vår undersøkelse og bidra til praktiske avklaringer rundt gjennomføringen. Når det gjelder sistnevnte formål, var tidsberegning og en felles forståelse av gjennomførelse elementer vi ønsket å avklare. Vi noterte derfor tidsbruk på hver av undersøkelsens tre instrumenter. Erfaringen fra piloteringen var at undersøkelsen tok i underkant av 60 minutter å gjennomføre. Samtlige av studentene i forskningsgruppa SESAM var til stede under piloteringen, samt under plotting av datamaterialet i etterkant. Dette var for å styrke forskningens reliabilitet ved å hindre feilkilder i datamaterialet. Det at dette ble gjort i fellesskap, bidro også til økt reliabilitet i form av at videre datainnsamling og plotting kunne bli utført så likt som mulig. Etter piloteringen noterte vi oss ting som vi tenkte burde informeres tydeligere om i forkant av undersøkelsen. Dette var blant annet for å tydeliggjøre elevenes rolle som deltaker i undersøkelsen, samt hindre feiloppfatninger rundt informasjonen som ble gitt.

Pilotering av undersøkelsen ga oss også et grunnlag for å gjøre Rasch-analyser som vi gjennomførte i samarbeid med vår veileder. Vi utførte analyser både på instrumentet som skulle måle mestringsforventning og det oversatte instrumentet til Usher og Pajares (2009) som skulle måle kildene til mestringsforventning. På grunnlag av analysene av pilotdataene endret vi på og fjernet noen av itemene i førstnevnte instrument. I min masteroppgave velger jeg å ikke gå videre i detaljer knyttet til dette instrumentet, da jeg ikke benytter meg av det videre i analysene. Sistnevnte instrument er derimot viktig for mine analyser. I utgangspunktet er jo dette instrumentet allerede validert, og antas derfor å måle det som det er intendert å gjøre. Oversettelsen av instrumentet skaper likevel en viss usikkerhet, og vi utførte derfor Rasch-analyser på pilotdataene for å sikre validiteten. Siden Usher og Pajares (2009) sitt instrument er ment for å måle fire ulike kilder til mestringsforventning, og RSM fordrer endimensjonalitet, analyserte vi disse kildene hver for seg. Outfit MNSQ-verdiene for de fire kildene var hovedsakelig tilfredsstillende og innenfor anbefalt intervall på 0.6 – 1.4 (Bond & Fox, 2015). Til tross for at noen verdier var utenfor grensene, så vi ikke noen grunn til å endre på oversettelsen av instrumentet. Argumentene for denne avgjørelsen var at det opprinnelige instrumentet allerede er validert fra før, i tillegg til at piloten jo var utført med relativt få respondenter. En detaljert rapportering av instrumentets konstruktvaliditet skjer i kapittel 4.2

med utgangspunkt i hoveddataene. I kommende kapittel gjør jeg rede for hvordan denne datainnsamlingen foregikk.

3.5 Datainnsamling

I dette delkapittelet vil jeg først beskrive den endelige undersøkelsen som ble benyttet under datainnsamlingen. Deretter vil jeg redegjøre for gjennomføringen av datainnsamlingen og utvalget undersøkelsen ble gjort i.

3.5.1 Den endelige undersøkelsen

Den endelige undersøkelsen brukt under datainnsamlingen (se vedlegg 2) ble distribuert i form av et papirhefte bestående av tre instrumenter. På forsiden etterspurte vi informasjon om respondentenes kjønn, fødselsår, fødselsmåned og morsmål. Vi innhentet de tre førstnevnte karakteristikkene fordi vi trengte informasjonen for videre analyser av datamaterialet, mens sistnevnte ble innhentet fordi det ga oss en mulighet til å utelukke språklige ujevnheter.

Instrument 1 (del 1 av undersøkelsen) skulle måle elevers mestringsforventning i matematikk. Elevene ble stilt spørsmålet «*Hvor trygg er du på om du greier å løse ulike matteoppgaver? Svar på en skala fra 1-6, hvor 1 betyr «veldig utrygg» og 6 betyr «veldig trygg»*. Elevene fikk ikke matematikkoppgavene skriftlig; disse ble kun vist elevene på en prosjektor i klasserommet i form av en PowerPoint-presentasjon. Instrumentet bestod av 12 matematikkoppgaver, 3 oppgaver fra hvert av de fire temaene *tall, geometri, statistikk* og *algebra.*

Instrument 2 (del 2 av undersøkelsen) var en oversatt versjon av instrumentet til Usher og Pajares (2009). Som tidligere nevnt, bestod instrumentet av 24 items, og påstandene skulle besvares ved hjelp av en Likert-skala fra 1 *(stemmer ikke i det hele tatt)* til *6 (stemmer veldig godt).* Det var 6 items som omhandlet hver av de fire kildene til mestringsforventning.

Instrument 3 (del 3 av undersøkelsen) var utformet som en typisk matematikkprøve der elevene skulle regne de matematikkoppgavene som ble vist dem i instrument 1. De 12 oppgavene var fordelt over to A4-sider, og inneholdt kun plass til å føre inn svaret på oppgaven. Utregningene kunne elevene gjøre på et kladdeark, som ikke ble samlet inn. Rekkefølgen på oppgavene var ikke den samme som rekkefølgen i det første instrumentet.

3.5.2 Gjennomføring av datainnsamling

Vi gjennomførte datainnsamlingen ved at samtlige av studentene i forskningsgruppa SESAM sendte forespørsler til skoler i sin geografiske nærhet. Vi innhentet data hver for oss (i tråd med gjennomføringsformen vi kom fram til under piloten), og flettet resultatene sammen til slutt. Vi

anser samarbeidet i gruppen SESAM som tidsbesparende, motiverende og mer reliabelt. Samarbeidet ga oss mulighet til å få et større datagrunnlag, og innhente data fra ulike geografiske plasser i Norge. I avsnittene nedenfor vil jeg redegjøre for den praktiske gjennomføringen i klasserommene vi innhentet data fra. Jeg vil også trekke frem enkelte beslutninger vi tok på grunnlag av piloteringen.

Fordi vi var tre studenter som bidro til datainnsamlingen valgte vi, basert på piloteringen, å skrive et dokument med retningslinjer for hva vi skulle si under gjennomføringen i klasserommene (se vedlegg 3). Dette var et bevisst valg for å sikre en så lik gjennomførelse i alle klasser som mulig. Retningslinjene var også en hjelp for å tydeliggjøre de elementene vi gjennom piloten hadde sett et behov for å tydeliggjøre, samt huske å informere om det vi skulle. Viktige punkter som fremgikk av retningslinjene og som ble tydeliggjort under gjennomførelsen var for eksempel undersøkelsens anonymitet (for å hindre at elevene skrev navnet sitt på arket), instrukser for avkrysningen (for å hindre at elever krysset av mellom to ruter og sikre entydige svar) og at det ikke var lov til å samarbeide om oppgavene.

Dessuten presiserte vi at undersøkelsen ikke var en prøve, at de ikke skulle få karakter på den, og at vi ikke skulle teste hvor mye matematikk elevene kunne. Dette for å hindre at situasjonen skulle oppfattes som en testsituasjon. Haciomeroglu (2019), som har forsket på forholdet mellom elevers mestringsemosjoner og deres mestringsforventninger, påpeker at elevers mestringsforventninger må undersøkes i ulike settinger. Han nevner for eksempel at bekymring kan påvirke elevenes mestringsforventninger i en prøvesituasjon. Kontekstuelle faktorer kan ha betydning for hvordan elever tolker informasjonen fra de ulike kildene (Bandura, 1997; Zimmerman, 1995). Vi ønsket derfor å være bevisst på hvilken kontekst undersøkelsen ble gjort i, og utføre undersøkelsen i en vanlig klasseromssituasjon.

Etter å ha informert elevene om undersøkelsen, og kort om undersøkelsens tre instrumenter (definert som del 1, 2 og 3 for elevene), delte vi ut heftene og ba elevene om å fylle ut avkrysningene omhandlende kjønn, fødselsår, fødselsmåned og morsmål. Deretter forklarte vi hvordan del 1 skulle foregå, viste med en eksempeloppgave, og satte i gang med del 1. Hver matematikkoppgave ble lest høyt, før elevene fikk ca. 5 sekunder til å se på oppgaven. Deretter bladde vi videre til en blankt ark i presentasjonen, slik at elevene fikk tid til å sette et kryss i heftet sitt for den aktuelle oppgaven. Vi valgte å gi elevene forholdsvis kort tid til å studere oppgaven, for å hindre at elevene skulle begynne å regne ut oppgaven. Vi presiserte også muntlig at elevene ikke skulle regne ut oppgavene, men derimot se på oppgaven og vurdere

hvor trygge de følte seg på at de kunne regnet den ut hvis de skulle ha gjort det. Vi brukte ordet *magefølelse* for å gi elevene en forståelse av denne vurderingen.

Når del 1 var ferdig, informerte vi elevene om del 2 og 3, og sa at elevene kunne gå direkte videre til del 3 når de var ferdige med del 2. Dette fungerte godt, og alle kunne vurdere og fylle ut påstandene i del 2 i sitt eget tempo. Når elevene begynte på del 3, der de skulle regne ut ulike oppgaver, fikk elevene utdelt et kladdeark de kunne bruke til utregninger. Heftet skulle kun inneholde selve svaret, og vi tok ikke inn kladdearkene når elevene var ferdig. Elevene fikk bruke så lang tid de ønsket på denne delen, og kunne levere når de følte seg ferdig. Undersøkelsen kunne ikke være tidsbegrenset fordi tidsbruk da ville blitt en forstyrrende faktor (Bond & Fox, 2015). Et fåtall elever ble ikke ferdig med undersøkelsen innen den praktiske tidsbegrensningen av en skoletime. Disse besvarelsene markerte vi med en x, og ble ikke tatt med som respondenter i analysene.

3.5.3 Utvalg

Utvalget i denne studien består av 583 elever på 9.trinn i Norge. *Populasjonen*, altså de enhetene problemstillingen gjelder for, er elever på 9.trinn i Norge (Christoffersen & Johannessen, 2012). I og med at vi ikke har hatt mulighet til å innhente svar fra hele populasjonen, har vi gjort et utvalg. For at man med størst mulig sikkerhet kan si noe om hele populasjonen på grunnlag av et mindre utvalg, bør utvalget være representativt for populasjonen. Ved et sannsynlighetsutvalg vil sannsynligheten for å få et representativt utvalg være stor. Dette ville imidlertid krevd en tilfeldig utvelgelse, for eksempel en trekning blant alle landets 9.klassinger, noe som i denne studien ikke ville vært mulig. For å kunne utføre undersøkelsen korrekt og på samme måte i alle klasser, var vi avhengig av at en i forskningsgruppa var til stede under gjennomføringen. Samtlige av studentene i SESAM har besøkt skoler i sin geografiske nærhet, noe som har medført at elever ved skoler fra både Sør-, Øst- og Vestlandet er representert i datamaterialet. Tilgjengelighet ble dermed en viktig faktor i utvalgsmetoden. Det var heller ikke noe selvfølge å få tillatelse til å gjennomføre undersøkelsen på alle skoler, da dette krevde at vi fikk en skoletime til rådighet. Utvalgsmetoden i denne studien er et bekvemmelighetsutvalg, og dermed ikke et sannsynlighetsutvalg (Johannessen et al., 2016). Dette er av rent praktiske årsaker, som blant annet studiens tidsbegrensning og omfang. Det er imidlertid ikke kjente faktorer som skulle tale mot at utvalget er representativt for populasjonen. Man skal likevel være forsiktig med å generalisere funn fra utvalget til hele populasjonen, da man ikke kan være sikker på at utvalget er representativt for populasjonen (Postholm & Jacobsen, 2011).

I utgangspunktet ønsket vi i SESAM å innhente svar fra opp mot 1000 respondenter. Situasjonen rundt COVID-19 gjorde imidlertid prosessen med å innhente data noe utfordrende. Coronarestriksjonene gjorde det vanskelig å få tillatelse til å komme rundt på skoler, samt at det tidvis var hjemmeskole for mange elever. Innsamlingen av data ble dermed brått vanskeligere, og vi ble derfor nødt til å si oss fornøyd med våre 583 respondenter. Totalt kontaktet vi 38 skoler, og fikk lov til å gjennomføre undersøkelsen på 13 av skolene. For å få tilgang til skolene har vi kontaktet rektorer, trinnledere eller lærere ved skolene.

Når jeg i første del av mine analyser vurderer hvorvidt instrument 2 er valid, gjøres det på grunnlag av undersøkelsens 583 respondenter. Senere i analysene ser jeg på høytpresterende og lavtpresterende elever. Det vil si at jeg i den delen av studien ikke baserer mine analyser på hele utvalget på 583 respondenter. Utvelgelsen av de to undergruppene har foregått på bakgrunn av instrument 3, som gir en pekepinn på elevenes prestasjon i form av en typisk matematikkprøve. I kapittel 4.2 validerer jeg instrument 2, som måler elevenes kilder til mestringsforventning, ved hjelp av Rasch-analyser. Dette instrumentet brukes i begge de to overnevnte delene av analysen. Det ville imidlertid blitt for omfattende å også skulle validere instrument 3, som måler elevenes prestasjon i matematikk. Min utvelgelse av høyt- og lavtpresterende elever tar derfor utgangspunkt i poengscore målt ved det tredje instrumentet av undersøkelsen. Jeg anser inndeling av undergruppene etter poengscore som godt nok for denne studiens hensikt. I tillegg støtter jeg meg til at dette er vanlig praksis i skolen også - elevene blir ikke gitt prøver med vektede poeng, og det er kanskje nettopp på denne måten man som lærer vil definere høyt- og lavtpresterende elever. I de neste avsnittene vil jeg forklare nærmere hvordan inndelingen av de to undergruppene ble gjort.

Av de 12 oppgavene i del 3 av undersøkelsen er det 18 respondenter som ikke besvarte noen av de disse korrekt. Det kan være flere grunner til dette. Typisk for disse besvarelsene var at elever hadde skrevet spørsmålstegn på alle oppgavene. Det gir grunn til å tro at noen elever kan ha manglet motivasjon for å gjennomføre undersøkelsen, og dermed ga seg fort når de stod fast på oppgavene. Jeg som forsker hadde ikke kjennskap til elevene fra før og visa versa. Altså hadde elevene ingen relasjon til meg som skulle tilsi at de av den grunn ønsket å yte sitt beste av ren «tillit» eller for å «ivareta relasjonen». De fikk også beskjed om at det kun var jeg og forskningsgruppa mi som skulle se besvarelsen deres, og at dette ikke var en matematikkprøve. Til tross for lærerens tilstedeværelse under gjennomførelsen, kunne heller ikke elevenes relasjon til læreren bidra til større innsats. Dessuten var flertallet av våre respondenter vant med å få karakter på prøvene de gjorde. Karakter kan vanligvis være en viktig faktor for

motivasjonen bak å prestere godt (Skagen, 1996). Elevene visste at denne undersøkelsen ikke ville påvirke deres karakter i matematikk, noe som kan ha bidratt til å senke elevenes motivasjon for å gjøre sitt beste. I tillegg var undersøkelsen anonym, som også utelukker at elevenes motivasjon kan ha blitt økt på grunn av et ønske om å ivareta faglig status eller liknende. Vi informerte om at deltakelse på undersøkelsen var frivillig, men fordi læreren hadde godtatt og satt av en skoletime til gjennomførelse av undersøkelsen, vil jeg si at det likevel lå en slags forventing til at elevene skulle delta. Alle disse faktorene gir grunn til å tro at det ikke nødvendigvis er elevenes lave evne i matematikk som gjorde at de fikk score 0 på del 3 av undersøkelsen. Nettopp fordi dette er svært uvisst, og vanskelig å si noe sikkert om, har jeg vurdert det slik at jeg velger å ikke bruke disse besvarelsene i undergruppen for lavtpresterende elever.

I inndelingen av høyt- og lavtpresterende elever har jeg fulgt anbefalingen til Johannessen et al. (2016) om at antallet i sentrale undergrupper bør være representert med over 100 stk. Det var 37 elever som fikk 1/12 korrekte svar og 76 elever som fikk 2/12 korrekte svar. Til sammen utgjør dette 113 elever, og det er altså disse elevene som inngår i gruppen *lavtpresterende elever*. Det var ingen elever som fikk 12/12 rette på del 3, men et antall på 6, 10, 18, 27, 49 som fikk henholdsvis 11/12, 10/12, 9/12, 8/12 og 7/12 rette svar. Elever som fikk en score på 7-11 poeng utgjør altså gruppen med *høytpresterende elever*, som til sammen blir på 110 elever.

Ut ifra antallet elever som fikk til de ulike oppgavene, kan det tyde på at nivået på oppgavene var litt høyt i forhold til elevenes kunnskap. Dette gjør det forholdsvis enkelt og valid å sortere ut høytpresterende elever. Disse blir svært synlige i datamaterialet. Men det at veldig mange har en svært lav score på testen, medfører at det jeg kategoriserer som lavtpresterende elever egentlig kan være en blanding av elever på middels og lavt nivå. I min studie har jeg grunnet oppgavens omfang ikke hatt mulighet til å sortere mer detaljert. Jeg kan heller ikke utelukke at noen av de lavtpresterende elevene kan ha matematikkvansker eller andre diagnoser. Ifølge Ostad (2010, gjenngitt av (Tryggestad & Eldevik, 2015)) anslås det at rundt 10% av elever har matematikkvansker. Det vil altså være naturlig å anta at en andel av mine lavtpresterende respondenter vil kunne ha matematikkvansker. Dette er en faktor jeg ikke tar hensyn til i denne studien, da dette er utenfor min kontroll. Begrepet om høyt- og lavtpresterende elever i denne studien kan kanskje sies å være noe uklart. Samtidig vil jeg peke på at tidligere forskning og litteratur ikke har en klar definisjon på hva som kan regnes som høyt- og lavtpresterende elever. Dette vil altså være noe forskjellig i alle studier. I min studie tror jeg likevel ikke at den valgte

inndelingen av disse vil begrense studiens potensiale til å besvare problemstillingen, men derimot bidra til å påpeke nyttige aspekt ved den.

3.6 Klassisk testteori

I kapittel 3.3 har jeg begrunnet hvorfor jeg ønsker å ta i bruk Rasch-analyser i denne studien. Bruken av Rasch-analyser erstatter imidlertid ikke analysemetoder innen klassisk testteori, men vil snarere ment for å supplere og komplementere disse analysemetodene. I dette delkapittelet vil jeg derfor først forklare hvordan jeg kombinerer Rasch-analyser med klassisk testteori og deretter redegjøre for de ulike analysene innen klassisk testteori som benyttes i studien.

Som tidligere nevnt vil Rasch-analyser være et nyttig verktøy for å koordinere og tilpasse et instrument til en spesifikk kontekst (Cavanagh & Fischer, 2018). Rasch-analyser vil også bidra til å omforme en variabel på ordinalnivå til en kontinuerlig variabel. Dette er kanskje det som av flere forskere fremheves som det viktigste argumentet for at Rasch-analyser gir et mer nyansert og korrekt resultat enn klassisk testteori (Bond & Fox, 2015; Boone, 2016; Cavanagh & Fischer, 2018). Bond og Fox (2015) kritiserer den klassiske testteorien for å kun foreta beskrivelser av rådata, og at testscores summeres opp og behandles som kontinuerlige data, til tross for at målingene egentlig ble gjort på ordinalnivå, eksempelvis ved bruk av en Likert-skala. Dette er en begrensning for klassisk testteori, fordi det trolig medfører unøyaktighet og målefeil (Cavanagh & Fischer, 2018). Når jeg i denne studien benytter klassisk testteori for å gjennomføre analyser av datamaterialet, har jeg imidlertid allerede benyttet Rasch-analyser til å omforme målenivået på de innsamlede dataene, og analysene innenfor klassisk testteori kan gjøres på de kontinuerlige dataene. På denne måten bidrar kombinasjonen av klassisk testteori og Rasch-analyser til en solid behandling av datamaterialet.

Analyser i klassisk testteori som benyttes i denne oppgaven er hypotesetester som t-test og ANOVA. SPSS 26.0 (Statistical Packages for the Social Sciences) ble benyttet som analyseprogram. I de kommende avsnitt vil jeg redegjøre for de nevnte analysene, samt z-scores. Først jeg vil jeg imidlertid kort redegjøre for prinsippene bak hypotesetesting.

3.6.1 Hypotesetester

Statistiske analyser benyttes i denne studien for å vurdere generaliseringsstyrken på funnene i utvalget. Som tidligere nevnt benyttes et utvalg på N=583 norske 9.trinnselever, og det jeg ønsker å teste er hvorvidt forskjeller mellom høyt- og lavtpresterende elevers kilder til mestringsforventning i dette utvalget kan generaliseres til hele populasjonen. Dersom det eksisterer en forskjell mellom de høyt- og lavtpresterende elevenes kilder til

mestringsforventning i mitt utvalg elever, vil min hypotese være at denne forskjellen også eksisterer i hele populasjonen. Nullhypotesen blir da at forskjellen ikke eksisterer i hele populasjonen. Siden det alltid vil være en viss sjanse for at det er tilfeldigheter i mitt utvalg som viser en forskjell mellom undergruppene, velger jeg et signifikansnivå (α) som angir hvor stor feilmargin jeg kan operere med. Signifikansnivået settes til $\alpha = 0.05$. Dersom testen viser at $p<0.05$ er forskjellen signifikant, og medfører at jeg kan forkaste nullhypotesen og konkludere med at forskjellen observert i utvalget kan generaliseres til hele populasjonen. Det betyr at det ville vært under 5% sjanse for å få et liknende resultat blant et annet utvalg elever, dersom denne forskjellen ikke hadde vært til stede i hele populasjonen. Det betyr også at jeg med 95% sikkerhet vil kunne si at skjevheten i utvalget også eksisterer i hele populasjonen. Med et signifikansnivå på $\alpha = 0.05$, betyr altså at jeg godtar en feilmargin på 5% når jeg vurderer om en forskjell i undergruppene i mitt utvalg kan generaliseres til hele populasjonen.

En hypotese kan være ensidig eller tosidig. En tosidig hypotese er ikke retningsbestemt, og i mitt eksempel ovenfor vil hypotesen være at det er en forskjell mellom høyt- og lavtpresterende elevers rapportering av kildene (Nardi, 2018). Dersom teori og empiri gir en tydelig retning for hva man kan forvente, vil man kunne benytte en ensidig hypotese. Denne er retningsbestemt og vil i mitt tilfelle være en hypotese om at høytpresterende elever rapporterer å kjenne seg mer igjen i situasjoner som antas å føre til høyere mestringsforventning enn lavtpresterende elever.

3.6.2 Uavhengig t-test

T-tester benyttes for å teste om det er signifikant forskjell mellom gjennomsnittet i to grupper. I denne studien er det høyt- og lavtpresterende elever som utgjør de to gruppene. Da må jeg benytte en uavhengig t-test, fordi datasettene kommer fra to forskjellige utvalg der ingen av elevene er representert i begge undergruppene. For å kunne bruke en t-test må den avhengige variabelen være kontinuerlig (Nardi, 2018). Som tidligere nevnt, sørget Rasch-analysen for at variabelen som opprinnelig var på ordinalt målenivå ble konvertert til en kontinuerlig variabel. For å utføre t-testen benyttet jeg meg av de to undergruppenes gjennomsnitt og standardavvik, samt utvalgsstørrelsen på de to undergruppene. T-testen gir testobservatoren T i retur, sammen med sannsynligheten p for å kunne få en slik testobservator i et utvalg med den gitte størrelsen og de gitte standardavvikene, ved tilfeldigheter, til tross for at skjevheten ikke skulle eksistere i hele populasjonen (Nardi, 2018). Dersom $p<\alpha$ (0.05) forkaster jeg altså nullhypotesen, og kan konkludere med at forskjellen i de to undergruppene er signifikante, og dermed kan generaliseres til hele populasjonen. Dersom man tester en ensidig hypotese i t-testen, må p-verdien som SPSS gir (som antar en tosidig hypotese), deles på to (IBM, 2020). Med et

konfidensintervall på 95%, vil dette bety at 2.5% i hver retning medfører signifikans. Siden hypotesen er ensidig, kan p-verdien derfor deles på to.

For å kunne uttale meg om effektstørrelsen på eventuelle forskjeller, gjør jeg en utregning av *Cohen's d,* basert på gruppenes utvalgsstørrelse, gjennomsnitt og standardavvik. En tommelfingerregel for effektstørrelsen sier at d=0.20 indikerer en svak sammenheng, d=0.50 indikerer en moderat sammenheng og d=0.80 indikerer en sterk sammenheng (van den Berg, u.å.). Effektstørrelsen forteller størrelsen på effekten som den uavhengige variabelen har på den avhengige variabelen (Cohen, Manion & Morrison, 2007). I min studie vil effektstørrelsen si noe om hvilken grad forskjellene mellom høyt- og lavtpresterende elevers kilder til mestringsforventning kan forklares av prestasjonsnivået.

3.6.3 One-way repeated measure ANOVA

ANOVA (Analysis of Variance) benyttes for å sammenligne gjennomsnittet i mer enn to grupper samtidig. I likhet med t-testen, fordrer ANOVA at den avhengige variabelen er kontinuerlig, slik at gjennomsnittet og standardavvik kan beregnes (Nardi, 2018). ANOVA-analysen undersøker om forskjellen innad i en gruppe er større eller mindre enn forskjellene mellom gruppene. I denne studien benyttes ANOVA til å sammenligne forskjellene mellom de ulike kildene innenfor hver av de to undergruppene. Dette medfører altså analyser av høyt- og lavtpresterende elevers kilder hver for seg, der det er de fire kildene innenfor hver elevgruppe som utgjør undergruppene jeg vil sammenligne gjennomsnittet i. Siden det er de samme elevene som er representert i de fire undergruppene (hver elev har svart på påstander innenfor alle de fire kildene), benytter jeg meg av det som kalles *one-way repeated measure ANOVA*.

For å kunne ta i bruk one-way repeated measure ANOVA stilles det 5 krav til datamaterialet (StatsTest, u.å.). (1) Den avhengige variabelen må være kontinuerlig, noe den er grunnet omgjøringen i Rasch-analysen (beskrevet i kapittel 6.2.1) og en omgjøring til z-scores (beskrevet i påfølgende delkapittel). (2) Variabelen må være normalfordelt. (3) Dataene må komme fra et tilfeldig utvalg (sannsynlighetsutvalg). Utvalgsmetoden benyttet i denne studien er ikke et sannsynlighetsutvalg. Som nevnt i kapittel 3.5.3 er det imidlertid ikke kjente faktorer som skulle tilsi at utvalget ikke er representativt. Jeg argumenterer derfor for at også dette kravet er oppfylt. (4) For å benytte denne ANOVA-analysen bør det være flere enn 5 respondenter i hver undergruppe. Dette kravet er tilfredsstilt, med N=110 høytpresterende elever og N=113 lavtpresterende elever. (5) Sfærisitet må være tilfredsstilt. Kravet om sfærisitet handler om at variansen til differansen mellom alle mulige parkombinasjoner av kildene bør være tilnærmet

lik (Bergsaker, 2019). I analysen kjører jeg også en test for å sjekke dette, kalt *Mauchly's test of sphericity*. Hvis denne viser $p<0.05$, kan vi forkaste nullhypotesen om at variansen er lik (Lærd Statisics). Dette medfører at kravet om sfærisitet ikke er tilfredsstilt. Hvis dette er tilfelle, kan man benytte en Greenhouse-Geisser-korrigering av frihetsgrader, som også rapporteres i ANOVA-analysen utført i SPSS. Dersom Mauchly's test of sphericity viser $p>0.05$, kan man gå ut ifra at sfærisiteten er tilfredsstilt.

Signifikansveriden *p* i en one-way repeated measure ANOVA gir informasjon om forskjellen mellom gjennomsnittet i undergruppene, men sier ikke noe om hvor stor en eventuell forskjell viser seg å være (Nardi, 2018). Dersom analysen viser signifikant forskjell mellom gruppene, kan det derfor være interessant å vite hvor stor denne forskjellen er. Dette ser jeg ved å se på *partial eta squared* (η^2). For å uttale meg om effektstørrelsen i resultatet av ANOVA-analysen (kapittel 4.2.2), benytter jeg en tommelfingerregel om at $\eta^2 = 0.01$ indikerer en svak sammenheng, $\eta^2 = 0.06$ indikerer en moderat sammenheng og $\eta^2 = 0.14$ indikerer en sterk sammenheng (van den Berg, u.å.)

Et signifikant resultat fra one-way repeated measure ANOVA, betyr at det generelt er en forskjell mellom gjennomsnittet i undergruppene. Analysen viser imidlertid ikke hvilke grupper disse forskjellene eksisterer mellom. Har man fått et signifikant resultat av analysen, kan det være lurt å kjøre en *Bonferroni post hoc – test* for å avgjøre hvor forskjellene ligger (Lærd Statisics). Denne testen viser parvise sammenligninger av alle undergruppene.

3.6.4 Omgjøring til z-scores

For å kunne sammenligne fenomener med ulike måleenheter kan man standardisere målene ved å bruke z-scores (Nardi, 2018). Z-scores kalles standardiserte fordi gjennomsnittet settes lik 0 og standardavviket lik 1 (Foldnes, Grønneberg & Hermansen, 2018). Z-scoren angir hvor mange standardavvik en observasjon ligger fra gjennomsnittet. En z-score fås ved å subtrahere gjennomsnittet i gruppen fra hver enkelt respondents score, og deretter dele på standardavviket.

3.7 Forskningsetikk

Som forsker har man et ansvar for å sikre at forskningen skjer i henhold til forskersamfunnets grunnleggende normer og verdier (NESH, 2016). Forskningsprosjekter skal meldes til Norsk senter for forskningsdata (NSD) og er pliktet til å følge kravene NSD stiller til personvern. I forkant av vår studie kontaktet vi NSD og sendte inn et meldeskjema hvor vi beskrev prosjektet. Undersøkelsen vi ville foreta skulle være anonym, men krevde fødselsår og fødselsmåned,

kjønn og morsmål (norsk eller annet) fra respondentene våre. Dette informerte vi om i meldeskjemaet. Vi fikk da tilbakemelding om at prosjektet ikke behøvde en vurdering fra NSD, fordi vi ikke skulle behandle personopplysninger (se vedlegg 1).

I NESHs forskingsetiske retningslinjer står det at forskeren ved behandling av personopplysninger og sensitiv informasjon har en lovfestet informasjonsplikt og må innhente samtykke fra deltakerne i forskningen (NESH, 2016). Siden vårt forskningsprosjekt ikke behandler personopplysninger, var det ikke behov for å innhente samtykke fra elevene eller deres foreldre. I forkant av undersøkelsen informerte vi klart og tydelig om at undersøkelsen var både frivillig og anonym, og at det kun var vi som forskergruppe som ville kunne se deres besvarelser. De innsamlede dataene er konfidensielt behandlet, og det vil ikke være mulig å identifisere respondentene i etterkant. Jeg vil altså konkludere med at respondentenes integritet er ivaretatt.

Jeg har nå redegjort for metoden som er benyttet i denne studien. På grunnlag av dette vil jeg videre i kapittel 4.0 benytte RSM som analyseverktøy, i tillegg til analysene beskrevet i kapittel 3.6, for å besvare mine forskningsspørsmål. Jeg vil både undersøke kvaliteten til instrument 2, og analysere datamaterialet med formålet om å kunne si noe om høyt- og lavtpresterende elevers kilder til mestringsforventning.

4.0 Resultater

Resultatkapittelet er todelt. I første del vil jeg presentere resultatet av det analysearbeidet som ble foretatt i forlengelsen av arbeidet og analysene som ble gjort i piloten (og beskrevet i kapittel 3.4). Denne delen vil altså omhandle forskningsspørsmålet om hvordan et instrument for å måle elevenes kilder til mestringsforventning kan oversettes og valideres for norske elever. Instrumentet det her er snakk om er det som i forrige kapittel ble definert som instrument 2. I tråd med Bond og Fox (2015) vil jeg altså først bruke tid på å undersøke instrumentet, før jeg deretter, i del 2, benytter instrumentet til å undersøke om det er forskjeller i høyt- og lavtpresterende elevers rapportering av kilder til mestringsforventning. I inneværende kapittel velger jeg for enkelhets skyld å benytte meg av forkortelser for de fire kildene til mestringsforventning: Mestringserfaringer (ME), vikarierende erfaringer (VE), oppmuntring, støtte og overtalelse fra andre (OA) og psykologiske og fysiologiske tilstander (PFT).

4.1 Instrumentets validitet og målingenes reliabilitet

I denne studien har jeg adaptert instrumentet *Sources of Middle School Mathematics Self Efficacy Scale*, utviklet av Usher og Pajares (2009). Usher og Pajares validerte instrumentet for 6. - 8.trinnselever, stort sett hvite, med middels og høy sosioøkonomisk status i sørøstlige deler av USA. Min intensjon ved å ta i bruk dette instrumentet var å kunne vurdere norske 9.trinns-elevers kilder til mestringsforventning. Som en del av denne studien ønsker jeg derfor å undersøke om en oversatt versjon av instrumentet er valid for et utvalg som er både geografisk, etnisk og sosiokulturelt forskjellig fra utvalget i valideringsstudien (Usher & Pajares, 2009). Valideringsprosessen nedenfor gjelder det oversatte versjonen (se del 2, vedlegg 2) av instrumentet som i sin opprinnelige form gjengis i vedlegg 5. Jeg vil også sjekke instrumentets reliabilitet. Formålet med å sjekke validitet og reliabilitet, er å forsikre meg om at instrumentet fungerer slik det er tenkt. God reliabilitet og validitet er viktig for å videre kunne benytte meg av dataene innhentet ved instrumentet, og diskutere og trekke konklusjoner på grunnlag av dets analyser.

Som allerede nevnt i kapittel 3.5.1 måler instrumentet til Usher og Pajares (2009) fire ulike kilder til mestringsforventning. Teori om de fire kildene (Bandura, 1997) gir ikke grunnlag for å tro at et instrument som måler de fire kildene samlet, skal måle et endimensjonalt konstrukt. Teorien skiller jo tydelig de fire ulike kildene fra hverandre. I valideringen behandles derfor kildene separat (som i piloten), slik at itemene som omhandler hver av kildene undersøkes som egne delinstrumenter:

1. Mestringserfaringer (ME-instrument): Item 1 – 6
2. Vikarierende erfaringer (VE-instrument): Item 7 – 12
3. Oppmuntring, støtte og overtalelse fra andre (OA-instrument): Item 13 - 18
4. Psykologiske og fysiologiske tilstander (PFT-instrument): Item 19 - 24

Jeg vil rapportere både reliabilitet og validitet. For å kunne si noe om validiteten til itemene i instrumentet vil jeg se på konstruktvaliditeten og følger samme prosess som brukes av Pampaka et al. (2012) og Bjerke og Eriksen (2016), som igjen følger retningslinjene til Wolfe og Smith (2007). Siden jeg i mine analyser behandler hver kilde som et instrument i seg selv, vil jeg i realiteten ende opp med å rapportere fire sett med analyser. Jeg kommer derfor først til å forklare hvilke analyser jeg gjør for hvert delinstrument. I det følgende vil jeg beskrive analyser av reliabilitet, outfit MNSQ, item measure, item polarity, item characteristics curves og differential test functioning. Etter det kommer resultatene av analysene knyttet til hvert delinstrument.

Reliabilitet

Reliabilitet rapporterer reproduserbarheten av scorene innhentet ved instrumentet. Reliabiliteten har altså ikke noe å gjøre med kvaliteten på instrumentet, men sier noe om målingenes pålitelighet (Linacre, 2012b). Rasch-analysene gir to reliabilitetsmål, både *person reliability* og *item reliability*. *Person reliability* angir hvor reproduserbar rekkefølgen på personenes gjenkjennelighetsgrad er for det bestemte utvalget på de gitte itemene (Linacre, 2012b). Dette målet på reliabilitet er det samme som man vanligvis rapporterer i klassisk testteori (Linacre, u.å). Winsteps rapporterer i tillegg *item reliability*, som sier noe om hvor reproduserbar rekkefølgen på itemenes vanskelighetsgrad er for det gitte settet med items for det bestemte utvalget. Linacre (2012b) sier som en tommelfingerregel at dersom *item reliability* er mindre enn 0.8, trenger man et større utvalg og at dersom *person reliability* er mindre enn 0.8, trenger man flere items.

Outfit MNSQ

Når hver av kildene i Usher og Pajares' (2009) instrument analyseres for seg, bør vi kunne forvente at hvert av delinstrumentet er endimensjonalt. Fremgår noe annet av analysene bør dette undersøkes nærmere. Tilpasningsstatistikk undersøkes for å avdekke om det skjuler seg flere dimensjoner i ME-instrumentet, og testes ved hjelp av fit-analyser. Som tydeliggjort i kapittel 3.3.3 rapporterer jeg outfit MNSQ-verdier, som er en chi-kvadrat-statistikk og en vanlig verdi å rapportere (Linacre, 2012a). Har itemene i instrumentet outfit MNSQ-verdier mellom 0.6 og 1.4, er det et tegn på at konstruktet er endimensjonalt (Linacre, u.å.).

Item measure

I Rasch-analysene er de ulike itemene presentert i rekkefølge basert på itemenes measure. Hvert item i Rasch-analyser blir målt av logit-skalaen, og spredningen på itemene kan si noe om hvordan itemene bidrar til å måle elevene i undersøkelsen. Er det store sprik i itemenes measure, vil det indikere at elever med en gjenkjennelsesgrad i dette spriket ikke blir målt så nøyaktig av instrumentet. Dersom flere items har omtrent samme measure, vil det indikere at det er flere items som måler de samme elevene. Dette vil bety at itemene overlapper hverandre noe.

Item polarity

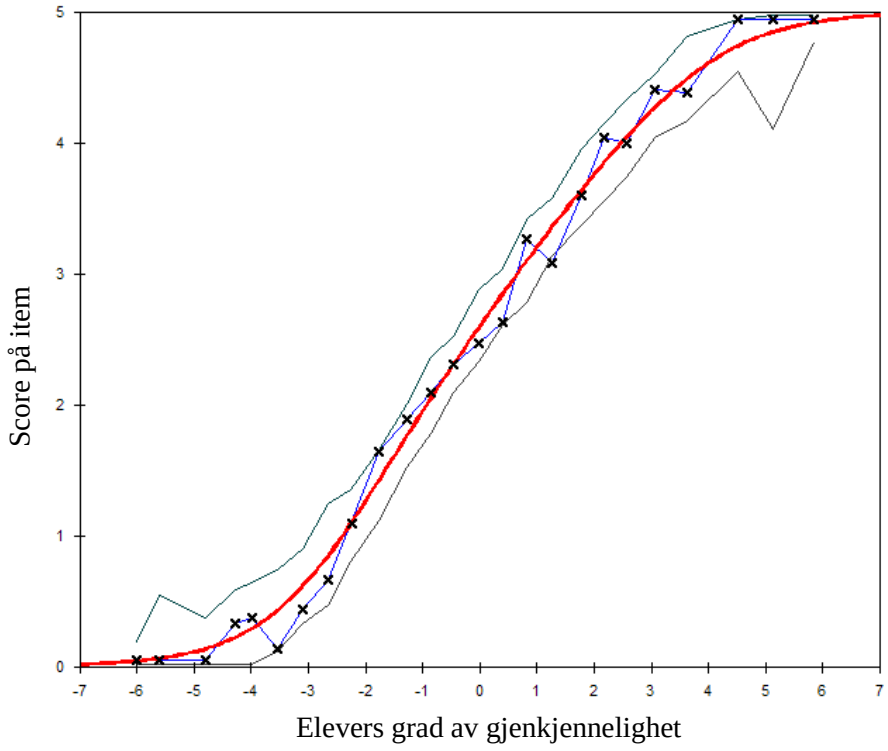
Jeg vil også undersøke «item polarity». Hensikten med dette er å sjekke om responsen på ett bestemt item stemmer overens med det totale responsmønsteret, altså om hvert enkelt item samsvarer med elevenes grad av gjenkjennelighet (Linacre, 2012a). RSM forventer at (1) en elev med høyere grad av gjenkjennelighet vil rapportere itemene til å stemme bedre, og (2) når itemene rapporteres til å stemme godt, har elevene høyere grad av gjenkjennelighet. Item polarity rapporterer graden av hvor godt dette stemmer for hvert item. I resultattabellene som rapporteres i de kommende delkapitlene ser vi dette av kolonnen «observert korrelasjon», som uttrykker korrelasjonen mellom hvor godt det bestemte itemet er rapportert til å stemme for elevene og elevenes grad av gjenkjennelighet. Vi ønsker å se høye, positive korrelasjoner, da dette betyr at responsen på itemet stemmer godt overens med elevenes grad av gjenkjennelighet. Er verdiene negative eller nær null, indikerer det problematiske items som ikke er konsistent med resten av konstruktet (Bond & Fox, 2015). En negativ verdi betyr at itemet motstrider forventningen om at jo høyere grad av gjenkjennelighet en elev er målt til, jo mer sannsynlig er det at itemet stemmer bedre for eleven (Linacre, 2012a). Negative korrelasjoner bør undersøkes nærmere, da disse itemene vil ødelegge for målingen. En negativ korrelasjon kan komme av at ordlyden i itemet går motsatt vei av de resterende itemene. Som nevnt i kapittel 3.2.3 har noen av itemene i instrumentet en ordlyd som går motsatt vei av de resterende itemene, noe som er hensyntatt ved å reversere disse itemene i analysen. Reverseringen vil altså korrigere for items som ellers ville resultert i negative korrelasjoner i denne analysen. Dersom korrelasjonen er nær null, kan det skyldes at itemet enten er svært lett eller svært vanskelig å kjenne seg igjen i, og dermed føre til at itemet ikke samsvarer så godt med elevenes grad av gjenkjennelighet. Er man usikker på om en korrelasjon er for lav, kan verdien sammenlignes med den korrelasjonen RSM forventer. Dette fremgår av resultattabellene i kolonnen «forventet korrelasjon».

Ved item polarity vil jeg, som nevnt i kapittel 3.3.5, også sjekke at de ulike svarkategoriene kommer i riktig rekkefølge for alle items. Dette vil være et tegn på at elevenes grad av

gjenkjennelighet stemmer overens med hvor godt påstandene stemmer. Disse analysene kan ikke leses av resultattabellene, men er lagt ved som vedlegg.

Item Characteristics Curves

Jeg inspiserer også Item Characteristics Curves (ICC) for hvert item i instrumentet. Dette vises i en graf som kan se som i figur 2, og som kan bidra til å avdekke eventuelle tilleggsdimensjoner i instrumentet. Den røde grafen viser den forventede interaksjonen mellom personer og items dersom dataene hadde passet Rasch-modellen perfekt (Bond & Fox, 2015). Den blå punktlinjen viser den faktiske interaksjonen. De to grå-grønne linjene på hver side av den røde grafen viser konfidensintervallet, der den vertikale avstanden fra den røde linjen er 1.96 standardavvik (Linacre, 2012a). Når punktlinjen ligger innenfor konfidensintervallet og tett opptil den røde grafen for alle items, er det ikke grunn til å mistenke en tilleggsdimensjon, fordi itemene oppfører seg slik modellen forventer.

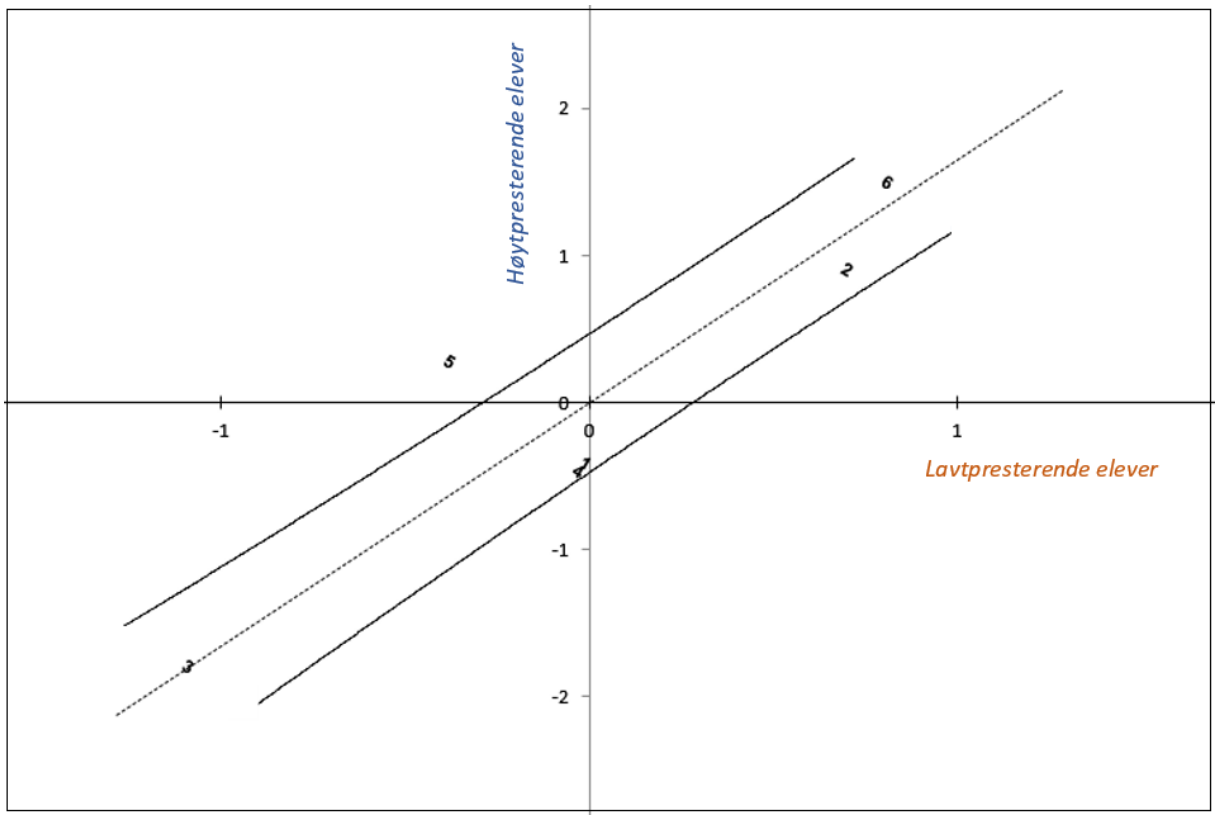


Figur 2. *Eksempel på Item Characteristic Curve (ICC) for et item.*

Differential Test Functioning

For å sikre at instrumentet blir oppfattet på samme måte i de to undergruppene i denne studien, vil jeg gjøre Differential Test Functioning-analyser (DTF-analyser), også beskrevet i kapittel 3.3.4 (Linacre, 2002). Analysene fremstilles i form av et spredningsdiagram, slik figur 3 er et eksempel på. I min studie er det lavtpresterende elever og høytpresterende elever som utgjør de

to undergruppene. Denne analysen er relevant og viktig å foreta, da jeg ønsker at instrumentet skal oppfattes likt i begge undergruppene. Hvis instrumentet konsekvent oppfattes annerledes for høytpresterende elever enn for lavtpresterende elever, vil resultatene trolig ikke kunne sammenlignes (Linacre, 2012c). X-aksen uttrykker hvordan itemene rapporteres av lavtpresterende elever og y-aksen uttrykker hvordan itemene rapporteres av høytpresterende elever. Aksene kan dermed sies å representere de to elevgruppenes oppfattelse av itemenes vanskelighetsgrad, i form av hvor lett eller vanskelig det er å kjenne seg igjen i itemene. Den stiplede linjen i diagrammet blir gjerne kalt *«the line of commonality»* (en linje med fellestrekk), fordi den i like stor grad forutser x-verdier basert på y-verdier som omvendt (Linacre, 2012c). De to heltrukne linjene utenfor den stiplede linjen viser 95% konfidensintervall. Videre ser vi itemenes plassering i diagrammet, og det ønskelige er at alle itemene plasserer seg innenfor konfidensintervallet. Dersom de to undergruppene opplever vanskelighetsgraden på et item som svært ulikt, vil dette synliggjøres ved at itemet ligger utenfor konfidensintervallet.



***Figur 3**. Eksempel på en DTF-analyse for et instrument med 6 items.*

I de kommende delkapitlene vil jeg systematisk analysere hvert delinstrument etter punktene beskrevet ovenfor.

4.1.1 ME-instrument

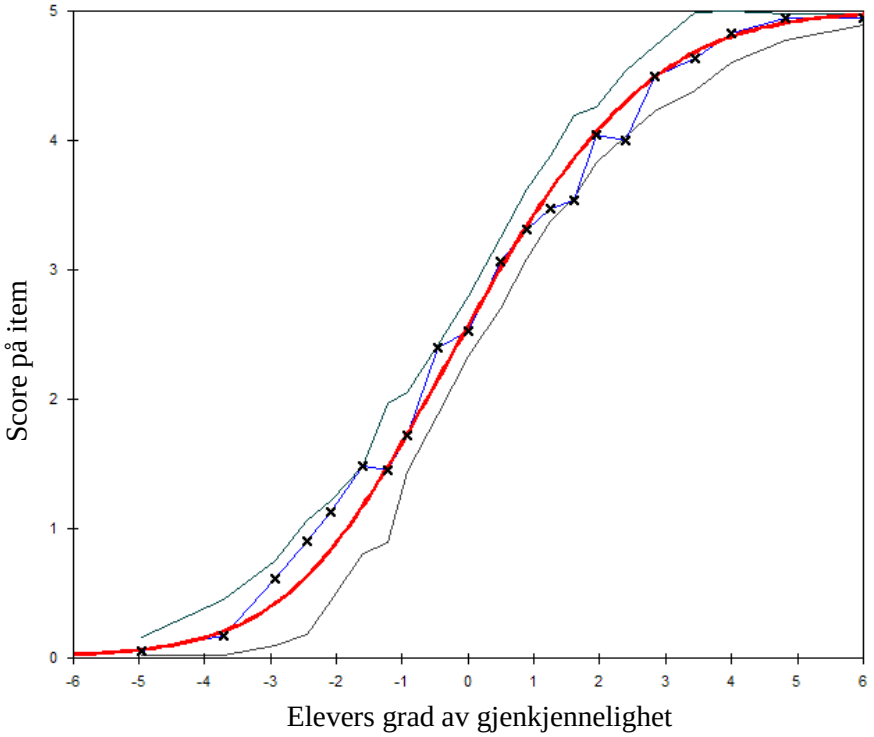
Etter å ha utført første analyse av ME-instrumentet, kunne jeg se at særlig item 3 skilte seg ut med en høy Outfit MNSQ-verdi [MNSQ=1.88, ZSTD=9.9]. Dette indikerer underfit og betyr at variasjonen på dette itemet var større enn forventet (Bond & Fox, 2015). Item 3 er reversert i analysene, fordi ordlyden i itemet går motsatt vei av de resterende itemene, se tabell 1. For item 3 vil altså høyere score indikere negativ påvirkning på mestringsforventning, mens en høy score for de resterende itemene i delinstrumentet vil bety positiv påvirkning på mestringsforventningen. Det er tenkelig at det plutselig reverserte itemet kan ha skapt forvirring hos elevene. Av den grunn har jeg foretatt en datarensking. Renskingen ble gjort ved å ta ut en rapport på de mest uventede responsene, der svært mange av disse omhandlet nettopp item 3. Jeg gikk deretter inn i datamaterialet og endret de mest uventede svarene på item 3 til koden 8 for «ubesvart». Man kan spørre seg om det er uetisk å endre elevenes svar på denne måten, men her støtter jeg meg til Linacre (2012a) som tydelig argumenterer for at dette er riktig prosedyre. Han påpeker at slike endringer gjøres for å kunne måle elever på en meningsfylt måte, og ikke på grunn av et ønske om å gi elever en høyere score. Etter å ha foretatt denne renskingen ser tilpasningsstatistikken bedre ut, se tabell 1.

Tabell 1. *Resultater for Rasch-analyse av ME-instrumentet.*

Resultater for ME-instrumentet ***N=583 | Person rel.: 0.90 | Item rel.: 0.99***				
Påstand	**Measure**	**Outfit MNSQ**	**Observert korrelasjon**	**Forventet korrelasjon**
6. Jeg får til å løse selv de vanskeligste oppgavene i matte.	1.09	1.00	0.86	0.86
2. Jeg har alltid lykkes i matte.	0.73	0.95	0.87	0.86
5. Jeg er flink til å løse ulike matteoppgaver.	-0.06	0.84	0.87	0.85
1. Jeg får gode karakterer på matteprøver.	-0.15	0.65	0.89	0.85
4. Jeg fikk god karakter i matte til sommeren i 8.klasse.	-0.24	1.10	0.85	0.85
3. Selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte.	-1.37	1.29	0.78	0.82

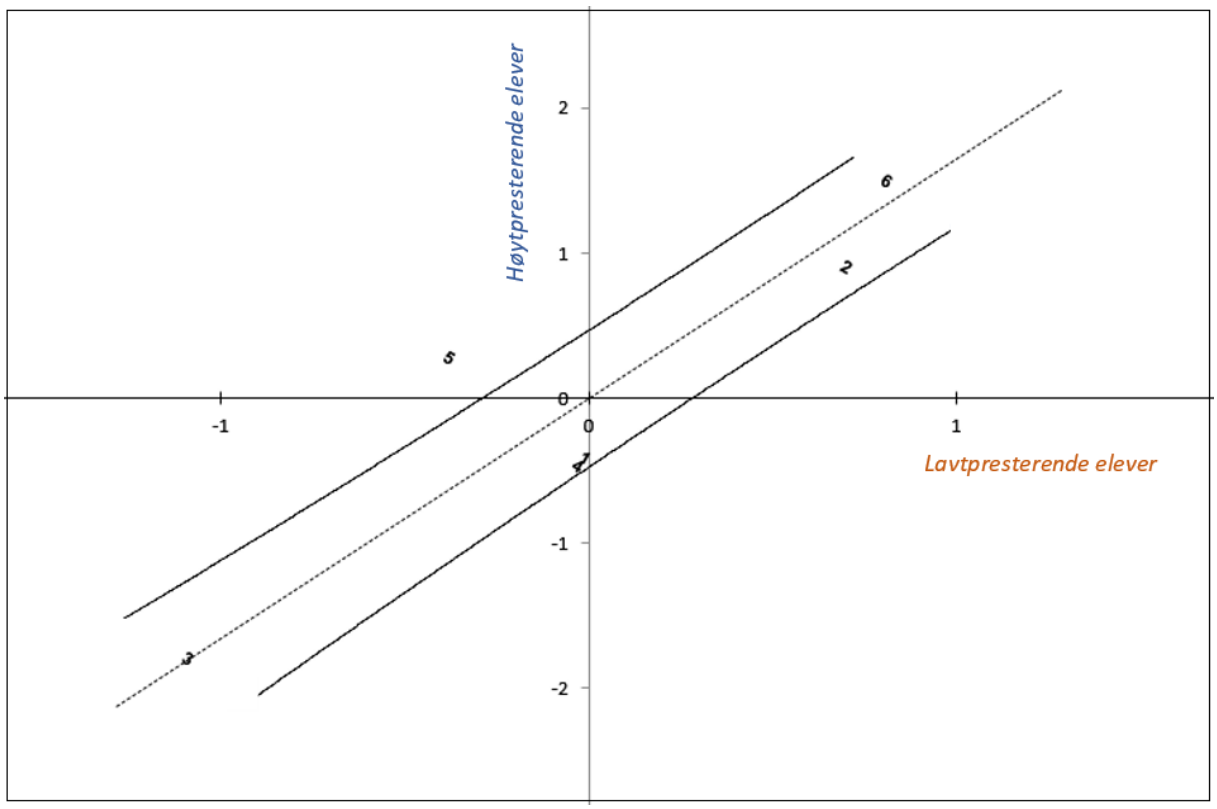
Rasch-analysen av itemene i ME-instrumentet er representert i tabell 1. Reliabiliteten kan sies å være god i henhold til Linacres (2012b) anbefaling (> 0.8) med *person reliability* på 0.90 og *item reliability* på 0.99. Outfit MNSQ-verdiene for alle items i delinstrumentet ligger mellom 0.65 og 1.29, som er innenfor anbefalt intervall (0.6 < outfit MNSQ > 1.4) og er tilfredsstillende (Bond & Fox, 2015). I tabellen er itemene rangert etter *measure*, som uttrykker hvordan itemene måles av instrumentet. Vi kan se at itemene er målt fra -1.37 til 1.09 og ser generelt ut til å ha

god spredning. Dette er ønskelig, da itemene vil dekke flere av respondentene. Likevel vil jeg kommentere at noen sprang er større enn andre. Vi ser for eksempel at mellom item 3 og 4 er det et stort sprang, og medfører at elever med en gjenkjennelighetsgrad mellom -0.24 og -1.37 ikke blir målt så nøyaktig av instrumentet. Item 1, 4 og 5 ser imidlertid ut til å ligge forholdsvis tett, som medfører at det er flere items som måler omtrent de samme elevene, og dermed overlapper hverandre litt. Fordi en jevnere spredning av itemene ville ført til en mer nøyaktig måling av elevene, er det nyttig å være bevisst på dette i den videre tolkningen av resultatene. Kolonnen «observert korrelasjon» viser *item polarity*, og vi ser at analysen for ME-instrumentet kun viser høye positive korrelasjoner mellom itemenes vanskelighetsgrad og elevenes grad av gjenkjennelighet, noe som er ønskelig. Jeg har også sjekket at svarkategoriene forekommer i riktig rekkefølge. Disse analysene ligger i vedlegg 6. Begge de to sistnevnte rapporteringene er altså tegn på at elevenes grad av gjenkjennelighet står i forhold til hvor godt påstandene generelt stemmer for elevene i delinstrumentet.



Figur 4. *ICC for item 2.*

Av figur 4, som viser ICC for item 2, kan vi se at den blå punktlinjen ligger tett opptil den røde grafen, som betyr at den observerte interaksjonen mellom items og elever ligger nær Rasch-modellens forventning. Observert interaksjon er også innenfor 95 % konfidensintervall. ICC for alle items i ME-instrumentet ser omtrent ut som den er vist i figur 4. Det er altså ikke grunn til å tro at det eksisterer tilleggsdimensjoner.



***Figur 5**. DTF-analyse av ME-instrumentet.*

Til slutt vil jeg vise til figur 5, der en DTF-analyse er fremstilt i et spredningsdiagram. Vi ser at itemene i ME-instrumentet i all hovedsak plasserer seg innenfor konfidensintervallet. Item 5 er imidlertid et unntak og ligger utenfor konfidensintervallet på venstre side. Det betyr at de lavtpresterende elevene rapporterer item 5 til å stemme dårligere enn forventet, sett i forhold til de høytpresterende elevene.

4.1.2 VE-instrument

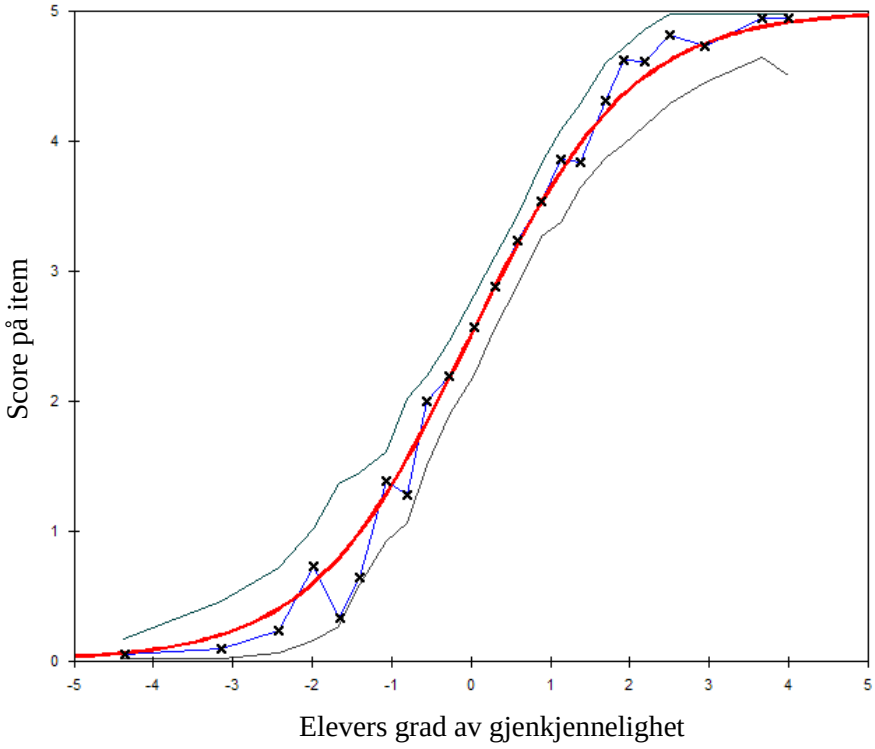
Analysen Rasch-analysen av VE-instrumentet blir representert i tabell 2. Med *person reliability* på 0.83 og *item reliability* på 0.99 konkluderer jeg med god reliabilitet i henhold til Linacres (2012b) anbefaling (>0.8). Outfit MNSQ varierer fra 0.69 til 1.25, og jeg konkluderer med at verdiene er tilfredsstillende (0.6 < outfit MNSQ > 1.4) (Bond & Fox, 2015). Itemenes *measure* varierer fra -0.26 til 0.74. Jeg vil kommentere at det ser ut til at item 11, 10, 8 og 12 ligger ganske tett, og vil dermed være noe overlappende når elevene måles. Dette er likevel ikke et stort problem, da det kun medfører at elever med en gjenkjennelighetsgrad på mellom -0.16 og -0.26 blir er målt mer nøyaktig. Elevene med en gjenkjennelighetsgrad på mellom 0.13 og 0.74 blir ikke målt like nøyaktig. Spriket er imidlertid ikke så stort at det vil hensyntas spesifikt videre. *Item polarity* er også tilfredsstilt, da vi finner høye og positive korrelasjoner nær den

forventede korrelasjonen (Linacre, 2012a). Også for VE-instrumentet forekommer svarkategoriene i riktig rekkefølge (se vedlegg 6).

Tabell 2. *Resultater av Rasch-analyse på VE-instrument.*

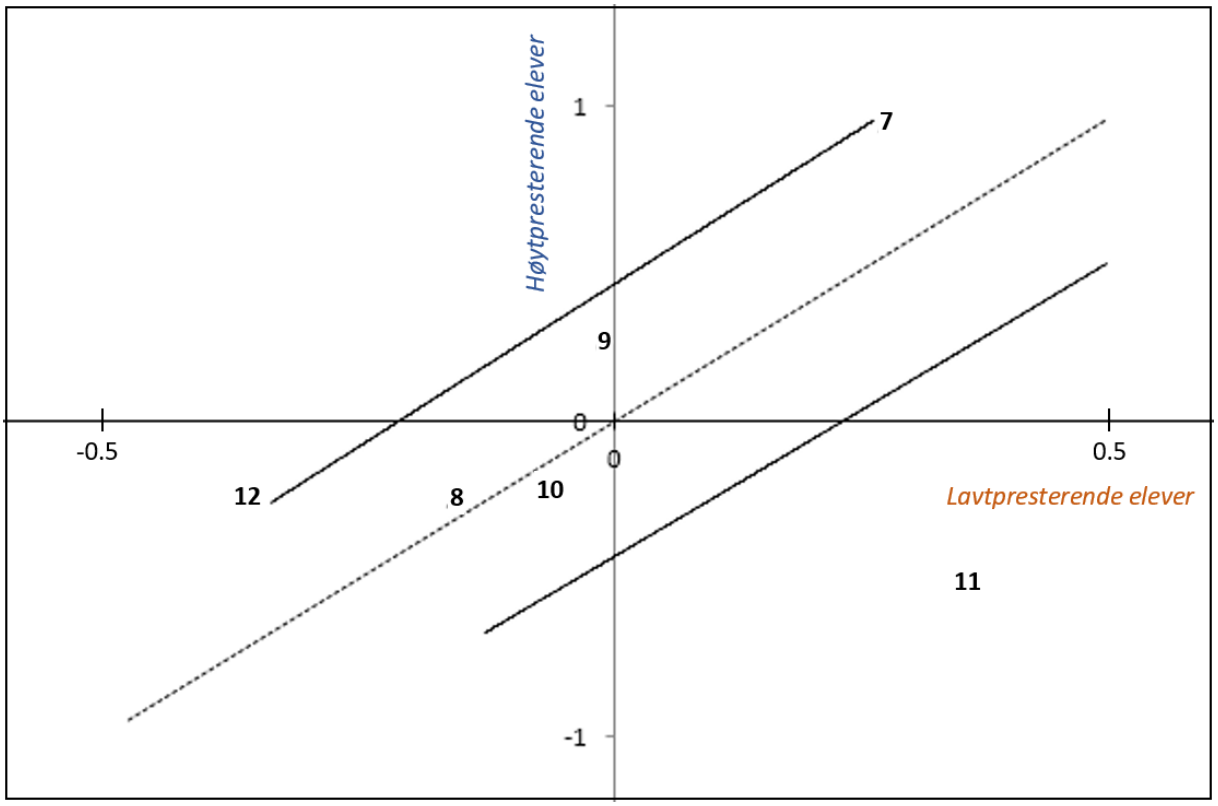
Resultater for VE-instrumentet *N=583 | Person rel.: 0.83 | Item rel.: 0.99*				
Påstand	**Measure**	**Outfit MNSQ**	**Observert korrelasjon**	**Forventet korrelasjon**
7. Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre.	0.74	1.25	0.69	0.73
9. Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre.	0.13	1.12	0.75	0.76
11. Jeg kan se for meg at jeg jobber med vanskelige matteoppgaver og lykkes med å løse dem.	-0.16	0.99	0.77	0.77
10. Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.	-0.21	0.69	0.82	0.77
8. Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.	-0.25	0.78	0.81	0.77
12. Jeg konkurrerer med meg selv i matte.	-0.26	1.22	0.74	0.77

ICC for alle items i delinstrumentet viser seg omtrent lik som ICC for item 8 (vist i figur 6). Grafen viser at den blå linjen som viser den faktiske interaksjonen mellom itemenes vanskelighetsgrad og elevenes gjenkjennelighetsgrad ligger innenfor konfidensintervallet og nær Rasch-modellens forventning. Dette støtter opp om delinstrumentets endimensjonalitet.



Figur 6. *ICC for item 8.*

DTF-analysen er fremstilt i figur 7 og viser at item 11 ligger langt utenfor konfidensintervallet på venstre side. Dette indikerer at lavtpresterende elever rapporterer at item 11 stemmer bedre enn forventet, sammenlignet med høytpresterende elever.



***Figur 7**. DTF-analyse for VE-instrumentet.*

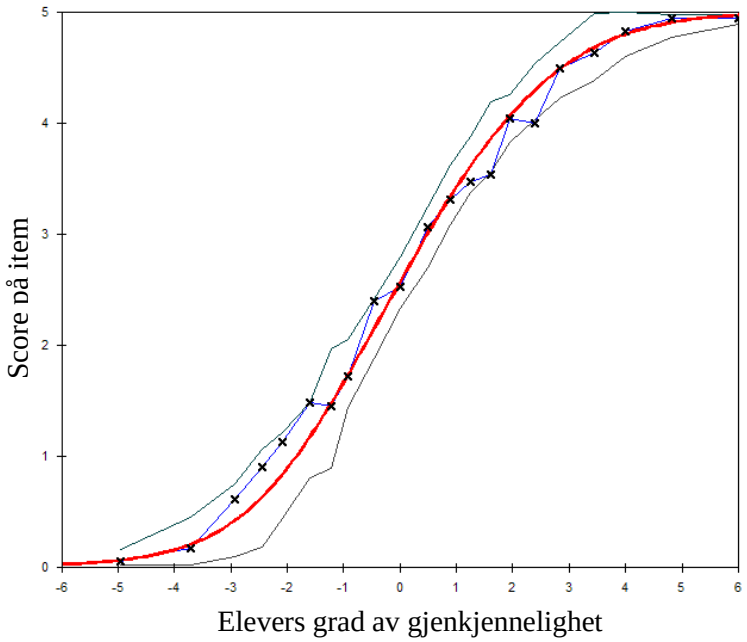
4.1.3 OA-instrument

Rasch-analysen for OA-instrumentet er presentert i tabell 3, og viser god reliabilitet (>0.8), med *person reliability* på 0.89 og *item reliability* på 0.99 (Linacre, 2012b). Outfit MNSQ varierer mellom 0.73 og 1.25 for alle items, og alle items innenfor Bond og Fox' (2015) anbefalinger (0.6 < outfit MNSQ > 1.4). Itemenes *measure* varierer fra -1.18 til 0.91 der alle 6 itemene ser ut til å ha god og jevn spredning. Dette indikerer at elevene blir målt godt og nøyaktig av instrumentet. Observert korrelasjon for alle items er over 0.80 og medfører at *item polarity* er tilfredsstilt (Linacre, 2012b). Svarkategoriene for alle items i instrumentet dukker opp i riktig rekkefølge og tyder på konsistens i forholdet mellom hvor godt påstandene stemmer og elevenes gjenkjennelighetsgrad (se vedlegg 6).

Tabell 3. Resultater fra Rasch-analyse på OA-instrumentet.

Resultater for OA-instrumentet *N=583 | Person rel.: 0.89 | Item rel.: 0.99*				
Påstand	**Measure**	**Outfit MNSQ**	**Observert korrelasjon**	**Forventet korrelasjon**
18. Klassekameratene mine liker å jobbe med meg i matte, fordi de syns jeg er god.	0.91	1.16	0.83	0.84
17. Medelevene mine har fortalt meg at jeg er flink i matte.	0.56	0.84	0.86	0.85
16. Jeg har fått mye skryt for mine evner i matte.	0.28	0.80	0.87	0.85
14. Andre har fortalt meg at jeg er veldig flink i matte.	-0.12	0.73	0.87	0.84
13. Mattelærerne mine har fortalt meg at jeg er flink i matte.	-0.45	1.25	0.81	0.84
15. Voksne i familien min har fortalt meg at jeg er flink i matte.	-1.18	1.04	0.81	0.83

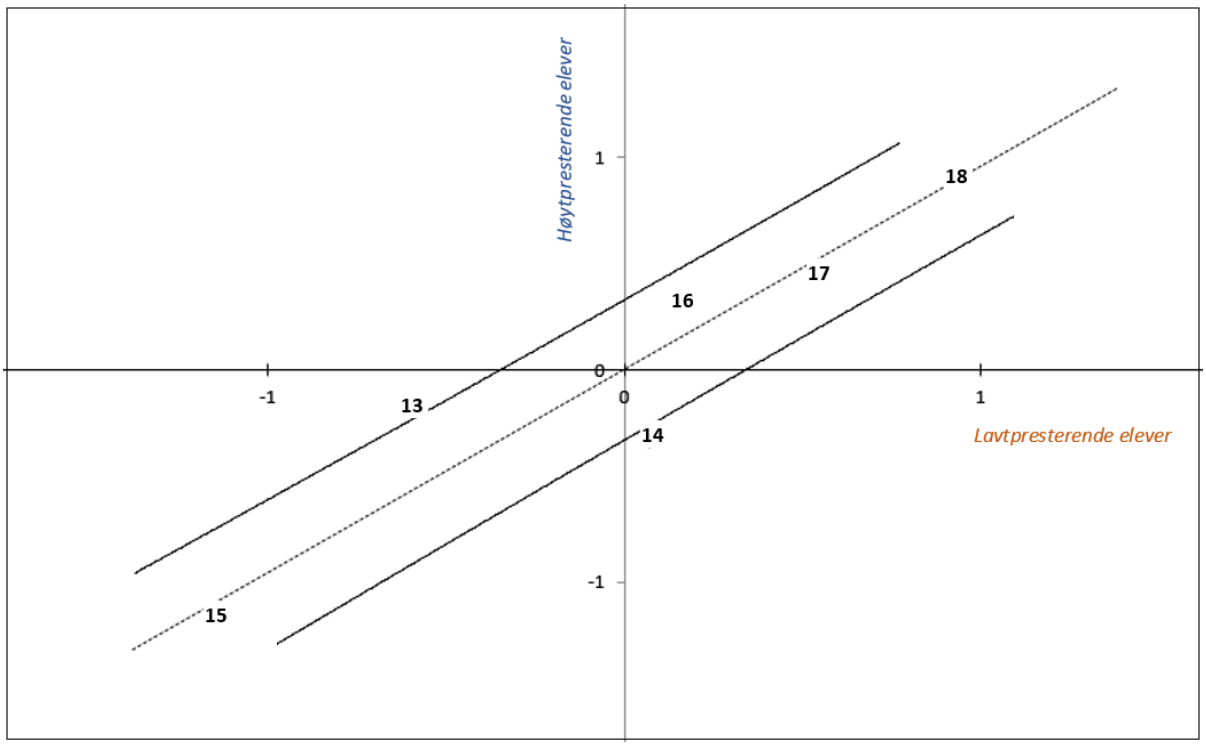
Figur 8 viser ICC for item 13, der grafen for de empiriske dataene (blå punktlinje) ligger tett opptil grafen som viser Rasch-modellens forventning (rød linje). ICC for de resterende itemene i instrumentet ser omtrent lik ut, noe som tyder på et endimensjonalt konstrukt.



Figur 8. *ICC for item 13.*

Av DTF-analysene for dette instrumentet, vist i figur 9, ser vi at de fleste itemene ligger innenfor 95 % konfidensintervall. Item 13 ligger så vidt utenfor på venstre side, og betyr lavtpresterende elever rapporterer lavere grad av gjenkjennelighet i itemet enn forventet, sammenlignet med høytpresterende elever. Tilsvarende finner vi item 14 så vidt utenfor på høyre side, som betyr at lavtpresterende elever rapporterer dette itemet til å stemme bedre enn

forventet, sammenlignet med høytpresterende elever. For dette instrumentet er marginene så små at jeg velger å overse det.



***Figur 9**. DTF-analyse for OA-instrumentet.*

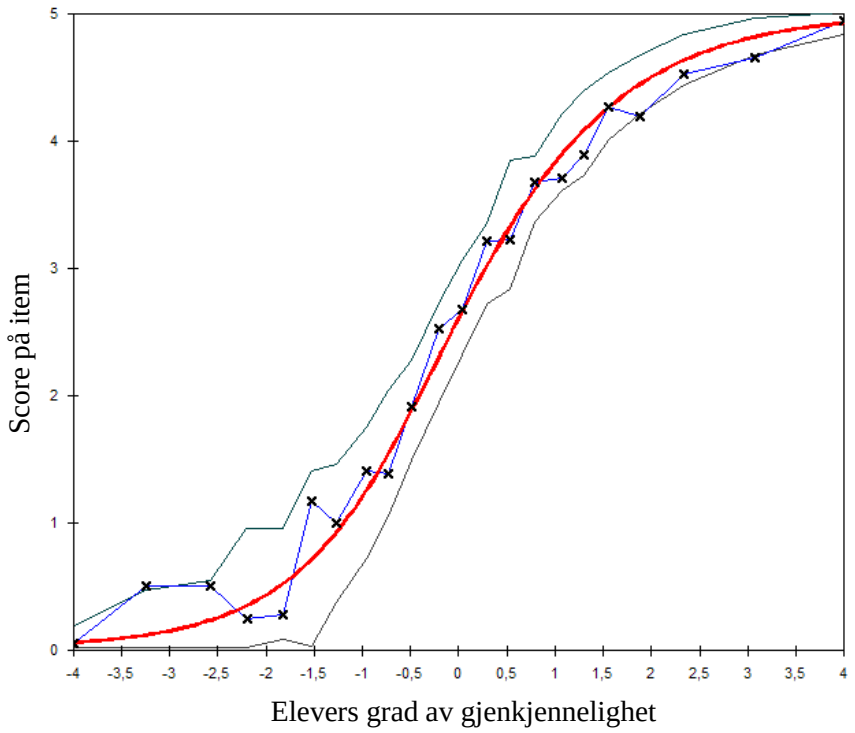
4.1.4 PTF-instrument

PFT-instrumentets Rasch-analyse presenteres i tabell 4. *Item reliability* er på 0.99 og er tilfredsstillende. *Person reliability* for dette instrumentet er 0.80 og er akkurat på grensen til hva Linacre (2012b) anbefaler (>0.80). Dette kan ha sin forklaring i at alle disse itemene er reverserte, som medfører at jo høyere opp på Likert-skalaen eleven svarer, jo sterkere påvirkes elevens mestringsforventning i negativ retning. Itemene i de andre delinstrumentene (med unntak av item 3) antas derimot å påvirke mestringsforventningen i positiv retning. Reverserte items kan skape forvirring hos respondentene, fordi de plutselig må tenke motsatt vei, og dermed påvirke svarmønsteret til respondentene noe (Usher & Pajares, 2009). Reliabiliteten er likevel så høy at det ikke er problematisk (Linacre, 2012b), og jeg lar derfor dette kun bli en bemerkning. Outfit MNSQ varierer fra 0.62 til 1.23, og avdekker dermed ikke så stort *over-* eller *underfit* på noen items at det er problematisk (Bond & Fox, 2015). Itemenes *measure* ligger på mellom -0.52 og 0.86, med en forholdsvis jevn spredning. *Item polarity* ser også ut til å være tilfredsstillende, med korrelasjoner på mellom 0.73 og 0.80, som også er nær de forventede korrelasjonene. Rekkefølgen for svarkategoriene er også for dette instrumentet korrekt (se vedlegg 6).

Tabell 4. Resultater av Rasch-analyse på PFT-instrument.

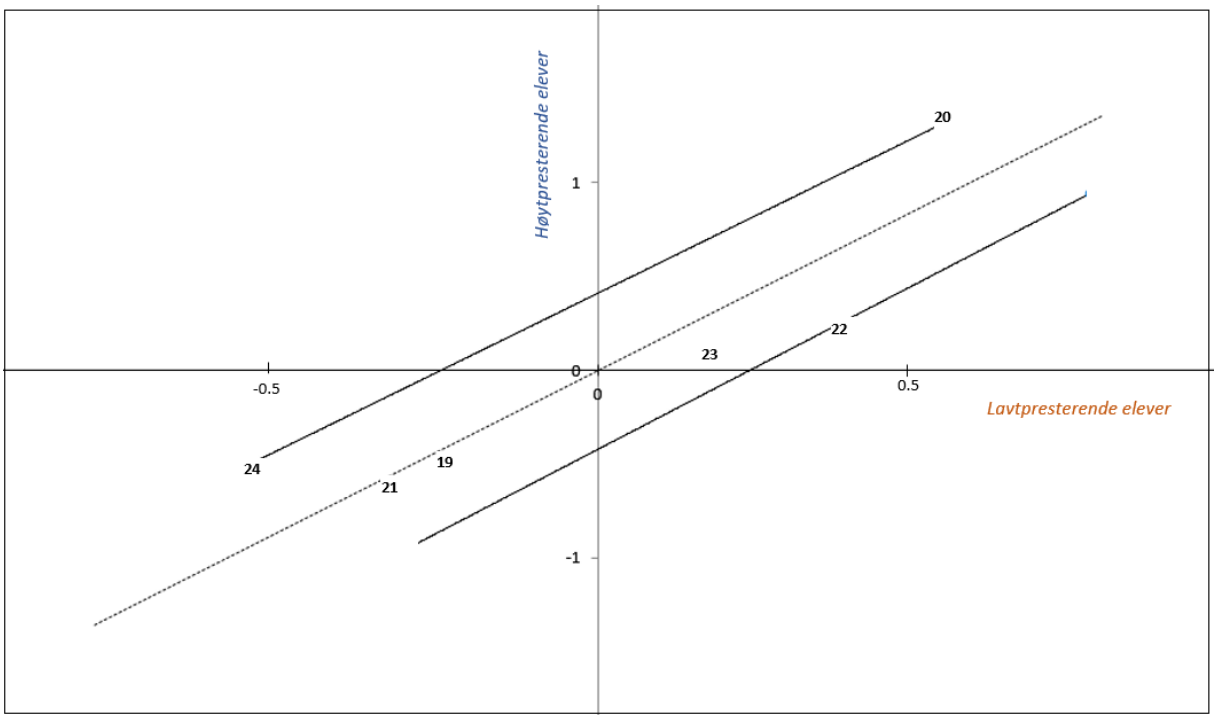
Resultater for PFT-instrumentet *N=583 | Person rel.: 0.80 | Item rel.: 0.99*				
Påstand	**Measure**	**Outfit MNSQ**	**Observert korrelasjon**	**Forventet korrelasjon**
20. Å jobbe med matte tapper meg for energi.	0.86	1.22	0.77	0.80
23. Jeg blir deppa av tanken på å gjøre matte.	0.25	1.06	0.78	0.79
22. Jeg får jernteppe og klarer ikke å tenke klart når jeg gjør matteoppgaver.	0.22	0.97	0.79	0.80
19. Jeg føler meg stresset og nervøs bare ved å være i mattetimene.	-0.38	1.01	0.76	0.76
21. Jeg blir stresset med en gang jeg begynner å jobbe med matte.	-0.42	0.62	0.80	0.76
24. Jeg blir anspent i hele kroppen når jeg må jobbe med matte.	-0.52	1.23	0.73	0.75

ICC for alle items i PFT-instrumentet ligner figur 10, som viser item 22. Det betyr at det ikke er tegn til tilleggsdimensjoner.



Figur 10. ICC for item 22.

Til slutt viser jeg til DTF-analysen for delinstrumentet i figur 11, der de fleste itemene plasserer seg fint innenfor konfidensintervallet. Item 20, 22 og 24 ligger i ytterkanten av konfidensintervallet, men avvikene her er så små at jeg velger å konkludere med at dette ikke gir noen praktiske begrensninger for videre analyser.



***Figur 11**. DTF-analyse for PFT-instrumentet.*

4.1.5 Et valid instrument

Etter å ha utført ovenstående analyser for alle delinstrumentene vil jeg konkludere med at både reliabiliteten og validiteten i studien er god. Oversettelsen av det allerede validerte instrumentet til Usher og Pajares (2009) ser altså ut til å fungere i en norsk kontekst blant et utvalg av norske 9.klassinger. DTF-analysene, som er utført med den hensikt å undersøke instrumentets oppfattelse blant de to undergruppene jeg vil benytte meg av videre i analysene, viser imidlertid at det er noen items som oppfattes ulikt. Dette er i hovedsak item 5 (vist i figur 5) og item 11 (vist i figur 7). Førstnevnte påstand rapporteres av lavtpresterende elever til å stemme dårligere enn forventet, sammenlignet med høytpresterende elever. Sistnevnte rapporteres av lavtpresterende elever til å stemme bedre enn forventet, sammenlignet med høytpresterende elever. Jeg kunne valgt å fjerne de problematiske itemene fra analysen og kjørt nye analyser, men på bakgrunn av at det originale instrumentet allerede er validert (Usher & Pajares, 2009) velger jeg å ikke la dette veie for tungt, annet enn at jeg hensyntar de spesifikke itemene i videre analyser. Fordi jeg nå vet at item 5 og item 11 oppfattes ulikt av de to undergruppene, velger jeg å være forsiktig med å trekke konklusjoner basert på disse itemene når jeg i diskusjonskapittelet ser nærmere på hvert enkelt item.

4.2 Resultater av studien

Nå som jeg vet at instrumentet for å måle elevenes kilder til mestringsforventning er valid og at målingene gjort med instrumentet er reliable, vil jeg se nærmere på resultatene av undersøkelsen. Studiens andre forskningsspørsmål vil i dette delkapittelet vies oppmerksomhet. I forskningsspørsmålet etterspørres det om det er noen forskjeller i hvordan høyt- og lavtpresterende elever rapporterer kildene til mestringsforventning, og om hvilke mønstre som eventuelt viser seg.

Når jeg fremlegger resultatene av undersøkelsen og rapporterer ulike gjennomsnitt i form av logits, er det viktig med en entydig forståelse av hva tallene er et mål på. Påstandene i det benyttede instrumentet vurderes av elevene på en Likert-skala med 6 svarkategorier. Ved bruk av skalaen rapporterer elevene i hvor stor grad påstandene stemmer for dem, og sier dermed noe om hvor stor grad av gjenkjennelighet elevene opplever i påstandene. En elev som eksempelvis rapporterer høy grad av gjenkjennelighet i påstandene i ME-instrumentet, vil antakelig påvirkes mer av denne kilden, enn en elev som rapporterer lav grad av gjenkjennelighet i påstandene.

4.2.1 Sammenligning av høyt- og lavtpresterende elever innenfor hver enkelt kilde

Siden dette instrumentet er delt inn i fire delinstrumenter under analysene, og hvert delinstrument er analysert for seg, kan jeg ikke sammenligne gjennomsnittsmålene til de ulike kildene i tabell 5 med hverandre. Dette fordi det i analysene dannes egen logit-skala basert på de importerte dataene for hvert delinstrument, og vi kan si at hvert delinstrument måles med sin egen «linjal». Dette medfører at en logit i analysen for ME-instrumentet ikke nødvendigvis tilsvarer en logit i VE-instrumentet og så videre. Tabell 5 viser gjennomsnittet for høyt- og lavtpresterende elever i hver av de fire kildene. Tabellen kan kun benyttes til å sammenligne vannrett, altså sammenligne høyt- og lavtpresterende elevers gjennomsnittlige score innenfor hver enkelt kilde.

Tabell 5. *Gjennomsnittet (logits) for hver av de to elevgruppene for hver kilde.*

Kilde	Gjennomsnitt (logits) for **høytpresterende elever** N_H=110	Gjennomsnitt (logits) for **lavtpresterende elever** N_L=113
ME	3.033	-1.403
VE	0.851	-0.897
OA	1.951	-1.146
PFT	1.966	-0.080

For alle kildene ser det ut til å være en forskjell i gjennomsnittlig score mellom de to elevgruppene basert på det utvalget elever jeg har i denne studien. Høytpresterende elever i mitt utvalg har høyere gjennomsnittlig score i alle kildene enn lavtpresterende elever. For å avgjøre om det er tilfeldig eller ikke at jeg i dette utvalget har fått slike forskjeller, velger jeg å gjøre tester for å sjekke om forskjellene er signifikante. Som forklart i kapittel 3.3.6 vil en høyere score bety at eleven rapporterer en høyere grad av gjenkjennelighet i items som bidrar til høyere mestringsforventning. Samtidig viser teori og tidligere forskning at elever som har høyere mestringsforventning presterer bedre (Bandura, 1977; Galyon et al., 2012; Salazar & Hayward, 2018; Usher & Pajares, 2008). Dette gir grunn til å forvente en bestemt retning på forskjellene i elevenes rapportering av kilder, og jeg ønsker derfor å teste en ensidig hypotese om at høytpresterende elever har en gjennomsnittlig høyere score enn lavtpresterende elever. Signifikansnivået er satt til $\alpha=0.05$. Nedenfor vil jeg først teste om forskjellen i kilden ME er signifikant, og deretter forklare hvordan jeg har utført disse testene i SPSS. Videre vil jeg kun rapportere resultatene fra de andre kildene, som utføres på tilsvarende måte.

Basert på forskjellene i gjennomsnittelig score i utvalget mitt (tabell 5) kjørte jeg en uavhengig t-test på ME-instrumentet for å se om høytpresterende elever har en signifikant høyere score enn lavtpresterende elever. Resultatet viser $p<\alpha$ og medfører at det er en signifikant forskjell mellom høyt- og lavtpresterende elevers gjennomsnittelige score [$T_{221}=-18{,}26$, $p<0.001$]. Cohen's d viser at sammenhengen er sterk (d=2,45), og betyr at forskjellen i kilden ME i stor grad kan forklares av prestasjonsnivået. Med andre ord kan jeg nå konkludere med høytpresterende elevers grad av gjennkjennelighet i påstandene i kilden ME er signifikant høyere enn de lavtpresterende elevenes grad av gjenkjennelighet.

På samme måte som for ME-instrumentet har jeg utført tester for å sjekke signifikans også for forskjellene vi ser i de andre kildene. Resultatene viser at høytpresterende elever rapporterer signifikant ($p<\alpha$) større grad av gjenkjennelighet enn lavtpresterende elever både i VE-instrumentet [$t_{221}=-9{,}69$, $p<0.001$, $d=1.30$], OA-instrumentet [$T_{221} = -11.20$, $p<0.001$, $d=1.50$] og for PFT-instrumentet [$T_{221}=-9.334$, $p<0.001$, $d=1.25$]. For alle kildene viser *Cohen's d* er sterk sammenheng mellom elevenes gjenkjennelighetsgrad i påstandene og deres prestasjonsnivå.

4.2.2 Sammenligning av de ulike kildene innenfor hver av elevgruppene

For å kunne sammenligne de ulike kildene opp mot hverandre innenfor hver undergruppe, har jeg foretatt en z-standardisering av hver enkelt persons gjenkjennelighetsgrad. Denne standardiseringen setter gjennomsnittet lik 0 og standardavviket lik 1, og er nærmere forklart

kapittel 3.6.4 (Nardi, 2018). Tabell 6 viser høyt- og lavtpresterende elevers gjennomsnittlige z-score for hver av kildene.

Tabell 6. *Gjennomsnittlig z-scores for høyt- og lavtpresterende elever for alle kildene*

Kilde	Gjennomsnittlig z-scores for **høytpresterende elever** N_H=110	Gjennomsnittlig z-scores for **lavtpresterende elever** N_L=113
ME	0.964	-0.869
VE	0.632	-0.518
OA	0.767	-0.560
PFT	0.599	-0.584

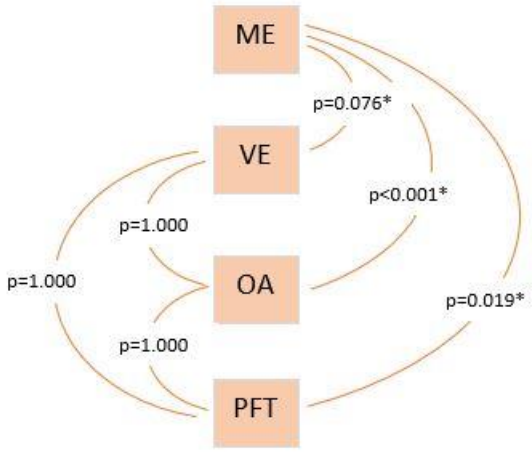
Ut ifra tabell 6 ser det ut til at høyt- og lavtpresterende elever rapporterer ulik grad av gjenkjennelighet i de ulike kildene. For begge elevgrupper ser det ut til at gjenkjennelighet i kilden ME skiller seg mest ut i forhold til de andre kildene. Høytpresterende elever ser ut til å kjenne seg mest igjen i denne kilden, deretter OA, VE og PFT. Lavtpresterende elever ser ut til å kjenne seg mindre igjen i ME enn de andre kildene. Denne elevgruppen rapporterer høyest grad av gjenkjennelighet i VE, deretter OA, PFT og ME.

One-way repeated measure ANOVA ble benyttet for å undersøke hypotesen om at gjennomsnittlig z-score i de ulike kildene er forskjellig. Analysen ble foretatt både i gruppen for høytpresterende elever og gruppen for lavtpresterende elever. SPSS ble benyttet som analyseverktøy, og signifikansnivået er α=0.05.

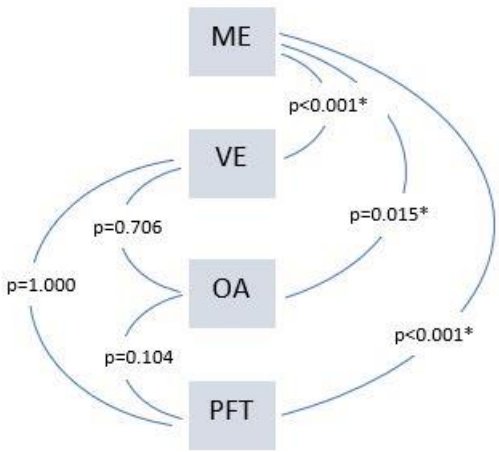
Som nevnt i kapittel 3.6.3 er ett av kravene for å kunne ta i bruk denne type analyse er at sfærisitet er tilfredsstilt. Mauchly's test of sphericity viser $p < .001$ og medfører at nullhypotesen om at variansen er lik kan forkastes. Dette medfører at kravet om sfærisitet ikke er tilfredsstilt. Derfor benytter jeg meg av en Greenhouse-Geisser-korrigering av frihetsgrader.

Resultatene for høytpresterende elever viste $p<\alpha$ og støtter dermed hypotesen om en signifikant forskjell i gjennomsnittlig z-scores for de fire kildene [$F(2.54,276.48)=7.27$, $p<0.001$]. Sammenhengen var moderat [$\eta^2=0.06$] (van den Berg, u.å.). Bonferroni Post hoc-testen viste at de høytpresterende elevene rapporterte om en signifikant **større** grad av gjenkjennelse i kilden ME, sammenlignet med hver av de andre kildene [ME-VE: $p<0.001$, ME-OA: $p=0.02$, ME-PFT: $p<0.001$]. Forskjellene i gjennomsnittlig z-score mellom de andre kildene er imidlertid ikke signifikante [VE-OA: $p=0.71$ VE-PFT: $p=1.00$, OA-PFT: $p=0.10$], se figur 13.

Resultatene for lavtpresterende elever viste p<α og dermed en signifikant forskjell i gjennomsnittlig z-scores for de fire kildene [F (2.40, 268.27) = 3.61, p< .001]. Sammenhengen ser ut til å være moderat [η^2 = .05] (van den Berg, u.å.). Bonferroni Post hoc-testen viste, på samme måte som for de høytpresterende elevene, signifikante forskjeller mellom kilden ME og alle de andre kildene [ME-VE: p=0.08, ME-OA: p<0.001, ME-PFT: p=0.02]. De lavtpresterende elevene rapporterte imidlertid signifikant **mindre** grad av gjenkjennelse i denne kilden sammenlignet med de andre kildene. Forskjellene mellom de andre kildene er imidlertid ikke signifikante [VE-OA: p=1.00, VE-PFT=1.00, OA-PFT=1.00], se figur 12.



Figur 12. *Signifikansverdier for kildene blant lavtpresterende elever*



Figur 13. *Signifikansverdier for kildene blant høytpresterende elever*

Signifikanstestene utført ovenfor viste at elevenes gjennomsnittlige z-score for kilden ME er signifikant forskjellig fra elevenes gjennomsnittlige z-score for alle de andre kildene. Dette gjelder både for høytpresterende elever og for lavtpresterende elever. I begge elevgruppene er elevenes gjennomsnittlige z-score for kildene VE, OA og PFT imidlertid ikke signifikant forskjellig fra hverandre. Differansen mellom høyt- og lavtpresterende elevers gjennomsnittlige z-score innenfor hver kilde er visualisert i figur 14. Disse forskjellene vil være signifikante som følge av signifikanstesten utført i kapittel 4.2.1 (de signifikante forskjellene er bevart i z-standardiseringen).

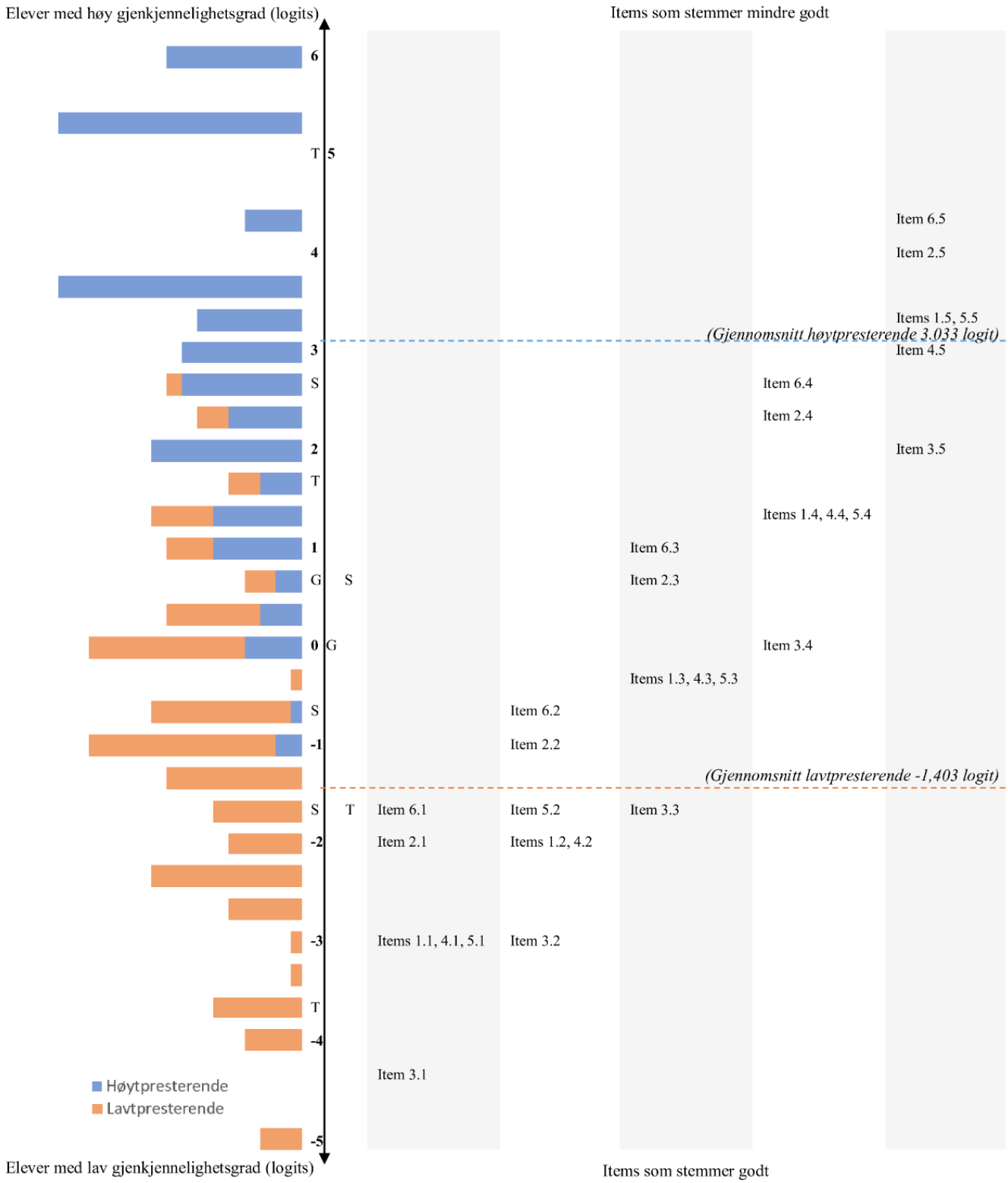
Figur 14. *En visualisering av høyt- og lavtpresterende elevers gjennomsnittlige z-score, samt differansen i z-score for hver kilde.*

4.3.3 Terskelverdier for hvert item

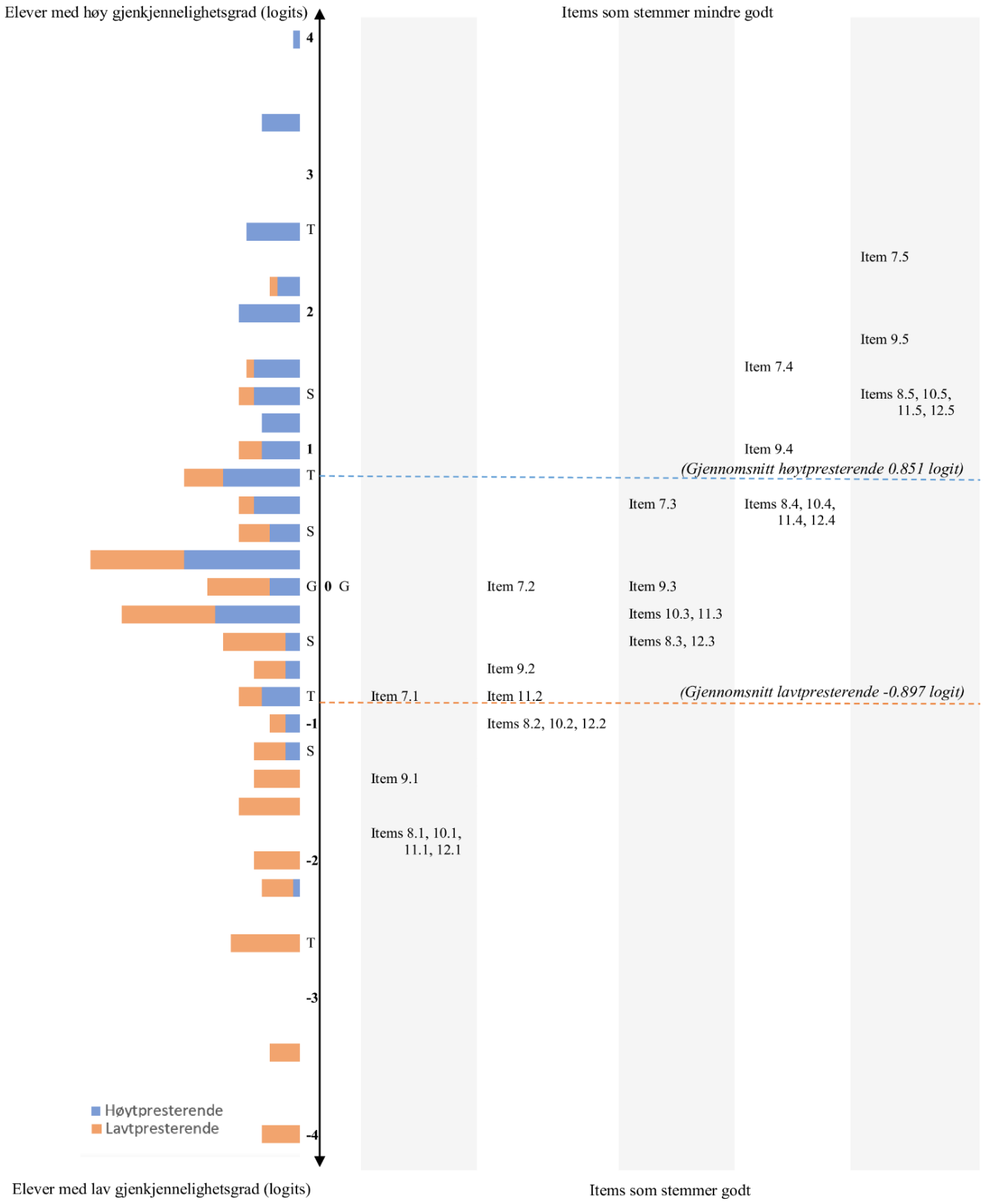
I dette delkapittelet ønsker jeg fremlegge og forklare modeller som representerer elevenes terskelverdier for hvert enkelt item, som er forklart i kapittel 3.3.5. Disse modellene er utarbeidet for hver av de fire kildene, og skal både gi et bilde av hvilke svarkategorier på Likert-skalaen som er benyttet for hvert item, og gi informasjon om hvor vi finner de høyt- og lavtpresterende elevene blant hele utvalget av elever. Modellene vil bli benyttet for videre tolkning i diskusjonskapittelet (kapittel 5.0) og vil derfor kun bli forklart her.

Venstre side av modellen synliggjør høyt- og lavtpresterende elever på en logit-skala over elevenes gjenkjennelsesgrad. Gjenkjennelighetsgraden er forklart nærmere i kapittel 3.3.6 og er definert ut ifra hvilke svarkategorier elevene har brukt til å angi svarene sine. Elever med en høy gjenkjennelighetsgrad rapporterer i større grad om tilstedeværelse av situasjoner som antas å påvirke mestringsforventningen i positiv retning, og tilsvarende fravær av situasjoner som påvirker mestringsforventningen i negativ retning, enn det elever med en lav gjenkjennelighetsgrad rapporterer om. De oransje vannrette søylene viser antall lavtpresterende elever ved logit-skalaens ulike nivåer av gjenkjennelighet og består totalt av N_L=113 elever. De blå vannrette søylene viser tilsvarende for høytpresterende elever, og utgjør totalt N_H=110 elever. De resterende elevene er fjernet fra den visuelle søylefremstillingen for å lettere kunne identifisere de to undergruppene jeg setter søkelys på i denne studien. De stiplede linjene, oransje og blå, viser henholdsvis lavtpresterende og høytpresterende elevers gjennomsnittlige gjenkjennelsesgrad. På venstre side av logit-skalaen er det markert gjennomsnitt (G), standardavvik (S) og to standardavvik (T) for elevenes gjenkjennelsesgrad for hele utvalget på N=583 elever. Tilsvarende informasjon finner også på høyre side av logit-skalaen, som omhandler itemenes vanskelighetsgrad.

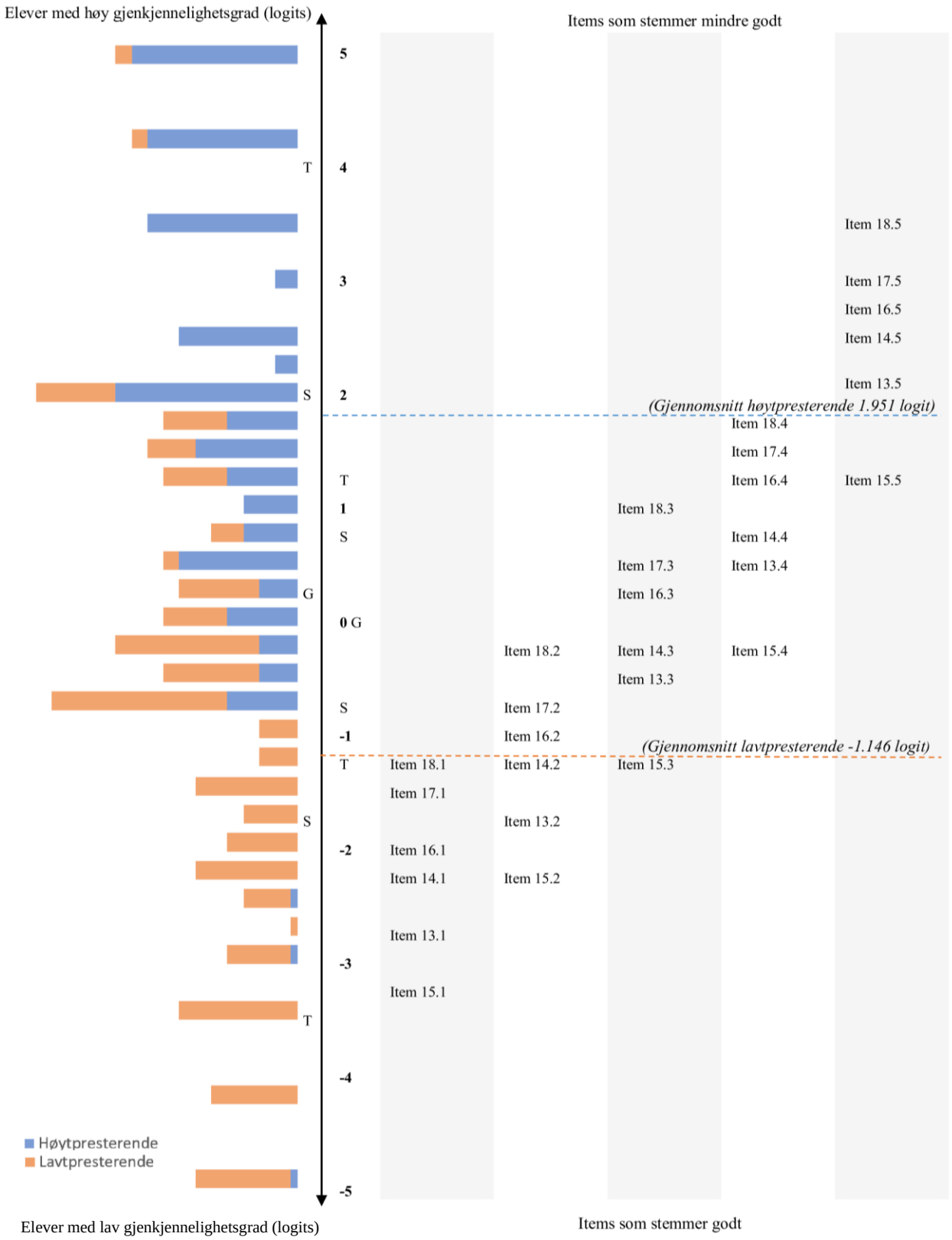
Høyre side av modellen viser itemenes terskelverdier (Andrich Treshold) for alle items i delinstrumentet fordelt på elevenes gjenkjennelighetsgrad. Her vises altså elevers gjenkjennelighetsgrad når sannsynligheten for å svare i én svarkategori er like stor som sannsynligheten for å svare i den overliggende svarkategorien på et item. De fem kolonnene er terskelverdier, der items i terskelverdi 1 (første kolonne) viser gjenkjennelighetsgraden til elever som har 50% sannsynlighet for å svare i svarkategori 1 og 50% sannsynlighet for å svare i svarkategori 2. Tilsvarende for de andre terskelverdiene. Items i terskelverdi 5 (femte kolonne) vil dermed markere elevenes gjenkjennelighetsgrad når det er like stor sannsynlighet for å svare i svarkategori 5 som 6. Items som er plassert lenger opp i kolonnene, viser at elever i større grad har rapportert itemene til å stemme dårligere og vi kan si at det generelt er vanskeligere å kjenne seg igjen i itemet. For disse itemene kreves det en høyere gjenkjennelighetsgrad for å velge å svare i høye svarkategorier. Item som er plassert lenger ned, er items som elever i større grad har rapportert til å stemme godt. Det skal altså mindre til for å svare i høyere svarkategorier på disse itemene, og det er lettere å kjenne seg igjen i dem. For de reverserte itemene vil dette bli motsatt. Dette vil selvsagt bli hensyntatt i videre tolkning av modellene.



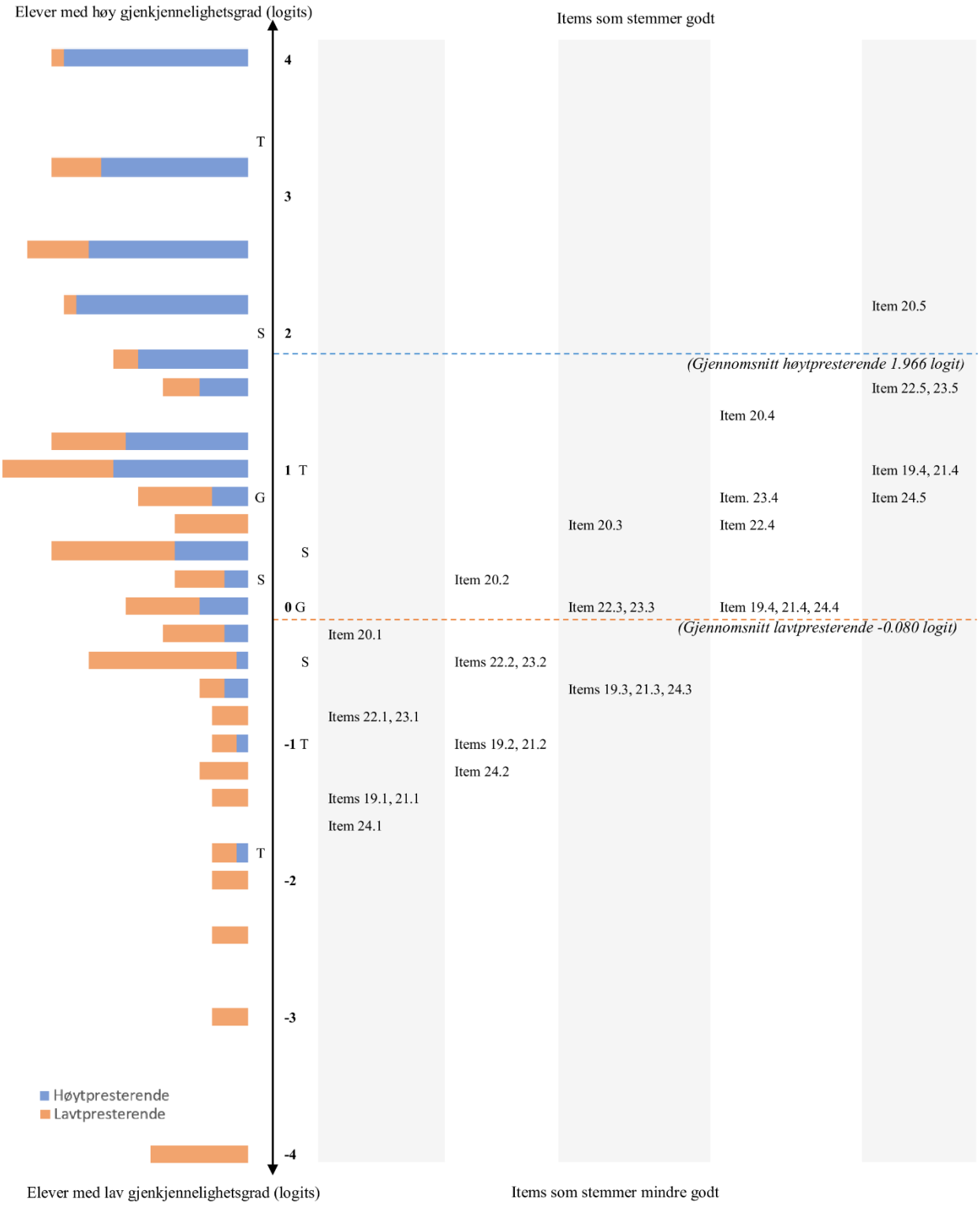
Figur 15. *Wright Map for ME-instrumentet (N=583, N_H=110, N_L=113)*



Figur 16. *Wright Map for VE-instrumentet (N=583, N_H=110, N_L=113)*



***Figur 18**. Wright Map for OA-instrumentet (N=583, N_H=110, N_L=113)*



Figur 19. *Wright Map for PFT-instrumentet (N=583, N_H=110, N_L=113)*

5.0 Diskusjon

I denne delen av oppgaven vil jeg diskutere resultatene presentert i forrige kapittel opp mot teori og tidligere forskning som ble presentert i kapittel 2.0. Jeg vil forsøke å besvare studiens forskningsspørsmål, for på den måten å kunne belyse den overordnede problemstillingen best mulig. Først vil jeg, på bakgrunn av kapittel 3.0 og 4.0, diskutere det første forskningsspørsmålet, som dreier seg om instrumentet som er benyttet i studien. Dette vies oppmerksomhet i kapittel 5.1. Deretter vil jeg i kapittel 5.2 diskutere det andre forskningsspørsmålet som dreier seg om forskjeller i høyt- og lavtpresterende elevers i kilder til mestringsforventning. Til slutt vil jeg i kapittel 5.3 drøfte hvilke praktiske implikasjoner disse funnene kan ha for læreren i klasserommet. Avslutningsvis i dette kapittelet vil jeg sammenfatte de viktigste poengene fra diskusjonen deretter gå videre til en konklusjon og avsluttende kommentarer i påfølgende kapittel.

5.1 Oversettelse og validering av instrumentet

Det første forskningsspørsmålet om hvordan et instrument for å måle elevers kilder til mestringsforventning i matematikk kan oversettes og valideres er besvart både i metodekapittelet (piloten) og resultatkapittelet (hovedstudien), der jeg samlet viser til utviklingen av og validiteten og reliabiliteten til det benyttede instrumentet. Med støtte i Bond og Fox (2015), som hevder at vi må bruke mer tid på å undersøke skalaene våre enn å undersøke med skalaene, vil jeg påpeke viktigheten av å gjennomføre en slik grundig valideringsprosess.

Det er mange faktorer som har vært med på å bidra til at det benyttede instrumentet nå kan sies å være valid for mitt utvalg av norske 9.klassinger, slik jeg konkluderte med i kapittel 4.1.5. Instrumentet *Sources of Middel School Self-Efficacy Scale*, utviklet av Usher og Pajares (2009), er allerede validert. Man kan derfor lure på om en ny valideringsrunde grunnet en oversatt versjon ville være nødvendig. Jeg vil i kommende avsnitt argumentere for at dette er nødvendig.

For det første er konstruktet mestringsforventninger situasjons- og kontekstavhengig, og medfører at en elevs mestringsforventninger kan variere fra en kontekst til en annen (Bandura, 1997). Konteksten kan påvirke hvordan mestringsforventninger skapes og oppfattes (Bandura, 1997; Haciomeroglu, 2019). Jeg tolker det slik at også elevers kilder til mestringsforventninger kan variere fra situasjon til situasjon. For det andre benyttes instrumentet i min studie i et utvalg som både er geografisk, etnisk og sosiokulturelt forskjellig fra utvalget som Usher og Pajares (2009) validerte instrumentet for. Elevene i denne studiens utvalg er dessuten trolig noe eldre

enn elevene i utvalget instrumentet ble validert for (Usher & Pajares, 2009). Usher og Pajares (2009) oppfordrer eksplisitt til å benytte og validere bruken av disse itemene i andre kontekster og emner. I denne studien benyttes itemene fremdeles innen en matematikkfaglig skolekontekst, men det er andre faktorer, nevnt ovenfor, som er forskjellig fra konteksten instrumentet opprinnelig ble validert for. For det tredje påpeker Usher og Pajares (2009) noe som særlig blir viktig når instrumentet også har blitt oversatt. De påpeker nemlig at ordlyden til itemene kan påvirke resultatet. Dette gir meg grunn til å tro at ordvalg og nyanseringer i oversettelsen kan påvirke resultatet, noe som belyser viktigheten av et grundig arbeid med oversettelsen og validering av instrumentet.

Som også tidligere nevnt, vil jeg påstå at samarbeidet i forskningsgruppa SESAM var viktig, særlig for oversettelsen av instrumentet. Med tanke på viktigheten av ordlyd og ordvalg, ser jeg det som svært nyttig å diskutere dette i samråd med andre. Etter at instrumentet var oversatt, ble instrumentet analysert ved hjelp av RSM på bakgrunn av innsamlet pilotdata, i tråd med de retningslinjene flere forskere gir (Bond & Fox, 2015; Boone, 2016; Linacre, 1994). Piloteringen ble utført i et utvalg bestående av 41 respondenter, noe som er innenfor anbefalingen Linacre (1994) gir, der nedre grense for antall respondenter er satt til N=30.

I kapittel 3.3.1 har jeg argumentert for hvorfor RSM er valgt som analyseverktøy i denne studien. Med støtte i Cavanagh og Fischer (2018) vil jeg påstå at Rasch-analyser er et godt verktøy for å administrere og skreddersy måleinstrumenter. Ved å benytte Rasch-analyser har jeg også kunnet konvertere opprinnelig ordinale data til kontinuerlige, noe som bidrar til å berike datamaterialet og øke kvaliteten på målingen. Jeg vil derfor påstå at dette vil bidra til å øke kvaliteten på de slutningene jeg videre vil trekke på bakgrunn av dataene innsamlet ved instrumentet (Usher & Pajares, 2009).

Jeg vil videre benytte meg av dataene innhentet ved instrumentet til å diskutere det andre forskningsspørsmålet mitt. Dette etterspør om det viser seg å være forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning, og hvilke mønstre som eventuelt viser seg. Diskusjonen har sitt faglige grunnlag i litteratur og forskning redegjort for i kapittel 2.0.

5.2 Forskjeller mellom høyt- og lavtpresterende elevers rapportering av kildene

I dette delkapittelet vil jeg se på hvilke forskjeller vi ser mellom høyt- og lavtpresterende elevers rapportering av kildene til mestringsforventning. I videre diskusjon vil jeg ikke gå inn på hver enkelt kilde og diskutere disse kronologisk. Jeg vil imidlertid trekke frem tre hovedforskjeller som jeg mener bør påpekes ut ifra resultatene i denne studien, og argumentere for hvorfor jeg

mener disse bør fremheves. Samtidig vil jeg søke etter forklarende faktorer som kan bidra til å belyse hvorfor disse forskjellene viser seg. Jeg vil også trekke inn og benytte meg av informasjonen innhentet for de ulike kildene der jeg mener det er relevant for diskusjonen.

5.2.1 Signifikant forskjell mellom høyt- og lavtpresterende elever i alle kilder

Først og fremst vil jeg ta tak i et sentralt funn fra analysene. Det er en signifikant forskjell mellom høyt- og lavtpresterende elever i alle kilder. Analysen viser at høytpresterende elever i større grad rapporterer at situasjonene beskrevet i instrumentets påstander stemmer for dem enn det de lavtpresterende elevene rapporterer. Tabell 5 i kapittel 4.2.1 viser gjennomsnittlig score, oppgitt i logits, for høyt- og lavtpresterende elever i alle de fire kildene. Dette er det jeg i denne studien har kalt for elevenes gjenkjennelighetsgrad, og er en score basert på hvilke svarkategorier elevene har svart i for itemene. Tabellen viser signifikante forskjeller mellom de høyt- og lavtpresterende elever i alle kildene. Disse forskjellene blir tydelig synliggjort i Wright Map's-figurene for hver kilde, hvor man kan se at lavtpresterende elever svarer i lavere svarkategorier og høytpresterende elever svarer i høyere svarkategorier, se figur 15, 16, 17 og 18.

Som tidligere nevnt er det mange forskere som har funnet at elevers mestringsforventning har en predikativ effekt på deres prestasjon (Bandura, 1977; Galyon et al., 2012; Salazar & Hayward, 2018; Usher & Pajares, 2008). Elever med høyere mestringsforventning presterer altså bedre. Forskning peker også på at de fire kildene har en predikativ effekt på elevers mestringsforventning (Chen, 2003; Phan, 2012; Usher, 2009; Yurt, 2014). Det er fristende å tenke at elever med høyere gjenkjennelighetsgrad i denne studien, altså elever som har svart i høyere svarkategorier for positivt rettede items og motsatt for negativt rettede items, nødvendigvis har en høyere mestringsforventning. Den norske oversettelsen av *Sources of Middel School Self-Efficacy Scale* gir imidlertid ikke et mål på elevenes mestringsforventning. Instrumentet gir derimot informasjon om hvor stor grad av gjenkjennelighet elevene rapporterer for situasjoner som antas å gi høyere mestringsforventning. Vi kan si at det er graden av elevenes tilgjengelighet på kildene som rapporteres, snarere enn hvor mye kilden påvirker deres mestringsforventning. På bakgrunn av denne studiens benyttede instrument, kan jeg dermed ikke konkludere med at høyere gjenkjennelighetsgrad betyr høyere mestringsforventning. Men det er faktorer som likevel gjør det fornuftig å anta at denne sammenhengen eksisterer. Siden høyere mestringsforventning medfører høyere prestasjon (Bandura, 1977), vil det være rimelig

å anta at mine høytpresterende elever også har høyere mestringsforventning[1]. Dette vil igjen utgjøre en mulig naturlig årsak til at de høytpresterende elevene i denne studien kjenner seg mer igjen i situasjoner som er antatt å gi høyere mestringsforventning.

Basert på teori og tidligere forskning (Bandura, 1977; Chen, 2003; Galyon et al., 2012; Phan, 2012; Salazar & Hayward, 2018; Usher, 2009; Usher & Pajares, 2008; Yurt, 2014), vil jeg si det er forventet at det eksisterer forskjeller mellom høyt- og lavtpresterende elevers rapportering av de ulike kildene, som også blir signifikant synlig i denne studien. Selv om forskjellen til en viss grad er forventet, er det interessant at forskjellen mellom høyt- og lavtpresterende elever viser seg i alle kildene. Dessuten er det et interessant funn fordi det åpner opp for en mer nyansert diskusjon rundt forskjellene vi ser i den ene elevgruppen sammenlignet med den andre. I det videre vil jeg trekke frem to av kildene som jeg mener viser forskjeller mellom elevgruppene, basert på størrelsen på forskjellene mellom høyt- og lavtpresterende elever og rekkefølgen de ulike kildene får innenfor hver av de to elevgruppene. Dette fremgår av tabell 6 og figur 14, der det er foretatt en z-standardisering som omgjør gjenkjennelighetsgraden oppgitt i logits til z-scores og dermed gjør det mulig å sammenligne på tvers av de ulike kildene. Wright Map's presentert i figur 15, 16, 17 og 18 vil også benyttes som verktøy for videre diskusjon.

5.2.2 Høyt- og lavtpresterende elevers rapportering av mestringserfaringer

Av figur 14 ser vi at den største forskjellen mellom de to elevgruppene finnes i kilden mestringserfaringer. Denne kilden ser ut til å skille seg ut både for høytpresterende elever og for lavtpresterende elever. De høytpresterende elevene rapporterer påstandene fra denne kilden til å stemme bedre enn de andre kildene. Og for de lavtpresterende ser vi det stikk motsatte; påstandene fra denne kilden er de påstandene som rangeres lavest. Denne forskjellen blir særlig interessant i lys av at denne kilden av Bandura (1997) blir fremhevet som den viktigste kilden for mestringsforventning, da det er det mest autentiske målet på egen evne. Tross mye variasjon i forskning på mestringsforventninger, fremgår mestringserfaringer som den kilden det er størst enighet rundt. Med visshet om viktigheten av mestringserfaringer som kilde til mestringsforventning, og det store spriket mellom høyt- og lavtpresterende elevers rapportering av kilden, ser jeg det særlig relevant å undersøke dette nærmere. Undergruppenes rapportering av mestringserfaringer er derfor det andre funnet jeg vil trekke frem.

[1] Denne sammenhengen er utenfor denne studiens omfang. En forskerkollega i SESAM (Elisabeth Terjesen) har imidlertid funnet at høytpresterende elever har signifikant høyere mestringsforventning enn lavtpresterende elever, ved bruk av del 1 og del 3 (vedlegg 2) i vår felles spørreundersøkelse. Masteroppgaven er per i dag ikke publisert, men er inne til sensur.

For at det skal være nyttig å se på denne forskjellen, er det først og fremst avgjørende å forsøke å forstå hvordan denne forskjellen kan tolkes. Det vi vet er at elevene gjennom det benyttede instrumentet har vurdert hvor godt påstandene stemmer for seg. Det kan imidlertid tenkes at elevene rapporterer at påstander stemmer dårlig av flere grunner. Det mest åpenbare vil være å tenke at elever som rapporterer at en påstand stemmer dårlig, mangler denne type erfaring. En elev som eksempelvis rapporterer item 2 *(Jeg har alltid lykkes i matte*) til å stemme dårlig, mangler kanskje erfaringer med å lykkes i matematikk. Dette er ganske åpenbart, og er en fornuftig forklaring på at det er lavtpresterende elever som rapporterer denne til å stemme dårligst.

Videre er det opplagt at elevenes egne tolkninger og vurderinger av situasjonen har påvirkning på elevenes rapportering av mestringserfaringer. Usher og Pajares (2008) har, i overensstemmelse med Bandura (1997, 2006), kritisert forskning som innhenter data om kilden mestringserfaringer ved å ta utgangspunkt i elevenes karakter i faget. Mestringserfaringer handler om elevenes egen vurdering av det å mestre, noe som også ble presisert underveis i denne studiens datainnsamling. Dette åpner altså opp for at elevenes tolkning av begreper som «gode karakterer», «alltid lykkes», «flink» og «får til å løse» (ordlyd hentet fra itemene i ME-instrumentet) påvirker hvor godt elevene vurderer påstandene til å stemme. Sett fra dette perspektivet vil man kunne tro at nettopp denne kilden, som innebærer stor grad av individuell tolkning, burde være noe mere uavhengig av de to elevgruppene. Resultatet viser imidlertid heller det motsatte. Videre vil jeg se på faktorer som kan bidra til å belyse og forklare nettopp dette.

Bandura (1997) sier at suksess bygger en robust tro på egne evner og dermed bygger mestringsforventninger, mens feil undergraver dem. Han trekker imidlertid frem en rekke faktorer som spiller inn for hvorvidt man tolker erfaringer som suksessfulle eller ikke. Erfaringene i seg selv er ikke alltid så betydningsfulle. Fordi samme grad av oppnådd suksess kan påvirke elevers mestringsforventninger på ulik måte, er det ikke tilstrekkelig å tro at mestringserfaringer kan måles ved hjelp av tidligere prestasjoner. Faktorer som forhåndsoppfatning av egne evner, vanskelighetsgrad på oppgaven, mengden hjelp mottatt og tidspunkt for suksess/feil i forhold til det totale mønsteret av erfaringer får betydning for innvirkningen mestringserfaringene får på mestringsforventningen (Bandura, 1997). Med andre ord er kilden mestringserfaringer kompleks og det er tilknyttede faktorer som påvirker hvor kraftfull denne kilden er. Bandura (1997) hevder at kunnskap om hvordan disse faktorene

påvirker den kognitive prosesseringen av erfaringene, vil bidra til å forstå hvordan man kan få mest mulig ut av sine egne mestringserfaringer.

Slik jeg tolker Bandura (1997) vil de overnevnte faktorene (1) kunne påvirke hvorvidt eleven selv oppfatter å ha mestret eller feilet, uavhengig av faktisk prestasjon, og (2) ha betydning for hvor mye erfaringen (enten det er mestring eller nederlag) faktisk påvirker deres mestringsforventning. Ved hjelp av ME-instrumentet innhentes det ikke informasjon for å vurdere punkt (2). Punkt (1) er derimot svært relevant, fordi bakenforliggende faktorer kan ha påvirket elevenes oppfatning av det å mestre eller feile, og vil derfor kunne ha preget resultatene av denne studien. Videre vil jeg argumentere for at et historisk perspektiv på matematikkfaget også kan bidra som en forklarende faktor til forskjellen denne studien viser i kilden mestringserfaringer.

Figur 14, som tydelig visualiser en forskjell mellom høyt- og lavtpresterende elevers rapportering av kilden mestringserfaringer, kan sies å gjenspeile oppfatningen om at matematikkfaget ses på som noe man enten får til eller ikke får til, nevnt innledningsvis. For de lavtpresterende elevene rapporteres denne kilden lavere enn de andre kildene, og motsatt for de høytpresterende elevene. Oppfatningene av å mestre eller feile kan ha røtter i matematikkfagets historie, i den forstand at faget lenge har vært preget av et sterkt resultatfokus, påpekt i kapittel 2.3.1 (Lampert, 1990; Skott et al., 2018; Stodolsky, 1985). Resultatfokuset tillegger resultatet større betydning, mens prosessen tilsvarende blekner. Et for stort fokus på instrumentell matematikk, og fravær av relasjonell matematikk, bygger opp under jakten på det korrekte svaret (Skemp, 1976). Jeg vil derfor argumentere for at den tradisjonelle skolematematikkens fokus på rett og galt svar kan påvirke elevenes tolkning av mestring og nederlag.

Som påpekt i dette delkapittelet er det mange faktorer som kan påvirke og forklare elevenes vurdering av mestringserfaringer. Det medfører at det ikke er en enkel oppgave å tolke hva det kommer av at denne studiens resultat viser at høytpresterende elever har en signifikant høyere gjenkjennelighetsgrad i kilden mestringserfaringer enn de lavtpresterende elevene. Den store forskjellen mellom høyt- og lavtpresterende elevers rapportering av denne kilden, tyder i alle fall på at det eksisterer et stort potensial i å fremme mestringserfaringer blant de lavtpresterende elevene. Dette potensialet vil jeg komme tilbake til i kapittel 5.3. Videre vil jeg diskutere forskjellen mellom høyt- og lavtpresterende i kilden vikarierende erfaringer.

5.2.3 Vikarierende erfaringer for høyt- og lavtpresterende elever

I forrige delkapittel diskuterte jeg kilden mestringserfaringer, som var den kilden med størst forskjell mellom høyt- og lavtpresterende elever gjenkjennelighetsgrad. Denne kilden var også mest iøynefallende med tanke på rekkefølgen for kildene i de to elevgruppene. Av figur 14 kan vi se at vikarierende erfaringer er den kilden der studien viser minst forskjell mellom høyt- og lavtpresterende elever. Likevel vil jeg påpeke at denne forskjellen er signifikant, og ikke betydelig mindre enn for de resterende kildene. Når jeg i forrige avsnitt diskuterte mestringserfaringer, tolket jeg det slik at en stor forskjell mellom elevgruppene kunne tyde på et stort potensial i å fremme de lavtpresterende elevenes mestringserfaringer. På den ene siden kunne jeg derfor valgt å tolke kilder med mindre forskjell mellom de to elevgruppene som mindre oppsiktsvekkende, og argumentere for at det ikke finnes et like stort potensial for «forbedring» i kilden VE. På den andre siden er det denne kilden de lavtpresterende elevene rapporterer størst grad av gjenkjennelighet i. Av den grunn velger jeg å trekke frem forskjellene mellom høyt- og lavtpresterende elever i kilden vikarierende erfaringer som mitt tredje funn.

Når jeg her refererer til rekkefølge, baserer denne seg på elevenes rapportering av de ulike situasjonene i itemene i instrumentet. Først i rekkefølgen finner vi kilder der elevene har rapportert større gjenkjennelighet i situasjoner som antas å påvirke mestringsforventningen i positiv retning. Funnene fra kapittel 4.2 viser at rekkefølgen for de høytpresterende elevene viser at mestringserfaringer kommer først, deretter oppmuntring, støtte og overtalelse fra andre, vikarierende erfaringer og psykologiske og fysiologiske tilstander (Se tabell 7). Det er kun mestringserfaringer som er signifikant forskjellig fra de andre kildene (Se figur 13). For de lavtpresterende elevene er rekkefølgen slik: vikarierende erfaringer, oppmuntring, støtte og overtalelse fra andre, psykologiske og fysiologiske tilstander og til slutt mestringserfaringer (se tabell 7), der det kun er mestringserfaringer som er signifikant forskjellig fra de andre kildene (se figur 12).

***Tabell 7**. Rekkefølge over hvilke kilder elever rapporterer størst grad av gjenkjennelighet i.*

Rekkefølge	Høytpresterende elever	Lavtpresterende elever
1	Mestringserfaringer	Vikarierende erfaringer
2	Oppmuntring, støtte og overtalelse fra andre	Oppmuntring, støtte og overtalelse fra andre
3	Vikarierende erfaringer	Psykologiske og fysiologiske tilstander
4	Psykologiske og fysiologiske tilstander	Mestringserfaringer

Vikarierende erfaringer er den av kildene der situasjoner som antas å fremme mestringsforventningen rapporteres til å stemme best for de lavtpresterende elevene. Til tross for at denne kilden ikke er signifikant forskjellig fra kildene oppmuntring, støtte og overtalelse fra andre og psykologiske og fysiologiske tilstander for de lavtpresterende elevene, velger jeg likevel å kommentere kort hva som kan være grunnen til at vikarierende erfaringer rapporteres så høyt for disse elevene. Jeg finner det relevant å se denne kilden i sammenheng med mestringserfaringer. Bandura (1997) trekker nemlig frem vikarierende erfaringers betydning for elever som har mange nederlagserfaringer, og påpeker at en vikarierende suksessfull erfaring kan bidra til at eleven tillegger egne tidligere nederlagserfaringer mindre vekt. Han hevder at slike omstendigheter kan medføre at vikarierende erfaringer blir en sterkere kilde til mestringsforventning enn mestringserfaringer, til tross for at sistnevnte vanligvis anses for å være den viktigste kilden. Yurt (2014) bygger på Bandura og legger til at vikarierende erfaringer har større effekt på prestasjon når personen har begrenset tilgang på egne erfaringer med å mestre eller ikke har grunnlag for å vurdere sin egen evne til å utføre oppgaven. I denne studien rapporteres kilden mestringserfaringer blant de lavtpresterende elevene signifikant lavere enn de andre kildene, og signifikant lavere enn de høytpresterende elevenes mestringserfaringer. Dette tyder på at de lavtpresterende elevene trolig har færre erfaringer med å mestre, og på den måten kan innlemmes blant dem Bandura (1997) hevder at vikarierende erfaringer kan bli særlig viktige for.

De tre hovedfunnene fra undersøkelsen presentert i dette delkapittelet er altså (1) at høytpresterende elever rapporterer en signifikant høyere gjenkjennelighetsgrad enn lavtpresterende elever i alle kildene, (2) det er en signifikant forskjell i hvordan kilden mestringserfaringer plasserer seg i rekkefølgen for kilder de kjenner seg best igjen i for høyt- og lavtpresterende elever, og (3) det er kilden vikarierende erfaringer som rapporteres til å stemme best for de lavtpresterende elevene, til tross for at det er i denne kilden forskjellen mellom høyt- og lavtpresterende elever er minst. I det kommende delkapittelet vil jeg bygge videre på disse funnene, og drøfte hvilken praktisk betydning de kan ha.

5.3 Praktiske implikasjoner

Når jeg nå i de tre forrige delkapitlene har trukket frem det jeg mener er de viktigste forskjellene som viser seg mellom høyt- og lavtpresterende elever i denne studien, er det relevant å stille spørsmål om hvilke praktiske implikasjoner disse funnene kan gi. En bevisstgjøring av eksisterende forskjeller i elevgruppenes rapportering av kilder til mestringsforventning, bør også gi relevante implikasjoner for læreren i klasserommet. I dette delkapittelet vil jeg derfor

diskutere hvilken betydning funnene kan ha for læreren i praksis, og hvorfor det er viktig å være bevisst på at disse forskjellene eksisterer.

Av de tre funnene påpekt i kapittel 5.2, som også er delvis overlappende, velger jeg først å rette søkelyset mot kilden mestringserfaringer, da denne kilden på mange måter utpeker seg. For det første er det i denne kilden det åpenbart er størst forskjell mellom de høyt- og lavtpresterende elevenes rapportering. For det andre vektlegger teori og tidligere forskning konsekvent denne kilden som den viktigste for mestringsforventninger (Bandura, 1997; Usher & Pajares, 2008). For det tredje rapporteres påstandene i ME-instrumentet til å være signifikant mindre gjenkjennelig enn påstandene i de andre kildene for de lavtpresterende elevene, og signifikant mer gjenkjennelig enn påstandene i de andre kildene for de høytpresterende elevene. Disse faktorene gjør det altså ekstra relevant å se på hvilke praktiske implikasjoner denne bevisstgjøringen kan ha for undervisning og læreren i klasserommet. Til tross for at studien viste signifikante forskjeller i alle kilder, vil jeg ta utgangspunkt i kilden mestringserfaringer i den videre diskusjonen, da jeg mener denne kilden tydeligst viser ulikheter i mønsteret for de to elevgruppenes kilder til mestringsforventning. I kapittel 5.3.2 vil jeg benytte instrumentets items og elevenes rapportering av disse til å kommentere faktorer som ser ut til å være viktig innenfor flere av kildene. Formålet med dette vil være å skape bevissthet rundt faktorer som kan bidra til at elever får mer nytte av kildene til mestringsforventning.

5.3.1 Et stort potensial for å fremme lavtpresterende elevers mestringserfaringer

Avslutningsvis i kapittel 5.2.2 nevnte jeg at den store forskjellen i kilden mestringserfaringer for de to elevgruppene, kan være et tegn på at det er et stort potensial for å fremme lavtpresterende elevers mestringserfaringer i matematikk. Jeg vil her trekke opp en diskusjon omhandlende hvordan lærere best mulig kan fremme elevers mestringserfaringer, og da særlig med blikket rettet mot de lavtpresterende elevene.

(...) teachers of mathematics should take care to deliver instruction in such a way as to maximize the opportunity for «authentic mastery experiences», however small.
- (Usher, 2009, s. 308)

Usher (2009) sitt sitat over uttrykker at lærerens instruksjon må gjøres på en slik måte at det kan maksimere muligheten for mestringserfaringer. Dette anser jeg som interessant, fordi forfatteren viser til lærerens instruksjon som avgjørende. Hampton (1998) forsket på kildene til mestringsforventning hos elever med lærevansker, og fant at det er særlig viktig å fremme suksessfulle erfaringer blant disse elevene. Hun viser til Bandura (1993) og oppfordrer rektorer

til å legge bedre til rette for at alle elever kan oppleve mestringserfaringer, til tross for eksisterende lærevansker. Det blir altså tydelig at det er grep som kan gjøres for å fremme elevers mestringserfaringer. Denne studien viser at det eksisterer et stort potensial når det gjelder lavtpresterende elevers mestringserfaringer. Spørsmålet blir derfor, hvordan kan man som lærer i klasserommet legge til rette for at lavtpresterende elever opplever autentiske mestringserfaringer? For å best mulig kunne besvare dette spørsmålet, vil jeg først og fremst trekke frem litteratur og forskning som utdyper hva som påvirker mestringserfaringene, samt vise tilbake til de forklarende faktorene jeg la frem i kapittel 5.2.2.

Pajares (2006) observerte unge menneskers varierende tolkninger av mestringserfaringer, og peker på situasjoner der uheldige tolkninger ødela for den mestringsforventningen som ville pushet dem videre. Vi forstår altså at elevenes tolkning av egne erfaringer er avgjørende for nytteverdien av erfaringene. En av de tidligere nevnte faktorene for hvordan elevene tolker sine erfaringer, er elevenes oppfatning av egen evne (Bandura, 1997). Synet på evne kan blant annet være avhengig av om elevene har et dynamisk eller et statisk tankesett (Boaler, 2013). Det essensielle fra teorien om tankesett, er at et statisk tankesett vil medføre at elevene tror deres evner er medfødt og uforanderlig, mens et dynamisk tankesett innebærer en forståelse av at evner kan utvikles gjennom innsats (Wæge & Nosrati, 2018). Jeg vil vise til et item fra undersøkelsen, som jeg mener kan si noe om elevenes syn på nettopp dette.

Av figur 15 kan vi se at item 3 *(selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte)* i undersøkelsen er den påstanden som ligger nederst mot «items som stemmer godt». Siden item 3 er reversert, betyr dette det motsatte; at mange elever mener dette ikke stemmer. Det vil ikke være veldig nyttig å sammenligne det reverserte itemet med de andre itemene i delinstrumentet fordi den reverserte ordlyden kan påvirke resultatene noe, slik også tidligere nevnt (Usher & Pajares, 2009). Men jeg vil likevel argumentere for at det er betryggende å se tendensen blant elevene. En gjennomsnittlig lavtpresterende elev vil trolig svare i kategori 3 eller 4 på påstanden, mens en gjennomsnittlig høytpresterende elev vil trolig svare i kategori 1 for påstanden (figur 15 viser svarkategoriene i motsatt rekkefølge for item 3, grunnet at dette er reversert i analysene). Vi ser også at det er svært få elever som vil svare i svarkategori 6 for denne påstanden. Elevene svarer altså generelt i lavere svarkategorier på denne påstanden. Dette tyder på at elevene til en viss grad ser verdien av å nedlegge en innsats, og jeg tolker det slik at mange elever har tro på at det å jobbe veldig hardt også vil påvirke hvordan de gjør det i matematikk.

Til tross for at det tilsynelatende ser ut som mange elever innerst inne forstår at innsatsen vil påvirke prestasjonen, vil jeg påpeke at elever med et statisk tankesett ofte tar i bruk unngåelsesstrategier for å slippe å oppleve å mislykkes (Bandura, 1993). Elevenes oppfatning av det å gjøre feil avhenger nemlig også av elevenes tankesett og kan påvirke elevenes vurdering av egne evner (Wæge & Nosrati, 2018). Med et dynamisk tankesett vil feil være en naturlig del av prosessen, mens man med et statisk tankesett oftere vil forsøke å unngå å feile, nettopp fordi dette kan oppfattes som et tegn på lav evne (Bandura, 1993). Dette kan medføre at elever unngår å legge ned stor innsats, nettopp fordi fallhøyden kan oppleves som enda større dersom elevene skulle mislykkes (Wæge & Nosrati, 2018). Når man ikke engang har prøvd å få til noe, vil nederlaget sannsynligvis ikke påvirke vurderingen av egne evner (Bandura, 1997). Slike unngåelsesstrategier vil frarøve elevene muligheten til å få gode mestringserfaringer. Mestringserfaringer der individet legger ned en stor innsats vil særlig for barn medføre at erfaringen kan få større innvirkning på mestringsforventningene. Dersom innsatsen derimot senkes av frykt for å mislykkes, vil dette trolig resultere i en ond sirkel, der eleven får færre autentiske opplevelser av å mestre grunnet den lave innsatsen. Det er kanskje nettopp derfor læreplanen (Kunnskapsdepartementet, 2019), som påpekt i kapittel 2.3.1, oppfordrer skolen til å skape et miljø der prøving og feiling blir en naturlig del av prosessen.

Jeg vil her vise tilbake til kapittel 5.2.2 hvor jeg argumenterte for at den tradisjonelle skolematematikkens fokus på rett og galt svar kan være en forklarende årsak til at mange elever enten sier de får til eller ikke får til matematikk. Denne enten/eller -holdningen gjenspeiles også i denne studiens resultater, særlig i kilden mestringserfaringer. Den nyeste læreplanen , LK20, viser imidlertid et solid forsøk på å komme vekk fra den oppfatningen av matematikkfaget (Kunnskapsdepartementet, 2019). Elevene skal legge mer vekt på strategier og fremgangsmåter enn på selve løsningene. Læreplanens overordnede del bygger også opp under dette, ved å påpeke at forståelse rundt egne læreprosesser og egen faglig utvikling vil bidra til selvstendighet og mestringsfølelse (Kunnskapsdepartementet, 2019). Det trekkes frem at skolens forventninger til hver enkelt elevs innsats og mestring er betydningsfull for elevens tro på egne evner. Det fremheves også at prøving og feiling kan være en kilde til læring, og at skolen derfor skal oppfordre og gi elevene trygghet til å prøve seg på oppgaver, også når de ikke er sikre på å lykkes. Dette vil trolig bidra til en kultur der feiling blir en naturlig del av prosessen, og dermed ikke i like stor grad vil tolkes som et nederlag eller tegn på lav evne av elevene.

En annen faktor som antas å være av betydning for om man tolker erfaringene sine som mestring eller nederlag, er oppgavens vanskelighetsgrad (Bandura, 1997). Er oppgavene svært lite

krevende, vil mestring av oppgaven trolig ikke tilføre ny informasjon om personens evne. Er det derimot en krevende oppgave som eleven mestrer, kan det bli en mestringserfaring som får større betydning for mestringsforventningene til eleven.

Frem til nå har jeg trukket frem flere faktorer av betydning for elevenes mestringserfaringer. Elevenes tolkning av egne erfaringer er essensielt, og vil blant annet påvirkes av vurderingen av egne evner og oppgavens vanskelighetsgrad. Elevenes tankesett, som påvirker deres syn på feil og innsats og dermed også syn på evne, vil også kunne påvirke tolkningen av egne erfaringer. I tillegg trekker Bandura frem kontekst og tidspunkt i ferdighetsutviklingen som faktorer av betydning for mestringserfaringene. Videre vil jeg sette søkelyset på læreren, og diskutere hvordan læreren gjennom utvikling av gode sosiomatematiske normer, beskrevet i kapittel 2.5, kan skape et miljø som kan fremme elevers mestringserfaringer.

Bandura (1993) antyder at sosiale sammenligninger ofte vektlegges i miljøer der evner tolkes som en ferdighet som kan utvikles. Fordi de fleste aktiviteter ikke har en standardisert skala for å vurdere sine egne evner, er det vanlig å sammenligne seg med andre. Hvilke personer man sammenligner seg med påvirker også hvordan man tolker sine egne evner. Bandura (1993) påpeker imidlertid at det særlig er sammenligning med seg selv, altså personlig framgang og prestasjoner, som er godt egnet for å fremme mestringsfølelse. Typisk for elever med et dynamisk tankesett er å heller sammenligne seg med egen forbedring enn andres prestasjoner. Han setter lys på tilbakemeldinger, og hevder at tilbakemeldinger med fokus på oppnådd progresjon understreker de personlige evnene, mens tilbakemeldinger som fokuserer på feil og mangler fremhever de personlige utilstrekkelighetene. Han spesifiserer at tilbakemeldinger omhandlende evne, for eksempel at elevers framgang viser at de har evne for faget, er særlig virkningsfullt i begynnelsen av ferdighetsutviklingen (Bandura, 1997).

I lys av Bandura (1993) tolker jeg det slik at lærerens tilbakemeldinger til elevene kan være av stor betydning for hvilken sammenligning elevene foretar seg, og dermed av stor betydning for om ulike erfaringer tolkes som mestring eller nederlag. Hvis læreren i sin tilbakemelding klarer å rette fokuset mot progresjon i forhold til tidligere prestasjoner i lignende situasjoner, vil man kanskje også kunne flytte elevens fokus fra å sammenligne seg med andre til å sammenligne seg med seg selv.

Ut ifra hvordan mestring legges frem i overordnet del av LK20, ser læreren ut til å ha en sentral rolle for å fremme elevenes mestring, særlig i form av å skape gode sosiomatematiske normer i et trygt miljø (Kunnskapsdepartementet, 2019). Hvis læreren klarer å etablere en

klasseromskultur der elevene verdsetter prosessen rundt matematikken fremfor svaret, så er det også rimelig å tro at elevene lettere vil kunne tolke en prestasjon som suksess, til tross for at det endelige svaret ikke nødvendigvis var korrekt. Et større fokus på den matematiske prosessen, kan også muligens bidra til at elevene får et bedre innblikk i oppgavens vanskelighetsgrad. Da vil elevene kunne vurdere sine egne evner på et bredere grunnlag. For å oppnå dette, kan det kanskje være nyttig å trekke frem «transmissionist»- og «connectionist»-undervisning. Førstnevnte vil være mer motivert av selve skoleringen, og prestasjoner og karakterer er målet. En slik undervisning vil ikke bidra til å underbygge viktigheten av den matematiske prosessen (Williams et al., 2010). «Connectionist»-undervisning vil derimot være motivert av å bygge forståelse hos elevene, og en slik type undervisning vil i større grad bidra til å fremme prosessen fremfor svaret.

For å oppsummere dette delkapittelet, vil jeg vise til Bandura (1997), som hevder at det kan være vanskelig å sile ut effekten fra en enkelt kilde, da informasjon fra en kilde kan påvirke de andre kildene. At lærerens tilbakemeldinger kan være av betydning for elevenes tolkning av egne evner, kan kanskje nettopp være et eksempel på at kilden oppmuntring, støtte og overtalelse fra andre kan påvirke elevenes mestringsforventning gjennom kilden mestringserfaringer. Tilsvarende kan vi si at kilden vikarierende erfaringer kan bidra til flere mestringserfaringer ved å kaste lys over oppgavens vanskelighetsgrad, som kan føre til en mer realistisk tolkning av egen erfaring.

I avsnittene over har jeg argumentert for at læreren, gjennom tilbakemeldinger og annen samhandling med elevene, kan påvirke hvordan elevene tolker sine erfaringer. Siden denne tolkningen både påvirkes av hvem elevene sammenligner seg med og hvordan de ser på egne evner, kan læreren gjennom å bygge opp gode sosiomatematiske normer i klasserommet bidra positivt til elevenes tolkning av egne erfaringer. Jobber læreren med å fremme et dynamisk tankesett, kan dette virke positivt på elevenes vurderinger fordi elevene blir bevisste på at evne er noe som kan utvikles. Dette vil trolig også innby til å sammenligne seg med seg selv og sin egen utvikling. En lærer som bevisst jobber for å bygge opp sosiomatematiske normer som fremmer verdien av den matematiske prosessen, kan bidra til å gi elevene en forståelse av at det å feile er en del av prosessen. Lærerens instruksjoner ser altså ut til å være svært sentrale for å gi elevene en sunn forståelse og tolkning av sine egne erfaringer, og vil trolig også resultere i at elevene kan få flere mestringserfaringer. Usher og Pajares (2008) påpeker at de forklaringene som signifikante andre gir til ungdom om sine egne nederlags- eller mestringsopplevelser

påvirker måten ungdom forklarer sine nederlags- eller mestringsopplevelser. Dette underbygger viktigheten av lærerens egne holdninger og lærerens bevissthet rundt dem.

Når Usher (2009) oppfordrer til å maksimere mulighetene for autentiske mestringsopplevelser, advarer hun samtidig mot overfladiske forsøk på å skape mestring. Dette kan være å gi enkle oppgaver til elevene, for at de skal oppleve å få til oppgaven og unngå nederlag. Dette vil likevel ikke ha stor effekt, fordi oppgavens vanskelighetsgrad også er med å påvirke hvorvidt erfaringene får betydning for mestringsforventningene. Når jeg har diskutert hvordan læreren kan maksimere muligheten for autentiske mestringsopplevelser, er det imidlertid viktig at det ikke forveksles med å hindre ethvert nederlag. Pajares (2006) påpeker nemlig at en av de viktigste karakteristikkene til suksessfulle mennesker er at feiling og motgang ikke undergraver mestringsforventningen deres. Han hevder at mestringsforventning ikke handler så mye om hvordan man skal lykkes, men snarere om å lære å holde ut når man ikke lykkes.

5.3.2 Hvordan fremme mestringsforventningene i lys av instrumentenes items?

Etter å ha viet kilden mestringserfaringer mye oppmerksomhet, vil jeg nå også kommentere noen av de andre kildene. Slik jeg påpekte i kapittel 5.2.1, er det en signifikant forskjell mellom høyt- og lavtpresterende elevers rapportering av alle de fire kildene til mestringsforventning. Selv om dette til en viss grad kan sies å være forventet, bør det likevel gi praktiske implikasjoner i form av et særlig fokus på å fremme kildene hos lavtpresterende elever. Når vi vet at mestringsforventningene utgjør et viktig bidrag for akademisk prestasjon, blir det også naturlig å diskutere hvordan læreren kan bidra til å fremme de ulike kildene hos sine elever. Dette er et stort spørsmål, og det er også mye teorien om mestringsforventning kan si om dette. Jeg vil imidlertid benytte meg av de innhentede dataene til å si noe om hvordan elevene i mitt utvalg rapporterer de ulike itemene i del-instrumentene. Med utgangspunkt i kildenes Wright Map's, presentert i figur 15, 16, 17 og 18, vil jeg se nærmere på itemene, for å videre kunne si noe generelt om kildene.

Først vil jeg vise til Wright Map for VE-instrumentet (se figur 16), der vi kan se at særlig item 7 (*Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre)* og item 9 *(Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre)* er påstander som elevene rapporterer til å stemme mindre godt enn de andre itemene i delinstrumentet. De resterende itemene rapporteres omtrent likt av elevene. Jeg vil derfor diskutere hva som kan være årsaken til dette. En gjennomsnittlig lavtpresterende elev i mitt utvalg vil sannsynligvis svare i svarkategori 1 eller 2 på item 7 og sannsynligvis i svarkategori 2 på item 9. En

gjennomsnittlig høytpresterende elev vil sannsynligvis svare i kategori 3 eller 4 på item 7 og kategori 4 eller 5 på item 9. I VE-instrumentet er det to items som omhandler vikarierende erfaringer fra voksne. Det ene er item 7 og det andre er item 8 *(Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte).* Det samme gjelder for vikarierende erfaringer fra medelever, der vi i tillegg til item 9 har item 10 *(Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte).* Item 8 og 10 er forholdsvis like, og innebærer at eleven allerede har fått se en løsningsmetode på en oppgave. Dette kan kanskje være årsaken til at disse påstandene rapporteres til å stemme bedre enn item 7 og 9. Bandura (1997) trekker modellering ved hjelp av stedfortredende mestringserfaringer frem som positivt, både fordi eleven kan lære effektive strategier og metoder, men også fordi eleven kan bli overbevist om selv å lykkes med samme oppgave. Item 7 og 9 inneholder ikke direkte informasjon om situasjoner der modellering nødvendigvis foregår. Dette belyser viktigheten av at lærere legger til rette for at strategier og metoder fremheves i situasjoner der modeller utfører en oppgave foran andre.

Virkningen av vikarierende erfaringer avhenger av individets vurdering av likhetstrekk til personen som utfører oppgaven og individets oppfattelse av personens evner (Wæge & Nosrati, 2018). Formuleringen fra item 9, «(…) *gjør det bedre enn meg i matte (...)»,* kan kanskje tolkes som medelever med større evner for matematikk, og dermed påvirke vurderingen om hvorvidt dette er motiverende for egen innsats til å gjøre det bedre. En oppfattelse av stor forskjell i evner, kan bidra til at elevene ikke like godt kan relatere seg til personen, og dermed ha en mindre virkning på elevenes mestringsforventning. Item 9 kan trolig være mindre gjenkjennelig for elevene enn item 7 fordi det kan være vanskeligere å identifisere seg selv med en voksen enn en medelev. Jo mer eleven kan identifisere seg med modellen, jo mer overbevisende oppleves det at eleven selv kan få til det samme (Bandura, 1997).

Andre forskere har kommentert de begrensningene noen av itemene i instrumentet har, blant annet at item 7 *(Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre)*, krever at respondenten både må svare på hvorvidt vikarierende erfaringer forekommer, men også virkningen av erfaringen (Ahn, Bong & Kim, 2017). Det samme gjelder for item 9. Dette kan muligens også påvirke elevenes rapportering noe.

Jeg vil påpeke at informasjonen om itemene i VE-instrumentet underbygger teorien om at vikarierende erfaringer blir mer betydningsfulle når eleven kan identifisere seg med den som utfører oppgaven (Bandura, 1997). Elevenes oppfattelse av personens evner, vil også ha betydning. For læreren i klasserommet kan dette kanskje belyse viktigheten av å sørge for at

også lavtpresterende elever får se eksempler der personer de kan relatere seg til utfører oppgaver. Rent praktisk kan dette for eksempel bety at læreren har fokus på at også lavtpresterende elever utfører matematikkoppgaver på tavla foran resten av klassen.

Jeg vil også vise til Wright Map for OA-instrumentet (se figur 18), der vi ser at det er item 18 (*Klassekameratene mine liker å jobbe med meg i matte, fordi de syns jeg er god)* som generelt rapporteres til å stemme dårligst. Deretter følger item 17 *(Medelevene mine har fortalt meg at jeg er flink i matte).* Disse itemene omhandler jevnaldrende. Samtidig ser vi at item 15 *(voksne i familien min har fortalt meg at jeg er flink i matte)* er det itemet som rapporteres til å stemme best. Verdien av oppmuntring, støtte og overtalelse blir ekstra stor når det kommer fra signifikante andre (Wæge & Nosrati, 2018). Lærere og foreldre er eksempler på dette. I ungdomsskolealder vil også jevnaldrende være særlig betydningsfulle (Utdanningsdirektoratet, 2016). Itemene omhandlende jevnaldrende i OA-instrumentet (item 17 og 18) rapporteres imidlertid til å stemme dårligere enn de andre itemene. Jeg vil derfor argumentere for at en bevisstgjøring av dette, kanskje kan bidra til at lærere i større grad kan skape et klassemiljø der elevene også kan rose hverandre og gi hverandre tilbakemeldinger. Det er imidlertid viktig at oppmuntring, støtte og overtalelse alltid er realistisk. Dessuten skal det ikke være overdrevet, på den måten at man roser elever med lavere matematiske evner mer enn andre som gjør det like bra. Forskning viser at dette kan føre til at elevene senker sin oppfatning av egne evner (Lord, Umezaki & Darley, 1990; Meyer, 1992, gjengitt i (Bandura, 1997)).

På bakgrunn av hvordan elevene i denne studiens utvalg har besvart itemene, ser det ut til at modellering av strategier og løsningsmetoder er viktig for virkningen av vikarierende erfaringer. I tillegg synliggjøres betydningen av at elevene kan identifisere seg med modellen som utfører den vikarierende erfaringen. Når det gjelder overtalelse, støtte og oppmuntring fra jevnaldrende, viser undersøkelsen at dette er situasjoner mange elever ikke kjenner seg like godt igjen i. Jeg mener en bevisstgjøring av disse tre rapporteringene kan bidra til å gi læreren nyttig informasjon om hvordan elever benytter seg av de ulike kildene. Informasjonen danner videre et grunnlag for lærerens arbeid med å legge til rette for at elevene skal få mest mulig ut av kildene.

Til slutt vil jeg påpeke at dette delkapittelet statuerer et eksempel på hvordan lærere kan benytte instrumentet til Usher og Pajares (2009) som en inngang til mer informasjon om sine elevers oppfattelse av kildene. Dette kan danne et grunnlag for å bedre tilpasse undervisningen til den aktuelle elevgruppen. Nettopp dette var også innlemmet i forfatternes intensjon for videre bruk av itemene i instrumentet.

5.4 Oppsummering

I kapittel 5.0 har jeg diskutert resultater av undersøkelsen opp mot teori og tidligere forskning, med et mål om å forsøke å besvare forskningsspørsmålene. I kapittel 5.1 har jeg påpekt viktigheten av en grundig valideringsprosess, og oppsummerende beskrevet hvilke faktorer som ble viktig for å oversettelsen og valideringen av instrumentet benyttet i denne studien. I kapittel 5.2 har jeg videre løftet frem tre hovedfunn fra undersøkelsen, med fokus på forskjeller i høyt- og lavtpresterende elevers kilder til mestringsforventning. På bakgrunn av disse funnene argumenterte jeg for at det eksisterer et stort potensial for å fremme lavtpresterende elevers mestringserfaringer, og diskuterte deretter i kapittel 5.3 den praktiske betydningen av disse funnene. Spørsmålet om hvordan læreren kan arbeide for å maksimere mulighetene for at lavtpresterende elever kan oppleve autentiske mestringserfaringer i matematikk ble diskutert. I tillegg benyttet jeg meg av denne studiens innsamlede data til å se nærmere på elevenes rapportering av de enkelte itemene innenfor flere kilder og til å påpeke enkelte faktorer som kan bidra til at læreren bedre kan tilrettelegge for elevenes mestringsforventninger.

6.0 Avslutning

Avslutningsvis vil jeg først vise til en konklusjon der jeg forsøker å besvare forskningsspørsmålene problemstillingen reiser. Deretter vil jeg gjøre noen kritiske refleksjoner ved å påpeke svakheter ved studien og dens anvendelse. Jeg vil også komme med innspill til videre forskning, før jeg sier noe om masteroppgavens praktiske nytteverdi for meg i mitt fremtidige lærervirke.

6.1 Konklusjon

I denne masteroppgaven har jeg sett på høyt- og lavtpresterende elevers forskjeller i kilder til mestringsforventning og tatt utgangspunkt i to forskningsspørsmål. Det første forskningsspørsmålet lyder slik: «*Hvordan kan et instrument som måler elevers kilder til mestringsforventning oversettes og valideres for bruk på norske elever?*». I denne studien har jeg benyttet meg av et allerede validert instrument, *Sources of Middle School Mathematics Self Efficacy Scale*, utviklet av Usher og Pajares (2009). Jeg har videre oversatt instrumentet og validert det for bruk på norske 9.klassinger. Prosessen med å oversette instrumentet er beskrevet i kapittel 3.2.3 og prosessen med å validere instrumentet er både representert i kapittel 3.0 (piloten) og kapittel 4.0 (hoveddata). I kapittel 5.1 begrunner jeg behovet for å utføre en grundig valideringsprosess av det oversatte instrumentet, og påpeker elementer som var viktige for prosessen. Jeg trekker også frem Rasch-analysens bidrag til instrumentets kvalitet.

Det andre forskningsspørsmålet lyder slik: «*Er det forskjeller i hvordan høyt- og lavtpresterende elever rapporterer de ulike kildene til mestringsforventning, og eventuelt hvilke mønstre viser seg?*». Undersøkelsens resultat viser signifikante forskjeller mellom høyt- og lavtpresterende elevers kilder til mestringsforventning. Høytpresterende elever rapporterer signifikant høyere grad av gjenkjennelse i påstander som antas å påvirke mestringsforventningen positivt, sammenlignet med lavtpresterende elever. Dette gjelder for alle kildene. Den største forskjellen mellom de to elevgruppene finner vi i kilden mestringserfaringer. Av de fire kildene, er det denne kilden de lavtpresterende rapporterer minst gjenkjennelighet i. For de høytpresterende ser vi det motsatte. Disse elevene rapporterer høyest grad av gjenkjennelighet for nettopp denne kilden. Det er dessuten bemerkelsesverdig at påstandene fra kilden mestringserfaringer rapporteres til å stemme signifikant dårligere enn påstandene fra de resterende kildene for lavtpresterende elever, mens påstandene fra kilden mestringserfaringer rapporteres til å stemme signifikant bedre enn påstandene fra de andre kildene for høytpresterende elever.

Videre har jeg drøftet hvilken praktisk betydning disse funnene kan få for lærerens undervisning i klasserommet. Først og fremst mener jeg at en bevisstgjøring av forskjeller i høyt- og lavtpresterende elever kilder til mestringsforventning er viktig for prinsippet om tilpasset opplæring (Opplæringslova, 1998). Den tydelige forskjellen i høyt- og lavtpresterende elevers rapportering av kilden mestringserfaringer, kombinert med vissheten om denne kildens betydning for mestringsforventning (Bandura, 1997), gjorde det relevant å drøfte hvordan man som lærer kan bidra til å fremme de lavtpresterende elevenes muligheter for å oppleve autentiske mestringserfaringer. I lys av et historisk perspektiv på matematikkundervisning, og i lys av de matematikkfaglige linjer vi ser nyere læreplaner trekke opp (Kunnskapsdepartementet, 2019), har jeg argumentert for at lærerens holdninger til evner og feil i matematikk kan spille en rolle for elevenes tolkning av sine erfaringer. Jeg har også trukket frem lærerens instruksjoner og sosiomatematiske normer i klasserommet som viktige hjelpemidler for å fremme verdien av den matematiske prosessen. Dette kan igjen bidra til at elever i større grad kan tolke sine erfaringer som mestring, til tross for at resultatet ikke nødvendigvis er korrekt. Forhåpentligvis kan det også bidra til å komme mer bort fra det tradisjonelle rett/galt-faget, nevnt innledningsvis i denne masteroppgaven. Jeg har også trukket frem at mestringsforventning handler om å holde ut når man ikke lykkes, snarere enn å lykkes (Pajares, 2006) Å jobbe for at feiling og motgang ikke undergraver elevenes mestringsforventning, vil også være i samsvar med begrepet om livsmestring i LK20, der det fremgår at målet med livsmestring blant annet er at elevene skal lære å håndtere både medgang og motgang på best mulig måte (Kunnskapsdepartementet, 2017).

6.2 Kritiske refleksjoner

I dette delkapittelet vil jeg påpeke noen svakheter ved studien. Først vil jeg trekke frem hvordan høyt- og lavtpresterende elever ble delt inn. Som forklart i kapittel 3.2.2 ble det ikke utført Rasch-analyser av instrumentet som målte elevenes prestasjon (instrument 3). Dette medfører at vanskelighetsgraden til de 12 oppgavene ikke ble hensyntatt når elevene fikk sin score. Til tross for at vi benyttet oppgaver hentet fra TIMMS-undersøkelsen, erfarte vi at oppgavene generelt var på et høyt nivå for vårt utvalg elever. Dette kunne vi se av elevenes poengscore, der en langt større andel av elevene fikk til en eller to oppgaver, mens ingen av elevene klarte alle 12 oppgavene. Hadde jeg også utført Rasch-analyser på dette instrumentet etter piloteringen, og deretter justert oppgavenes vanskelighetsgrad, kunne jeg sannsynligvis fått en jevnere spredning av elever som fikk til de ulike oppgavene. Instrumentet kunne da målt elevenes prestasjon mer nøyaktig. Grunnet masteroppgavens tidsramme og omfang, ble dette

imidlertid ikke mulig, og jeg gjorde derfor utvalget basert på poengscore. I kapittel 3.5.3 argumenterte jeg for at dette er godt nok for min studie. Jeg ønsker likevel å påpeke dette som en svakhet ved studien.

I kapittel 4.1.5 har jeg konkludert med at instrumentet benyttet til å måle elevers kilder til mestringsforventning er valid. Likevel vil jeg også vise til enkelte svakheter som kom til syne i analyseprosessen og satte noen begrensninger for den videre tolkningen av resultatene. Itemenes *measure*, rapportert i delkapitlene 4.1.1 – 4.1.4, viser for noen av delinstrumentene varierende sprang mellom itemene. Et stort sprang mellom itemenes *measure*, tilsier at noen av elevene ikke blir målt like nøyaktig ved instrumentet. Dette så vi et eksempel på i ME-instrumentet rapportert i tabell 1. Her vises også et eksempel på det motsatte, nemlig at det er flere av itemene som måler de samme elevene og på den måten overlapper hverandre. Dette vil bety at noen av itemene blir overflødig. OA-instrumentet, rapportert i tabell 3, er et godt eksempel på jevn spredning i itemene. Til tross for at jeg ikke mener dette har store konsekvenser for min tolkning av resultatene, er det likevel verdt å påpeke som en svakhet.

Jeg vil også rette et kritisk blikk mot utvalgsmetoden brukt i denne studien. Som beskrevet i 3.5.3 ble det gjort et bekvemmelighetsutvalg. Fordi denne utvalgsmetoden ikke er et sannsynlighetsutvalg kan jeg ikke være sikker på at utvalget er representativt for hele populasjonen, noe som er en forutsetning for å gjøre hypotesetestene utført i kapittel 4.2. Dersom utvalget ikke er representativt, vil resultatene fra utvalget ikke kunne generaliseres til hele populasjonen (Postholm & Jacobsen, 2011). Jeg vil imidlertid nevne at det ikke er kjente faktorer som skulle tilsi at utvalget ikke er representativt. Fordi jeg ikke kan være sikker på at utvalget er representativt, vil jeg derfor påpeke dette som en begrensning ved studien. Til tross for at jeg i denne studien finner signifikante forskjeller mellom høyt- og lavtpresterende elevers rapportering av kildene til mestringsforventning, vil jeg ikke uttale meg bastant om at dette er generaliserbart til hele populasjonen.

Utvalgsstørrelsen kan også kommenteres i denne sammenhengen. Jo større utvalget er, jo mindre blir feilmarginene når vi undersøker om resultatene er signifikante. Som tidligere nevnt satte situasjonen rundt COVID-19 en begrensning for datainnsamlingen, som medførte at jeg og medstudenter i SESAM ikke fikk samlet inn like mye data som først planlagt.

Til slutt vil jeg nevne at manglende motivasjon og villighet blant elevene i vårt utvalg kan ha påvirket datamaterialet. Til tross for at undersøkelsen var frivillig, hadde læreren i de enkelte

klassene ofte en forventning om at alle elevene skulle delta. Dette kan sies å være et resultat av måten dataene ble samlet inn på. Siden vi var avhengige av at en fra forskningsgruppa SESAM besøkte skolen, ble vi gitt en skoletime til rådighet for å utføre undersøkelsen. Undersøkelsen ble gjort i helklasse, og det medførte derfor en slags forventning om at alle elever skulle delta. Lite interesserte elever kan derfor være en svakhet ved datamaterialet som ble benyttet i studien.

6.3 Innspill til videre forskning

Med støtte i Gao (2019) vil jeg påpeke kompleksiteten til mestringsforventning og dets kilder. Jeg vil hevde at denne kompleksiteten, samt at mestringsforventning er situasjons- og kontekstavhengig, skaper et behov for flere studier som omhandler elevenes kilder til mestringsforventninger i matematikk. Yildiz og Özdemir (2019) etterlyser flere kvantitative studier på mestringsforventning som inkluderer prestasjonsnivå som en faktor. Jeg mener resultatene av denne masteroppgaven synliggjør et ytterligere behov for forskning på høyt- og lavtpresterende elevers kilder til mestringsforventning i matematikk. Resultatene viser at det er tydelige forskjeller i de to elevgruppenes rapportering av kildene, og det trengs derfor mer forskning til for å undersøke disse forskjellene. I denne studien har jeg drøftet hvordan man som lærer kan fremme lavtpresterende elevers mestringserfaringer, som følge av forskjellen i undergruppenes rapportering av denne kilden. Mer systematisk forskning, særlig på denne kilden, mener jeg bør være gjenstand og søkelys for videre forskning på feltet. Dette finner jeg også støtte til av Bandura (1997), som hevder at kunnskap om hvordan ulike faktorer påvirker den kognitive prosesseringen av elevenes erfaringer, vil bidra til å forstå hvordan man kan få mest mulig ut av sine egne mestringserfaringer.

6.4 Praktisk nytteverdi

I masteroppgavens innledning tydeliggjorde jeg at et ønske for masterstudiet var å få innsikt i matematikkfaglig forskning som kunne gi meg praktisk nytte i mitt fremtidige lærervirke. En fordypning i teori og forskning om mestringsforventning i matematikk har ført til at jeg i større grad ser viktigheten av å fremme dette hos mine fremtidige elever. Resultatene av denne studien har ført til en bevisstgjøring av forskjellene i høyt- og lavtpresterende elevers rapportering av kildene til mestringsforventning, og særlig innen kilden mestringserfaringer. Dette medfører at jeg som lærer vil gjøre en ekstra innsats for å fremme lavtpresterende elevers mestringsforventning i matematikkfaget, og også være oppmerksom på de faktorene som fremmer mestringserfaringer. Jeg ønsker derfor å skape gode sosiomatematiske normer som kan bidra til å fremheve den matematiske prosessen som verdifull og fremme et dynamisk

tankesett. Samtidig vil jeg hjelpe elevene med tolkningen av sine egne erfaringer, og oppmuntre til utholdenhet når de opplever motgang. På denne måten kan jeg også bidra til å fremme livsmestring i matematikkfaget (Kunnskapsdepartementet, 2017).

Referanser

Ahn, H. S., Bong, M. & Kim, S.-i. (2017). Social models in the cognitive appraisal of self-efficacy information. *Contemporary Educational Psychology, 48*, 149 - 166.

Andrich, D. (1978). A Rating Formulation for Ordered Response Categories. *Psycometrika, 43*(4), 561-573.

Andrich, D. (2010). Understanding the Response Structure and Process in the Polutomous Rasch Model. I M. L. Nering & R. Ostini (Red.), *Handbook of Polytomous Item Respons Theory Models*. New York: Routledge.

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review, 84*(2), 191-215.

Bandura, A. (1993). Perceived Self-Efficacy in Cognitive Development and Functioning. *Educational Psychologist, 28*(2), 117-148.

Bandura, A. (1997). *Self-Efficacy, The Exercise of Control*. United States of America: W.H. Freeman and Company.

Bandura, A. (2006). Guide for constructing self-efficacy scales. I F. Pajares & T. Urdan (Red.), *Self-Efficacy Beliefs of Adolescents* (s. 307-337). Information Age Publishing.

Bergsaker, A. S. (2019). Statistisk Analyse i SPSS. Hentet 11.03 2021 fra https://www.uio.no/for-ansatte/kompetanse/tema/data/it-forskning/spss/spss2019_oppf.pdf

Bjerke, A. H. & Eriksen, E. (2016). Measuring pre-service teachers' self-effiacy in tutoring children in primary mathematics: an instrument. *Research in mathematics education, 18*(1), 61-79.

Boaler, J. (2013). Ability and Mathematics: the mindset revolution that is reshaping education. *Forum, 55*(1), 143-151.

Bond, T. & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental Measurement in the Human Schiences, Third Edition*. New York: Routledge.

Boone, W. J. (2016). Rasch Analysis for Instrument Development: Why, When, and How? *CBE-Life Sciences Education, 15*(4). https://doi.org/10.1187/cbe.16-04-0148

Cavanagh, R. F. & Fischer, W. P. (2018). The science in human schience research: The case for Rasch measurement in learning environment research. *Journal of Physics: Conference Series, 1065*(7). Hentet fra https://iopscience.iop.org/article/10.1088/1742-6596/1065/7/072010/pdf

Chen, P. P. (2003). Exploring the accuracy and predictability of the self-efficacy beliefs of seventh-grade mathematics students. *Learning and Individual Differences, 14*, 79-92.

Christoffersen, L. & Johannessen, A. (2012). *Forskningsmetode for lærerutdanningene*. Oslo: Abstrakt forlag.

Cohen, L., Manion, L. & Morrison, K. (2007). *Research Methods in Education*. London and New York: Routledge.

Foldnes, N., Grønneberg, S. & Hermansen, G. H. (2018). *Statistikk og dataanalyse - En moderne innføring*. Oslo: Cappelen Damm Akademisk.

Galyon, C. E., Blondin, C. A., Yaw, J. S., Nalls, M. L. & Williams, R. L. (2012). The relationship of academic self-efficacy to class participation and exam performance. *Social Psychology of Education, 15*(2), 233 - 249.

Gao, J. (2019). Sources of Mathematics Self-Efficacy in Chinese Students: A Mixed-Method Study with Q-sorting Procedure. *International Journal of Science and Mathematics Education, 18*(4), 713-732.

Haciomeroglu, G. (2019). The Relationship between Elementary Students' Achievement Emotions and Sources of Mathematics Self-efficacy. *International Journal of Educational Research and Science, 5*(2), 12.

Hampton, N. Z. (1998). Sources of Academic Self-Efficacy Scale: An Assessment Tool for Rehabilitation Counselors. *Rehabilitation Counseling Bulletin, 41*(4).

Haavik, A.-B. (2015, 4.12.2020). TIMSS 2015, Frigitte oppgaver i matematikk for ungdomstrinnet. Hentet 10.09 2020 fra https://www.uv.uio.no/ils/forskning/prosjekter/timss/2015/frigitte-oppgaver/t15_g8_hefte_matte_211217.pdf

IBM. (2020). One-tail test for T-test of independent samples. Hentet 13.05 2021 fra https://www.ibm.com/support/pages/one-tail-test-t-test-independent-samples

Johannessen, A., Tufte, P. O. & Christoffersen, L. (2016). *Introduksjon til samfunnsvitenskapelig metode*. Oslo: Abstrakt forlag AS.

Kunnskapsdepartementet. (2006). *Læreplanverket for Kunnskapsløftet : grunnskolen* (978-82-8372-053-2). Oslo: Pedlex.

Kunnskapsdepartementet. (2011). *Fra matteskrekk til mattemestring*. Hentet fra https://www.regjeringen.no/globalassets/upload/kd/vedlegg/grunnskole/strategiplaner/matematikk_aug_2011.pdf?fbclid=IwAR1al8pWo8T7qT_g_lBgB7WbaTl8Zk4s-LTwZ51A3hVcrKLSmnCOaEQvJAA

Kunnskapsdepartementet. (2017). Verdier og prinsipper for grunnopplæringen - overordnet del av læreplanverket. Hentet 14.05. 2021 fra https://www.regjeringen.no/no/dokumenter/verdier-og-prinsipper-for-grunnopplaringen/id2570003/

Kunnskapsdepartementet. (2019). *Læreplan i matematikk 1. - 10. trinn (MAT01-05). Fastsatt som forskrift. Læreplanverket for Kunnskapsløftet 2020.* Hentet fra https://data.udir.no/kl06/v201906/laereplaner-lk20/MAT01-05.pdf?lang=nob

Lampert, M. (1990). When the Problem Is Not the Question and the Solution Is Not the Answer: Mathematical Knowing and Teaching. *American Educational Research Association, 27*(1), 29-63.

Linacre, J. M. (1994). Sample Size and Item Calibration Stability. Hentet 02.03 2021 fra https://www.rasch.org/rmt/rmt74m.htm

Linacre, J. M. (2000). Comparing and Choosing between "Partial Credit Models" (PCM) and "Rating Scale Models" (RSM). Hentet 16.03 2021 fra https://www.rasch.org/rmt/rmt143k.htm

Linacre, J. M. (2002). Differential Item Functioning and Differential Test Functioning (DIF & DTF). Hentet 15.03 2021 fra https://www.rasch.org/rmt/rmt163g.htm

Linacre, J. M. (2012a). Winsteps Rasch Tutorial 2. Hentet 22.02 2021 fra https://www.winsteps.com/a/winsteps-tutorial-2.pdf

Linacre, J. M. (2012b). Winsteps Rasch Tutorial 3. Hentet 22.02 2021 fra https://www.winsteps.com/a/winsteps-tutorial-3.pdf

Linacre, J. M. (2012c). Winsteps Rasch Tutorial 4. Hentet 22.01 2021 fra https://www.winsteps.com/a/winsteps-tutorial-4.pdf

Linacre, J. M. (2014). Winsteps (Versjon 3.81.0). Beaverton, Oregon: winsteps.com.

Linacre, J. M. (u.å). Reliability and separation of measures. Hentet 09.05 2021 fra https://www.winsteps.com/winman/reliability.htm

Linacre, J. M. (u.å.). Fit diagnosis: infit outfit mean-square standardized. Hentet 21.01 2021 fra https://www.winsteps.com/winman/misfitdiagnosis.htm?fbclid=IwAR2dk-AzZbn1GXUateq7dGaceohfZjJ4vdapJCD2HQXmZqDxOitx4kP6MTo

Lopez, F. G. & Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *Career Development Quarterly, 41*, 10.

Lopez, F. G., Lent, R. W., Brown, S. D. & Gore, P. A. (1997). Role of Social-Cognitive Expectations in High School Students' Mathematics-Related Interest and Performance. *Journal of Counseling Psychology, 44*(1), 44-52.

Lærd Statisics. Sphericity. Hentet 13.05. 2021 fra https://statistics.laerd.com/statistical-guides/sphericity-statistical-guide.php

Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifactors? *Journal of Personality and Social Psychology, 70*(4), 810-819. https://doi.org/10.1037/0022-3514.70.4.810

Mozahem, N. A., Boulad, F. M. & Ghanem, C. M. (2020). Secondary school students and self-efficacy in mathematics: Gender and age differences. *International Journal of Educational Research*.

Nardi, P. M. (2018). *Doing Survey Research : A Guide to Quantitative Methods*. Milton: Routledge.

NESH. (2016). Forskningsetiske retningslinjer for samfunnsvitenskap, humaniora, juss og teologi. Hentet 04.03 2021 fra https://www.forskningsetikk.no/retningslinjer/hum-sam/forskningsetiske-retningslinjer-for-samfunnsvitenskap-humaniora-juss-og-teologi/

NOU 2016:14. (2016). *Mer å hente - Bedre læring for elever med stort læringspotensial*. Hentet fra https://www.regjeringen.no/no/dokumenter/nou-2016-14/id2511246/sec6

Opplæringslova. (1998). *Lov om grunnskolen og den vidaregåande opplæringa (opplæringslova)*. Hentet fra https://lovdata.no/dokument/NL/lov/1998-07-17-61

Pajares, F. (2006). Self-Efficacy during Childhood and Adolescence - Implications for Teachers and Parents. I F. U. Pajares, Tim (Red.), *Self-Efficacy Beliefs of Adolescents*. Greenwich, CT: Information Age.

Pajares, F. & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology, 24*.

Pampaka, M., Williams, J., Hutcheson, G., Wake, G., Black, L., Davis, P. & Hernadez-Martinez, P. (2012). The association between mathematics pedagogy and learners' dispositions for university study. *British Educational Research Journal, 38*(3). Hentet fra https://www.jstor.org/stable/23211458

Phan, H. P. (2012). Relations between informational sources, self-efficacy and academic achievement: a developmental approach. *Educational Psychology, 32*(1), 81-105.

Postholm, M. B. & Jacobsen, D. I. (2011). *Læreren med forskerblikk*. Kristiansand: Høyskoleforlaget AS - Norwegian Academic Press.

Rentzsch, K., Schütz, A. & Schröder-Abé, M. (2011). Being Labeled *Nerd*: Factors That Influence the Social Acceptance of High-Achieving Students. *The Journal of Experimental Education, 79*, 143-168.

Salazar, L. R. & Hayward, S. L. (2018). An Examination of College Students' Problem-Solving Self-Efficacy, Academic Self-Efficacy, Motivation, Test Performance, and Expected Grade in Introductory-Level Economics Courses. *Decision Sciences Journal of Innovative Education, 16*(3), 217 - 240.

Skagen, K. (1996). *Karakterboka. Om karakterer og vurdering i ny skole.* Oslo: Universitetsforlaget AS.

Skemp, R. R. (1976). Relational Understanding and Instrumental Understanding. *Mathematics Teaching in the Middle School, 12*(2), 8.

Skott, J., Skott, C. K., Jess, K. & Hansen, H. C. (2018). *Matematikk for lærerstuderende, Delta 2.0, Fagdidaktikk 1.-10. klasse*. København: Samfundslitteratur.

Skovsmose, O. (1998). *Undersøgelseslandskaber*. Bergen: Landslaget for matematikk i skolen (LAMIS): Rapport for Lamis 1.sommerkurs 1998.

Skaalvik, E. M., Federici, R. A. & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research, 72*, 129-136.

Skaalvik, E. M. & Skaalvik, S. (2011). Self-Concept and Self-Efficacy in Mathematics: Relation with Mathematics Motivation and Achievement. *Journal of Educational Research, 5*(3/4).

StatsTest. (u.å.). One-Way Repeated Measures ANOVA. Hentet 13.05 2021 fra https://www.statstest.com/one-way-repeated-measures-anova/

Stevens, T., Olivarez Jr., A., Lan, W. Y. & Tallent-Runnels, M. K. (2004). Role of Mathematis Self-Efficacy and Motivation in Mathematics Performance Across Ethnicity. *The Journal of Educational Research, 97*(4), 208-222.

Stigler, J. W. & Hiebert, J. (2009). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York: Free Press.

Stodolsky, S. S. (1985). Telling Math: Origins of Math Aversion and anxiety. *Educational Psychologist, 20*(3), 9.

Street, K. E. S., Malmberg, L.-E. & Stylianides, G. J. (2017). Level, strength, and facet-specific self-efficacy in mathematics test performance. *ZDM Mathematics Education, 49*, 379-395.
Stuart, Y. S. (2017). *Solving the Achievement Gap. Overcoming the Structure of School Inequality.* Minneapolis: Palgrave Macmillan US.
Thrane, C. (2018). *Kvantitativ metode. En praktisk tilnærming.* Oslo: Cappelen Damm AS.
Toland, M. D. & Usher, E. L. (2016). Assessing Mathematics Self-Efficacy: How Many Categories Do We Really Need? *The Journal of early adolescence, 36*(7), 932-960. https://doi.org/10.1177/0272431615588952
Tryggestad, H. & Eldevik, S. (2015). Opplæring for elever med matematikkvansker. Hentet 06.01 2021 fra https://utdanningsforskning.no/artikler/opplaring-for-elever-med-matematikkvansker/
Tufte, P. O. (2011). Kvantitativ metode. I K. Fangen & A.-M. Sellerberg (Red.), *Mange ulike metoder*. Oslo: Gyldendal Norsk Forlag AS.
Usher, E. L. (2009). Sources of Middle School Students' Self-Efficacy in Mathematics: A Qualitative Investigation. *American Educational Research Journal, 46*(1), 275-314.
Usher, E. L. & Pajares, F. (2008). Sources of Self-Efficacy in School: Critical Review of the Literature and Future Directions. *Review of Educational Research, 78*(4), 751-796.
Usher, E. L. & Pajares, F. (2009). Sources of self-efficacy in mathematics: A validation study. *Contemporary Educational Psychology, 34*, 89-101.
Utdanningsdirektoratet. (2016). Læringsmiljø og relasjoner mellom elever. Hentet 10.05 2021 fra https://www.udir.no/laring-og-trivsel/skolemiljo/sosial-laring-gjennom-arbeid-med-fag/Relasjoner-mellom-elever/Laringsmiljo-og-relasjoner-mellom-elever/
Utdanningsdirektoratet. (2019, 29.05.2019). Mål 2 - færre barn og unge på lavt nivå i realfag. Hentet 04.01 2021 fra https://www.udir.no/tall-og-forskning/publikasjoner/realfagsbarometeret/mal-2/
Utdanningsdirektoratet. (2020). TIMSS. Hentet 04.01 2021 fra https://www.udir.no/tall-og-forskning/internasjonale-studier/timss/#157829
van den Berg, R. G. (u.å.). Effect Size - A Quick Guide. Hentet 15.03 2021 fra https://www.spss-tutorials.com/effect-size/
Williams, J., Black, L., Davis, P., Hernadez-Martinez, P., Hutcheson, G., Nicholson, S., … Wake, G. (2010). Pedagogies for social diversity and difference. Keeping open the door to mathematically demanding programmes in further and higher education: a cultural model of value. I M. David (Red.), *Improving Learning by Widening Participation in Higher Education*. London and New York: Routledge.
Wolfe, E. W. & Smith, E. V. J. (2007). Instrument development tools and activities for measure validation using Rasch models: part 1 - instrument development tools. *Journal of Applied Measurement, 8*(1), 97-123.
Wæge, K. & Nosrati, M. (2018). *Motivasjon i matematikk*. Oslo: Universitetsforlaget.
Yackel, E. & Cobb, P. (1996). Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. *Journal for Research in Mathematics Education, 27*(4), 458-477.
Yildiz, P. & Özdemir, E. Y. (2019). Mathematics self-efficacy beliefs and sources of self-efficacy: A Descriptiv Study with two Elementary School Students. *International Journal of Progressive Education, 15*(3), 194-206.
Yurt, E. (2014). The Predictive Power of Self-Efficacy Sources for Mathematics Achievement. *Education and Science, 39*(176), 159-169.
Zimmerman, B. J. (1995). Self-efficacy and educational development. I A. Bandura (Red.), *Self-efficacy in changing societies*. California: Cambridge University Press.

Vedlegg 1: Godkjennelse fra NSD

NSD NORSK SENTER FOR FORSKNINGSDATA

NSD sin vurdering

Prosjekttittel

SESAM - Self-efficacy, it's sources and achievement in mathematics

Referansenummer

152185

Registrert

22.09.2020 av Marie Larsen [redacted]

Behandlingsansvarlig institusjon

OsloMet – storbyuniversitetet / Fakultet for lærerutdanning og internasjonale studier / Institutt for grunnskole- og faglærerutdanning

Prosjektansvarlig (vitenskapelig ansatt/veileder eller stipendiat)

Annette Hessen Bjerke,[redacted]

Type prosjekt

Studentprosjekt, masterstudium

Kontaktinformasjon, student

Marie Larsen,[redacted]

Prosjektperiode

24.08.2020 - 15.06.2021

Status

30.09.2020 - Vurdert anonym

Vurdering (1)

30.09.2020 - Vurdert anonym

Det er vår vurdering at det ikke skal behandles direkte eller indirekte opplysninger som kan identifisere enkeltpersoner i dette prosjektet, så fremt den gjennomføres i tråd med det som er dokumentert i meldeskjemaet den 30.09.2020 med vedlegg, samt i meldingsdialogen mellom innmelder og NSD.

Prosjektet trenger derfor ikke en vurdering fra NSD.

HVA MÅ DU GJØRE DERSOM DU LIKEVEL SKAL BEHANDLE PERSONOPPLYSNINGER?
Dersom prosjektopplegget endres og det likevel blir aktuelt å behandle personopplysninger må du melde dette til NSD ved å oppdatere meldeskjemaet. Vent på svar før du setter i gang med behandlingen av

 personopplysninger.

VI AVSLUTTER OPPFØLGING AV PROSJEKTET
Siden prosjektet ikke behandler personopplysninger avslutter vi all videre oppfølging.

Lykke til med prosjektet!

Kontaktperson hos NSD: Kajsa Amundsen
Tlf. Personverntjenester: 55 58 21 17 (tast 1)

Vedlegg 2: Spørreundersøkelse til hoveddatainnsamling

Kjønn	☐ Jente	☐ Gutt
Fødselsår	☐ 2006	☐ Annet
Morsmål	☐ Norsk	☐ Annet

Fødselsmåned

☐ Januar	☐ Februar	☐ Mars	☐ April
☐ Mai	☐ Juni	☐ Juli	☐ August
☐ September	☐ Oktober	☐ November	☐ Desember

DEL 1

Hvor trygg er du på om du greier å løse ulike matteoppgaver?

Svar på en skala fra 1-6, hvor 1 betyr «*veldig utrygg*» og 6 betyr «*veldig trygg*».

	1	**2**	**3**	**4**	**5**	**6**
Oppgave 1						
Oppgave 2						
Oppgave 3						
Oppgave 4						
Oppgave 5						
Oppgave 6						
Oppgave 7						
Oppgave 8						
Oppgave 9						
Oppgave 10						
Oppgave 11						
Oppgave 12						

DEL 2

Hvor godt stemmer påstandene for deg? Gi svar på en skala fra 1 – 6.

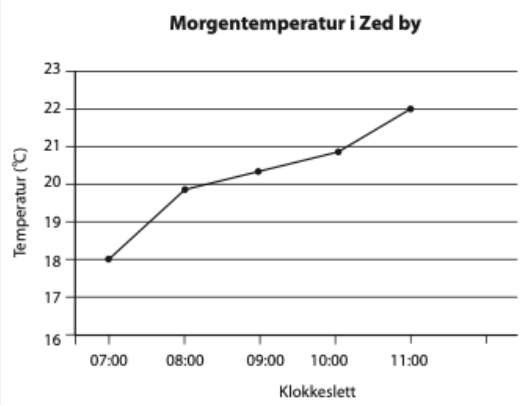
1 betyr «*stemmer ikke i det hele tatt*» og 6 betyr «*stemmer veldig godt*».

	1	2	3	4	5	6
1. Jeg får gode karakterer på matteprøver.						
2. Jeg har alltid lykkes i matte.						
3. Selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte.						
4. Jeg fikk god karakter i matte til sommeren i 8.klasse.						
5. Jeg er flink til å løse ulike matteoppgaver.						
6. Jeg får til å løse selv de vanskeligste oppgavene i matte.						
7. Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre.						
8. Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
9. Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre.						
10. Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
11. Jeg kan se for meg at jeg jobber med vanskelige matteoppgaver og lykkes med å løse dem.						
12. Jeg konkurrerer med meg selv i matte.						
13. Mattelærerne mine har fortalt meg at jeg er flink i matte.						
14. Andre har fortalt meg at jeg er veldig flink i matte.						
15. Voksne i familien min har fortalt meg at jeg er flink i matte.						
16. Jeg har fått mye skryt for mine evner i matte.						
17. Medelevene mine har fortalt meg at jeg er flink i matte.						
18. Klassekameratene mine liker å jobbe med meg i matte, fordi de syns jeg er god.						
19. Jeg føler meg stresset og nervøs bare ved å være i mattetimene.						
20. Å jobbe med matte tapper meg for energi.						
21. Jeg blir stresset med en gang jeg begynner å jobbe med matte.						
22. Jeg får jernteppe og klarer ikke å tenke klart når jeg gjør matteoppgaver						
23. Jeg blir deppa av tanken på å gjøre matte.						
24. Jeg blir anspent i hele kroppen når jeg må jobbe med matte.						

DEL 3

Regn ut oppgavene nedenfor.

Oppgave 1



Grafen viser temperaturen hver time fra klokka 07:00 til klokka 11:00.
Anslå temperaturen klokka 09.30.

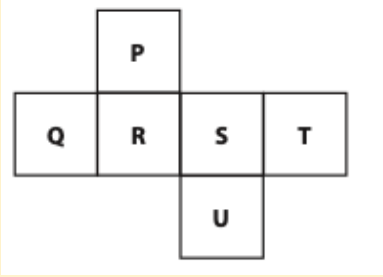
Svar: _____°C

Oppgave 2

Hva er størst, $\frac{7}{12}$ eller $\frac{2}{3}$?

Svar: _____

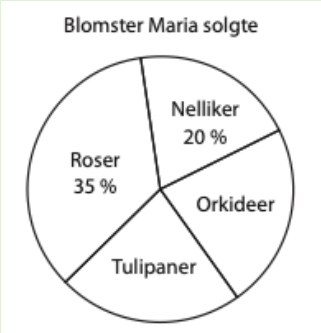
Oppgave 3



Lisa har bygd en terning ved å brette figuren over. Hvilken sideflate står på motsatt side av Q?

Svar: _____

Oppgave 4



Maria solgte 4 typer blomster. Hun solgte like mange tulipaner og orkideer.

Hvor mange prosent av blomstene som ble solgt, var tulipaner?

Svar: _____%

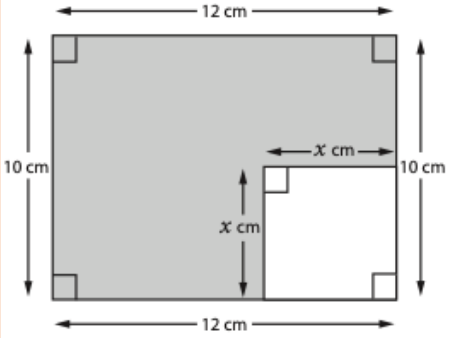
Oppgave 5

Skriv dette som et desimaltall.

$$8 + 50 + \frac{3}{100} + \frac{1}{10}$$

Svar:_____

Oppgave 6



Skriv et uttrykk (ved hjelp av x) for arealet av den **gråfargede** delen av figuren.

Svar: _____cm^2

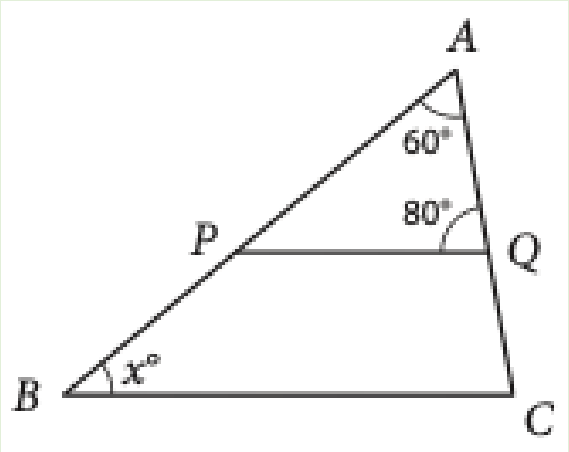
Oppgave 7

En maskin har 7 forskjellige farger på tyggegummiballer. Liv la merke til at folk kjøpte 306 tyggegummiballer, og at 23 av de var blå.

Hva er sannsynligheten for at den neste tyggegummiballen vil være blå?

Svar: _____

Oppgave 8



Linjene PQ og BC er parallelle.
Hvor stor er x?

Svar: _____

Oppgave 9

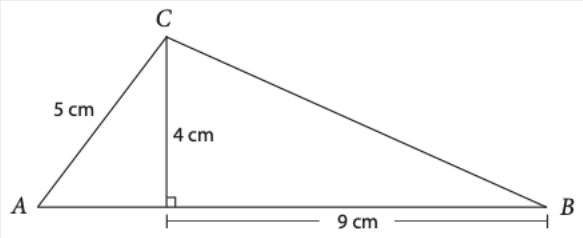
Finn verdiene av x og y slik at begge likningene er sanne.

$$3x + y = 13$$

$$5x - y = 27$$

Svar: x = _____ y =_____

Oppgave 10



Hva er arealet til trekanten ABC?

Svar: _____cm^2

Oppgave 11

Sektordiagrammet viser hvilken type TV-program 240 elever likte best. Hvor mange elever likte programmer om historie best?

Svar: _____

Oppgave 12

a = 5 og b = 2

Hva er verdien av $a^2b - 3(a - b)$?

Svar:_____

Vedlegg 3: Retningslinjer for gjennomføringen

Presentere hvem vi er, og hvorfor vi er kommet til deres klasse.

Vi presenterer undersøkelsen:

- Undersøkelsen tar ca. en time. Vi setter pris på at du svarer så godt du kan. Påpeke at vi skal ikke teste hvor mye matematikk elevene kan, så ingen trenger å få nerver.
- Undersøkelsen er helt anonym, så ikke skriv navnet ditt på noen av arkene du får utdelt. Det er frivillig å delta, og du kan trekke deg ved å si ifra til meg eller ved å ikke levere inn undersøkelsesheftet.
- Heftet består av tre deler. Vi går gjennom de ulike delene, slik at de vet hva som kommer. Si at dette er et individuelt arbeid, så de kan ikke samarbeide med andre.

Vi deler ut heftene, ber elevene finne frem en blyant og fylle ut forsiden.

Del 1

- Vi tar opp presentasjonen på tavla og går igjennom et eksempel. Her presiserer vi at elevene ikke må krysse mellom rutene, men at de må velge en av rutene. Gå gjennom eksempel, presiser at dette er samme oppsett som resten av DEL 1. Presiser at elevene ikke skal regne ut, men krysse av for trygg de er på at de kan løse oppgaven. Hver oppgave blir lest høyt før vi går rett til hvitt ark og elevene får tid til å krysse. Når vi er ferdig med del 1 presiseres det at det ikke er lov til å bla tilbake i heftet for å endre svarene sine.

Del 2 og del 3

- Vi ber elevene bla om og informerer elevene om at de skal sette ett kryss pr. påstand. Her er det også viktig at elevene ikke krysser mellom rutene. Må også informere om at det her er en "ny skala" der 1 er "stemmer ikke i det hele tatt" og 6 er "stemmer veldig godt".
  Gi beskjed om at elevene kan fortsette på del 3 når de er ferdig, og begynne å regne oppgavene. Informere om at de får et kladdeark, der de kan gjøre utregninger hvis de trenger, og fylle inn svaret i heftet. Rekke opp hånda når man er ferdig med å svare på alt.

Spørsmål vi fikk i pilot-undersøkelsen

- "Hva regnes som gode karakterer?" - vi må presisere at det er elevenes egen mening som gjelder.
- Hva er morsmål - det språket som blir snakket hjemme
- Spørsmål om hjelp til utregning – vi kan ikke hjelpe elevene.

<u>Vi må sjekke at forsiden er utfylt når vi samler inn undersøkelsen.</u>

Vedlegg 4: Spørreundersøkelse til pilot

Hei!

Denne undersøkelsen gjennomføres av forskningsgruppa SESAM fra OsloMet. Vi er tre masterstudenter, som sammen med vår veileder, samler inn data på 9. trinn i Norge.

Vi setter stor pris på at du svarer så godt du kan. Det er frivillig å delta i undersøkelsen. Svaret ditt er anonymt, så du skal ikke skrive navnet ditt på arket. Det er kun forskere i SESAM som skal se besvarelsen din.

Undersøkelsen består av tre deler – noe avkrysning og noe regning.

Kryss av for det alternativet som stemmer for deg:

Kjønn

☐ Jente ☐ Gutt

Fødselsår

☐ 2006 ☐ Annet

Fødselsmåned

☐ Januar ☐ Februar ☐ Mars ☐ April
☐ Mai ☐ Juni ☐ Juli ☐ August
☐ September ☐ Oktober ☐ November ☐ Desember

Morsmål

☐ Norsk ☐ Annet

DEL 1

Hvor trygg er du på om du greier å løse ulike matteoppgaver?

Svar på en skala fra 1-6, hvor 1 betyr «*veldig utrygg*» og 6 betyr «*veldig trygg*».

Oppgavene blir presentert på tavla og lest høyt.

	Veldig utrygg					*Veldig trygg*
	1	**2**	**3**	**4**	**5**	**6**
Oppgave 1						
Oppgave 2						
Oppgave 3						
Oppgave 4						
Oppgave 5						
Oppgave 6						
Oppgave 7						
Oppgave 8						
Oppgave 9						
Oppgave 10						
Oppgave 11						
Oppgave 12						
Oppgave 13						
Oppgave 14						
Oppgave 15						
Oppgave 16						

DEL 2

Hvor godt stemmer påstandene for deg? Gi svar på en skala fra 1-6,
der 1 betyr «*stemmer ikke i det hele tatt*» og 6 betyr «*stemmer veldig godt*».

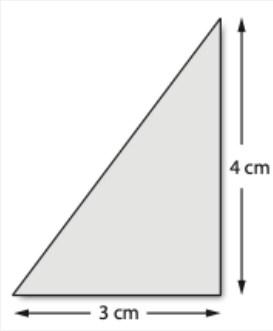
	1	2	3	4	5	6
Jeg får gode karakterer på matteprøver						
Jeg har alltid lykkes i matte						
Selv når jeg jobber veldig hardt, gjør jeg det dårlig i matte						
Jeg fikk god karakter i matte til sommeren i 8.klasse						
Jeg er flink til å løse ulike matteoppgaver						
Jeg får til å løse selv de vanskeligste oppgavene i matte						
Å se voksne gjøre det bra i matte, pusher meg til å gjøre det bedre.						
Når jeg ser mattelæreren min løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
Å se andre elever gjøre det bedre enn meg i matematikk, pusher meg til å gjøre det bedre.						
Når jeg ser en medelev løse en oppgave, kan jeg se for meg at jeg selv kan løse en slik oppgave på samme måte.						
Jeg kan se for meg at jeg jobber med vanskelige matteoppgaver og lykkes med å løse dem.						
Jeg konkurrerer med meg selv i matte.						
Mattelærerne mine har fortalt meg at jeg er flink i matte.						
Andre har fortalt meg at jeg er veldig flink i matte.						
Voksne i familien min har fortalt meg at jeg er flink i matte.						
Jeg har fått mye skryt for mine evner i matte.						
Medelevene mine har fortalt meg at jeg er flink i matte.						
Klassekameratene mine liker å jobbe med meg i matte, fordi de syns jeg er god.						
Jeg føler meg stresset og nervøs bare ved å være i mattetimene.						
Å jobbe med matte tapper meg for energi.						
Jeg blir stresset med en gang jeg begynner å jobbe med matte.						
Jeg får jernteppe og klarer ikke å tenke klart når jeg gjør matteoppgaver						
Jeg blir deppa av tanken på å gjøre matte.						
Jeg blir anspent i hele kroppen når jeg må jobbe med matte.						

DEL 3

Regn ut oppgavene nedenfor.

Oppgave 1

Andreas har et rektangulært papir.
Han deler det i to og står igjen med denne papirtrekanten.



Beregn lengden av den tredje siden av trekanten.

Svar: _____cm

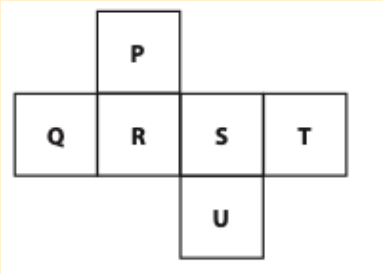
Oppgave 2

$a = 1 + x$ og $b = 1 - x$

Hva er $a - b$?

Svar: _____

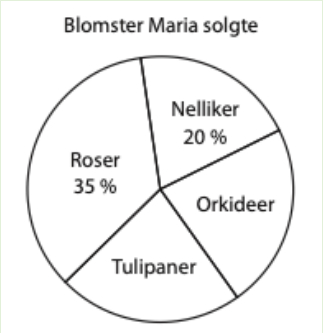
Oppgave 3



Lisa har bygd en terning ved å brette figuren over. Hvilken sideflate står på motsatt side av Q?

Svar: _____

Oppgave 4



Maria solgte 4 typer blomster. Hun solgte like mange tulipaner og orkideer.

Hvor mange prosent av blomstene som ble solgt, var tulipaner?

Svar: _____%

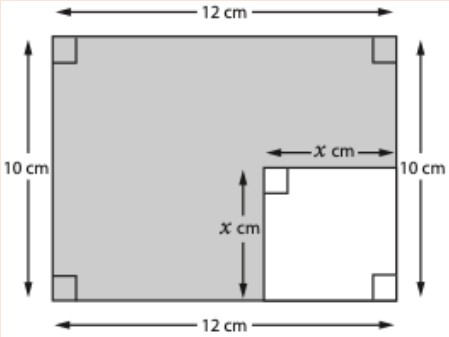
Oppgave 5

Skriv dette som et desimaltall.

$$8 + 50 + \frac{3}{100} + \frac{1}{10}$$

Svar:_____

Oppgave 6



Skriv et uttrykk (ved hjel av x) for arealet av den **gråfargede** delen av figuren.

Svar: _____cm^2

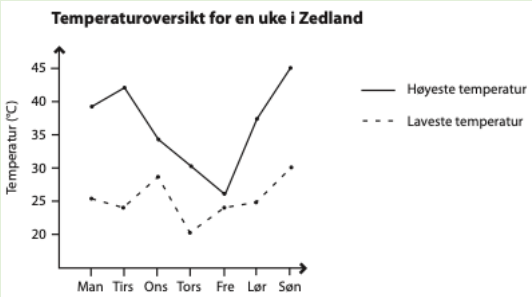
Oppgave 7

En maskin har 7 forskjellige farger på tyggegummiballer. Liv la merke til at folk kjøpte 306 tyggegummiballer, og at 23 av de var blå.

Hva er sannsynligheten for at den neste tyggegummiballen vil være blå?

Svar: _____

Oppgave 8



Grafene over viser høyeste og laveste temperatur hver dag i løpet av en uke i Zedland. Hvilken dag var forskjellen mellom den høyeste og den laveste temperaturen 10˚C?

Svar: _____

Oppgave 9

Finn verdiene av x og y slik at begge likningene er sanne.

$$3x + y = 13$$

$$5x - y = 27$$

Svar: x= _____ y=_____

Oppgave 10

Du har en kakeoppskrift som krever 2 egg og 0,3 liter melk. Du vil lage en så stor kake som mulig, og du har 5 egg.
Hvor mange liter melk trenger du da?

Svar: _____liter

Oppgave 11

Sektordiagrammet viser hvilken type TV-program 240 elever likte best. Hvor mange elever likte programmer om historie best?

Svar: _____

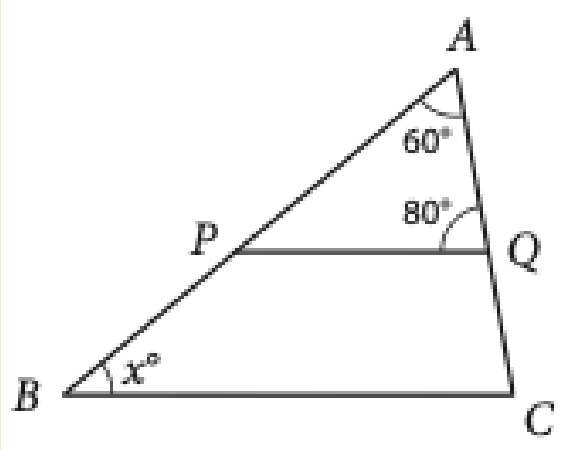
Oppgave 12

a = 5 og b = 2

Hva er verdien av $a^2b - 3(a - b)$

Svar:_____

Oppgave 13



Linjene PQ og BC er parallelle.
Hvor stor er x?

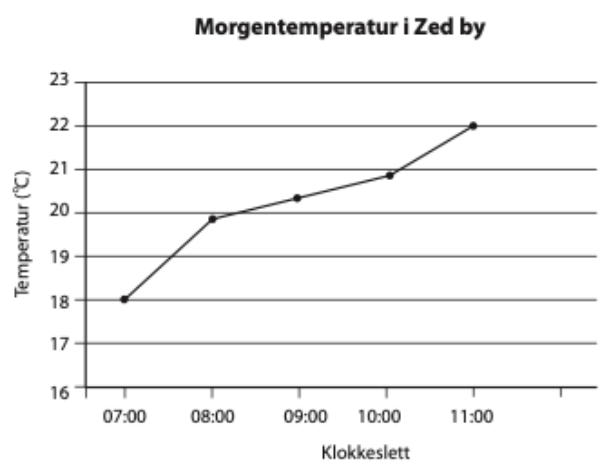
Svar: _____

Oppgave 14

Hva er størst, $\frac{7}{12}$ eller $\frac{2}{3}$?

Svar: _____

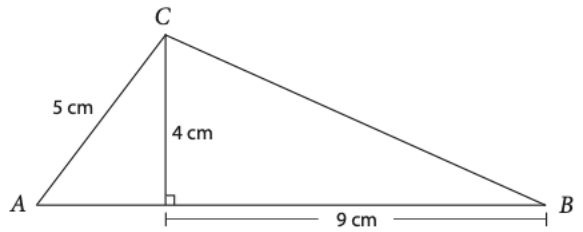
Oppgave 15



Grafen viser temperaturen hver time fra klokka 07:00 til klokka 11:00.
Anslå temperaturen klokka 09.30.

Svar: _____°C

Oppgave 16



Hva er arealet til trekanten ABC?

Svar: _____

Vedlegg 5: Sources of Middle School Mathematics Self Efficacy Scale

1. I make excellent grades on math tests (ME-1)[2]
2. I have always been successful with math (ME-3)[2]
3. Even when I study very hard, I do poorly in math (ME-6)[*1]
4. I got good grades in math on my last report card (ME-8)[1]
5. I do well on math assignments (ME-9)[1M]
6. I do well on even the most difficult math assignments (ME-12)[3]
7. Seeing adults do well in math pushes me to do better (VA-4)[2]
8. When I see how my math teacher solves a problem, I can picture myself solving the problem in the same way (VA-6)[2]
9. Seeing kids do better than me in math pushes me to do better (VP-1)[1]
10. When I see how another student solves a math problem, I can see myself solving the problem in the same way (VP-9)[2]
11. I imagine myself working through challenging math problems successfully (VS-4)[1]
12. I compete with myself in math (VS-5)[3]
13. My math teachers have told that I am good at learning math (P-4)[1M]
14. People have told me that I have a talent for math (P-5)[3]
15. Adults in my family have told me what a good math student I am (P-7)[2]
16. I have been praised for my ability in math (P-13)[1M]
17. Other students have told me that I'm good at learning math (P-14)2M
18. My classmates like to work with me in math because they think I'm good at it (P-16)[1M]
19. Just being in math class makes feel stressed and nervous (PH-2)[*1M]
20. Doing math work takes all of my energy (PH-3)[*2]
21. I start to feel stressed-out as soon as I begin my math work (PH-5)[*1]
22. My mind goes blank and I am unable to think clearly when doing math work (PH-7)[*1]
23. I get depressed when I think about learning math (PH-9)[*1M]
24. My whole body becomes tense when I have to do math (PH-12)[*1]

Hentet fra (Usher & Pajares, 2009, s. 98)

Vedlegg 6: Svarkategorienes rekkefølge for delinstrumentene

Svarkategorier ME-instrument:

TABLE 26.3 Kilder ZOU461WS.TXT Feb 9 2021 16:17
INPUT: 583 Person 24 Item REPORTED: 583 Person 6 Item 6 CATS WINSTEPS 3.91.2

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: CORRELATION ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item	
3	5	0	32	6	-2.11	1.81	.33	2.8	3.6	-.32	03 Item3ME	5 VT Veldig trygg
	4	1	43	8	-1.28	1.04	.16	2.4	2.5	-.26		4
	3	2	65	11	-.91	1.06	.13	1.6	1.5	-.26		3
	2	3	89	16	-.60	.92	.10	.9	.7	-.25		2
	1	4	146	26	.72	1.30	.11	1.5	1.2	.04		1
	0	5	193	34	2.38	1.80	.13	1.8	1.7	.62		0 VU Veldig utrygg
	MISSING ***		15	3#	.75	1.55	.41			.01		
4	0	0	35	6	-2.63	1.41	.24	1.3	1.4	-.40	04 Item4ME	0 VU Veldig utrygg
	1	1	68	12	-1.52	.96	.12	1.0	1.0	-.37		1
	2	2	110	19	-.69	.73	.07	.7	.6	-.30		2
	3	3	124	21	.38	.89	.08	.7	.8	-.05		3
	4	4	120	21	1.51	1.00	.09	.8	.7	.23		4
	5	5	123	21	3.11	1.61	.15	1.4	1.3	.64		5 VT Veldig trygg
	MISSING ***		3	1#	-.26	2.79	1.97			-.03		
5	0	0	32	6	-2.81	1.11	.20	.7	.7	-.40	05 Item5ME	0 VU Veldig utrygg
	1	1	53	9	-1.60	.90	.12	.8	.8	-.34		1
	2	2	115	20	-.78	.87	.08	.8	.8	-.33		2
	3	3	173	30	.42	.92	.07	.9	.8	-.05		3
	4	4	140	24	1.84	1.15	.10	.9	.9	.35		4
	5	5	67	12	3.99	1.33	.16	.8	.7	.60		5 VT Veldig trygg
	MISSING ***		3	1#	1.98	2.46	1.74			.05		
6	0	0	92	16	-2.13	1.06	.11	.7	.7	-.57	06 Item6ME	0 VU Veldig utrygg
	1	1	96	17	-.73	.84	.09	.8	.8	-.28		1
	2	2	116	20	.14	.79	.07	.7	.6	-.11		2
	3	3	135	23	1.16	1.03	.09	.9	.9	.16		3
	4	4	105	18	2.78	1.27	.12	1.0	1.1	.51		4
	5	5	34	6	3.90	1.75	.31	1.7	1.8	.40		5 VT Veldig trygg
	MISSING ***		5	1#	1.67	2.42	1.21			.05		
2	0	0	73	13	-2.11	1.36	.16	1.4	1.3	-.50	02 Item2ME	0 VU Veldig utrygg
	1	1	82	14	-1.03	.97	.11	1.0	1.0	-.32		1
	2	2	125	22	-.09	.83	.07	.7	.7	-.18		2
	3	3	136	24	.82	.89	.08	.8	.7	.06		3
	4	4	91	16	2.31	.90	.09	.6	.6	.36		4
	5	5	66	12	4.09	1.22	.15	.7	.8	.61		5 VT Veldig trygg
	MISSING ***		10	2#	-.14	1.23	.41			-.05		
1	0	0	33	6	-2.64	1.20	.21	.9	.9	-.38	01 Item1ME	0 VU Veldig utrygg
	1	1	51	9	-1.92	.78	.11	.4	.5	-.38		1
	2	2	121	21	-.81	.70	.06	.5	.5	-.35		2
	3	3	153	26	.34	.71	.06	.5	.4	-.07		3
	4	4	131	23	1.88	.82	.07	.4	.4	.34		4
	5	5	89	15	3.63	1.50	.16	1.2	1.0	.63		5 VT Veldig trygg
	MISSING ***		5	1#	.08	2.46	1.23			-.02		

Svarkategorier VE-instrument

TABLE 26.3 Kilder ZOU681WS.TXT Feb 9 2021 17:18
INPUT: 583 Person 24 Item REPORTED: 583 Person 6 Item 6 CATS WINSTEPS 3.91.2

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: CORRELATION ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item	
7	0	0	140	24	-1.50	1.67	.14	1.7	1.6	-.52	07 Item7VEv	0 VU Veldig utrykk
	1	1	137	24	-.48	.86	.07	1.0	1.1	-.14		1
	2	2	126	22	.09	.75	.07	1.2	1.1	.07		2
	3	3	96	17	.69	.83	.08	1.1	1.5	.23		3
	4	4	55	10	1.42	.91	.12	1.0	1.0	.32		4
	5	5	24	4	2.48	1.04	.22	.9	.9	.35		5 VT Veldig trygg
	MISSING	***	5	1#	-.31	.71	.35			-.01		
12	0	0	83	14	-2.27	1.50	.17	1.2	1.3	-.58	12 Item12VEs	0 VU Veldig utrykk
	1	1	72	12	-1.07	1.00	.12	1.1	1.2	-.24		1
	2	2	105	18	-.31	.67	.07	1.0	.9	-.06		2
	3	3	104	18	.13	.68	.07	.9	.9	.07		3
	4	4	106	18	.69	.84	.08	1.1	1.1	.25		4
	5	5	110	19	1.33	1.19	.11	1.4	1.5	.46		5 VT Veldig trygg
	MISSING	***	3	1#	1.03	.17	.12			.05		
9	0	0	105	18	-1.94	1.64	.16	1.6	1.6	-.57	09 Item9VEm	0 VU Veldig utrykk
	1	1	118	20	-.74	.81	.07	1.0	1.0	-.21		1
	2	2	94	16	-.16	.68	.07	.9	.9	-.01		2
	3	3	95	16	.21	.57	.06	.8	.8	.09		3
	4	4	88	15	.87	.66	.07	.8	.8	.27		4
	5	5	83	14	1.77	1.03	.11	1.0	1.1	.50		5 VT Veldig trygg
11	0	0	71	12	-2.61	1.45	.17	1.0	1.0	-.61	11 Item11VEs	0 VU Veldig utrykk
	1	1	77	13	-1.05	.75	.09	.7	.7	-.24		1
	2	2	117	20	-.29	.77	.07	1.1	1.2	-.06		2
	3	3	130	22	.22	.66	.06	.8	.8	.11		3
	4	4	115	20	.83	.89	.08	1.0	1.1	.31		4
	5	5	71	12	1.61	1.16	.14	1.3	1.2	.42		5 VT Veldig trygg
	MISSING	***	2	0#	-.08	.19	.19			.00		
8	0	0	71	12	-2.71	1.34	.16	.7	.8	-.64	08 Item8VEv	0 VU Veldig utrykk
	1	1	71	12	-1.06	.77	.09	.8	.8	-.23		1
	2	2	103	18	-.42	.62	.06	.7	.7	-.10		2
	3	3	140	24	.16	.63	.05	.8	.8	.10		3
	4	4	112	19	.71	.71	.07	.8	.8	.26		4
	5	5	85	15	1.72	1.09	.12	1.0	.9	.50		5 VT Veldig trygg
	MISSING	***	1	0#	1.04	.00				.03		
10	0	0	59	10	-2.87	1.41	.18	.8	.8	-.61	10 Item10VEm	0 VU Veldig utrykk
	1	1	80	14	-1.18	.78	.09	.7	.7	-.28		1
	2	2	111	19	-.45	.63	.06	.7	.6	-.11		2
	3	3	154	26	.19	.62	.05	.8	.7	.12		3
	4	4	107	18	.85	.69	.07	.7	.6	.30		4
	5	5	71	12	1.85	1.10	.13	.9	.8	.48		5 VT Veldig trygg
	MISSING	***	1	0#	-.28	.00				.00		

Missing % includes all categories. Scored % only of scored categories

Svarkategorier OA - instrument

```
↑TABLE 26.3 Kilder                                ZOU346WS.TXT  Feb 10 2021 14:58
INPUT: 583 Person  24 Item  REPORTED: 583 Person  6 Item  6 CATS WINSTEPS 3.91.2
-------------------------------------------------------------------------------

         Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES:  CORRELATION ORDER

-----------------------------------------------------------------------------------------
|ENTRY   DATA  SCORE |     DATA    |      ABILITY     S.E.  INFT OUTF PTMA |              |
|NUMBER  CODE  VALUE |  COUNT   %  |    MEAN    P.SD  MEAN  MNSQ MNSQ CORR.| Item         |
|--------------------+-------------+--------------------------------------+--------------|
|   15   0         0 |     43   7  |   -4.09     1.52  .23  1.7  1.8  -.51 |15  Item15OS  | 0 VU   Veldig utrykk
|        1         1 |     44   8  |   -2.41     1.11  .17  1.0  1.0  -.31 |              | 1
|        2         2 |     64  11  |   -1.20      .90  .11  1.0  1.0  -.20 |              | 2
|        3         3 |    104  18  |    -.49      .85  .08   .9   .7  -.13 |              | 3
|        4         4 |    153  26  |     .56     1.06  .09  1.2   .9   .10 |              | 4
|        5         5 |    171  30  |    2.43     1.85  .14  1.3  1.2   .63 |              | 5 VT   Veldig trygg
|        MISSING *** |      4   1#|     .25     1.66  .96             .00 |              |
|                    |             |                                      |              |
|   13   0         0 |     63  11  |   -3.43     1.85  .23  2.1  2.3  -.54 |13  Item13OS  | 0 VU   Veldig utrykk
|        1         1 |     57  10  |   -1.64     1.15  .15  1.2  1.4  -.25 |              | 1
|        2         2 |    107  19  |    -.84     1.11  .11  1.4  1.4  -.20 |              | 2
|        3         3 |    105  18  |     .02      .93  .09  1.0  1.0  -.03 |              | 3
|        4         4 |    131  23  |    1.08      .93  .08   .8   .7   .21 |              | 4
|        5         5 |    115  20  |    3.02     1.84  .17  1.5  1.3   .61 |              | 5 VT   Veldig trygg
|        MISSING *** |      5   1#|     .02     1.25  .62            -.01 |              |
|                    |             |                                      |              |
|   18   0         0 |    135  23  |   -2.44     1.78  .15  1.4  1.3  -.61 |18  Item18OS  | 0 VU   Veldig utrykk
|        1         1 |    105  18  |    -.81      .99  .10  1.0   .9  -.19 |              | 1
|        2         2 |    114  20  |     .15      .95  .09  1.0  1.2   .00 |              | 2
|        3         3 |    104  18  |    1.15      .94  .09   .9   .9   .20 |              | 3
|        4         4 |     78  14  |    2.52     1.25  .14   .9  1.0   .40 |              | 4
|        5         5 |     39   7  |    4.20     1.88  .31  2.4  2.8   .47 |              | 5 VT   Veldig trygg
|        MISSING *** |      8   1#|    1.07      .63  .24              .05 |              |
|                    |             |                                      |              |
|   17   0         0 |    112  19  |   -2.75     1.71  .16  1.2  1.2  -.61 |17  Item17OS  | 0 VU   Veldig utrykk
|        1         1 |    103  18  |   -1.04      .93  .09   .8   .8  -.24 |              | 1
|        2         2 |    110  19  |    -.06      .87  .08   .9   .8  -.04 |              | 2
|        3         3 |    103  18  |     .83      .75  .07   .7   .6   .14 |              | 3
|        4         4 |     78  14  |    1.86      .90  .10   .7   .7   .29 |              | 4
|        5         5 |     69  12  |    4.03     1.39  .17   .9   .9   .61 |              | 5 VT   Veldig trygg
|        MISSING *** |      8   1#|     .57     1.07  .41              .02 |              |
|                    |             |                                      |              |
|   16   0         0 |     96  17  |   -3.03     1.60  .16  1.1  1.0  -.61 |16  Item16OS  | 0 VU   Veldig utrykk
|        1         1 |     85  15  |   -1.34      .80  .09   .5   .6  -.27 |              | 1
|        2         2 |    122  21  |    -.17      .73  .07   .7   .6  -.08 |              | 2
|        3         3 |     98  17  |     .74      .72  .07   .6   .5   .11 |              | 3
|        4         4 |     99  17  |    1.67     1.09  .11  1.1  1.0   .29 |              | 4
|        5         5 |     77  13  |    3.70     1.62  .19  1.1  1.1   .59 |              | 5 VT   Veldig trygg
|        MISSING *** |      6   1#|    -.84     2.12  .95            -.04 |              |
|                    |             |                                      |              |
|   14   0         0 |     79  14  |   -3.38     1.53  .17  1.0  1.0  -.60 |14  Item14OS  | 0 VU   Veldig utrykk
|        1         1 |     84  15  |   -1.49      .85  .09   .7   .7  -.29 |              | 1
|        2         2 |     97  17  |    -.49      .65  .07   .6   .5  -.12 |              | 2
|        3         3 |    102  18  |     .38      .80  .08   .7   .7   .05 |              | 3
|        4         4 |    107  19  |    1.16      .76  .07   .8   .7   .20 |              | 4
|        5         5 |    109  19  |    3.36     1.54  .15   .9   .9   .66 |              | 5 VT   Veldig trygg
|        MISSING *** |      5   1#|     .29     1.77  .89              .01 |              |
-----------------------------------------------------------------------------------------
 # Missing % includes all categories. Scored % only of scored categories
```

Svarkategorier PFT-instrument

TABLE 26.3 Kilder ZOU394WS.TXT Feb 10 2021 14:54
INPUT: 583 Person 24 Item REPORTED: 583 Person 6 Item 6 CATS WINSTEPS 3.91.2

Item CATEGORY/OPTION/DISTRACTOR FREQUENCIES: CORRELATION ORDER

ENTRY NUMBER	DATA CODE	SCORE VALUE	DATA COUNT	%	ABILITY MEAN	P.SD	S.E. MEAN	INFT MNSQ	OUTF MNSQ	PTMA CORR.	Item	
24	5	0	37	6	-2.41	1.50	.25	1.6	1.8	-.51	24 Item24PFT	5 VT Veldig trygg
	4	1	24	4	-1.12	.79	.16	.8	.8	-.25		4
	3	2	42	7	-.22	.93	.14	1.3	1.8	-.19		3
	2	3	89	15	.28	.80	.09	1.1	1.3	-.16		2
	1	4	123	21	.83	.76	.07	.8	.8	-.03		1
	0	5	265	46	2.05	1.41	.09	1.2	1.2	.59		0 VU Veldig utrygg
	MISSING	***	3	1#	-.05	1.31	.93			-.04		
19	5	0	45	8	-2.24	1.39	.21	1.2	1.3	-.54	19 Item19PFT	5 VT Veldig trygg
	4	1	30	5	-.63	.73	.14	1.2	1.3	-.21		4
	3	2	58	10	-.21	.59	.08	.8	.8	-.23		3
	2	3	64	11	.22	.69	.09	.8	.9	-.15		2
	1	4	116	20	.89	.77	.07	1.0	.8	-.02		1
	0	5	264	46	2.13	1.35	.08	1.1	1.1	.63		0 VU Veldig utrygg
	MISSING	***	6	1#	-.38	1.65	.74			-.08		
20	5	0	93	16	-1.14	1.51	.16	1.2	1.1	-.53	20 Item10PFT	5 VT Veldig trygg
	4	1	81	14	-.04	.78	.09	1.0	1.0	-.23		4
	3	2	87	15	.54	.75	.08	1.2	1.0	-.10		3
	2	3	117	20	1.14	.81	.08	1.2	1.3	.05		2
	1	4	104	18	1.68	.95	.09	1.4	1.6	.20		1
	0	5	91	16	3.30	1.31	.14	1.4	1.5	.59		0 VU Veldig utrygg
	MISSING	***	10	2#	-.17	1.69	.56			-.08		
23	5	0	73	13	-1.42	1.57	.19	1.5	1.7	-.52	23 Item23PFT	5 VT Veldig trygg
	4	1	51	9	-.29	.92	.13	1.4	1.5	-.22		4
	3	2	70	12	-.11	.67	.08	.8	.8	-.22		3
	2	3	87	15	.63	.68	.07	.8	.8	-.08		2
	1	4	122	21	1.21	.73	.07	.9	.8	.08		1
	0	5	174	30	2.67	1.26	.10	1.0	.9	.66		0 VU Veldig utrygg
	MISSING	***	6	1#	.85	.95	.43			-.01		
22	5	0	52	9	-1.89	1.40	.20	.8	.8	-.52	22 Item22PFT	5 VT Veldig trygg
	4	1	57	10	-.60	.92	.12	.9	1.1	-.30		4
	3	2	65	11	.15	.72	.09	1.0	1.1	-.16		3
	2	3	102	18	.57	.70	.07	.9	.8	-.10		2
	1	4	173	30	1.39	.86	.07	.9	1.1	.17		1
	0	5	132	23	2.81	1.38	.12	1.1	1.0	.59		0 VU Veldig utrygg
	MISSING	***	2	0#	-1.32	2.68	2.68			-.08		
21	5	0	38	7	-2.37	1.44	.24	1.2	1.2	-.51	21 Item21PFT	5 VT Veldig trygg
	4	1	29	5	-1.31	.80	.15	.6	.6	-.30		4
	3	2	45	8	-.43	.45	.07	.5	.4	-.23		3
	2	3	82	14	.14	.47	.05	.5	.4	-.19		2
	1	4	136	24	.88	.62	.05	.6	.5	-.02		1
	0	5	242	42	2.29	1.30	.08	.9	.9	.67		0 VU Veldig utrygg
	MISSING	***	11	2#	.66	1.04	.33			-.02		

Missing % includes all categories. Scored % only of scored categories