

Prestasjonsmotivasjon i en low-stakes testkontekst

Analyser av de norske 3MX-elevenes motivasjon og innsats på matematikktesten i TIMSS-Advanced 2008

Tatjana Hillervik



RDID 4190 – Masteroppgave i realfagdidaktikk
Lærerutdanning i realfag (LAP-90)
Institutt for lærerutdanning og skoleforskning
Det utdanningsvitenskapelige fakultet

UNIVERSITETET I OSLO

Våren 2011

Prestasjonsmotivasjon i en low-stakes testkontekst - *analyser av de norske 3MX-elevenes motivasjon og innsats på matematikktesten i TIMSS-Advanced 2008*

Masteroppgave i realfagdidaktikk

Institutt for lærerutdanning og skoleforskning

UNIVERSITETET I OSLO

Våren 2011

© Tatjana Hillervik

2011

Prestasjonsmotivasjon i en low-stakes testkonteks

Analyser av de norske 3MX- elevenes motivasjon og innsats på matematikktesten i TIMSS
Advanced 2008

Tatjana Hillervik

<http://www.duo.uio.no/>

Trykk: Grafisk, Skedsmo videregående skole

Sammendrag

Målet med denne masteroppgaven er å undersøke norske elevers selvrapporterte motivasjon og innsats på matematikktesten i TIMSS Advanced 2008. TIMSS (Trends in International Mathematics and Science Study) Advanced er en internasjonal komparativ undersøkelse av matematikk- og fysikkspesialistene i det siste året i videregående skole. Over 1900 norske 3MX-elever deltok i undersøkelsen i 2008.

Testresultater i TIMSS Advanced påvirker ikke elevenes karakterer i faget, og testtakerne har ikke noe personlig utbytte av testen fordi testresultater for hver enkelt elev ikke presenteres for verken eleven eller hans/hennes lærer. Slike tester betegnes som *low-stakes* tester for elever, samtidig som testresultater har en stor betydning for deltakerlandene på politisk nivå fordi testresultatene kan ha innvirkning på utdanningspraksis og utdanningspolitikk. Det er en generelt utbredt bekymring for at elever ikke er motivert til å gjøre sitt beste på tester som ikke har direkte konsekvenser for dem, og at elevenes manglende motivasjon kan føre til testresultater som underestimerer elevenes dyktighet. Tross denne bekymringen finnes det få empiriske studier som har undersøkt sammenheng mellom elevenes prestasjon og motivasjon for å gjøre sitt beste i kontekst av storskala internasjonale komparative undersøkelser, og vi vet foreløpig lite om denne sammenhengen og elevenes testmotivasjon i slike kontekster. Måling av testmotivasjon hos elevene ble gjort i TIMSS Advanced for første gang i 2008, hvor elevene i et spørreskjema i etterkant av testen ble bedt om å angi hvor motiverte de var for å gjøre sitt beste og hvor god innsats de gjorde i arbeidet med oppgavene på testen.

I min studie undersøker jeg norske 3MX-elevers motivasjon og innsats på matematikktesten i TIMSS Advanced, basert på deres svar på spørsmål i spørreskjemaet. Jeg undersøker også sammenhengen mellom elevenes prestasjoner og deres angitte motivasjon på testen, og i tillegg trekker jeg inn i analysene bakgrunnsvariabler som elevenes begrunnelser for valget av fordypning i matematikk. Videre gjør jeg noen sammenlikninger for å undersøke om det er forskjell mellom jenter og gutter på noen av de ovennevnte punktene, og gjør også noen sammenlikninger mellom Norge og Sverige.

Jeg har gjennomført en kvantitativ undersøkelse innen rammene for prestasjonsmotivasjonsteori om forventning og verdisseting, som er presentert i et eget kapittel i denne oppgaven. Teoriens paradigme sier at innsatsen som individer yter for å utføre en bestemt oppgave, er påvirket av to faktorer: deres forventning til suksess

(forventningsfaktor) og deres vurdering av oppgavens verdi (verdifaktor), og disse faktorene utformes i en sosial kontekst. Testmotivasjonsmålinger er et relativt nytt felt i skoleforskningen, og i teorikapitlet tar jeg også opp noen problemstillinger rundt de teoretiske rammene for testmotivasjonskonstruktet.

Analyser gjennomført i denne masteroppgaven viser at bare under halvparten av norske 3MX-elever sier seg helt enige eller litt enige i at de var motivert for å gjøre sitt beste på TIMSS-testen, og i at de arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige. Deres svar på enkeltpåstandene om motivasjon og innsats i elevspørreskjemaet kan tyde på at elevene i gjennomsnitt hadde noe lavere innsatsnivå på matematikktesten i TIMSS Advanced enn på andre tester de har ellers i løpet av året, men jeg vil påpeke at mine analyser av datamaterialet i TIMSS Advanced ikke gir mulighet til å trekke slutninger om i hvilken grad lavere testmotivasjonsnivå hos norske 3MX-elever kan forklares med hvor viktig de vurderte testen til å være.

Når det gjelder sammenhengen mellom prestasjon og motivasjon for å gjøre sitt beste på testen, er det ved en korrelasjonsanalyse i denne oppgaven påvist en signifikant positiv sammenheng mellom den målte testmotivasjonen hos norske elever og deres prestasjon på matematikktesten i TIMSS Advanced, men denne sammenhengen er ikke sterk. Det finnes få studier som tidligere har undersøkt sammenhengen mellom testmotivasjon hos elever og deres prestasjon på testen i komparative undersøkelser som TIMSS, og konklusjonene i disse undersøkelsene er ikke entydige. Jeg peker dermed på behovet for videre studier for å kunne få en bedre forståelse av sammenhengen mellom prestasjon og motivasjon hos elever i slike testkontekster.

Videre viser analysene i denne oppgaven at det er ingen signifikante forskjeller i det målte testmotivasjonsnivået mellom jenter og gutter blant norske 3MX-elever i TIMSS Advanced. Det er heller ikke påvist kjønnsforskjeller i sammenhengen mellom elevenes prestasjon på testen og deres testmotivasjon.

Sammenlikninger av elevenes angitte motivasjon og innsats på matematikktesten i TIMSS Advanced mellom Norge og Sverige tyder på omtrent samme testmotivasjonsnivå hos elevene i disse to landene, noe som føyer seg til likhetstrekkene mellom Norge og Sverige i TIMSS Advanced 2008.

Gjennom min oppgave håper jeg å ha gitt et lite bidrag til at man ser betydningen av å forske ikke bare på elevenes prestasjoner i internasjonale komparative studier, men også å se på andre faktorer som motivasjon og innsats på testen.

Forord

For halvannet år siden fikk jeg mulighet til å komme i gang med studier innen realfagdidaktikk ved ILS, og denne masteroppgaven er den avsluttende fasen i disse studiene. Fra tidligere har jeg bakgrunn i studier ved Moskva statlige universitet med diplomoppgave innen anvendt matematikk og kybernetikk, og de siste 14 årene har jeg jobbet som matematikklærer i videregående skole i Norge. Som student ved Universitetet i Oslo har jeg fått mange spennende studieopplevelser i møte med forskningen innen matematikdidaktikk. Det har vært en lærerik prosess på mange plan.

Jeg vil uttrykke min store takknemlighet overfor Liv Sissel Grønmo og Torgeir Onstad for deres støtte og veiledning underveis i arbeidet med denne masteroppgaven. Jeg vil også rette en stor takk til min tidligere lærerkollega og nåværende doktorgradsstipendiat ved ILS Ida Friestad Pedersen for hennes tilbakemeldinger på mine spørsmål underveis i skriveprosessen.

Jeg vil også takke min mann Bjørn og mine barn Oksana og Alexander, for deres støtte og tålmodighet gjennom en hektisk periode av mitt liv, hvor de måtte vende seg til at mine studier og jobbforpliktelser ofte gikk foran alt annet, selv om jeg prøvde å vise min familie mest mulig oppmerksomhet, spesielt under den tøffe sykdomsperioden for Bjørn.

Jeg vil også takke mine venner og kollegaer for deres oppmuntrende kommentarer underveis.

Fetsund, mai 2011

Tatjana Hillervik

Innholdsfortegnelse

1	Innledning	1
1.1	Bakgrunn for valg av tema	1
1.2	Problemstilling	2
1.2.1	Begrepsavklaringer	2
1.2.2	Low-stakes tester og testmotivasjon	4
2	TIMSS Advanced	7
2.1	Historikk	7
2.2	Mål og problemstillinger	9
2.3	Matematikkprestasjoner i TIMSS Advanced 2008	11
2.4	Kort om elevspørreskjema og noen bakgrunnsvariabler	15
2.5	Måling av testmotivasjon i TIMSS Advanced	16
3	Teori.....	18
3.1	Motivasjonsteori	18
3.1.1	Teori om forventning og verdisetting	20
3.2	Testmotivasjon	23
3.2.1	Måling av testmotivasjon	27
3.2.2	Noen resultater fra tidligere empiriske studier av testmotivasjon	30
4	Metode	33
4.1	Data	33
4.2	Den psykometriske skalaen for konstruktet ”Testmotivasjon” i TIMSS Advanced 2008	34
4.3	Databehandling	36
4.3.1	Behandling av samlevariabler	36
4.3.2	Prestasjonsskår i matematikktesten	36
4.3.3	Bortfalte verdier	37
4.3.4	Statistiske analyser	38
5	Analyser og resultater	42
5.1	Testmotivasjon hos de norske 3MX-elevne i TIMSS Advanced 2008	42

5.1.1	Elevenes svar på påstandene i elevspørreskjema	42
5.1.2	Testmotivasjon som samlev variabel.....	47
5.2	Kjønnsforskjeller i testmotivasjon blant norske elever	48
5.2.1	Kjønnsforskjeller i elevenes svar på påstandene i elevspørreskjema	48
5.2.2	Kjønnsforskjeller i testmotivasjon som samlev variabel.....	51
5.3	Norske elevers angivelse av motivasjon og innsats på testen sett i sammenheng med deres begrunnelser for valget av fordypning i matematikk	53
5.4	Prestasjonsresultater i matematikktesten i TIMSS Advanced sett i sammenheng med elevenes selvrapporterte testmotivasjon	59
5.5	Testmotivasjon hos norske elever på matematikktesten i TIMSS Advanced 2008 sammenliknet med svenske elever.....	62
6	Oppsummering	66
6.1	Elevenes svar på påstander om motivasjon og innsats på TIMSS Advanced testen	66
6.2	Testmotivasjon og noen bakgrunnsvariabler.....	67
6.3	Sammenheng mellom testmotivasjon og testprestasjon.....	68
6.4	Kjønnsforskjeller.....	69
6.5	Sammenlikning av resultater for Norge og Sverige	69
6.6	Veien videre	70
	Litteraturliste.....	72

1 Innledning

1.1 Bakgrunn for valg av tema

Internasjonale komparative undersøkelser (som for eksempel PISA og TIMSS¹) har ingen direkte konsekvenser for deltakende elever, og dermed representerer slike undersøkelser såkalte "low-stakes" tester² for elevene (Baumert & Demmrich, 2001; Wise & DeMars, 2008; Eklöf, 2010b), samtidig som testresultater har en stor betydning for deltakerlandene (testarrangørene). På politisk nivå betraktes slike undersøkelser som "high-stakes"³ tester fordi resultatene kan ha innvirkning på utdanningspraksis og utdanningspolitikk (Eklöf, 2010b ref. Robitaille, Beaton & Plomp, 2000).

Internasjonale komparative undersøkelser blir ofte kritisert for at testsituasjonen kan ha vært forskjellig i de forskjellige deltakerlandene. Denne kritikken tar utgangspunkt i at mentale og fysiske handlinger er situert i kontekster og er en del av sosiale praksiser. Hvis elevene i noen land hadde, av politiske eller kulturelle grunner, bedre motivasjon til å yte sitt beste på slike tester enn elevene i andre land, så kan det være vanskelig å sammenlikne elevprestasjonene i de forskjellige landene." *It is possible that some of the differences between students from different nations could be explained by differences in test-taking motivation.*" (Eklöf, 2010b, s.354). Tatt i betraktning utbredelsen av internasjonale komparative studier, er det viktig å kunne forstå bedre hvilken rolle elevenes motivasjon har i kontekst av slike tester (Barry m.fl., 2010).

Det er en generelt utbredt bekymring for at elevenes manglende motivasjon på tester som ikke har direkte konsekvenser for dem, kan føre til resultater som underestimerer elevenes dyktighet (Baumert & Demmrich, 2001; Thelk m.fl., 2009). I hvilken grad elever er motivert for å ta en prøve og hvordan motivasjon i testsituasjon påvirker prestasjoner på testen har vært gjenstand for forskning i flere land. For eksempel viser studier i USA og Tyskland at man så langt ikke har klart å påvise at manglende testmotivasjon kan årsaksforklare dårlige resultater i TIMSS og PISA (Kjærnsli & Roe, 2010). I Sverige har det vært forsket på elevenes motivasjon i TIMSS 2003, og i TIMSS Advanced 2008. Resultatene viser at de svenske

¹ Programme for International Student Assessment og Trends in International Mathematics and Science Study

^{2,3} Omtales nærmere i 1.2.1 Begrepsavklaringer

elevene i stor grad var motivert for å gjennomføre TIMSS 2003-undersøkelsen, mens foreløpige resultater fra TIMSS Advanced 2008 avdekket et lavere nivå av motivasjon og innsats på testen. Dessuten er det i disse undersøkelsene funnet en signifikant positiv sammenheng mellom elevenes rapporterte innsats og deres prestasjoner på testen (Eklöf, 2007, 2010a; Skolverket 2009). I Sverige er det også blitt sammenliknet elevenes testmotivasjon under en nasjonal prøve i naturfag og TIMSS Advanced 2008-testen, som ble vurdert som henholdsvis high-stakes test og low-stakes test for elevene, og resultatene viste at elevene som deltok i den nasjonale testen betraktet den som viktig for dem og har vist god innsats på prøven, mens i TIMSS-testen har elevene rapportert om et relativt lavt nivå av motivasjon og innsats på testen (Eklöf, 2010a).

Spørsmål om testmotivasjon er nytt i TIMSS Advanced 2008, sammenliknet med den tidligere undersøkelsen i 1995. Dette spørsmålet ble tatt med i elevspørreskjema i Slovenia, Russland, Sverige og Norge. Sverige er så langt det eneste landet som har lagt fram foreløpige analyser av svarene på motivasjonsspørsmålet. I Norge har det vært lite forskning på testmotivasjon og hvordan testresultater kan ses i sammenheng med testmotivasjon (Kjærnsli & Roe, 2010). De innsamlede data i TIMSS Advanced 2008 gir meg muligheten til å få innblikk i dette forskningsfeltet, og jeg vil undersøke hvordan norske elever angir sin innsats og motivasjon for å gjøre sitt beste på TIMSS-testen og forsøke å drøfte relasjon mellom prestasjon og motivasjon på testen med utgangspunkt i motivasjonsteori og det empiriske datamaterialet fra TIMSS Advanced 2008.

1.2 Problemstilling

1.2.1 Begrepsavklaringer

Problematikken i denne masteroppgaven er tett knyttet til teststatus og motivasjon i testsituasjon. I engelskspråklig faglitteratur brukes de veletablerte begrepene ”low-stakes test” og ”high-stakes test” om teststatus. For å finne gode oversettelser av disse, er det søkt i norsk faglitteratur, men det ser ut som om disse begrepene ikke har et kortfattet navn på norsk. For eksempel i den norske rapporten PISA 2009 ”På rett spor”, under drøfting av motivasjon blant norske elever, omtales PISA-prøven som en prøve, hvor resultatene ikke har betydning for elevenes karakter (Kjærnsli & Roe, 2010). Et mulig forslag til norsk oversettelse av ”low-

stakes test” kunne, for eksempel, være ”test av liten betydning” (og tilsvarende ”test av stor betydning” for ”high-stakes test”), men det virker både noe upresist og tungt å bruke.

Dessuten er det flere faktorer som legges til grunn for vurdering av teststatus. Jeg velger dermed å beholde den engelskspråklige terminologien om teststatus og forklarer begrepene nærmere.

Teststatus refererer til konsekvensene testresultatet har. Avhengig av konsekvensvurdering, deles tester i to typer tester: low-stakes og high-stakes.

Low-stakes test for elevene har følgende kjennetegn: testresultater påvirker ikke elevenes karakterer i faget, og testtakere har ingen personlig utbytte av testen fordi testresultater for hver enkelt elev ikke presenteres for verken eleven eller hans/hennes lærer, og kan dermed ikke brukes til å finne ut hva eleven kan og ikke kan, slik at det kan iverksettes tiltak som kan forbedre elevenes kunnskaper.

High-stakes test for elevene har klare konsekvenser for dem. Når vurderingen av testens resultater har stor betydning på systemnivå slik at den kan danne et grunnlag for viktige beslutninger, anses testen til å være high-stakes test på systemnivå. “Certain uses of achievement test results are termed "high stakes" if they carry serious consequences for students or for educators.” (AERA, 2000, s. 4).

Én og samme test kan ha forskjellig status på forskjellige nivåer, for eksempel low-stakes test for testtaker kan være high-stakes for testarrangør.

Motivasjon er et aspekt som sterkt preger utdannings- og læringsprosesser (Schunk m.fl., 2010). Viktige begreper innen motivasjonsteori blir nærmere omtalt i kapittel 3, men det synes å være passende å gi en kort introduksjon av noen sentrale begrep i dette kapitlet.

Pintrich and Schunk (2010) har definert motivasjon som en prosess, hvor en målrettet aktivitet settes i gang og deretter opprettholdes. I en testkontekst vil den målrettede aktiviteten bety å gjøre sitt beste, slik at man viser nøyaktig hva man kan innenfor det kognitive domenet som testen setter på prøve (Wise & DeMars, 2005). En slik motivasjon defineres innen rammene for teori om forventning og verdi og kalles for *testmotivasjon*. Det ligger implisitt i definisjonen at testmotivasjon er situasjonsbetinget og den kan variere, avhengig av hvilken sosiokulturell kontekst testen gjennomføres i. (Baumert & Demmrich, 2001).

1.2.2 Low-stakes tester og testmotivasjon

I dag gjennomføres det mange tester i skolen, hvor teststatus på testtakernivå og på systemnivå er ganske forskjellige fra hverandre. For eksempel, har resultatene i internasjonale komparative undersøkelser ingen konsekvenser for hver enkelt elev som deltar, men har store konsekvenser for vurdering av nasjonale utdanningssystemer og institusjoner. Målet med slike undersøkelser er å estimere og sammenlikne gjennomsnittlig kompetansenivå blant elever fra forskjellige land (Wise, 2009). Tolkningen av testresultater i slike undersøkelser har dermed en sentral rolle i vurdering av testens validitet (Baumert & Demmrich, 2001; Eklöf, 2010b ref. Kane, 2009). Testvaliditet defineres som ”*integrative evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores and other modes of assessment*” (Eklöf, 2006b ref. Messick, 1989, s.11). Det betyr at for å kunne trekke gyldige slutninger om testresultater må vi vite om vi har målt det som testen er konstruert for å måle.

Validiteten av testresultater bygger på følgende premiss: den demonstrerte kunnskapen = den reelle kunnskapen. For å tilegne seg kunnskap trengs det en viss grad av motivasjon og innsats for å lære. På samme måte trengs det en viss grad av testmotivasjon for å kunne demonstrere den tilegnede kunnskapen i en testsituasjon (Eklöf, 2010b). I vurderingskontekster er det viktig å forstå motivasjonsnivå blant testdeltakere for å vite hvorvidt vi har testet kunnskap eller deres vilje til å vise denne kunnskapen (Wise & DeMars, 2008; Thelk m.fl., 2009; Eklöf, 2010b ref. AERA, APA, NCME⁴, 1999). Testmotivasjon kan sees i sammenheng med teststatus. Low-stakes tester, hvor testresultater verken har konsekvenser eller gir noe personlig utbytte for testdeltakere, kan forbindes med en lav testmotivasjon hos en del av dem (Barry & Finney, 2009; Eklöf, 2010b). Resultater av teoretisk og empirisk forskning tyder også på at lav testmotivasjon ofte fører til redusert innsats og prestasjon på testen (Wise & DeMars, 2005). “Without adequate effort, test performance is likely to suffer, resulting in the examinee’s test score underestimating his or her actual level of proficiency.” (Wise & DeMars, 2008, s.3). Lav testmotivasjon kan dermed føre til variasjon i testresultater, som ikke gir et riktig bilde av de reelle kunnskapene hos testtakere. Med andre ord, når testen er, sett fra testtakernes synspunkt, en low-stakes test, kan manglende

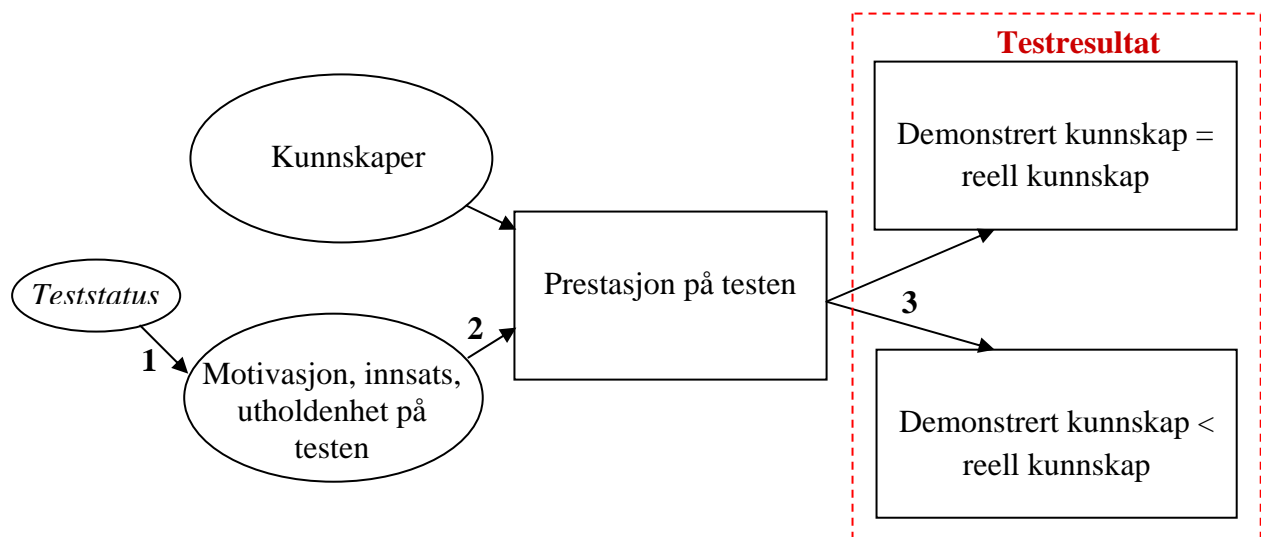
⁴ American Educational Research Association, American Psychological Association, National Council on Measurement in Education

drøftinger av deres testmotivasjon ved evaluering av testresultatene true validiteten av testresultattolkningene (Wise & DeMars, 2008; Thelk m.fl., 2009).

Den ovennevnte argumentasjon peker ut tre punkter som bør fokuseres på i forbindelse med resultatevaluering av low-stakes tester:

1. Testmotivasjonsnivå hos testtakere, i det lav testmotivasjon hos den enkelte testtaker kan ha sammenheng med den "lave" teststatusen.
2. Sammenheng mellom testtakernes testmotivasjon og deres prestasjon på testen.
3. Validiteten av testresultater ut i fra vurderingskonteksten.

Det ovennevnte kan settes sammen og illustreres i en modell, vist på figur 1.1.



Figur 1.1 Testresultat sett i sammenheng med teststatus og testmotivasjon. Bearbeidet og forenklet utgave av modellen for faktorer som påvirker prestasjonsresultater på testen (Eklöf, 2010b)

I denne masteroppgaven vil jeg undersøke norske elevers selvrappørterte nivå av motivasjon og innsats på matematikktesten i TIMSS Advanced 2008, og med utgangspunkt i modellen i figur 1.1 kan problemstillingen i oppgaven uttrykkes ved følgende forskningsspørsmål:

1. Hvor høy motivasjon og innsats oppgir norske elever at de hadde på matematikktesten TIMSS Advanced 2008?
2. Hvordan er sammenhengen mellom norske elevers angivelse av egen innsats og motivasjon og deres selvvurdering i og verdisetting av faget matematikk?

3. Hvordan er sammenhengen mellom norske elevers prestasjoner på testen og deres angivelse av egen motivasjon og innsats?
4. Er det forskjeller mellom norske jenter og gutter på noen av disse punktene?
5. Er det forskjeller mellom de norske og de svenske elevene på noen av disse punktene?

For å belyse forskningsspørsmålene i denne oppgaven gjennomfører jeg en kvantitativ undersøkelse basert på det empiriske datamaterialet fra elevspørreskjema og prestasjonsresultatene i TIMSS Advanced 2008.

2 TIMSS Advanced

2.1 Historikk

TIMSS (Trends in International Mathematics and Science Study) Advanced er en internasjonal komparativ undersøkelse av matematikk- og fysikkspesialistene i det siste året i videregående skole, som føyer seg i rekken av TIMSS studier i regi av en internasjonal organisasjon for utdanningsforskning, IEA (International Association for the Evaluation of Educational Achievement). Denne organisasjonen ble dannet i forbindelse med den første studien FIMS (First International Mathematics Study, 1967) som ble etterfulgt av studier i seks andre fag (Angell m.fl., 1999). *”The founders of IEA viewed the world as a natural educational laboratory, where different school systems experiment in different ways to obtain optimal results in the education of their youth.”* (IEA, 2007a, online http://www.iea.nl/brief_history_of_iea.html). IEA’s studier er empiriske undersøkelser som er designet i form av både tverrsnittsundersøkelser og longitudinelle undersøkelser⁵ (trendstudier), hvor datainnsamling skjer gjennom hovedsakelig kvantitative metoder med representativt utvalg av elever fra deltakerlandene (IEA, 2007b).

I begynnelsen av 80-årene ble det gjennomført to store separate undersøkelser av realfagene i skolen: SIMS (Second International Mathematics Study) i matematikk og SISS (Second International Science Study) i naturfagene. I stedet for å gjennomføre nye separate studier, ble det i 90-årene satt i gang en kombinert studie i matematikk og naturfag med navnet TIMSS (Third International Mathematics and Science Study). *“Med 45 deltakerland og nesten en million elever fra 15000 skoler, er TIMSS uten tvil det største og mest ambisiøse komparative prosjektet som er gjennomført innen utdanning.”* (Angell m. fl.,1999, s.7). De neste TIMSS undersøkelsene fortsatte med samme forkortelsen men med den nye, nåværende betydningen ”Trends in International Mathematics and Science Study” ettersom studiene var designet ikke bare for å kunne sammenlikne mellom land men også for å sammenlikne over tid (Grønmo m.fl., 2010).

⁵ Tverrsnittsundersøkelse gjennomføres på ett tidspunkt. Longitudinell undersøkelse (her som trendstudie) er studie hvor undersøkelse av samme tema gjennomføres på flere tidspunkter, og det er forskjellige personer som deltar på forskjellige tidspunkt (Johannesen m.fl., 2009)

De tidligere studiene undersøkte to populasjoner: klassetrinnene med flest 9-åringer og klassetrinnene med flest 13-åringer (i dag omfatter dette 4. og 8. trinn i grunnskolen i Norge). Av studiene gjennomført før 90-tallet, deltok Norge bare i SISS. Fra 1995 ble TIMSS studiene utvidet og de skulle omfatte også elever i avgangsklassene i videregående skole (Grønmo m.fl., 2010 ref. Mullis m.fl., 1998). Det ble definert tre populasjoner innen denne elevgruppen: generalistene (alle elever i samtlige studieretninger), fysikkspesialistene (elever som tok høyeste spesialisering i fysikk) og matematikkspesialistene (elever som tok høyeste spesialisering i matematikk). Generalistene ble testet i allmenne matematikk- og naturfagkunnskaper, mens fysikkspesialistene og matematikkspesialistene ble testet i henholdsvis bare fysikk og matematikk. Det er undersøkelsene av fysikk- og matematikkspesialistene i TIMSS 1995 som har blitt videreført som TIMSS Advanced 2008 (Grønmo m.fl. 2010, s.237). Det er relativt få land som har deltatt i TIMSS Advanced 2008 sammenliknet med tidligere TIMSS undersøkelser. Liste over deltakere i matematikkundersøkelsen er vist i tabell 2.1.

Tabell 2.1 Deltakerland i matematikkundersøkelsen i TIMSS Advanced 2008.

Deltok både i 1995 og i 2008*	Deltok bare i 2008
Italia	Armenia
Norge**	Filippinene
Russland	Iran
Slovenia	Libanon
Sverige	Nederland
* Deltakelse i begge studiene gir trenddata som viser utviklingen over tid	
** Se om norsk deltakelse nedenfor	

Norsk deltakelse i TIMSS Advanced

I 1995 deltok Norge bare i undersøkelsene av generalistene og fysikkspesialistene. Våren 1998, etter ønske fra norske myndigheter, ble også matematikkspesialistene undersøkt, og det ble gjennomført den samme matematikkundersøkelsen i Norge som hadde vært gjennomført internasjonalt i 1995⁶ (Angell m.fl., 1999). Utsettelsen av matematikkundersøkelsen i Norge førte til at de norske resultatene ikke ble med i grunnlaget for den standardiserte skalaen og

⁶ Samtidig ble det gjennomført en gjentakelse av generalisttesten for å undersøke hvilke endringer i elevenes kunnskaper som kunne registreres etter innføringen av Reform 94 (Angell m.fl, 1999, s.7)

beregningen av det internasjonale gjennomsnittet. Men selv om det resulterer i noen usikkerhet forbundet med norske data fra 1998, finnes det grundig gjennomførte analyser (Angell, Kjærnsli & Lie, 1999) som gir et godt grunnlag for å vurdere hvordan matematikkprestasjonene har forandret seg fra 1998 til 2008 (Grønmo m. fl., 2010).

I TIMSS Advanced 2008 ble fysikk- og matematikkspesialistene i Norge definert som de elevene som tok henholdsvis 3FY og 3MX våren 2008 (Grønmo m.fl., 2010).

2.2 Mål og problemstillinger

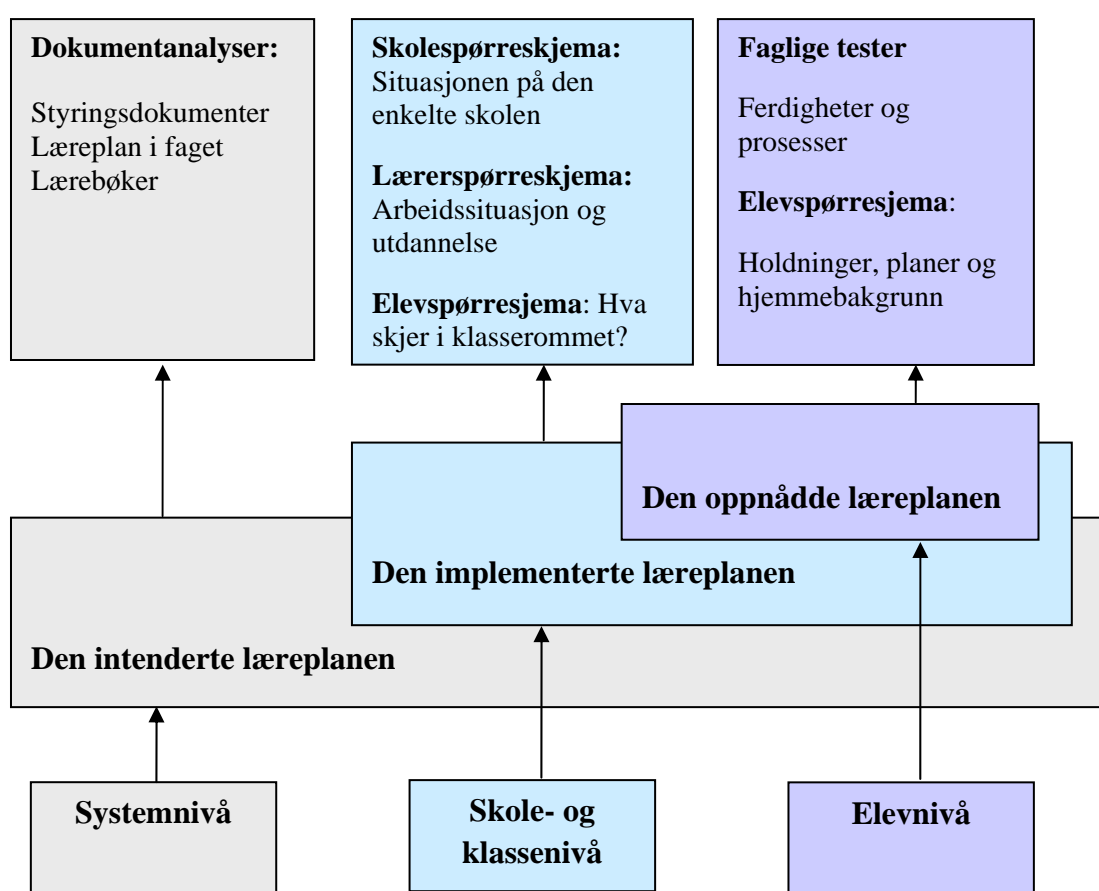
Valget av populasjon i TIMSS Advanced er begrunnet med den store betydningen en velfungerende videregående utdanning, spesielt innen realfag, har for videreutvikling av det moderne samfunnet: *”Many believe that the future security and wellbeing of their societies are strongly linked to the quality and quantity of well educated citizens graduating from their secondary schools, particularly those with strong backgrounds and career interests in fields related to mathematics, science, engineering, and technology.”* (Mullis m.fl., 2009, s.7). På grunn av den hierarkiske naturen av realfagene hvor utvikling av kompetansen på høyere nivå forutsetter at man har tilegnet seg de nødvendige kunnskapene på lavere nivå, er kompetansen i avansert matematikk og fysikk fra videregående skole nødvendig for å kunne studere matematikk og naturvitenskap videre.

Som det er nevnt tidligere i dette kapitlet er TIMSS-studiene designet for å kunne sammenlikne utdanningsresultater mellom land, og for å kunne måle utvikling over tid ved å fremskaffe gode trenddata for de enkelte deltakerlandene. Kort oppsummert består målene for TIMSS Advanced i å

- undersøke kunnskapene til elever som tar full fordypning i matematikk (eller fysikk) på siste trinn i videregående skole
- studere hvordan disse elevenes prestasjoner henger sammen med ulike faktorer som kjønn, faglig selvtillit og holdninger
- undersøke lærernes bakgrunn og tilretteleggingen av undervisningen
- sammenlike prestasjoner og bakgrunnsfaktorer mellom land
- studere utvikling over tid (trendstudier)

- prøve å identifisere faktorer, nasjonalt og internasjonalt, som fremmer god læring og en positiv utvikling innen matematikk (og fysikk) i skolen (Grønmo m.fl. 2010, s.13)

TIMSS Advanced, i likhet med TIMSS, kan karakteriseres som en læreplanbasert undersøkelse. Med "læreplanen" menes da det som står i fagplaner og lærebøker og analyse av læreplanene i deltakerlandene står sentralt i TIMSS fordi man har som mål at oppgavene i testen i stor grad er tilpasset det som undervises i majoriteten av deltakerlandene. Dataene i TIMSS-undersøkelsene analyseres på tre nivåer (TIMSS Norge⁷). Skjematisk fremstilling av analysenivåene er vist på figur 2.1.



Figur 2.1 Analysenivåene i TIMSS. Skjematisk fremstilling av hva som måles og hvordan det måles (TIMSS Norge⁸).

Systemnivå gjelder utdanningssystemet og dreier seg om hva slags utdanningstilbud samfunnet og myndighetene ønsker og planlegger at elevene skal få. Det dreier seg om organisering av skoletilbudet, rammefaktorer, ressurstilgang og elevenes muligheter til skole- og fagvalg. Ikke minst dreier det seg om læreplaner og vurderingsformer, og med den

⁷ http://www.timss.no/timss05_bakgrunn.html

⁸ Kilde: http://www.timss.no/timss05_bakgrunn.html (noe bearbeidet)

intenderte læreplanen menes det den tilsiktede læreplanen som er innfelt i fagplan og lærerbøker. På klassenivå undersøker man hvordan intensjonene fra systemnivå er omsatt i praksis. Det omhandler undervisningen og læringsmiljøet i klasserommet. Elevnivå dreier seg om hva som er oppnådd. På dette nivået undersøkes det hvilke fagkunnskaper og holdninger til faget elevene har (Grønmo m.fl., 2010).

For hvert enkelt land foretas det analyser av sammenhengen mellom de nevnte tre læreplannivåene. ”På den måten kan en se i hvilken grad "systemets" og lærerens mål og intensjoner blir oppfylt. I tillegg skal selvfølgelig også landene sammenlignes med hverandre. Forholdene i hvert land kan da studeres i et internasjonalt perspektiv.” (TIMSS Norge⁹).

2.3 Matematikkprestasjoner i TIMSS Advanced 2008

Det er gjort grundige analyser av resultatene i TIMSS Advanced 2008, og de er presentert både i de nasjonale rapportene (i Norge ved Grønmo m.fl., 2010) og i den internasjonale rapporten (Mullis m.fl., 2009). I dette avsnittet presenteres det noen av resultatene som synes å være relevante for analysene i denne masteroppgaven. De utvalgte resultatene viser hvordan de norske 3MX-elevne skårer på matematikktesten og også noen bakgrunnsvariabler basert på deres svar på elevspørreskjema. I tillegg til dette trekkes det inn i en viss grad de svenske resultatene.

Matematikktesten har to dimensjoner. Den ene dimensjonen dreier seg om faglige innholdskategorier som skal testes. Innholdskategorier for TIMSS Advanced 2008 er algebra, kalkulus og geometri. Den andre dimensjonen dreier seg om kognitive kategorier, som består av kunnskap, anvendelse og resonnement. Fordelingen mellom de kognitive kategoriene på testen er slik at omtrent 65 – 70 prosent av testens oppgaver skulle måle elevenes prestasjoner innen kategoriene anvendelse og resonnement (Mullis m.fl., 2009). Testresultatene måles på en standardisert skala hvor det internasjonale gjennomsnittet fra 1995 har blitt standardisert til 500 med et standardavvik på 100 (mer om nærmere beskrivelse av fordelingen mellom de forskjellige innholdskategoriene og de kognitive kategoriene på testen og om skaleringen kan leses i Grønmo m.fl., 2010).

⁹ http://www.timss.no/timss05_bakgrunn.html

Bak gjennomsnittsskår kan det skjule seg store variasjoner i elevprestasjoner. Sammenlikning av gjennomsnittlige resultater på tvers av landene kan derfor gi et veldig forenklet bilde av elevenes kunnskapsnivå i de forskjellige landene. For å unngå dette ble det dermed beskrevet såkalte kunnskapsnivåer i TIMSS Advanced. Disse nivåene er utviklet internasjonalt med hensyn til de to dimensjonene i testen. Det er definert tre kompetansenivåer. De oppgis ved såkalte *benchmark levels* eller kritiske nivåverdier som tilsvarer den nedre poenggrensen for hvert nivå (mer om utviklingen av ”benchmarking” kan leses i Arora, m.fl., 2009).

Kompetansenivåene deles inn i avansert, høyt og middels. Kompetansen til elever som skårer under middels nivå er ikke definert internasjonalt, men i den norske rapporten er det omtalt som lavt kompetansenivå. I figur 2.2 er det angitt de kritiske nivåverdiene samt en kort beskrivelse av de fire kompetansenivåene.

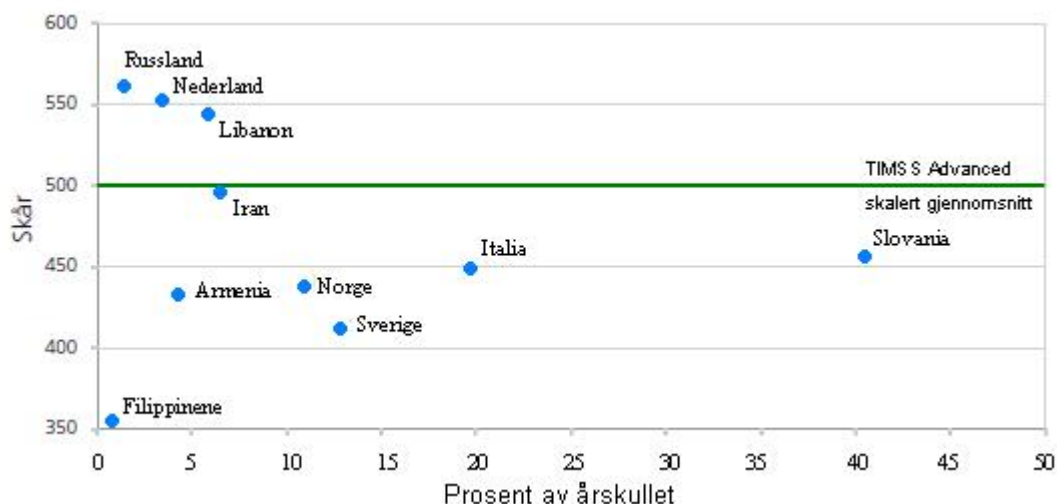
<p>Avansert nivå (625)</p> <p>Elevene viser begrepsforståelse og behersker prosedyrer. De demonstrerer evne til å gjennomføre resonnementer i algebra, trigonometri, geometri og differensial- og integralregning, og bruker dette til å løse problemer i komplekse situasjoner.</p>
<p>Høyt nivå (550)</p> <p>Elevene kan bruke sin kjennskap til matematiske begreper og prosedyrer i algebra, kalkulus, geometri og trigonometri for å analysere og løse både rutinepregede og ikke rutinepregede flertrinnsoppgaver.</p>
<p>Middels nivå (475)</p> <p>Elevene kan bruke sin kjennskap til begreper og prosedyrer i algebra, kalkulus og geometri for å løse rutinepregede oppgaver.</p>
<p>Lavt nivå (mindre enn 475 poeng)</p> <p>Elevene har ikke nådd middels nivå. Det kan for eksempel bety at de ikke kan utføre grunnleggende algebraiske operasjoner, at de ikke kan derivere enkle funksjoner eller at de ikke kan addere og subtrahere vektorer på koordinatform.</p>

Figur 2.2 ”Korte beskrivelser av de ulike kompetansenivåene i matematikk i TIMSS Advanced.” (Grønmo m.fl., 2010, s.47- 49)

”De norske 3MX-elevene presterer svakere enn elever i de fleste land det er naturlig å sammenlikne med. Unntaket er Sverige, som har en enda større tilbakegang i prestasjoner fra den forrige studien, og hvor gjennomsnittsskåren for prestasjoner ligger under den norske. Likevel er det likheten mellom de norske og de svenske resultatene som ofte framstår som slående.” (Grønmo m.fl., 2010¹⁰, s.7). Vi kan se landenes matematikkskår satt sammen med den prosentvise andelen av årskullet på figur 2.2. Denne andelen består av elevene som landene har definert som sine matematikkspesialister. De norske og de svenske elevene har

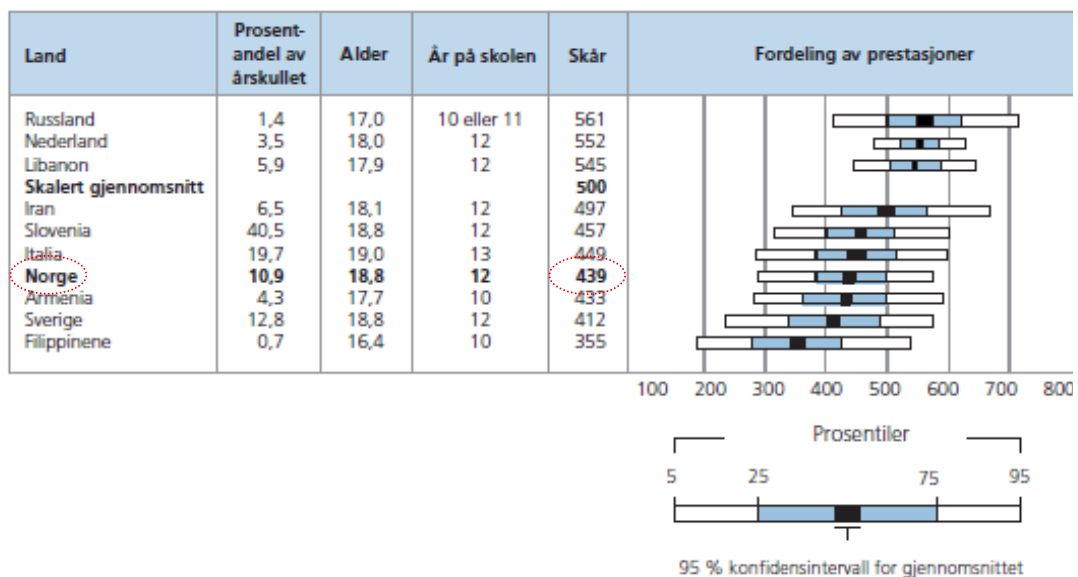
¹⁰ Matematikk i motvind, kortrapport, kap.2, s.2, lastet ned fra <http://www.timss.no/timss2008.html>

nøyaktig samme gjennomsnittsalder, og de utgjør omtrent like stor andel av årskullet i sitt land, med 13 % i Sverige mot 11 % i Norge.



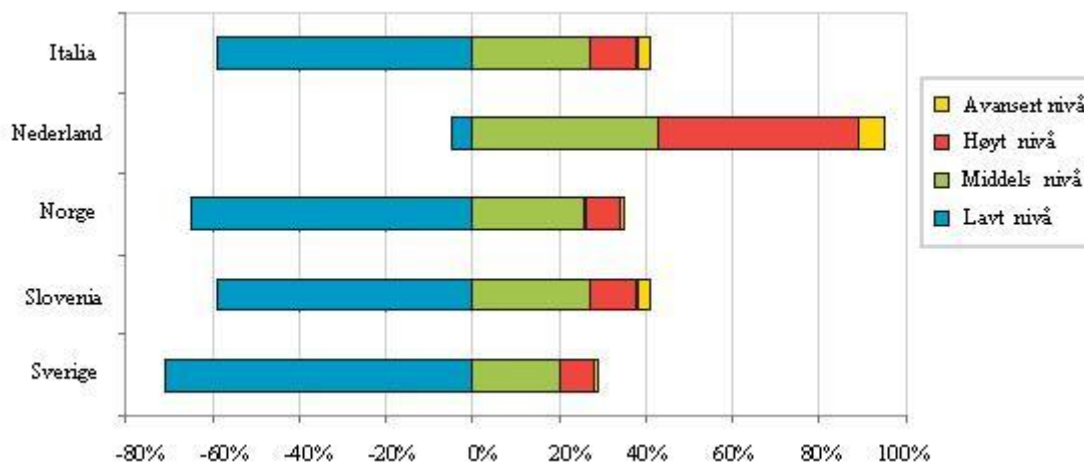
Figur 2.3 "Sammenhengen mellom landenes matematikkskår og andel av årskullet som tar fordypning i matematikk (dekningsgrad)" (Grønmo m.fl., 2010, s.16).

Som det er vist videre på figur 2.4 presterer både Norge og Sverige under det skalerte gjennomsnittet (forskjellen fra gjennomsnittet er signifikant for begge landene).



Figur 2.4. "Hovedresultater i matematikk for alle landene som deltok i TIMSS Advanced i 2008." (Grønmo m.fl., 2010, s.15).

Fordelingen på kompetansenivåene blant elevene er også ganske lik i Norge og i Sverige og fremstår som "bemerkelsesverdig svak" (Grønmo m.fl., 2010), se figur 2.5.



Figur 2.5 "Fordelingen av elevene på kompetansenivåene i matematikk i TIMSS Advanced. Prosentandeler til høyre for null angir hvor mange prosent av elevene på kurset som når de internasjonalt definerte nivåene i TIMSS Advanced. Prosentandeler til venstre for null angir hvor mange prosent av elevene på kurset som ligger under laveste definerte kompetansenivå, altså på det vi har kalt lavt nivå." (Grønmo m.fl., 2010, s.50)

Som figur 2.5 viser at Norge har nesten ingen elever (bare 1 prosent) som presterer på avansert nivå og det er svært få av de norske elevene når høyt nivå, mens hele 65 prosent av dem ligger på lavt nivå. Norske elever presterer svakt på alle rapporteringsområdene i TIMSS Advanced 2008: både algebra, kalkulus og geometri.

Norge har sammen med Sverige den største tilbakegangen i prestasjoner sammenliknet med den første undersøkelsen TIMSS Advanced, og dette er samtidig som antall elever i matematikkspesialist-populasjonene har gått ned i begge landene. Norske elever har "gått fra å prestere i nærheten av det internasjonale, standardiserte gjennomsnittet til å prestere signifikant lavere enn dette" (Grønmo m.fl., 2010, s.58). Nedgangen i matematikkprestasjoner hos norske og svenske elever fra 1995/1998 samsvarer med resultater fra tidligere TIMSS-undersøkelser i matematikk av elever i grunnskolen som viste en markert tilbakegang i deres matematikkprestasjoner fra 1995 til 2003 (Grønmo m.fl., 2010; Skolverket, 2009). I den norske rapporten bemerkes det at elever som er undersøkt i TIMSS Advanced 2008 i Norge tilhører samme årskull elever som ble undersøkt på 8.trinn i TIMSS 2003. Det kan også nevnes at både i Norge og i Sverige ble det i løpet av 1990-tallene gjennomført store utdanningsreformer. For grundige analyser av prestasjonsresultater i TIMSS Advanced 2008 og de mulige årsakene til den negative utviklingen i Norge og i Sverige henvises det til Grønmo m.fl., 2010 og Skolverkets Huvudrapport, 2009.

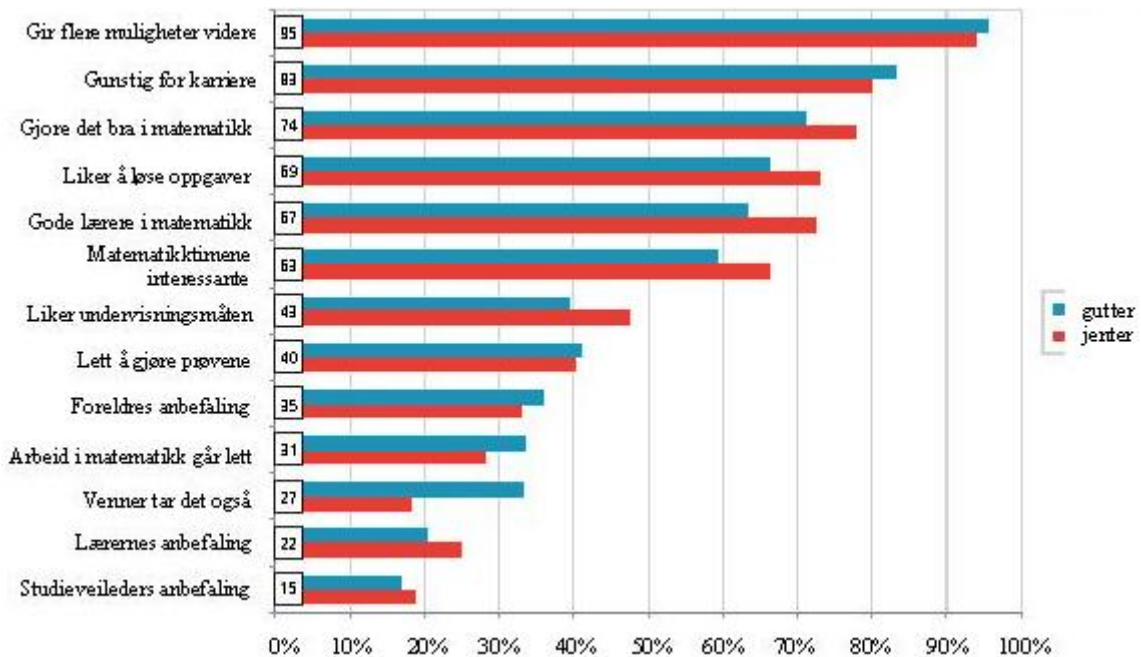
Det er ikke funnet signifikante kjønnsforskjeller i matematikkprestasjoner hos de norske elevene verken totalt sett på testen eller innenfor de rapporteringsområdene hver for seg (Grønmo m.fl., 2010).

2.4 Kort om elevspørreskjema og noen bakgrunnsvariabler

I etterkant av prøven besvarte elevene et omfattende spørreskjema. Spørsmålene i spørreskjemaet berører blant annet elevenes sosiokulturelle bakgrunn, undervisningssituasjon i matematikktimene, elevenes holdninger i matematikkfaget og videre studieplaner. Det er gjort mange interessante analyser knyttet til sammenhenger mellom elevenes prestasjoner på matematikktesten i TIMSS Advanced og deres svar på spørsmålene i spørreskjemaet, og resultatene av disse analysene er presentert i den norske rapporten (se Grønmo m.fl., 2010). Jeg har valgt her å presentere analyser av elevenes begrunnelser for valget av fordypning i matematikk, fordi datamaterialet knyttet til dette er brukt som bakgrunnsvariabler i analyser i denne oppgaven.

I TIMSS Advanced er undersøkelsen av faktorer som påvirker elevenes valg av fordypning i matematikk begrunnet med følgende: *”Since it is very important to attract these students to study advanced mathematics in the first place, and then to retain them for tertiary-level study of mathematics and a career involving mathematics, it is useful to know what factors attracted them to the study of mathematics.”* (Mullis, m.fl., 2009, s.135). I spørreskjemaet ble elevene bedt om å ta stilling til en rekke utsagn som beskriver ulike grunner for å velge fordypning i matematikk. For hver grunn skulle de angi om denne hadde vært ”veldig viktig”, ”viktig”, ”uviktig” eller ”veldig uviktig”. Disse utsagnene kan knyttes til tre hovedkategorier av fagvalg beskrevet av I.Ramberg (2006): profesjonsorientert fagvalg, helgardert fagvalg og interessebasert fagvalg. (Grønmo m.fl., 2010). Profesjonsorientert og helgardert fagvalg baseres på elevenes planer om bestemte utdannings- og yrkesplaner og ønsker om å holde mange muligheter åpne for videre studier og yrkesliv. Interessebasert fagvalg bygger på faglige interesser, og elever som begrunner valg av fordypning i matematikk med interesse for faget kjennetegnes ved høy faglig selvtilitt og mestringsoppfatning, og de har gode erfaringer med matematikk fra tidligere skolegang. Norske elevenes svar på påstandene om valgbegrunnelse tyder på at mange 3MX-elever har gjort interessebasert fagvalg (Grønmo m.fl., 2010). Over 60 prosent av elevene begrunnet fagvalget sitt med faglig interesse (”Jeg

liker å løse matematiske problemer” - 69 % og ”Matematikktimene er interessante” - 63%). Det var også mange elever som har baserte fagvalget på faglig selvtillitt og mestringsoppfatning (”Jeg gjør det vanligvis bra i matematikk” - 74 %, ”Arbeidet i matematikk går lett” - 31%).



Figur 2.6 Prosentandelen av gutter (blått) og jenter (rødt) i 3MX som oppgir at de ulike grunnene har vært ”Veldig viktig” eller ”Viktig” for deres valg av matematikk. Tallene i rammer er tilsvarende prosentandelen hvis vi ikke skiller mellom gutter og jenter. (Grønmo m.fl., 2010, s.199)

Analyser av sammenhenger mellom norske elevers begrunnelser for valg av matematikk og deres prestasjoner på TIMSS-testen i matematikk viser at ”norske elever som presterer bra på matematikktesten i TIMSS Advanced, kjennetegnes ved å ha en positiv holdning til faget og undervisningen, ved å ha en høy selvoppfatning i faget” (Grønmo m.fl., 2010, s.204).

2.5 Måling av testmotivasjon i TIMSS Advanced

Spørsmål om testmotivasjon er nytt i TIMSS Advanced 2008, sammenliknet med den tidligere undersøkelsen i 1995. Dette spørsmålet ble tatt med i elevspørreskjema i Slovenia, Russland, Sverige og Norge. I etterkant av matematikktesten ble elevene bedt om å angi i hvilken grad de var enige eller uenige med gitte påstander om deres motivasjon og innsats på testen. Nærmere informasjon om disse påstandene er gitt i kapittel 4.

Motivasjonsskalaen i TIMSS Advanced 2008 er forankret i prestasjonsmotivasjonsteori om forventning og verdisetting (som omtales nærmere i kapittel 3), og bygger på forskningen og de empiriske dataene knyttet til anvendelser av motivasjonsskala SOS¹¹ utviklet av Donna L.Sundre (1999), og flere skalaer av Hanna Eklöf (2006a, b; 2010a) utviklet og testet i forbindelse med studier av testmotivasjon hos både svenske elever i 8.klasse som deltok i TIMSS 2003 og svenske gymnaselever i 12.klasse som deltok i pilotstudien TIMSS Advanced 2008¹². Gjennom empiri og statistiske analyser av skalaene brukt i Sverige ble det vist at de hadde tilfredsstillende reliabilitet og validitet ved testmotivasjonsmålinger i low-stakes testkontekster (Eklöf, 2006b, 2010a), og basert på disse resultatene valgte enkelte deltakerland i TIMSS Advanced 2008 å inkludere en 4-punktsskala med seks påstander som måler elevenes motivasjon og innsats på testen, i deres elevspørreskjema. Foreløpige analyser av elevenes svar på spørsmål om deres motivasjon og innsats på testen er lagt fram bare i Sverige.

¹¹ SOS (The Student Opinion Scale) er utviklet og brukt ved James Madison University, USA, og består av 10 variabler likt fordelt på to faktorer: den ene måler testtakerens selvrapporert innsats og utholdenhet på testen og den andre måler testtakerens oppfatninger av hvor stor betydning testen har for han eller henne. Skalaen kan lastes ned fra <http://www.jmu.edu/assessment>. I sin helhet passer ikke SOS for tester hvor testresultatene ikke leveres tilbake til testtakerne. (Eklöf, 2010a)

¹² Pilotstudien ble gjennomført blant elever i 12.klasse (i aldersgruppen 18-19 år) ved 20 utvalgte svenske videregående skoler i forkant av hovedstudien TIMSS Advanced 2008. For nærmere omtale - se kapittel 4.

3 Teori

3.1 Motivasjonsteori

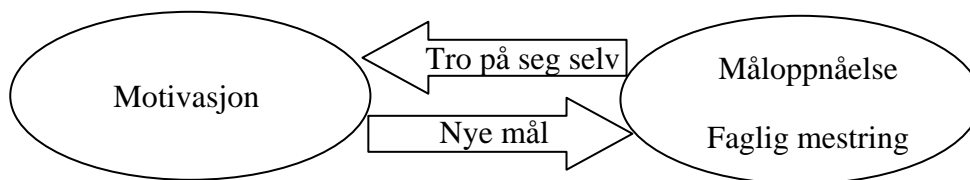
I begynnelsen av 1900-tallet var ikke studier av motivasjon et eget forskningsområde, slik som det er i dag (Schunk m.fl., 2010). På den tiden var motivasjon underordnet den nyetablerte disiplinen psykologi, og synet på motivasjon, i likhet med psykologien generelt, var forankret i filosofi. Motivasjon var begrepsfestet gjennom vilje/lyst og instinkt. Mange av de gamle teoriene om motivasjon regnes ikke som holdbare lenger, men flere teoretiske prinsipper innen den moderne motivasjonsteori kan spores tilbake til disse. Den moderne forskningen om motivasjon fokuserer på de underliggende mentale prosesser som bestemmer motivasjon og motivasjonens rolle i prestasjonskontekster. Utviklingen gjennom årene har ført til etablering av eget forskningsfelt som studerer motivasjon i utdanningsprosesser (Schunk m.fl., 2010).

Motivasjon er et teoretisk konstrukt, med andre ord en abstrakt variabel, som defineres gjennom teori og observasjoner, avhengig av det valgte teoretiske perspektivet og den empiriske konteksten (Eklöf, 2010b). For eksempel, i det generelle kognitive perspektivet understrekes betydningen av kognitive (mentale) prosesser, oppfatninger, affekter og verdier for motivasjon hos et individ. Det finnes mange forskjellige konseptuelle modeller og teoretiske perspektiver i motivasjonsteori: teori om forventning og verdisetting (forventning-verdi teori), teori om veloverveid handling (attributt-teori), sosial-kognitiv teori, teori om indre og ytre motivasjon, teori om målorientering og affekt-teori. Selv om viktige forskjeller skiller mellom disse teoriene, bygger de på samme forutsetninger: motivasjon er en mental prosess; motivasjon er ikke synonym med prestasjon, men det eksisterer et gjensidig avhengighetsforhold mellom disse; motivasjon forandrer seg med individets utvikling; motivasjon gjenspeiler individuelle og kulturelle forskjeller, samt forskjeller mellom grupper (Schunk m.fl., 2010).

Motivasjonsteori beskriver motivasjon som en aktiv prosess, rettet mot et bestemt mål. Denne prosessen er viktig for å kunne sette i gang og opprettholde en aktivitet som fører til måloppnåelsen. *“Motivation is the process whereby goal-directed activity is instigated and*

sustained.” (Schunk m.fl., 2010, s.4). Man kan sammenlikne motivasjon med et drivstoff som får et kjøretøy til å bevege seg i en bestemt retning (Eklöf, 2010b).

I læringskontekster eksisterer det et avhengighetsforhold mellom motivasjon på den ene siden og læring og prestasjon på den andre siden. Motivasjon fremmer læring og prestasjoner, samtidig som resultater av det eleven lærer og gjør, påvirker hans eller hennes motivasjon til å lære mer. Ved å lykkes med oppsatte mål i læringsprosessen og øke faglig mestring vil eleven kunne se at han eller hun har de nødvendige evnene til å lære. Dette vil gi eleven motivasjon til å sette opp nye mål for læring. (Schunk m.fl., 2010). En skjematisk illustrasjon av sammenhengen mellom motivasjon og læring/prestasjon kan uttrykkes i følgende figur:



Figur 3.1 Gjensidig avhengighetsforhold mellom motivasjon og læring/prestasjon.

Det er ikke mulig å måle motivasjonsprosessen direkte, vi kan bare måle forskjellige faktorer knyttet til denne prosessen (Eklöf, 2006b). Vi kan tyde motivasjon gjennom behavioristiske indikatorer som Pintrich, Schunk & Meece (2010) kaller for motivasjonsindekser. Det er definert fire slike indekser.

Tabell 3.1 Motivasjonsindekser (Schunk, Pintrich & Meece, 2010).

Indeks	Sammenheng med motivasjon
Oppgavevalg	I situasjoner hvor en elev kan velge fritt sine arbeidsoppgaver, indikerer elevenes valg at han eller hun har motivasjon for å utføre oppgavene
Innsats	God innsats, spesielt når det gjelder vanskelige oppgaver, tyder på motivasjon hos eleven
Utholdenhet	Bevaring av arbeidsinnsats over lengre tid, spesielt når oppgaven byr på utfordringer, forbindes med høyere nivå av motivasjon
Prestasjonsresultater	Valgmuligheter, innsats og utholdenhet øker prestasjon

Valg av arbeidsoppgaver peker ut elevens interesser. Når elever har valgmuligheter, vil deres valg av arbeidsoppgaver indikere hva som motiverer dem. Men det er ofte slik at vi ikke kan benytte denne indeksen, fordi i klasserommet er arbeidsoppgaver ofte bestemt på forhånd. Det gjelder spesielt testsituasjoner.

Den andre motivasjonsindeksen er *innsats*. God innsats, spesielt når det gjelder vanskelige arbeidsoppgaver, indikerer motivasjon hos elever. Men her må vi ta i betraktning at effekten av denne indeksen kan "forstyrres" av dyktighetsnivå hos elever: jo mer dyktig eleven er, jo desto enklere det er for han eller henne å prestere bedre med mindre innsats. Dermed er nytteverdien av innsats som motivasjonsindeks er begrenset.

Den tredje motivasjonsindeksen er *utholdenhet*. Utholdenheten er viktig for å lykkes med en bestemt oppgave, fordi det krever å legge ned en viss mengde arbeid i det, samtidig som resultatet ikke nødvendigvis kommer raskt. Arbeid over lengre tid, uten å gi opp selv om man støter på problemer og utfordringer, forbindes med høyere motivasjonsnivå.

Den siste, og den indirekte motivasjonsindeksen er elevens *prestasjon*. Mange studier har påvist positiv sammenheng mellom prestasjon og de tre andre motivasjonsindekser: valg av oppgaver, innsats og utholdenhet. Elever som engasjerer seg i en bestemt oppgave, yter innsats og viser utholdenhet i sitt arbeid, vil mest sannsynlig oppnå bedre resultater enn elever som ikke gjør det (Schunk m.fl., 2010 ref. Pintrich, 2003), og dårlige prestasjoner kan indikere lav motivasjon og innsats hos elever.

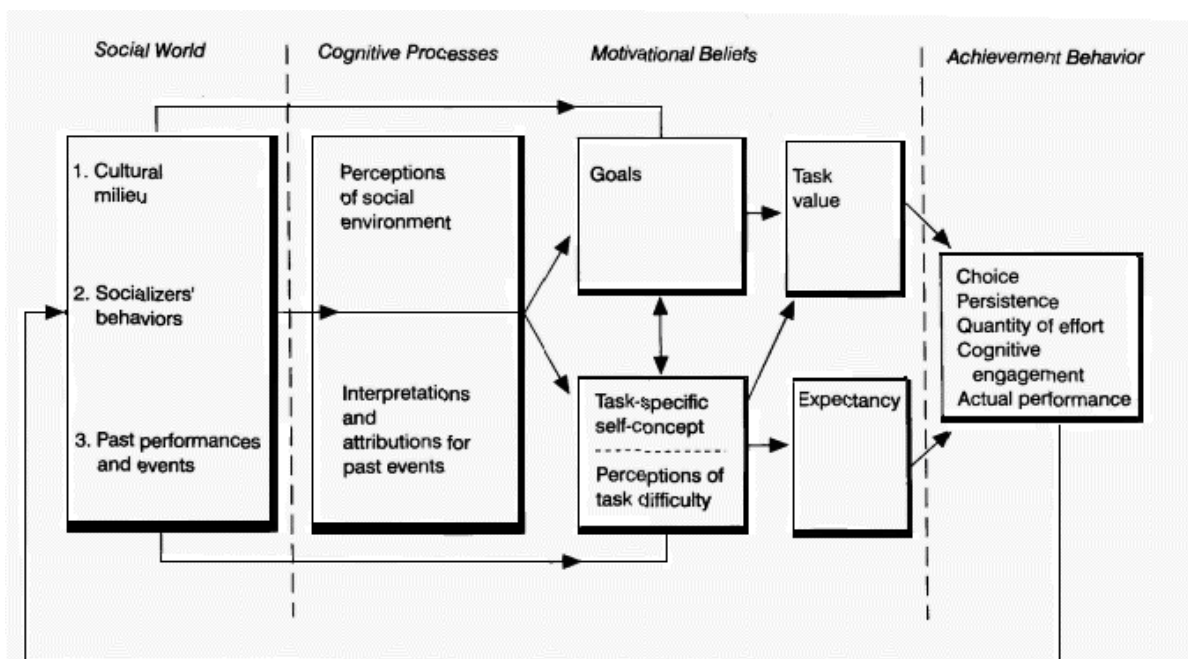
Motivasjon knyttet til forskjellig prestasjonsrelatert atferd, kalles for prestasjonsmotivasjon (Wigfield & Eccles, 2002).

3.1.1 Teori om forventning og verdisetting

"Achievement motivation theorists attempt to explain people's choice of achievement tasks, persistence on those tasks, vigour in carrying them out, and quality of task engagement." (Wigfield & Eccles, 2002, s.1).

Mange teorier om prestasjonsmotivasjon ser på motivasjon i et kognitivt perspektiv, som forutsetter at sammensettingen av individets kognitive faktorer påvirker hans/hennes motivasjon. De mest moderne teoriene går videre og trekker inn også det sosiokulturelle perspektivet. Disse teoriene erkjenner rollen sosiokulturelle kontekster har i utforming av

individets motivasjon (Eklöf, 2006b). Mange av disse omfatter forventning og verdisetting (Schunk m.fl., 2010), men den mest anvendte av dem i dag har opphav i arbeidet til Eccles, Wigfield og deres kolleger (Schunk m.fl., 2010; Eklöf, 2006b; Barry & Finney, 2009) og kalles for forventning-verdi teori (videre omtalt som FVT). Denne teorien har ganske omfattende teoretiske rammer som binder sammen mange psykologiske og sosiokulturelle komponenter med utgangspunkt i antakelsen om at individets forventninger og verdier utformes i en sosial kontekst. En skjematisk modell av FVT til Eccles og Wigfield er vist på figur 3.2.



Figur 3.2 A social cognitive expectancy-value model of achievement motivation (Schunk, Pintrich & Meece, 2010, s.51)

Teoriens paradigme sier at innsatsen som individer yter for å løse en bestemt oppgave, er påvirket av to faktorer: (1) deres forventninger til suksess (forventningsfaktor), og (2) individets vurdering av oppgavens verdi (verdifaktor).

Forventningsfaktoren, kalt "Expectancy" i figur 3.2, defineres av spørsmålet "Hvor sannsynlig er at jeg klarer å løse denne konkrete oppgaven?" (Schunk m.fl., 2010 ref.: Eccles, 1983, 2005; Pintrich 1988; Wigfield, 1994 og Wigfield & Eccles, 1992). Denne faktoren er sterkt relatert til vurdering av egne evner (Eklöf, 2010b) og er bestemt av individets selvoppfatning og hans/hennes forventning til å lykkes.

Individets selvoppfatning er synet på egen kompetanse og består av to komponenter: mestringsforventning og selvvurdering (Kjærnsli, 2007). *Mestringsforventning* er situasjonsbetinget og gjelder forventning om å lykkes i den konkrete utfordringen individet står overfor. (Schunk m.fl., 2010; Wigfield & Eccles, 2002). ”Elevens forventning om å mestre et spesifikt problem har avgjørende betydning for den innsatsen og utholdenheten som investeres. Forskning viser at det er en sammenheng mellom mestringsforventning og elevenes prestasjoner på et spesifikt fagområde.” (Kjærnsli, 2007, s.11). Selvvurdering er domenebetinget og gjelder vurdering av sin egen kompetanse innen et bestemt fagområde¹³. (Schunk m.fl., 2010 ; Wigfield & Eccles, 2002). Positiv selvvurdering kan sees som et svært ønskelig resultat, og oppnåelse av dette kan i seg selv være et mål for utdanningen (Kjærnsli, 2007 ref. Branden, 1994). Selvvurdering kan ha flere dimensjoner: elevene vurderer egen prestasjon ved å relatere seg både i forhold til andre elever og i forhold til hvordan de selv presterer i andre skolefag. Det er også slik at deres selvvurdering på forskjellige områder kan være relativt uavhengige av hverandre. (Kjærnsli, 2007 ref.: Marsh & Shavelson, 1985 og Marsh, Byrne & Shavelson, 1988).

Verdifaktoren til innsatsen som ytes for å løse en bestemt oppgave (”Task value” i figur 3.2), bestemmes av spørsmålet ”Ønsker jeg å gjøre denne oppgaven, og hvorfor gjør jeg det?” (Schunk m.fl., 2010 ref.: Eccles, 1983, 2005; Pintrich 1988; Wigfield, 1994 og Wigfield & Eccles, 1992, 2002). Verdisetting av oppgaven har fire komponenter: (1) *oppnåelsesverdien*: verdien av evt. oppnådd suksess, det vil si hvor viktig det er å gjøre bra; (2) *den indre verdien*: den subjektive interessen i oppgaven; (3) *nytteverdien*: vurdering av oppgavens nytte; (4) *kostnadsverdien*: vurdering av påkrevet innsats (Wigfield & Eccles, 2002). FVT predikerer at når testtakere ikke ser en verdi i testen de skal gjennomføre, vil verdifaktoren tilsi en lav motivasjon og innsats på testen (Wise & DeMars, 2005; Thek m.fl. 2009).

Prestasjonsmotivasjon kan måles på forskjellige generaliseringsnivåer. Det mest vanlige er å måle domenebestemt motivasjon, det vil si motivasjon innen et bestemt fagområde, for eksempel i matematikk eller i naturfag. Den andre typen prestasjonsmotivasjon er situasjonsbestemt, og den måler motivasjon for å gjøre sitt beste i en bestemt situasjon. Empirisk forskning har vist at de teoretiske rammene som er omtalt ovenfor er holdbare både i studier av domenebestemt og situasjonsbestemt prestasjonsmotivasjon (Eklöf, 2010b).

¹³ I rapporten ”Matematikk i motvind brukes det både ”faglig selvoppfatning” og ”faglig selvillitt” om synet på egen kompetanse i faget. (se Grønmo, Onstad & Pedersen, 2010)

3.2 Testmotivasjon

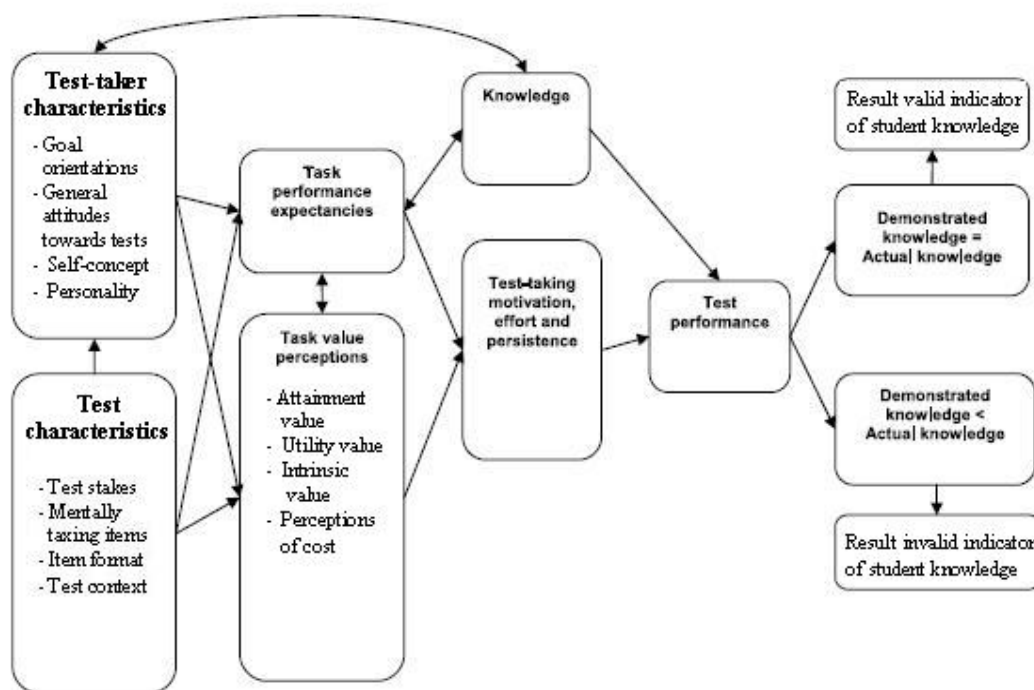
Testmotivasjon er situasjonsbestemt i det den er knyttet til en gitt utfordring, eller en bestemt test (Eklöf, 2010b). Baumert og Demmrich (2001) definerer testmotivasjon som viljen til å engasjere seg i arbeid med testoppgavene og deretter vise innsats og utholdenhet i dette arbeidet. Dette kan utdypes videre. Som det er nevnt tidligere, beskrives motivasjon som en aktiv prosess, rettet mot et bestemt mål. Den målorienterte aktiviteten i en testsituasjon består av å gjøre sitt beste med sikte på å vise sine kunnskaper og ferdigheter innen det som testes. Testmotivasjon skal igangsette og opprettholde denne aktiviteten slik at testen gir muligheten til å fastslå med validitet ens sanne ferdigheter og kunnskap (Wise & DeMars, 2003).

Det er flere psykologiske faktorer ved deltakelse på en test som for eksempel generelle holdninger; holdninger til faget og skoletester, både til tester generelt og til den aktuelle testen; selvoppfatning; personlige mål og egenskaper; angst. Basert på disse personrelaterte faktorene og testens parametre (som vanskelighetsgrad, utforming av oppgavene, teststatus osv), vil testtaker gjøre et grovt overslag og stille seg spørsmål som: ”Hvor sannsynlig er det at jeg klarer å lykkes på denne testen?”, ”Er dette en viktig test for meg?”, ”Hvilke konsekvenser kan testresultater ha for meg?”, ”Virker testen krevende?”, ”Virker denne testen spennende?”, ”Hvor mye innsats kreves det for å gjøre det bra på testen?”, ”Hvor mye innsats er det verdt å legge inn i arbeid med testen?”. Testtakerens oppfatninger rundt disse spørsmålene vil påvirke hans eller hennes innsats og utholdenhet på testen (Eklöf, 2010b).

Som det er nevnt tidligere i avsnitt 3.1, er innsats og utholdenhet definert som motivasjonsindekser, og det eksisterer en sammenheng mellom disse indeksene og prestasjon (se tabell 3.1). Samspillet mellom testtakerens primære kunnskaper og den kognitive evnen til å omsette informasjon på den ene siden, og hans/hennes testmotivasjon på den andre siden, vil påvirke testtakerens prestasjoner på testen (Baumert & Demmrich, 2001; Eklöf, 2010b). Empirisk forskning har vist at høyere vanskelighetsgrad på oppgaver i testen er forbundet med lavere testmotivasjon hos testtakere (Barry & Finney, 2009 ref.: Bovaird, 2002, Wise, 2006 og Wolf m.fl., 1995). I FVT's perspektiv begrunnes denne sammenhengen med at vanskelige oppgaver krever mer av elevene, som fører til høyere kostnadsverdi og lavere forventning om suksess, og dermed lavere testmotivasjon hos elevene. Gjennom empiri er det også påvist at testtakere har tendens til å ha lavere testmotivasjon når de skal ”produsere” et svar i åpne oppgaver fremfor å velge et riktig svaralternativ i flervalgsoppgaver, fordi åpne

oppgaver har en tendens til å måle høyere nivå av de kognitive sammensettingene, og dermed krever mer av elevene. Dette fører igjen til lavere forventning om å lykkes, som igjen påvirker elevens testmotivasjon. (Barry & Finney, 2009 ref.: DeMars, 2000 og Sundre & Kitsantas, 2004). Manglende motivasjon på testen kan føre til testresultater som underestimerer testtakerens kompetansenivå (Wise, 2009).

Basert på FVT og den eksisterende teoretiske og empiriske forskningen om testmotivasjon, har Eklöf (2010b) tegnet en modell, som viser antatte sammenhenger mellom testtakerens evner og egenskaper, testens parametre, testmotivasjon og testresultater (Fig.3.3).



Figur 3.3. Factors in the assessment situation affecting student test-taking motivation and the effect of test-taking motivation on performance (Eklöf, 2010b, s.348).

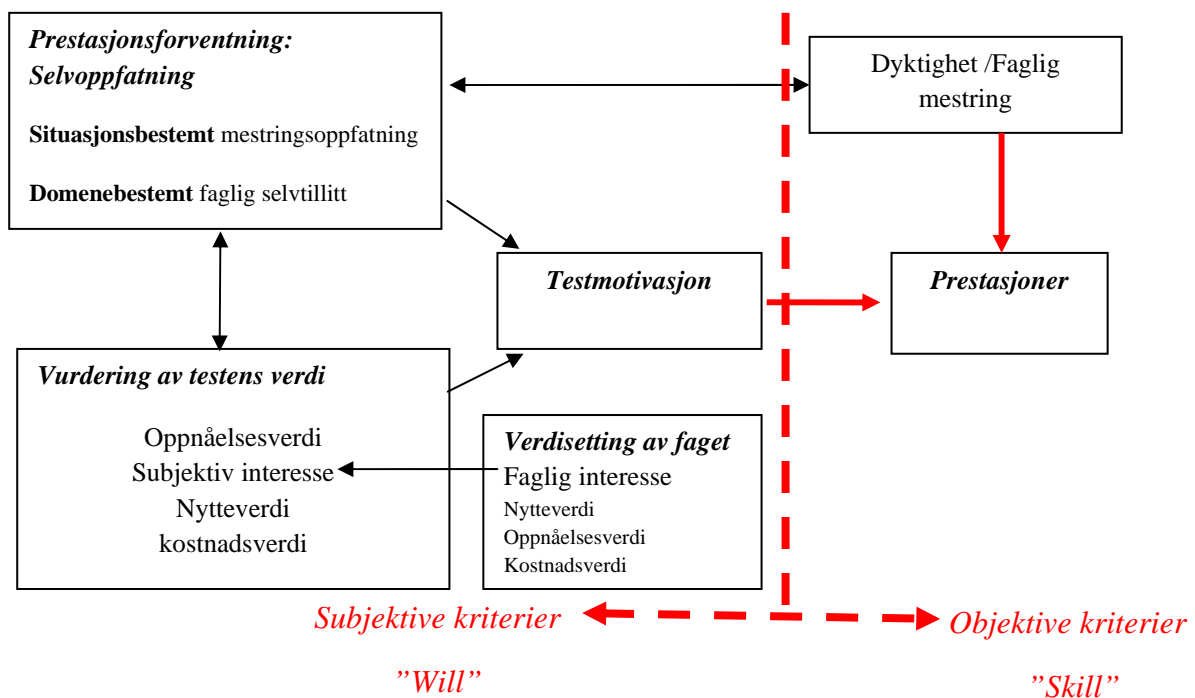
Vi skal se nærmere på denne modellen med utgangspunkt i teststatus. Som det er forklart i avsnitt 1.2.1 har low-stakes tester ingen direkte konsekvenser for testtakerne. Det vil mest sannsynlig føre til at oppnåelsesverdien og nytteverdien av testen blir vurdert som lave av testtakerne. Hvis kostnadsverdien på samme tid blir vurdert som høy, vil det komme i konflikt med de førstnevnte verdifaktorene. Denne uoverstemmelsen vil åpenbart påvirke vurderingen av testens indre verdi i negativ retning (Eklöf, 2010b), og vi kan anta at testtaker ikke vil verdisettesten høyt. I perspektivet av forventning-verdi teori vil verdifaktoren i en slik

situasjon tilsi lav motivasjon og innsats på testen. Altså, forbindelsen mellom verdifaktoren ("Task value perceptions" på fig.3.3) og motivasjon på testen er veldig sterk (Eklöf, 2010b), samtidig som verdifaktoren er sterkt relatert til teststatus.

Forbindelsen mellom forventningsfaktoren ("Task performance expectations" på fig.3.3), testmotivasjon og teststatus synes å være mer komplisert. Sammenliknet med verdifaktoren er forventningsfaktoren ikke like åpenbart relevant for situasjonsbestemt motivasjon (Eklöf, 2010b). Som det ble påpekt tidligere har elevenes prestasjonsresultater sammenheng med subjektive faktorer som motivasjon, faglig selvtilitt og verdisetting i både læringskontekster og i testkontekster. På tester måles elevenes kunnskaper i faget gjennom deres prestasjonsresultater ved hjelp av objektive kriterier (som testskår og lignende). For å kunne trekke gyldige konklusjoner om prestasjonsresultater på testen må vi vite i hvor stor grad vi har testet de reelle kunnskapene hos elever. *"Students need to have both the 'will' and the 'skill' to be successful in classrooms. Clearly, students with the skill but lacking the will to diligently complete assessment tasks will produce gross underestimates of student abilities and learning growth over time"* (Sundre & Moore, 2002, s.3). Med utgangspunkt i modell på figur 3.3 kan vi illustrere sammenhengene mellom de subjektive og objektive kriteriene i vurderingskontekster (se figur 3.4).

Som sagt kan mangel på viljen hos elevene til å yte sitt beste på testen true validiteten av testresultater. For å kunne fange opp den eventuelt manglende testmotivasjon er det viktig å få vite om den gitte situasjonskonteksten har gitt utslag på motivasjonsfaktorer som påvirker denne viljen. Når motivasjonsmålinger er begrunnet med bekymringen for at den lave teststatusen kan føre til lav testmotivasjon, kan det reises spørsmål om i hvilken grad kan prestasjonsforventning som motivasjonsfaktor bli påvirket av testkonteksten og om elevenes prestasjonsforventning ville være annerledes om testen for dem hadde vært "high-stakes" i stedet for "low-stakes". Studier har vist at i praksis er det vanskelig å skille mellom situasjonsbetinget mestringsforventning (forventning om å lykkes i den konkrete faglige utfordringen) og domenebestemt faglig selvtilitt, selv om teoretisk er disse begrepene adskilt fra hverandre (Eklöf, 2010b). Dessuten er det også vist at *"examinees appear to be willing to put forth effort in low-stakes contexts, but are more willing on tests that are not exceedingly difficult; at some point, if the test is too difficult, effort decreases"* (Barry & Finnley, 2009, s.33). Sammenhengen mellom elevens kognitive evner og faglig mestring på den ene siden og oppgavens format og vanskelighetsgrad på den andre siden synes å være avgjørende for

elevens vilje til å yte innsats på testen samtidig som testen er ment til å måle nettopp den samme sammenhengen. Dette gjenspeiles i figur 3.4, hvor faglig mestring, prestasjonsforventning, testmotivasjon og prestasjonsresultater danner en ”lukket ring” samtidig som vi ved motivasjonsmålinger er vi interessert i å skille ”det subjektive” fra ”det objektive”.



Figur 3.4 Motivasjon og prestasjon sett i sammenheng med faglig selvtillitt og faglig interesse. Denne modellen er rammet inn i elevens primære kognitive evner og personlige karakteristikk og situasjonskontekst (”test-taker characteristics” og ”test characteristics” på figur 3.3).

De uthevede linjene er knyttet til validitesproblematikken innen resultattolkninger i testkontekster (som er omtalt innledningsvis i oppgaven).

Det ovennevnte synes å vise at selv om faglig selvtillitt og mestringsforventning har i følge prestasjonsmotivasjonsteori sin rolle i utformingen av prestasjonsmotivasjon hos elevene, kan det være behov for videre oppklaring av FVT’s faktormodellen for testmotivasjonskonstruktet når det gjelder elevenes holdninger og prestasjon i low-stakes testkontekster. For eksempel har Barry & Finnley i sin studie av testmotivasjon blant collegestudenter konkludert med at “personality characteristics may be more useful than motivational variables (e.g., expectancies, goal orientation) in explaining motivation in low-stakes context” (Barry & Finnley, 2009, s. 35).

3.2.1 Måling av testmotivasjon

Som det er sagt tidligere i dette kapitlet, er motivasjon en abstrakt variabel, som defineres gjennom teori og observasjoner, avhengig av det valgte teoretiske perspektivet og den empiriske konteksten. ”A valid assessment of a construct is an assessment that sufficiently captures the construct of interest while not measuring other, irrelevant, constructs” (Eklöf, 2006b ref. Messick, 1989). Begrepsfesting, operasjonalisering, måling og tolkning av de målte resultatene står sentralt i forskningen om motivasjon (Schunk m.fl., 2010).

Motivasjon kan måles på flere måter. Mest vanlig er å måle motivasjon enten gjennom direkte observasjon av behavioristiske indekser (se tabell 3.1 i avsnitt 3.1) eller gjennom verbal karakterisering. Verbal karakterisering kan gjøres gjennom selvrapportering og/eller andre personers vurdering. (Schunk m.fl., 2010; Eklöf, 2010b). Alle disse målemetodene har sine svakheter (Eklöf, 2010b). Testmotivasjon er situasjonsbestemt prestasjonsmotivasjon, og valg av målemetoder vil være avhengig av rammene rundt testsituasjonen.

Direkte observasjoner av behavioristiske indekser fokuserer på hvordan innsats, utholdenhet og prestasjoner kommer til uttrykk gjennom atferd, men samtidig overser de underliggende kognitive og affektive prosesser. Denne målemetoden har flere svakheter og synes å være mindre egnet til å måle testmotivasjon. For det første tar manglende fokus på kognitive og affektive prosesser bort muligheten til å forstå nettopp de underliggende faktorene for motivasjon, noe som opptar mange forskere som studerer motivasjon (Schunk m.fl., 2010). For det andre kan direkte observasjon lett forveksle innsats, utholdenhet og prestasjon med andre variabler som har en betydning i den gitte situasjonen og som ikke er så lett å gardere seg mot (Eklöf, 2010b). For eksempel ved en prøve eller test kan en slik variabel være elevens dyktighetsnivå eller hvor fort eleven pleier å jobbe i testsituasjoner. I tillegg kan man påpeke at denne metoden kan bare brukes på få elever om gangen og ikke vil egne seg for å måle motivasjon i for eksempel store undersøkelser som TIMMS.

Selvrapportering fanger opp individets meninger og uttalelser om seg selv. Ved selvrapportering benyttes det spørreskjemaer, intervjuer, stimulerte tilbakeblikk, høyttenkning og dialoger. Vi kan ta for oss en kort beskrivelse av disse målemetodene.

Spørreskjema inneholder spørsmål eller påstander med på forhånd gitte svaralternativer (prekodete skjemaer). Som eksempler på slike spørreskjemaer kan det nevnes the Motivated Strategies for Learning Questionnaire (Schunk m.fl., 2010 ref. Pintrich, Smith, Garcia &

McKeachie 1993) og The Student Opinion Scale (SOS) (Sundre & Moore, 2002; Thelk m.fl.2009 ref. Sundre, 1997). SOS inneholder 5 spørsmål om testtakerens innsats på testen og 5 spørsmål om hvor viktig testen synes å være for testtakeren, og dette spørreskjemaet er blitt utviklet for å måle testtakernes motivasjon i low-stakes tester. Selv om det er vist at SOS skala har høy reliabilitet og validitet, er det ikke egnet for å måle motivasjon og innsats på tester hvor testresultater ikke utleveres til testtakerne (som det er tilfelle med TIMSS-testene) (Eklöf, 2010b). Prekodete spørreskjemaer er den vanligste måten å samle inn kvantitative data på. De gir muligheten til se på likheter og variasjoner i svarene, undersøke utbredelse av fenomener og generalisere resultater fra utvalg til populasjon (Johannessen m.fl., 2009).

Intervju er en samtale med en struktur og et formål (Johannessen m.fl., 2009 ref. Kvale, 1997). Intervjuer kan brukes som supplerende metode ved for eksempel en storskala undersøkelse for å utdype spesifikke problemstillinger i undersøkelsen.

For å bruke *stimulerte tilbakeblikk*, må elevenes arbeid filmes først. Videre skal videoopptaket brukes til å få elevene til å forsøke å forklare deres tanker ved forskjellige tidspunkt. Forskeren velger hvilke tidspunkt det gjelder og stopper videoopptaket underveis for å stille spørsmål som vil hjelpe elevene med å rekonstruere sine tanker. En slik metode gir muligheten til å belyse spesifikke situasjoner og prøve å forstå de underliggende prosessene, men ulempen ved den er at den er avhengig av respondentenes hukommelse. Derfor er det spesielt viktig å ikke la tiden gå for lenge mellom opptaket og rekonstruksjonen (Schunk m.fl., 2010).

Høyttenkning refererer til elevenes verbale kommentarer om deres tanker, aktiviteter og følelser mens de jobber med en oppgave. Forskeren er interessert i hvordan de kognitive strategiene forandres i arbeidsprosessen, og hvordan det påvirker motivasjon.

Måleresultatene i de to siste metodene er avhengige av forskerens tolkning av den ”selvrapporterte” motivasjonen, hvilken kan være kilde for resultatenes feiltolkning.

Selvrapportering er som regel lett å administrere og gjennomføre. Problemer kan oppstå når resultatene skal tolkes. For det første er resultatenes validitet avhengig av måleskalaens reliabilitet (Schunk m.fl., 2010). Selvrapportering gjenspeiler ofte subjektive oppfatninger som kan være vanskelig å vurdere i forhold til objektive kriterier (Eklöf, 2010b). For det andre er det viktig at respondentene svarer ærlig, at de forstår deres oppgaver og spørsmålene

som stilles, og at de er i stand til gjøre og svare på det de er bedt om (Schunk m.fl., 2010; Eklöf, 2010b). For eksempel vil elever som ikke var motivert til å gjøre sitt beste på en bestemt test trolig være lite motivert til å yte innsats og svare ærlig på spørsmål om motivasjon (Eklöf, 2010b). Det kan også tenkes at elever som opplevde testen som vanskelig og/eller hadde lav selvoppfatning, i etterkant av testen vil prøve å bortforklare de eventuelt dårlige testresultatene de forventer å få med lav motivasjon og dårlig innsats på testen. Men til tross for de nevnte svakhetene har empirisk forskning vist at selvrapporing gir en god indikasjon på elevens motivasjon (Eklöf, 2010b ref. Wise, Wise & Bhola, 2006).

Andres vurdering kan gjøres gjennom et vurderingsskjema som fylles ut av en observatør som for eksempel lærer, forelder eller forsker. Observatøren skal gi sin vurdering av forskjellige aspekter som indikerer hvorvidt eleven er motivert. Det er en relativt mer objektiv målemetode sammenliknet med selvrapporing. Dessuten skal en slik vurdering forsøke å fange opp de underliggende prosessene som påvirker motivasjon hos elevene, noe som ikke kan gjøres ved direkte observasjoner. Svakheten i denne målemetoden er at den forutsetter at observatøren husker alt elevene gjør under observasjonen. På grunn av den selektive naturen av hukommelsen, kan denne målemetoden for motivasjon føre til ukorrekte slutninger (Schunk m.fl., 2010).

Tatt i betraktning de ovennevnte fordelene og svakhetene til de forskjellige målemetodene for motivasjon, kan det være hensiktsmessig å bruke kombinasjon av disse for eksempel ved å bruke observatører i tillegg til selvrapporing ved hjelp av spørreskjemaer og intervjuer. En slik kombinasjon ble for eksempel brukt i PISA 2009 (Kjærnsli & Roe, 2010).

Som det er sagt tidligere er motivasjon en latent konstrukt som kan forstås og konseptualiseres gjennom forskjellige teorier, avhengig av hvilke psykologiske prosesser som det fokuseres på. Det er også slik at teoretiske rammer som er gyldige i bestemte kontekster ikke nødvendigvis kan anvendes i noen andre kontekster (Eklöf, 2010b ref. Kane, 2006).

Vurderingen av de psykometriske egenskapene til den valgte målemetoden har altså en sentral plass i validitetsvurdering av motivasjonsmålinger (Schunk m.fl., 2010; Eklöf, 2006b, 2010b). Selv om det er i en viss grad referert til empirisk forskning underveis i dette kapitlet, vil jeg avslutningsvis gi en liten oversikt over noen tidligere resultater fra empirisk forskning av testmotivasjon.

3.2.2 Noen resultater fra tidligere empiriske studier av testmotivasjon

Resultater av mange empiriske studier har gitt støtte til en teoretisk begrunnet antakelse om en eksisterende sammenheng mellom teststatus, testmotivasjon og prestasjon på testen.

For eksempel har Wolf & Smith (1995) i en studie med eksperimentell design vist at heving av testens status førte til økning av testmotivasjon hos collestudenter, og økningen tilsvarte en forskjell med effektstørrelse på 1,45, og det var også registrert en økning av testskår med effektstørrelse på 0,26. Sammenheng mellom rapportert testmotivasjon og prestasjon hadde korrelasjonskoeffisient på 0,35 når testen var betraktet av studentene som high-stakes test, mens tilsvarende korrelasjonseffekt under "low-stakes" test var på 0,23. Wise & DeMars (2005) har gjennom oppsummering av resultater fra en serie på 12 studier vist at forskjellen i testprestasjoner mellom motiverte elever og ikke-motiverte elever på testen hadde effektstørrelse på en halv standardavvik i favør av velmotiverte elever. Deres studier var basert på en metode som "*involves either identifying or establishing groups of examinees that should logically differ in motivation, and then comparing their levels of test performance. For example, for one group of examinees test performance may contribute to a course grade, while for another group does it not.*" (Wise & DeMars, 2008, s. 4)

Barry & Finnley (2009) har undersøkt om det finnes bestemte profiler av testtakere blant collestudenter ved low-stakes tester. Det har konkludert med at i motsetning til studentenes prestasjonsforventning (som hadde relativt liten sammenheng med studentenes motivasjonsnivå) hadde testtakernes personlige karakteristikk, som for eksempel å være samvittighetsfulle, en stor betydning for deres innsats og motivasjon på testen (Barry & Finnley, 2009).

Studier har også vist at testkarakteristikk og oppgaveformat har en viss sammenheng med motivasjon og innsats på testen. For eksempel, er det vist at oppgaver av høy vanskelighetsgrad er assosiert med lavere testmotivasjon og innsats (Eklöf, 2010b ref. Wolf, Smith & Birnbaum, 1995), og at åpne oppgaver hører sammen med lavere motivasjon- og innsatsnivå hos testtakere i større grad enn flervalgsoppgaver (Eklöf, 2010b ref. DeMars, 2000).

Det er også blitt rapportert om kjønnsforskjeller i testmotivasjon (Eklöf, 2006b, Kjærnsli & Roe, 2010) og raseforskjeller i holdninger under testing (Eklöf, 2010b ref. Chan m.fl. 1997).

Til slutt må det også nevnes at noen studier har funnet at heving av teststatus ikke hadde noen betydning for prestasjoner på testen (Kjærnsli & Roe, 2010; Eklöf, 2010b).

Undersøkelser av testmotivasjon i internasjonale komparative studier (TIMSS og PISA)

I flere land har det vært gjennomført studier som skulle undersøke om testmotivasjon hos elevene på de store internasjonale undersøkelsene kunne forklare testresultatene i landet. Her er det en kort oversikt over disse studiene.

USA: For å undersøke om manglende testmotivasjon kunne forklare de svare resultatene blant amerikanske elever i TIMSS 2003, ble det gjennomført en undersøkelse hvor hensikten var å undersøke i hvilken grad belønning kan påvirke elevenes prestasjoner. Undersøkelsen var basert på frigitte TIMSS-oppgaver, og elevene ble premiert med 10 dollar for hver oppgave de klarte å løse riktig. Resultatet av undersøkelsen viste at pengegaven hadde ingen effekt, og det ble konkludert med at manglende testmotivasjon på TIMSS-testen kunne ikke være en mulig årsak til det svake resultatet blant amerikanske elever i TIMSS 2003 (Eklöf, 2006b ref. O'Neil m.fl., 2004).

Tyskland: Som følge av kritikken mot storskalaundersøkelser hvor det ble hevdet at tyske elever ikke ville være motivert for å gjøre sitt beste på tester hvor de ikke fikk karakter, har Baumert & Demmrich (2001) undersøkt om heving av teststatus kan føre til forbedring av testresultater ved en eksperimentell studie hvor 467 elever i 9.klasse ble bedt om å løse enkelte matematikkoppgaver fra PISA-testen under fire ulike betingelser. Den ene elevgruppen fikk en introduksjon tilsvarende den som ble gitt til elevene under PISA-testen. De tre andre fikk enten informative tilbakemeldinger på sine prestasjoner, karakter på testen eller en pengegave. Hovedfunnet i studien var at de tre sistnevnte betingelsene (som økte teststatus overfor elevene) ikke hadde påvirkning på elevenes testmotivasjon, og som følge av dette var det ikke funnet forskjeller i verken brukt innsats på prøven eller testresultater. Dermed ble det konkludert med at manglende testmotivasjon hos tyske elever i storskalaundersøkelser kunne ikke forklare elevenes dårlige testresultater (Baumert & Demrich, 2001).

Sverige: Testmotivasjon hos de svenske elevene ble testet i forbindelse med TIMSS 2003 i en studie som omfattet et utvalg på 350 av de totalt 4210 TIMSS-testtakere. Utvalget besto av

177 gutter og 173 jenter i 14 – 15-årsalder fra 18 tilfeldig valgt klasser. I forkant av TIMSS-testen besvarte elevene et eget spørreskjema kalt The Test-Taking Motivation Questionnaire som besto av 24 forskjellige spørsmål og påstander. Resultatet av studien viste at de fleste elever var ganske motiverte for å gjøre sitt beste på testen selv om testen ikke hadde personlig konsekvenser for hver enkelt elev, og sammenhengen mellom testskår og den rapporterte testmotivasjon hadde korrelasjonskoeffisient $r = 0,25$, $p < 0,05$. Det ble konkludert med at *”the results from this study suggest that the Swedish TIMSS 2003 results were not biased by low motivation among the students. Still, there was a positive and statistically significant relationship between motivation and performance, and further studies in a more international context are warranted.”* (Eklöf, 2010b, s.350).

Norge: I PISA 2009 ble det undersøkt elevenes testmotivasjon gjennom selvrapporing i elevspørreskjema og intervjuer i etterkant av testen, samt gjennom observasjoner fra gjennomføringen av testen. Analysene viste at norske elever var i stor grad motivert for å gjøre sitt beste på testen, og den undersøkte sammenhengen mellom testmotivasjon og matematiskskår på testen hadde korrelasjonskoeffisient $r = 0,20$, $p < 0,01$. Det ble konkludert med at funn i Norge er i tråd med forskningen i USA, Tyskland og Sverige (Kjærnsli & Roe, 2010).

Vi bør bemerke her at disse studiene omfattet grunnskoleelever i ca.15-årsalder. I TIMSS Advanced er testmotivasjonsmålinger foretatt for første gang i 2008, men bare i fire deltakerland – Norge, Sverige, Russland og Slovenia. Foreløpig er det bare i Sverige det er gjort noen analyser av testmotivasjonsmålingene på testen. I forkant av TIMSS Advanced 2008 ble det gjennomført en pilotstudie i Sverige, hvor hensikten var å teste testmotivasjonsskala i forkant av hovedstudien. Resultater fra denne studien er omtalt i kapittel 4 og 5.

4 Metode

4.1 Data

Datamaterialet som er benyttet i denne masteroppgaven består av norske elevers prestasjonsresultater fra matematikktesten TIMSS Advanced 2008 og deres svar på spørsmålene i elevspørreskjema som ble avgitt i etterkant av testen. I spørreskjemaet ble elevene bedt om blant annet å angi i hvilken grad de var enige eller uenige med gitte påstander om deres motivasjon og innsats på testen. Dette datamaterialet oppbevares i flere datafiler lagret i SPSS, PASW Statistics Data Editor, som jeg fikk utlevert fra ILS.

Elevene som ble testet i TIMSS Advanced 2008 i hvert deltakerland var trukket ut etter bestemte statistiske regler og prosedyrer, og utvalgene tilfredstilte representativitetskravene i TIMSS Advanced (mer om dette kan leses i Grønmo m.fl., 2010). Av de 240 aktuelle videregående skoler i Norge ble 120 trukket ut til å delta i matematikk, og de andre 120 skoler i fysikk. Det var 107 av de 120 utvalgte skolene som valgte å delta i matematikktesten. På de 107 skolene deltok til sammen 1932 3MX-elever i undersøkelsen (Grønmo m.fl., 2010). Datamaterialet i denne masteroppgaven omfatter elever som både har gjennomført matematikktesten og besvart elevspørreskjemaet i etterkant av testen. Dette utgjør 1912 norske elever som tok 3MX våren 2008 (i aldersgruppen 18-19 år).

I tillegg til det ovennevnte datamaterialet ble det i denne masteroppgaven brukt bearbejdede resultater fra to studier i Sverige: pilotstudie og hovedstudie TIMSS Advanced 2008. Pilotstudien ble gjennomført blant elever i 12.klasse (i aldersgruppen 18-19 år) ved 20 utvalgte svenske videregående skoler i forkant av hovedstudien. Blant disse skolene ble det valgt ut fem skoler til å delta i en undersøkelse av elevenes oppfatninger av testens betydning samt deres innsats og motivasjon for å gjøre sitt beste på testen. Denne undersøkelsen omfattet 163 svenske elever, alle definert som matematikkspecialister og/eller fysikkspecialister. Ved to av de fem utvalgte skolene har elevene gjennomført TIMSS-test i matematikk, og ved de tre andre skolene har elevene gjennomført TIMSS-test i fysikk (utvalget ble gjort av Skolverket¹⁴, som har ansvaret for gjennomføringen av TIMSS studier i Sverige). På bakgrunn av resultatene fra denne undersøkelsen har flere deltakerland i TIMSS Advanced

¹⁴ Den sentrale forvaltningsmyndighet for det offentlige skolevesenet i Sverige

2008 (Sverige, Norge, Slovenia og Russland) valgt å inkludere spørsmål om motivasjon og innsats i deres elevspørreskjema. Selve hovedstudien TIMSS Advanced 2008 i matematikk i Sverige omfattet 2244 matematikkspecialister og 2250 fysikkspecialister, som i 2008 var gymnaselever i 12.klasse (i aldersgruppen 18-19 år). Resultatene fra pilotstudien og foreløpig analyser av testmotivasjon hos svenske elever på selve TIMSS Advanced 2008 testen er rapportert i artiklene til Hanna Eklöf (2010a, 2010b).

4.2 Den psykometriske skalaen for konstruktet "Testmotivasjon" i TIMSS Advanced 2008

Testmotivasjon hos elevene i TIMSS Advanced 2008 ble målt ved selvrappotering i etterkant av matematikktesten. Elevene skulle svare på spørsmål "Angi i hvilken grad du er enig i påstandene nedenfor om din egen innsats på oppgavene i den undersøkelsen du har vært med på" som inneholder 6 utsagn med firepunkts Likert-skala fra 1 = "helt enig" til 4 = "helt uenig". De avgitte svarene kan behandles som om vi anvendte en intervallskala med de verdiene 1, 2, 3 og 4 som koder for de fire svaralternativene. De seks påstandene i spørreskjemaet er:

- P1. Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen.
- P2. Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene.
- P3. Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året.
- P4. Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige.
- P5. Jeg var motivert for å gjøre mitt beste på disse oppgavene.
- P6. Jeg kunne ha anstrengt meg mer da jeg arbeidet med disse oppgavene.

I disse påstandene kan vi gjenkjenne motivasjonsindekser omtalt i kapittel 3. De referer til innsats (påstandene P1, P2, P3 og P6) og utholdenhet (P4), og i tillegg gir P5 en direkte verbalisering av elevens motivasjon på testen.

Påstandene kan deles inn i to grupper. Den ene gruppen (videre kalt "Gruppe 1") har positivt betonet ordlyd, enten fonetisk som "Jeg gjorde mitt beste.." og "Jeg var motivert.." eller innholdsmessig, som "Jeg arbeidet uten å gi opp..". I denne gruppen finner vi påstandene P1, P4 og P5. Den andre gruppen (videre kalt "Gruppe 2") har negativt betonet ordlyd, enten fonetisk som "Jeg var ikke helt konsentrert..", "Jeg anstrengte meg mindre.." eller innholdsmessig som "Jeg kunne ha anstrengt meg mer". Påstandene P2, P3 og P6 tilhører gruppe 2.

Responskalaen "virker" på disse to gruppene i motsatte retninger: valg av "Helt enig" på påstand i gruppe 1 vil indikere høyt nivå av innsats og motivasjon på testen, mens i gruppe 2 vil valg av "Helt enig" indikere det motsatte, nemlig lavt innsats- og motivasjonsnivå. En slik sammensetting av variabelverdiene i et konstrukt krever omkodning av svaralternativene i noen av påstandene (variablene) for å kunne teste deres indre konsistens og for å beregne selve konstruktverdien. Ved hjelp av transformasjonsfunksjonen "Recode" i SPSS ble dermed variabelverdiene i gruppe 1 omkodet slik at lave verdier på svaralternativene indikerer lavt nivå av motivasjon og brukt innsats på testen. (1 tilsvarer det laveste nivået og 4 - det høyeste nivået).

Sammen utgjør de seks påstandene (P1 – P6) en samlevariabel, som jeg kaller videre for testmotivasjonskonstrukt. *"En samlevariabel er en slags sammenfatning av flere variabler. Etablering av en samlevariabel er en omfattende prosess som baserer seg både på faglig innsikt og på statistiske metoder. Med bakgrunn i erfaring og tidligere forskning vil man ofte anta at flere variabler måler aspekter av samme fenomen," og " ... de sammen danner et konstrukt."* (Grønmo m.fl., 2010, s.265). Det er flere fordeler ved å sette sammen flere enkelte variabler i ett konstrukt (Eklöf, 2006b). For å kunne ha nytte av disse fordelene er det viktig at variablene har høy indre konsistens og at det sammensatte konstruktet har validitet som betyr at det kan brukes som mål for fenomenet som skal undersøkes (har validitet).

Testmotivasjonskonstruktet i TIMSS Advanced 2008 er forankret i prestasjonsmotivasjonsteori om forventning og verdisetting, og bygger på forskningen og de empiriske dataene ved bruk av skalaen SOS¹⁵ utviklet av Donna L.Sundre (1999), og skalaene

¹⁵ The Student Opinion Scale er utviklet og brukt ved James Madison University, USA og består av 10 variabler likt fordelt på to faktorer: den ene måler testtakerens selvrapportert innsats og utholdenhet på testen og den andre måler testtakerens oppfatninger av hvor stor betydning testen har for han eller henne. Skalaen kan lastes ned fra <http://www.jmu.edu/assessment>. I sin helhet passer ikke SOS for tester hvor testresultatene ikke utleveres tilbake til testtakerne. (Eklöf, 2010a)

av Hanna Eklöf (2006, 2010a) utviklet i forbindelse med studier av testmotivasjon hos både svenske elever i 8.klasse som deltok i TIMSS 2003 og svenske gymnaselever i 12.klasse som deltok i pilotstudien TIMSS Advanced 2008. Gjennom empiri og statistiske analyser av disse skalaene er det fastslått at de har tilfredsstillende reliabilitet og validitet (Eklöf, 2010a).

4.3 Databehandling

Alle statistiske analyser av datamaterialet er gjennomført med SPSS, PASW Statistics 18.

4.3.1 Behandling av samlevariabler

Hvis det stilles flere spørsmål om det samme fenomenet, som måles ved Likert-skala, kan det konstrueres en samleindeks for dette fenomenet ved å beregne gjennomsnittet av verdiene som representerer de valgte svaralternativene¹⁶ (Johannessen m.fl., 2009). Denne metoden ble brukt av Hanna Eklöf ved beregning av testmotivasjon som en samlevariabel i undersøkelsen av testmotivasjon hos de svenske elevene i TIMSS Advanced 2008 (se Eklöf, 2010b). Jeg har valgt å bruke samme metode i denne oppgaven og beregner testmotivasjonskonstruktet som et gjennomsnitt av variabelverdiene til de seks påstandene om motivasjon og innsats, videre kalt for testmotivasjonsskår. En viktig forutsetning her er at alle de enkelte variabelverdiene i konstruktet tolkes på én og samme måte: like verdier indikerer samme nivå av motivasjon og innsats. Det er valgt å bruke verdien 1 som indikasjon på det laveste nivået av innsats og motivasjon, og 4 som indikasjon på det høyeste nivået. Derfor er det brukt de omkodete variabelverdiene i påstandsguppe 1 (P1, P4 og P5, se avsnitt 4.2), som sammen med påstandene i den andre gruppen (P2, P3 og P6) dannet bergningsgrunlaget for konstruktet ”Testmotivasjon”. Som resultatet av påstandenes metriske skala rangeres testmotivasjonsskår fra 1 på det laveste til 4 på det høyeste, hvor de lave verdiene hører sammen med lavt nivå av motivasjon og innsats.

4.3.2 Prestasjonsskår i matematikktesten

¹⁶ Dette kalles for additiv skala, og indeksen regnes ut ved å summere skårene på alle svarene på de gitte spørsmålene, og deretter dividere dette på antall spørsmål og på antall respondenter (Johannessen m.fl., 2009)

Matematikkoppgavene i TIMSS Advanced utgjør en stor oppgavebank som er fordelt i 7 blokker, og i TIMSS Advanced 2008 ble et bestemt utvalg av oppgaver fra de forskjellige blokkene fordelt på fire forskjellige hefter slik at heftene overlapper hverandre med en blokk.

Tabell 4.1 Fordeling av blokker i hefter (Grønmo m.fl., 2010, s.257).

Hefter	Blokker		
Hefte 1	M1	M2	M3
Hefte 2	M4	M3	M5
Hefte 3	M6	M4	M1
Hefte 4	M2	M5	M7

Hver elev fikk dermed ett hefte med utvalgte oppgaver som utgjør bare en del av oppgavebanken. Det er umulig å bruke elevens resultatskår på oppgavene i heftet til å sammenlikne hans eller hennes prestasjoner med prestasjonene til elever som fikk et annet hefte. ”Avanserte statistiske metoder er brukt for å behandle dataene på en måte som muliggjør sammenlikninger” (Grønmo m.fl., 2010, s.262). Som resultatet av dette ble det regnet ut fem enkeltskår (kalt ”plausible values”) for hver enkelt elev. Disse enkeltskårene er statistiske verdier som viser hvordan eleven sannsynligvis ville ha prestert på de andre heftene enn det eleven fikk på testen, og beregningene av disse verdiene er basert på kunnskapen om hvordan den konkrete eleven presterte på oppgavene fra en bestemt blokk i sitt hefte og hvordan elevene som fikk et annet hefte med den ”overlappende” blokken presterte på denne blokken. Videre ble disse enkeltskårene skalert slik at det internasjonale gjennomsnittet ble 500 poeng med standard avvik på 100 poeng. I denne masteroppgaven ble det regnet ut prestasjonsskår for hver enkelt elev som gjennomsnittet av de fem enkeltskårene. Disse enkeltskårene er som nevnt tidligere i dette kapitlet hentet fra databasen oppbevart ved instituttet.

4.3.3 Bortfalte verdier

Det er lavt antall bortfalte verdier i svarene på motivasjonsspørreskjema:

Tabell 4.2 Prosentvis bortfall av variabelverdiene i de seks påstandene ($n = 1912$)

P1	P2	P3	P4	P5	P6
1,7%	1,4%	1,6%	1,8%	1,9%	1,4%

Bortfalte verdier fordeler seg på forskjellige variabler og forskjellige elever. Ved å filtrere bort elever med minst én bortfalt verdi ble antall caser redusert fra 1912 til 1853, som tilsvarer 3,1% bortfall, hvor 3,6% utgjør bortfall blant jenter og 2,7% er bortfall blant gutter. (Tatt utgangspunkt i det norske bruttoutvalget på 1932 elever er bortfallet på 4,1 %). Bortfall på 5-10 prosent av bruttoutvalget er regnet som lite og vi kan ha ganske stor tillitt til generalisering av resultatene (Johannessen m.fl., 2009).

Det finnes flere måter å estimere bortfalte verdier på. Jeg har valgt å estimere de bortfalte verdiene ved å erstatte dem med variabelens gjennomsnittsverdi fordi denne metoden ble brukt i analyser av elevenes respons på motivasjonsspørreskjema i Sverige (Eklöf, 2010a) og disse analyseresultatene er brukt i mine analyser som referanse.

Dessuten er det tre elever som ikke svarte på spørsmål om kjønn. Blant disse er det én elev som ikke svarte på noen spørsmål i spørreskjemaet i det hele tatt, mens de to andre har svart på alle de seks spørsmålene om motivasjon på testen. Disse tre casene er filtrert bort ved analyser av samvariasjon mellom motivasjon og prestasjon fordi disse analysene undersøker bl.a. kjønnsforskjeller.

4.3.4 Statistiske analyser

Reliabilitet av den konstruerte skalaen for testmotivasjon

Reliabiliteten til enkeltvariablene i et konstrukt kalles for variabelenes indre konsistens og omhandler i hvilken grad de ulike variablene måler det samme fenomenet. I denne oppgaven representerer de seks påstandene om motivasjon og innsats en samlevariabel for testmotivasjon. Påstandenes reliabilitet er målt ved Cronbach's koeffisient alfa.

Tilfredsstillende verdier av alfa har 0,7 som nedre grense, men det er ønskelig med alfaverdier over 0,8. Høy alfakoeffisient viser at svarene på påstandene i spørreskjemaet ikke er et resultat av tilfeldigheter i deres oppsett og at de danner et konstrukt.

Deskriptiv statistikk

Testing av normalitet

Det er gjort analyser av skjevhet og kurtose (univariat skewness & kurtosis) av de enkelte variablene i motivasjonskonstruktet for å undersøke deres normalitet, fordi mange statistiske tester (som for eksempel Pearsons produktmoment-korrelasjon) forutsetter at variablene som testes er normalfordelt.

Skjevhet viser hvor symmetrisk en klokkeformet fordelingskurve er. Når skjevheten er lik null er fordelingsformen symmetrisk klokkeformet. Positiv skjevhet tilsvarer høyre skjevhet med hale mot høye verdier. Medianen i en slik fordeling er lavere enn gjennomsnittet. Negativ skjevhet tilsvarer venstre skjevhet med hale mot lave verdier. Medianen i en slik fordeling er høyere enn gjennomsnittet. Kurtose viser hvor spiss klokkeformet fordelingskurve er. Positiv kurtose indikerer om spiss kurve (leptokurtisk kurve), negativ kurtose – flat kurve (platokurtisk kurve).

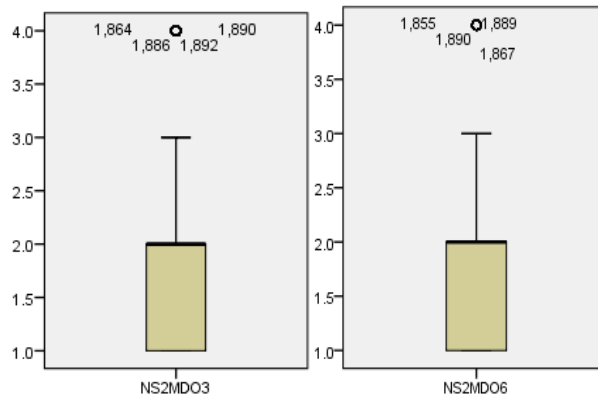
Tommelfingerregelen sier at når beregnet skjevhet og kurtose ligger mellom ± 1 , og forholdet mellom deres verdier og deres standardfeil er mindre enn 3.3 (såkalt z-verdi), er variabelen normalfordelt. For høy z-verdi kan indikere tilstedeværelse av ekstreme verdier i variabelmålinger (såkalt ”outliers”). Men variabelfordelinger i store utvalg er ikke ”følsomme” for z-verdier og vi kan betrakte variablene som normalfordelt bare ut ifra selve skjevhets- og kurtoseverdiene. La oss se på hvordan dette er relatert til de seks variablene i spørsmål om motivasjon og innsats.

Deskriptiv analyse av de seks variablene i motivasjonskonstruktet viser at de kan anses som normalfordelt, men påstand 3 og 6 har skjevhetsverdi på grensen av det akseptable nivået. Resultatet av skjevhets- og kurtoseanalyse er vist i tabell 4.2.

Tabell 4.2 SPSS-analyse av skjevhet og kurtose i fordelingene av variabelverdiene i de seks påstandene i spørreskjemaet.

	Skewness		Kurtosis	
	Statistic	Std. Error	Statistic	Std. Error
P 1	.396	.056	-.551	.113
P 2	.483	.056	-.553	.113
P 3	.987	.056	.170	.113
P 4	-.144	.056	-.865	.113
P 5	-.057	.057	-1.011	.113
P 6	1.033	.056	.408	.113

Vi sjekker de seks påstandene for ekstreme variabelverdier, og analysen (Explore "outliers" i SPSS) viser at det er noen få av dem i P3 og P6, med fire caser i hver. Resultatet er vist i figur 4.1.



Figur 4.1 Caser med ekstreme verdier i påstand 3 og 6.

De fire ekstreme variabelverdiene påvirker ikke den deskriptive statistikken for disse variablene, fordi det er ganske få av dem tatt i betraktning størrelsen på utvalget ($n = 1912$), og vi kan dermed akseptere at variablene i motivasjonskonstruktet er tilnærmet normalfordelt.

Sentraltendens og spredningsmål i variablenes fordelinger i analysene er vist ved gjennomsnittet og standardavviket for variabelen.

Frekvenstabeller

Det er undersøkt prosentvise andeler av elever som sier seg enige eller uenige i påstandene i spørreskjemaet. Resultatene er brukt både til å vise hvordan elevsvarene fordeler seg på de ulike påstandene og til å sammenlikne fordelingene innen de ulike elevgruppene.

Sammenlikningene er gjort både for å undersøke kjønnsforskjeller blant norske elever og for å sammenlikne resultatene mellom Norge og Sverige.

Samvariasjon

Samvariasjon mellom variabler ble målt ved bivariat korrelasjonsanalyse med Pearsons produktmoment-korrelasjon som mål, også kalt Pearsons r . En positiv korrelasjon mellom to variabler viser at høye verdier på den ene variabelen hører sammen med høye verdier på den andre, og omvendt. Negativ korrelasjon viser at høye verdier på den ene variabelen går sammen med lave verdier på den andre, og omvendt. Pearsons r angir både type samvariasjon (positiv, negativ eller fraværende), og hvor sterk den er. Pearsons r er en standardisert

koeffisient som varierer mellom -1 og $+1$, hvor 0 er et uttrykk for at det ikke eksisterer korrelasjon mellom variablene, mens 1 viser er det fullstendig positiv sammenfall mellom verdiene på variablene, og tilsvarende med -1 : fullstendig negativt sammenfall. ”Det finnes ikke noe fasitsvar på hva som er en høy korrelasjon. Det er blant annet avhengig av hva som undersøkes, og hvor sterk korrelasjon man forventer. I samfunnsvitenskapelig forskning regnes Pearsons r opp til $0,20$ som en svak samvariasjon, $0,30 - 0,40$ som relativt sterk og over $0,50$ som meget sterk.” (Johannessen m.fl., 2009, s. 259).

Korrelasjonens signifikans viser hvor sannsynlig det er at samvariasjon ikke er bare resultat av tilfeldighet. Den måles enten på $0,05$ -nivå eller på $0,01$ -nivå. Når korrelasjon er signifikant på $0,05$ -nivå, betyr det at det er mindre enn 5% sannsynlighet for at det egentlig ikke er noen korrelasjon mellom variablene i populasjonen, og at den beregnede r -verdien bare er resultat av tilfeldighet. Ved $0,01$ -nivå er denne sannsynligheten redusert til 1% . Verdier over $0,05$ tolkes som at korrelasjon ikke er signifikant.

Sammenlikning av statistiske mål

I samfunnsforskning er vi ofte interessert i å sammenlikne grupper for å finne ut om det er forskjeller mellom dem (Johannessen m.fl., 2009). I analysene i denne oppgaven er det undersøkt kjønnsforskjeller i testmotivasjon ved t -test, testing av differanser mellom gjennomsnitt kalt for hypotesetesting. Det sentrale spørsmålet ved hypotesetesting er om resultatet er statistisk signifikant eller ikke (Johannessen m.fl., 2009). Signifikansverdi viser beregnet usikkerhet for hva som kan være en faktisk forskjell. Det tolkes på samme måte som det er beskrevet ovenfor med den forskjellen at det gjelder forskjeller mellom grupper, og ikke samvariasjon mellom variabler.

Men de målte forskjellene kan være signifikante fordi de er store og/eller fordi det testede utvalget er stort. Derfor er det vanlig i tillegg til t -test å bruke en såkalt effektstørrelse for å kunne tolke de eventuelle forskjellene mellom grupper. Effektstørrelsen bruker standardavviket for hele populasjonen som mål og beregnes ved å dividere forskjellene i gruppene gjennomsnittsverdier med populasjonens standardavvik. På denne måten får man inntrykk av hvor stor forskjellen mellom gruppene er i forhold til spredningen i utvalget i sin helhet.

5 Analyser og resultater

5.1 Testmotivasjon hos de norske 3MX-elevene i TIMSS Advanced 2008.

5.1.1 Elevenes svar på påstandene i elevspørreskjema

Reliabilitetsanalyse av de seks variablene i motivasjonsspørsmålet 22 i spørreskjema viser høy indre konsistens med $\alpha = 0,82$ (0,7 regnes i mange sammenheng som nedre grense for en samlev variabel, som skal brukes til å sammenlikne grupper innen testpopulasjon). Dette betyr at de kan brukes som mål for ett og samme konstrukt, i kapittel 4 kalt for testmotivasjonskonstrukt. Før vi tar for oss analyser som omhandler dette konstruktet, skal vi se nærmere på de seks enkeltvariablene i konstruktet.

Svaralternativene for de seks utsagnene i spørreskjema var gitt på en firepunktsskala, og for hvert utsagn skulle elevene angi om de var ”Helt enig”, ”Litt enig”, ”Litt uenig” eller ”Helt uenig” med påstanden. Tabell 5.1 viser prosentvis andel av elevene som sier seg ”helt enig” eller ”litt enig” i hver påstand.

Tabell 5.1 Prosentandel av elevene som sier seg ”helt enig” eller ”litt enig” i hver av de seks påstandene i konstruktet ”Testmotivasjon, sortert etter de utsagnene elevene er mest enige i. (Hver påstand er merket med nummer i parentesene etter påstandenes rekkefølge i spørreskjema.)

	Prosentandel (av totalt 1912 elever)
Jeg kunne ha anstrengt meg mer da jeg jobbet med disse oppgavene (P6)	84
Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året (P3)	82
Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene (P2)	68
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen (P1)	68
Jeg var motivert for å gjøre mitt beste på disse oppgavene (P5)	48
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige (P4)	44

Utsagnene i de tre øverste radene i tabellen har et negativt betonet ordlyd i motsetning til utsagnene i de tre nederste radene, og valg av enig eller litt enig i de tre første påstandene vil indikere negativ vurdering av egen motivasjon og innsats på testen. Det er en generelt større andel elever som var enig eller litt enig i de tre øverste påstandene enn i de tre andre, og det første inntrykket av tallene i tabellen synes å være at elevene i stor grad hadde en negativ oppfatning av egen motivasjon og innsats på testen.

Som det er nevnt tidligere er motivasjon en abstrakt variabel som måles gjennom behavioristiske indikatorer. Innsats og utholdenhet er slike indikatorer. Høyt nivå av innsats, spesielt når oppgavene er vanskelig og byr på utfordringer, indikerer høyere motivasjon hos elevene (Schunk m.fl., 2010).

Når vi ser på påstandene i spørreskjema i perspektivet av denne teorien, kan det tenkes at påstandene P1, P2, P3 og P6 representerer innsatsindikatorer og at P4 representerer utholdenhetsindikator. I tillegg kan påstand P3 gi en indikasjon på hvordan elevene oppfattet viktigheten av denne prøven for dem. Det kan tenkes at elever som anstrengte seg mindre på TIMSS-testen enn det de gjør ellers på de ordinære skoleprøvene, ser på TIMSS-testen som mindre viktig for dem. Påstanden P5 gir en verbalisering av elevenes motivasjon for å gjøre sitt beste på testen og elevenes respons på denne påstanden forteller om hvordan elevene selv oppfattet sitt motivasjonsnivå.

En relativt stor elevandel (68 %) (tatt i betraktning elevenes respons på andre påstander) sier seg helt enig eller litt enig i å gjøre sitt beste på testen samtidig som omtrent 4 av 5 sier at de kunne ha anstrengt seg mer og at de anstrengte seg mindre på denne testen enn på prøver ellers i året. Dessuten sier like mange elever (68 %) at de ikke var helt konsentrert under testen. Dette kan tyde på at mange elever hadde et lavere innsatsnivå på denne testen.

Vi kan også bemerke at bare 44 prosent av elevene som sier at de arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige. I følge motivasjonsteori, vil bevaring av arbeidsinnsats over lengre tid, spesielt når oppgaven byr på utfordringer, bety god utholdenhet i arbeidet man gjør. God utholdenhet forbindes med høyere nivå av motivasjon. Når det er under halvparten av elevene som prøvde så godt de kunne selv om oppgavene var vanskelig, kan det tyde på at mange elever ikke var nok motivert for å yte sitt beste. Det kan i så fall støttes av elevenes rapportering om sin motivasjon på testen i påstand 5. Det er mindre enn halvparten av elevene som sier seg enig eller delvis enig i å være motivert for å gjøre sitt beste på oppgavene i testen.

Sammenheng mellom elevenes respons på påstander om innsats og utholdenhet på testen og deres rapportering om sin motivasjon for å gjøre sitt beste på testen er undersøkt ved korrelasjonsanalyse og resultatet er vist i tabell 5.2.

Tabell 5.2 Korrelasjon mellom elevenes svar på påstanden om god motivasjon på testen og påstandene om å gjøre sitt aller beste og å vise utholdenhet. To stjerner indikerer at korrelasjon er signifikant på 0,01-nivå.¹⁷

	Jeg gjorde mitt aller beste på oppgavene i denne testen	Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige
Jeg var motivert for å gjøre mitt beste på oppgavene i testen	$r = 0,57^{**}$	$r = 0,51^{**}$

De positive verdiene av korrelasjonskoeffisienten r viser at å velge "helt enig" på påstandene om innsats og utholdenhet på TIMSS-testen hører sammen med å velge "helt enig" på påstanden om å være motivert for å gjøre sitt beste på testen. Denne sammenhengen kan sies å være sterk men tolkningen av dens betydning er avhengig av hvor sterk sammenheng mellom variablene man *forventer* å se (Johannesen m.fl., 2009).

Testmotivasjon er definert som viljen til å engasjere seg i arbeid med testoppgavene for å gjøre sitt beste med siktet på å vise sine kunnskaper og ferdigheter innen det som testes. Testmotivasjon skal igangsette og opprettholde dette arbeidet, med andre ord skal føre til innsats og utholdenhet i dette arbeidet (Baumert og Demmrich, 2001). Når eleven påstår å være motivert for å gjøre sitt beste på testen synes det å være rimelig å forvente at han eller hun *gjør* sitt beste og *viser* utholdenhet i sitt arbeid. Korrelasjonskoeffisienten $r = 0,57$ i tabell 5.2 viser at det er ca.33 prosent av variansen i den rapporterte innsatsen som kan knyttes til elevenes rapporterte motivasjon. Tilsvarende kan ca. 26 prosent av variansen i den rapporterte utholdenheten "forklares" med elevenes motivasjon. Hva kunne forklare at samvariasjonen mellom elevenes svar på disse påstandene ikke er større enn det vi ser her? Pintrich og Schunk har antydnet at tolkningen av innsats som en motivasjonsindeks kan "forstyrres" av blant annet dyktighetsnivå hos eleven (Schunk m.fl., 2010). Barry og Finnley (2009) har undersøkt testmotivasjon hos collestudenter gjennom studie med fem tester satt sammen i en serie, og har vist at innsatsen blant studentene var lavest ved kognitivt vanskelige tester som krevde evner til å resonnerer (noe vi ofte forbinder med ikke-rutinepregete oppgaver). De

¹⁷ Positiv korrelasjonskoeffisienter viser at valg av lave verdier ("Helt enig" og "Litt enig") på den ene variabelen hører sammen med valg av lave verdier på den andre variabelen.

konkluderte blant annet med at studenter er mindre villige til å yte innsats på vanskeligere tester, og at *”examinees appear to be willing to put forth effort in low-stakes contexts, but are more willing on tests that are not exceedingly difficult; at some point, if the test is too difficult, effort decreases.”* (Barry & Finnley, 2009, s.33). Med støtte i deres funn kan det tenkes at noe av variasjonen i norske elevers innsats og utholdenhet på TIMSS-testen kan, i tillegg til motivasjon, også knyttes til de kognitive faktorene. Det kan i så fall bety at noen elever rapporterte om dårlig innsats på testen (selv om de oppga å være motivert for å gjøre sitt beste) fordi de oppfattet testen som krevende og vanskelig for dem. I så fall kan lavere innsats og utholdenhet på testen hos disse elevene ikke forklares bare med manglende vilje til å engasjere seg i arbeid med oppgavene, men også forholdet mellom deres forutsetninger (som kunnskaper og kognitive evner) og testens karakteristikker (innholds- og/eller kognitivmessig). Som det er nevnt i kapittel 2 har analysene av prestasjonsresultater i TIMSS Advanced 2008 vist at Norge har svært få elever som når avansert nivå på matematikktesten, og hele 65 prosent av norske elever ligger på lavt nivå.

Som det er vist i kapittel 3 gjenspeiler testmotivasjon et samspill mellom mange forskjellige faktorer. Fra hypotesene i prestasjonsmotivasjonsteori vet vi at vurdering av testens betydning er en viktig komponent i verdifaktoren som påvirker testmotivasjon, og angivelig finnes det en sammenheng mellom testens verdivurdering hos elevene og deres motivasjon for å yte sitt beste på testen (Wise & DeMars, 2005; Barry & Finney, 2009; Eklöf, 2010b). I TIMSS-undersøkelsen ble elevene ikke spurt direkte om deres oppfatninger av hvor viktig testen var for dem, men påstanden *”Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året”* kan gi oss en indirekte indikasjon på elevenes holdninger til testen. 82 prosent av elevene (over 4 av 5) sier seg helt enig eller delvis enig i denne påstanden. Dette kan gi inntrykk av at mange elever vurderte TIMSS-testen som mindre viktig for dem, og dermed anstrengte de seg mindre enn det de gjør på andre prøver ellers. TIMSS-test er en så kalt *”low-stakes”* test for elever fordi den ikke har direkte konsekvenser for hver enkelt av dem, og dette kunne tenkes som en mulig forklaring på at det var mange elever som ikke engasjerte seg i arbeid med testoppgavene på samme måte som de gjør ellers på skoleprøver. Jeg har undersøkt sammenheng mellom elevenes rapportering om egen innsats på TIMSS-testen sammenliknet med andre skoleprøver, og deres respons på noen andre påstander i spørreskjemaet. Resultater av korrelasjonsanalyse er vist i tabell 5.3.

Tabell 5.3 Korrelasjon mellom elevenes vurdering av egen innsats på TIMSS-testen sammenliknet med prøver ellers, og deres respons på påstandene om motivasjon og innsatsnivå på testen. To stjerner indikerer at korrelasjon er signifikant på 0,01-nivå.

	Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året
Jeg var motivert for å gjøre mitt beste på disse oppgavene	$r = -0,44^{**}$
Jeg kunne ha anstrengt meg mer da jeg jobbet med disse oppgavene	$r = 0,52^{**}$
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen	$r = -0,50^{**}$

Den negative verdien av korrelasjonskoeffisienten r viser at å velge "helt enig" på påstanden om dårligere innsats på TIMSS-testen, sammenliknet med andre prøver, hører sammen med å velge "helt uenig" på påstanden om å være motivert og å ha gjort sitt beste på denne testen. Tallene i tabellen viser at elevenes eventuelt dårligere innsats på TIMSS-testen sammenliknet med andre prøver har en negativ sammenheng med deres motivasjon for å gjøre sitt beste og investerte innsats på testen, men en positiv sammenheng med deres respons på påstanden om å kunne ha anstrengt seg mer. Korrelasjonskoeffisientene viser at 19% av variasjon i angivelse av motivasjon for å gjøre sitt beste og 27% av variasjon i responsene på "kunne ha anstrengt meg mer" kan forklares med hva elevene angir om sin innsats på denne testen sammenliknet med andre tester. Med forbehold om gyldigheten til antakelsen om at den sistnevnte påstanden kan indikere hvor viktig TIMSS-testen anses av elevene å være for dem, kan resultatene i denne analysen tyde på at motivasjon for å gjøre sitt beste og innsats på testen hos norske 3-MX elever i TIMSS Advanced har en sammenheng med deres vurdering av testens status. For å kunne gjøre noen slutninger om en slik sammenheng trenges det videre undersøkelser. For eksempel påstander som "Det var viktig for meg å gjøre det bra på denne testen", "Denne testen var viktig for meg" og "Jeg bryr meg ikke om hvordan jeg skårer på denne testen", brukt i Student Opinion Survey (SOS) for å måle testtakerens oppfatninger av hvor stor betydning testen har for han eller henne (se kapittel 4), kan da eventuelt inkluderes i elevspørreskjema ved en ny TIMSS Advanced undersøkelse.

Til slutt vil jeg påpeke at selvrapporing ofte gjenspeiler subjektive oppfatninger som kan være vanskelig å sette opp mot objektive kriterier (Eklöf, 2010b), og at validiteten av resultatene ved selvrapporing er avhengig av blant annet at elevene svarer ærlig. Hvordan

tolker forskjellige elever påstand om å gjøre sitt aller beste? Det kan for eksempel tenkes at noen elever syntes at oppgavetyperne og vanskelighetsgraden på oppgavene var slik at det ikke ga mulighet til å *vise* sitt aller beste, eller at noen andre oppfattet testen som ikke krevende i det hele tatt og syntes at man ikke trengte å ”*ta fram*” sitt beste. For flinke elever er det enklere å prestere bedre med mindre innsats og de er ofte mer motivert for å avlegge en prøve enn svake elever (Kjærnsli & Roe, 2010 ref.: Weiner, 2007 og Boekaert, 2010). Jeg vil også understreke at elevspørreskjemaet ble besvart av elevene i etterkant av testen, og det kan hende at elever som følte at de ikke fikk det til på testen, kanskje vil i etterkant av testen svare at de ikke har gjort sitt beste. For bedre å kunne forstå testmotivasjon hos norske elever i kontekst av TIMSS-tester kan det være viktig i de kommende undersøkelsene å kombinere elevenes selvrapporing i etterkant av testen med andre målemetoder som for eksempel intervjuer med utvalgte elever og/eller observasjoner av elevenes behavioristiske faktorer under gjennomføring av testen (de forskjellige metodene for testmotivasjonsmålinger er omtalt i kapittel 3).

For å oppsummere resultater av matematikkelevenenes selvrapporing om motivasjon og innsats på TIMSS-testen 2008 skal vi nå se på konstruktet testmotivasjon.

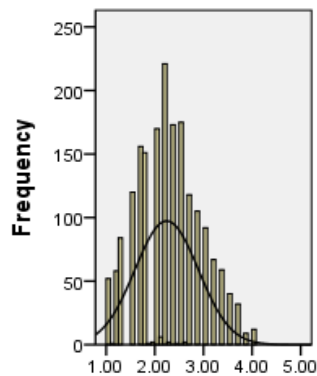
5.1.2 Testmotivasjon som samlev variabel

Som det er forklart i kapittel 4 er det regnet ut testmotivasjonsskår basert på de registrerte variabelverdiene av de seks påstandene i elevspørreskjema. Testmotivasjonsskår rangeres fra 1 til 4, hvor lavere verdier indikerer lavere motivasjonsnivå. Det synes å være passende å tolke verdiene i skalaen på følgende måte: gjennomsnittet på 3 eller høyere tilsvarer høyt nivå, 2 og 3 avgrenser middelsnivå, mens 2 eller lavere betyr lavt nivå.

Deskriptiv analyse av testmotivasjonsskår viser akseptable verdier av skjevhet og kurtose (mellom -1 og 1), som indikerer at variabelen er tilnærmet normalfordelt, og at motivasjonsskårens gjennomsnittsverdi ligger på 2,24 med standardavvik på 0,65.

Tabell 5.3 Deskriptiv statistikk for konstruktet Testmotivasjon.

Min.verdi	Maks.verdi	Gjennomsnittet	St.avvik	Skjevhet	Kurtose
1.00	4.00	2.24	0.65	0.28	-0.38



Figur 5.1 Histogram for testmotivasjonskonstruktet

Ut i fra høyreskjevhet¹⁸ i fordelingen og gjennomsnittsverdien på 2,24, som kan assosieres med i underkant av middelnivå, synes det riktig å konkludere med at norske elever som deltok i matematikktesten i TIMSS Advanced 2008 hadde i gjennomsnitt noe lavere testmotivasjonsnivå.

5.2 Kjønnsforskjeller i testmotivasjon blant norske elever

5.2.1 Kjønnsforskjeller i elevenes svar på påstandene i elevspørreskjema

Det var tre elever blant alle som har fylt ut elevspørreskjema som hadde blankt svar på spørsmål om kjønn, og deres svar på motivasjonsspørsmål har jeg derfor tatt bort. Jeg har brukt frekvenstabeller, og resultatene for hvert enkelt utsagn blant jenter og gutter er vist i tabell 5.4.

Tallene i tabellen synes å gi et inntrykk av at jentene er generelt noe mindre positive i deres selvrapping enn guttene.

¹⁸ Tyngden i tallmaterialet tenderer mot lavere verdier. Medianen i tallmaterialet er lavere enn gjennomsnittsverdien.

Tabell 5.4 Prosentandel av elever som sier seg ” helt enig” eller ”litt enig” i enkeltutsagnene i konstruktet ”Testmotivasjon, fordelt på jenter og gutter. (sorteringen i tabellen tilsvarer tabelloppsettet i 5.1.)

	Jenter (n = 729)	Gutter (n = 1180)
Jeg kunne ha anstrengt meg mer da jeg jobbet med disse oppgavene (P6)	85,7	83,6
Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året (P3)	83,7	81
Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene(P2)	72,2	64,2
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen (P1)	67,1	66,2
Jeg var motivert for å gjøre mitt beste på disse oppgavene (P5)	48,7	46
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige (P4)	41,2	44,9

For å se om det kan finnes noe mønster i elevenes respons på påstandene skal vi se på gjennomsnittsverdiene og forskjellens effektstørrelse for hver påstand. Effektstørrelsen er regnet ut som $\frac{m_g - m_j}{SD}$, hvor m_g og m_j er gjennomsnittsverdiene for henholdsvis gutter og jenter, og SD er standardavviket for alle elever. Positiv verdi av effektstørrelsen indikerer at forskjellen er i guttenes favør, dvs at guttene er generelt mere positive enn jentene, mens negativ verdi indikerer det motsatte, nemlig at jentene er mere positive enn guttene. For påstandene P1, P4 og P5 er det brukt en ”snudd” skala sånn som det er beskrevet i kapittel 4¹⁹. Resultatet er vist i tabell 5.5.

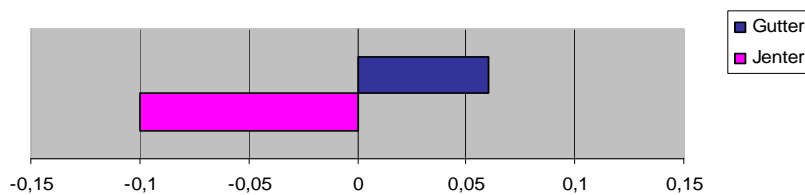
Tallene i tabellen viser at jentene er noen grad mindre positive i rapportering av sin innsats og utholdenhet på testen, men mer positive i rapportering av motivasjon for å gjøre sitt beste på testen (P5). I påstand 2 (”Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene”) kan vi registrere en merkbar høyere effektstørrelse i guttenes favør enn i de andre påstandene.

Tabell 5.5 Kjønnforskjeller for enkeltpåstandene i konstruktet ”Testmotivasjon”.

¹⁹ Lavere m-verdier i tabellen indikerer mer negativ respons i forhold til motivasjon og brukt innsats på testen.

	Gjennomsnittet		St.avvik	Effektstørrelse
	Jenter	Gutter		
P1	2,82	2,83	0,89	0,01
P2	2,14	2,29	0,93	0,16
P3	1,77	1,79	0,89	0,03
P4	2,37	2,44	0,94	0,07
P5	2,52	2,46	0,98	-0,06
P6	1,71	1,73	0,84	0,03

Vi kan utdype tolkningen av effektstørrelse ved å ta utgangspunkt i P2 og se på de standardiserte verdiene av svarene på denne påstanden²⁰. Gjennomsnittet av de standardiserte verdiene er på $-0,1$ for jenter og $0,06$ for gutter. Negativ verdi viser ikke at man er negativ men at man er *mindre positiv* enn gjennomsnittet, som er illustrert på figur 5.2.



Figur 5.2 z-skår for gjennomsnittet av variabelverdiene for påstand 2 fordelt på kjønn.

Altså kan vi se at jentene er i gjennomsnittet mindre positive enn guttene i deres respons på påstand 2, og forskjellen utgjør $0,16$ på den standardiserte skalaen, med andre ord noe mindre enn en femte del av standardavviket. Ser vi på effektstørrelsene i de andre påstandene ser vi at det er ganske lite som skiller mellom jenter og gutter i disse påstandene.

T-test bekrefter at jenter og gutter skiller seg fra hverandre statistisk signifikant i svarene på påstand 2. Den viser at det er mindre enn 1 % sannsynlighet for at forskjellen i respons på påstand 2 blant jenter og gutter skyldes en ren tilfeldighet ($p = 0,001$). Når det gjelder de andre påstandene viser de beregnede p-verdier at det er veldig høy sannsynlighet for at evt. kjønnsforskjeller skyldes en ren tilfeldighet, og dermed kan vi konkludere med at det finnes ingen signifikante kjønnsforskjeller i elevenes respons på verken påstand 1, 3, 4, 5 eller 6. Resultatet av t-test er vist i tabell 5.6.

²⁰ På en standardisert skala ligger gjennomsnittet i populasjonen på 0, og standardavvik er lik 1.

Tabell 5.6 Resultatet av en t-test av kjønnsfordelingen i responsene på de seks påstandene om motivasjon og innsats.

		Levene's Test for Equality of variance		t-test for Equality of Means				
		F	Sig.	t	df	Sig. (2-tailed)	Mean Differ.	Std. Error Diff
P1	Equal variances assumed	4.277	.039	.29	1876	.775	.012	.042
	Equal variances not assumed			.29	1577.71	.772	.012	.042
P2	Equal variances assumed	20.619	.000	-3.41	1881	.001	-.149	.044
	Equal variances not assumed			-3.47	1597.15	.001	-.149	.043
P3	Equal variances assumed	3.848	.050	-.65	1878	.517	-.027	.042
	Equal variances not assumed			-.66	1590.91	.511	-.027	.042
P4	Equal variances assumed	4.507	.034	1.50	1874	.134	.067	.045
	Equal variances not assumed			1.52	1560.52	.130	.067	.044
P5	Equal variances assumed	.241	.623	-1.31	1871	.192	-.061	.047
	Equal variances not assumed			-1.31	1533.25	.191	-.061	.047
P6	Equal variances assumed	1.731	.188	-.61	1881	.545	-.024	.040
	Equal variances not assumed			-.61	1566.87	.541	-.024	.039

5.2.2 Kjønnsforskjeller i testmotivasjon som samlevariabel

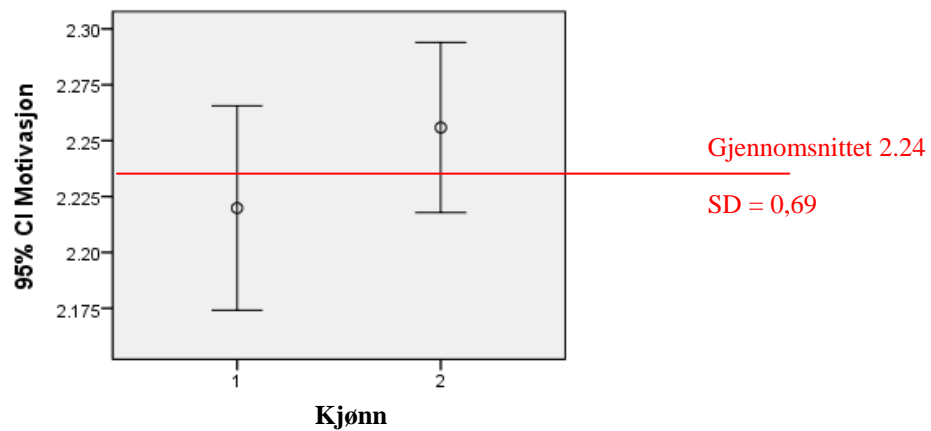
Når vi ser på konstruktet "Testmotivasjon" viser t-test at det er ingen signifikant forskjell i testmotivasjonsskår mellom jenter og gutter: t-testen viser at det er over 24 % sannsynlighet for at forskjellen i testmotivasjon blant jenter og gutter skyldes en ren tilfeldighet ($p = 0,24$). For å utdype dette resultatet kan vi se på effektstørrelsen (se tabell 5.7).

Tabell 5.7 Kjønnsforskjeller for konstruktet "Testmotivasjon".

	Gjennomsnittet		St.avvik	Effektstørrelse	z-skår	
	Jenter	Gutter			Jenter	Gutter
Testmotivasjonsskår	2,82	2,83	0,89	0,06	-0,03	0,02

Vi kan se at jentene er noe mindre positive på konstruktet testmotivasjon enn guttene, men "avstanden" mellom jentene og guttene på den standardiserte skala utgjør bare en 20-del av

standardavviket. Grafisk framstilling av 95% konfidensintervaller i figur 5.3 kan illustrere denne situasjonen:



Figur 5.3 Error bar for motivasjonsskår med 95 % konfidensintervall, splittet på kjønn (jenter - 1, gutter - 2). Det markerte gjennomsnittet og standardavviket gjelder hele utvalget.

Analysene presentert i dette avsnittet viser at det finnes ingen statistisk signifikante kjønnsforskjeller i testmotivasjonskonstruktet blant elevene i denne undersøkelsen, men jenter er i gjennomsnittet noe mindre positive enn guttene i deres respons på enkelte påstander i spørreskjemaet. Mens jentene er mer positive i deres respons på påstand om å være motivert for å gjøre sitt beste på testen, er de mer negative i rapportering av sin innsats og utholdenhet på testen, og dessuten skiller jentene seg signifikant fra guttene i deres angivelse av hvor konsentrerte de var da de jobbet med oppgavene på testen. En mulig forklaring på dette kan være at i matematikk stiller jenter strengere krav til sin innsats enn guttene gjør. Det kan i så fall støttes av flere studier av kjønnsforskjeller i matematikk. For eksempel har en 4-årig studie av kjønnsforskjeller²¹ i holdninger i matematikk blant elever i 15- til 17-årsalderen i Sverige vist at det eksisterer en oppfatning blant elevene om at jenter bekymrer seg mer enn gutter, og at når man mislykkes i sine prestasjoner tror jenter i større grad enn gutter at det skyldes for lite arbeid. I USA viste en studie (2003) av kvinnelige bachelor studenter med høye prestasjonsresultater i matematikk at de forklarte sine gode resultater med bare hardt arbeid og de følte seg skuffet når deres innsats ikke ble verdsatt i utdanningskontekster (Brandell & Staberg, 2008). Det kan også bemerkes at i motsetning til denne undersøkelsen har i PISA 2009 i Norge vist at jentene var generelt mer positive enn guttene i rapportering av testmotivasjon på testen, hvor kjønnsforskjellen i testmotivasjonsskår var på kvart standardavvik i jentenes favør. Den eventuelle tendensen til endring i jentenes vurdering av

²¹ GeMa (Gender and Mathematics) - studie av svenske ungdomselever i perioden 2001 – 2004.

sin motivasjons og innsats på testene i forhold til guttene i TIMSS Advanced sammenliknet med undersøkelser i ungdomsskole, kan mulig ikke ensforklares med alderen men hvilket fag som elevene testes i, og at elevene som tar skolens mest avanserte matematikkurs er en utvalgt gruppe. Mens TIMSS og PISA gjelder "felles" fag som alle elever har på skolen, gjelder TIMSS Advanced avansert matematikk på videregående skole nivå, et fag som blant andre realfagene betraktes ofte som guttefag. Det kan tenkes at mens det å være flittig i sitt arbeid hører naturlig til jentenes selvbilde i grunnskole, betraktes god innsats av jentene som en nødvendig forutsetning for å lykkes i fordypning i matematikk. Ved en nylig gjennomført statistisk undersøkelse av kjønnsforskjeller i realfagene i norsk videregående skole ble det blant annet analysert valg av matematikkfagene i det første året, og resultatene viser at jenter må ha langt sterkere grunnskolerresultater enn gutter før de velger den teoretiske varianten²² (Bjørkeng, 2011). Også analysene i TIMSS Advanced 2008 viser at ved valg av fordypning i matematikk (3MX) er gode prestasjoner ("Jeg gjør det vanligvis bra matematikk") noe viktigere for jentene enn for guttene (se kapittel 2). Vi skal komme tilbake til kjønnsforskjeller senere i dette kapitlet, der hvor vi skal drøfte sammenheng mellom testmotivasjon og prestasjonsresultater på testen.

5.3 Norske elevers angivelse av motivasjon og innsats på testen sett i sammenheng med deres begrunnelser for valget av fordypning i matematikk

Motivasjon betraktes som et sammensatt fenomen med mange dimensjoner, og naturen av dette fenomenet er ennå ikke fullt forstått (Giota, 2006). På grunn av den teoretiske og latente karakteren av testmotivasjonskonstruktet står utvikling av konstruktets målemetoder sentralt i forskningen av testmotivasjon. Erfaringer fra empirisk forskning og tolkningen av de empiriske data anses som viktige kilder for å drive denne prosessen videre (Eklöf, 2010b). Den omfattende datamengden som er samlet inn i TIMSS Advanced gir muligheten til å undersøke forskjellige sammenhenger, og det synes å være interessant å trekke inn noen bakgrunnsvariabler som er påvist til å ha en relativt sterk sammenheng med testprestasjoner inn i analyser av elevenes testmotivasjon.

²² Den teoretiske matematikkvarianten etter Reform 2006 gir grunnlaget til å ta avansert matematikk det siste året på videregående skole.

I ”Matematikk i motvind” er det rapportert om positiv signifikant korrelasjon mellom norske elevers matematikkprestasjoner i TIMSS Advanced og deres fagvalgbe grunnelser som ”Gjør det bra i matematikk” ($r = 0,37$), ”Liker å løse oppgaver” ($r = 0,33$), ”Arbeid i matematikk går lett” ($r = 0,25$) og ”Matematikktimene er interessante” ($r = 0,22$)²³. I rapporten konkluderes det med at norske elever som presterer bra på matematikktesten i TIMSS Advanced, kjennetegnes ved å ha en positiv holdning til faget og undervisningen og ved å ha en høy selvoppfatning i faget (Grønmo m.fl., 2010). Tilsvarende sammenheng er også funnet internasjonalt i TIMSS Advanced: *”Across the participating countries, students who considered a positive orientation to be important for choosing to study mathematics had higher average mathematics achievement than students who thought it less important.”* (Mullis m.fl., 2009, s.137).

Faglig mestring og trivsel er grunnleggende faktorer ikke bare for elevenes valg av faglig fordypning men også for elevenes faglige utvikling (Grønmo m.fl., 2010). Denne utviklingen skjer gjennom læringsprosess hvor det i følge motivasjonsteori om forventning og verdisetting, eksisterer et gjensidig avhengighetsforhold mellom motivasjon på den ene siden og faglig mestring på den andre siden. Dette avhengighetsforholdet forklares med at gjennom mestring i faget vil eleven utvikle faglig selv tillitt og tro på seg selv, og dette vil bidra til motivasjon både for å lære mer og for å yte sitt beste i læringsprosessen²⁴. Som det er antydnet i kapittel 3 er faglig selv tillitt og tro på seg selv viktige faktorer for prestasjonsmotivasjon ikke bare i (domenebestemte) læringskontekster men også i (situasjonsbestemte) testkontekster. I læringskontekster vil elever som interesserer seg for faget mest sannsynlig ha høyere motivasjon for å lære mer. I testsituasjoner kan faglig interesse bidra til å øke elevenes verdisetting av testen fordi det kan tenkes at oppgavene i seg selv kan virke interessante. Dette kan gjennom økt motivasjon føre til bedre innsats og dermed bedre prestasjon. Med utgangspunkt i denne argumentasjonen har jeg valgt å undersøke sammenheng mellom faglig interesse, faglig selv tillitt som fagvalgbe grunnelse og elevenes angivelser av sin motivasjon og innsats på TIMSS-testen. Jeg har også valgt å trekke inn ”Gode lærere” som fagvalgbe grunnelse i analysene i dette avsnittet, fordi det er en relativt stor andel av norske elever som oppga det som viktig for dem.

²³ Alle korrelasjonene er signifikante (Grønmo m.fl., 2010)

²⁴ Som det er nevnt i kapittel 3 er vurdering av sin egen kompetanse (faglig selv tillitt) og tro på seg selv i en faglig utfordring (mestringsforventning) de to komponentene som danner forventningsfaktoren i prestasjonsmotivasjon.

Faglig interesse

Sammenheng mellom elevenes faglig interesse som fagvalgbe grunnelse og elevenes svar på de seks påstandene i motivasjonskonstruktet er testet ved korrelasjonsanalyse og resultatene viser noe nevneverdig sammenheng (statistisk signifikant med r -størrelsesorden over 0,1) mellom faglig interesse og fire av de seks påstandene. Resultatene er presentert i tabell 5.8.

Tabell 5.8 Korrelasjon mellom faglig interesse som begrunnelse for å velge fordypning i matematikk og ulike påstander om motivasjon og innsats. Det er tatt med bare sammenhenger hvor korrelasjonskoeffisienter er større enn 0,1. To stjerner indikerer at korrelasjon er signifikant på 0,01- nivå.

	Liker å løse matematiske problemer	Matematikktimene er interessante
Jeg var motivert for å gjøre mitt beste på disse oppgavene	$r = 0,19^{**}$	$r = 0,19^{**}$
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige	$r = 0,18^{**}$	$r = 0,18^{**}$
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen	$r = 0,11^{**}$	$r = 0,13^{**}$
Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn på prøver ellers i året	$r = -0,12^{**}$	$r = -0,11^{**}$

Positive r -verdier i de tre første radene viser at å velge ” helt enig ” på påstanden om innsats og motivasjon hører sammen med å oppgi fagvalgbe grunnelsen som veldig viktig. Negative r -verdier i den siste raden i tabellen viser at å velge ” helt uenig ” på påstanden om innsats hører sammen med å oppgi fagvalgbe grunnelsen som veldig viktig. Det kan tenkes at faglig interesse kan påvirke vurdering av den såkalte indre verdien (den subjektive verdien) av testen, fordi å løse oppgaver kan oppfattes som interessant i seg selv. I følge FVT vil denne subjektive verdien av testen i samspill med andre komponenter danne grunnlaget for den totale verdivurderingen av testen og deretter påvirke testmotivasjon hos elevene. Som det ble antatt tidligere i dette kapitlet, kan elevenes respons på påstand om innsats på TIMSS-testen sammenliknet med andre tester indikere at mange elever anså testen som mindre viktig for dem sammenliknet med andre skoleprøver. Det kan i så fall bety at elevene vurderte den ytre verdien av testen som lav, og sammen med vurderingen av påkrevet innsats på testen kan dette ”motvirke” elevenes eventuelt positive vurdering av testens subjektive verdi. Disse antakelsene kunne, i så fall, til en viss grad forklare de lave r -verdiene i tabell 5.8. Men

samtidig er det viktig å påpeke at svak korrelasjon mellom variablene kan være et tilfeldig utslag av deres variasjon i det utvalget i denne undersøkelsen er stort.

Til slutt skal vi undersøke sammenheng mellom testmotivasjonskonstruktet og faglig interesse som fagvalgbe grunnelse. Korrelasjonsanalyse viser at denne sammenhengen er statistisk signifikant, men i samsvar med resultatene i tabell 5.8 er den svak. Resultatet av korrelasjonsanalysen er vist i tabell 5.9.

Tabell 5.9 Korrelasjon mellom faglig interesse som begrunnelse for å velge fordypning i matematikk og motivasjonsskår. To stjerner indikerer at korrelasjon er signifikant på 0,01-nivå.

	Liker å løse matematiske problemer	Matematikktimene er interessante
Testmotivasjonsskår	$r = 0,18^{**}$	$r = 0,17^{**}$

Positiv r-verdi viser at å oppgi faglig interesse som veldig viktig fagvalgbe grunnelse hører sammen med høyere testmotivasjonsnivå, med andre ord har faglig interesse en positiv sammenheng med testmotivasjon. Det er vanskelig å trekke slutninger om denne sammenhengen fordi den er ganske svak og kan være et tilfeldig resultat av variasjonen i datamaterialet fordi utvalget er stort. Dessuten er faglig interesse er en av flere faktorer som kan påvirke vurderingen av testens verdi hos elevene og dermed deres testmotivasjon, og forholdet mellom de ulike verdissettingskomponentene er komplisert. For å belyse dette samspillet trenges det videre undersøkelser, og for å kunne si noe om for eksempel hvor stor betydning faglig interesse har i forhold til de andre verdissettingsfaktorer som påvirker testmotivasjon hos norske elever i kontekst av TIMSS Advanced tester, trenges det metode som kan måle dette. Med forbehold om at de påviste sammenhengene er svake kan jeg oppsummere resultatene ved å antyde at matematikkelever som er interessert i faget og liker å løse matematiske problemer vil sannsynlig ha høyere testmotivasjon og vil vise bedre innsats og utholdenhet på en matematikktest, selv om de anser testen som mindre viktig for dem. Men det gjenstår å undersøke dette videre.

Faglig selvtillitt

Korrelasjonsanalysen av sammenhengene mellom elevenes faglig selvtillitt og de seks påstandene om deres motivasjon og innsats på testen viser noe nevneverdig sammenheng bare mellom "Gjør det bra i matematikk" som fagvalgbe grunnelse og to av de seks påstandene i testmotivasjonskonstruktet. Resultatene av korrelasjonsanalysen er presentert i tabell 5.10. I

den samme tabellen er det også oppgitt korrelasjonskoeffisienten for ”Gjør det bra i matematikk” og testmotivasjonskonstruktet som samlevariabel.

Tabell 5.10 Korrelasjon mellom faglig selvtillitt som begrunnelse for å velge fordypning i matematikk og ulike påstander om motivasjon og innsats og motivasjonsskår. Det er tatt med bare sammenhenger hvor korrelasjonskoeffisienter er større enn 0,1. To stjerner indikerer at korrelasjon er signifikant på 0,01- nivå.

	Gjør det bra i matematikk
Jeg var motivert for å gjøre mitt beste på disse oppgavene	$r = 0,14^{**}$
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige	$r = 0,13^{**}$
Testmotivasjonsskår (samlevariabel)	$r = 0,13^{**}$

Korrelasjonskoeffisientene i tabellen viser at å oppgi fagvalgbe grunnelsen ”Gjør det bra i matematikk” har positive sammenhenger med utholdenhet på testen, motivasjon for å gjøre sitt beste og også testmotivasjonskonstruktet. Selv om sammenhengene er ganske svake er det ikke uventet at når korrelasjonsanalyse først viser noe nevneverdig sammenheng mellom faglig selvtillitt og motivasjonsvariablene i spørreskjema, så gjelder det utholdenhet på testen og motivasjon for å gjøre sitt beste. Marit Kjærnsli har påpekt at ”elevens forventning om å mestre et spesifikt problem har avgjørende betydning for den innsatsen og utholdenheten som investeres” (Kjærnsli, 2007, s.11). Det kan tenkes at elever som har høy faglig selvtillitt har høyere forventning om å lykkes med oppgavene i testen og dermed viser høyere utholdenhetsnivå. Men til tross for dette resonnermentet synes det å være vanskelig å trekke noen slutninger rundt sammenhengene mellom faglig selvtillitt som fagvalgbe grunnelse og testmotivasjon hos norske matematikkelever på TIMSS-testen. I analysene i TIMSS Advanced 2008 er det vist at faglig selvtillitt som fagvalgbe grunnelse hos de norske elevene hadde en relativt sterk korrelasjon med deres prestasjon på TIMSS-testen (Grønmo m.fl., 2010), men analysene i dette avsnittet viser at sammenhengen mellom faglig selvtillitt som be grunnelse for fagvalget hos norske elever og deres testmotivasjon i TIMSS Advanced er ikke tydelig. Som det er forklart i kapittel 3, er forventningsfaktorens rolle for testmotivasjonskonstruktet i low-stakes testkontekster ennå ikke er helt forstått, og resultater av empirisk forskning kan tyde på at personlighetstrekk hos elever (som for eksempel å være samvittighetsfull eller flittig) kan ha større betydning for testmotivasjon i low-stakes tester enn for eksempel deres forventning om å lykkes med testoppgavene (Barry & Finnley, 2009).

Jeg vil jeg også påpeke at begrunnelsen av fagvalget med faglig selvtilitt gir bare en indirekte indikasjon på elevenes faglig selvtilitt. TIMSS Advanced 2008 ble gjennomført i slutten av skoleåret, og på det tidspunktet kunne faglig selvtilitt i matematikk 3MX hos elevene være noe annerledes enn den selvoppfatningen i matematikk elevene hadde og vektlagt ved valget av 3MX. Det kan hende at bakgrunnsvariabler som måler elevenes selvoppfatning i faget direkte ville ha gitt et klarere bilde av relasjon mellom faglig selvtilitt hos elevene og deres testmotivasjon.

Lærerens rolle

Det er ikke funnet sammenheng mellom elevenes matematikkprestasjoner og deres begrunnelser av valgfaget med gode lærere, verken i Norge eller internasjonalt. *”There was no consistent relationship across countries between mathematics achievement and reporting that good teaching was an important reason for studying advanced mathematics.”* (Mullis m.fl., 2009, s.138). Men fordi det er en relativt stor andel av norske elever som har lagt vekt på at *”Det er gode lærer i matematikk”* ved sitt valg av faget (Grønmo m.fl., 2010) synes det allikevel å være interessant å undersøke om motivasjon på testen har en sammenheng med denne fagvalgbe grunnelsen.

Jeg har utført en korrelasjonsanalyse, og i likhet med faglig selvtilitt, har fagvalgbe grunnelsen *”Gode Lærere”* en sammenheng med motivasjon for å gjøre sitt beste og utholdenhet på testen. Korrelasjonsanalyse viser signifikante men (igjen) ganske svake sammenhenger mellom disse variablene. Resultatet er vist i tabell 5.11, sammen med korrelasjonskoeffisienten for testmotivasjonsskår.

Tabell 5.11 Korrelasjon mellom ”Gode lærere” som begrunnelse for å velge fordypning i matematikk motivasjon og innsats på testen. Det er tatt med bare sammenhenger hvor korrelasjonskoeffisienter er større enn 0,1. To stjerner indikerer at korrelasjon er signifikant på 0,01- nivå.

	Gode lærere
Jeg var motivert for å gjøre mitt beste på disse oppgavene	$r = 0,13^{**}$
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige	$r = 0,11^{**}$
Testmotivasjonsskår (samlevariabel)	$r = 0,12^{**}$

Disse resultatene viser at å oppgi gode lærere som veldig viktig begrunnelse for fagvalget har positive sammenhenger med utholdenhet på testen, motivasjon for å gjøre sitt beste og også testmotivasjonskonstruktet. Disse positive sammenhengene kan mulig forklares med at gode lærere bidrar til å fremme interesse for og engasjement i faget hos elevene, og analysene i dette avsnittet viser tendens til at faglig interesse i matematikk hos norske elever har en positiv sammenheng med deres motivasjon og utholdenhet på TIMSS-testen. Antakelsen om at faglig interesse kan forbindes med gode lærere kan i en viss grad støttes av en korrelasjonsanalyse som jeg har gjennomført og som viser positive sammenhenger med $r = 0,44^{**}$ mellom "Gode lærere" og "Liker å løse oppgaver" og $r = 0,38^{**}$ mellom "Gode lærere" og "Matematikktimene er interessante". Men tatt i betraktning at korrelasjonskoeffisientene i tabell 5.11 er ganske svake, synes det vanskelig å trekke noen slutninger om resultatene i denne korrelasjonsanalysen.

En kort oppsummering:

Det er vist at norske elever som presterer bra på matematikktesten i TIMSS Advanced, kjennetegnes ved blant annet å ha positiv holdning til og høy selvoppfatning i faget (Grønmo m.fl., 2010). Analyseresultatene i dette avsnittet kan mulig tyde på at norske matematikkelever som er interessert i faget og har høy faglig selvtilitt har høyere testmotivasjon og viser bedre innsats og utholdenhet på testen. Vi kan også antyde at gode lærere kan gjennom utvikling av interesse for matematikk og faglig selvtilitt hos elevene bidra til å styrke elevenes faglig standhaftighet som kan føre til bedre utholdenhet i faglige utfordringer. Men tatt i betraktning at de påviste sammenhengene i analysene er ganske svake, er det vanskelig å trekke noen slutninger og det bør gjennomføres flere undersøkelser for å kunne oppklare sammenheng mellom elevenes holdninger og selvoppfatning i matematikk og deres motivasjon og innsats i TIMSS Advanced kontekst.

5.4 Prestasjonsresultater i matematikktesten i TIMSS Advanced sett i sammenheng med elevenes selvrapporterte testmotivasjon

Dette avsnittet dreier seg om følgende forskningsspørsmål i problemstillingen i denne masteroppgaven: hvordan presterer norske elever sett i sammenheng med deres

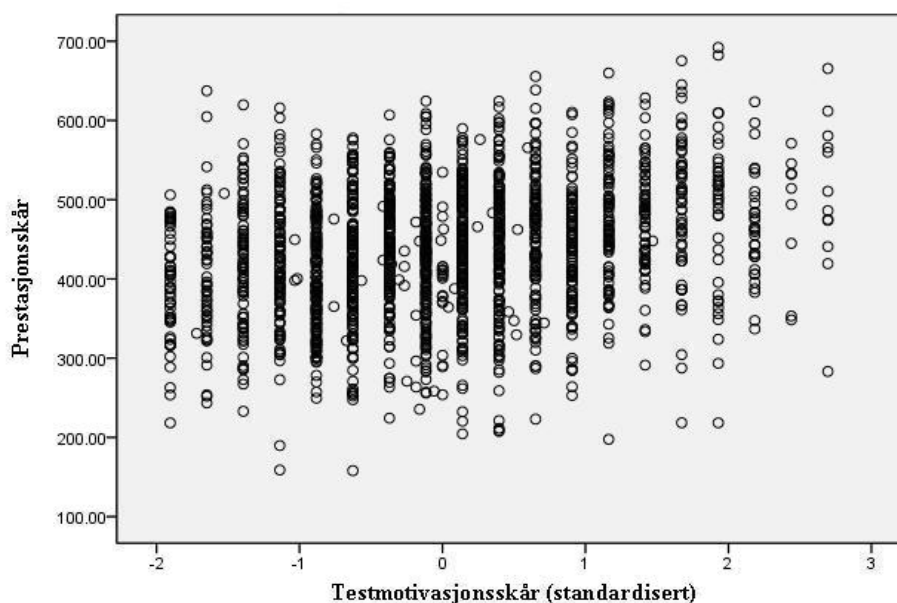
selvrapporterte motivasjon på testen? Sammenhengen er undersøkt ved korrelasjonsanalyse, og resultatet er vist i tabell 5.12.

Tabell 5.12 Korrelasjon mellom prestasjon på matematikktesten og testmotivasjon. To stjerner indikerer at korrelasjon er signifikant på 0,01- nivå.

	Testmotivasjonskår
Testskår	$r = 0,26^{**}$

Den positive korrelasjonskoeffisienten viser at høye prestasjonsskår på matematikktesten hører sammen med høyere testmotivasjonskår. Denne sammenhengen er signifikant, og $r = 0,26$ viser at omtrent 6,8 prosent av variansen i testskårverdien kan ”forklares” av elevenes testmotivasjon. I de andre storskalastudiene hvor det ble foretatt testmotivasjonsmålinger, er de tilsvarende korrelasjonskoeffisientene: 0,21 ($p < 0,01$) i TIMSS 2003 i Sverige, 0,20 ($p < 0,01$) i PISA 2009 i Norge og 0,30²⁵ i TIMSS Advanced 2008 i Sverige (se henholdsvis Eklöf, 2006b, Kjærnsli & Roe, 2010 og Eklöf, 2010b).

Som vi ser av r-verdien er sammenhengen mellom prestasjonsskår og testmotivasjonskår svak, og vi kan illustrere denne sammenhengen ved et plot i figur 5.5. Vi kan se at både elever med høy og lav testskår rapporterer om lav testmotivasjon og det er stor spredning i testskår hos forskjellige elever med like testmotivasjonskår.



Figur 5.5 Samvariasjon mellom testmotivasjon og prestasjonsskår.

²⁵ Sammenhengen er signifikant, men signifikansnivå er ikke oppgitt (se Eklöf, 2010a)

Selv om analysen i dette avsnittet viser en positiv og signifikant sammenheng mellom elevenes matematikkprestasjoner og deres testmotivasjon i TIMSS Advanced, er sammenhengen på ingen måte tydelig. For å kunne trekke noen slutninger er det åpenbart et behov for grundigere analyser som undersøker hvordan sammenheng mellom norske elevers matematikkprestasjoner og deres testmotivasjon i TIMSS Advanced 2008 er, noe som rammene for denne oppgaven ikke gir rom for.

Kjønnsforskjeller

I Norge er det ikke funnet signifikante kjønnsforskjeller i matematikkprestasjoner i TIMSS Advanced 2008 (Grønmo m.fl., 2010). Som det er vist tidligere i dette kapitlet, er det heller ikke funnet signifikante kjønnsforskjeller i testmotivasjon hos norske elever, men i gjennomsnitt er guttene noe mer positive i sin rapportering av innsats og utholdenhet på testen enn jentene. Sammenhengen mellom testskår og testmotivasjonsskår hos jenter og gutter er undersøkt ved korrelasjonsanalyse og resultatet er vist i tabell 5.13.

Tabell 5.13 Korrelasjon mellom prestasjon og motivasjon, splittet på kjønn. To stjerner indikerer at korrelasjon er signifikant på 0,01- nivå.

	Jenter	Gutter
Testmotivasjonsskår	$r = 0,23^{**}$	$r = 0,28^{**}$

Av korrelasjonskoeffisientene i tabellen ser vi at hos jentene kan 5,3% av testskår knyttes til variasjon i deres testmotivasjonsskår, mens hos guttene er den tilsvarende samvariasjonen på 7,8%. Forskjellen er ganske liten, og kan være et tilfeldig utslag av variablenes variasjon i det store utvalget som ble testet i TIMSS Advanced 2008. Analysene i dette kapitlet tilsier at det ikke er funnet kjønnsforskjeller i testmotivasjon hos norske matematikkløvere i TIMSS Advanced 2008, og kjønnsmessig er sammenhengen mellom elevenes testmotivasjon og prestasjon på testen tilnærmet lik. Foreløpig vet vi lite om kjønnsforskjeller når det kommer til sammenhengen mellom testmotivasjon og prestasjon (Eklöf, 2006b, 2010b), og behovet for videre forskning innen dette område synes å være tydelig.

5.5 Testmotivasjon hos norske elever på matematikktesten i TIMSS Advanced 2008 sammenliknet med svenske elever

Som det er nevnt tidligere er det fire deltakerland i TIMSS Advanced 2008 som har valgt å inkludere spørsmål om testmotivasjon i sine nasjonale elevspørreskjemaer. Så langt er det lagt fram foreløpige analyser av elevenes svar bare i Sverige. Resultatene er publisert i flere artikler av H.Eklöf (2010a, 2010b), og testmotivasjon er også tatt med som bakgrunnsvariabel i analyser av faktorer som hadde innvirkning på elevenes testprestasjoner i den offentlige rapporten TIMSS Advanced 2008 i Sverige (Skolverket, 2009). Artiklene til Eklöf omfatter både pilotstudien og hovedstudien TIMSS Advanced 2008, og begge studiene er omtalt tidligere i kapittel 4. Resultatene fra pilotstudien gjelder både de elevene som tok matematikktesten og de som tok fysikktesten, uten at de er behandlet hver for seg. Allikevel synes det å være relevant å trekke inn disse resultatene som referanse når vi sammenlikner testmotivasjon hos norske og svenske elever på matematikktesten i hovedundersøkelsen, fordi analyser og resultater fra pilotstudien ble lagt til grunn ved utformingen av motivasjonsskalaen i elevspørreskjema i TIMSS Advanced 2008.

Resultatene fra pilotstudien er blant annet omtalt som prosentandeler av elever som sier seg ”helt enig” eller ”litt enig” i hver av de seks påstandene i spørreskjemaet (se Eklöf, 2010a), som gjør det mulig å sammenlikne disse resultatene med tilsvarende ”norske” prosentandeler i tabell 5.1. Tabell 5.16 viser de norske og de svenske prosentandelene i hver påstand.

I den offentlige hovedrapporten i Sverige kommenteres det at 60 prosent av de svenske elevene som tok matematikktesten i TIMSS Advanced 2008 (hovedstudien) svarte at de gjorde sitt beste på testen, samtidig som 3 av 4 sa at de anstrengte seg mindre på TIMSS-testen enn de ellers gjør på andre skoleprøver (Skolverket, 2009). Sammenliknet med prosentandelene i tabell 5.16 for P1 og P3, synes disse tallene å kunne indikere at fordelingen av elevresponsene på de enkelpåstandene i spørreskjema i hovedstudien har omtrent det samme mønsteret som i pilotstudien. Deskriptiv statistikk, vist i tabell 5.17, ser ut til å bekrefte denne antakelsen. Den viser også at elevene i TIMSS Advanced-hovedstudien i Sverige rapporterte om omtrent samme motivasjons- og innsatsnivå på testen som elevene i

pilotstudien (Eklöf, 2010a). Deskriptiv statistikk ved gjennomsnittsverdier og standardavvik for Norge og Sverige er vist i tabell 5.17.

Tabell 5.16 Prosentandel av norske og svenske elever som sier seg "helt enig" eller "litt enig" i hver av de seks påstandene i elevspørreskjema. Sorteringen i tabellen tilsvare oppsettet i tabell 5.1. (Hver påstand er merket med nummer i parentesene etter påstandenes rekkefølge i spørreskjema.)

	Norge (n = 1912)	Sverige ²⁶ (n = 163)
Jeg kunne ha anstrengt meg mer da jeg jobbet med disse oppgavene (P6)	84	82
Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året (P3)	82	78
Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene (P2)	68	over 50 ²⁷
<hr/>		
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen (P1)	68	57
Jeg var motivert for å gjøre mitt beste på disse oppgavene (P5)	48	42
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige (P2)	44	34

Tallene i tabell 5.17 viser at norske elever har noe høyere gjennomsnittsverdier i responsene på påstandene med positiv ordlyd og noe lavere verdier i påstandene med negativ ordlyd. Det er vanskelig å kommentere denne forskjellen ytterligere, fordi testmotivasjonsmålinger på tvers av landene er nytt i TIMSS Advanced og vi har foreløpig lite erfaring med og lite kunnskap om elevenes testmotivasjon i store internasjonale komparative undersøkelser. Kulturelle forskjeller kan være én av de potensielle årsakene til eventuelle forskjeller i responsmønstrene mellom land. *"There seem to be cultural differences in how students treat the response scales, where students in some countries are reluctant to use the extreme ends of the scales"* (Eklöf, 2006b, s.17). For å skaffe forståelse av testmotivasjonsfenomenet og kunnskap om de forskjellige aspektene ved testmotivasjonsmålinger i internasjonale undersøkelser på tvers av landene synes det å være viktig at ved de kommende TIMSS-undersøkelsene blir de første testmotivasjonsmålingene i TIMSS Advanced 2008 fulgt opp videre og den kvantitative metoden av testmotivasjon hos elevene som ble brukt i undersøkelsen i 2008, blir utvidet med kvalitativ forskning.

²⁶ Pilotstudien, se Eklöf 2010a

²⁷ Den eksakte prosentandelen er ikke oppgitt: "more than half of the students claimed that they were not concentrated while working with test items." (Eklöf, 2010a, s.6).

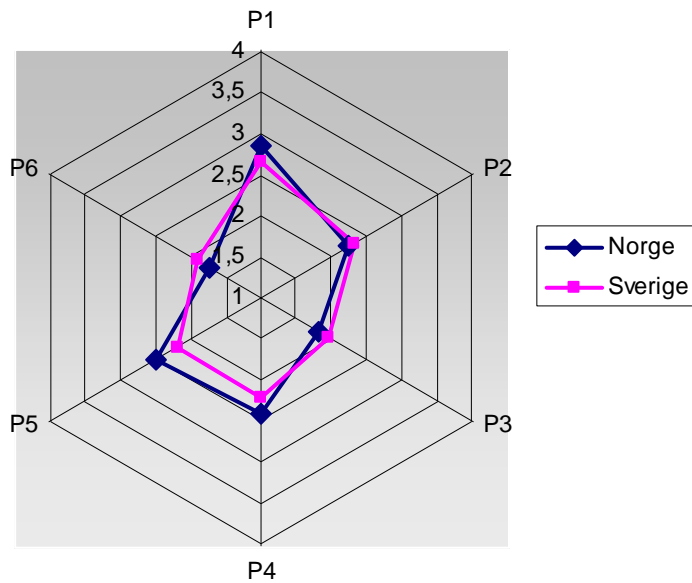
Tabell 5.17 *Deskriptiv statistikk av testmotivasjonsmålinger i Norge og i Sverige. Tallene i kolonne for Sverige viser resultater fra hovedstudien av matematikkelever og fra pilotstudien. Lavere gjennomsnittsverdier (M-verdier) i tabellen indikerer mer negativ respons i forhold til motivasjon og innsats på testen. (1 er lavest, og 4 er høyest)*

	Norge (n = 1912)		Sverige ²⁸			
			Hovedstudien (n = 2244)		Pilotstudien (n = 163)	
	M	SD	M	SD	M	SD
Jeg gjorde mitt aller beste på oppgavene i denne undersøkelsen (P1)	2.83	0.89	2.65	0.82	2.61	0.83
Jeg var motivert for å gjøre mitt beste på disse oppgavene (P5)	2.48	0.98	2.18	0.88	2.30	0.83
Jeg arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige(P4)	2.41	0.94	2.21	0.83	2.26	0.82
<i>* Verdiene for de tre siste utsagnene er ” snudd”, se kapittel 4.</i>						
Jeg var ikke helt konsentrert da jeg arbeidet med disse oppgavene (P2)	2.23	0.93	2.31	0.80	2.44	0.85
Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året (P3)	1.79	0.89	1.95	0.87	1.88	0.87
Jeg kunne ha anstrengt meg mer da jeg jobbet med disse oppgavene (P6)	1.73	0.84	1.90	0.85	1.88	0.76

Tallene i tabell 5.17 er illustrert i figur 5.6, hvor det er satt opp diagram og 95 % - konfidensintervaller for elevenes respons på enkeltpåstandene i spørreskjema i TIMSS Advanced 2008 i Norge og i Sverige (hovedstudie).

Det er ikke oppgitt den gjennomsnittelige testmotivasjonsskåren for svenske elever i kildene som er brukt i denne oppgaven. Jeg kan bare referere til H.Eklöf (2010b) som konkluderer med at svenske elever i både pilotstudie og hovedstudie TIMSS Advanced 2008 har rapportert om lavere nivå av motivasjon og innsats på testen: ”... *the Grade 12 students participating in the Main Study also reported a rather low level of effort on the test*” (Eklöf, 2010b, s.6) og “*preliminary results suggest that the Swedish students participating in the Main Study were not more motivated and did not spend more effort than the students participating in the field-test*” (Eklöf, 2010b, s.9).

²⁸ Se Eklöf 2010a



95% -konf.intervall

P	Norge	Sverige
1	2,79 – 2,87	2,63 – 2,68
2	2,19 – 2,27	2,27 – 2,35
3	1,69 – 1,83	1,91 – 1,99
4	2,37 – 2,45	2,18 – 2,24
5	2,44 – 2,52	2,14 – 2,22
6	1,69 – 1,77	1,86 – 1,94

Figur 5.6 Diagram over gjennomsnittet i norske og svenske elevers respons på hver av de seks enkelpåstandene om motivasjon og innsats. Lavere gjennomsnittsverdier på diagrammet indikerer mer negativ respons i forhold til motivasjon og innsats på testen. (1 er lavest, og 4 er høyest). Til høyre er det beregnet 95% - konfidensintervall for hver påstand i TIMSS Advanced 2008.

Analysene i dette kapitlet tyder på at norske matematikkelever hadde i underkant av middels testmotivasjonsnivå, og av tabell 5.17 og figur 5.6 ser vi at det er ganske små forskjeller mellom de norske og de svenske matematikkelevers angivelser av motivasjon og innsats på TIMSS-testen, og dette synes å indikere om omtrent samme testmotivasjonsnivå i TIMSS Advanced 2008 hos matematikkelevne i Norge og i Sverige.

6 Oppsummering

Målet med denne masteroppgaven er å undersøke norske 3MX-elevens motivasjon og innsats på matematikktesten i TIMSS Advanced 2008 basert på hva de har svart på spørsmål om dette etter at de var ferdig med testen. Jeg undersøker også sammenhengen mellom det elevene svarer om egen innsats og motivasjon i TIMSS Advanced og hva de har svart på spørsmål om valget av fordypning i matematikk. Jeg gjør i tillegg noen sammenlikninger hvor jeg undersøker om det er forskjell mellom jenter og gutter på de ovennevnte punktene, og gjør også noen sammenlikninger mellom Norge og Sverige.

6.1 Elevenes svar på påstander om motivasjon og innsats på TIMSS Advanced testen

Analysene av elevenes angivelse av sin motivasjon og innsats på testen viser at bare under halvparten av 3MX-elevene sier seg helt enige eller litt enige i at de var motivert for å gjøre sitt beste på TIMSS-testen, og i at de arbeidet uten å gi opp med alle oppgavene selv om de var vanskelige. Omtrent 4 av 5 elever sier at de kunne ha anstrengt seg mer og at de anstrengte seg mindre på denne testen enn på prøver ellers i året. Mange elever (68 %) sier at de ikke var helt konsentrert under testen, samtidig som en relativt stor elevandel (68 %) (tatt i betraktning elevenes respons på de allerede nevnte påstandene) sier seg helt enig eller delvis enig i at de gjorde sitt beste på testen.

Elevenes svar på enkeltpåstandene om motivasjon og innsats i elevspørreskjema kan tyde på at norske 3MX-elever i gjennomsnitt hadde noe lavere testmotivasjonsnivå på matematikktesten i TIMSS Advanced enn på andre tester de har i løpet av året. Jeg vil påpeke at mine analyser av datamaterialet i TIMSS Advanced ikke gir mulighet til å trekke noen slutninger om i hvilken grad lavere testmotivasjonsnivå hos norske 3MX-elever kan forklares med hvor viktig de vurderte testen til å være.

Resultater i internasjonale undersøkelser som TIMSS Advanced har ikke direkte konsekvenser for de elevene som tar testen. At de ble lagt inn spørsmål til elevene i noen land som Norge om deres motivasjon og innsats er blant annet begrunnet med bekymringen for at den lave statusen testen har for elevene kan føre til lav motivasjon hos dem for å gjøre sitt beste. Det kan forårsake testresultater som underestimerer elevenes kunnskapsnivå (Eklöf,

2006b; 2010a, b). I TIMSS Advanced ble ikke elevene spurt direkte om deres oppfatninger av hvor viktig testen var for dem. Påstanden ” Jeg anstrengte meg mindre da jeg arbeidet med disse oppgavene enn det jeg gjør på prøver ellers i året” synes å kunne gi oss en indirekte indikasjon på elevenes holdninger til testen. 82 prosent av elevene sier seg helt enig eller delvis enig i denne påstanden. Dette kan gi inntrykk av at mange elever vurderte TIMSS-testen som mindre viktig for dem, og at de derfor anstrengte seg mindre enn det de gjør på andre prøver. Korrelasjonsanalysene av elevenes respons på enkeltpåstandene i testmotivasjonskonstruktet tyder på at sammenhenger mellom de underliggende faktorer som påvirker norske elevers motivasjon og innsats på TIMSS-testen er kompliserte. Ca. 30 prosent av variasjon i elevenes rapportering av lavere innsats og utholdenhet på testen kunne forklares med hva de svarte om egen motivasjon til å gjøre sitt beste. Ca. 20 prosent av variasjon i elevenes rapportering av egen motivasjon for å gjøre sitt beste kunne knyttes til deres svar på påstanden om at de anstrengte seg mindre på TIMSS-testen enn på andre tester.

For bedre å kunne forstå testmotivasjon hos norske elever i kontekster av typen TIMSS trengs det mer forskning på området. I de kommende TIMSS undersøkelsene kunne man kombinere kvantitative målemetoder, som elevenes selvrapporing i etterkant av testen, med kvalitative målemetoder som for eksempel intervjuer og observasjoner av elevenes innsats mens de jobbet med testen. ”*There is no best method for assessing test-taking behaviour. As most methods for measuring constructs such as motivation have drawbacks, using more than one measure and more than one method may lend credibility to score interpretation and use and might add to the quality of the assessment.*” (Eklöf, 2010b, s.354). Det bør også påpekes at elevenes oppfatninger av hvordan testen har gått, kan påvirke hvordan de i etterkant av testen svarer på spørsmål om motivasjon og innsats. Det kan derfor være hensiktsmessig å måle testmotivasjon også i forkant av en test.

6.2 Testmotivasjon og noen bakgrunnsvariabler

Til tross for at analysene i TIMSS Advanced både i Norge og internasjonalt har påvist en signifikant og relativt sterk sammenheng mellom elevenes prestasjoner på matematikktesten og deres begrunnelser for valg av faget med interesse for og høy selvoppfatning i faget, viser ikke analysene i denne oppgaven noen tydelig sammenheng mellom elevenes angitte motivasjon og innsats på testen og de ovennevnte fagvalgbe grunnelsene. Det kan være flere grunner til dette.

Selvoppfatning i faget og faglig interesse er i følge det teoretiske fundamentet for prestasjonsmotivasjon viktige for elevenes forventning om å lykkes med og verdisetning av en arbeidsoppgave, noe som påvirker elevenes motivasjon og innsats på oppgaven.

Spørreskjemaet i TIMSS Advanced måler ikke elevenes selvoppfatning i og interesse for faget direkte. Undersøkelsen ble gjennomført i slutten av skoleåret, og på det tidspunktet kunne elevenes selvoppfatning i og interesse for faget matematikk 3MX være noe annerledes enn den selvoppfatningen i matematikk elevene hadde og vektla ved valg av 3MX. Det kan være at bakgrunnsvariabler som måler elevenes selvoppfatning i og interesse for faget mer direkte, kunne gitt et klarere bilde av sammenheng mellom faglig selvtilitt og interesse hos norske 3MX-elever og deres testmotivasjon på TIMSS-testen.

Jeg vil også påpeke at samspillet mellom de forskjellige faktorer som påvirker prestasjonsmotivasjon i low-stakes testkontekster ennå ikke er helt forstått (Eklöf, 2010b) og at resultater av empirisk forskning kan tyde på at personlighetstrekk hos elever (som for eksempel å være samvittighetsfull eller flittig) kan ha større betydning for testmotivasjon i low-stakes tester enn for eksempel forventninger om å lykkes og verdisetning (Barry & Finnley, 2009). Behovet for videre forskning av testmotivasjonsfenomenet i low-stakes testkontekster synes å være ganske tydelig.

6.3 Sammenheng mellom testmotivasjon og testprestasjon

Ved korrelasjonsanalyse er det påvist en signifikant positiv sammenheng mellom den målte testmotivasjonen hos norske elever og deres prestasjon på matematikktesten i TIMSS Advanced, men denne sammenhengen er ikke sterk.

Det finnes få studier som tidligere har undersøkt sammenheng mellom testmotivasjon hos elevene og deres testprestasjon i komparative undersøkelser som TIMSS (Eklöf, 2006 b, 2010 b; Kjærnsli & Roe, 2010). Konklusjonene i disse undersøkelsene er ikke entydig. I TIMSS 2003 i Sverige ble det vist at når domenebestemte konstrukter (som selvoppfatning i og verdigsetting av faget) var holdt fast, ble ”virkningen” av testmotivasjonskonstruktet på variasjon i testskår ubetydelig, og det ble konkludert med at elevenes prestasjoner ikke kunne forklares med elevenes testmotivasjon (Eklöf, 2006 b). I PISA 2009 i Norge ble det konkludert med at funnene i denne undersøkelsen var i tråd med bl.a. resultatene i TIMSS

2003 i Sverige. I den offentlige rapporten TIMSS Advanced 2008 utgitt av Skolverket i Sverige, ble innsatsindeks basert på elevenes svar på spørreskjemaet, inkludert i analyser av faktorer som påvirket svenske elevers prestasjon i matematikktesten. Denne indeksen var en av de relativt få variabler som hadde innvirkning på elevenes prestasjonsskår (Skolverket, 2009).

Det synes å være behov for flere statistiske analyser av de empiriske dataene, som er samlet inn i TIMSS Advanced og andre studier for å kunne få en bedre forståelse av sammenhengen mellom prestasjonsskår og målt testmotivasjon hos norske elever, noe rammene i denne oppgaven ikke gir rom for.

6.4 Kjønnforskjeller

Analysene i denne oppgaven viser ingen signifikante kjønnforskjeller i testmotivasjonsskår blant norske 3MX-elever i TIMSS Advanced. Det er heller ikke påvist kjønnforskjeller i analysene av sammenheng mellom testprestasjon og testmotivasjon. Vi vet foreløpig lite om kjønnforskjeller når det kommer til måling av testmotivasjon hos norske elever, både generelt og i kontekst av internasjonale komparative undersøkelser. Testmotivasjonsmålinger er ganske nytt felt i skoleforskningen, og man trenger flere studier av dette både nasjonalt og internasjonalt.

6.5 Sammenlikning av resultater for Norge og Sverige

De selvrapporterte nivåene i motivasjon og innsats på TIMSS Advanced testen for matematikkelevne i Norge er ganske likt det man har funnet i Sverige. En oppsummering av resultater i TIMSS Advanced i Sverige basert på data fra pilotstudien og hovedstudien, påpekes det at “preliminary findings from the subsequent Swedish TIMSS Advanced 2008 Main Study also reveal a modest level of motivation and effort among the students, and a significant relation between reported level of effort and achievement on the test” (Eklöf, 2010a, s.1). Det samsvarer godt med det jeg har funnet for Norge.

Vi har lite forskning på elevenes testmotivasjon i store internasjonale komparative undersøkelser, det gjelder både innad i land og mellom land. Sammenlikninger av

testresultater mellom deltakerland står sentralt i alle slike undersøkelser. Analysene av prestasjonsresultater i TIMSS Advanced viser at svenske resultater i stor grad likner på de norske. Det samsvarer godt med andre sammenlikninger som er foretatt mellom disse to landene i den norske rapporten. ”Sverige framstår som et land som har mye til felles med Norge når det gjelder mange av de faktorene” som er drøftet i den norske rapporten (Grønmo m.fl., 2010, s. 207). Testmotivasjon ser dermed ut til å føye seg i rekken av likhetsfaktorene mellom Norge og Sverige i TIMSS Advanced. Det vil være veldig interessant å se på resultater av testmotivasjonsmålinger i andre land. Russland og Slovenia stilte de samme spørsmålene til sine elever i TIMSS Advanced, men ingen resultater er så langt publisert. Ikke minst det at Russland og Slovenia er land som presterer mye bedre enn de to skandinaviske landene, gjør en utvidelse av analysene i TIMSS Advanced som også inkluderer disse landene interessant. Sammenlikninger og analyser av responsmønstrene på tvers av land og kulturer vil kunne gi viktig kunnskap om elevenes holdninger og ”testadferd” i kontekst av komparative internasjonale undersøkelser. Det vil kunne styrke validiteten av prestasjonssammenlikninger på tvers av deltakerland.

6.6 Veien videre

Gjennom min oppgave håper jeg å ha gitt et lite bidrag til at man ser betydningen av å forske, ikke bare på elevenes prestasjoner i internasjonale komparative studier, men også å se på andre faktorer som testmotivasjon og innstas. Det er flere aspekter ved videre forskning som synes å peke seg ut. Man kan vurdere å ta i bruk forskjellige metoder for å måle testmotivasjon. I tillegg til selvrapportering i etterkant av testen slik som i TIMSS Advanced 2008, kunne man gjøre observasjoner underveis mens elevene arbeider med testen, eller man kunne stille spørsmål i forkant av testen. Studier som bruker flere metoder for å måle motivasjon og innsats vil kunne bidra med informasjon som kan forbedre måten dette måles på. Det kan føre til utvikling av mer pålitelige måleinstrument for testmotivasjon som kan implementeres i internasjonale komparative studier. Det er viktig med mer kunnskap om sammenhengen mellom teststatus, elevenes testmotivasjon og deres prestasjon på testen, ikke minst gjelder det i store komparative internasjonale undersøkelser. Også det å inkludere flere land i testmotivasjonsmålinger enn de fire som gjorde dette i TIMSS Advanced 2008 vil kunne gi viktige bidrag til forskningen på dette feltet. Man kan da tenke seg at alle deltakerland har med denne typen spørsmål. Analyser på tvers av land og kulturer kan gi

viktige bidrag i denne forskningen. Jeg vil avslutte denne oppgaven med følgende sitat: "*It is suggested that measurement practitioners routinely apply motivation filtering whenever the data from low-stakes tests are used to support program decisions.*" (Wise & DeMars, 2008, s.2)

Litteraturliste

- AERA (American Educational Research Association) (2000). *Position statement of the American Educational Research Association concerning high-stakes testing in PreK12 education*. Educational Researcher, Vol.29, No.8.
- Angell C., Kjærnsli M. & Lie S. (1999). *Hva i all verden skjer i realfagene i videregående skole?* Oslo, Universitetsforlaget.
- Arora, A., Foy, P., Martin, M.O., & Mullis, I.V.S. (Eds.) (2009). *TIMSS Advanced 2008 Technical Report*. Chestnut Hill, MA, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. Lynch school of education, Boston College. [Online] (Hentet fra: http://timssandpirls.bc.edu/timss_advanced/downloads/TA08_Technical_Report.pdf 17.02.11)
- Barry C. & Finney S. (2009). Does it Matter How Data are Collected? A Comparison of Testing Conditions and the Implications for Validity. *Research & Practice in Assessment* 3(1), [Online] (Hentet fra: <http://www.virginiaassessment.org/rpa/4/index.php> 10.02.11)
- Barry C., Horst S.J., Finney S., Brown A. & Kopp J. (2010). Do examinees have similar test-taking effort? A high-stakes question for low-stakes testing. *International journal of testing*, 10(4), 342 – 363
- Baumert J. & Demmrich A. (2001). Test Motivation in the Assessment of Students Skills: The Effects of Incentives on Motivation and Performance. *European Journal of Psychology of Education*, XVI(3), 441-462.
- Bjørkeng B. (2011). *Jenter og realfag i videregående opplæring*. Statistisk sentralbyrå, 2011, Rapport 3/11 [Online] (Hentet fra: http://www.ssb.no/emner/04/02/30/rapp_201103/, 19.04.11)

- Brandell G. & Staberg E.-M. (2008). Mathematics: a female, male or gender-neutral domain? A study of attitudes among students at secondary level. *Gender and Education*, 20(5), 495 – 509.
- Eklöf H. (2006a). Development and validation of scores from an instrument measuring student test-taking motivation. *Educational and Psychological Measurement*, 66(4), 643 – 656.
- Eklöf H. (2006b). *Motivational Beliefs in the TIMSS 2003 Context*. PhD-avhandling. Department of Educational Measurement, Umeå University.
- Eklöf H. (2007). Testtaking Motivation and Mathematics Performance in TIMSS 2003. *International Journal of Testing*, 7(3), 311 – 326.
- Eklöf H. (2010a). *Student Motivation and Effort in the Swedish TIMSS Advanced Field Study*. Conference Paper, The 4th IEA International Research Conference. [Online] (Hentet fra: http://www.iea-irc.org/index.php?id=timss_advanced, 09.02.11)
- Eklöf H. (2010b). Skill and will: test-taking motivation and assessment quality. *Assessment in Education: Principles, Policy & Practice*, 17(4), 345 -356.
- Giota J.(2006). Om elevers motivation, kompetens och prestasjoner i skolan. *Pedagogisk Forskning i Sverige*, 11(2), 94–115.
- Grønmo L.S., Onstad T. & Pedersen I.F. (2010). *Matematikk i motvind. TIMSS Advanced 2008 i videregående skole*. Oslo, Unipub.
- IEA (©2007a). [Online] Lastet ned 19.04.11 http://www.iea.nl/brief_history_of_iea.html
- IEA (©2007b). [Online] Lastet ned 19.04.11 <http://www.iea.nl/studies.html>
- Johannesen A., Tufte P.A. & Kristoffersen L. (2009). *Introduksjon til samfunnsvitenskapelig metode*. Oslo, Abstrakt forlag.
- Kjærnsli M. (2007). Hva får vi vite fra PISA 2006? *Acta Didactica Norge*, 1(1), [Online] (Hentet fra: <http://adno.no/index.php/adno/article/view/22/63> 13.02.11)

- Kjærnsli M. & A.Roe (2010). *På rett spor. Norske elevers kompetanse i lesing, matematikk og naturfag i PISA 2009*. Oslo, Universitetsforlaget.
- Mullis, I.V.S., Martin, M.O., Robitaille, D.F. & Foy, P. (2009). *TIMSS Advanced 2008 International Report: Findings from IEA's Study of Achievement in Advanced Mathematics and Physics in the Final Year of Secondary School*. Chestnut Hill, MA, TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College. [Online] (Hentet fra: http://timssandpirls.bc.edu/timss_advanced/downloads/TA08_International_Report.pdf 17.02.11)
- Sundre, D. L (1999). *Does examinee motivation moderate the relationship between test consequences and test performance?* Paper presented at the annual meeting of the American Educational Research Association, Montreal, Quebec, Canada. [Online] (Hentet fra: www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED432588, 17.02.11)
- Sundre, D. L., & Moore, D. L. (2002). The Student Opinion Scale: A measure of examinee motivation. *Assessment Update*, 14, 8-9.
- Schunk, D.H., Pintrich, P.R. & Meece, J.L. (2010). *Motivation in education: Theory, Research, and Applications*. 3rd ed. Prentice Hall. Pearson Education International.
- Skolverket (2009). *TIMSS Advanced 2008. Huvudrapport*. Stockholm: Skolverket [Online] (Hentet fra: <http://www.skolverket.se/publikationer?id=2291> 10.02.11)
- Thelk A.D., Sundre D.L., Horst S. & Finney S. (2009). Motivation Matters: Using the Student Opinion Scale to Make Valid Inferences about Student Performance. *The journal of general education*, 58(3), 129 – 151.
- TIMSS Advanced 2008. *Elevspørreskjema. Matematikk*. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.
- UDIR (2009). *Sluttrapport, Oppdragsbrev nr. 6 – 2007 om tiltak knyttet til individvurdering i skole og fag- og yrkesopplæring*. Oslo: Utdanningsdirektoratet.

- Wigfield A. (1994). Expectancy-value theory of achievement motivation: a developmental perspective. *Educational Psychology Review* 6(1), 49 - 78.
- Wigfield, A. & Eccles, J.S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In *Development of achievement motivation*, ed. A.Wigfield & J.Eccles, 92-120. New York, NY: Academic Press.
- Wise S.L.(2009). Strategies for managing the problem of unmotivated examinees in low-stakes testing programs. *The journal of general education*, 58(3), 152 – 166.
- Wise S.L. & DeMars Ch.E.(2003). *Examinee Motivation in Low-Stakes Assessment: Problems and Potential Solutions*. Paper presented at the annual meeting of the American Association of Higher Education Assessment Conference, Seattle, WA [Online] (Hentet fra: www.jmu.edu/assessment/wm_library/Examinee_Motivation.pdf 17.02.11)
- Wise, S.L., & DeMars Ch.E.(2005). Low examinee effort in low-stakes assessment: Problems and possible solutions. *Educational Assessment*, 10(1), 1 – 17.
- Wise S.L. & DeMars Ch.E.(2008). Examinee Non-Effort and the Validity of Program Assessment Results. Paper presented at the annual meeting of the National Council on Measurement in Education, New York. [Online] (Hentet fra: www.jmu.edu/assessment/SLW%20Impact%20study%20ms%20NCME.pdf 27.02.11)
- Wise, S.L. & Kong X. (2005). Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163 – 183.
- Wolf L.F. & Smith J.K. (1995). The consequence of consequence: Motivation, anxiety, and test performance. *Applied Measurement in Education*, 8(3), 227 – 242.