

2 MÅL, METODER OG GJENNOMFØRING

Etter at vi i det første kapitlet har vist noen hovedresultater, er det nå tid for å forklare hvor alle disse dataene kommer fra, hva slags undersøkelse dette er, og hvordan den er gjennomført. Dette kapitlet dreier seg derfor om å beskrive mål og metoder for TIMSS-undersøkelsene generelt og TIMSS 2003 spesielt. Det er mange detaljer i dette kapitlet, detaljer som hver for seg er viktige for å gi et helhetlig bilde av undersøkelsen og kvaliteten på dataene. Imidlertid er ikke alle disse detaljene nødvendige for å kunne forstå resultatene slik de presenteres i denne rapporten.

2.1 Hva er TIMSS?

2.1.1 Historikk

TIMSS (Trends in International Mathematics and Science Study) er en internasjonal undersøkelse som handler om matematikk og naturfag i skolen. Undersøkelsens kanskje viktigste mål er å beskrive og sammenlikne elevprestasjoner, så vel nasjonalt som internasjonalt, for deretter søke å forklare og forstå forskjeller i prestasjoner ut fra andre data i undersøkelsen. Man ønsker slik å kunne si noe om hvilke faktorer som best fremmer læring. Flere hundre tusen elever i vel 50 land er nå med i undersøkelsen. Dette gjør TIMSS til historiens mest omfattende komparative forskningsprosjekt som er gjennomført innen utdanning.

TIMSS administreres av den internasjonale organisasjonen IEA (International Association for the Evaluation of Educational Achievement), som er et nettverk for utdanningsforskning. IEA ble etablert i 1959. Dette skjedde i forkant av FIMS (First International Mathematics Study, se Husén 1967) som fant sted i begynnelsen av 1960-årene. Den ble etterfulgt av FISS (First International Science Study, se Comber & Keeves 1973) tidlig i 1970-årene. De påfølgende studiene SIMS (Second International Mathematics Study, se Travers & Westbury 1989, Robitaille & Garden 1989, Burstein 1992) og SISS (Second International Science Study, se IEA 1988, Postlewaite & Wiley 1992, Rosier & Keeves 1991) ble gjennomført i første halvdel av 1980-årene. Av disse deltok Norge kun i SISS (se for eksempel Sjøberg 1986). Det er vanskelig å knytte bestemte årstall til disse første store internasjonale undersøkelsene, da de ble gjennomført til forskjellige tider i de ulike landene som deltok.

Arbeidet med TIMSS tok til i 1991 under navnet ”Third International Mathematics and Science Study”, men selve undersøkelsen fant sted i 1995 med Norge som et av deltakerlandene. I 1999 ble ”TIMSS-Repeat” gjennom-

ført, men da uten norsk deltakelse. Fra og med undersøkelsen i 2003 ble TIMSS omdøpt til "Trends in International Mathematics and Science Study", og man tar nå sikte på å gjennomføre undersøkelser hvert fjerde år, det vil si i 2003, 2007, 2011 osv.

I stedet for å gjennomføre nye separate matematikk- og naturfagstudier i 1990-årene ble altså TIMSS etablert som et integrert realfagsprosjekt. Dermed kan TIMSS-studien bidra til at både sammenhenger og viktige forskjeller mellom matematikk og naturfagene i skolen blir grundig belyst.

2.1.2 *Organisering og viktige mål i TIMSS*

Det internasjonale prosjektsenteret ligger ved Boston College i USA. Andre sentre med delansvar for TIMSS-studien er IEA Data Processing Center i Hamburg og Statistics Canada i Ottawa. I hvert land er det et nasjonalt senter som har ansvar for all tilrettelegging og gjennomføring i eget land. Det norske prosjektsenteret er lagt til Institutt for lærerutdanning og skoleutvikling (ILS) ved Universitetet i Oslo (UiO). Det er naturlig nok et tett samarbeid mellom prosjektsenteret i Boston, delseentrene som er nevnt ovenfor, og de nasjonale prosjektgruppene i deltakerlandene.

Målene for TIMSS er kort beskrevet å

- undersøke elevenes kunnskaper i matematikk og naturfag
- studere hvordan kunnskaper henger sammen med faktorer som for eksempel holdninger, kjønn, hjemmebakgrunn, skolearbeid, fritidssysler og undervisningens innhold og organisering
- gjøre sammenlikninger mellom land
- studere utvikling over tid ved å sammenlikne nye resultater med resultater fra tidligere TIMSS-undersøkelser
- prøve å finne fram til faktorer, nasjonalt og internasjonalt, som fremmer god læring og en positiv utvikling innen realfagene i skolen

2.1.3 *Analysen av data*

Data analyseres på følgende tre "nivåer":

1. nivå: Systemet

Det første nivået gjelder systemet slik det legges til rette av myndighetene og samfunnet. Vi tenker da på det som står i fagplaner og lærebøker. Vi snakker her om den intenderte læreplanen ("intended curriculum"). Systemets rammefaktorer, skolesystemets struktur og elevenes muligheter for valg av skole og fag er også analysert. Informasjon på dette området blir hentet inn ved hjelp av spørreskjemaer som besvares av lærere og skoleledere, fortrinnsvis rektorer eller inspektører. Opplysninger hentes også inn fra den nasjonale prosjektgruppa.

2. nivå: Klasserommet

Neste nivå er klasserommet. Dette nivået handler om selve undervisningen og læringsmiljøet i klassen. Hvordan blir den intenderte læreplanen satt ut i livet? Vi snakker her om den implementerte læreplanen ("implemented curriculum"). Både elever og lærere svarer på spørreskjemaer om hva som skjer i klasserommet. Viktige bakgrunnsfaktorer man her vil ha fatt i, vil være elevenes hjemmebakgrunn, klassemiljøet, hva de gjør i fritiden, og hvor mye tid de bruker på hjemmearbeid. Likedan vil man for lærernes del vite noe om deres utdanning og holdninger, undervisningsmetoder, hvilke emner som blir undervist, og deres arbeidssituasjon.

3. nivå: Elevene

Det tredje og siste nivået er knyttet til det vi kaller den resulterte læreplanen ("attained curriculum"). Eksempler på viktige problemstillinger på dette nivået er:

- Hvordan varierer elevenes faglige prestasjoner mellom kjønn, klasser, skoler og land?
- Hvilke holdninger til realfagene har elevene utviklet?
- Hvordan varierer disse holdningene med alder og kjønn og mellom land?
- Kan noen av de andre variablene i undersøkelsen, for eksempel hjemmebakgrunn (primært familiens sosioøkonomiske status) "forklare" disse variasjonene?

For hvert enkelt land foretas det analyser av sammenhengen mellom disse tre nivåene av "læreplanen". Ut fra dette forsøker man så å trekke slutninger om i hvilken grad "systemets" og lærerens mål og intensjoner blir nådd. Landene sammenliknes også innbyrdes slik at forholdene kan studeres i et internasjonalt perspektiv.

2.1.4 *Populasjoner i TIMSS 1995, 1999 og 2003*

TIMSS 1995

TIMSS fokuserte i 1995 på tre ulike alderstrinn. Populasjonene var definert slik:

- **Populasjon 1:** De to klassetrinnene med flest 9-åringer
- **Populasjon 2:** De to klassetrinnene med flest 13-åringer
- **Populasjon 3:** Det siste året i videregående skole

Norge deltok i alle tre populasjonene. De elevene som var med i den norske delen av undersøkelsen, kom fra 2. og 3. klasse, fra 6. og 7. klasse og fra det tredje året i videregående skole, alle studieretninger. Klassebetegnelsen er her de som gjaldt før L97.

TIMSS 1999

Denne undersøkelsen ble som nevnt kalt "TIMSS-Repeat", og populasjonen var begrenset til det som var øverste klassetrinn i populasjon 2 i 1995. Norge deltok ikke i denne undersøkelsen.

TIMSS 2003

TIMSS 2003 rettet seg mot to klassetrinn. I vårt land (og de fleste andre) dreide det seg om 4. klasse og 8. klasse. Populasjonene var definert slik:

Populasjon 1: Det øverste av de to klassetrinnene med flest 9-åringer

Populasjon 2: Det øverste av de to klassetrinnene med flest 13-åringer

Imidlertid ble det godkjent at noen land, særlig enkelte utviklingsland, deltok med eldre elever enn definisjonen skulle tilsi. Det som er felles for alle deltakerlandene, er følgelig i større grad klassetrinnet enn elevenes alder. Derfor er populasjonene i de internasjonale rapportene betegnet etter klassetrinn. Vi har i denne boka brukt Populasjon 1/2 eller 4./8. klasse om hverandre som betegnelser for de to populasjonene.

Populasjon 1 svarer til 4. klasse i de fleste landene, også i vårt land. For Norges vedkommende kan vi dermed sammenlikne resultatene for jevngamle elever, testet i henholdsvis 1995 og 2003. De som ble testet i 2003, har imidlertid fulgt ny læreplan og gått ett år mer på skolen enn elever som deltok i 1995.

Populasjon 2 svarer til 8. klasse i de fleste landene, deriblant vårt land. Imidlertid er det viktig å huske at våre ”åttendeklassinger” egentlig gikk i sitt sjuende skoleår fordi de hoppet over ”2. klasse” da L97 ble innført. Vi har data fra elever i 7. klasse i 1995, og vi kan da gjøre sammenlikninger mellom elever som var like gamle, og som hadde gått like mange år på skole i 1995 og 2003. Elevene som deltok i 2003, hadde imidlertid fulgt den nye læreplanen, L97, siden 3. klasse. Selv om vi ennå ikke kan få en fullstendig vurdering av effektene av reformen, gir resultatene mange viktige indikasjoner på virkningene av reformen når det gjelder realfagene.

2.2 Rammeverk og instrumenter for TIMSS 2003

2.2.1 Hensikt og utforming

Utviklingen av rammeverket for TIMSS 2003 (Mullis mfl. 2003) bygger videre på Curriculum Frameworks for Mathematics and Science (Robitaille 1993), som dannet basis for undersøkelsene som ble gjennomført i 1995 og 1999. Revisjonen reflekterer endrede læreplaner og endret undervisningspraksis i deltakerlandene gjennom det siste tiåret. Man ønsker samtidig å opprettholde muligheten til å studere trender, og mange oppgaver fra 1995 og 1999 ble derfor gitt på ny i 2003. Det er lagt vekt på bred internasjonal deltakelse i utformingen av rammeverket. Et sammensatt panel av matematikk- og naturfageksperter fra flere land har bidratt til utformingen av de delene av rammeverket som er direkte fagrelaterte. Hensikten med dette var å forsikre seg om at mål som blir ansett for å være viktige i realfagundervisningen i deltakerlandene, ble vurdert og eventuelt inkludert i rammeverket. Dersom det skal tas hensyn til resultatene fra TIMSS-undersøkelsen i vurderingen av realfagenes situasjon i de nasjonale skolesystemene, er det viktig at elevene blir testet i lærestoff de ifølge læreplanen og rapportert undervisningspraksis i hovedsak skal kunne.

Det ble tatt hensyn til følgende faktorer ved den endelige utformingen av innholdskategoriene og utvelgelsen av oppgaver til testen (Mullis mfl. 2003):

- at innholdet var inkludert i læreplanen hos majoriteten av deltakerlandene
- at innholdet sannsynligvis kan regnes som viktig for matematikk- og naturfagundervisning i framtiden
- at oppgavene var godt tilpasset de deltakende elevers alderstrinn
- at oppgavene måtte fungere teknisk godt i en storskalaundersøkelse
- at hver oppgave medvirket til en balansert test med hensyn til innholdskategorier og kognitive nivåer

Det siste punktet er ikke minst viktig ettersom det er en uttalt målsetting at man fra og med 2003 i TIMSS gradvis vil legge mer vekt på spørsmål og oppgaver som gir bedre innsikt i elevers analytiske og utforskende evner. For å kunne nå dette målet vil problemløsningsoppgaver få en stadig større plass i testen.

Resultater fra TIMSS 2003 kan benyttes til ulike formål av forskjellige aktører. Politikere, undervisningspersonale og forskere vil få data som viser elevenes kunnskaper og ferdigheter i matematikk og naturfag i 2003. I tillegg vil data fra følgende områder bli lagt fram:

- trenddata fra 1995 via 1999 til 2003. (Ettersom Norge ikke deltok i 1999, vil det selvsagt ikke foreligge norske data fra denne undersøkelsen. Trendanalyser fra 1995 til 2003 vil vi imidlertid kunne foreta også her i Norge.)
- sammenliknbare data fra andre land med analyser som antyder mulige årsaker til forskjeller i skåre
- data som muliggjør analyse av likheter og forskjeller i undervisningen mellom deltakerlandene

2.2.2 *Rammeverket for matematikk*

Rammeverket for matematikk (Mullis mfl. 2003, s. 7–34) har to organiserende dimensjoner, en innholdsdimensjon og en kognitiv dimensjon. Hver dimensjon har flere kategorier. Innholdsdimensjonen definerer hvilke temaområder som dekkes av testen, og den kognitive dimensjonen definerer den typen kognitiv atferd som forventes av elevene i møtet med de aktuelle oppgavene. De ulike kategoriene er igjen delt inn i en rekke underkategorier, for eksempel er tall delt inn i hele tall, brøk og desimaltall, forhold, proporsjonalitet og prosent. Ved rapportering internasjonalt er det innholdsdimensjonen som brukes til å rapportere mer detaljerte resultater.

Som det framgår av tabellene 2.1 og 2.2, er det prosentvis satt av noe ulik testtid for de utvalgte innholdskategoriene og kognitive kategoriene i 4. og 8. klasse. Det er kun foretatt enkelte små justeringer i disse kategoriene i forhold til dem man benyttet i 1995 og 1999, slik at grunnlaget for å sammenlikne prestasjoner ikke er forrykket.

Tabell 2.1 *Fordeling av matematikkoppgaver etter innholdsdimensjon*

	4. klasse	8. klasse
Tall	40 %	30 %
Algebra/Mønstre*	15 %	25 %
Målinger	20 %	15 %
Geometri	15 %	15 %
Datarepresentasjon	10 %	15 %

* For 4.klasse benytter man betegnelsen Mønstre som mer dekkende for innholdet

Tabell 2.2 *Fordeling av matematikkoppgaver etter kognitiv dimensjon*

	4. klasse	8. klasse
Kjenne fakta og prosedyrer	20 %	15 %
Bruke begreper	20 %	20 %
Løse rutineproblemer	40 %	40 %
Gjennomføre resonnement	20 %	25 %

Generelt vil man kunne si at den kognitive kompleksiteten til en oppgave øker fra en kognitiv dimensjon til den neste, altså ovenfra og ned i tabellen. Det vil være en progresjon fra det å kjenne fakta og prosedyrer til å bruke begreper og videre til å benytte denne kunnskapen til å løse et problem. Likeledes vil det være mindre krevende å bruke matematisk kunnskap innenfor relativt kjente og ukompliserte situasjoner i forhold til å benytte disse kunnskapene til systematisk resonnering innenfor nye og til dels ukjente problemområder. Denne kognitive kompleksiteten må likevel ikke forveksles med oppgavens vanskelighetsgrad. Det er fullt mulig å lage både lette og vanskelige matematikkoppgaver innenfor alle de nevnte kognitive dimensjonene. Faktisk etterstreber man i TIMSS nettopp en slik variasjon i vanskelighetsgrad i forhold til både innholdsdimensjonene og de kognitive dimensjonene. Internasjonalt er det bestemt å gjøre nye beregninger og publisere en ny rapport der landene og elevgrupper innad i land sammenliknes også langs de kognitive dimensjonene.

At elever skal kunne kommunisere matematiske ideer og prosesser, er ansett som svært viktig i moderne matematikkundervisning. Begrunnelsen for dette er særlig at det å kunne representere, modellere og tolke informasjon framstår som helt nødvendig for å kunne orientere seg i dagens samfunn. I TIMSS ser man på denne evnen til kommunikasjon som et svært viktig resultat av matematikkundervisningen i skolen, men den er ikke inkludert som en egen kognitiv dimensjon. I stedet ser man på det som en type virksomhet som spenner over alle de matematiske innholdskategoriene og prosessene. Prinsipielt bør det derfor være mulig å måle elevenes evne til kommunikasjon gjennom åpne oppgaver knyttet til alle de foreliggende avgrensede dimensjonene.

2.2.3 *Rammeverket for naturfag*

I likhet med rammeverket for matematikk har rammeverket for naturfag to organiserende dimensjoner, en innholdsdimensjon og en kognitiv dimensjon. Begge har flere elementer som definerer henholdsvis de spesifikke naturfaglige fagområdene i undersøkelsen og den ulike kognitive atferden som forventes av

elevene i møte med oppgavene. Det er en noe ulik vektning av disse elementene i testene for 4. og 8. klasse, og dette er vist i tabellene 2.3 og 2.4.

Tabell 2.3 Fordeling av naturfagoppgaver etter innholdsdimensjon

	4. klasse	8. klasse
Biologi	45 %	30 %
Fysikk/Kjemi	35 %	*
Kjemi	*	15 %
Fysikk	*	25 %
Geofag**	20 %	15 %
Miljølære	*	15 %

* Kjemi og fysikk er slått sammen til én kategori i 4. klasse. Miljølære er ikke egen kategori for dette klasse-trinnet.

** Med Geofag menes det som på engelsk kalles "earth science", emner innen astronomi, naturgeografi, geologi og geofysikk.

Vi ser av tabell 2.3 at det vil være fem rapporteringskategorier for 8. klasse og tre for 4. klasse. Som for matematikk vil resultatene her bli rapportert separat for hvert element i innholdsdimensjonen og for faget totalt.

Tabell 2.4 Fordeling av naturfagoppgaver etter kognitiv dimensjon

	4. klasse	8. klasse
Faktakunnskap	40 %	30 %
Begrepsforståelse	35 %	35 %
Resonnement og analyse	25 %	35 %

Som vi ser av tabell 2.4, er det tre kategorier for den kognitive dimensjonen for begge klassetrinnene med noe ulik vektning. Internasjonalt blir det imidlertid ikke lagt opp til rapportering etter denne dimensjonen.

I tillegg til innholdsdimensjonen og den kognitive dimensjonen vil man ha ytterligere en kategori, nemlig naturvitenskapelig arbeidsmetode. Naturvitenskapelig arbeidsmetode på dette nivået innebærer at elevene forventes å vise at de kan

- formulere spørsmål og hypoteser
- planlegge undersøkelser
- samle inn og representere data
- analysere og tolke data
- trekke konklusjoner og skrive forklaringer

Selv om disse fasene i naturvitenskapelig arbeidsmetode er passende for elever i både 4. og 8. klasse, vil kravene til kompleksitet øke med høyere klassetrinn i takt med elevenes kognitive utvikling.

Det er særlig gjennom oppgaver som krever at elevene benytter sine kunnskaper og sin forståelse i en praktisk kontekst, at man vil kunne måle deres evne til å benytte naturvitenskapelig arbeidsmetode. De aktuelle oppgavene er derfor konstruert på en måte som skal gjøre det mulig å vurdere i hvilken grad

elevene har den nødvendige forståelsen og innsikten som skal til for å bruke naturvitenskapelig arbeidsmetode

2.2.4 Kontekstuekt rammeverk

For å få en dypere forståelse av hva resultatene fra TIMSS 2003 innebærer, og hvordan de kan brukes for å forbedre undervisning og læring i matematikk og naturfag, er det viktig å forstå elevenes læringskontekst. Parallelt med den faglige testen blir det derfor samlet inn opplysninger om blant annet

- læreplaner
- skoler
- læreres utdanning og deres undervisning
- klasseromsaktiviteter og karakteristiske trekk ved disse
- elevers hjemmebakgrunn og holdninger

Det blir lagt vekt på å undersøke hvordan man gjennom de nasjonale undervisningssystemene søker å nå de læreplanmålene man har satt seg. I likhet med hvordan rammeverket for matematikk og naturfag definerer hvilke områder som elevene blir testet i, vil det kontekstuelle rammeverket avgrense hvilke forhold elevene blir spurt om når det gjelder undervisning og læring.

2.2.5 TIMSS og PISA

TIMSS kan karakteriseres som en læreplanbasert undersøkelse. Et av de viktigste kriteriene for utvelgelse av oppgaver er at de er relevante i forhold til hva som undervises i majoriteten av deltakerlandene. Man ønsker altså å måle det som kan betegnes som ”skolekunnskap”. I PISA, en komparativ internasjonal undersøkelse innenfor utdanning i regi av OECD, har man valgt å bruke et annet utgangspunkt (Lie mfl. 2001, Kjærnsli mfl. 2004). Her har man utviklet et rammeverk ut fra en konsensus blant deltakerlandene på politisk nivå om hvilke kunnskaper og ferdigheter som det vil være nødvendig å inneha for å kunne delta som fullverdige borgere i framtidens samfunn. Det fokuseres først og fremst på de nyttige sidene ved fagene, og man søker å måle elevenes evne til aktivt å bruke sine kunnskaper og erfaringer. OECD ønsker gjennom PISA-undersøkelsen å få et svar på hvor godt skolen forbereder elevene på å møte de utfordringene de trolig vil møte i framtiden. Selv om man slik ser at utgangspunktet for de to undersøkelsene er forskjellig, vil det overordnede målet i både PISA og TIMSS være å framskaffe veldokumenterte forskningsresultater som bidrar til at deltakerlandene skal kunne ha bedre grunnlag for å ta avgjørelser om skolepolitiske spørsmål. For ytterligere informasjon om mål og perspektiver i PISA, se Kjærnsli mfl. 2004.

2.2.6 TIMSS og L97

Opgavene i TIMSS er laget med utgangspunkt i deltakerlandenes læreplaner og er basert på en betydelig grad av enighet om hva som anses som en viktig

del av skolens matematikk og naturfag. I gjennomgangen av rammeverket for TIMSS 2003 har vi sett at undersøkelsen dekker over et vidt spekter av faglige emner og typer av kognitive ferdigheter innenfor matematikk og naturfag. For å kunne si noe om hvorvidt TIMSS kan sies å representere en relevant evaluering av norske elever i 2003 og 1995, vil vi måtte undersøke i hvilken grad den i rimelig grad stemmer overens med henholdsvis M87 og L97 når det gjelder faglige mål og prioriteringer. I kapitlene 3 og 5 vil vi komme nærmere inn på dette. En internasjonal undersøkelse må nødvendigvis strebe etter å dekke mer enn det som er felles for alle lands læreplaner, og følgelig vil elever i alle land møte utfordringer de ikke har møtt i undervisningen tidligere. Målet er at utvalget av emner og oppgavetyper framstår som ”like urettferdig for alle land” (ifølge et vanlig, men uoffisielt slagord). I den forstand er det vanskelig å se at TIMSS-testene i matematikk og naturfag på noen spesiell måte har gitt fordeler eller ulemper for norske elever.

Rammeverket for TIMSS 2003 er en videreføring av de rammeverkene man utarbeidet for TIMSS 1995 og 1999. Den utviklingen som har skjedd, gjenspeiler endringer i fagdidaktiske tanker og teorier i samme tidsrom og av den grunn noe avvikende prioriteringer av hva som bør testes. En del såkalte ”trendoppgaver” er blitt beholdt, mens en del nye oppgaver er kommet til. Både læreplanene våre og TIMSS-undersøkelsen har vært påvirket av de til enhver tid rådende fagdidaktiske syn og den utviklingen som der finner sted. Vår konklusjon er at man med stor tyngde kan hevde at de faglige emnene og perspektivene i TIMSS 2003 representerte et relevant utgangspunkt for å måle norske elevers faglige nivå i 2003 og endringen i nivå siden 1995.

2.2.7 Oppgaver

Dersom elever skulle bli testet i samtlige emner innenfor matematikk og naturfag, ville dette medført minst en sju timers prøve på 8. klassetrinn og mer enn fem og en halv times testing på 4. klassetrinn. En slik testtid er selvsagt helt uakseptabel. Oppgavene blir derfor fordelt på et visst antall hefter etter et nøye utarbeidet mønster, og elevene får så ett hefte hver. Det er i alt 12 hefter for hver av de to populasjonene, og disse roteres innenfor alle deltakende klasser. I en og samme klasse kan altså maksimalt tre elever få det samme oppgaveheftet.

Tradisjonelt har det i store undersøkelser vært brukt flervalgsoppgaver hvor elevene skal krysse av for riktig svar. Flervalgsoppgavers store styrke er at de generelt øker reliabiliteten til en undersøkelse, rett og slett ved at det tar mindre tid å svare på dem, og at man derfor kan inkludere flere oppgaver i testen. Dessuten kan svarene rettes objektivt, noe som gir høy sensorreliabilitet. I tillegg blir rettingen mindre tidkrevende, testen kan lett standardiseres, og administrasjonskostnadene blir lavere. For en mer utførlig drøfting av egenskaper ved flervalgsoppgaver, se Olsen mfl. (2001).

Tidligere har vi beskrevet hvordan man i TIMSS også ønsker å måle elevenes evne til å kommunisere matematiske ideer og prosesser og deres innsikt i og bruk av naturvitenskapelig arbeidsmetode. Til dette formålet vil åpne opp-

gaver egne seg langt bedre enn flervalgsoppgaver. Allerede i 1995 bestemte man seg i TIMSS for å inkludere åpne oppgaver i testingen av elever. Dette ble deretter fulgt opp i både 1999 og i 2003. Fordelingen mellom de to oppgavetyperne i 2003 var omtrent 60 prosent flervalgsoppgaver og 40 prosent åpne oppgaver i begge populasjonene.

Prosessen med å velge ut de matematikk- og naturfagoppgavene som ble med i hovedundersøkelsen, var svært omfattende. En del av oppgavene er såkalte trendoppgaver, det vil si oppgaver som også ble benyttet i 1995 og/eller i 1999. Hensikten med å bruke disse er at man ønsker å kunne måle hvorvidt det over tid er skjedd endringer i elevenes kunnskaper og ferdigheter i realfag (se mer om dette i 2.3.4).

De nye oppgavene ble plukket ut etter å ha vært igjennom en grundig utprøving på flere hold. For det første er de anbefalt som gode oppgaver av de faglige ekspertkomiteene. Dernest ble de benyttet i en omfattende generalprøve i alle deltakerlandene året før hovedundersøkelsen, der man blant annet undersøker om de fungerer godt ut fra psykometriske kriterier. Oppgavene ble i tillegg gjenstand for kritisk gjennomgang av de nasjonale komiteene i deltakerlandene, som vurderte hvorvidt de passer rimelig bra for deres elever ut fra landets nasjonale læreplan. Som vi tidligere har sett, skulle også oppgavene til sammen dekke forskjellige faglige emner, det skulle være en riktig fordeling mellom flervalgsoppgaver og åpne oppgaver, og det måtte også være en god balanse mellom de ulike oppgavenes vanskelighetsgrad. Den prøven som til slutt ble brukt i hovedundersøkelsen, holdt derfor høy faglig standard, og den representerer på en god måte rammeverket for prøven.

2.2.8 Koder for åpne oppgaver

Som vi har nevnt, besto testene i TIMSS 2003 av både flervalgsoppgaver og åpne oppgaver. Det er en stor utfordring å oppnå en høy reliabilitet, både nasjonalt og internasjonalt, ved retting av de åpne oppgavene. I TIMSS har man gjort mye for å løse dette ved å avholde internasjonale retteseminarer hvor alle deltakerlandene pliktet å møte. Her ble så alle åpne oppgaver nøye drøftet. Flere besvarelser på hver oppgave ble vurdert av deltakerne, og man oppnådde en konsensus om hvordan man skulle kode ulike typer svar. Denne konsensusen ble så grunnlaget for trening av sensorer på nasjonalt nivå. Målet var å oppnå en høy sensorreliabilitet både nasjonalt og internasjonalt på de åpne oppgavene.

Ett av målene med TIMSS er å få innsikt i elevers kunnskap og forståelse av viktige begreper innen sentrale områder i matematikk og naturfag. For å få innsikt i hvordan elevene har tenkt, hvilke løsningsstrategier de har brukt, og eventuelle alternative forestillinger, er det viktig at besvarelsene ikke bare blir kodet for riktig eller galt. Da ville mye av den diagnostiske informasjonen gå tapt. Med bakgrunn i dette er det derfor blitt utviklet et spesielt kodesystem som man anvender i rettingen av de åpne oppgavene. Den norske prosjektgruppen har vært sentral i utviklingen av dette kodesystemet, som første gang

ble benyttet i TIMSS 1995 (Kobberstad mfl. 1994, Lie mfl. 1996, Angell 1996). Kodesystemet er tosifret, og det første sifferet indikerer om svaret er riktig, delvis riktig eller galt. Det andre sifferet refererer til typer av svar, for eksempel hvilke tanker elevene har om et begrep, eller hvilken metode de har brukt i løsningen. Koder mellom 20 og 29 vil si at det er gitt to poeng for svaret, mens koder mellom 10 og 19 betyr at svaret er honorert med ett poeng. Koder mellom 70 og 79 viser at svaret er feil og følgelig gir null poeng. Dermed eleven overhodet ikke har forsøkt å svare på oppgaven, gis koden 99. Basert på dette generelle systemet ble det utviklet en detaljert rettemal for hver eneste åpen oppgave. Hver kode var angitt med detaljert beskrivelse av kjennetegn ved slike svar og med autentiske eksempler på elevsvar.

Basert på norske og internasjonale elevsvar på oppgavene i TIMSS 1995 har den norske prosjektgruppa publisert mange fagdidaktiske analyser der de tosifrete kodene er trukket inn. I to bøker, en for hvert av fagene, presenterte vi alle de frigitte TIMSS-oppgavene med svarfordeling og fagdidaktiske kommentarer (Brekke mfl. 1998, Kjærnsli mfl. 1999). For øvrig vil detaljerte data om elevenes svar på enkeltoppgaver danne grunnlag for flere oppfølgende studier framover, slik det også ble gjort for data fra TIMSS 1995 (for eksempel Angell 1996, Kjærnsli mfl. 2002) og for flere hovedoppgaver i realfagdidaktikk ved ILS).

2.2.9 Spørreskjemaer

Det ble lagt ned mye arbeid i å utvikle spørreskjemaer til elever, lærere og skoleledere med spørsmål som gjør det mulig å analysere og belyse interessante sammenhenger mellom prestasjoner og ulike bakgrunnsvariabler. I likhet med oppgavene er spørreskjemaene også et resultat av en lang prosess med mange revisjoner. Alle deltakerlandene har hatt mulighet til å vurdere, revidere og reorganisere spørsmålene i disse skjemaene flere ganger underveis i utviklingen. Det var også mulig med bakgrunn i særegne kulturelle forhold å utelate, eventuelt modifisere enkelte spørsmål, men dette måtte da begrunnes spesielt og godkjennes av TIMSS sentralt i hvert enkelt tilfelle. For å kunne foreta sammenlikninger internasjonalt er det selvsagt nødvendig at deltakerlandene benytter identiske spørsmål.

Et spesielt problem gjelder i naturfag, siden dette i noen land, slik som i Norge, representerer ett integrert skolefag på de aktuelle klassetrinnene. I andre land, derimot, undervises det i fysikk, kjemi, biologi og eventuelt "earth science" som separate fag. Følgelig måtte spørreskjemaene utformes i forskjellige varianter når det gjaldt holdninger til og undervisning i "science". Derfor kan norske data om dette bare sammenliknes med noen av deltakerlandene.

Elevspørreskjema

Dette skjemaet inneholdt en rekke spørsmål for å kartlegge ulike bakgrunnsvariabler. Heftet som ble brukt i populasjon 2, var noe mer omfattende enn det for populasjon 1. Elevene ble blant annet spurt om holdninger til faget, foreldres utdanning, antall bøker i hjemmet, hva som gjøres i matematikk- og natur-

og miljøfagtimene, bruk av datamaskin hjemme og på skolen, det sosiale miljøet på skolen og fritidsaktiviteter. Hensikten er så å analysere disse opplysningene på ulike måter, blant annet for å kunne avdekke interessante sammenhenger med faglige prestasjoner.

Lærerspørreskjema

Mens man i populasjon 1 kun benyttet ett lærerspørreskjema, hadde man for populasjon 2 et skjema for lærere i matematikk og et annet for lærere i natur- og miljøfag. Det var den læreren (eventuelt en av lærerne) som underviste TIMSS-klassen i de aktuelle fagene, som skulle besvare lærerspørreskjemaet. En del av spørsmålene i alle disse skjemaene var like, for eksempel de som var knyttet til lærerens egen utdanning, antall elever i klassen, samarbeid med andre lærere, undervisningsmetoder og arbeidsmåter, etter- og videreutdanning, skolens befatning osv. Spørsmål som gikk på det faglige innholdet i timene, var naturlig nok forskjellige, alt etter hvilket fag eller trinn det handlet om. Hensikten med disse spørsmålene er todelt: både å gi deskriptive data om realfaglærere og realfagundervisning og å avdekke mulige sammenhenger mellom disse dataene og elevprestasjoner.

Skolespørreskjema

Dette skjemaet skulle besvares av rektor eller en annen representant for administrasjonen. Her ble det spurt blant annet om antall elever på skolen, lærernes, elevenes og foreldrenes holdninger og engasjement, rektors arbeidsområde, skolens tilbud om etterutdanning til lærere, eventuell mangel på ulike typer undervisningsmateriale, PC-tilgjengelighet og hyppighet av problematferd blant elevene.

2.2.10 Oversettelse

Alle de internasjonale instrumentene var i utgangspunktet utviklet på engelsk og måtte så oversettes til forskjellige språk. Allerede ved utviklingen og utvelgelsen av oppgaver ble det lagt vekt på at det ikke forekom ord og uttrykk som favoriserer eller disfavoriserer elever ut fra nasjonale, kulturelle eller sosiale forhold eller ut fra landets geografiske beliggenhet. Fra tidligere studier har man likevel erfart at språklige forskjeller og kulturelle tilpasninger har vært et stort problem. Samme tekst kan oversettes på mange ulike måter. Alle kan språklig sett være "riktige", men oppgavens vanskegrad kan variere ut fra hvilken oversettelse man velger. For å imøtekomme så mye som mulig av den kritikken som har vært reist mot komparative studier nettopp når det gjelder oversettelse, ble det i TIMSS satt svært strenge krav til oversettelsesprosedyrene.

De nasjonale sentrene i hvert deltakerland var ansvarlig for å oversette og ferdigstille instrumentene. Kravene til dette arbeidet har som overordnet utgangspunkt at for at man skal kunne gjøre valide sammenlikninger, må instrumentene være så like som mulig i alle deltakerland. Vanskegraden, lesbarheten, tabeller, layout osv., alt måtte gjøres så likt som mulig. Alle oppgaver og spørsmål måtte oversettes uavhengig av minst to personer med god språklig og

faglig kompetanse og med god kjennskap til de aktuelle elevgruppene. En tredje person måtte være med i diskusjonen om den endelige versjonen.

Etter generalprøven i 2002 ble det utarbeidet en oversikt over de få oppgavene som fungerte litt ”unormalt”, for eksempel ved at påfallende få elever svarte riktig i et land. Slike oppgaver måtte man gå nøye igjennom på nytt for eventuell forbedring av oversettelsen.

Her i Norge ble instrumentene oversatt til både bokmål og nynorsk. Vi knyttet til oss språklige konsulenter som bisto oss i deler av dette arbeidet. De oversatte instrumentene ble så sendt til det internasjonale senteret, som ved hjelp av sine språkekspertene ga oss verdifulle tilbakemeldinger. Revisjoner av den opprinnelige teksten ble så gjennomført. Det falt heldig ut at vi kunne forholde oss til to forskjellige språkekspertene, en for bokmål og en for nynorsk, ettersom vi da fikk ulike språklige innspill. Dette kunne vi dra nytte av i den endelige versjonen. Vi la for øvrig stor vekt på å gjøre bokmåls- og nynorsk-utgaven av instrumentene så like som overhodet mulig når det gjaldt bruk av fagterminologi og krav til leseferdighet.

2.3 Gjennomføring av undersøkelsen i Norge

I det følgende gir vi en kort beskrivelse av hvordan undersøkelsen ble gjennomført i praksis. En mer detaljert beskrivelse av dette finnes i en egen rapport (TIMSS 2003).

2.3.1 Tidspunkt for gjennomføring

Hovedundersøkelsen i TIMSS 2003 ble internasjonalt gjennomført på to forskjellige tidspunkter. Land på den sørlige halvkula gjennomførte undersøkelsen i august–september 2002, mens land på den nordlige halvkula gjennomførte testen i mars–april 2003. De ulike tidspunktene skyldes hovedsakelig ulik organisering av skoleåret. I vårt land ble undersøkelsen gjennomført på skolene i løpet av april måned. Skolene valgte selv dag for gjennomføringen.

2.3.2 Utvalg av skoler og elever

Det var strenge internasjonale kriterier for hvilke elever som kunne ekskluderes fra utvalget på klassenivå. I hovedsak var dette elever som har en psykisk eller fysisk funksjonshemming som gjør dem ute av stand til å svare på oppgavene, eller fremmedspråklige elever som ikke kan tilstrekkelig norsk til å forstå teksten i heftene. Her i landet ble ca. 1,7 % av elevene ekskludert av slike årsaker, omtrent den samme prosentdelen som i de fleste andre land. Dette betyr at selv om vi har mange integrerte elever i vårt land, er de samme ekskluderingskriteriene benyttet her som i alle de andre deltakerlandene. Den største forskjellen er trolig at mens vi ekskluderer enkelte elever, er det i en del andre land hele spesialklasser eller spesialskoler som blir ekskludert. Resultatet er uansett at utvalget av elever er sammenliknbart i alle deltakerlandene.

Et utvalg på 150 skoler ble trukket for hvert klassetrinn for seg ut fra en liste over alle skoler med elever på de aktuelle trinnene. Videre ble én klasse (på små skoler) eller to klasser (på større skoler) trukket til å være med fra hver av de uttrukne skolene. Tabell 2.5 viser antallet av skoler, klasser og elever som faktisk deltok i undersøkelsen. Frafallet av skoler og av enkeltelever lå henholdsvis på litt under og litt over 10 prosent, noe som holdt seg godt innenfor det som var akseptabelt internasjonalt.

Uten å gå for mye i detalj vil vi her konstatere at utvalget av respondenter er gjort slik at de er påvist å representere hele populasjonen av skoler, klasser og elever på en god måte. Og siden utvelgingen er foretatt på en probabilitistisk (tilfeldig) måte, kan vi bruke statistiske metoder til å estimere hvor nøyaktig vi kan slutte fra resultater i utvalgene til hele populasjonen på hvert klassetrinn. I denne boka vil vi ofte angi 95 % konfidensintervaller eller feilmarginer for gjennomsnittsverdier for hele populasjonen.

Tabell 2.5 *Antall skoler, klasser og elever som deltok i TIMSS 2003*

Populasjon	Antall skoler	Antall klasser	Antall elever
Populasjon 1	139	228	4342
Populasjon 2	138	179	4133

Under prosessen med å trekke skoler ble skolene delt inn i ulike strata for at man skulle kunne gjøre et mer effektivt utvalg av skoler. Blant annet ble det trukket ut forholdsvis få små skoler for å unngå for mye administrasjon i forhold til tilgjengelige data. Men det gjør at små skoler er underrepresentert i vårt utvalg, og dette må kompenseres ved at elever på slike skoler tillegges en høyere vekt. Ved beregning av gjennomsnitt for land eller for grupper av elever er elever og skoler gitt ulik vekt i henhold til hvor stor sannsynlighet de hadde for å bli trukket ut. På den måten kan vi sørge for å gi gode estimater for slike gjennomsnittsverdier.

2.3.3 Gjennomføring på skolene

På hver skole var det en skolekoordinator som hadde ansvaret for gjennomføringen av undersøkelsen på sin skole. Uttrukne lærere i de respektive fagene, naturfag og matematikk, ble bedt om å fylle ut et lærerspørreskjema, og rektor eller inspektør fikk i oppgave å besvare et skolepørreskjema. Koordinatoren hadde også ansvaret for at undersøkelsen ble gjennomført nøyaktig etter reglene i hver av klassene, eventuelt ved hjelp av en egen testadministrator. I tillegg til et oppgavehefte fikk hver elev et elevspørreskjema de skulle besvare.

Det var 12 forskjellige oppgavehefter som alle inneholdt både matematikk- og naturfagoppgaver. Disse ble fordelt på elevene etter et på forhånd tilfeldig bestemt mønster. Alle heftene hadde to bolker, og begge var omtrent like arbeidskrevende. Elevene i populasjon 1 fikk nøyaktig 72 minutter til å besvare oppgavene i sitt hefte, 36 minutter til hver bolke, med en innlagt pause på ca. 10 minutter midtveis. Samme opplegg fulgte elevene i populasjon 2, men her

hadde elevene anledning til å bruke 2 x 45 minutter på oppgaveheftet. Etter at de var ferdig med de faglige oppgavene, fikk de en ny pause, og deretter svarte de på spørsmålene i elevspørreskjemaet.

Personer fra den nasjonale TIMSS-gruppa besøkte noen skoler som ledd i vår kontroll av prosedyrene. I tillegg fikk noen skoler uanmeldt besøk av en internasjonal kontrollør (se avsnitt 2.3.5).

2.3.4 Databehandling

Alle oppgavehefter og spørreskjemaer ble etter gjennomføringen på skolene sendt til ILS. Etter at alle dataene var tastet inn og rensset nasjonalt ved hjelp av programmer laget spesielt for TIMSS, ble de sendt til IEA Data Processing Center i Hamburg. Der ble alle lands data nøye gjennomgått. Alle avvik fra de internasjonale instrumentene ble identifisert, feil i dataene ble rettet opp så langt det var mulig, og detaljert dokumentasjon på kvaliteten av dataene ble laget. Alle land fikk under denne prosessen gjentatte muligheter til å kontrollere sine data. Hvert land fikk også statistikk for hver oppgave og hvert spørsmål for å undersøke om det av ulike grunner kunne være noe som ikke hadde fungert som forventet. For eksempel kan en påfallende lav andel riktige svar for flinke elever på en oppgave være en indikasjon på trykkfeil, at svaralternativene er listet opp i feil rekkefølge eller liknende. Det er svært viktig at slike typer feil blir oppdaget og korrigert. I verste fall kan det være at en oppgave har fungert så dårlig at den ikke kan telle med ved beregning av elevenes faglige skåre.

Med utgangspunkt i elevenes svar på oppgavene ble det regnet ut to samlede skårer for hver elev, en i matematikk og en i naturfag. Som vi tidligere har nevnt, har ikke alle elevene fått de samme oppgavene, det var i alt 12 ulike oppgavehefter. Blant annet for å kunne sammenlikne elevens skåre, selv om de har besvart ulike oppgavehefter, benyttes moderne og anerkjente modeller (såkalt ”Tre-parameter Item Response Theory”, se for eksempel Crocker & Algina 1986 for en beskrivelse). I tillegg til en overordnet skåre i hvert av de to fagene ble det også regnet ut delskårer innenfor hvert av de faglige emnene som er vist i tabellene 2.1 og 2.3.

Noen viktige sammenlikninger mellom land er presentert i de første internasjonale rapportene (Mullis mfl. 2004, Martin mfl. 2004). Noen slike resultater har vi kunnet hente direkte derfra, og videre har vi ut fra data om faglig skåre og svar på spørreskjemaer kunnet gjøre egne analyser både på nasjonalt og internasjonalt nivå.

2.3.5 Kvalitetskontroll

I en så ambisiøs studie som TIMSS er det svært viktig at dataene er av høy kvalitet, og at de er sammenliknbare deltakerlandene imellom. En av måtene å sikre dette på er at alle trinn i undersøkelsen kvalitetssikres. Det ble i TIMSS utarbeidet standardiserte og detaljerte manualer for dette. Disse manualene danner i ettertid en klar og eksplisitt dokumentasjon av alle de operasjonelle

prosedyrene. Som tidligere nevnt var det også knyttet streng kontroll til oversettelsesprosedyrene og rettingen av åpne oppgaver.

Også selve gjennomføringen av undersøkelsen på de enkelte skolene ble kvalitetskontrollert. Et bestemt antall skoler i hvert land fikk uanmeldt besøk av en observatør for å se at alle retningslinjer ble overholdt. I tillegg til dette ble det i hvert land utnevnt en uavhengig kontrollør som kontrollerte prosjektledelsens arbeid. Svært mange forholdsregler er således tatt for å sikre at dataene i TIMSS-studien er sammenliknbare på tvers av land.

2.4 Hovedfunn fra TIMSS 1995

Som beskrevet tidligere deltok vårt land i alle tre populasjonene i TIMSS 1995, og dette innebar to klassetrinn i både populasjon 1 og populasjon 2. I TIMSS 2003 er det bare elever i det øverste av disse to klassetrinnene som er med, resultater fra 2003 vil derfor bli sammenliknet med elever fra det øverste trinnet i 1995. Vi har nedenfor oppsummert de viktigste funnene fra TIMSS 1995 for disse elevene. For å kunne diskutere realfagene i norsk skole i et større perspektiv har vi også tatt med noen grunnleggende funn fra de andre klassetrinnene som deltok i 1995.

I øverste klasse i populasjon 2 i TIMSS 1995, som vi vil sammenlikne med 8. klasse i 2003, kan hovedfunnene oppsummeres slik (Lie mfl. 1997a):

- Elevene i Norge skåret signifikant lavere enn gjennomsnittet av land i matematikk og omtrent som gjennomsnittet i naturfag. I begge fag var Sverige klart det beste nordiske landet. (Finland deltok ikke.)
- Norske elever gjorde det forholdsvis bedre på oppgaver som handler om å bruke matematikk i praktiske sammenhenger enn på oppsatte regnestykker.
- I naturfag presterte norske elever spesielt dårlig i kjemi og bedre enn gjennomsnittet i geofag, biologi og miljøspørsmål.
- Det var ikke signifikant forskjell i vårt land mellom kjønnene når det gjaldt prestasjoner i matematikk, men i naturfag skåret guttene signifikant høyere.
- Det var stor forskjell i guttenes favør både når det gjaldt positiv holdning til og selvtillit i fagene.
- Norge var blant de land som hadde størst framgang mellom de to klassetrinnene som var med i populasjon 2.
- Det så ikke ut til at det spilte noen vesentlig rolle om oppgavene ble gitt som flervalgsoppgaver eller som åpne oppgaver som elevene selv skal formulere svaret på. Oppgaveformatet så heller ikke ut til å favorisere det ene kjønn på bekostning av det andre.

I øverste klasse i populasjon 1 i TIMSS 1995, som vi vil sammenlikne med 4. klasse i 2003, kan hovedfunnene oppsummeres slik (Lie mfl. 1997b):

- Norske elever gjorde det svært dårlig sammenliknet med elever på tilsvarende alderstrinn i de fleste andre landene som deltok, men noe