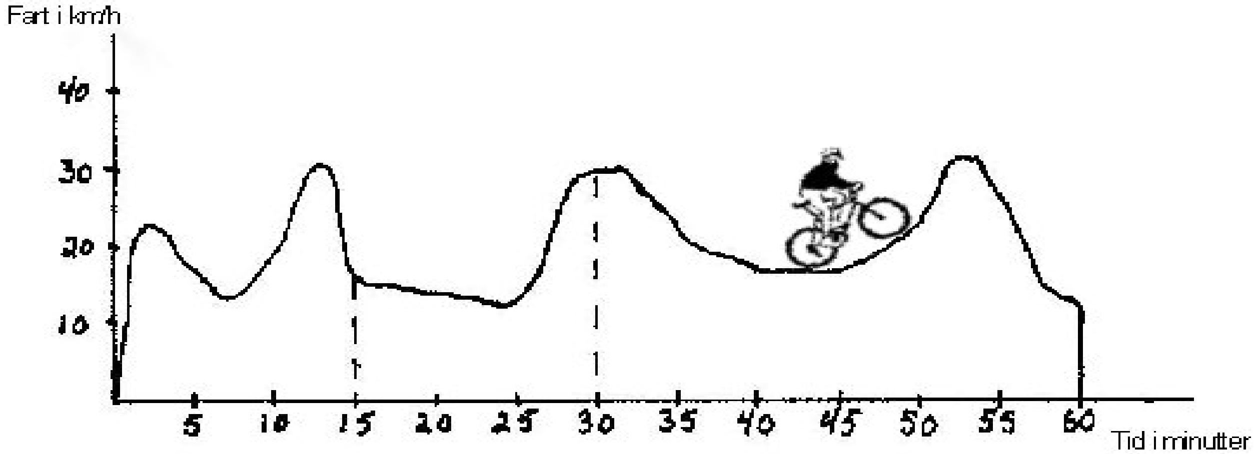
# Utvikling av matematikkoppgaver i PISA

*Hovedoppgave i realfagdidaktikk*
*av*
*Ole Kr. Bergem*




**Institutt for lærerutdanning og skoleutvikling**

**Universitetet i Oslo**

**Oktober 2002**

*Udfordringen er at søge en matematikundervisning der ikke blot udgjør en indføring i den matematiske tankeverden, men som samtidig danner grundlag for refleksioner over den rolle matematikken spiller og kan spille som element i en teknologisk fantasi (Skovsmose 2002, s.26).*

# Forord

Da jeg bestemte meg for å ta hovedfag i realfagsdidaktikk, var jeg heldig å bli invitert til å være med i PISA prosjektet. Mitt bidrag skulle være å delta i arbeidet med å utvikle matematikkoppgaver beregnet på 15 åringer. Disse kunne jeg så bruke som utgangspunkt for min hovedoppgave. I løpet av våren 2001 materialiserte dette arbeidet seg i form av et hefte bestående av 7 temaoppgaver, med underliggende delspørsmål, totalt 27.

I mai 2001 gjennomførte jeg en pilotundersøkelse ved fire skoler i østlandsområdet. 152 elever fra 7 klasser deltok, og svarene som framkom har blitt kodet og systematisert. Oppgavene ble så, sammen med deler av det statistiske materialet, sendt inn til PISA sentralt.

Våren 2002 ble det i PISA's regi gjennomført en generalprøve i deltakerlandene, for blant annet å avgjøre hvilke oppgaver som skal benyttes i den endelige undersøkelsen i 2003. I denne generalprøven er noen av mine oppgaver, med relativt moderate endringer, kommet med. For meg personlig har dette vært en betydelig inspirasjonskilde og i skrivende stund synes jeg fremdeles det er spennende å se om jeg får med noen av "mine" oppgaver i 2003.

Underveis i den prosessen jeg har vært gjennom, har jeg mottatt så vel konstruktiv kritikk, som verdifulle råd, knyttet til utformingen av oppgavene og gjennomføringen av pilotundersøkelsen, fra mine veiledere Liv Sissel Grønmo og Svein Lie, samt fra øvrige deltakere i prosjektgruppa for PISA, i særdeleshet Rolf Olsen og Are Turmo. Jeg vil derfor benytte anledningen til å utrykke min takknemlighet for denne støtten og bistanden. I tillegg vil jeg rette en takk til min mor, som med stor tålmodighet har lest korrektur.

# INNHOLDSFORTEGNELSE

# 1. Innledning

Min hovedoppgave vil være bygd opp rundt de oppgavene jeg har laget til PISA prosjektet. Etter å ha presentert og begrunnet mine forskningsspørsmål, vil jeg i kapittel 2 foreta en oppsummering av de store internasjonale, komparative undersøkelsene knyttet til matematikkfaget, som tidligere har blitt gjennomført. Jeg vil også redegjøre for noe av kritikken som fra ulike hold er blitt reist mot dem.

I kapittel 3 beskriver jeg kort PISA, og de helt sentrale tankene bak denne undersøkelsen. I kapittel 4 forsøker jeg først å vise hvordan matematikkfaget har blitt legitimert opp gjennom historien. Dette som en bakgrunn for en kritisk drøfting av konstruktet "*mathematical literacy*", som er svært sentralt i PISA. Som vi straks skal se får dette konstruktet også stor betydning for forskningsspørsmålene og dermed innholdet i min hovedoppgave. I dette kapitlet vil jeg videre knytte forbindelser mellom "*mathematic literacy*" og "*literacy*"-begrepet innenfor filologi og pedagogikk. Jeg vil også presentere utvalgte matematikk - didaktikeres syn på hva som bør stå i fokus for framtidens oppgaveutviklere, og deres begrunnelser for dette.

I kapittel 5 beskriver jeg organiseringen av rammeverket for matematikk i PISA, og utdyper noen helt sentrale kategorier. I kapittel 6 refererer jeg de viktigste kravene til oppgaver som skal benyttes som måleinstrumenter i denne undersøkelsen, slik disse framkommer i rammeverket i PISA. Jeg redegjør dessuten for prosedyrene for utvelgelse av oppgaver til PISA 2003. Kapittel 7 er tilegnet metodeteori, og jeg går her gjennom de testteoretiske begreper, som jeg senere i min hovedoppgave benytter. Jeg vil her særlig legge vekt på å drøfte validitetsbegrepet relativt inngående.

I kapittel 8, som er den empiriske delen av min hovedoppgave, presenterer og drøfter jeg de matematikkoppgavene jeg har utviklet, og resultatene fra pilotundersøkelsen jeg gjennomførte. Dette relateres til de tidligere refererte oppgavekravene i PISA, og til "*mathematical literacy*". Her vil jeg også legge fram norske data fra generalprøven i PISA, avholdt i april 2002. Disse dataene, som kun gjelder mine oppgaver, forelå på et såpass sent tidspunkt i forhold til framdriften i mitt arbeid, at drøftingen av dem ikke har kunnet bli særlig omfattende, men jeg har altså likevel valgt å ta dem med.

Kapittel 9 er viet en sammenfattende vurdering av "testens"[1] reliabilitet og oppgavenes validitet. I kapittel 10 trekker jeg de endelige konklusjoner av mitt arbeid, og forsøker å dokumentere hvorledes jeg mener å ha besvart mine forskningsspørsmål.

## 1.1 Problemstillinger

Under utarbeidelsen av oppgavene brukte jeg en god del tid på å prøve å forstå hva som ligger i konstruktet "*mathematical literacy*", som definert i rammeverket i Pisa.

1. Oppgavene jeg har utviklet er ikke ment å skulle utgjøre en *test*. Dette vil jeg redegjøre nærmere for senere.

Også under bearbeidelsen av svarene fra den pilotundersøkelsen jeg gjennomførte, har jeg funnet det nødvendig å gå tilbake og reflektere nærmere over ulike problemstillinger knyttet til forståelsen av dette konstruktet. Jeg bestemte meg således relativt tidlig for at en drøfting av "*mathematical literacy*" burde få en vesentlig plass i den teoretiske delen av min oppgave.

Hovedproblemstillingen min vil derfor være :

**Hva er "*mathematical literacy*" og hvordan konstruere matematikkoppgaver som kan benyttes til å teste "*mathematical literacy*" på en adekvat måte?**

For å besvare dette vil jeg foreta en kritisk - teoretisk analyse av konstruktet "*mathematical literacy*", slik det blir utviklet i PISA, og samtidig forsøke å sette dette konstruktet inn i en faglig historisk - didaktisk sammenheng. Jeg vil dessuten redegjøre for de kravene man i PISA stiller til oppgaver som skal brukes for å måle "*mathematical literacy*".

Med bakgrunn i empiri fra min egen pilotundersøkelse og den statistiske bearbeidelsen av resultatene fra den, vil jeg så forsøke å drøfte følgende problemstillinger :

**Har jeg lykkes i å utvikle oppgaver som er i samsvar med PISA's krav slik de kommer til uttrykk i rammeverket for undersøkelsen ?**

**Vil disse oppgavene kunne brukes til å måle i hvilken grad 15 år gamle elever har blitt "*mathematic literate*" individer, dvs. er de valide i forhold til konstruktet "*mathematical literacy*" ?**

Disse to problemstillingene er som man ser nær knyttet til hverandre. De konkrete kravene til oppgavene er jo utformet med tanke på at de aktuelle matematikkoppgavene, som testinstrumenter, på best mulig måte skal kunne måle "*mathematical literacy*".

Angående den første problemstillingen, kan det synes noe overflødig å gå gjennom alle kravene referert i kapittel 3 ett for ett for hver enkelt oppgave, ettersom dette etter min oppfatning vil føre til unødvendig mange gjentakelser. Jeg har derfor heller valgt å poengtere de trekkene ved de ulike oppgavene som jeg mener er mest interessante, og av størst betydning for å avgjøre hvorvidt de er i overensstemmelse med PISA's ulike krav.

Den sistnevnte problemstillingen vil jeg forsøke å belyse gjennom en kvalitativ analyse av de forskjellige oppgavene, og en statistisk analyse av de foreliggende elevsvarene. I den avsluttende drøftingen av hver oppgave forsøker jeg å argumentere for oppgavenes validitet utfra rammeverket i PISA, da særlig i forhold til "*mathematical literacy*".

# 2. Historikk

I dette kapitlet vil jeg, gjennom et historisk tilbakeblikk på de store komparative, internasjonale undersøkelsene som har blitt gjennomført i løpet av de siste decennier, gjøre rede for hvilken forskningstradisjon PISA står i, og til dels er en videreføring av. For en mer utførlig behandling av emnet, se Kobberstad (1991), Kind, Kjærnsli, Lie, Turmo (1999) og Lie, Kjærnsli, Roe, Turmo (2001).

**Internacional Association for the Evaluation of Educational Achievement (IEA)** er en internasjonal organisasjon for utdanningsforskning. Organisasjonen ble grunnlagt i sammenheng med undersøkelsen som senere er blitt omtalt som den første internasjonale studie i matematikk, **FIMS – First International Mathematics Study**. Den ble gjennomført i regi av UNESCO i 1959-61, og omfattet ca. 12000 elever, fordelt på 12 land. Hensikten med undersøkelsen var å finne ut om det metodisk sett var mulig å få verdifull informasjon om kunnskapsnivået i forskjellige fag, i ulike land, gjennom slike komparative studier. (Kobberstad 1991) Elevprestasjoner i matematikkfaget ble ansett som enklest å måle i en slik undersøkelse, og disse ble derfor brukt som et generelt mål for utbyttet av skolens virksomhet. Dette medførte hard kritikk fra matematikkdidaktisk hold. Den anerkjente nederlandske matematikkdidaktikeren Hans Freudental, hevdet at han ikke anså FIMS for å være en matematikkundersøkelse, ettersom det var få matematikere og matematikk - didaktikere med på gjennomføringen av undersøkelsen. Undersøkelsen ble også kritisert fordi fagplanene i deltakerlandene, i liten grad, ble trukket inn i analysen av resultatene.

På tross av denne kritikken, hersket det i fagkretser en utstrakt oppfatning om at internasjonale, komparative undersøkelser i matematikk, kunne gi oss verdifull og viktig informasjon. I tillegg, ble det på 60- og 70- tallet, gjennomført store kostnadskrevende reformer innenfor undervisningssektoren, hvor særlig realfagene ble viet stor oppmerksomhet. Det var derfor også fra politisk hold interesse for å evaluere disse tiltakene.

> *Å finne ut hvilke faktorer som er av betydning for å øke kunnskapsnivået i matematikk, ble ansett som spesielt viktig i et stadig akselererende teknologisk samfunn (Kobberstad 1991, s.14).*

Dette medførte at man i 1976 startet arbeidet med å gjennomføre IEA's andre internasjonale matematikkundersøkelse, **SIMS – Second International Mathematics Study**. Man gjorde her nytte av de erfaringene man satt igjen med etter FIMS, og tok også hensyn til den kritikken som hadde framkommet i etterkant av den undersøkelsen. Matematikere og fagdidaktikere ble i sterkere grad trukket med i planleggingen og gjennomføringen av undersøkelsen. Læreplaner, fagplaner og undervisningspraksis, ble viet langt større oppmerksomhet, i den grad at man sier at undersøkelsen var ***curriculum basert.*** (Hva dette innebærer vil jeg komme nærmere tilbake til nedenfor, under omtalen av TIMSS).

Kritiske røster hevdet at man i slike store komparative internasjonale undersøkelser som FIMS og SIMS, kun testet *om,* og *hvordan,* de nasjonale læreplanene stemte overens med den læreplanen som forskerne hadde konstruert, og som dannet utgangspunktet for undersøkelsene. Utover dette var alle konklusjoner ugyldige (Keitel & Kilpatrick 1999).

Videre gikk kritikken ut på at:

> *No allowance is made for different aims, issues, history and context across the mathematics curricula of the systems beeing studied. No one really addresses how well the students in a system are learning the mathematics curriculum that their system has offered them (Ibid, s.243).*

Ettersom man ikke tok nok hensyn til de kvalitative aspekter ved undervisning, som nevnes i sitatet, ble forskerne beskyldt for å sammenligne det usammenlignbare (Ibid).

Norge deltok ikke verken i FIMS eller SIMS, men derimot i SISS (Second International Science Study), som ble gjennomført parallelt med SIMS.

I stedet for å gjennomføre separate studier i de to realfagene, bestemte man seg på 90 -  tallet for å slå sammen disse, og foreta en kombinert studie i matematikk og naturfag. Den fikk navnet **TIMSS – Third International Mathematics and Science Study**.

> *Prosjektets mål er å beskrive de ulike aspekter av "curriculum" i realfagene og sammenhengen mellom disse. Sentralt står sammenlikning mellom de ulike land, for på den måten å komme fram til hvilke faktorer som best fremmer læring (Lie, Kjærnsli & Brekke 1997, s.8).*

I TIMSS bygget man videre på erfaringer fra FIMS, SIMS og SISS. Man forsøkte å komme kritikerne i møte ved å komplementere de innsamlede ordinære dataene, med studier av selve den konkrete undervisningspraksisen i utvalgte deltakerland, og en grundigere analyse av nasjonale læreplaner. (Keitel & Kilpatrick, 1999)

De innsamlede dataene analyserte man nå på tre forskjellige nivåer, illustrert gjennom begrepene den intenderte læreplan (intended curriculum), den implementerte læreplanen (implemented curriculum) og den oppnådde/resulterte læreplan (attained curriculum). Målet for hele undersøkelsen var å finne sammenhenger mellom elevprestasjoner, og ulikheter på de tre curriculumnivåene, slik at man kunne trekke gyldige slutninger om hvilke faktorer som fremmet god læring i realfagene.(Lie, *et. al.* 1997)

Kritikken forstummet imidlertid ikke. Fremdeles behandlet man ifølge kritikerne, læreplan som en konstant og ikke som en variabel. Kravet var at man gjennomførte en dypere og mer seriøs eksaminasjon av læreplanene i matematikk, knyttet til deltakerlandenes ulike matematikkdidaktiske ideologi og praksis. Slik mente man at evalueringen og vurderingen av elevenes matematiske kunnskaper og ferdigheter kunne bli interpretert på et mer produktivt og valid sett. (Keitel & Kilpatrick 1999)

Som en kort oppsummering kan man si at et kjennetegn ved noen av de første internasjonale komparative undersøkelsene i matematikk, som SIMS og TIMSS, var at de tok utgangspunkt i deltakerlandenes læreplaner, og derfor fokuserte på det vi kan kalle for "skolekunnskaper". Noe av kritikken rettet seg da mot at oppgavene i undersøkelsene, i ulik grad, var i overensstemmelse med deltakerlandenes læreplaner, og at en sammenlikning av elevresultatene derfor ikke var berettiget.

En annen type kritikk som hadde blitt rettet mot FIMS, SIMS og SISS, var at det kanskje viktigste aspektet ved kunnskaper, evnen til å kunne anvende dem, ikke ble fanget opp av denne type undersøkelser. Dette er noe av bakgrunnen for at man i regi av IEA, gjennom TIMSS, parallellt  med den curriculum baserte hovedundersøkelsen, bestemte seg for å utføre en ”*mathematics and science literacy study*”. Denne var rettet mot elever på siste trinn i videregående skole.

Man ville prøve å finne et svar på spørsmålet om hvilke matematiske og naturfaglige kunnskaper elevene sitter igjen med etter 12 års skolegang, og i hvilken grad de er i stand til å anvende denne kunnskapen. Undersøkelsen kom som en følge av den pågående debatt:

> *..about appropriate goals for mathematic and science education, about the needs of young people leaving school and entering a complex world dominated by technology (Orpwood & Garden 1998, s.18).*

Jeg vil forøvrig utdype dette nærmere i kapittel 4.

PISA-undersøkelsen i regi av OECD bygger videre på denne type vinkling. Man fokuserer her på:

> *...elevenes evne til aktivt å bruke kunnskaper og erfaringer og hvordan de forholder seg til emner som trolig vil være relevante for framtiden. PISA – undersøkelsen er basert på en internasjonal konsensus på politisk nivå blant OECD – land om hva som anses å være viktig å kunne i årene framover (Lie, et.al. 2001, s.11).*

Ideene bak PISA vil jeg redegjøre nærmere for i neste kapittel.

# 3. Programme for International Student Assessment (PISA)

Programme for International Student Assessment, heretter omtalt som PISA, er en internasjonal komparativ undersøkelse i regi av OECD. Den har som mål å finne ut hvorvidt 15 åringer i de deltakende land har ervervet de kunnskaper som er nødvendig for å kunne delta i samfunnet på en fullverdig måte. Grunnen til at denne aldersgruppa er valgt ut, er at man i svært mange av deltakerlandene ved dette tidspunkt, avslutter den obligatoriske skolegangen.

Elevene blir testet i oppgaver knyttet til fagene matematikk, naturfag og eget morsmål, dvs for vårt lands vedkommende, faget norsk. Planen er å gjennomføre disse undersøkelsene hvert tredje år. Alle tre fagene er med i hver undersøkelse, men man alternerer på hvilket av dem som står spesielt i fokus. I år 2000 lå hovedvekten, dvs. ca.2/3 av testen, på norskfaget. Ved neste korsvei, i 2003, er det matematikk som blir viet mest oppmerksomhet. Min hovedoppgave vil utelukkende være knyttet til utprøving av oppgaver til denne delen av PISA.

Man har i PISA utviklet et rammeverk, "***Draft Mathematics Framework for OECD/PISA***", heretter forkortet til DMF (OECD PISA 2001a), hvor man eksplisitt gir uttrykk for rammene for undersøkelsen innenfor faget matematikk. Her blir innledningsvis, i tråd med det ovenfornevnte, hovedmålet for prosjektet beskrevet som det å utvikle :

> *..indicators of the extent to which the educational system in participating countries have prepared 15-years old to play constructive roles as citizens in society. The assessment…..focus on determining if students can use what they have learned (OECD PISA 2001, s.5).*

Man ønsker altså å måle elevenes evne til å gjøre bruk av sine matematiske kunnskaper og ferdigheter, i møte med dagliglivets ulike utfordringer, både de av praktisk art,og de mer teoretiske.

Dette blir ytterligere understreket noe senere, når man fastslår at man ønsker å finne ut hvorvidt elevene "*can use the mathematics they have been taught to help make sense of the situations they confront*" (Ibid, s.5).

Den bakenforliggende årsaken til at dette blir ansett som svært viktig, kan kort beskrives på følgende måte:

Daglig blir man bombardert av informasjon knyttet til ulike matematiske begreper. Det er derfor svært relevant å undersøke om elever, som er ferdig med den obligatoriske skole – gangen, kan *nyttiggjøre* seg den matematiske kunnskapen de her har ervervet, og om de evner å orientere seg i møte med den stadig økende informasjonsflommen.

På politisk nivå er det bred enighet blant de deltakende OECD land, om at det vil være ønskelig at ens borgere i framtiden er i besittelse av en slik matematisk kompetanse.

I hvilken grad man på nåværende tidspunkt har lykkes med å utvikle den, er hva PISA-undersøkelsen i regi av OECD, ønsker å finne et svar på.

Resultatene av testen vil også, dersom den avslører åpenbare og spesielle svakheter i populasjonen til et deltakerland, senere kunne danne utgangspunkt for en revidering av nasjonale læreplaner og undervisningsopplegg (Lie et. al. 2001).

Det fins mange fellestrekk i grunnsynet på matematikk mellom PISA og L97. Begge steder vektlegges nytteaspekter ved faget, samtidig som man ønsker å ivareta det almenndannende perspektivet. For PISA er dette momenter jeg senere i min hovedoppgave vil utdype nærmere. Som en følge av slike generelt sammenfallende ideer, kan det hevdes at PISA-testen, prinsipielt, burde passe bra for norske elever (Ibid).

I kapittel 4 vil jeg redegjøre nærmere for det i PISA sentrale konstruktet ”*mathematical literacy*”, mens jeg i kapittel 6 vil gå mer detaljert inn på de viktigste elementene i PISA's rammeverk.

# 4. Hvorfor lære matematikk ?

Både nasjonalt og internasjonalt blir matematikk ansett for å være et av skolens "kjernefag", og det blir i den offentlige debatt sjelden stilt spørsmålstegn ved viktigheten av å lære faget. Man er derfor i den priviligerte situasjon at legitimeringsbehovet utad, er nærmest ikke-eksisterende.

Hva er årsaken til dette? Hvorfor og hvordan har matematikkfaget oppnådd slik en høy status innenfor undervisning? Hvilke type argumenter og tenkemåter har implisitt fått en slik aksept, at de stort sett aldri blir utfordret i den offentlige diskurs omkring skolens innhold?

Dette er problemstillinger som har stor aktualitet innad i vårt fag, og som er blitt gjort til gjenstand for inngående drøftinger av flere framtredende matematikkdidaktikere. I min hovedoppgave har spørsmålene relevans, sett på bakgrunn av den sentrale posisjonen konstruktet "*mathematical literacy*", har i rammeverket i PISA. Jeg vil derfor redegjøre for enkelte utvalgte sider av denne problematikken, og også vie dens historiske genese og utvikling oppmerksomhet. For en mer utførlig og inngående drøfting, se Howson & Mellin-Olsen 1986a og 1986b, Niss 1996, Ernest 2000 og Gjone 1994.

Allerede i oldtiden finner vi ulike årsaker og begrunnelser for læring av matematikk. For fem tusen år siden, i Mesopotamia, var det de herskendes behov for skrivere, til å beregne skatter og avgifter og regulere handel, som førte til opprettelsen av skriverskoler hvor matematikk -faget ble undervist (Ernest 2000). Nærmere starten av vår tidsregning fikk Pythagoreerne æren av å ha befridd aritmetikken fra kun å tjene handelsstanden, og for å ha utviklet geometri til en høyere vitenskap, kort sagt, løsrevet matematikken fra det rent matnyttige (Grønmo 1991).

Allerede tidlig i matematikkens historie, ser vi altså her to kvalitativt ulike begrunnelser for faget, en knyttet til dets nytteverdi, den andre til dets egenverdi.

Niss(1996) hevder at det utfra analyser av matematikkundervisning, i lys av historiske kilder og kontemporære perspektiver, kun kan sies å eksistere noen få typer begrunnelser for faget. De inkluderer at matematikk:

> *(1) contributes to the technological and socio-economic development of society at large*
>
> *(2) contributes to society's political, ideological and cultural maintenance and development*
>
> *(3) provides individuals with prerequisites which may help them to cope with life (Niss 1996, s.13)*

Argument 1 og 3, er klart knyttet til nytteaspektet ved faget, mens argument 2, er av en litt annen karakter, og av Niss blir kalt for "*demokratiargumentet*" (Ibid).

Selv innenfor den vestlige kulturkrets, var det ikke før i det nittende århundre at undervisning, som offentlig anliggende, ble tilbudt et bredere lag av befolkningen. Før dette, var en slik skolering, kun for de få utvalgte og priviligerte.

Den matematikkundervisning som eksisterte i tidlig middelalder, må kunne sies først og fremst å være knyttet til begrunnelse (2) over, ettersom samfunnet på den tid, hadde små muligheter til å nyttiggjøre seg matematikken i nevneverdig grad. I sen-middelalderen og begynnelsen av moderne tid, dvs. fra midten av det attende århundre, ble begrunnelse (3) stadig mer brukt. Dette var relatert til de økonomiske, finansielle og industrielle stenders økende betydning fra dette tidsrom av, og til framveksten av populistiske, demokratiske bevegelser (Ibid).

Eksplisitte uttrykte mål for matematikkundervisningen var nærmest totalt fraværende i læreplaner på 1800 tallet. Hovedvekten i undervisningen ble lagt på at elevene skulle kunne utføre matematiske prosedyrer og metoder, og være i stand til å løse standard matematikk - oppgaver korrekt. Dette ble så testet i skriftlige eksaminasjoner. Selv om det i skrifter fra departementalt hold kunne bli hevdet at målene for undervisningen gikk utover det å tilegne seg denne etablerte kunnskapen, var dette vanskelig å operasjonalisere uten en fagplan å forholde seg til. Det ble imidlertid hevdet, at matematikkundervisningen utviklet elevenes generelle evne til logisk tenkning, og dette ble vurdert som en spesielt positiv egenskap for faget.

I begynnelsen av det tyvende århundre, oppsto det innenfor fagdidaktikken,[2] mer pragmatiske, nytteorienterte bevegelser, som satte spørsmålstegn ved datidens etablerte sannheter, om matematikkfagets evne til å utvikle og forbedre elevenes kognitive ferdigheter. Denne begrunnelsen for faget, som kalles det formative argumentet (Grønmo 1991), har likevel blitt brukt for å legitimere matematikkundervisning helt fram til våre dager.

På nittenhundretallet finner vi videre at alle de tre argumenter Niss nevner, med vekslende styrke, er blitt brukt for å begrunne matematikkundervisning. I politiske og økonomiske nedgangstider, for eksempel i mellomkrigstiden, har nytteaspektet ved faget blitt aksentuert. Demokratiargumentet har hatt større gjennomslagskraft i økonomiske oppgangstider, som på 50- og 60-tallet (Niss 1996).

Legitimeringen av undervisningen i matematikk her i landet, har selvsagt vært influert av internasjonale trender. I almuesskolen, fra midten av det attende århundre, og fram mot slutten av det nittende, skulle matematikk, i likhet med de andre fagene, være et ”*formativt dannelsesmiddel*” (Gjone 1994). Nytteaspektet ved faget var svært lite framtredende. Det vil si, at det man lærte på skolen, hadde man lite glede av i det praktiske liv. Dette ble endret ved innføringen av folkeskolen, hvor nytteverdien ved faget, som her ble kalt regning, ble eksplisitt understreket. Etter at lovene om pengevesenet og om metriske mål trådte i kraft i 1875, ble dette ytterligere aksentuert (Ibid).

Fram til 2.verdenskrig hadde vi derfor to ulike faglige matematiske tradisjoner i skolen, folkeskolen med regning og realskole/gymnas med matematikk. Den førstnevnte knyttet til nytteaspektet ved faget, den sistnevnte til dannelse.

2. Strengt tatt er matematikkdidaktikk som fag av langt nyere dato (se Gjone 2001), men jeg velger likevel her å benytte ordet ”fagdidaktikk”, ettersom det er snakk om tanker og ideer knyttet til undervisning.

I tiden etter 2. verdenskrig, har tanken om "Enhetsskolen", vært en dominerende faktor og en drivende kraft i norsk og nordisk skoleutvikling. Det er naturlig å knyttet dette til den sterke posisjonen sosialdemokratiet har hatt i Skandinavia. Fremdeles står enhetstankene sterkt i norsk skolepolitikk, noe ikke minst R94 og L97 er illustrerende eksempler på.

Gjone (1994) hevder at:

> *…det meste av det som har skjedd i etterkrigstidens matematikkundervisning kan beskrives som en svingning mellom to matematikksyn. På den ene siden er matematikk nyttig, på den andre siden er den også ledd i en danningsprosess (Gjone 1994, s.1).*

At alle elever skulle bli tilbudt den samme undervisning[3], medførte på femtitallet her til lands, at matematikkfagets nytteaspekt ble nedtonet, og dannelsesmotivet ble mere framtredende.

Internasjonalt var, som påpekt av Niss, nytteaspektet ved faget dominerende i 1930-åra. Reaksjonen mot dette kom etter 2.verdenskrig. "Moderne matematikk", er betegnelsen man oftest bruker på reformbevegelsen som vokste fram mot midten av 50-tallet, og som varte i ca.20 år. Dannelsesaspektet var igjen det dominerende, samtidig som man gjennom undervisningen la vekt på de strukturelle sider ved faget. Selv om man uttrykte ønske om, og hadde tro på, at også fagets nytteverdi ble tatt vare på gjennom den moderne matematikken, var kritikken som etter hvert reiste seg mot bevegelsen, rettet mot nettopp den påståtte mangelen på dette. Nytteaspektet ved matematikkundervisningen, ble altså etterlyst.

Den nye bevegelsen, som oppsto i kjølvannet av denne kritikken, fikk navnet "*Back to basics*", og man skulle nå konsentrere seg om å utvikle de grunnleggende regneferdighetene.

Denne bevegelsen ble ikke like sterk i Norge som i mange andre land. Mønsterplanen, som var ferdig i 1974, understreket likevel klart nytteaspektet ved faget. Dette var også det dominerende perspektivet utover på 1980-tallet, og i mønsterplan-revisjonen i 1987.(Ibid)

I løpet av de siste decennier har målene for matematikkundervisningen i grunnskolen og videregående utdanning, blitt utvidet betraktelig. Følgende generelle mål har fått sin tilslutning fra det store flertall av land:

- (1) "matematikkundervisning for alle", dvs. alle bør få mulighet til å utvikle en viss matematisk kompetanse
- (2) undervisning må differensieres utfra den enkelte elevs evner og forutsetninger
- (3) deltakelse og samarbeid mellom elever for å løse felles oppgaver må stimuleres
- (4) vurderingen må skje i overensstemmelse med overordnede mål for undervisningen

3. Den obligatoriske skolen har her i landet i etterkrigstiden blitt utvidet fra 7 til 10 år. Over 80% av årskullene har de siste årene i tillegg fullført videregående skole.

- (5) fokus bør være på den enkelte elevs behov og interesser, med tanke på mulighetene til aktivt å kunne delta i det sosiale og offentlige liv
- (6) utvikle elevenes personlighet ved å bygge opp selvtillit, uavhengighet og selvstendig tenkning
- (7) legge vekt på matematisk aktivitet og matematiske prosesser
- (8) utvikle matematisk tenkning og kreativitet, og evne til problemløsning
- (9) bidra til at elevene verdsetter fagets egenart, og samtidig dets styrke for modeller og modellbygging
- (10) gjøre elevene i stand til å kunne kritisere og vurdere bruken av faget i ekstra-matematiske kontekster (Niss 1996, s.32-33, min oversettelse).

Selv om mange land altså vil bifalle denne type mål, er det store forskjeller på hvordan de blir tolket, og hvilke former for undervisning man praktiserer. Disse forskjellene finner man både intranasjonalt og internasjonalt.

> *…in a great many places there is a marked gap between society's acknowledged goals of mathematics education and the reality of teaching and learning in the same society (Ibid, s.33).*

De*t*te er noe av bakgrunnen for at man i TIMSS, som jeg nevnte i kapittel 2, legger stor vekt på læreplananalyser, og der skiller mellom intendert, implementert og oppnådd læreplan (Robitaille & Garden 1996).

Flere av de ovenfor beskrevede målene for matematikkundervisning, som det altså internasjonalt er bred enighet om, er nært knyttet til konstruktet "*mathematical literacy*" som utviklet i TIMSS og siden videre elaborert i PISA.

## 4.1 Mathematical Literacy

Som nevnt i forrige kapittel, hersker det i dag en utbredt oppfatning om at kunnskaper og ferdigheter i matematikk, generelt sett, er viktig. Samtidig kan det helt klart stilles spørsmålstegn ved i hvilken grad, eller hvor mye matematisk kunnskap man trenger.

Forskning har vist at det i yrkeslivet er slik at nødvendigheten av å beherske matematiske teknikker, relatert til den jobben man utfører, for det meste består av repetisjoner av noen få veldefinerte oppgaver, innenfor et smalt ferdighetsområde, som for det meste er aritmetisk. I økende grad er det mulighet for å løse denne aritmetikken ved hjelp av kalkulator. (Dørfler & McLone 1986)

I et nytteperspektiv behøver man strengt tatt først og fremst ulike tallferdigheter, dvs.

- kunne telle
- lese tall
- kunne klokken
- tyde rutetabeller
- forstå enkle grafiske fremstillinger o.l. (Ibid)

Nyttesynet på behovet for matematikkundervisning i skolen, kan oppsummeres som et argument for en større vektlegging av de ferdigheter som er nødvendige for å beherske dagliglivet. Elever som av forskjellige årsaker ønsker kunnskaper utover dette, kan ifølge et slikt syn tilbys en mer spesialisert matematikkundervisning.

Ernest(2000) går så langt som å hevde at:

> *….although there is undoubtedly an information revolution taking place, increased mathematical knowledge is not needed by most of the population to cope with their new roles as regulated subjects, workers and consumers (Ernest 2000, s.3).*

Dette har av Niss(1994) blitt kalt "*relevans paradokset*", på grunn av matematikkens samtidige objektive relevans og subjektive irrelevans.

Ernest, i likhet med en rekke andre framtredende matematikkdidaktikere, mener at det å legitimere faget kun ut fra nyttebetraktninger, derfor er uheldig. Han argumenterer for at matematikkundervisning heller bør begrunnes i:

> *The appreciation of mathematics as making a unique contribution to human culture with special concepts and a powerful aesthetic of its own (Ibid, s.7).*

Han plasserer seg dermed innenfor dannelsestradisjonen i faget.

På tross av ulike oppfatninger om hvordan matematikkundervisning bør legitimeres, og hvilke kunnskaper og ferdigheter det er viktig at elevene behersker, har det i kjølvannet av de store internasjonale komparative undersøkelsene, som har blitt gjennomført innenfor faget, både i utdanningspolitiske kretser og innenfor fagdidaktiske miljøer, blitt stadig økende interesse knyttet til spørsmålene om "*the residue of conceptual learning*" (Orpwood & Garden 1998, s.11).

Hvilke kunnskaper og ferdigheter sitter elevene igjen med etter fullført skolegang? Er de i stand til å anvende disse kunnskapene i det virkelige liv? Slike problemstillinger har man gjennom forskning ønsket å undersøke nærmere, og om mulig antyde svar på.

Første gang denne type forskningsspørsmål ble forsøkt belyst i en større internasjonal sammenheng var, som nevnt i kapittel 2, i regi av TIMSS. Parallellt med den ordinære curriculum-baserte undersøkelsen i 1996, gjennomførte man altså "*The Mathematics and Science Literacy Study*". Målet ble formulert slik:

> *The role of the literacy study within TIMSS, is to ask whether school leavers can remember the mathematics and science they have been taught and can therefore apply this knowledge to the challenges of life beyond school (Ibid, s. 11).*

Undersøkelsen rettet seg mot populasjon 3 i TIMSS, det vil si elever i siste året i videregående skole. Man ønsket altså å bidra til å besvare spørsmålene omkring hva som "sitter igjen", etter mange års undervisning, i matematikk og naturfag.

I rammeverket for undersøkelsen definerte man komponentene til "*mathematic literacy*" og "*science literacy*" for å være:

- mathematical content
- science content
- reasoning and social utility in mathematics, science and technology (Ibid)

Som man ser var hovedfokus her delvis rettet mot konteksten til fagene.

Oppgavene i matematikkdelen av undersøkelsen ble utviklet og utvalgt fra innholdsområdene:

- number sense
- algebraic sense
- measurement and estimation (Ibid)

Et ytterligere kriterium i utvelgelse av oppgaver var at de omhandlet:

> *… the sort of mathematics question that could arise in real-life situations and that they be contextualized accordingly (Ibid, s.38).*

Selv om undersøkelsen knyttet til "*mathematics and science literacy*"- problematikken kun utgjorde en begrenset del av TIMSS, vakte den stor interesse blant så vel politikere, presse og media som fagdidaktikere.

Ettersom det internasjonale matematikkdidaktiske fagmiljøet rent kvantitativt er nokså begrenset, vil omfattende internasjonale forskningsprosjekter knyttet til faget påvirke målene for senere undersøkelser. Det synes klart at denne undersøkelsen, i regi av IEA, var noe av bakgrunnen for at man senere, i regi av OECD, gjennom PISA, bestemte seg for å gjennomføre en ny og lignende undersøkelse, nettopp med fokus på "*mathematical literacy*".

Målgruppa for undersøkelsen var nå elever som avsluttet den obligatoriske skolegangen, dvs.15 åringer.

Gjennom PISA undersøkelsen ønsker man, som jeg redegjorde for i kapittel 3:

> *…to develop indicators of the extent to which the educational systems in participating countries have prepared 15-year-olds to play constructive roles as citizens of society (OECD PISA 2001a, s.5).*

Vi ser at denne målformuleringen ligger nær opp til de refererte målene i TIMSS, og at man, på tross av at IEA og OECD er totalt uavhengige organisasjoner, må tillate seg å kunne påpeke kontinuiteten i den nye forskningsproblematikken.

I DMF blir det, som tidligere omtalt, understreket at fokus er rettet mot hvorvidt elevene kan <u>bruke</u> det de har lært, og i hvilken grad 15-åringene følgelig kan bli ansett som "*informed and reflective citizens, and intelligent consumers*" (Ibid, s.5).

Det er altså ikke "rene" matematiske ferdigheter og kunnskaper, som man tradisjonelt har vært mest opptatt av i tidligere internasjonale matematiske undersøkelser som FIMS og SIMS, man ønsker å måle. Derimot fokuseres det på :

> *..mathematical knowledge put into functional use in a multitude of different contexts in varied, reflective and insight-based ways ( Ibid, s.5).*

Dette er bakgrunnen for at konstruktet "*mathematical literacy*", står sentralt i den teoretiske drøftingen av målene i PISA, slik de presenteres i dets rammeverk. "*Mathematical literacy*" blir definert slik :

> *Mathematical Literacy is an individual`s capacity to identify and understand the role that mathematics play in the world, to make well-founded judgements and to engage*

*in mathematics, in ways that meet the needs of that individual`s life as a constructive, concerned and reflective citizen (Ibid, s.5).*

For å kunne nyttiggjøre seg matematisk kunnskap, er det selvsagt en forutsetning at man er i besittelse av den. "Mathematical literacy" forutsetter derfor kjennskap til matematisk terminologi, fakta og prosedyrer, men viser utover disse kunnskapene isolert sett. Til grunn for denne vinklingen i PISA, ligger overbevisningen om at hovedmålet for matematikk – undervisningen, i alle land, bør være å lære elevene å *matematisere*. Dette begrepet knyttes i DMF til det som kjennetegner en matematisk tilnærmingsmåte til autentiske problemer, og som kan karakteriseres på følgende måte :

1. Man starter med et problem som er knyttet til virkeligheten.
2. Man organiserer problemet i henhold til matematiske begreper og identifiserer den relevante matematikken.
3. Gradvis transformerer man det virkelige problemet til et matematisk problem, gjennom å gjøre antagelser, generalisere og formalisere.
4. Det matematiske problemet løses.
5. Løsningen knyttes til det opprinnelige problemet, slik at dette framstår som løst. Samtidig identifiserer man løsningens begrensninger (Ibid, s.16-17, min oversettelse).

Tanken er at dersom elever lærer å arbeide på denne måten, allerede mens de går på skolen, vil de kunne trekke veksler på det senere i livet. Grafisk framstilles dette slik:

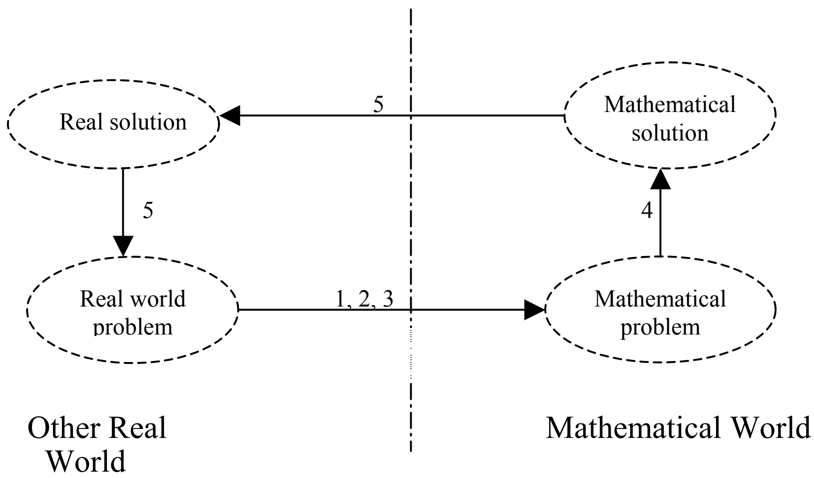



**Fig.1**

Som det går fram av fig.1 fordeles de fem punktene i tre grupper. Den første er knyttet til det å oversette det foreliggende virkelige problemet til et matematisk problem. Den andre består i å løse problemet matematisk, "*the deductive part of the modelling cycle*" (OECD PISA 2001a, s.18). Det siste trinnet er relatert til refleksjonen over hele den matematiske prossessen man har vært igjennom og det resultatet man har kommet fram til.

Fokuseringen på "*mathematical literacy*", i PISA, begrunnes avslutningsvis i to hovedpunkter.  Det første er at for å orientere seg i den eksponensielt økende flommen av informasjon som omgir oss, og samtidig være i stand til å ta avgjørelser og felle dommer over sannhets - gehalten i den, bør man som en aktivt, deltakende samfunnsborger, være en ”*mathematical literate*” person. Konsekvensene av ikke å være det, kan åpenbart være fatale, ettersom man da kan svindles og duperes på et endeløst antall måter.

Det andre punktet er knyttet til at våre samfunn i uoverskuelig framtid, og i takt med den økende graden av teknifisering, vil ha behov for en stadig mer kompetent arbeidskraft. I flere og flere yrker vil det derfor kreves evne til å beherske prosedyrer knyttet til, og basert på, matematisk tenkning.

Selv om nyttebetraktninger er sterkt framtredende i begrunnelsen for fokuseringen på ”*mathematical literacy*”, ser vi, ikke minst i definisjonen av konstruktet, at man implisitt distanserer seg fra en ensidig og smal generell nyttelegitimering av faget. En formulering som:

> *Mathematical literacy is an individuals capacity to identify and understand the role that mathematics play in the world.. (Ibid, s.5).*

er klart nærmere begrunnelser hørende hjemme innenfor dannelsestradisjonen i faget. Det kan derfor hevdes at man i PISA, gjennom utviklingen av dette konstruktet, forsøker å lage en syntese av nyttetradisjonen og dannelsestradisjonen innenfor matematikkfaget.

## 4.2 Fra ”literacy” til ”mathematical literacy”

”*Mathematical literacy* ” konstruktet er åpenbart sprunget ut av ”*literacy* ”- begrepet innenfor filologi og pedagogikk. Dette blir også understreket i PISA’s rammeverk.

> *For a person to be literate in language implies that the person knows many of the design resources of the language and is able to use those resources for several different social functions. Analogously, considering mathematics as a language implies that students not only must learn the design features involved in mathematical discourse,… they also must learn to use such ideas to solve non-routine problems in a variety of situations defined in terms of social functions (OECD PISA 2001a, s.6-7).*

Innenfor filologi og pedagogikk er det lange, faglige tradisjoner for drøfting av de ulike aspekter ved viktigheten av at alle mennesker er i besittelse av ferdigheter i lesing og skriving. Politisk har man, ikke minst i land med demokratiske styresett, stort sett fått fullt gjennomslag for argumentasjonen om at det er en fundamental menneskerett å lære og lese og skrive. I tråd med dette, har det blitt ansett som en høyt prioritert statlig oppgave å ta ansvar for en slik opplæring. Dette ansvaret har så blitt delegert videre til læreplanutviklere og praktiserende pedagoger.

Et av argumentene med bredest appell har vært at dersom et demokratisk styresystem skal fungere, er det nødvendig at samfunnsmedlemmer holder seg informert og orientert. Følgelig må de være lese/skrive - kyndige. En person som ikke er i besittelse av disse ferdighetene, vil også marginaliseres i arbeidsmarkedet, ettersom det store flertallet av yrker forutsetter gode lese- og skriveferdigheter.

Like åpenbart er det kanskje ikke at man, for å kunne delta som fullverdig samfunnsmedlem, bør være matematisk ”*literate*”, men svært mange matematikkdidaktikere argumenterer for et slikt syn. Det er vel riktig å si at man nå også er i ferd med å oppnå politisk gjennomslag for slike tanker, både internasjonalt og nasjonalt, noe ikke minst den brede deltakelsen i

PISA er en bekreftelse på. Likeledes har vi på nasjonalt nivå de siste åra, opplevd at det med lignende begrunnelser er blitt satt et betydelig og spesielt fokus på realfagenes plass i skolen.

Jeg vil senere komme tilbake til denne diskusjonen. Først vil jeg utdype slektskapet mellom konstruktet ”*mathematical literacy*” og ”*literacy*”- begrepet, slik dette av enkelte pedagoger, er blitt kritisk analysert og vurdert.

Den brasilianske pedagogen Paolo Freire understreket at det å lære og lese er en politisk handling. I hans hjemland, hvor analfabetismen på 50-60 tallet var svært høy, utviklet han pedagogisk materiale som gjorde at voksne kunne lære å lese på 30-40 timer. Det geniale i hans metode, var at han tok utgangspunkt i kursdeltakernes egne erfaringer, språk og kultur. Gjennom sitt opplegg viste han at deres tolkning og opplevelse av virkeligheten, var like viktig og riktig som maktelitens. Disse beskrivelsene ble derfor brukt som utgangspunkt for kritiske diskusjoner, knyttet til den politiske hverdag. Resultatmessig var hans kampanje en stor suksess, men hans radikale budskap skapte stor frykt i politisk konservative kretser, og Freire ble i 1964 tvunget i eksil.

Freires kunnskapssyn er konstruktivistisk. Han understreker at kunnskap ikke er statisk, men kontinuerlig skapes gjennom at mennesker reflekterer og handler. Kunnskap krever subjekter, og eksisterer ikke utenfor den menneskelige bevissthet, hevder han. Hensikten med kunnskapen, mener Freire, er frigjøring fra politisk, økonomisk og kulturell dominans. Dette oppnås ved at man gjennom ”*literacy*”, altså ved å bli en ”*literate*” person, klarere kan forstå hvordan ens liv er skapt av verden, men at det også er mulig å forandre den. I sin egen undervisning av voksne analfabeter tok han utgangspunkt i distinksjonen mellom begrepene natur og kultur. Gjennom å illustrere hvordan vår kultur er menneskeskapt, og det som først og fremst skiller oss fra dyreverdenen, lyktes han i å få sine elever til å innse at mulighetene til å påvirke og endre sine livsbetingelser derfor var mange. Å bli en ”*literate*” person er i våre moderne samfunn en nødvendig betingelse for å kunne delta i ”*making decisions and sharing power*” (Brown 1987, s.215).

Freire var opptatt av at undervisningen bør ha en dialogisk form, hvor lærer og elever i fellesskap undersøker virkelige problemer knyttet til hverdagen. Det å kunne formulere og analysere problemene, er like viktig som det å løse dem, mente han. Ja, i virkelighetens komplekse verden, er det faktisk ofte slik at problemer ikke blir løst, men at man kun kan oppnå en større innsikt i hva de består i.

Freire's epistemologiske innsikter bør også kunne overføres til matematikkundervisningen, og dette har da også blitt forsøkt gjort av flere matematikkdidaktikere.

D`Ambrosio (1985), inspirert av sin landsmann Freires arbeider innenfor pedagogikken, analyserer det han kaller en interkulturell overføring av matematisk kunnskap. Gjennom å introdusere et kulturbegrep som krever en analyse av individuell og sosial adferd, ses etnomatematikk som en form for strukturert kunnskap og ”*matheracy*” som en karakteristisk type menneskelig adferd. ”*Matheracy*” – begrepet står for den egne matematikken elevene har med seg til skolen, og som, i likhet med det språket de evner å kommunisere med, fungerer for dem til daglig. D`Ambrosio understreker derfor viktigheten av at matematikk - undervisning må ta utgangspunkt i elevenes etablerte begrepsmessige forståelse (deres forforståelse), knyttet til deres opplevelse av virkeligheten. Dersom man ikke makter dette, vil man kun tjene den herskende klasses interesse, hevder han.

Hvilken relevans har så disse synspunkter i forhold til PISA ?

I rammeverket for PISA understrekes det at man ønsker å måle hvorvidt elevene kan anvende sine matematiske kunnskaper, og at man ønsker å gjøre dette i tilnærmede, autentiske "*real - life*"- situasjoner. Det blir også påpekt at språket i oppgavene bør være lett forståelig. Dette er altså helt i tråd med ideene til D'Ambrosio som jeg refererte ovenfor. Vanskelighetene med å gjennomføre dette er særlig knyttet til autentisitet - problematikken.

> *This authenticity in the use of mathematics is an important aspect of the design and analysis of items for PISA, strongly related to the definition of mathematical literacy (OECD PISA 2001a, s.12).*

Det blir her understreket at autentisitets-aspektet er svært viktig både for utviklingen og analysen av instrumenter. I PISA har man således høye ambisjoner om å kunne måle "*mathematical literacy*" utfra slike perspektiver. I hvilken grad har man så mulighet til å lykkes med dette ?

Skolen er på mange måter en kontekst hvor problemer blir oppkonstruert. For å lykkes fullstendig med det som er PISA`s intensjon, måtte vi ideelt sett vært "flue på veggen" i, for elevene, reelle dagligdagse situasjoner utenfor skolen. Dette ville åpenbart vært svært vanskelig å gjennomføre i en større kvantitativ undersøkelse, som jo krever et utvalg av en viss størrelse, for at eventuelle funn regnes som interessante, men kunne kanskje i en mer begrenset skala, vært et fruktbart utgangspunkt for en kvalitativ undersøkelse.

Ettersom mange av verdens ledende matematikkdidaktikere deltar i utformingen og gjennomføringen av PISA, er man selvsagt oppmerksom på en del av begrensningene som ligger i undersøkelsens format.

> *The operational problem faced by PISA is how to assess whether 15-years-old students are mathematically literate in terms of their ability to mathematise. Unfortunately, in a timed assessment this is difficult because for most complex real situations the full process of proceeding from reality to mathematics and back often involves collaboration, involves locating appropriate resources, and takes considerable time (Ibid, s.8).*

For i det hele tatt å få gjennomført denne type kvantitativ, komparativ undersøkelse, er man derfor tvunget til å være pragmatisk, og inngå kompromisser i forhold til det absolutt ideelle. Man legger derfor i PISA stor vekt på at oppgavene skal ha en autentisk kontekst og ligge så nær opp til "*real life*" som mulig. Slik håper man, gjennom PISA 2003, å kunne bidra til å øke vår felles kunnskap om i hvilken grad undervisningen i de ulike deltakerland bidrar til å gjøre 15-åringer til "*mathematical literate*" individer. I tillegg har man ambisjoner om å peke på utslagsgivende bakgrunnsfaktorer for dette, slik at det er mulig å omstrukturere undervisningsopplegg og sette inn krefter der det er størst behov. Om man vil foreta en slik eventuell allokering av ressurser må selvsagt avgjøres på nasjonalt nivå.

## 4.3 "Mathematical literacy" og reflektiv viten

Den danske matematikkdidaktikeren Ole Skovsmose drøfter i flere av sine artikler ulike aspekter ved konstruktet "*mathematical literacy*". I samsvar med rammeveket i PISA, begrunner han nødvendigheten av at mennesker må være matematisk "*literate*", med samfunnets behov for en fungerende arbeidsstokk, og dets ønske om å kunne informere folk om deres forpliktelser. Han er inspirert av Freires teorier og arbeider, og hans analyse av "*mathematical literacy*" inneholder, i tillegg til det ovenfornevnte, en kritisk dimensjon som jeg er tilbøyelig til å mene i kanskje litt for liten grad vektlegges i PISA. (Dette er en problemstilling jeg senere vil komme tilbake til).

Skovsmose (1992) knytter "*mathematical literacy*", til Freires kritiske og frigjørende "*literacy*" - prosjekt, og hevder at det som et radikalt konstrukt, må grunnes i en kritisk holdning til sosiale maktstrukturer, med fokus på de muligheter det gir folk til å delta i det politiske arbeidet med å endre samfunnet. Han hevder at dersom vi opererer med et utvidet demokratibegrep, som innebærer aktivt kritiske aktører, medfører dette at samfunnet må legge forholdene til rette, slik at dets medlemmer tilegner seg den matematiske kompetansen som er nødvendig, for at de skal kunne framstå som "*mathematical literate*" personer. Matematikkundervisningen i skolen, vil i et slikt perspektiv, åpenbart spille en svært viktig rolle. Skovsmose (1994) beskriver selv et undervisningsopplegg han med hell har vært med på å utvikle og gjennomføre på ungdomsskolenivå i Danmark, hvor man tok utgangspunkt i følgende problemstillinger:

> *Is it possible in elementary mathematics education to make any sense of seeing mathematics as a principle for design? Is it possible to create situations in which the students are using competencies, developed in mathematics education, in making interpretations of social phenomena? Is it possible to involve students in a process of system development in order to make them understand something about conditions of mathematical based management? (Skovsmose 1994, s.39).*

Gjennom å peke på mulighetene som ligger i kritisk matematikkundervisning, svarer han langt på vei bekreftende på disse spørsmålene.

Demokratiproblemet i våre høyt utviklede teknologiske samfunn er, ifølge Skovsmose, i stor grad knyttet nettopp til det å utvikle en generell, kritisk kompetanse, som kan matche den sosiale og teknologiske utviklingen.

Hovedbudskapet hans er altså, kort oppsummert, at skal demokratiet fungere, må dets aktører være "*mathematical literate*" individer.

For å lykkes med å utvikle en slik kompetanse, kreves det at man gjennom undervisningen i skolen makter å integrere matematisk, teknologisk og reflektiv kunnskap. Dette blir av Skovsmose knyttet til tre ulike typer viten, som han hevder at matematikkundervisning således kan orienteres mot, nemlig:

1. Matematisk viten (mestring av algoritmer, teoremer, bevis etc.).

2. Teknologisk viten (evner til å anvende matematisk kompetanse i modellbygging).

3. Reflektiv viten (Kompetanse i å reflektere over, og evaluere bruk av matematikk. Refleksjon er knyttet til evaluering av konsekvensene av teknologiske foretak). (Skovsmose 1992)

Skovsmose utdyper forskjellen mellom teknologisk og reflektiv viten ved å understreke at teknologisk viten ikke er i stand til å vurdere resultatene av dens egen produksjon, da behøves refleksjon. Kort sagt, hevder han, at teknologisk viten har som mål å løse et problem, mens reflektiv viten, evaluerer løsningen.

Videre forklarer han, at matematikk bidrar med å formatere vår oppfatning av samfunnet og de sosiale prosesser som her foregår. Han understreker at matematikk ikke bare består i beskrivelser og håndtering av ulike typer problem, men at den i likhet med for eksempel ideologi, på en fundamental måte, påvirker og styrer vår virkelighetsoppfatning. Dette kan på mange måter, ses på som en videreutvikling av Freires tanker om den sosiokulturelle og politiske betydningen av å bli en "*literate*" person, nå knyttet til "*mathematical literacy*".

Hvis mennesker ikke bare skal være passive mottagere av informasjon, men også delta aktivt i utformingen av vårt samfunn, må de oppnå en forståelse av samfunnets grunnleggende strukturerende prinsipper. Både språkbruk og matematikkbruk, er formaterende krefter i samfunnet, og båndene til Freires "*literacy"* - prosjekter er derfor naturlige å knytte. I likhet med disse prosjektene, kan og bør derfor matematikkundervisningen, ifølge Skovsmose, spille en kritisk rolle i dagens samfunn.

## 4.4 PISA og reflektiv viten

Flere andre matematikkdidaktikere har, i slektskap med Freire's kritiske epistemologi knyttet til "*literacy*"- konstruktet, forsøkt å utvikle lignende teorier for matematikk, og analysert de ulike konsekvensene som derav følger for matematikkundervisningen. Her vil jeg kort presentere noen av disse, og forsøke å peke på hvilke utfordringer slike perspektiver kan gi oss som matematikkdidaktikere, med et særlig blikk på oppgaveutvikling. Jeg vil også i dette, og det påfølgende delkapittel, knytte noen kritiske kommentarer til PISA-undersøkelsen og dens versjon av "*mathematical literacy*", utfra en slik vinkling. Disse kan muligens fortone seg i overkant skeptiske til hele PISA-undersøkelsen som ide, men det er langt fra min intensjon å tilkjennegi den type synspunkter. Snarere synes jeg det er et svært spennende prosjekt, som det har vært et privilegium å få delta i. Enkelte av de matematikkdidaktikere jeg støtter meg til i mine kritiske bemerkninger i min hovedoppgave, for eksempel Niss og Kilpatrick, har selv vært med på å utvikle rammeverkene i PISA. Kritiske bemerkninger betyr derfor selvsagt ikke en desavuering av prosjektet som sådan. I følge anerkjente vitenskapsteoretikere som for eksempel Popper (1959), Kuhn (1970) og Lakatos (1979), er det jo nettopp slik at den konvensjonelle forskningen drives framover i spenningsfeltet mellom det kreative og det kritiske, i en slags dialektisk utvikling.

I forhold til PISA - undersøkelsen synes det naturlig å spørre om i hvilken grad PISA, gjennom sitt rammeverk og sine intensjoner, makter å stimulere den viktige reflektive kunnskapen. Er det overhodet mulig å oppnå dette gjennom denne type undersøkelser, eller er det kun det Skovsmose kaller den teknologiske komponenten i "*mathematical literacy*" som måles, og som man derfor strengt tatt markedsfører?

Før jeg gir meg inn på denne diskusjonen, vil jeg gjerne presentere andre matematikk - didaktikeres tanker knyttet til matematikkens samfunnsmessige og virkelighetsproduserende betydning.

John Abraham og Neil Bibby hevder at matematikkundervisning i tillegg til å relatere til elevenes erfaring, må inneholde en kritisk dimensjon som nettopp er orientert mot en forståelse av hvordan våre erfaringer er sosialt influert og determinert. I likhet med Skovsmose, understreker de, at matematikk strukturerer våre erfaringer og vurderinger. Det er derfor viktig at elever gjennom matematikkundervisningen i skolen oppnår en forståelse av matematikk som en sosialt organisert aktivitet, som bidrar til å forme vårt virkelighetssyn. De mener videre, at elever derfor må konfronteres med ulike måter å bruke matematikken på, og hvilke sosiale konsekvenser det kan få (Abraham & Bibby 1988).

For kort å vende tilbake til spørsmålene jeg knyttet til PISA, kan det synes som om man, gjennom denne type undersøkelser, i liten grad evner å stimulere til en type matematikk –

undervisning som i tilstrekkelig grad evner å integrere det ovenfornevnte aspekt med den mer tradisjonelle formidlingen av faget.

Selve rammene for undersøkelsen ved deltakerskolene gjør dette vanskelig. Oppgavene skal for eksempel besvares i løpet av en dobbeltime, og det er ikke lagt opp til noen drøfting av alternative løsningsmuligheter i plenum. Det er jo også vanskelig å se hvordan dette skulle kunne la seg kombinere med testens krav til reliabilitet.

I tillegg vil korte tidsrammer selvsagt virke begrensende på den type oppgaver det er mulig å bruke i undersøkelsen. Her spiller også reliabilitetskrav inn, ettersom man for å tilfredsstille dem, må sørge for at hver elev besvarer et relativt stort antall oppgaver. En enkelt oppgave bør derfor ikke være så tidkrevende, at elevene bruker en for stor del av sin tid på å løse den.

I rammeverket i PISA gir man også uttrykk for å se enkelte begrensninger testformatet utøver på mulige oppgavetyper:

> *Ideally, to judge whether 15-year-old students can make use of their accumulated mathematical knowledge to solve mathematical problems confronted as they interact with the world, one would collect information about their capability of mathematising a number of (such) complex situations. Clearly this is impractical. Instead, PISA has chosen to prepare items to assess different parts of this process (OECD PISA 2001a, s.9).*

Nå er det jo slik at dersom elevene skal oppnå en forståelse av matematikk som en sosialt organisert aktivitet som bidrar til å forme vårt virkelighetssyn (Abraham & Bibby 1988), krever det nettopp en type oppgaver som ofte vil være tidkrevende, i og med at de blant annet vil måtte basere seg på muntlige drøftinger underveis og i etterkant av elevenes arbeid med dem.

Den anerkjente danske matematikkdidaktikeren Mogens Niss hevder at hovedproblemet for den trenden i dagens matematikkundervisning, som jeg er tilbøyelig til å mene PISA står for, med vekt på anvendt matematikk knyttet til autentiske oppgaver, er at det fokuseres for mye på elevenes nærmiljø. Problemer av større sosial betydning, men lenger unna deres daglige liv, og således mindre "relevante" og "autentiske", kommer følgelig i skyggen. Han konkluderer med at vi må investere mye anstrengelse i å finne signifikante problemer av større samfunnsmessig og strukturell betydning, hvor matematikk kan og må benyttes for å finne en god løsning (Niss 1987).

## 4.5 "Reflective citizens" som kritiske aktører.

Selv om ordet "kritisk" ikke benyttes i PISA's definisjon av "*mathematical literacy*", velger jeg å tolke uttrykket "*reflective citizen*" , som finnes der, i en retning hvor det inneholder et kritisk element. Man kan etter min mening nemlig vanskelig kalles en reflekterende samfunnsborger, dersom man ikke har evnen til kritisk tenkning. I PISA ønsker man altså at matematikkundervisningen skal bidra til at elevene utvikler en slik kritisk evne.

Hva "kritisk evne" så består i, og hvordan den best utvikles, vil det selvsagt kunne være delte meninger om.

I DMF er man forsiktig med å trekke fagpolitiske slutninger over hva det innebærer å være en "*reflective citizen*". Det blir først og fremst knyttet til den personlige interessesfære, individets egeninteresse av å kunne orientere seg i et stadig mer komplekst og uoversiktelig informasjonssamfunn. Likevel finnes det også antydninger til politiske implikasjoner, blant annet sies følgende :

> *Citizens are currently being bombarded with information on issues such as "global warming and the greenhouse effect", "population growth", "oil slicks and the seas", "the disappearing countryside" (OECD PISA 2001a, s.5 ).*

Og videre:

> *One example of the need for a citizen to have mathematical literacy is the frequent demand for individuals to make judgments and assess the accuracy of conclusions and claims about information from surveys and studies. Being able to judge the soundness of the claims from such arguments is, and increasingly will be, a critical aspect of being a responsible citizen (Ibid, s.8).*

I et så omfattende samarbeidsprosjekt som PISA er man selvsagt avhengig av å finne formuleringer som alle deltakende land kan gi sin tilslutning til, for i det hele tatt å få utført noe som helst. Det er åpenbart mange hensyn som må tas, og man har helt klart lagt ned svært mye arbeide i å oppnå konsensus rundt de sentrale problemstillinger og språklige formuleringer i rammeverket for undersøkelsen.

Det er muligens en av årsakene til at matematikkfagets og matematikkundervisningens kritiske dimensjon ikke blir eksplisitt uttalt. Jeg velger likevel å tolke ovenstående sitater i en retning som innebærer at man vedgår at matematikkundervisningen ***har*** politiske implikasjoner. Her er man altså helt på linje med flere av de matematikkdidaktikere jeg tidligere har omtalt, som eksempelvis D'Ambrosio, Abraham, Bibby, Skovsmose og Niss. De har igjen, som jeg tidligere har beskrevet, tatt opp arven etter Freire. Han var helt klar på at utdanning er betinget av sosiale, historiske og politiske faktorer og at det ikke finnes noen objektiv og universell uforanderlig kunnskap og undervisningstype. Som lærer er man også en politisk aktør.

> *We must always ask ourselves : In favor of whom and of what do we use our technical competence (Freire 1987, s. 212).*

Med støtte i de tidligere refererte matematikkdidaktikeres arbeider, vil jeg hevde at Freires metodologi for å utvikle kritisk bevissthet, også bør være relevant innenfor matematikk - didaktikken. Ved utelukkende å drive tradisjonell matematikkundervisning, støtter man opp under samfunnets gjeldende ideologi, og bidrar derfor i liten grad til å utvikle kritisk reflekterende individer. Det bør derfor være en utfordring for matematikkundervisningen å sette spørsmålstegn ved etablerte "sannheter", som blir tatt for gitt i vårt samfunn, godt understøttet av ulike medieaktører.

Dette kan gjøres ved at man i Freires ånd, og som flere av de ovenfornevnte fagdidaktikerne gir sin tilslutning til, i fellesskap med elevene drøfter ulike aspekter ved anvendt matematikk. Eksempelvis kan man sammen analysere statistiske undersøkelser, for å finne ut hvilke antagelser som implisitt ligger til grunn for dem, og om det eventuelt kan tenkes andre måter å presentere de samme dataene på.

Gjennom utvikling av mer strukturelt samfunnsrelaterte matematikkoppgaver, vil vi som matematikkdidaktikere kunne bidra til at elevene utvikler sin egen kritiske evne, og blir i stand til å sette spørsmålstegn ved overleverte, etablerte sannheter. Dermed vil de på en bedre måte være rustet:"*to deal with a very complex and rapidly changing society*" (OECD PISA 2001a, s.8).

De vil da også i større grad kunne framstå som det som i PISA's definisjon av "*mathematic literacy*", kalles "*reflective citizens*".

Hvordan man skal kunne integrere en slik type oppgaver i framtidige PISA-undersøkelser, bør være en stor utfordring for involverte matematikkdidaktikere å finne et svar på.

# 5. Organisering av rammeverket for matematikk i PISA

Rammeverket for matematikk i PISA danner grunnlaget for, og beskrivelsen av:

> *..an assessment of the extent to which 15-year-olds can handle mathematics in a well-founded manner when confronted with real world problems, or in more general terms: an assessment of how mathematically literate 15-year-olds are (OECD PISA 2001a, s.9).*

Denne vurderingen knyttes til følgende tre elementer, som jeg i dette kapitlet nærmere vil beskrive:

- matematiske kompetanser
- situasjoner og kontekster
- sentrale ideer

Fig.2 er en visuell framstilling av disse tre nøkkelkomponentene i rammeverket for matematikk.

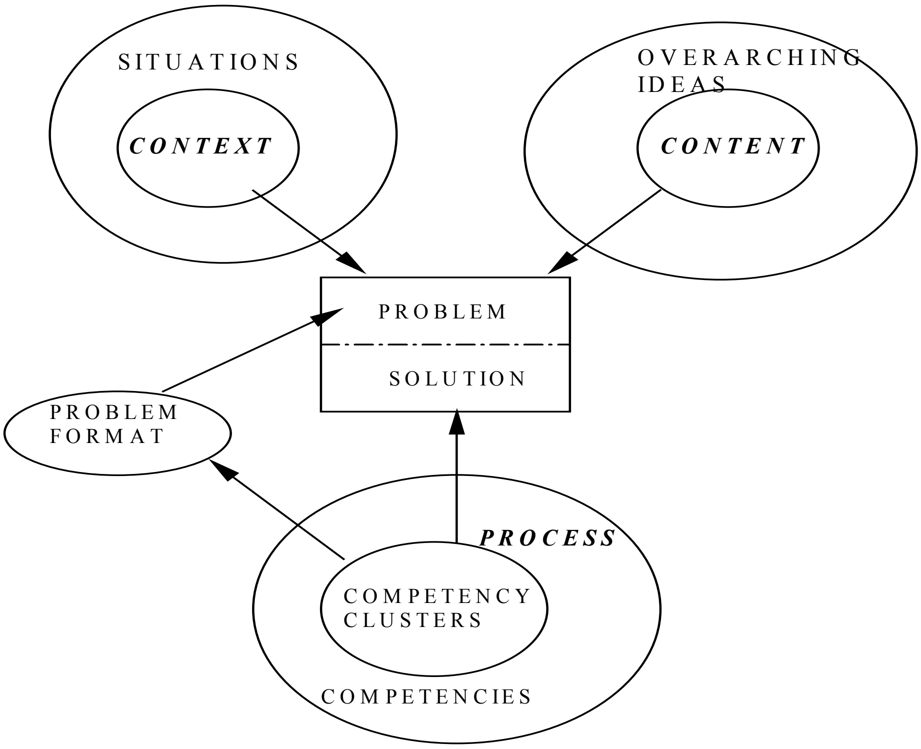



**Fig.2**

Vi ser av figuren at elementene som definerer det matematiske problemet, er knyttet til kontekster, innhold og problemformatet, mens løsningen springer ut av de matematiske kompetansene, "*the core of mathematical literacy*" (OECD PISA 2001a, s.11). I dette kapitlet vil jeg gi en forklaring av de begrepene som benyttes i denne figuren, og forholdet mellom dem.

## 5.1 Matematiske kompetanser

Det er vanlig å skille mellom matematikk som produkt og som prosess. Matematikk som produkt består da av det formelle systemet av aksiomer, teorier, modeller og formler som kan sies å utgjøre fagets byggverk.

Prosessdimensjonen i matematikk er derimot knyttet til å *utføre* de aktuelle operasjonene som kreves for å løse en oppgave, altså selve den matematiske aktiviteten.

Det er vel riktig å si at man i de senere årene er blitt stadig mer opptatt av det prosessuelle aspektet ved faget. Dette henger særlig sammen med at det konstruktivistiske læringssynet har vært det dominerende paradigme i matematikkdidaktiske fagmiljøer i dette tidsrommet. Elevenes egen konstruksjon av kunnskap er blitt ansett som det sentrale. Økt grad av elevaktivitet er nøkkelbudskapet, og det som ifølge konstruktivismen bør prioriteres i læringssituasjonen. Det er altså fagets prosessuelle dimensjon man tar til orde for å vie mest oppmerksomhet.

I PISA knytter man denne dimensjonen av faget til det å inneha matematisk kompetanse. En tenker seg da at det kreves ulik matematisk kompetanse for å løse forskjellige typer matematiske problemer. Man har valgt å skille mellom tre kompetanseklasser, slik at hver oppgave tilordnes kun en kompetanseklasse. For å være i stand til å beskrive og rapportere elevers matematiske kapasitet, har man valgt å gruppere disse kompetansene i grupper (clusters), basert på typen av kognitive krav/aktiviteter som behøves for å løse forskjellige matematiske problem (OECD PISA 2001a).

Disse kompetanseklassene avgrenses slik :

- Reproduksjonsklassen (the reproduction cluster): Reproduksjon, definisjoner og beregninger.
- Forbindelsesklassen (the connections cluster): Se forbindelser og kunne integrere informasjon som grunnlag for problemløsing.
- Refleksjonsklassen. (the reflection cluster): Matematisering, matematisk tenkning og generalisering (Ibid, s.19 – 22, min oversettelse).

***Reproduksjonsklassen*** er knyttet til elevers bruk av faktakunnskaper og standardalgoritmer. Oppgavene knyttet til denne kategorien, krever at elevene kan manipulere uttrykk som inneholder matematiske symboler og formler, og at de kan bruke disse i sine utregninger. Problemløsing av den enkleste typen hører også til her. Problemene kan være uttrykt verbalt, men konteksten er matematisk, slik at "problemet" er eksplisitt gitt. Vi ser av fig.3 på neste side, at "*routine*" er et begrep som går igjen for å beskrive kompetansene knyttet til reproduksjonsklassen.

Selv om ikke kompetanseklassene er ment å ha en hierarkisk struktur, vil man nok lett sitte igjen med inntrykk av at oppgaver tilhørende denne kategorien, er de letteste. Det blir likevel understreket at for elever som ikke er kjent med de relevante algoritmene, vil oppgaver framstå som vanskelige uansett hvilken kompetanseklasse de tilhører.

***Forbindelsesklassen*** kjennetegnes ved at elever for å kunne løse oppgaver herfra, må evne å se sammenhenger mellom ulike deler av matematikken, og kunne beherske fagets formelle språk. Det vil likevel gå relativt tydelig fram av oppgavene hva det matematiske problemet består i, elevene må altså ikke selv definere dette. Som det heter i DMF :

*The Connections Cluster of competencies  reflects students' abilities to choose and develop strategies, to choose mathematical tools, to use multiple methods or to apply multiple steps in the mathematisation process. These competencies include student`s abilities to interpret the meaning of a solution and to check the validity of their work (Ibid, s.20).*

***Refleksjonsklassen*** er på sett og vis en videreføring av den foregående. Problemtypene vil her være mer sammensatte, og kreve at eleven i tillegg til å matematisere evner å utvikle originale løsningsstrategier.

*Items measuring the Reflection Cluster of competencies should reflect students' abilities to analyse, to interpret, to reflect, to explain, and to present mathematical generalisations, arguments and proofs. Students should be able to pose problems in addition to solving problems (Ibid, s.22).*

Refleksjonsklassen skiller seg fra de foregående ved at elevene *selv* må finne fram til hva som er oppgavens matematiske problem. De må *matematisere*, altså først finne problemet, oversette det i relevante former til et matematisk problem, løse dette, for så å gå tilbake og knytte løsningen til det opprinnelige problemet. I tillegg må de kunne formulere og kommunisere de matematiske problemene, og knytte disse til relevante matematiske ideer. Oppgaver tilhørende denne klassen vil måtte betraktes som spesielt krevende, og de kognitive matematiske prosesser som er involvert er knyttet til begreper som kritisk tenkning, analyse og refleksjon. Følgende diagram benyttes i DFM for å belyse forholdet mellom de tre kompetanseklassene :

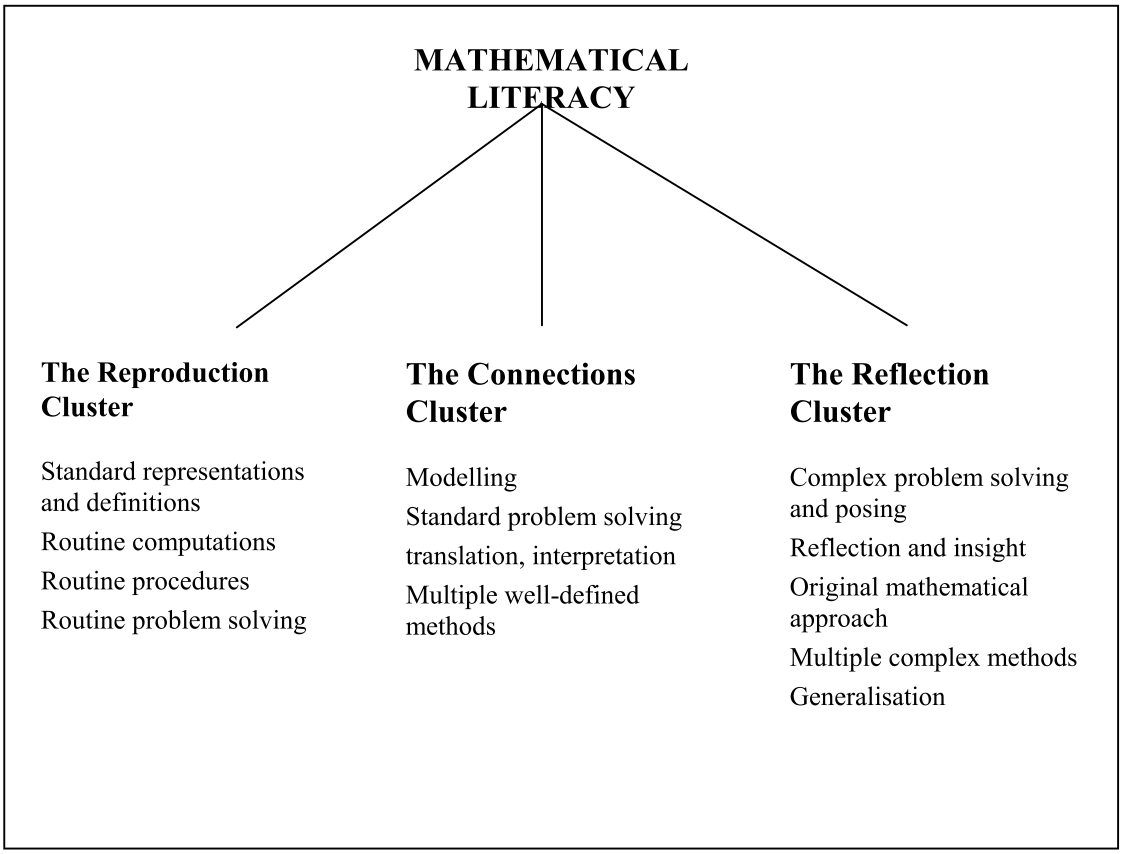



**Fig.3**

Vi ser av fig.3, at man altså definerer de tre klassene ved hjelp av ulike typer kompetanse, på en slik måte at ingen formuleringer direkte gjentas. De enkleste oppgavene å klassifisere, vil nok være de som involverer kompetanser som hører inn under reproduksjonsklassen. På tross av de ulike formuleringene av hvilke kompetanser som hører inn under de to andre klassene, kan det i praksis være noe vanskeligere å skille mellom dem. Som det uttrykkes i DMF:

> *Strategy development by students is needed for assessment items that involve competencies in both Connections and Reflection competency clusters (Ibid, s.24).*

Forskjellen mellom de to er knyttet til "*the level of complexity or originality*" (Ibid, s.25).

Det vil mellom disse to kompetanseklassene altså på sett og vis være en mer flytende overgang. Å avgjøre hvilken av dem en oppgave hører inn under, vil følgelig i noen tilfeller fortone seg problematisk, og derfor måtte bli gjenstand for nøye vurdering. Det utslagsgivende vil da være graden av, eller om man vil, kompleksiteten av den matematiseringen som kreves for å løse oppgaven.

## 5.2 Situasjoner og kontekster

Situasjoner og kontekster er også dimensjoner knyttet til "*mathematical literacy*", som man i PISA benytter i klassifikasjon av oppgaver. Valg av matematisk metode og representasjon vil være avhengig av i hvilken kontekst og i hvilken situasjon oppgaven presenteres for elevene.

Situasjonen er "*that part of the real world of the student in which the problem is situated*" (Ibid,s.33).

Situasjoner klassifiseres ut fra distansen til elevene på følgende måte :

1 Personlig liv (personal)

2 Utdannende (educational)

3 Arbeid og fritid (occupational)

4 Nabolag og lokalsamfunn (public)

5 Vitenskap (scientific)

Det vil i PISA bli lagt vekt på å ha oppgaver hentet fra alle disse situasjonene.

Den verbale formuleringen av en oppgave, eller annen type visuell ikke-matematisk informasjon, setter oppgavene inn i en spesiell kontekst. Dersom den ikke refererer til noe utenom den rene matematiske verdenen, kalles konteksten "intra-matematisk". Er imidlertid det motsatte tilfelle, det vil egentlig si i alle andre tilfeller, kalles den "ekstra-matematisk".

I PISA skiller man også mellom oppgaver med en relevant kontekst og de med kvasirelevant verbal informasjon. De førstnevnte blir selvsagt foretrukket og regnet som mest adekvate for PISA's formål: Å måle og vurdere "*mathematical literacy*".

Det man ønsker er å ha oppgaver, som i tillegg til å være hentet fra en autentisk situasjon, også har en autentisk kontekst.

> *In summary PISA values most highly tasks that could be encountered in one of a variety of real-world situations, and that have a context for which the use of mathematics to solve the problem would be authentic. Problems with extra-mathematical contexts that influence the solution and its interpretation are preferred*

*as a vehicle for assessing mathematical literacy, since these problems are most like those encountered in day to day life (Ibid, s.12).*

## 5.3 Sentrale ideer (*Overarching Ideas*)

Som vitenskap er matematikk oppdelt i relativt strengt atskilte disipliner. Dette har opp gjennom årene hatt stor innvirkning på hvordan læreplaner og lærebøker har blitt utformet.

I internasjonale matematikkundersøkelser er det også tradisjon for å anvende emne - inndelinger. Den bakenforliggende ideen er at disse vil være interessante kategorier for å undersøke forskjeller i elevers prestasjoner.

Nedenfor viser jeg eksempelvis hvilke faglige emneinndelinger man benyttet i henholdsvis SIMSS og TIMSS:

SIMSS

1 Aritmetikk

2 Algebra

3 Geometri

4 Deskriptiv statistikk

5 Måleenheter

TIMSS

1 Tallregning

2 Målinger

3 Geometri

4 Proporsjonalitet

5 Likninger og funksjoner

6 Datarepresentasjon

7 Elementær analyse

8 Begrunnelse og struktur

9 Annet

Mot slutten av det forrige århundre ble det imidlertid reist en god del kritikk mot det vitenskapssynet som ligger til grunn for en slik inndeling. Fra konstruktivistisk hold ble det påpekt at denne fagoppdelingen antydet at det eksisterte flere matematiske områder, eller "virkeligheter", som hadde lite med hverandre å gjøre. Særlig på grunnskolenivå mente man at det var viktig at elevene opplevde matematikken enhetlig, som en type prosess eller aktivitet.

Så vel internasjonalt som nasjonalt førte denne kritikken til at problemløsning ble lansert som en viktig del av matematikkundervisningen. Her var nettopp meningen at elevene skulle dra nytte av alle sine matematiske kunnskaper og ferdigheter for å finne fram til vellykkede løsningsstrategier. Problemløsning ble her i landet, gjennom M87, lansert som et eget emneområde.

I L97 har man bygget videre på dette konstruktivistisk orienterte, prosessuelle faglige synet, ved å vektlegge elevenes egenaktivitet. Man har i tillegg fokusert sterkere på matematikkens praktiske dimensjon, ved blant annet å lansere "matematikk i dagliglivet", som et eget hovedemne. Det blir understreket at :

> *Elevene skal lære å bruke sine kunnskaper i matematikk som et nyttig redskap i oppgaver og problemer i dagliglivet og samfunnslivet (L97, s.166).*

I forhold til tidligere legger man altså større vekt på at matematikkunnskaper skal kunne anvendes i det praktiske liv, i møte med dagliglivets ulike utfordringer.

I PISA forsøker en, i tråd med et slikt syn, og delvis i motsetning til tidligere internasjonale undersøkelser, å fokusere på et bredt og integrert spekter av kunnskaper, ferdigheter og holdninger. Man opererer derfor med en annen type emneinndeling enn man har gjort i tidligere tester. Dette blir i DMF begrunnet ut fra at fenomener i virkelighetens verden, som kan behandles matematisk, sjelden er knyttet til kun ett emneområde.

> *Rarely do problems arise in ways and contexts that allow their understanding and solution to be achieved through an application of knowledge from a single content strand (OECD PISA 2001a, s.13).*

Ettersom PISA har som mål å vurdere elevers kapasitet til å løse autentiske problemer, finner man det derfor riktig å bruke en fenomenologisk tilnærmingsmåte til oppgavenes innholdsside.

> *This means describing content in relation to the phenomena and the kinds of problems for which it was created (Ibid, s.13).*

På denne måten oppnår man samsvar mellom målene for undersøkelsen og oppgave - kategoriene. I stedet for emner kaller man kategoriene nå ideer, og man har samtidig lykkes i å redusere antallet til kun fire.

Disse fire "***sentrale ideene***" (*Overarching Ideas),* er definert så pass romslig at de inkluderer alle typiske emner som finnes i tidligere internasjonale undersøkelser og i nasjonale læreplaner.

De er:

1 Kvantitativt resonnement (*Quantity*).

2 Rom og form (*Space and shape*).

3 Forandring og sammenheng (*Change and relationships*).

4 Usikkerhet (*Uncertainty*).

Mens de matematiske kompetanseklassene er knyttet til oppgavenes prosessuelle aspekt, kan man si at de sentrale ideene dreier seg om oppgavenes innhold. En oppgave må imidlertid ikke nødvendigvis i sin helhet falle inn under kun en av disse sentrale ideene. Ettersom de har mange tilknytningspunkter til hverandre, ja nærmest flyter over i hverandre, ville dette være et nærmest umulig krav å forholde seg til for en oppgaveutvikler. Man ønsker likevel i PISA, at alle de fire ideene representeres i de anvendte oppgaver i et noenlunde likt antall. For å sikre en jevn fordeling, blir det derfor foretatt analyser av oppgavene, i den hensikt å klassifisere dem i forhold til "***kompetanseklasser***" og "***sentrale ideer***".

Som oppgaveinnsender kreves det også at man eksplisitt gjør det klart hvilke kategorier ens oppgaver hører inn under.

Her følger en utdyping av hva man i PISA forstår med de fire sentrale ideene :

### 5.3.1 Kvantitativt resonnement

Denne sentrale ideen er knyttet til vårt behov for å kunne forstå og bruke kvantitative beregninger i vårt daglige liv og virke. Det inkluderer :

- tallforståelse
- å forstå meningen med operasjoner på tallstørrelser
- å ha en følelse for "*the magnitude of numbers*"
- elegante utregninger
- mental aritmetikk
- evne til å utføre overslag (*estimations*)(Ibid, s.33, min oversettelse).

Slike typer ferdigheter anses som helt nødvendige for at man skal kunne holde seg orientert og delta i samfunnslivet på en adekvat måte, altså som et fullverdig samfunnsmedlem.

### 5.3.2 Rom og form

Denne sentrale ideen er relatert til vårt behov for å kunne orientere oss i rommet, og gjenkjenne ulike mønstre og konstruksjoner.

> *….we must be able to understand the properties of objects, and the relative positions of objects. We must be aware of how we see things and why we see them as we do (Ibid, s.15).*

Vi må som mennesker, for å leve et sosialt fullverdig liv, kunne orientere oss i den visuelle verden, og evne å dekodifisere og tolke den informasjonen vi mottar. De viktigste aspektene ved denne sentrale ideen er:

- å kunne gjenkjenne former og mønstre
- å kunne beskrive, kode og dekodifisere visuell informasjon
- å forstå formers dynamiske endringer
- å gjenkjenne likheter og ulikheter
- å forstå relative posisjoner
- å kunne forholde seg til to- og tre dimensjonale representasjoner og forholdet mellom dem

- å kunne navigere i rommet
  (Ibid, s.37, min oversettelse).

Denne kategorien er nært knyttet til emnet geometri, men viser altså utover dette, til forståelse av for eksempel relasjoner mellom objekter. Den har derfor nære forbindelser til en annen sentral ide, "***forandring og sammenheng***".

### 5.3.3 Forandring og sammenheng

I DMF blir det hevdet at :

> *Functional thinking, that is thinking in terms of and about relationships, is one of the most fundamental disciplinary aims of the teaching of mathematics (Ibid, s.16).*

Det blir påpekt at forandring er noe som manifesterer seg i alle naturlige fenomener, og at det er noe vi daglig blir konfrontert med gjennom de observasjoner vi gjør. Dette er tanker som har røtter helt tilbake til antikken. "Alt flyter" var den greske filosofen Heraklits berømte tese[4]. Den har inspirert mange arbeider opp gjennom historien, både av litterær og filosofisk art. At alle fenomener er i stadig endring, og at dette bør gjenspeiles i vår erkjennelse, er altså ikke helt nye ideer, men det betyr ikke at de er utdaterte. PISA skal ha æren for å ha påpekt relevansen disse tankene har for matematikken, både i oppgavesammenheng og som vurderingsgrunnlag i internasjonale komparative undersøkelser.

Matematiske forhold presenteres ofte i formen likheter eller ulikheter, men PISA's kategori er ment å vise utover dette, til også å innbefatte for eksempel ekvivalens, delbarhet, inkludering o.s.v.

Forhold kan bli representert på mange ulike måter, grafisk, symbolsk, geometrisk eller algebraisk for å nevne noen. Det er viktig at man evner å oversette det matematiske innholdet fra en representasjon til en annen. Oppgaver som får fram dette forholdet, vil være svært relevante i PISA.

### 5.3.4 Usikkerhet

Denne kategorien er knyttet til statistikk og sannsynlighet, og står kanskje særlig nært selve hovedideen i hele PISA-undersøkelsen.

Flommen av informasjon som vi daglig blir overøst med, er ofte preget av usikkerhet. Værprognoser som ikke slår til, ustabile gallupmålinger, fluktuerende aksjeindekser, tvetydige og motstridende ekspertuttalelser er eksempler på dette.

I det hele tatt kan vi vel hevde at selve kjennetegnet på vår postmoderne tilværelse er at vår tidligere autoritetstro, og vår trygge og sikre opplevelse av virkeligheten, nå er erstattet av utbredt tvil og usikkerhet innenfor mange av livets domener. Uforutsigbarhet og fragmentering er her nøkkelord. Den eneste stabile tilstanden er tilstanden av instabilitet. At elever erverver kunnskaper og ferdigheter som gjør at de bedre forstår hvordan dette ”henger sammen”, og at de kan trekke slutninger og gjøre seg opp meninger på bakgrunn av usikre

4. Arne Næss (1978) mener Heraklits utsagn heller bør oversettes med ”alt strømmer”, ”alt renner” eller ”alt er i bevegelse”, for å understreke at utsagnets mening ikke er knyttet til noe fast som beveger seg i et flytende medium. Ut fra

en slik eksegese av Heraklits skrifter, er meningen at alt er i uoppløselig forandring. Desto tettere ligger utsagnet opp til helt sentrale tanker i PISA.

empiriske data, er derfor viktig. Innsamling av data, bearbeidelse, analyse og presentasjon av funn knyttet til variasjon, er matematiske aktiviteter som står sentralt innenfor denne kategorien.

PISA vil, gjennom sin vektlegging av usikkerhet som en sentral ide, bidra til at denne type matematikk blir fokusert og prioritert i samtlige deltakerland.

### *5.3.5 Forholdet mellom de fire sentrale ideene*

De fire sentrale ideene er avgrenset på en måte som gjør at de skal være nyttige som kategorier i forhold til å forstå og analysere virkelige fenomener, og vår begrepsbruk knyttet til disse.

Det er selvsagt umulig å inndele virkelighetens fenomener i skarpt avgrensede kategorier, til det er verden for komplisert og sammensatt. Grensen mellom ideene i PISA vil på tilsvarende måte være flytende og gå over i hverandre. Man kan si at :

> *..each of them represents a certain perspective, or point of view, which can be thought of as possessing a core, a centre of gravity, and somewhat blurred outskirts that allow for intersection with other overarching ideas (Ibid, s.14).*

# 6. Retningslinjer for utvikling av matematikkoppgaver til PISA

Oppgavene som utvikles for å bli brukt i PISA må tilfredsstille en rekke krav.

Disse konkrete kravene er tilkjennegitt i heftet "***Mathematics Item Development for PISA 2003 and Item Submission Guidelines***", heretter forkortet til MID (OECD PISA 2000).

Her følger en gjennomgang av noen av de viktigste punktene der.

Innledningsvis blir det slått fast at oppgavene bør være tilpasset 15 åringer, og de må være originale, dvs. ikke hentet fra for eksempel, tidligere utgitte oppgavehefter.

Videre heter det at dersom en oppgave inneholder flere spørsmål, bør det være slik at det å besvare ett spørsmål, ikke forutsetter at man har maktet å besvare det foregående.

Oppgavene bør bestå av et innledende stimulus materiale/informasjon, etterfulgt av tilknyttede relevante spørsmål. Autentiske kontekster er også ønskelig, noe som er en direkte følge av ambisjonene om å måle "*mathematical literacy*".

Det forlanges videre at de innsendte oppgaver fra deltakerlandene, er testet i begrensede pilotundersøkelser på lokalt nivå, og at det foreligger statistisk bearbeidede data av svarene. Disse bør så brukes til å forbedre oppgavene, og til å utarbeide en rettemal.

Oppgavene bør kunne løses kun ved hjelp av det utstyret elevene normalt er i besittelse av, slik som linjal, passer, gradskive og kalkulator.

Et viktig og kanskje noe problematisk krav, er videre formulert på denne måten :

> *The level of reading required to successfully engage with an item should be considered very carefully. The wording of items should be as simple and direct as possible (OECD PISA 2000, s.12).*

Dette blir i den samme paragrafen begrunnet slik :

> *This is an assessment of mathematical literacy, not of reading ability. In addition, high demands in the formulation of items can have a differential impact on boys compared with girls at least in some countries (Ibid, s.12).*

Det er ikke vanskelig å ha sympati med et slikt krav og dets begrunnelse, men samtidig er det jo slik at målet i PISA er å undersøke elevenes funksjonelle matematikkunnskaper, og om de evner å nyttiggjøre seg disse i sitt daglige liv. For å kunne sies å være i besittelse av denne type kunnskaper, må man kunne håndtere og desiffrere den informasjonsflommen man utsettes for. Virkeligheten er dessverre ikke slik ordnet at matematikk og språk kommer i adskilte "pakker", men kjennetegnes tvert imot av at tekst og matematiske begreper er vevd tett sammen. Eksempelvis vil presentasjoner av statistikk og diagrammer i media alltid ledsages av en viss mengde verbale utsagn. Ettersom oppgavene i PISA helst skal knyttes til autentiske kontekster, kan det synes svært vanskelig å finne en enkel løsning på dilemmaet knyttet til hvor mye tekst de kan og bør inneholde.

Oppgavene må ikke ha noen kulturelle skjevheter (*bias*), ettersom de jo skal benyttes i et stort antall land, med ulike kulturelle og geografiske forutsetninger.

Oppgaver som krever svært komplekse rette-skjemaer bør unngås. Dette blir begrunnet med deltakerlandenes begrensede ressurser, og faren for en lavere sensorreliabilitet[5].

Et kanskje helt åpenbart krav er at oppgavene bør være så klare som mulig, slik at misforståelser unngås.

I en oppgave som består av flere delspørsmål, bør det være slik at den nødvendige informasjonen for å kunne besvare dem, befinner seg i umiddelbar forkant av det aktuelle spørsmålet.

Omtrent en tredjedel av oppgavene skal bestå av flervalgsoppgaver, en tredjedel av åpne oppgaver, og den siste tredjedelen av oppgaver som krever et kort, gitt svar, dvs. at de kan rettes av en "*data entry operator without expert judgement, and scored automatically*" (Ibid, s.11).

Gjennom arbeidet med å utvikle oppgaver var det viktig å ta hensyn til alle disse ulike kravene. Dette førte selvsagt til en god del omformuleringer av tekster og modifiseringer av spørsmålsstillinger underveis.

## 6.1 Oppgavetyper.

Oppgavene i PISA presenteres i egne hefter, som de deltakende elever får utdelt. Der gis innledningsvis en instruksjon i hva som forventes av eleven, og hvordan han/hun skal besvare de ulike oppgavetypene, som finnes i heftet. Oppgavene har disse formatene :

### *6.1.1 Flervalgsoppgaver (multiple choice items)*

Disse består igjen av to typer. I den ene presenteres først et stimulusmateriale og en oppgavetekst, etterfulgt av flere ulike svaralternativer, som regel fire eller fem. Eleven skal så sette ring rundt det svaret som han/hun mener er riktig. De uriktige svaralternativene kalles *distraktorer*. Når man analyserer elevsvarene er det viktig å undersøke hvordan de aktuelle distraktorene i oppgaven fungerer, og hvordan elevene sansynligvis benytter seg av dem i løsningsprosessen. Slik kan man oppnå økt innsikt i elevenes tenkemåter og kunnskaper, altså få verdifull diagnostisk informasjon[6] (Olsen, Turmo & Lie, 2001).

Den andre typen flervalgsoppgaver består av ulike påstander om et saksforhold. Eleven blir så bedt om å avgjøre om disse er riktig eller gale, ved å krysse av på alternativene ja/nei, eller riktig/galt.

En stor styrke ved flervalgsoppgaver er at de øker testens reliabilitet[7], både ved at det blir plass og tid til flere oppgaver, og ved at sensorreliabiliteten øker i forhold til om testen kun skulle inneholde åpne oppgaver.

5. ***Sensorreliabilitet*** vil si i hvor stor grad det er samsvar mellom to eller flere sensorer som retter en test, altså hvorvidt sensorenes vurderinger er like (Angell 1996).

6. Se kap.8.1 for forklaring av ***diagnostisk informasjon***.

7. ***Reliabilitet*** omtales i metode-teori kapitlet.

### 6.1.2 Oppgaver med korte, gitte svar (closed constructed response items)

Denne oppgavetypen kjennetegnes ved at eleven selv skal formulere sitt svar på en standardisert måte. Det kan da lett fastslås om svaret er rett eller galt, og dette kan derfor gjøres direkte av dataoperatøren, uten at matematikkeksperter er involvert.

Fordelen med å vurdere oppgaver med korte, gitte svar i forhold til flervalgsoppgaver, er at man unngår at eleven lykkes i å gjette riktig svar, og at man ikke trenger å bekymre seg for distraktorers eventuelle uheldige effekt (OECD PISA 2001a).

### 6.1.3 Åpne oppgaver (open constructed response items)

Her må elevene selv formulere svarene sine. I noen oppgaver vil det være korrekt å svare relativt kort, mens andre krever fyldigere svar med begrunnelser, utregninger, forklaringer og lignende. Av den engelske ordlyden, ser vi at forskjellen mellom denne type oppgaver og den forrige, er knyttet til distinksjonen *open response/closed response.* Selv om det også for oppgaver tilhørende denne kategorien vil være slik at et spesielt svar er det riktige, vil elevene kunne komme fram til dette svaret på ulike måter. Derav betegnelsen *open response*. Oppgavene vil generelt kreve en mer kompleks type *matematisering* enn de som tilhører den forrige kategorien.

Hensikten med mye av testmaterialet i PISA er å undersøke om elevene evner å reflektere og tenke aktivt rundt et tema, og om de er i stand til å kommunisere sin viten. For dette formålet egner åpne oppgaver seg bedre enn de ovenfornevnte (Olsen et.al.2001). Åpne oppgaver har sin store styrke i at de kan formuleres slik at de stimulerer elevene til å integrere kunnskaper og ferdigheter av forskjellig kompleksitet, fra ulike områder av faget (OECD PISA 2001a).

> *The key feature of open constructed response items is that they allow students to demonstrate their abilities by providing solutions at a range of levels of mathematical complexity (OECD PISA 2001a, s.27).*

Ulempen med åpne oppgaver i store internasjonale komparative undersøkelser, er blant annet at de er svært arbeidskrevende. Det må investeres mye tid og anstrengelser i å utarbeide kodeskjemaer og oversettelser, som er så entydige, at man oppnår en akseptabel sensorreliabilitet.

I PISA ønsker man å bruke alle disse tre oppgavetypene:

> *PISA will assess mathematical literacy through a combination of items with open-constructed response formats, closed-constructed response formats and multiple-choice formats. About equal numbers of each of these item-formats will be used in constructing the test instruments for PISA 2003 (Ibid, s.26).*

## 6.2 Utvelgelse av oppgaver til PISA 2003

Den endelige utvelgelsen av oppgaver til PISA 2003, er en arbeidskrevende prosess, hvor det blir lagt stor vekt både på detaljer og helhet. Min utlegning baserer seg hovedsakelig på dokumentet "***Mathematics Item Development and Selection***", heretter kalt MIS (OECD PISA 2001b).

Alle innsendte oppgaver fra deltakerlandene har vært gjennom en omfattende og rigorøs utvelgelsesprosess. Innledningsvis ble de forelagt et ekspertutvalg, som på grunnlag av oppgavenes "*face validity*" (se kap.7) enten avviste dem, eller godkjente dem for videre bearbeiding. Minst to uavhengige eksperter foretok denne vurderingen. Oppgavene ble så gruppert i "*item bundles*", for gjennomgang av eksperter tilknyttet nasjonale prosjekt - komiteer. Samtidig ble det utført begrensede pilotundersøkelser. I tillegg til dette ble oppgavene vurdert av et eget matematikkforum i PISA, og av deres "*Mathematics Expert Group*". Etter tilbakemelding fra disse instansene, og utfra piloteringen, ble oppgavene så revidert, eventuelt refusert. De gjenstående oppgaver ble nok en gang nøye evaluert av nasjonale prosjektkomiteer. Denne gang ble man bedt om å formidle sine synspunkter gjennom utfylling av spesielle vurderingsskjemaer, med sentralt angitte vurderingskriterier. Tilbakemeldingen som PISA mottok, fikk stor innflytelse på utvelgelsen/avvisningen av oppgaver på dette stadiet i prosessen.

Av en samling på i overkant av 400 oppgaver, plukket man nå ut 260 til å være med på generalprøven, våren 2002. Disse skal senere reduseres til det ønskede antall, ca.110, blant annet utfra statistisk bearbeiding av resultatene fra denne generalprøven, tilbakemeldinger fra deltakerlandene m.m. Til hovedundersøkelsen PISA 2003 ønsker man 120 - 130 oppgaver. Tyve av disse vil være oppgaver som er beholdt fra PISA 2000, utfra muligheten for testsammenlignende undersøkelser.

Den endelige avgjørelsen om hvilke oppgaver som skal være med i 2003 vil bli bestemt innen oktober 2002, og vil skje etter følgende kriterier:

- oppgaven må passe inn i rammeverket
- oppgaven må være av høyest mulig kvalitet
- oppgaven bør reflektere en blanding av typer og deltakerland, slik at faren for mulige skjevheter (*bias*) minimaliseres
- oppgavene bør variere bredt i vanskelighetsgrad (se under p-verdier i metodekapitlet)
- oppgavene skal representere den ønskelige blanding av matematisk innhold, matematiske prosesser, situasjoner og formater, utfra følgende tabell: (utgangspunktet er da 120 oppgaver)

| Oppgavetype | Kompetansegruppe | | | |
|---|---|---|---|---|
| | Reproduksjon | Forbindelse | Refleksjon | **Totalt** |
| Flervalgsoppgave | Ca.15 | Ca.20 | Ca. 5 | **Ca.40** |
| Korte, gitte svar | Ca.10 | Ca.25 | Ca. 5 | **Ca.40** |
| Åpen oppgave | Ca. 5 | Ca.15 | Ca.20 | **Ca.40** |
| **Totalt** | **Ca.30** | **Ca.60** | **Ca.30** | |

Som man forstår kan det være en rekke ulike grunner til at en oppgave blir refusert. En avvisning av en oppgave, særlig på et senere stadium i utvelgelsesprosessen, betyr derfor ikke nødvendigvis at det i PISA's øyne er en matematisk uinteressant eller dårlig oppgave. Det kan for eksempel skyldes at kompetansegruppa, eller formatet oppgaven tilhører, er tilfredsstillende dekket, og at man derfor prioriterer en annen type oppgave, selv om den sistnevnte, utfra andre utvalgskriterier, kan være en mindre god oppgave.

# 7. Metodeteori

Jeg vil i dette kapitlet gi en kort presentasjon av visse testteoretiske og statistiske begreper jeg benytter i min hovedoppgave. Ettersom validitetsbetraktninger er nært knyttet til mine forskningsspørsmål, vil drøftingen av dette begrepet bli viet størst oppmerksomhet, og følgelig få en dominerende plass i denne framstillingen.

For en bredere dekning av emnet, se Crocker og Algina (1986).

## 7.1 Testteori

I alle store kvantitative undersøkelser vil man ønske å foreta statistiske analyser av det innsamlede materialet. Statistikk og testteori vil også kunne være relevant på tidlige og forberedende stadier i undersøkelsen. I PISA vil det for eksempel være slik at statistiske resultater fra pilotundersøkelser og generalprøven, vil ha stor betydning ved den endelige utvelgelse av testoppgaver, til gjennomføring av hovedundersøkelsen.

**Testteorien** er historisk knyttet til faget psykologi og dets utvikling mot akademisk disiplin. Det sentrale innenfor testteori er å drøfte hvordan man kan måle psykologiske størrelser**,** og hvilke metoder man kan benytte for best mulig å lykkes med dette (Angell 1996).

Ettersom måling av psykologiske størrelser altså er det essensielle innenfor testteori, omtales den ofte som **psykometri**. Videre kaller man slike psykologiske størrelser for **konstrukter** (Ibid).

Målet med PISA - undersøkelsen er å finne ut i hvilken grad 15 - åringer er ”*mathematic literate*”. Oppgavene som blir plukket ut til å være med i den endelige testen, bør derfor alle bidra til å måle konstruktet ”*mathematical literacy*”.

I mitt arbeid med å utvikle matematikkoppgaver, var følgelig **”***mathematic literacy*”, det psykologiske konstruktet jeg måtte forholde meg til. Det er derfor viktig å avklare hva som ligger i dette konstruktet. I tillegg til å foreta en kvalitativ analyse av konstruktet, som jeg gjorde i kapittel 5, vil jeg gjennom en kvantitativ analyse av elevresultatene i min pilotundersøkelse, drøfte hvorvidt de matematikkoppgavene jeg har utviklet egner seg til å måle dette konstruktet. Jeg vil altså benytte både kvalitative og kvantitative metoder for å forsøke å gi et svar på mine innledende forskningsspørsmål.

## 7.2 Statistiske begreper

Testteori og statistikk er knyttet sammen ved at man i testteorien benytter begreper og modeller hentet fra det statistiske fagområdet. Jeg vil derfor kort redegjøre for de viktigste statistiske begrepene jeg benytter.

| Oppgave | P - verdi | Diskriminering (Pointbiserials) | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 1/2 | 0,4 | 0,69 | 16 | -0,57 | 0,22 | 0,92 | -0,98 |

Jeg har her laget en tabell hvor jeg benytter en del statistiske begreper som er svært relevante med tanke på en drøfting av hvorvidt en matematikkoppgave egner seg til å måle elevers prestasjoner, altså dens psykometriske egnethet. Denne type tabell vil jeg bruke i analysen av alle oppgavene.

**P- verdiene** viser hvor stor andel av elevene som har svart riktig på en oppgave, og ligger i intervallet 0,00 – 1,00. For oppgaver hvor elevene enten får 0 eller 1 poeng er det da slik at dersom 40% har svart riktig, blir p-verdien 0,4. Noen av oppgavene mine vurderes med 0, 1 og 2 poeng, det vil si ikke riktig/ikke svart, delvis riktig og fullstendig. Man finner da oppgavens p-verdi ved å addere prosenten av elevene som har svart fullstendig med prosenten av de som har svart delvis riktig, dividert på 2, for så å dividere denne summen med 100. Dersom 20 % har svart fullstendig og 20 % delvis riktig, får vi altså en p-verdi på 0,3.

Ved utvelgelse av testoppgaver vil p-verdier framkommet i pilotering være et av kriteriene det legges vekt på. Man vil i ordinære rangeringstester fortrinnsvis komponere testen slik, at hovedtyngden av oppgaver har p-verdier innenfor intervallet 0,3-0,7. Dette må ikke oppfattes som absolutte grenseverdier, men opererer man med mange testoppgaver utenfor dette "omtrentlige"intervallet, vil disse bli betraktet som "dyre" oppgaver i forhold til de stramme tidsrammer man som regel har. Årsaken er at de gir begrenset informasjon, ved å diskriminere enten kun mellom de svake elevene, eller kun mellom de sterke. Ettersom man imidlertid ofte også vil ønske å kunne diskriminere mellom elever tilhørende disse gruppae, vil man sørge for å ha enkelte testoppgaver med p-verdier utenfor det omtalte intervall.

Som en følge av at man i PISA undersøkelsen i hvert deltakerland opererer med store utvalg av elever, kan man her tillate seg å inkludere en god del såkalt "dyre" testoppgaver. Hovedtyngden av oppgaver vil imidlertid også i PISA ha p-verdier som ligger innenfor det ovenfornevnte intervallet.

**"Pointbiserial"** viser korrelasjonen mellom score på den enkelte oppgaven og totalscore, og beregnes ved hjelp av Pearson Product Moment Correlation Coefficient. Den gir et mål på graden av lineær sammenheng mellom to variabler, og ligger mellom –1,00 og 1,00 (Angell 1996).

Pointbiserial er det statistiske faguttrykket man benytter når den ene variabelen er dikotom, dvs. kun har verdiene 0 eller 1, eller kan tilordnes disse. For en del av mine oppgaver er det slik at elevene kan oppnå 0, 1 eller 2 poeng. Jeg har derfor valgt å kalle den beregnede korrelasjonskoeffisienten med totalscore for *diskriminering.* For at en oppgave skal fungere godt psykometrisk, kreves det at den diskriminerer tydelig mellom svake og sterke elever, vurdert ut fra hele testen. De som svarer riktig på oppgaven, bør altså ha en høyere totalscore enn de som svarer galt. Dette vil medføre at vi får en positiv verdi for *diskriminering*. Det er ikke slik at vi har noen absolutt grense som *diskrimineringsverdien* må overskride, men blir den for lav, for eksempel under 0,2, bør man enten omarbeide oppgaven, og foreta en ny pilotutprøving av den, eller simpelthen utelate den fra testen.

Det er viktig å være oppmerksom på at man for å kunne bruke ***p-verdier*** eller ***diskriminering*** til å velge ut oppgaver til en framtidig test, forutsetter at framtidens elever befinner seg på samme faglige nivå som de elever oppgavene er pilotert på (Kleve 1994). Oppgavers diskrimineringsevne vil nemlig være relativ til elevgruppas sammensetning. En oppgave som diskriminerer godt i en sterk elevgruppe, kan selvsagt diskriminere dårlig i en svak og vice versa. Man ønsker derfor alltid å pilotere oppgaver på et bredt, tilnærmet tilfeldig utvalg av elever. Økonomiske rammer kan gjøre dette vanskelig i praksis, men et noenlunde representativt utvalg av elever vil man alltid tilstrebe å oppnå.

Generalprøven i PISA,våren 2002, ble for eksempel her til lands av praktiske og økonomiske grunner, kun foretatt i østlandsområdet. Ved den endelige undersøkelsen i 2003, vil imidlertid utvelgelsen av deltakerskoler bli foretatt etter strenge kriterier, slik at utvalget blir tilfeldig. Dette er nødvendig for at man skal kunne trekke slutninger, generalisere eventuelle funn fra utvalget, til en større populasjon (Moore & McCabe 1999, Kap.3.3).

**Z-score** er en verdi som viser hvor mange standardavvik[8] over eller under gjennomsnittet en elevs score ligger. I tabellen over vil det si at de som får 0 poeng på dette spørsmålet, har en totalscore som ligger 0,57 standardavvik under gjennomsnittet, mens de som har fått to poeng har en totalscore som ligger 0,92 standardavvik over gjennomsnittet.

Multipliserer man z-score med antall elever som hører inn under vedkommende svar – kategori, for deretter å summere disse produktene, får man svaret 0. Det vil med andre ord si at den gjennomsnittlige z-score for hele utvalget er 0.

For at en oppgave skal fungere godt psykometrisk, må det være slik at z-score for de som svarer riktig på oppgaven, er høyere enn for de som svarer galt. Likeledes må z-score for de som svarer fullstendig, være høyere enn for de som bare svarer delvis riktig.

Vi ser altså at både ***p-verdi, diskriminering*** og ***z-score*** beregnet ut fra elevsvarene på de ulike oppgavene, gir oss en indikasjon på om oppgavene psykometrisk sett er gode.

I PISA 2003, ønsker man ikke kun å rapportere resultatene ved hjelp av måleskalaer, men også:

> *…å tilordne slike skalaer meningsfulle verbale beskrivelser, altså å forankre tallene på skalaen til meningsfulle utsagn (Olsen 2000, s.1).*

Utfra de oppgavene elevene evner å løse, har man som mål å kunne gi en beskrivelse av hvilke kompetanser de innehar. Motivet for å kommunisere funn ved hjelp av verbale beskrivelser, er at man slik antar det er mulig å unngå medienes og politikernes fokusering på internasjonale ”ligatabeller” ved offentliggjøringen av forskningsresultatene. Intensjonen med studiene går jo langt utover dette, til å kunne si noe om hvilke faktorer som ser ut til å fremme bedre læring. Å forankre skalaene i ord, vil derfor være et forsøk fra forskernes side på å ta større ansvar for konsekvensene av forskningen, ved å forbedre den eksterne kommunikasjonen (Ibid). Om man vil lykkes med dette ambisiøse rapporteringsprosjektet for PISA 2003 gjenstår å se.

8) **Standardavvik** er, i likhet med **varians,** et mye benyttet spredningsmål i kvantitative undersøkelser. Man finner variansen ved å kvadrere hver elevs avviksscore fra middelverdien, for så å beregne gjennomsnittet av disse kvadratiske avvikene. Standardavviket er kvadratroten av variansen (Angell 1996).

### 7.2.1 Reliabilitet

**Reliabilitet**t er knyttet til hvorvidt en elevs totale score ville bli den samme om man gjentok testen under helt like forhold, eller benyttet en parallell, alternativ test. Begrepet er altså knyttet til ønske om at resultatene fra en test skal være reproduserbare og konsistente (Kleve 1994).

En **reliabilitetskoeffisient** defineres som korrelasjonen mellom score på to parallelle tester. Den kanskje mest brukte metoden for å uttrykke denne, er å benytte **Cronbachs alfa.** Ved hjelp av den, kan vi beregne reliabiliteten til et sett av oppgaver gjennom reliabilitets - koeffisienten mellom enkeltoppgavene testen består av (Angell 1996).

> *Det betyr at vi betrakter Cronbachs alfa som den gjennomsnittelige korrelasjonen mellom en bestemt test og alle mulige tester med samme antall oppgaver i et univers av tester (Ibid, s.85).*

I praksis vil det være slik at man ikke gir parallelle tester, men at man gjennom kun <u>en</u> administrasjon av en test ønsker å estimere et bestemt konstrukt. Dette vil for eksempel være tilfelle i PISA. For å måle hvorvidt elever da svarer konsistent på en test, og dermed gir oss grunner til å anta at vi kan generalisere våre funn, trekke generelle slutninger ut fra gjennomføringen av kun den ene testen, kan man dele testen i to, for deretter å finne korrelasjonen mellom de to halvdelene. Dette kalles "*split half*"- metode. Ved hjelp av denne korrelasjonen, korrigert for antall oppgaver, vil man da få et estimat av testens reliabilitet. Cronbachs alfa kan betraktes som gjennomsnittet av alle mulige "*split half*"- koeffisienter (Ibid).

Selv om ikke hensikten med min utvikling av oppgaver var å komponere en matematikktest som kunne stå på egne ben, men å prøve ut enkeltoppgaver, vil jeg likevel benytte Cronbachs alfa i min analyse. Dette vil jeg gjøre for å undersøke hvorvidt oppgavene mine måler noen av de samme trekkene, altså for å få et mål på "testens" indre konsistens.

Det er flere faktorer man bør være oppmerksom på ved tolkning av Cronbachs alfa. De mest relevante i tilknytning til mitt materiale er følgende :

- Reliabiliteten er avhengig av antall oppgaver på en test.
- Cronbachs alfa forteller oss hvor høy reliabilitetskoeffisienten minst er, ikke hvor høy den egentlig er.
- Testtiden spiller en rolle for reliabiliteten. Dersom for eksempel de svakere elevene ikke blir ferdig med alle oppgavene, medfører dette at reliabiliteten øker, ettersom prestasjonen på de ubesvarte oppgavene da blir helt konsistent.
- Cronbachs alfa måler ikke endimensjonalitet. Ettersom den er knyttet til korrelasjon mellom oppgavene, og fordi denne korrelasjonen kan skyldes mer enn <u>en</u> underliggende faktor, kan vi ikke ut fra en høy alfa slutte oss til at testen måler vårt ene utvalgte og veldefinerte konstrukt.
- En høy alfa er en nødvendig, men ikke en tilstrekkelig forutsetning for at en test måler ett eneste underliggende konstrukt. (Ibid, s.86-87)

Det vil også være andre faktorer som influerer på reliabiliteten til en test, blant annet sensorreliabiliteten. Den er et uttrykk for i hvilken grad det er samsvar mellom to eller flere sensorer som retter en test. Oppgavetypene testen består av er med på å avgjøre hvor høy sensorreliabiliteten blir. Flervalgsoppgaver gir, som nevnt i kap.5.3.1, en langt høyere sensorreliabilitet enn åpne oppgaver. Dette er en av grunnene til at man i store internasjonale undersøkelser ofte foretrekker hovedsakelig å bruke de førstnevnte.

I PISA benytter man, som vi har sett, i tillegg til flervalgsoppgaver også åpne oppgaver, og oppgaver som krever et kort, gitt svar (se kap.6.1.2). Gjennom å arrangere treningssamlinger knyttet til rettingen av oppgavene i undersøkelsen, og hvor representanter fra alle deltakerland er med, forsikrer man seg om at sensorreliabiliteten blir optimal. Man gjennomfører også reliabilitetsstudier for å undersøke hvorvidt sensorreliabiliteten er så høy som ønskelig.

> *Because of the potential for disagreement between markers of these items, PISA will implement marker reliability studies to monitor the extent of disagreement (OECD PISA 2001a, s.27).*

I tillegg forsøker man å utvikle kodeskjemaer som er så entydige som mulig, og som blir nøye gjennomgått i forkant av sensureringen.

### 7.2.2 Validitet

Selv om jeg skulle finne at testen jeg gjennomførte hadde høy reliabilitet, og at oppgavene følgelig måler noen av de samme trekkene, betyr ikke dette at de nødvendigvis måler det psykologiske konstruktet jeg ønsker at de skal måle, ”*mathematic literacy*”. Mens reliabilitet uttrykker hvor nøyaktig vi måler det vi måler, er **validitet** knyttet til hvorvidt en test måler det den gir seg ut for å måle. Validitet refererer derfor alltid til noe utenfor selve testen, for eksempel til mål i en læreplan, praktiske ferdigheter, eller psykologiske konstrukter som ”*mathematical literacy*” (Angell 1996).

Jeg vil med en gang understreke at en test ikke <u>har</u> en viss validitet, men at denne validiteten er knyttet til en spesiell bruk av testen innenfor en spesiell sammenheng (Satterly 1989).

Det er også viktig å påpeke at validitet først og fremst har å gjøre med tolkning av resultatene fra testen. Validitet kan beskrives som prosessen knyttet til å samle evidens til støtte for de slutninger man trekker fra resultatscoren til en test (Crocker & Algina 86).

Også andre teoretikere understreker at validitet primært har å gjøre med meningen av test – score, og måten vi bruker denne testscoren til å ta avgjørelser (Ebel, Frisbie 1991). Ebel og Frisbie hevder videre at det derfor nettopp er brukerne av testscoren som må begrunne og rettferdiggjøre sine slutninger og: ”*the appropriateness of their use of the scores*” (Ebel, Frisbie 1991, s.102).

For å kunne lykkes med dette, må man være i stand til å produsere ”bevismateriale” som støtter opp under de slutninger man har trukket.

For at den akademiske diskurs ikke skal bli for spissfindig og uhåndterlig, er det likevel aksept for å tale om for eksempel *oppgavers validitet* i forhold til ulike kriterier, selv om altså dette er en talemåte som strengt tatt ikke er helt korrekt. Jeg har i min hovedoppgave valgt å følge en slik linje, og har derfor blant annet beholdt den opprinnelige formuleringen av mine forskningsspørsmål. Under drøftingene av oppgavene i tilknytning til elevsvarene, vil jeg også, i overensstemmelse med det ovenfornevnte, benytte uttrykk som *oppgavenes validitet.*

Gronlund summerer opp de viktigste aspekter ved validitetsbegrepet i følgende punkter:

1. *Validity is inferred from available evidence (not measured)*
2. *Validity depends on many different types of evidence*
3. *Validity is expressed by degree (high, moderate, low)*
4. *Validity is specific to a particular use*
5. *Validity refers to the inferences drawn, not the instrument*
6. *Validity is a unitary concept*
7. *Validity is concerned with the consequences of using the assessment*

*(Gronlund 1998, s.201).*

Mens reliabilitet kan måles eksakt ved hjelp av kvantitative, statistiske metoder, gjennom for eksempel beregning av Cronbachs alfa for en test, er altså validitetsbetraktninger langt mer kompliserte, og forbundet med større usikkerhet. I tillegg til å finne støtte for validitetskrav gjennom å referere til statistiske beregninger, vil man også som man ser måtte foreta mer kvalitative vurderinger.

Validitet kan drøftes ut fra ulike analytiske oppdelinger av begrepet. Jeg har valgt å benytte en tredeling som er relativt utbredt, og vil argumentere for validitet ut fra tre forskjellige aspekter av begrepet, nemlig: **Innholdsrelatert validitet**, **kriterierelatert validitet** og **konstruktrelatert validitet**. Dette må ikke anses som tre kvalitativt ulike typer validitet, men heller forstås som ulike typer validitets – evidens.

> *Validity is a single, unitary concept that is based on various forms of evidence, content – related evidence, criterion related evidence, construct – related evidence (Ibid, s.201).*

**Innholdsrelatert validitet** er knyttet til i hvilken grad det er mulig å trekke slutninger fra resultatene framkommet i en test, til emneområdet som oppgavene i testen er ment å dekke, og som de i prinsippet skal være et tilfeldig utvalg fra. Dersom man gir en matematikkprøve i emnet ”algebra” til en 8. klasse, og ønsker at resultatene fra prøven skal gi en indikasjon på elevenes kunnskapsnivå i algebra, må man sørge for at oppgavene til sammen dekker dette emnet, og at de nettopp kan sies å representere et tilfeldig utvalg av algebraoppgaver for dette trinnet. De nærmeste til å avgjøre om dette er tilfellet, er ”eksperter” på området. Dette vil i denne sammenheng si matematikklærere på 8. trinn, eventuelt læreplan - utviklere i matematikk.

Slike vurderinger vil ikke være av samme objektive karakter som beregning av reliabilitet, men ha innslag av subjektivt skjønn. Dette henger sammen med at man her er inne på et område knyttet til ”mening”, og hvor vurderingene man gjør er av mer kvalitativ karakter. Man må da på ulike måter finne gode grunner for de valg man gjør og de slutninger man trekker, og argumentere for at disse valgene/slutningene er berettigede og fornuftige med tanke på det man ønsker å undersøke. For å redusere den subjektive faktoren, ville det for eksempel synes rimelig, at man i algebraeksempelet over lot flere ”eksperter” vurdere om oppgavene var representative i forhold til det innholdet de var ment å dekke. Oppnåelse av intersubjektiv enighet vil i en slik sammenheng styrke påberopelsen av høy grad av *innholdsrelatert validitet*.

Som jeg gjentatte ganger har påpekt, har det aldri vært hensikten at de oppgavene jeg har utviklet tilsammen skal utgjøre en helhetlig test. Det kan derfor synes noe søkt at jeg her innfører begrepet *innholdsrelatert validitet,* med tanke på senere å bringe det inn i analysen av oppgavene. Jeg har valgt å gjøre dette utfra en argumentasjon om at mine oppgaver kan relateres til rammeverket i PISA, og at de dekker aspekter ved de sentrale ideer som der er beskrevet. Jeg skylder da å gjøre oppmerksom på at mine oppgaver selvsagt ikke fullstendig dekker de omtalte sentrale ideer, og heller ikke kan anses å utgjøre et slags representativt utvalg i forhold til disse. Min bruk av begrepet *innholdsrelatert validitet* vil defor kunne oppfattes som noe ukonvensjonell, men må altså betraktes på bakgrunn av en slik presisering.

**Face – validity**[9] er et validitetsbegrep som er beslektet med innholdsvaliditet. *Face – validity* er knyttet til den umiddelbare opplevelsen av validitet man opplever at en test har.

**Kriterierelatert validitet** er forholdet mellom testscore og et ytre kriterium. Noen ganger vil det være mulig å beregne korrelasjonen mellom dette ytre kriteriet og den testscore man har fått. Denne målbare størrelsen blir kalt for **validitetskoeffisienten,** og brukes som støtte for de slutninger man trekker av sitt tallmateriale.

Man skiller mellom to former for kriterierelatert validitet, nemlig **prediktiv validitet** og **samtidig validitet.**

**Prediktiv validitet** er knyttet til i hvilken grad testscoren kan predikere en framtidig prestasjon. Dette vil man spesielt være interessert i ved opptaksprøver, ved tester i forbindelse med jobbansettelser og lignende. I PISA - undersøkelsen er denne type validitet særlig aktuell. En av hovedmålsetningene der er nemlig at man gjennom tester, bestående av godt gjennomarbeidede og piloterte oppgaver, utfra elevenes testscore skal kunne si noe om i hvilken grad de i framtiden vil framstå som samfunnsnyttige og fullverdige borgere. Man ønsker altså at testresultatene skal ha en *prediktiv validitet*. På bakgrunn av dette, planlegger man i PISA 2003, gjennon telefonintervjuer, å følge opp hver enkelt elev som deltar i testen i flere år framover. Hvorvidt man vil lykkes med på denne måten å innhente de nødvendige data for å gjennomføre en *prediktiv validitetsanalyse* av testen, gjenstår å se.

**Samtidig validitet** finner man ved å beregne korrelasjonen mellom testscoren fra egen undersøkelse, og en annen måling utført ved det samme, eller et nærliggende tidspunkt. Dette vil selvsagt kun være interessant dersom den andre undersøkelsen er knyttet til måling av samme underliggende psykologiske konstrukt, hvilket bringer oss over i den siste typen validitet jeg her vil drøfte, **konstruktrelatert validitet**.

**Konstruktrelatert validitet** er opptatt med å etablere meningen til testscoren, det vil si å identifisere det psykologiske konstruktet som teoretisk sett er ansvarlig for forskjellen i test-score mellom de som har tatt testen (Satterly 1989).

9. Noen kritikere mener en slik *face – validity* har svært liten verdi, på grunn av dens subjektive og lite vitenskapelige karakter, og har derfor døpt den om til ”*faith – validity*”.(Satterly 1989)

Denne typen validitet vil spesielt være relevant når man ønsker å trekke slutninger fra test-scoren til et adferdsområde som ikke kan bli adekvat representert ved et enkelt kriterium, eller fullstendig definert gjennom et innholdsunivers (Ibid).

Vi ser igjen at meningen til test-scoren er det sentrale.

> *Construct validity is based on an integration of any evidence that bears on the interpretation of meaning of the test scores.(…….) The meaning of the measure, and hence its construct validity, must always be pursued – not only to support test interpretation but also to justify test use (Messick 1989, s.17).*

Det fins ulike prosedyrer for konstruktvalidering. Man kan for eksempel finne korrelasjonen mellom test-score og andre mål for det samme konstruktet, framkommet gjennom andre pålitelige og anerkjente undersøkelser knyttet til det samme konstrukt.

En innvending mot denne type korrelasjonsmetode er at den beskyldes for å ha en sirkulær karakter. Ved oppnåelse av høy korrelasjon har man kun ”funnet ut” det vi allerede vet gjennom den andre undersøkelsen. Hvis den nye testen altså kun gir samme informasjon som den gamle, har den åpenbart svært begrenset verdi.

Faktoranalyse er en annen teknikk man kan benytte for å undersøke konstruktrelatert validitet. Gjennom denne matematiske teknikken, søker man å redusere et komplekst system av innbyrdes oppgavekorrelasjoner til færre faktorer. Faktoranalyse kan derfor bidra til å identifisere underliggende konstrukter vi ikke er oppmerksomme på (Angell 1996). Problemet er hvorvidt disse konstruktene, empirisk identifisert gjennom faktoranalysen, korresponderer med de teoretiske konstruktene vi er opptatt av, og som testen ble utviklet for å måle.

Generelt kan man si at alle argumenter for konstruktrelatert validitet må baseres på teorier som ligger bak de aktuelle testoppgaver og variabler. Dette understreker viktigheten av klare konseptuelle definisjoner i forkant av prosessen med oppgaveutvikling.

> *It is from these conceptual definitions and their theoretical base that hypotheses are generated regarding the traits measured by the instrument ..(Gable & Wolf 1993, s.102).*

Det sistnevnte poeng viser hvor viktig det er å foreta kvalitative analyser av de konstruktene man ønsker å undersøke gjennom utprøving av oppgaver i en test.

I denne avhandlingen er ”*mathematical literacy*” det konstruktet jeg forsøker å relatere mine matematikkoppgaver til. Jeg har valgt å ikke benytte meg av faktoranalyse, men derimot lagt vekt på en begrepsmessig analyse av det nevnte konstrukt, for derigjennom å argumentere for oppgavenes konstruktrelaterte validitet.

### *7.2.3 Reliabilitet vs. Validitet – medspillere eller kombattanter ?*

Avslutningsvis vil jeg i dette kapitlet knytte et par betraktninger til i hvilken grad man kan si at reliabilitet og validitet hører sammen.

Ideelt sett ønsker man at en test har høy reliabilitet, og at man samtidig kan påberope seg gode grunner for å hevde at de slutninger man trekker ut fra anvendelsen av testen også har høy validitet. Her kan man imidlertid oppleve et dilemma, eller et visst motsetningsforhold, mellom hensynet til det ene og det andre.

Dersom en test har lav reliabilitet er den i prinsippet ubrukelig, og kan heller ikke ha høy validitet. Vi må derfor forsikre oss om at testen vår har høy reliabilitet. Dette kan vi gjøre gjennom for eksempel en utstrakt bruk av flervalgsoppgaver. Det gir høy sensorreliabilitet og gjør det mulig å ha med mange oppgaver, noe som altså øker muligheten for å oppnå en høy reliabilitet. Ensidig bruk av flervalgsoppgaver kan på den annen side redusere validiteten til de slutninger man kan trekke av resultatscoren til testen. Dette vil skje dersom man for eksempel ønsker å undersøke et konstrukt hvor formuleringsevne utgjør et viktig element. Da vil åpne oppgaver egne seg best. De har imidlertid, som vi har sett, den ulempen at sensorreliabiliteten blir langt lavere.

I store komparative internasjonale undersøkelser er man selvsagt avhengig av en høy reliabilitet, noe som har medført at man hovedsakelig har valgt å gjøre bruk av flervalgsoppgaver. Dette har ført til kritikk fra faglig kompetent hold om at testene har begrenset verdi, fordi de kun måler utvalgte, smale deler av elevenes kunnskaper. Det blir hevdet at de derfor i liten grad fanger opp prioriterte mål i enkelte deltakerlands læreplaner, som for eksempel evne til å formulere seg og kommunisere sine funn, samarbeidsevner, refleksjon og kritisk tenkning.

> *The multipel-choice format, furthermore, reinforces the idea that someone else already knows the answer to the question, so that original interpretations are not expected; the task, then, is to find or guess the "right" answer rather to engage in interpretavive activity (Resnick & Resnick 1992, s.47).*

Man anklager altså testene for å ha lav innholdsrelatert validitet, ved at de kun i beskjeden grad kan sies å være knyttet til eksplisitte mål i deltakerlandenes generelle læreplaner og fagplaner. I tillegg vil de psykologiske konstruktene man ut fra oppgaveresultatene kan si noe om være til dels uinteressante, da oppgavene altså kun måler svært begrensede sider ved elevers kompetanse. Den konstruktrelaterte validiteten vil altså heller ikke være tilfredsstillende.

Ettersom man i PISA har valgt å gjøre bruk av ulike oppgavetyper, og heller ikke knytter undersøkelsen til deltakerlandenes læreplaner, unngår man til en viss grad å bli rammet av en slik kritikk.

Kritikken som har blitt reist mot den utstrakte bruken av flervalgsoppgaver i denne type undersøkelser, har dessuten blitt imøtegått av blant annet Tamir (1990). Han hevder at flervalgsoppgaver, dersom de brukes riktig, kan være utmerkede diagnostiske redskaper for identifisering av elevers begrepsmessige forståelse, særlig når de blir bedt om å begrunne sine svarvalg. Olsen et. al. (2001) hevder i forlengelsen av denne argumentasjonen at flervalgsoppgaver til en viss grad:

> *…reflect a Vygotskian perspective of knowledge. If a question is stated and the student is left alone to answer this question in his or her own words there is a possibility that some misunderstanding will occur. By giving response alternatives, the student is provided with additional information on how to interpret the question. In this sense, the distractors have the same function as a conversation partner, hindering some of the possible misinterpretations (Olsen et.al.2001, s.408).*

Utfra referanser til empiriske studier utført i Norden (Gisselberg et.al. 1996), konkluderer de med at de ikke har funnet grunn til å anta at åpne oppgaver systematisk avdekker elevers kognitive forståelse på en bedre måte enn gode flervalgsoppgaver. De påpeker derimot at det er et komplekst samspill mellom karakteristiske trekk ved oppgaver og elevsvar. (Olsen et. al. 2001).

Fortrinnet med gode flervalgsoppgaver i forhold til åpne oppgaver, er altså at de bidrar til å gi tester høy reliabilitet, samtidig som de kan gi god diagnostisk informasjon (Ibid).

I store internasjonale undersøkelser mener Olsen et. al. (2001) det er viktig å ha et langt sterkere fokus på den potensielle diagnostiske informasjonen i benyttede oppgaver, og at dette diagnostiske perspektivet bør bli ansett som en:

> *..integrated part of the research aims for future international large-scale assessments (Ibid, s.417).*

For dette formålet vil gode flervalgsoppgaver kunne være utmerkede instrumenter.

Vi har sett at samtidige ønsker om høy reliabilitet og høy grad av validitet for en undersøkelse, delvis kan stå i et motsetningsforhold til hverandre. Likevel vil det være riktig å si at de først og fremst ”spiller på samme lag”, og at man i alle typer undersøkelser, ved hjelp av nøye gjennomtenkte vurderinger og avveininger, må sørge for at hensynet til begge faktorer blir ivaretatt.

# 8. Presentasjon og drøfting av oppgavene

I dette kapitlet vil jeg presentere de oppgavene jeg har utviklet og pilotert i tilknytning til PISA - prosjektet. Jeg vil videre drøfte disse oppgavene i lys av de statistisk bearbeidede elevresultatene, oppgavekriteriene i PISA, og konstruktet ”*mathematical literacy*”.

Ettersom det er selve matematikkoppgavene som står i fokus i min hovedoppgave, har jeg valgt å inkludere dem i sin helhet i dette kapitlet. De følger altså ikke med kun som vedlegg.

Til hvert spørsmål har jeg utviklet egne koder. Disse kodene, sammen med utvalgte statistiske kategorier, blir brukt under drøftingen av elevsvarene. Jeg vil nok engang understreke at elevresultatene, for mitt vedkommende, hovedsaklig er interessante i den grad de kan gi meg nyttig informasjon om oppgavene, og hvordan disse fungerer for mitt utvalg.

## 8.1 Koder

Jeg har under rettingen av de åpne oppgavene valgt å bruke tosifrede koder, etter modell fra kodeskjemaene utviklet i TIMSS (Brekke, Kobberstad, Lie & Turmo, 1999).

Dette falt naturlig, først og fremst fordi det er et kodesystem som er bra, og som man i TIMSS har gode erfaringer med. I tillegg ble hele kodesystemet nærmest unnfanget av TIMSS - gruppa ved ILS i Oslo, og jeg har derfor muligheten til å konsultere de absolutte spesialister på området, dersom jeg skulle ha behov for det.

Bakgrunnen for å ta i bruk tosifrede koder, er at man gjennom dem kan skaffe seg langt mer informasjon om elevers kunnskap og tenkemåter, enn man gjør ved kun å bruke kategoriene riktig/galt. For eksempel kan de avsløre hvilke løsningsstrategier elevene har benyttet, og hvilke misoppfatninger de eventuelt har. Det sistnevnte kalles innenfor matematikk - didaktikken for **diagnostisk informasjon** (Ibid).

Det første sifferet i kodene forteller om svar er avgitt, om det er riktig eller galt, eventuelt delvis riktig. Det andre sifferet refererer til typer av svar, for eksempel hvilken løsningsmetode elevene har benyttet, forklaringstypen, eller eksemplene de har anvendt, eller typen av feil/misoppfatning de har avslørt.

> *The score (the dimension of correctness) is thus linked to the other integrated aspects in such a way that the data can be analysed both for correctness and for diagnostic information (Angell et.al. 2000, s.161).*

Koder mellom 20 og 29 innebærer at eleven har fått to poeng for svaret, mens koder mellom 10 og 19 medfører ett poeng. Koder mellom 70 og 79 betyr at svaret er feil, og kode 99 vil si at eleven ikke har avgitt svar. I tillegg til disse, benyttet jeg i utgangspunktet kodene 97 og 98 for ”tull og tøys” og ”vet ikke”. Empirisk viste det seg imidlertid i mitt utvalg at disse kategoriene ble svært små, og derfor uinteressante. Jeg har på grunn av dette, innlemmet disse ”svarene” under kode 79, ”øvrige svar”. Unntaket er de to siste oppgavene, hvor jeg har valgt å beholde disse to kodene, ettersom det her var flere elever som avga den slags ”svar”.

# WIMBLEDON

Den første Wimbledon – turneringen i tennis ble arrangert i året 1877. Å gå seirende ut av denne tradisjonsrike og prestisjefylte turneringen er noe av det største en tennisspiller kan oppleve i sin karriere. Kvinneklassen ble opprettet allerede i 1884, og av kun tretten spillere stakk Spencer Gore av med seieren. Senere er turneringen blitt betraktelig utvidet, og i våre dager er 128 kvinner kvalifisert til å delta i kampen om det gjeve seierstrofeet.

Turneringen arrangeres som cup, dvs. at taperen av hver match blir utslått. Det blir ikke avholdt bronsefinale.

**Spørsmål 1:** WIMBLEDON

Hvor mange kamper må nå spilles for å kåre vinneren? ...........kamper

**Spørsmål 2:** WIMBLEDON

Vis eller forklar hvordan du kom fram til svaret.

## 8.2 Oppgave 1

Oppgave 1 er knyttet til en konkret årlig tennisturnering, Wimbledon, og har således en autentisk kontekst. Den inneholder, i likhet med flere av mine øvrige oppgaver, relativt mye tekst, noe som ikke er uproblematisk, ettersom det heter i MID:

> *The level of reading required to successfully engage with an item should be considered very carefully. The wording of items should be as simple and direct as possible. This is an assessment of mathematical literacy, not of reading ability (OECD PISA 2000, s.11-12).*

Kravene om autentisk kontekst og begrenset mengde tekst er, som jeg tidligere har vært inne på, et eksempel på to krav som kan være vanskelig å etterkomme samtidig. Det blir, som man ser, uttrykkelig understreket i MID at oppgaveteksten må være så enkel som mulig, selvsagt på bakgrunn av at oppgavene bør være valide i forhold til det konstruktet vi ønsker å måle, ”*mathematical literacy*”. Blir leseferdighetene en for betydelig faktor i løsningen av matematikkoppgaver, vil det man måler i for stor grad være "*reading literacy*", noe som helt klart ikke er ønskelig (se også kap.9.1.1).

Dette har jeg lagt vekt på å forholde meg til under utarbeidelsen av alle oppgavene.

**KODER FOR SPØRSMÅL 1/2**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **49** | **32,0** | |
| 20 | 128 spillere – en ut i hver kamp gir 127 kamper | 4 | 2,6 | 1,08 |
| 21 | Addert 64+32+16+….+1 = 127 | 45 | 29,4 | 0,85 |
| **Delvis riktig** | | **25** | **16,3** | |
| 10 | Riktig oppsatt, småfeil i utregning | 23 | 15,0 | 0,33 |
| 11 | 127 kamper uten eller med mangelfull forklaring | 2 | 1,3 | -0,79 |
| **Ikke riktig** | | **54** | **35,4** | |
| 70 | 128 kamper uten forklaring | 3 | 2,0 | -0,41 |
| 71 | 63/64/65 kamper, dvs. dividert 128 med 2 | 17 | 11,1 | -0,27 |
| 73 | 6/12/13 kamper | 8 | 5,2 | -1,01 |
| 74 | 7 kamper, blander sammen begrepene "runder" og "kamper" | 5 | 3,3 | -0,33 |
| 79 | Øvrige svar | 21 | 13,8 | -0,84 |
| **Ikke svart** | | **25** | **16,3** | |
| 99 | Blankt | 25 | 16,3 | -0,62 |

Under rettingen fant jeg ut at det var hensiktsmessig å slå sammen spørsmål 1 og 2, slik at svarene ble vurdert ut fra den begrunnelsen elevene hadde gitt. Det var blant annet en del elever som åpenbart blandet sammen begrepene "kamper" og "runder". Ettersom jeg ønsket å holde muligheten åpen for å analysere svarkategoriene separat, fikk disse egen kode.

Jeg har av psykometriske årsaker valgt å gjøre rettemalen til denne oppgaven relativt omfattende, med mange ulike koder. Dette gir meg større muligheter til gjennom statistiske analyser å oppdage interessante sammenhenger og korrelasjoner.

Som tabellen viser, svarer omtrent halvparten av elevene fullstendig, eller delvis riktig på denne oppgaven. Jeg har ikke gitt poeng for elever som har svart 128 kamper uten begrunnelse (kode 70), mens de som har svart det samme med en fornuftig begrunnelse, er plassert under kode 10. Av z-verdiene til de to kategoriene ser man at jeg psykometrisk sett har gode grunner for å gjøre dette, ettersom elevene med kode 10 er vesentlig flinkere enn de i kode 70. (En annen sak er at så få elever havner i kode 70, at det knapt kan forsvares å opprettholde den som egen kategori).

Elevene plassert under kode 11 har en z-score på -0,79, og dette taler isolert sett for at de burde plasseres i en 70 - kode. Imidlertid dreier dette seg kun om to elever, og det kan synes noe meningsløst i det hele tatt å opprette en egen kode for et så lavt antall. Alternativt kan man derfor slå sammen denne koden med en annen. Problemet er at det ikke faller naturlig å slå den sammen med en 70 kode, ettersom svaret i og for seg er riktig, bortsett fra at det ikke inneholder en tifredsstillende forklaring. En bedre løsning ville derfor være å inkludere disse svarene i kode 10.

Som tidligere nevnt, hadde man i generalprøven i PISA med en oppgave som var så godt som identisk med denne. Her opererte man med følgende kodeskjema:

**TENNIS TOURNAMENT SCORING**

**Full credit**

**Code 11**: 127 matches. Calculated by subtracting1(the tournament winner) from 128(the number of players). Every game has exactly one loser, and all players apart from the eventual winner lose one game.

**Code 12**: 127 matches. Calculated by summing the number of matches in each round (64+32+16+8+4+2+1=127).

**Code 13**: 127 matches. Calculated by any other method, or with no method shown.

**No Credit**

**Code 00**: Other responses

**Code 99:** Missing (OECD PISA 2002, s.20)

Som vi ser er dette et mye mindre elaborert kodeskjema, som inneholder få koder, faktisk kun en for feilsvar. Dette skyldes nok i hovedsak at det her var snakk om en pilotering. I den endelige PISA - undersøkelsen i 2003, vil trolig kodeskjemaet være en god del mer omfattende. Den diagnostiske informasjonen man gjennom svarene kan motta, ville ved benyttelse av det ovenstående kodeskjemaet, være svært begrenset.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 1/2 | 0,40 | 0,67 | 16,3 | -0,61 | 0,24 | 0,87 | -0,62 |

Kategoriene i denne tabellen ble presentert i metodeteori kapitlet. Av ovenstående tabell ser vi at denne oppgaven psykometrisk sett fungerer bra. P-verdien viser at noe under halvparten av elevene får til oppgaven, mens vi ser at ca. 1/6 av elevene ikke besvarer spørsmålet. De øvrige verdiene indikerer at oppgaven diskriminerer godt mellom elevene.

**Spørsmål 3:** WIMBLEDON

I år 2000 vant Venus Williams denne turneringen.

Hvor mange kamper måtte hun spille før hun mottok det ettertraktede seierstrofeet?

……….kamper

**KODER FOR SPØRSMÅL 3**

| Koder | Svaralternativer | Frekvens | Prosent | Z -verdi |
|---|---|---|---|---|
| **Fullstendig** | | **26** | **17,0** | |
| 20 | 7 kamper | 26 | 17,0 | 0,80 |
| **Delvis riktig** | | **48** | **31,4** | |
| 10 | 6 kamper, blant annet p.g.a. følgefeil fra forrige oppgave | 48 | 31,4 | 0,52 |
| **Ikke riktig** | | **43** | **26,1** | |
| 70 | 64/65 kamper | 6 | 3,9 | -0,56 |
| 71 | 127/128 kamper | 11 | 7,2 | 0,07 |
| 79 | Øvrige svar | 26 | 17,0 | -0,78 |
| **Ikke svart** | | **36** | **23,5** | |
| 99 | Blankt | 36 | 23,5 | -0,58 |

Spørsmål 2 og 3 henger åpenbart nært sammen, slik at om man har svart riktig på oppgave 2, burde man også kunne lykkes i å besvare den påfølgende oppgaven riktig. Likevel ser man at det er langt færre som har svart fullstendig på oppgave 3, enn oppgave 2.

Svært mange har her svart 6, i stedet for 7 kamper. Jeg valgte allerede tidlig i rettingen å gi disse egen kode, slik at det på et senere tidspunkt skulle være mulig å analysere denne svarkategorien separat.

Jeg var derimot noe mer usikker på om jeg skulle belønne dette svaret med poeng, ettersom det strengt tatt er feil.

Av den statistiske bearbeidelsen ser vi at z-verdien til elevene tilhørende denne svar – kategorien, er forholdsvis høy, det vil si at det er relativt flinke elever som har avgitt dette svaret. Ut fra psykometriske betraktninger, vil det derfor være riktig å gi denne gruppa ett poeng på denne oppgaven, både for å skille dem fra 2 - poengs elevene som er sterkere, og fra 0 - poengs elevene, som ut fra z - verdiene åpenbart er svakere.

For poenggivende koder vil det imidlertid være slik, at man først og fremst tar hensyn til om elevene har gjort noe riktig i besvarelsen, deretter sjekker man gjennom z-verdiene hvordan oppgaven fungerer. Man kan selvsagt på bakgrunn av disse z-verdiene foreta endringer i kodingen, men z-verdier alene kan ikke bestemme hvorvidt man knytter poeng til de ulike kodene. En slik praksis ville kunne få svært underlige og uønskede konsekvenser.

Det aktuelle svaret, 6 kamper, har flere elever kommet fram til på grunn av følgefeil fra forrige spørsmål. Jeg mener derfor de på sett og vis har tenkt riktig, har avgitt delvis riktig svar, og at de bør belønnes med ett poeng. Z-verdiene gir meg støtte for en slik vurdering.

Et av de vanligste feilsvarene på dette spørsmålet er knyttet til kode 71, altså 63/64/65 kamper. Svaret kan i første omgang virke noe underlig, men det er mulig at selve cup-systemet er ukjent for en del elever, slik at de rett og slett ikke forstår oppgaven.

De samme refleksjonene gjelder selvsagt overfor de som har svart 127/128 kamper, kode 72. Selv om disse elevene er en god del flinkere enn de jeg omtalte i avsnittet over, er det noe med kamper kontra runder i cuper som åpenbart faller vanskelig.

I likhet med i forrige oppgave, konstaterer jeg også her at "ikke svart"- prosenten, er temmelig høy, nemlig 23,5. Dette kan synes å bekrefte at selve oppgavekonteksten er fremmed og uforståelig for en del elever.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 3 | 0,33 | 0,58 | 23,5 | -0,57 | 0,52 | 0,80 | -0,58 |

Vi ser av de ulike verdiene i denne tabellen, at også denne oppgaven psykometrisk sett fungerer bra. Litt uvanlig er det kan hende at z-verdien for de som ikke har besvart oppgaven, Z(9), er nærmest identisk med Z(0), feilsvarene. Dette var også tilfelle på den første oppgaven.

Ofte vil man se at det er de aller svakeste elevene som overhodet ikke gir noe svar. At det er annerledes her, kan bero på tilfeldigheter, grunnet mitt relativt beskjedne utvalg, eller det kan muligens skyldes motivasjonsproblemer. Jeg har en mistanke om at endel elever som normalt ville havnet i en av de andre kategoriene generelt har tatt noe lett på denne testen, simpelthen latt være å svare på en del oppgaver, ettersom prestasjonen på testen ikke ga uttelling på karakteren i faget. Dette kan selvsagt lyde noe spekulativt, men da jeg administrerte testen, observerte jeg ved selvsyn at motivasjonen blant elevene var noe varierende.

De tilsvarende verdiene for majoriteten av spørsmålene i denne testen, er for øvrig mer "normale", det vil si at Z(9) for det meste er lavere enn Z(0).

**Spørsmål 4:** WIMBLEDON

Vi forutsetter at turneringen også i 1884 ble arrangert som cup (uten bronsefinale).

Hvor mange kamper ble da spilt i kvinneklassen for å kåre en vinner?

………kamper

**KODER FOR SPØRSMÅL 4**

| Koder | Svaralternativer | Frekvens | Prosent | Z -verdi |
|---|---|---|---|---|
| **Riktig** | | **26** | **17,0** | |
| 10 | 12 kamper | 26 | 17,0 | 1,00 |
| **Ikke riktig** | | **78** | **51,0** | |
| 70 | 11 kamper | 4 | 2,6 | 0,23 |
| 71 | 13 kamper | 6 | 3,9 | 0,08 |
| 72 | 14 kamper | 7 | 4,6 | 0,10 |
| 73 | 6/7 kamper (Dividerer 13 med 2) | 12 | 7,8 | -0,47 |
| 79 | Øvrige svar | 46 | 30,1 | $6{,}0 \times 10^{-4}$ |
| **Ikke svart** | | **49** | **32,0** | |
| 99 | Blankt | 49 | 32,0 | -0,46 |

Riktig svar på oppgave 4 forutsetter at man har forstått prinsippet om at det alltid ved denne type turneringer, vil bli avholdt en kamp mindre enn det totale antall deltakere. I oppgave 1 / 2 var det kun 4 elever som eksplisitt ga uttrykk for at de forsto dette, men vi ser nå at så mange som 26 elever svarer riktig her. Verd å merke seg er at dette er flinke elever, med en z-verdi på 1,00. Slik bekreftes antagelsen om at spørsmålet er relativt vanskelig. Nær en tredjedel av elevene har for øvrig ikke besvart oppgaven.

Jeg ser ellers at alle 70 - kodene, bortsett fra en, har positiv z-verdi. Særlig gjelder dette de svarene som er i nærheten av det riktige. Dette gjør at jeg har vurdert muligheten av å belønne elevene som har svart riktig med to poeng, og en 20 - kode, for så å omgjøre kode 70,71 og 72 til 10-er koder. En slik endring av kodene for denne oppgaven, vil ha en holdbar psykometrisk begrunnelse i forhold til z-verdiene.

Imidlertid er, som vi har vært inne på, den viktigste grunnen for å gi elevene en ”delvis riktig”- kode, og ett poeng, at de har gjort noe riktig i sitt løsningsforsøk. I denne oppgaven var det ikke nødvendig å vise hvordan man hadde tenkt, ettersom elevene skulle finne svaret uten å utføre noen omfattende regneoperasjoner. De har altså ikke vist noen utregning, og det er derfor vanskelig å avgjøre om de har løst oppgaven på en delvis riktig måte eller ikke. Jeg valgte ut fra slike vurderinger å beholde mine opprinnelige koder.

Som en kuriositet kan nevnes at nøyaktig det samme antall elever som besvarte spørsmål 3 på en fullstendig måte, også  har fått til oppgave 4. Det er imidlertid ikke de helt samme elevene. Kun 9 elever svarte fullstendig på både oppgave 3 og oppgave 4. Av de øvrige 17 som svarte riktig på oppgave 4, hadde 11 også svart delvis riktig på oppgave 3, mens de siste 6 elevene svarte feil på oppgave 3. Verdt å merke seg er imidlertid at hele fem av disse seks elevene, fikk kode 71 på oppgave 3, hvilket kan tyde på at de til en viss grad forstår det generelle prinsippet for utregning av antall kamper i en cup - turnering, men sannsynligvis ikke helt har innsett forskjellen på begrepene ”runder” og ”kamper”. Av z-verdien for 70-kodene på oppgave 3 ser vi da også at elevene som har fått kode 71, er vesentlig flinkere enn de som har fått de andre feilkodene.

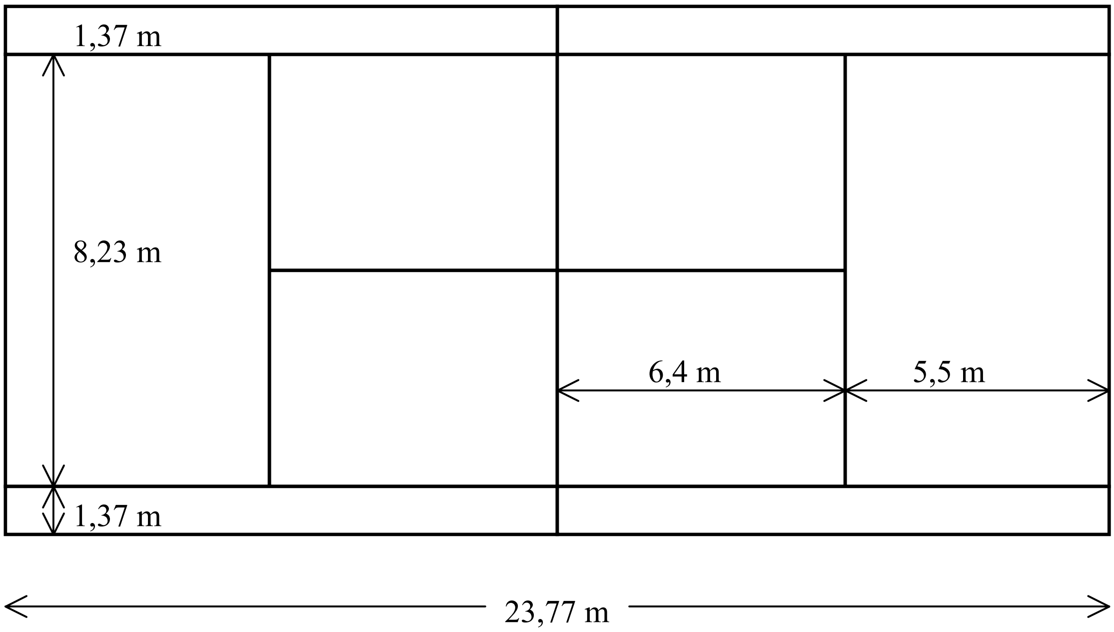
| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spørsmål 4 | 0,17 | 0,45 | 32,0 | -0,03 | 1,00 | -0,46 |

I forhold til de øvrige oppgavene jeg har gjennomgått, og sett i lys av at også denne oppgaven befant seg i begynnelsen av heftet, ser vi at "ikke svart"- prosenten er relativ høy, og p-verdien svært lav. At over 50% av elevene svarer feil, gjør også sitt til at z-verdien for 70 - kodene blir høyere enn på de foregående oppgavene. Ellers er de andre verdiene psykometrisk sett akseptable. Jeg ser ikke minst at oppgaven diskriminerer bra mellom elevene.

**Spørsmål 5:** WIMBLEDON

En tennisbane har mål som vist på figuren nedenfor.

Hva er arealet av hele banen? Vis utregningene dine.




**KODER FOR SPØRSMÅL 5**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **70** | **45,8** | |
| 20 | 260,7m$^2$/261m$^2$/264m$^2$ (avrunding til hele meter) | 70 | 45,8 | 0,62 |
| **Delvis riktig** | | **27** | **17,7** | |
| 10 | 195m$^2$ dvs. arealet av singelbanen | 18 | 11,8 | -0,04 |
| 11 | Desimalfeil | 2 | 1,3 | -0,39 |
| 12 | 232,9m$^2$,228,4m$^2$ utfra (1,37m+8,23m) x 23,77m | 3 | 2,0 | -0,07 |
| 13 | Riktig oppsatt, småfeil i utregning | 4 | 2,6 | 0,26 |
| **Ikke riktig** | | **39** | **25,6** | |
| 79 | Øvrige svar | 39 | 25,6 | -0,71 |
| **Ikke svart** | | **17** | **11,1** | |
| 99 | Blankt | 17 | 11,1 | -0,88 |

Spørsmål 5 er knyttet til utregning av areal, og er den delen av Wimbledon - oppgaven som flest elever har besvart riktig.

Noe av det mest iøynefallende her, sett på bakgrunn av at dette ikke er en flervalgsoppgave, er at "ikke svart"- prosenten på dette spørsmålet er svært lav, like i overkant av ti prosent. Dette indikerer at oppgaven omhandler et tema de fleste føler seg relativt trygge på. Noe som også understøtter denne antagelsen, er altså at bortimot to tredjedeler (63,5%) av elevene har svart fullstendig, eller delvis riktig på dette spørsmålet. Dette er markert høyere enn på de andre spørsmålene i denne oppgaven. Det er vel grunn til å tro at dette kan skyldes at spørsmålet er heller tradisjonelt, i den forstand at det ligger tett opp til en type oppgaver elevene er kjent med fra skolen.

Lengden av banen er for øvrig oppgitt på to ulike måter, slik at elevene selv kan velge hvilke av disse lengdene de vil benytte i sine utregninger. I følge regler om gjeldende siffer, ville 260 $m^2$ være det korrekte svar på denne oppgaven, dersom man velger å benytte de minst nøyaktige lengdeopplysningene. Jeg har imidlertid ikke valgt å la dette få konsekvenser for kodingen, men inkludert alle svar i overkant av 260 $m^2$ i kode 20.

Vi ser av tabellen at tre av fire 10-er koder har negativ z-verdi, noe som medfører at de som har fått ett poeng på denne oppgaven, også samlet får en svakt negativ z-verdi. Den er dog vesentlig høyere enn z-verdien for de som har svart galt, slik at jeg psykometrisk sett har gode grunner for poengmessig å skille mellom disse svarkategoriene. Også her er imidlertid den utslagsgivende årsaken til at disse elevene har fått poeng, at de har gjort noe riktig i sin besvarelse av spørsmålet.

Det som kan virke fornuftig å gjøre her, er å eliminere visse enkeltkoder, slik at kode 11, 12 og 13 slås sammen med kode 10. På den måten vil alle elevene i "delvis riktig"- kategorien samles under denne koden. Da unngås mange små og relativt uinteressante koder, og kodeskjemaet blir i tillegg enklere.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 5 | 0,55 | 0,63 | 11,11 | -0,71 | -0,03 | 0,62 | -0,88 |

Verdiene i ovenstående tabell viser at oppgaven psykometrisk sett fungerer bra.

**Spørsmål 6:** WIMBLEDON

Når Venus Williams server, går ballen ca. 17,5 meter gjennom lufta før den treffer bakken på motstanderens side etter ca 0,3 sekunder.

Hva er da ballens gjennomsnittshastighet i m/s ?

…………m/s

**KODER FOR SPØRSMÅL 6**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Riktig** | | **73** | **47,7** | |
| 10 | 58,3m/s evt. 58m/s | 73 | 47,7 | 0,56 |
| **Ikke riktig** | | **47** | **30,7** | |
| 70 | 17,8m/s, 17,2m/s (adderer/subtraherer 17,5 og 0,3) | 3 | 2,0 | -1,32 |
| 71 | 5,25m/s eller 52,5m/s | 13 | 8,5 | -0,14 |
| 72 | 0,0875m/s, (17,5 x 0,3) dividert med 60 eller 100 | 4 | 2,6 | -1,18 |
| 73 | 5,8m/s (Desimalfeil) | 4 | 2,6 | -0,31 |
| 79 | Øvrige svar | 23 | 15,1 | -0,31 |
| **Ikke svart** | | **33** | **21,6** | |
| 99 | Blankt | 33 | 21,6 | -0,67 |

Spørsmål 6 er også av en type som elevene bør gjenkjenne fra skolematematikken, ettersom det er knyttet til benevnelsen m/s. I underkant av halvparten av elevene svarer her korrekt. Det er ellers grunn til å merke seg at alle fire regnearter tas i bruk for å besvare denne oppgaven. Det riktige svaret framkommer selvsagt ved å benytte divisjon. Ved å addere eller subtrahere de to oppgitte sifferne i oppgaven, får man de svar som befinner seg i kode 70. Kun tre elever har funnet dette opportunt.

Flere har imidlertid valgt å benytte multiplikasjon, og man oppnår da resultatet 5,2 m/s. Enkelte elever har trolig innsett at dette har måttet være for lavt, og har derfor oppgitt 52,5m/s som svar. Dette er selvsagt et langt fornuftigere svar, faktisk relativt nær det riktige svaret, men framkommer først etter en viss "triksing" med utregningene.

Et svar som ikke er like rasjonelt, er 0,0875m/s, som har fått koden 72. Her har elevene først multiplisert de oppgitte sifferne, for så å dividere med 60. Riktignok er det bare snakk om fire elever, men man kan selvsagt undres over hvordan det er mulig å oppgi et slikt svar. Det er hvert fall åpenbart at begrepet m/s for disse elevene gir liten mening.

Fart oppgis jo i de fleste sammenhenger i km/h, og i min første formulering av dette spørsmålet skulle elevene oppgi svaret med denne benevnelsen. Dette krever imidlertid regneoperasjoner som anses for relativt vanskelige, og jeg bestemte meg derfor, etter å ha rådført meg med PISA - prosjektgruppa ved instituttet, for den oppgitte spørsmåls - formuleringen.

Muligens vil elevene ha et nærmere forhold til hastighetsangivelsen km/h, ettersom deres hastighetsoppfatning er knyttet til benevnelsen km/h gjennom sykkel- og bilkjøring. De vil dermed lettere kunne anslå om et svar virker fornuftig, dersom det oppgis med denne benevnelsen. Svaret 0,0875 m/s tilsvarer ca. 0,3 km/t, og det kan tenkes at det hadde vært lettere å innse at det sistnevnte ikke var et rimelig svar på farten til en serveball i tennis.

Hvorvidt jeg burde holdt fast ved den opprinnelige teksten er vanskelig å avgjøre, men det viser en type dilemma man står ovenfor ved utforming av oppgaver.

Ser vi på kodeskjemaet, kan det også her virke mest fornuftig å slå sammen enkelte 70 - koder. Både kode 70, 72 og 73 er åpenbart svært smale kategorier. Grunnen til at jeg har valgt å beholde dem, er at de viser hvor mange elever som har benyttet hver av de fire ulike regneartene, hvilket ga meg muligheten til å analysere dem separat.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spørsmål 6 | 0,48 | 0,54 | 21,6 | -0,40 | 0,56 | -0,67 |

Verdiene i ovenstående tabell bekrefter at oppgaven psykometrisk sett fungerer godt.

### *8.2.1 Wimbledon oppgaven, PISA og "mathematical literacy".*

I hvilken grad har jeg i Wimbledon-oppgaven lyktes med å innfri de tidligere omtalte oppgavekravene i PISA, og kan denne oppgaven benyttes til å måle aspekter ved konstruktet "*mathematical literacy*"?

Jeg vil gjøre oppmerksom på at dette var en av de oppgavene som ikke kom med i PISA's egen pilotundersøkelse. Dette behøver ikke nødvendigvis å indikere at den ikke tilfredsstilte de ulike oppgavekravene, men kan bl.a. skyldes at en oppgave som var svært lik, ja for enkelte spørsmål nærmest helt identisk, allerede hadde blitt vurdert og antatt, i forkant av at mine oppgaver ble innsendt. (Oppgaven het Australian Open !!!)

Som vi så i kapittel 4, har begrunnelsen for matematikkfaget historisk sett i hovedsak svingt mellom nytte- og dannelsesargumenter.

Svein Sjøberg (1998) hevder at enhver form for legitimering av vitenskap, kunst, kultur og kunnskap tradisjonelt har kommet til uttrykk gjennom disse to retningene. Den ene, som har sine røtter i den filosofiske arven fra det gamle Hellas, ser på kunnskap som et mål i seg selv. Refleksjon, erkjennelse, innsikt og forståelse blir her nærmest meningen med et meningsfylt liv, som Sjøberg uttrykker det. I den instrumentelle tradisjonen er det derimot nytten av den kunnskapen man erverver som rettferdiggjør satsningen på skole og utdanning.

Disse to tradisjonene omtales også av Sjøberg som henholdsvis dannelsesargumentet og nytteargumentet. Sjøberg påpeker at nytteargumentet er det som har stått sterkest i skoledebatten her til lands de siste årene. Hva skal vi med dette, hva skal vi bruke det til, hvilken nytte har vi av å lære dette, har vært spørsmål som stadig har blitt stilt av involverte aktører.

Vi så i kapittel 4, at det man i PISA mener bør være den sentrale ideen innenfor matematikk – undervisning, er å lære elevene å *matematisere*. Jeg beskrev i nevnte kapittel hvordan dette kan summeres opp i følgende punkter: Først tar man utgangspunkt i et virkelig problem, så transformerer man det til et matematisk og løser det, for så endelig å anvende løsningen på det virkelige problemet.

Intensjonen med å lansere denne framgangsmåten, er nok ikke at all matematikkundervisning nødvendigvis bør flyttes fra klasserommet og ut i den "virkelige" verden, selv om dette til tider utvilsomt kan være motiverende og virkningsfullt. Derimot tror jeg poenget er at man i matematikkoppgaver flittig *anvender* problemer knyttet til virkeligheten. Slik vil elevene allerede i ung alder kunne erfare *nytten* av å være i stand til å bruke matematiske algoritmer og tenkemåter, og derigjennom forhåpentligvis kunne innse viktigheten av å lære seg grunnleggende matematiske ferdigheter.

Selv om Sjøberg argumenterer for at dannelsesaspektet knyttet til kunnskapstilegnelse bør bli mer framtredende i den offentlige debatt, vil det muligens fremdeles overfor barn og unge være lettere å argumentere med matematikkens *nytteverdi*, for å øke motivasjonen og lærelysten.

Wimbledon oppgaven har fått en autentisk kontekst med en historisk innledning, og tilfredsstiller PISA's krav på dette feltet. Spørsmålene er også formulert slik at elevene ikke kan ty til ferdig innlærte algoritmer for å besvare dem, men må anvende sine akkumulerte matematiske kunnskaper. De vil derfor oppleve at de drar *nytte* av tidligere tilegnede matematiske ferdigheter for å løse denne oppgaven, som i tillegg, forhåpentligvis, framstår for dem som forholdsvis relevant og meningsfull.

Dette er med på å øke oppgavens validitet i forhold til konstruktet "*mathematical literacy*". Vi så jo i kapittel 4 hvordan:

> *…mathematical knowledge put into functional use in a multitude of different situations and contexts in varied, reflective and insight-based ways (OECD PISA 2001a, s.5).*

kjennetegner dette konstruktet. Det er altså den ***konstruktrelaterte validiteten*** til oppgaven jeg her argumenterer for. At en så godt som identisk oppgave ble brukt i generalprøven i PISA, understøtter også påberopelsen av en slik validitet.

For å kunne besvare spørsmålene i oppgave 1, må elevene ha en viss *tallforståelse* og kunne utføre *meningsfylte operasjoner* på de oppgitte data. Spørsmål tre inviterer også til å bruke *elegante beregninger*. For alle spørsmålene vil det være slik at elevene ved å benytte seg av *overslag*, kan undersøke rimeligheten av sitt svar.

Som man ser, har jeg dermed knyttet denne oppgaven til flere av elementene i den sentrale ideen ***kvantitativt resonnement***. Med dette har jeg forsøkt å gi en begrunnelse for oppgavens ***innholdsrelaterte validitet.***

En innvending mot denne oppgaven kan være at noen av spørsmålene er for nære knyttet til hverandre. Det vil for eksempel være usannsynlig at man skal kunne besvare spørsmål 3, dersom man ikke har fått riktig svar på spørsmål 1 / 2. Ved en gjennomgang av svarene viste det seg at dette stemts. Ingen som har svart uriktig på spørsmål 1 / 2, har svart korrekt på spørsmål 3. Noe av det samme kan hevdes om spørsmål 4 kontra spørsmål 1, men her viste det seg faktisk at noen elever har svart riktig på spørsmål 4, selv om de ikke har lyktes i å besvare spørsmål 1 korrekt.

## 8.3 OPPGAVE 2

I oppgaven med tittel ”Moped” skal elevene hente ut informasjon fra en autentisk tabell. Jeg har altså også her en ”*real life context*”.

Denne oppgaven kom i en lettere revidert utgave med i generalprøven i PISA. Man hadde da redusert den noe ved å fjerne spørsmål 7 og 8, og erstattet disse med et nyformulert spørsmål knyttet til utregning av renter. Spørsmål 9 i mitt opprinnelige utkast var beholdt uendret.

# MOPED

Mette ønsker å kjøpe seg en ny moped og har av banken fått innvilget en lånesøknad på 30000 kr.

Her ser du tilbakebetalingsplanen for dette lånet.

**BETALINGSPLAN**

| Termin | Terminbeløp | Avdrag | Renter | Gebyr | Restgjeld |
|---|---|---|---|---|---|
| 1 | 4 035 | 3 000 | 975 | 60 | 27 000 |
| 2 | 3 938 | 3 000 | 878 | 60 | 24 000 |
| 3 | 3 840 | 3 000 | 780 | 60 | 21 000 |
| 4 | 3 742 | 3 000 | 682 | 60 | 18 000 |
| 5 | 3 645 | 3 000 | 585 | 60 | 15 000 |
| 6 | 3 548 | 3 000 | 488 | 60 | 12 000 |
| 7 | 3 450 | 3 000 | 390 | 60 | 9 000 |
| 8 | 3 352 | 3 000 | 292 | 60 | 6 000 |
| 9 | 3 255 | 3 000 | 195 | 60 | 3 000 |
| 10 | 3 158 | 3 000 | 98 | 60 | 0 |

**Oversikt**

| | |
|---|---|
| Lånebeløp | 30 000 |
| Termingebyr | 60 |
| Etableringsgebyr | 1 000 |
| Nedbetalingstid (år) | 5 |
| Antall terminer per år | 2 |
| Rente I prosent | 6,5 |
| Avdragsfrie terminer | 0 |
| | |
| Sum termingebyr | 600 |
| Sum avdrag | 30 000 |
| Sum renter | 5 362 |
| Samlet kostnad | 36 962 |

Vi antar at Mette aksepterer lånebetingelsene, at terminbeløpene skal betales på datoene 01.01 og 01.07 årlig og at hun betaler punktlig.

Hun starter nedbetalingen av lånet 01.07.01.

**Spørsmål 7:** MOPED

Hvor stor er restgjelden hennes etter innbetalingen den 01.07.03?

……………kr

**KODER FOR SPØRSMÅL 7**

| **Koder** | **Svaralternativer** | **Frekvens** | **Prosent** | **Z - verdi** |
|---|---|---|---|---|
| **Riktig** | | **43** | **28,1** | |
| 10 | 15000 kr | 43 | 28,1 | 0,85 |
| **Ikke riktig** | | **85** | **55,7** | |
| 70 | 18000 kr | 9 | 5,9 | 0,13 |
| 71 | 12000 kr | 26 | 17,0 | -0,07 |
| 79 | Øvrige svar | 50 | 32,8 | -0,31 |
| **Ikke svart** | | **25** | **16,3** | |
| 99 | Blankt | 25 | 16,3 | -0,83 |

Jeg ser av kodeskjemaet og tabellen under, at denne oppgaven fungerer greit. Z-verdiene for de ulike hovedkategoriene viser at oppgaven diskriminerer godt, og kodene er her temmelig uproblematiske og ikke - kontroversielle. "Øvrige svar", som blir en slags sekkekategori, kan virke relativ stor, men det skyldes at elevsvarene spriket i ulike retninger, slik at det var vanskelig å skille ut flere klasser med svar som det falt naturlig å sette sammen under separate koder. Dette skyldes trolig at en god del elever enten rett og slett ikke forsto spørsmålet, eller at de ikke skjønte hva de skulle gjøre for å løse det. Blant annet var det flere som foretok helt unødvendige utregninger.

Det kunne derfor vært svært interessant å omgjøre denne oppgaven til en flervalgsoppgave, med distraktorene 9000 - 12000 - 18000 og 21000, i tillegg til det korrekte alternativet. Som jeg nevnte i metode-teori kapitlet har det, i sammenheng med forskning på resultater fra TIMSS, vist seg at oppgaveformatet og valg av distraktorer har stor innflytelse på p-verdiene som oppnås (Olsen et.al. 2001).

I denne aktuelle oppgaven er det stor sannsynlighet for at de ovenfornevnte svaralternativene ville hjulpet elevene til å forstå så vel spørsmålet, som hvilket type svar som var forventet. Dette ville i så fall medført en langt lavere "ikke-svart"- prosent, og trolig også en høyere p-verdi. Slik kunne altså en endring av oppgaveformatet penset en del elever inn på riktig spor, og rettledet dem i arbeidet med å finne den riktige løsningen. På den måten ville distraktorene hatt samme funksjon som en samtalepartner innenfor Vygotsky's læringsteori, slik jeg omtalte det i kapittel 7.2.3.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spørsmål 7 | 0,28 | 0,54 | 16,3 | -0,19 | 0,86 | -0,83 |

**Spørsmål 8:** MOPED

På hvilken dato vil Mette være gjeldfri?

………………………

**KODER FOR SPØRSMÅL 8**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Riktig** | | **47** | **30,7** | |
| 10 | 01.01.06 eller 02.01.06 | 47 | 30,7 | 0,89 |
| **Ikke riktig** | | **73** | **47,7** | |
| 70 | 01.07.05 | 24 | 15,7 | 0,09 |
| 71 | 01.07.06 | 15 | 9,8 | 0,03 |
| 72 | 01.07.10/01.01.10 ( 10 som år i stedet for termin) | 6 | 3,9 | -0,56 |
| 79 | Øvrige svar | 28 | 18,4 | -0,55 |
| **Ikke svart** | | **33** | **21,6** | |
| 99 | Blankt | 33 | 21,6 | -0,78 |

Ut fra de to tabellene ser vi at også denne oppgaven fungerer bra psykometrisk. Z-verdiene viser at den diskriminerer fint mellom svake og sterke elever, noe som også bekreftes av verdien for diskriminering.

Vi ser av 70 - kodene at det er forholdsvis stor forskjell mellom z-verdiene til kodene 70 og 71 på den ene siden, og kodene 72 og 79 på den andre. De som har avgitt svar plassert under de to første kodene har, som man ser, kun bommet med en termin på det korrekte svaret, mens elevene som havner under de andre kodene har avgitt mindre fornuftige svar. Jeg kunne ut fra en slik argumentasjon valgt å endre kodene, slik at kode 10 hadde blitt en 20-kode, og kodene 70 og 71 hadde blitt omgjort til poenggivende 10-er koder. Innvendingen er her igjen at svarene i kodene 70/71 er gale, og at de derfor ikke bør gi poeng.

På den annen side kan det hevdes at disse elevene har gjort ***noe*** riktig, men begått en tellefeil underveis, og dermed kommet fram til feil svar. Dette kan utgjøre et grunnlag for å gi dem ett poeng. En slik vurdering finner altså støtte i de beregnede z-verdiene.

Jeg har etter disse avveiningene likevel valgt å holde fast ved mine opprinnelige koder, fordi jeg mener det bør veie tyngst at de aktuelle svarene faktisk er gale.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spm.8 | 0,31 | 0,60 | 21,6 | -0,22 | 0,89 | -0,78 |

**Spørsmål 9:** MOPED

Hvorfor blir rentebeløpet stadig mindre?

……………………………………………………………………………………………………

**KODER FOR SPØRSMÅL 9**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **37** | **24,2** | |
| 20 | Eksplisitt forklare at rentebeløpet regnes av restgjelden. Når den avtar vil derfor også rentebeløpet synke. | 37 | 24,2 | 0,76 |
| **Delvis riktig** | | **52** | **34,0** | |
| 10 | Rentebeløpet blir mindre fordi restgjelden blir mindre.<br>Ingen forklaring av hvorfor det er slik. | 46 | 30,1 | 0,20 |
| 11 | Rentebeløpet blir mindre fordi man betaler renter av gjelden | 4 | 2,6 | 0,68 |
| **Ikke riktig** | | **31** | **20,3** | |
| 70 | Mangelfull forklaring som for eksempel :<br>- Rentebeløpet blir mindre p.g.a. terminbeløpet<br>- Fordi hun låner fra banken<br>- Fordi hun betaler punktlig/betaler tilbake etter hvert<br>- Fordi man får renter av pengene man har på konto | 31 | 20,3 | -0,40 |
| **Ikke svart** | | **33** | **21,6** | |
| 99 | Blankt | 33 | 21,6 | -0,81 |

Vi ser av kodeskjemaet og tabellen under, at oppgaven psykometrisk sett fungerer bra.

Et framtredende trekk ved svarene på spørsmål 9, er at mange elever er svært upresise når de skal begrunne med egne ord hvorfor rentebeløpet synker.

Som det går fram av kodene, har jeg valgt å skille mellom de som har en fullstendig begrunnelse for dette, og de som kun svarer at rentebeløpet blir mindre fordi restgjelden synker, uten å uttrykke eksplisitt at rentebeløpet regnes av restgjelden. Ser vi på z-verdiene for disse gruppae, finner vi at jeg psykometrisk sett har gode grunner for å gjøre dette, i og med at elevene under 20-koden har en langt høyere z-verdi enn de som havner i kode 10.

De få elevene som er samlet under kode 11, har imidlertid en z-verdi som er nærmest identisk med den til de som har svart fullstendig. Dette ville være et argument for å innlemme dem i 20-koden. På grunn av svarformuleringen har jeg imidlertid valgt å avstå fra dette, da de ut fra dette kriteriet, etter min oppfatning, hører hjemme der de allerede er plassert. Ettersom det dog er lite meningsfullt å opprettholde så små kategorier, vil det være mer riktig å heller legge disse svarene inn under kode 10.

Som nevnt i innledningen til oppgaven, ble dette spørsmålet beholdt uendret i generalprøven i PISA. Der hadde man imidlertid valgt å benytte et kodeskjema som var en god del enklere enn mitt. Det er imidlertid grunn til å påpeke at kodene som benyttes i generalprøven, ikke nødvendigvis er identisk med det man kommer til å benytte i den endelige undersøkelsen. Meningen er blant annet å bruke resultatene fra generalprøven til drøfting og elaborering av de endelige kodeskjemaene for PISA 2003.

Slik var PISA's kodeskjema med resultatene fra den norske generalprøven:

**KODER FOR SPØRSMÅL 9 FOR GENERALPRØVEN I PISA**

**(OECD PISA 2002, s.52)**

| **Koder** | **Svaralternativer** | **Frekvens** | **Prosent** | **Z - verdi** |
|---|---|---|---|---|
| **Fullstendig** | | **99** | **43,8** | |
| 1 | Code 1: Reason stating that the balance is decreasing.<br>• The interest decreases with time because there is less to pay off. | 99 | 43,8 | 0,70 |
| **Ikke riktig** | | **34** | **15,0** | |
| 0 | Code 0 : Other responses<br>• Because the value of the car goes down<br>• The interest decreases because the amount of each payment decreases. | 34 | 15,0 | -0,60 |
| **Ikke svart** | | **93** | **41,2** | |
| 99 | Code 9 : Missing | 93 | 41,2 | -0,52 |

I tabellen under har jeg i tillegg til verdiene fra min pilotundersøkelse, lagt inn de tilsvarende verdiene fra generalprøven i PISA, som jeg omtaler som PISA G. Parentesene indikerer at verdiene, av ulike grunner, ikke er helt sammenlignbare.

| Sp.m. 9 | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Min pil. | 0,41 | 0,56 | 21,6 | -0,40 | 0,24 | 0,76 | -0,81 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **PISA G.** | **(0,44)** | | **41,2** | **-0,60** | | **(0,70)** | **-0,52** |

Vi ser at den største forskjellen her faktisk er "ikke svart" - prosenten. Noe overraskende er den nesten dobbel så høy i PISA G. Å uttale seg om årsaker til dette kan lett fortone seg temmelig spekulativt, men det er nærliggende å anta at dette er knyttet til elevenes motivasjon for å besvare de forelagte spørsmålene. Ved tre av de fire skolene som deltok i min pilotundersøkelse, var jeg selv tilstede for å motivere elevene til innsats. På den fjerde spurte faglig lærer undertegnede om det var i orden at hun kikket gjennom elevbesvarelsene for eventuelt å vurdere dem karaktermessig, hvilket jeg bifalte. Denne opplysningen inspirerte nok også elevene til optimal innsats. Alt i alt var derfor elevene som deltok i min pilotundersøkelse, forholdsvis godt motivert til innsats. Det ble neppe, av de involverte lærere, lagt ned et tilsvarende arbeid for å motivere elevene som deltok i generalprøven i PISA. I hvertfall er det vanskelig å se andre umiddelbare årsaker til at forskjellen mellom de omtalte verdiene ble så vidt store.

### 8.3.1 Mopedoppgaven, PISA og "mathematical literacy"

Jeg hadde på forhånd antatt at denne oppgaven ville være relativt enkel, men det viste seg ved pilotutprøvingen at flere elever hadde problemer med å hente ut den nødvendige informasjonen fra tabellen, og følgelig besvarte oppgaven på en mangelfull måte. I tillegg var mange som nevnt svært upresise, når de skulle begrunne med egne ord hvorfor rentebeløpet stadig ble mindre. Selv om jeg mistenker flere av dem for å forstå bakgrunnen for dette, maktet de ikke å uttrykke det på en adekvat måte. Man kan selvfølgelig innvende at det her blir "*reading literacy*" jeg måler, og ikke "*mathematical literacy*". Igjen er vi da inne på problemstillinger knyttet til mye/lite oppgavetekst, og til hvorvidt det overhodet er mulig å måle "ren" matematisk kompetanse i "autentiske"oppgaver, ettersom det å uttrykke denne gjennom tilfredsstillende svar, åpenbart forutsetter både tekstforståelse og visse verbale formuleringsevner.

At denne oppgaven ble plukket ut til å være med i PISA's egen pilotutprøving, tilsier selvsagt at den holder seg innenfor de omtalte oppgavekravene. Jeg har i tillegg lyst til å påpeke at det å kunne forstå tabeller og diagrammer vel må kunne sies å bli bare viktigere og viktigere i dagens samfunn, i det vi blir presentert for den slags informasjon i stadig nye sammenhenger. Det å kunne tolke denne informasjonen, er absolutt nødvendig for å kunne holde seg orientert på helt sentrale områder i livet, og det utgjør derfor en svært viktig del av det å være en "*mathematical literate*" person og en "*reflective citizen*".

Lindenskov og Wedege(2000) har utviklet begrepet "*numeralitet*", som har mye til felles med "*mathematical literacy*". De gir følgende todelte definisjon av konstruktet:

> 1)*Numeralitet er funktionelle matematikkfærdigheter og -forståelser som alle mennesker principielt har brug for at have.*
>
> *2) Numeralitet ændrer seg med tid og sted; samfunsudvikling og teknologisk udvikling (Lindenskov og Wedege 2000, s.5).*

De påpeker hvor viktig det er at:

> *…befolkningen kan forstå og forholde seg til mediedebatten, både av hensyn til demokratiet som grunnleggende for dansk samfunnsliv og til den enkeltes liv (Ibid,s.3-4).*

Mopedoppgaven, som av PISA er omdøpt til Banklånoppgaven, knytter seg til en svært viktig del av så vel konstruktet "*numeralitet*" som "*mathematical literacy*", ved at den presenterer elevene for aspekter ved virkeligheten som de ikke kan unngå å bli konfrontert med i sitt daglige liv, og som det er viktig at de kan forholde seg til og mestre for å framstå som fullverdige samfunnsmedlemmer.

Jeg vil derfor hevde at den med hell må kunne brukes i en test hvor hensikten er å måle slike psykologiske konstrukter, og at den ***konstruktrelaterte validiteten*** til oppgaven følgelig er relativ høy.

Denne oppgaven faller, i likhet med Wimbledonoppgaven, inn under den sentrale ideen ***kvantitativt resonnement***. Det å kunne utføre *meningsfylte operasjoner* på de dataene man presenteres for, hører med til denne sentrale ideen (se kap.5.3.1), og er et viktig aspekt ved "*mathematical literacy*". For å lykkes med å besvare de ulike spørsmålene i Mopedoppgaven,vil det nettopp være nødvendig å makte dette. Følgelig vil jeg også argumentere for oppgavens ***innholdsrelaterte validitet***.

## 8.4 OPPGAVE 3

Oppgave 3 har i likhet med de to foregående også en autentisk kontekst. Den kom imidlertid ikke med blant de utvalgte oppgaver i generalprøven i PISA, muligens fordi kroppshøyde knyttet til folkeslag regnes for å være et upassende emne.

Den korte teksten som introduserer oppgaven reviderte jeg flere ganger, blant annet etter råd fra PISA - gruppa ved instituttet, nettopp for å unngå formuleringer som kunne oppfattes støtende på noen som helst slags måte. Personlig mener jeg derfor at den endelige ordlyden ble fullstendig nøytral med hensyn til dette, men det kan tenkes at man i PISA sentralt har vurdert det annerledes.

Det kan også ha vært en innvending mot denne oppgaven at Burundi tilhører et område i Afrika som i løpet av de siste årene er blitt rammet av så vel blodige etniske konflikter, som store tørkeperioder, med påfølgende hungersnød for sivilbefolkningen. Slike katastrofer har hatt som konsekvens at titusenvis av menneskeliv har gått tapt. Man kan derfor i PISA ha vurdert det slik at konteksten, i betydningen settingen til denne oppgaven, derfor var noe upassende i og med at den fokuserte på helt andre ting, og at den på en måte bekrefter matematikkens "kalde" og "livsfjerne" holdninger. Dette er en type beskyldninger som man innenfor faget sterkt ønsker å gjendrive, blant annet nettopp gjennom prioriteringen av autentiske og relevante oppgaver, og fokusering på konstrukter som "*mathematical literacy*".

Spørsmål 10 og 11 er for øvrig av typen "korte, gitte svar", som altså i prinsippet skal kunne rettes av en dataoperatør, mens spørsmål 12 er en flervalgsoppgave.

# BURUNDI

Steppefolket tutsiene, som lever i Sentral Afrika, er et av verdens høyeste folkeslag. Mange blir over 2 meter høye. Pussig nok er de naboer til pygmeene, som er verdens minste folk og svært sjelden blir over 1,40 meter høye.

I Burundi bor det mange tutsier. Landet har ca. 6,22 millioner innbyggere og av disse er 13,5 % tutsier.

**Spørsmål 10:** BURUNDI

Hvor mange tutsier bor det I Burundi ?

…………………tutsier

**KODER FOR SPØRSMÅL 10**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Riktig** | | 89 | 58,2 | |
| 10 | 0,84 millioner/839700 | 89 | 58,2 | 0,54 |
| **Ikke riktig** | | **49** | **32,1** | |
| 70 | Desimalfeil | 18 | 11,8 | -0,40 |
| 79 | Øvrige svar | 31 | 20,3 | -0,83 |
| **Ikke svart** | | **15** | **9,8** | |
| 99 | Blankt | 15 | 9,8 | -0,99 |

Vi ser av kodeskjemaet at nær 60% av elevene svarer riktig på dette spørsmålet, og at under 10% svarer blankt. Relativ enkel prosentregning er derfor åpenbart noe som elevene til en stor grad behersker. Det er dessuten grunn til å legge merke til at selv om p-verdien for denne oppgaven er forholdsvis høy, så viser verdiene i tabellen under at den diskriminerer godt.

Man ser også at elever som havner i kode 79, og som altså har begått noe annet enn en desimalfeil, er klart svakere enn de som "kun" har satt komma feil sted. De sistnevnte har altså brukt riktig algoritme, men likevel ikke kommet fram til det korrekte svaret. Sett på bakgrunn av at oppgaven er relativ enkel, at svaret de har avgitt er galt og z-verdien deres lav, er det her likevel ikke noen som helst grunn til å revurdere poenggivingen og kodingen.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Sp.m.10 | 0,58 | 0,64 | 9,8 | -0,67 | 0,54 | -0,99 |

**Spørsmål 11:** BURUNDI

Burundi har et flateinnhold på 27834 km$^2$ .

Hva er befolkningstettheten, målt i innbyggere/km$^2$?

……………… innbyggere/km$^2$

**KODER FOR SPØRSMÅL 11**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Riktig** | | **54** | **35,3** | |
| 10 | 223/224 innb./km$^2$ | 54 | 35,3 | 0,84 |
| **Ikke riktig** | | **54** | **35,3** | |
| 70 | 4475/447,5/44,75/0,4475/0,00475 innb./km$^2$ | 13 | 8,5 | 0,08 |
| 71 | Desimalfeil i forhold til riktig svar | 9 | 5,9 | -0,14 |
| 72 | 30,16 innb./km$^2$ (dividert 839700 : 27834) | 8 | 5,2 | 0,54 |
| 79 | Øvrige svar | 24 | 15,7 | -0,53 |
| **Ikke svart** | | **45** | **29,4** | |
| 99 | Blankt | 45 | 29,4 | -0,82 |

Som det framgår av kodeskjemaet for spørsmål 11, er elevene langt mer usikre på utregning knyttet til innb/km$^2$, enn prosentregning. Kun i overkant av en tredjedel har her kommet fram til riktig svar, mens nær 10% har byttet om dividend og divisor, og dermed fått feil svar. Ca. 30% av elevene har ikke besvart spørsmålet.

Elevene som har havnet i kode 72 har regnet ut hvor mange tutsier det bor i Burundi pr. km$^2$ . De har altså gjort en korrekt utregning, men tatt utgangspunkt i uriktige tall.

Z-verdien til denne gruppa er langt høyere enn for de andre feilsvarene, og det er grunn til å vurdere om jeg ikke burde endre kodene, slik at dette svaret ble poenggivende. Da vil altså 72 bli omgjort til en 10-er kode, og 10-er koden blir en 20 - kode. Man må kunne hevde at denne elevgruppa har gjort noe riktig, ettersom de har utført de korrekte regneoperasjoner, om enn på feil tall. Z-verdiene understøtter en slik endring av kodeskjemaet.

Slik ville mitt nye kodeskjema da bli:

**KODER FOR SPØRSMÅL 11 (revidert)**

| **Koder** | **Svaralternativer** | **Frekvens** | **Prosent** | **Z – verdi** |
|---|---|---|---|---|
| **Fullstendig** | | **54** | **35,3** | |
| 20 | 223/224 innb./km$^2$ | 54 | 35,3 | 0,84 |
| **Riktig** | | **8** | **5,2** | |
| 10 | 30,16 innb./km$^2$ (dividert 839700 : 27834) | 8 | 5,2 | 0,54 |
| **Ikke riktig** | | **48** | **30,1** | |
| 70 | 4475/447,5/44,75/0,4475/0,00475 innb./km$^2$ | 13 | 8,5 | 0,08 |
| 71 | Desimalfeil i forhold til riktig svar | 9 | 5,9 | -0,14 |
| 79 | Øvrige svar | 24 | 15,7 | -0,53 |
| **Ikke svart** | | **45** | **29,4** | |
| 99 | Blankt | 45 | 29,4 | -0,82 |

Verdiene i tabellen under er beregnet ut fra det ikke-reviderte kodeskjemaet til spørsmål 11. Det er først og fremst p-verdien som blir litt høyere utfra mitt nye kodeskjema, nemlig 0,38. Dessuten blir Z(0) noe lavere, og Z(2) kommer inn som egen kolonne. Jeg vil også kunne få en marginal endring i verdien for diskriminering. Vi ser for øvrig at også denne oppgaven diskriminerer bra.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spørsmål 11 | 0,35 (0,38) | 0,63 | 29,4 | -0,16 | 0,84 | -0,82 |

---

**Spørsmål 12:** BURUNDI

Hutuene er en annen stor folkegruppe i dette landet. De utgjør ca. 81%  av befolkningen.

Hva er forholdet mellom antall tutsier og hutuer? Sett en ring rundt bokstaven for det riktige alternativet.

A 1:4

B 1:5

C 1:6

D 1:8

**KODER FOR SPØRSMÅL 12**

| **Koder** | **Frekvens** | **Prosent** | **Z-verdi** |
|---|---|---|---|
| 1 (A) | 9 | 5,9 | -0,02 |
| 2 (B) | 23 | 15,0 | -0,49 |
| **3* (C)** | **81** | **52,9** | **0,37** |
| 4 (D) | 29 | 19,0 | -0,29 |
| **Ikke svart** | **11** | **7,2** | |
| 99 | 11 | 7,2 | -0,89 |

Dette er min første flervalgsoppgave i denne testen, og jeg ser umiddelbart at "ikke svart"-prosenten er langt lavere enn på de foregående oppgavene. En av de store fordelene ved flervalgsoppgaver, nemlig at svarprosenten ofte er svært høy, illustreres dermed godt.

I flervalgsoppgaver er det viktig å utvikle gode distraktorer, det vil si svaralternativer som er feil, men som kan avdekke grunnleggende misforståelser hos elevene, og slik sett gi god diagnostisk informasjon. Dersom man på et tidlig stadium piloterer oppgaven som en åpen oppgave, men senere ønsker å gjøre den om til en flervalgsoppgave, kan man benytte ofte forekommende feilsvar som slike distraktorer.

Gunvor Rand (1965) påpeker hvordan man i denne type oppgaver må sikre seg at:”*også de svarmuligheter som ikke er riktige, tiltrekker seg oppmerksomhet*” (Rand 1965, s.21).

Hun hevder at 0 - distraktorer, dvs. svar som ingen forsøkspersoner velger, like gjerne kan utelates.

Verdt å merke seg her er et alle distraktorene for spørsmål 12 er blitt benyttet, selv om de vel strengt tatt ikke alle har et ”pseudo-logisk” fundament. Det kan derfor være vanskelig å se grunner til at man her velger galt, bortsett fra at det er ren gjetning. Dette kan også anføres som en kritikk av oppgaven og de utvalgte distraktorene, altså at de ikke evner å avdekke noen grunnleggende misforståelser hos elevene. Muligens kan slike vurderinger være en medvirkende årsak til at denne oppgaven ikke ble utvalgt av PISA.

Vi ser av tabellen under at alle feilalternativene har negativ z-verdi, og at de som ikke har gjort noe forsøk på å besvare oppgaven er de desidert svakeste elevene. På tross av de ovenfor nevnte kritiske bemerkninger som kan anføres mot distraktorene, ser vi at oppgaven altså diskriminerer relativt bra, selv om verdien for diskriminering absolutt er noe lavere enn den har vært på de foregående oppgavene.

| Oppgave | P - verdi | Diskri - minering | Ikke svart | Z(alt.1) | Z(alt.2) | Z(alt3) | Z(alt.4 ) | Z(9) |
|---|---|---|---|---|---|---|---|---|
| Sp.m. 12 | 0,53 | 0,39 | 7,2 | -0,02 | -0,49 | 0,37 | -0,29 | -0,89 |

### 8.4.1 Burundioppgaven, PISA og "mathematical literacy".

Den statistiske analysen av elevsvarene til de tre spørsmålene i denne oppgaven, viser at den stort sett fungerer bra testteoretisk. Oppgaven har en autentisk kontekst, i og med at den tar utgangspunkt i biografiske og geografiske faktaopplysninger. I PISA-terminologi vil man si at konteksten er ***ekstra-matematisk***, og at situasjonen er ***utdannende.***

For å komme fram til de riktige løsningene i Burundioppgaven, kreves det at en kan håndtere utregninger med millioner i den ene faktoren, at man forstår og kan regne med benevningen innbyggere/km$^2$ , og at man kjenner til hva et forhold mellom to oppgitte størrelser innebærer. Dette er ferdigheter som krever at man forstår *operasjoners mening på tallstørrelser*, og oppgaven faller således klart inn under den sentrale ideen ***kvantitativt resonnement***. Jeg vil slik argumentere for oppgavens ***innholdsrelaterte validitet***.

De ferdighetene jeg beskriver i avsnittet over, mener jeg det er nødvendig å beherske dersom man skal gå for å være en "*mathematical literate*" person, og en "*informed and reflective citizen*" (OECD PISA 2001a, s.5).

Man kan derfor hevde at oppgaven ut fra statistiske data og ovenstående argumenter, egner seg bra for å måle enkelte sider ved "*mathematical literacy*"- konstruktet, altså at jeg også for denne oppgaven har oppnådd en betydelig grad av ***konstruktrelatert validitet***. Av en eller annen grunn kom altså likevel ikke denne oppgaven med i generalprøven i PISA, og det vil jeg gjerne knytte ytterligere noen kommentarer til.

Oppgavene som ble utviklet til PISA måtte, som jeg beskrev i kapittel 5, gjennom en rekke ”filtre” før de eventuelt ble godkjent for bruk i generalprøven. Det kan derfor synes temmelig meningsløst å spekulere i årsaker til at enkeltoppgaver ikke ble utvalgt. Likevel har jeg for Burundioppgaven lyst til å peke på to mulige avvisningsgrunner. En av dem kan være, som jeg allerede har vært inne på, og som er overveiende sannsynlig, at oppgaveteksten ble ansett som noe "touchy" og upassende. Den andre årsaken kan være knyttet til at oppgaven i tillegg muligens er litt tam og kjedelig, med relativt tradisjonelle, "skolematematiske" oppgaver. Etter å ha levd med disse oppgavene et år, og etter hvert forsøkt å se på de med et kritisk blikk, er dette en innvending mot Burundioppgaven som jeg selv mener er relevant. Det innebærer ikke at jeg er fryktelig misfornøyd med den, men jeg er tilbøyelig til å mene at den som sagt kanskje er noe ordinær og konvensjonell, selv om jeg synes selve innledningen er litt spennende og kreativ.

## 8.5 Oppgave 4

I oppgave 4, ”Sykkeltur med musikk”, består den innledende teksten av kun to relativt korte setninger. Likevel har jeg med denne introduksjonen åpenbart maktet å fange elevenes oppmerksomhet og interesse, ettersom det kun er 2% som ikke har besvart spørsmål 13. Svarprosenten på 98 er i særklasse den høyeste, alle spørsmålene i heftet tatt i betraktning. Ser man på PISA's definisjoner av ”situasjoner” og ”kontekster” (se kap.5.2), er spørsmål 13 knyttet til situasjonen ”*personlig liv*”, og med en ”*ekstra matematisk*” kontekst. Dette kan muligens være ekstra gunstige kategorier for å skape interesse og motivasjon for oppgavene blant elevene.

”Sykkeltur med musikk” kom også med i PISA's pilotutprøving, men da var kun spørsmål 15 og 17 beholdt.

## SYKKELTUR MED MUSIKK

John var i butikken og kjøpte seg to nye CD – plater. Spilletiden på den ene var oppgitt slik for de enkelte sporene (i minutter:sekunder) :

11 : 52

6 : 43

8 : 54

5 : 09

12 : 38

**Spørsmål 13:** SYKKELTUR MED MUSIKK

Hva var den totale spilletiden? Sett en ring rundt bokstaven for det riktige alternativet.

A 44:36

B 45:26

C 45:16

D 43:96

**KODER FOR SPØRSMÅL 13**

| Kode | Frekvens | Prosent |
|---|---|---|
| 1 | 13 | 8,5 |
| 2 | 27 | 17,6 |
| **3*** | **48** | **31,4** |
| 4 | 62 | 40,5 |
| **Ikke svart** | **3** | **2,0** |
| 99 | 3 | 2,0 |

Denne oppgaven var i utgangspunktet en åpen oppgave. På bakgrunn av de elevsvarene jeg mottok i min første pilotering i egen klasse, endret jeg den til en flervalgsoppgave. Jeg brukte da enkelte av de mest forekommende elevsvarene som distraktorer. Imidlertid satt jeg kun igjen med to distraktorer i tillegg til det korrekte svaret, slik at jeg selv måtte konstruere det siste alternativet, ettersom det minimum bør være tre distraktorer. Det var derfor knyttet en viss spenning til om dette konstruerte svaralternativet ville makte å lokke til seg noen "proselytter" i den videre pilotutprøving. Som man ser av tabellen har alle distraktorene blitt tatt i bruk.

Det er for øvrig verdt å legge merke til at det er alternativ D(4) som får størst tilslutning blant elevene. Dette resultatet får man dersom man summerer tidsangivelsene etter ordinært titalls mønster, uten å ta i betraktning at det er minutter og sekunder man har med å gjøre.

Bortimot 40 prosent av elevene avga dette svaret, og vi ser at z-verdien for denne koden faktisk er den laveste (bortsett fra Z(9)). Det er altså først og fremst den svakere delen av elevene som begår denne klassiske feilen.

Vi ser for øvrig at elever som har gitt sin tilslutning til alternativ B(2) har en klart positiv z-score, noe som kan forklares ved at disse elevene for så vidt har gjort noe riktig. Dette svaret tilsier nemlig at elevene har utført summeringen av minutter og sekunder separat, altså på en korrekt måte, og at de deretter har dividert antall sekunder med 60. Til slutt har de imidlertid oppgitt "resten" som sekunder, og dermed fått feil svar. De har derfor delvis benyttet riktig algoritme. Som man ser av z-verdien, er dette elever som skårer vesentlig høyere enn elever som har oppgitt andre feilsvar.

| Oppgave | P - verdi | Diskri - minering | Ikke svart | Z(alt.1) | Z(alt.2) | Z(alt.3) | Z(alt.4 ) | Z(9) |
|---|---|---|---|---|---|---|---|---|
| Spørsmål 13 | 0,31 | 0,35 | 2,0 | -0,20 | 0,15 | 0,52 | -0,39 | -0,86 |

**Spørsmål 14:** SYKKELTUR MED MUSIKK

På den andre CD – platen var oppgitt spilletid 48 min. John ville gjerne høre hele plata på sin Discman, mens han syklet de 12 km hjem fra butikken.

Hvilken gjennomsnittsfart målt i km/h måtte han holde for å være hjemme akkurat når siste låt var slutt?

……………km/h

**KODER FOR SPØRSMÅL 14**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Riktig** | | **31** | **20,3** | |
| 10 | 15 km/h | 31 | 20,3 | 1,15 |
| **Ikke riktig** | | **92** | **60,1** | |
| 70 | 25 km/h evt. 0,25 km/h (12 : 48) | 21 | 13,7 | 0,07 |
| 71 | 4 km/h (48 : 12) | 30 | 19,6 | -0,42 |
| 79 | Øvrige svar | 41 | 26,9 | -0,23 |
| **Ikke svart** | | **30** | **19,6** | |
| 99 | Blankt | 30 | 19,6 | -0,51 |

Av tabellene ser vi at p-verdien her er relativ lav, men at oppgaven diskriminerer svært godt. Z-verdiene avslører at de som har svart riktig er vesentlig flinkere enn de som har svart uriktig, eventuelt ikke svart. Dette viser at oppgaven må karakteriseres som vanskelig.

Kode 70 skiller seg noe ut blant feilsvarene, i og med at elevene i denne kategorien har fått en høyere z-verdi enn de andre. Svaret framkommer ved at man dividerer 12 med 48, hvilket selvsagt er noe mer fornuftig enn bare å dividere det største tallet med det minste, som mange elever her har gjort. Det kan likevel helt klart ikke gi poeng, ettersom det er direkte galt.

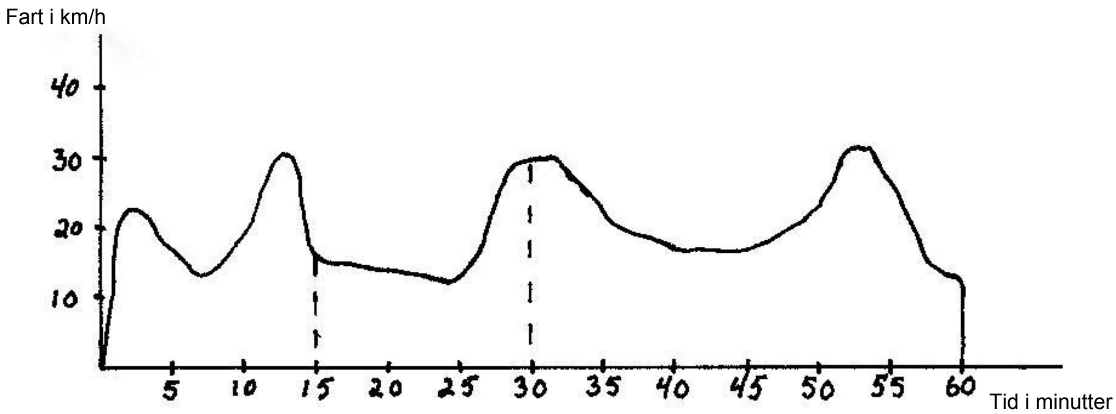
Det kunne vært interessant å gjøre om også denne oppgaven til en flervalgsoppgave. Aktuelle distraktorer hadde da vært 4km/h og 25km/h. Ved en slik transformasjon ville svarprosenten sannsynligvis blitt vesentlig høyere. Hvorvidt p-verdien også ville blitt høyere kan selvsagt ikke med sikkerhet avgjøres, men det er grunn til å legge merke til at hele 26,9% av elevene har avgitt svar som har havnet i kategori 79, øvrige svar. Olsen et.al.(2001) påpeker at dersom elevene i en flervalgsoppgave kun har 4-5 alternativer å forholde seg til, og ikke finner sitt eget svar blant disse, vil de ofte forsøke å utføre en ny utregning, for å komme opp med et nytt svar som stemmer overens med et av de gitte alternativene.

*Theoretically, the p-value (the percentage of students answering the item correctly) for the item should, for that reason, be somewhat higher..(for the multiple choice version)..than for the constructed response version (Olsen et.al. 2001, s.408).*

Det er derfor all grunn til å anta at jeg ved å endre oppgaveformatet, fra en åpen til en flervalgsoppgave, ville opplevd at p-verdien her hadde blitt høyere.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Spørsmål 14 | 0,20 | 0,58 | 19,6 | -0,22 | 1,15 | -0,51 |

Etter at han var kommet hjem tegnet John følgende graf for å illustrere denne sykkelturen :




---

**Spørsmål 15:** SYKKELTUR MED MUSIKK

Studer utsnittet av grafen som befinner seg mellom de stiplede loddrette linjene, altså mellom det 15. og 30. minuttet av Johns sykkeltur.

Hvilket av utsagnene nedenfor synes du best beskriver den veistrekningen John da sannsynligvis syklet. Sett en ring rundt bokstaven for det beste alternativet.

A John syklet først en slak nedoverbakke og så en bratt oppoverbakke.

B John syklet først en slak oppoverbakke og så en bratt nedoverbakke.

C John syklet først et flatt parti og så en bratt oppoverbakke.

D John syklet først rett fram, så svingte han brått til venstre, fortsatte i den retningen et stykke før han så svingte brått til høyre.

**KODER FOR SPØRSMÅL 15 (PISA GENERALPRØVEN I PARENTES)**

| Koder | Frekvens | Prosent | Z -verdi |
|---|---|---|---|
| 1 | 47 (61) | 30,7 (26,1) | -0,38 (-0,11) |
| **2*** | **85 (87)** | **55,6 (37,2)** | **0,44 (0,54)** |
| 3 | 10 (41) | 6,5 (17,5) | -0,66 (-0,30) |
| 4 | 2 (6) | 1,3 (2,6) | -0,73 (-0,86) |
| **Ikke svart** | **9 (39)** | **5,7 (16,7)** | |
| 99 | 9 (39) | 5,7 (16,7) | -1,25 (-0,57) |

Av tabellene ser vi at svarprosenten for denne oppgaven, som altså er en flervalgsoppgave, er vesentlig høyere enn vi har observert på de åpne oppgavene.

Også her har alle distraktorene blitt benyttet, men kun to elever har gitt sin tilslutning til alternativ D(4). Jeg har i denne oppgaven ikke benyttet tidligere framkommede elevsvar som distraktorer, da jeg i utgangspunktet allerede hadde bestemt meg for her å lage en flervalgsoppgave. Distraktorene er således egenproduserte. Det var imidlertid svært vanskelig å finne mer enn to gode distraktorer, slik at den siste vel var noe søkt.

Denne oppgaven ble beholdt uendret i generalprøven i PISA, og jeg registrerte da at de også hadde beholdt de samme distraktorene. Hvorvidt dette blir endret dersom oppgaven skulle komme med i den endelige hovedundersøkelsen, blir spennende å se.

Alle distraktorene har fått negativ z-verdi, mens det ut fra Z(9) er åpenbart at det er de aller svakeste elevene som ikke har besvart denne oppgaven.

Over halvparten av elevene har i min pilotundersøkelse svart riktig, og denne gruppa har samtidig den klart høyeste z-verdien. Oppgaven diskriminerer fint, og alle verdiene indikerer at den fungerer godt psykometrisk.

Av verdiene fra PISA G, ser vi at det igjen er ikke svart-prosenten som er mest bemerkelsesverdig. Det er vanskelig å finne noen god forklaring på at denne verdien skal være nærmere tre ganger så høy som den tilsvarende verdien i mitt utvalg, bortsett fra manglende motivasjon og motivasjonsarbeide.

Vi ser også at Z(9)-verdien er relativt avvikende. Dette må oppfattes som en konsekvens av at denne gruppa her utgjøres av langt flere elever, noe som trekker z-verdien i retning 0. For de øvrige z-verdiene, er det bare moderate forskjeller.

| Sp.m.15 | P - verdi | Diskri - minering | Ikke svart | Z(alt.1) | Z(alt.2) | Z(alt.3) | Z(alt.4 ) | Z(9) |
|---|---|---|---|---|---|---|---|---|

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Min pil. | 0,56 | 0,50 | 5,7 | -0,39 | 0,44 | -0,66 | -0,73 | -1,25 |
| **PISA G.** | **0,37** | | **16,7** | **-0,11** | **0,54** | **-0,30** | **-0,86** | **-0,59** |

---

**Spørsmål 16:** SYKKELTUR MED MUSIKK

Hva ble den virkelige gjennomsnittshastigheten for hele sykkelturen målt i km/h ?

…………km/h

**KODER FOR OPPGAVE 16**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **24** | **15,7** | |
| 20 | 12 km/h | 24 | 15,7 | 1,23 |
| **Delvis riktig** | | **34** | **22,2** | |
| 10 | 14 – 21 km/h | 34 | 22,2 | 0,27 |
| **Ikke riktig** | | **39** | **25,6** | |
| 70 | 5 km/h (60min. : 12 km) | 6 | 3,9 | 0,51 |
| 79 | Øvrige svar | 33 | 21,7 | -0,54 |
| **Ikke svart** | | **56** | **36,6** | |
| 99 | Blankt | 56 | 36,6 | -0,43 |

Det mest bemerkelsesverdige i denne kodeguiden er, at z-verdien for de som har avgitt feilsvaret 5 km/h er unormalt høy til å være en feilkategori. For å undersøke dette nærmere, har jeg gått gjennom elevbesvarelsene til de elevene som har avgitt dette svaret på nytt. Slik ønsket jeg å se om jeg kunne avdekke noen interessante sammenhenger. Jeg fant at 5 av disse 6 elevene hadde svært høy totalsum på testen, og gode standpunktkarakterer i matematikk (4,5,5,5,6), mens den siste eleven hadde en meget lav sammenlagt score. Tre av de fem flinke elevene hadde også svart feil på spørsmål 14 på en slik måte at det var konsistent med den feilen de her har begått, nemlig å dividere minutter med kilometer for å finne hastigheten. Kun en elev hadde svart riktig på spørsmål 14 og feil på spørsmål 16.

Konklusjonen for mitt vedkommende ble at denne z-verdien skyldes en ren tilfeldighet, som generelt har å gjøre med at utvalget mitt av elever rent antallmessig er lavt, og i dette tilfellet spesielt er forårsaket av at denne svarkategorien kun består av seks elever. Da kan man risikere å få slike anormale z-verdier på gale svar.

Ellers legger jeg merke til at p-verdien er relativt lav, og ”ikke-svart”- prosenten høy. Når jeg i tillegg registrerer at z-verdien for riktig svar er så vidt høy, trekker jeg den slutning at spørsmålet nok ble oppfattet som noe vanskelig.

Verdiene i tabellen under indikerer at oppgaven likevel fungerer bra psykometrisk, og at den blant annet diskriminerer godt.

Spørsmål 16 ville også med letthet kunne blitt omformet til en flervalgsoppgave. Jeg har i kodeskjemaet tilkjennegitt at svarene i ”delvis riktig”- kategorien hadde et relativt stort sprik, og at både den og ”øvrige svar”- gruppa hver for seg rommer over 20% av utvalget. I tillegg er ”ikke svart”- prosenten altså høy. Det er derfor grunn til å tro at jeg ved en slik transformasjon av formatet, noe som blant annet innebærer et begrenset antall svar - alternativer, ville kunne oppnådd helt andre statistiske verdier, som høyere p-verdi og lavere ”ikke svart”- prosent.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Sp.m. 16 | 0,27 | 0,60 | 36,6 | -0,38 | 0,27 | 1,24 | -0,43 |

**Spørsmål 17:** SYKKELTUR MED MUSIKK

Vil du ut fra grafen anta at John bor på toppen eller bunnen av en lang bakke ? Begrunn svaret.
…………………………………………………………………………………………………

**KODER FOR SPØRSMÅL 17 (PISA GENERALPRØVEN I PARENTES)**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **60 (37)** | **39,2 (15,8)** | |
| 20 | Toppen fordi det ifølge grafen går saktere mot slutten | 53 | 34,6 | 0,78 |
| 21 | Toppen fordi han begynner å sykle nedover evt. fordi farten raskt blir høy i starten | 7 | 4,6 | 0,27 |
| **Delvis riktig** | | **25** | **16,3** | |
| 10 | Toppen med uklar begrunnelse for eksempel:<br>a) fordi han hadde lav hastighet<br>b) fordi han sykler fortere jo høyere han kommer | 25 | 16,3 | -0,21 |
| **Ikke riktig** | | **52 (144)** | **34,1 (61,5)** | |
| 70 | Han bor "midt imellom", han avslutter midt i en nedoverbakke | 3 | 2,0 | -1,01 |
| 71 | Bunnen fordi grafen går rett nedover mot slutten | 24 | 15,7 | -0,27 |
| 72 | Bunnen med annen forklaring evt. uten forklaring | 16 | 10,5 | -0,66 |
| 79 | Øvrige svar | 9 | 5,9 | -0,51 |
| **Ikke riktig** | | **16** | **10,5 (22,6)** | |
| 99 | Blankt | 16 | 10,5 | -0,85 |

Det er særlig et par ting man legger merke til ved å studere kodeskjemaet. For det første er det relativt stor forskjell på z-verdiene mellom de som har fått kode 20, og de som har fått kode 21.

Jeg var lenge i tvil om hvorvidt jeg burde belønne elevene som har besvart dette spørsmålet med å henvise til åpningsfarten, med to poeng. Etter å ha konsultert personer i PISA - gruppa ved instituttet, fant vi i fellesskap ut at oppgaveteksten var noe uklar med hensyn på om grafen illustrerte kun hjemturen fra butikken, eller tur/retur butikken, og at den derfor kunne oppfattes på begge måter. I sistnevnte tilfelle er svaret med kode 21 faktisk korrekt. I ettertid ser vi at det altså er over et halvt standardavvik i forskjell på gjennomsnittlig totalscore for disse to gruppae.

Når vi samtidig registrerer at Z(1) er negativ, kunne det her være fristende å omgjøre kodeskjemaet slik at kode 21 ble en 10-er kode, og kode 10 en 70-kode. Z-verdiene ville da i større grad understøtte kodeskjemaet og poenggivningen, slik at denne var mer i overensstemmelse med testteoretiske prinsipper. På bakgrunn av at teksten var noe tvetydig, har jeg imidlertid valgt å beholde kodingen som vist. Denne omtalte uklarheten i oppgaveteksten ble for øvrig gjenstand for min oppmerksomhet allerede under avviklingen av min egen pilotundersøkelse, slik at jeg rakk å korrigere teksten før oppgaven ble innsendt til PISA.

De øvrige verdiene i tabellen under indikerer at oppgaven psykometrisk sett fungerer godt.

Igjen er andelen elever som ikke har besvart spørsmålet langt høyere i PISA G enn i mitt materiale. Vi ser i tillegg at p-verdien også er vesentlig lavere. Noe av forklaringen skyldes at kodeguidene ikke er helt sammenfallende, men selv om jeg korrigerer for dette, får jeg en langt høyere p-verdi i mitt utvalg (0,35). Det er vanskelig å se noen god grunn til at forskjellen skulle bli så vidt stor, men avviket er altså betydelig. Z(0) blir høyere i PISA G, ettersom en langt større del av elevene havner i denne kategorien.

| Sp.m.17 | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Min pil. | 0,47 | 0,59 | 10,5 | -0,48 | -0,21 | 0,72 | -0,84 |
| **PISA G**. | (0,16) | | 22,6 | (0,01) | | **0,73** | **-0,54** |

I og med at kodeskjemaene som nevnt ikke er helt sammenfallende, er ikke alle verdiene direkte sammenlignbare. Jeg har valgt å sette de ikke - sammenlignbare verdiene fra PISA G i parentes.

### *8.5.1 Sykkeltur med musikkoppgaven, PISA og "mathematical literacy".*

Som nevnt i presentasjonen av denne oppgaven, kom den i en noe redusert utgave med i generalprøven i PISA. Det var da kun de spørsmålene som direkte var koblet til sykkelturen og diagrammet som var beholdt.

Koblingen mellom musikk og sykkeltur falt muligens ikke helt i smak i PISA sentralt. Her må det imidlertid tilføyes at denne oppgaven omfatter hele fem spørsmål, mens oppgavene i PISA's generalprøve for det meste hadde mellom ett og tre. At "sykkeltur med musikkoppgaven" ble avkortet i PISA's versjon, impliserer derfor ikke nødvendigvis at de resterende spørsmålene ikke tilfredsstilte kravene i MID og DFM. Jeg vil likevel gjerne påpeke et par poeng knyttet til de ikke-utvalgte spørsmål.

Spørsmål 13 kan egentlig "stå på egne ben", i den forstand at det faktisk etter min mening er en komplett oppgave knyttet til utregning av tid. Av alle spørsmålene i mitt hefte som ikke kom med i generalprøven, er jeg mest ”misfornøyd” med at ikke dette kom gjennom nåløyet. Når dette er sagt, er jeg selvfølgelig klar over alle de ulike utvalgskriteriene man har i PISA (se kap.6.2), og at det at oppgaven ble avvist ikke nødvendigvis innebærer at den ble regnet som svak. Blant annet dekker jo en del av de innsendte oppgaver samme tema.

Poenget mitt er snarere at spørsmål 13 måler et aspekt ved ”*mathemetical literacy*” som helt udiskutabelt er svært sentralt og viktig, nemlig elevenes evner til å utføre tidsberegninger. Den ***konstruktrelaterte validiteten*** til dette spørsmålet må derfor kunne sies å være relativ høy.

I tillegg var dette, som jeg nevnte i innledningen til oppgaven, et spørsmål som oppnådde en svært høy svarprosent, og som derfor åpenbart appellerte til elevene.

Derimot er jeg tilbøyelig til å mene at spørsmål 14 har fått en noe uheldig formulering, og at spørsmålet framstår som noe oppkonstruert og kunstig. Det er heller uvanlig at man avpasser farten på syklingen etter lengden på en CD, slik som spørsmålsstillingen forutsetter. I ettertid ser jeg at jeg heller burde ha oppgitt gjennomsnittshastigheten til syklisten, for deretter å spørre om han kom hjem før, eller etter at CD-en var ferdigspilt. Da hadde spørsmålet unngått å bli rammet av følgende kritikk:

> *This can be contrasted with problems frequently seen in school mathematics texts where the main purpose is to practise the mathematics involved, rather than to use mathematics to solve a real problem (OECD PISA 2001a, s.12).*

Spørsmål 16 er også beheftet med en mangel. For å svare riktig her, må man bruke informasjon gitt i spørsmål 14, nemlig at det for John var 12 km fra butikken og hjem. I §59 i MID står det imidlertid, som jeg tidligere har vært inne på i kapittel 5:

> *The order of presentation of information is an… important matter. In general PISA items should follow these principles: first the stimulus, then an introduction with the information for the question as close as possible to the question itself.* ***If the item has several questions, information specifically for each question should be given just before that question*** *(OECD PISA 2000, s.12, min utheving).*

Det hadde selvsagt vært relativt enkelt å gjøre disse endringene, og oppgaven som helhet ville da avgjort vært mer i overensstemmelse med kravene formulert i MID og DMF.

Jeg vil i tillegg nevne en kritisk bemerkning jeg mottok av en observant elev som var med i pilotundersøkelsen. Han bemerket tørt at dersom kurven til diagrammet i spørsmål 15 var riktig, og avstanden mellom butikken og hjemmet kun var 12 km, ville syklisten bruke langt mindre enn en time på hjemturen. Utfra diagrammet var jo gjennomsnittshastigheten nærmere 20 km/h, mente han. Eleven hadde selvsagt rett, og jeg måtte revidere diagrammet før oppgavene ble innsendt til PISA. Jeg tar med historien for å illustrere at årvåkenhet og nøyaktighet er svært nyttige og nødvendige personlige egenskaper å ha, når man driver med oppgaveutvikling.

Som jeg påpekte under drøftingen av Mopedoppgaven, er det å kunne tolke tabeller og diagrammer en særdeles viktig del av det å være en ”*mathematical literate*” person. Den type informasjon er noe man til stadighet presenteres for, særlig gjennom media. En person som ikke er i stand til å vurdere denne informasjonen, vil i utallige sammenhenger heller ikke kunne gjøre seg opp egne meninger, og vil derfor ikke framstå som en fullverdig, deltakende samfunnsborger. Dermed vil vedkommende også kunne få problemer med å ta vare på sine rettigheter og utføre sine demokratiske plikter på en akseptabel måte.

På bakgrunn av en slik argumentasjon mener jeg derfor, at den ***konstruktrelaterte validiteten*** til oppgaven som helhet er høy.

De spørsmålene som er med i PISA's pilotundersøkelse, dvs. spørsmål 15 og 17, er av PISA plassert i kategorien "***forandring og sammenheng***". De øvrige spørsmålene faller også klart inn under denne sentrale ideen. Oppgaven sirkler jo rundt begrepet hastighet, og elevene må forstå, og kunne uttrykke, sammenhengen mellom anvendt tid og tilbakelagt distanse, for å besvare spørsmålene korrekt.

I DMF sies blant annet dette om denne kategorien:

> *Change and relationships involves functional thinking. For 15-year-olds this includes that students have a notion of rate of change,………and dependence of one variable on another. They should be able to make judgements about "how fast" processes are taking place, also in a relative way (OECD PISA 2001a, s.39).*

Sykkeltur med musikkoppgaven må kunne sies å være nært knyttet til denne sentrale ideen. På tross av at et par av spørsmålene med fordel kunne vært omformulert, og i sin nåværende form må sies å redusere oppgavens ***innholdsrelaterte validitet***, kan det likevel argumenteres for at den ***innholdsrelaterte validiteten*** til store deler av oppgaven som helhet er relativ høy.

## 8.6 Oppgave 5

Oppgave 5 er bygd opp rundt en situasjon mange elever kjenner til, det å gå med aviser. Elevene skal her blant annet vise at de forstår begrepene fastlønn/timelønn og forholdet mellom tidsbruk/timelønn, ved å erklære seg enig/uenig i tre oppgitte verbale utsagn som omhandler disse temaene. Deretter skal de gi sin tilslutning til en av fire grafiske framstillinger av forholdet mellom anvendt tid på avisruta, og oppnådd timelønn.

I tillegg til dette, blir elevene bedt om å løse to mer tradisjonelle utregningsoppgaver.

Jeg vil gjøre oppmerksom på at jeg har kalt de tre utsagnene i spørsmål 19 for henholdsvis 19a, 19b og 19c, og at jeg har valgt å kommentere tabellene med de ulike statistiske verdiene for disse samlet, ettersom de henger så tett sammen temamessig. Disse kommentarene er plassert umiddelbart etter tabellene for spørsmål 19c.

# AVISBUD

Maria går med aviser alle dager bortsett fra lørdag og søndag og tjener 400 kr i uka. Hun bruker daglig 1 time på ruta si.

**Spørsmål 18:** AVISBUD

Hva blir timelønna ? Vis utregning.

**KODER FOR SPØRSMÅL 18**

| Koder | Svaralternativer | Frekvens | Prosent | Z -verdi |
|---|---|---|---|---|
| **Fullstendig** | | **131** | **85,6** | |
| 20 | 80 kr med vist utregning | 131 | 85,6 | 0,15 |
| **Delvis riktig** | | **9** | **6,0** | |
| 10 | 80 kr uten å vise utregning | 3 | 2,0 | -0,82 |
| 11 | 57 kr (dividerer med 7) | 5 | 3,3 | -0,09 |
| 12 | 66 kr (dividerer med 6) | 1 | 0,7 | -1,01 |
| **Ikke riktig** | | **9** | **5,9** | |
| 79 | Øvrige svar | 9 | 5,9 | -1,08 |
| **Ikke svart** | | **4** | **2,5** | |
| 99 | Blankt | 4 | 2,5 | -1,52 |

Det jeg her først og fremst legger merke til, er at denne oppgaven har fått en p-verdi på hele 0,89. Dette vil altså testteoretisk være en "dyr" oppgave, i den forstand at den kun vil diskriminere mellom det aller svakeste segmentet av elever. Et argument for å ta med slike oppgaver i en test, er at enkle oppgaver kan øke elevers motivasjon til å gi seg i kast med også vanskeligere materiale. Opplevelsen av at man i det minste lykkes med noe, kan altså slik sett få positive ringvirkninger. Den endelige avgjørelsen om hvilke oppgaver man vil ha med i en test, må derfor også inkludere denne type vurderinger.

Som vi har sett i kapittel 6, legger man i PISA vekt på å ha med oppgaver innenfor et bredt spekter av p-verdier, fordi man i tillegg til å rapportere resultatene ved hjelp av måleskalaer, har som mål å kunne gi en beskrivelse av hvilke kompetanser elevene innehar (Olsen 2000). Representanter fra PISA - gruppa ved instituttet, som har deltatt på oppgavekonferanser i PISA - regi, understreker at man faktisk etterspør spesielt oppgaver med høye p-verdier, fordi man i utvalget (poolen) av oppgaver, har få som hører til denne kategorien. At dette spørsmålet i min pilotundersøkelse har fått en høy p-verdi, er derfor neppe årsaken til at denne oppgaven ikke kom med i generalprøven i PISA.

Vi ser at kategoriene under "delvis riktig" har samlet svært få elever, og at jeg derfor burde slå sammen disse til en kode. Av tabellen under ser vi at z-verdien for denne grupperingen samlet sett er blitt så vidt lav som -0,44. Det kan gi grobunn for kritikk mot min beslutning om å belønne disse svarene med poeng. Etter min oppfatning har elevene her imidlertid utført de riktige regneoperasjoner, og uttrykt forståelse for hvilken algoritme som var den riktige å benytte, men begått små feil på veien fram til avgitt svar. Dette er bakgrunnen for min poenggiving.

At så mange lykkes i å løse oppgaven, gjør sitt til at vi her har fått en påfallende lav verdi for Z(2).

Tatt i betraktning av at dette ikke er en flervalgsoppgave, er "ikke svart"- prosenten svært lav, kun 2,6. Dette henger åpenbart sammen med at oppgaven må kunne betegnes som enkel.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 18 | 0,89 | 0,40 | 2,6 | -1,07 | -0,44 | 0,15 | -1,52 |

---

**Spørsmål 19:** AVISBUD

Nedenfor følger noen påstander. Vis om du er enig eller uenig ved å sette en ring rundt 'enig' eller 'uenig' for hver av påstandene.

| **Påstand** | |
|---|---|
| Dersom Maria bruker lenger tid på avisruta, vil timelønna hennes øke. | Enig / Uenig |
| Lønna til Maria er uavhengig av hvor lang tid hun bruker på avisruta. | Enig / Uenig |
| Dersom Maria bruker kortere tid på avisruta, vil timelønna hennes bli lavere. | Enig / Uenig |

**KODER FOR SPØRSMÅL 19A**

| Koder | Frekvens | Prosent | Z-verdi |
|---|---|---|---|
| 1 | 30 | 19,6 | -0,58 |
| **2*** | **115** | **75,2** | **0,21** |
| **Ikke svart** | **8** | **5,2** | |
| 99 | 8 | 5,2 | -0,80 |

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Sp.m.19A | 0,75 | 0,36 | 5,2 | -0,58 | 0,21 | -0,80 |

**KODER FOR SPØRSMÅL 19B**

| Koder | Frekvens | Prosent | Z-verdi |
|---|---|---|---|
| **1*** | **98** | **64,1** | **0,30** |
| 2 | 48 | 31,4 | -0,48 |
| **Ikke svart** | **7** | **4,6** | |
| 99 | 7 | 4,6 | -0,93 |

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Sp.m.19B | 0,64 | 0,41 | 4,6 | -0,48 | 0,30 | -0,93 |

**KODER FOR SPØRSMÅL 19C**

| Koder | Frekvens | Prosent | Z-verdi |
|---|---|---|---|
| 1 | 31 | 20,3 | -0,53 |
| **2*** | **113** | **73,9** | **0,21** |
| **Ikke svart** | **9** | **5,9** | |
| 99 | 9 | 5,9 | -0,81 |

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(9) |
|---|---|---|---|---|---|---|
| Sp.m.19C | 0,74 | 0,35 | 5,9 | -0,53 | 0,21 | -0,81 |

Det første som slår en ved gjennomgangen av tabellene over, er at verdiene som framkommer for de tre deloppgavene er svært like.

Vi ser at p-verdiene er forholdsvis høye. Dette skyldes nok at spørsmålene er relativt enkle. I tillegg er dette en type flervalgsoppgave hvor man kun velger mellom to mulige alternativer, og dermed har 50% sjanse for å få riktig svar. (Sannsynligheten for å få riktig svar på alle tre spørsmålene ved en tilfeldighet, er selvsagt mindre.)

Ettersom p-verdiene er så vidt høye, blir Z(1) verdiene her forholdsvis lave. Vi ser likevel at oppgaven diskriminerer bra, og at de få elevene som ikke avgir svar er de desidert svakeste. "Ikke svart"- prosenten er for øvrig lav, kun ca.5,0.

---

**20:** AVISBUD

Maria ønsker seg en timelønn på 100 kr.

Hvor lang tid kan hun da bruke på avisruta si pr. dag? Vis hvordan du regner.

**KODER FOR SPØRSMÅL 20**

| **Koder** | **Svaralternativer** | **Frekvens** | **Prosent** | **Z -verdi** |
|---|---|---|---|---|
| **Fullstendig** | | **29** | **19,0** | |
| 20 | 48 min. | 29 | 19,0 | 1,29 |
| **Delvis riktig** | | **8** | **5,3** | |
| 10 | 48 min. uten utregning | 1 | 0,7 | 0,91 |
| 11 | 4 timer i uka | 2 | 1,3 | -0,11 |
| 12 | 45 min. | 5 | 3,3 | 0,73 |
| **Ikke riktig** | | **66** | **43,4** | |
| 70 | 80 min. | 5 | 3,3 | -0,22 |
| 71 | 72min./75min./1t15.min | 25 | 16,3 | -0,02 |
| 79 | Øvrige svar | 36 | 23,5 | -0,40 |
| **Ikke svart** | | **44** | **28,8** | |
| 99 | Blankt | 44 | 28,8 | -0,48 |

Av kodeskjemaet ser vi at kun 19% av elevene har svart fullstendig riktig på dette spørsmålet. Ut fra den høye verdien til Z(2), kan jeg slå fast at dette er svært flinke elever. 8

elever har svart delvis riktig, og det naturlige ville da være å plassere disse under en felles kode. Jeg har imidlertid valgt å beholde de kategoriene jeg opprinnelig utarbeidet, for derigjennom å framvise noe av progresjonen i kodearbeidet, men er altså likevel klar over at de strengt tatt burde slås sammen.

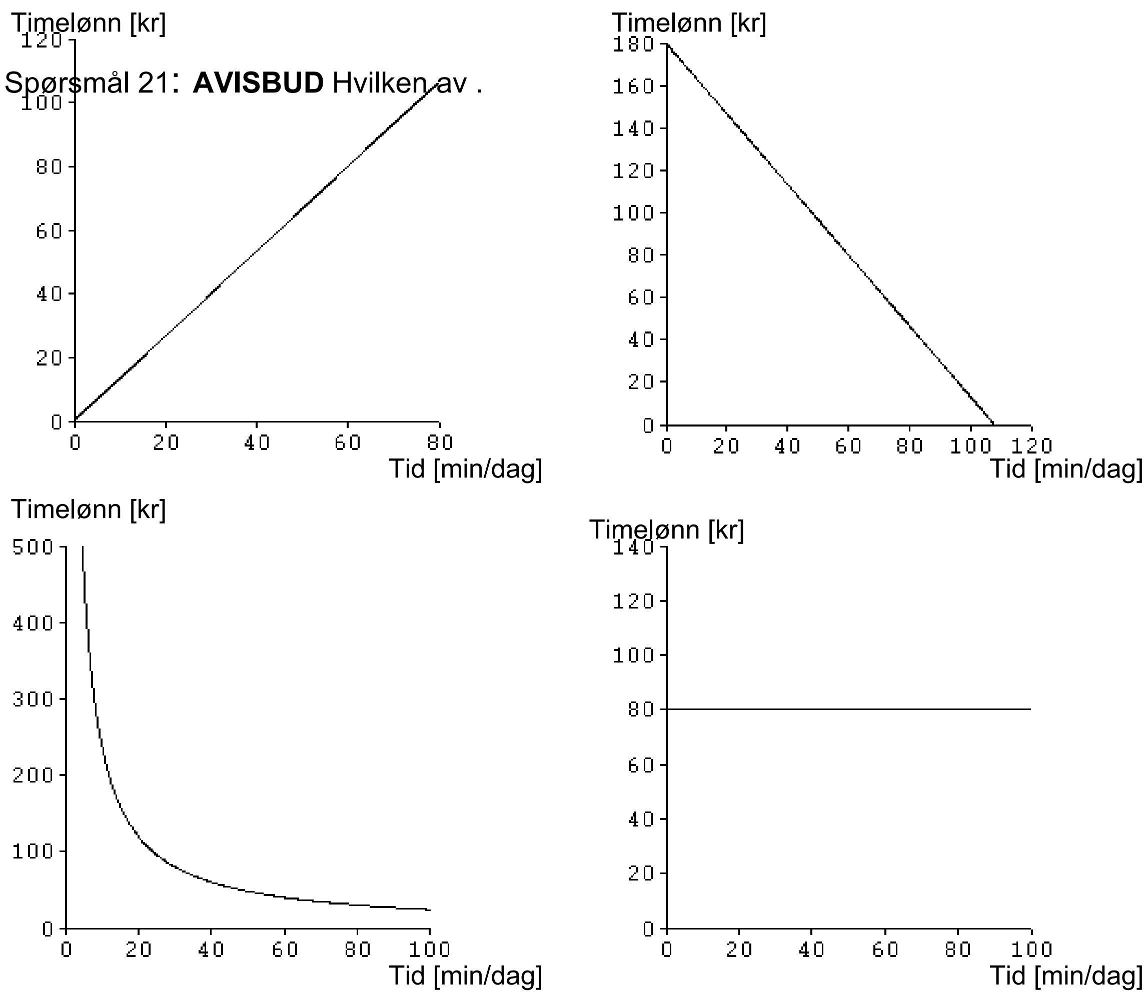
Jeg registrerer forøvrig at oppgaven diskriminerer godt, og at de statistiske verdiene samlet sett viser at oppgaven fungerer bra psykometrisk. (Det kan muligens innvendes at p-verdien er noe lav.)

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Sp.m. 20 | 0,22 | 0,66 | 28,8 | -0,24 | 0,54 | 1,29 | -0,48 |

## SPØRSMÅL 21 : AVISBUD

Hvilken av grafene nedenfor synes du best illustrerer forholdet mellom Marias daglige tidsbruk på avisruta og timelønna hun oppnår.

Vis ditt svar ved å sette en ring rundt bokstaven ved siden av den grafen du mener er riktig.



**KODER FOR SPØRSMÅL 21**

| Koder | Frekvens | Prosent | Z-verdi |
|---|---|---|---|
| 1 | 57 | 37,3 | 0,01 |
| 2 | 22 | 14,4 | 0,53 |
| **3*** | **8** | **5,2** | **0,45** |
| 4 | 40 | 26,1 | 0,08 |
| **Ikke svart** | **20** | **13,1** | |
| 99 | 20 | 13,1 | -0,81 |

Av tabellene kan vi se at denne oppgaven skiller seg radikalt ut fra de som jeg inntil nå har analysert. P-verdien er meget lav, kun 0,05, hvilket alene tilsier at spørsmålet egner seg dårlig i testsammenheng. Testteoretisk er det enda mer uheldig at z-verdien for det riktige svaret, alternativ 3, er lavere enn for alternativ 2, som altså er et galt svar. Z-verdiene for alle feilsvarene er for øvrig positive, noe som selvsagt heller ikke er ønskelig. Disse faktorene medvirker igjen til at verdien for diskriminering blir eksepsjonelt lav.

En nærliggende forklaring til at så få elever svarte riktig på dette spørsmålet, er at grafer og funksjoner utgjør et nokså begrenset emne på ungdomsskolen her til lands. På dette trinnet er elevene først og fremst blitt kjent med lineære funksjoner/grafer. De blir dessuten så vidt introdusert for parabler. Hyperbler, asymptoter og grenseverdier har de ingen erfaring med. Sett på bakgrunn av disse fakta, er det ikke så overraskende at p-verdien for dette spørsmålet ble så vidt lav. Som ansvarlig for oppgaveteksten vil jeg likevel understreke, at jeg nok hadde forventet at flere av de "flinke" elevene ville innsett at alternativene 1,2 og 4 åpenbart måtte være gale, og følgelig kunne elimineres, og at det riktige svaret, alternativ 3, derfor var det eneste mulige. Denne antagelsen slo dog definitivt ikke til.

Totalt sett blir derfor konklusjonen at spørsmål 21 ikke passer inn i denne ”testen”, gitt mitt utvalg av elever.

| Oppgave | P - verdi | Diskri - minering | Ikke svart | Z(alt.1) | Z(alt.2) | Z(alt.3) | Z(alt.4 ) | Z(9) |
|---|---|---|---|---|---|---|---|---|
| Sp.m.21 | 0,05 | 0,11 | 13,1 | 0,01 | 0,53 | 0,45 | 0,08 | -0,81 |

### *8.6.1 Avisbudoppgaven, PISA og "mathematical literacy".*

Denne oppgaven består av fire delspørsmål. Vi har sett at spørsmål 21, ut fra de statistiske dataene, egner seg dårlig som testoppgave for mitt utvalg.

Et interessant poeng er her at spørsmål 19 og 21 henger nøye sammen. I spørsmål 19 uttrykkes sammenhengen mellom tidsbruk og timelønn verbalt, mens det i spørsmål 21 presenteres en grafisk framstilling av dette forholdet. Vi ser at mange elever sier seg enig/uenig i de tre påstandene i spørsmål 19 på en korrekt måte, altså slik at de gir inntrykk av å forstå forholdet mellom anvendt tid/timelønn. Å overføre denne innsikten til den grafiske framstillingen i spørsmål 21, viser seg imidlertid å være vanskelig for de fleste. Svært mange velger da alternativ A, som gir uttrykk for at timelønnen stiger ved økt tidsbruk på avisruten. Langt færre hadde imidlertid sagt seg enig i denne påstanden uttrykt verbalt. Denne diskrepansen er det nærliggende å forklare med at graf A "ligner" mest på de grafene elevene har erfaring med fra skolen, nemlig grafer med positivt stigningstall som går gjennom origo. Slik bør en grafisk framstilling se ut, later mange til å mene.

Niss(1998) uttaler på generelt grunnlag følgende om dette saksforholdet:

> *For a student engaged in learning mathematics, the specific nature, content and range of a mathematical concept that he or she is acquiring or building up are, to a large part, determined by the set of specific domains in which that concept has been concretely exemplified and embedded for that particular student (Niss 1998, s.14).*

Han understreker hvor viktig det er at lærere varierer eksempelbruken i matematikk-undervisningen, slik at ikke elevers begrepsforståelse blir knyttet til enkeltstående eksempler. Likeledes hevder han at man, så vidt mulig, bør benytte ulike representasjoner, numeriske, verbale, symbolske, grafiske etc. Skal elevene erverve matematisk kompetanse, må de jo nettopp tilegne seg *generelle* faglige kunnskaper, slik at de er i stand til å bruke disse i relevante sammenhenger. Til så vel lærernes som elevenes forsvar, skal det dog sies at alternativ C i spørsmål 21, viser en type graf som er ukjent for elevene. Undertegnede må derfor ta hovedskylden for at dette spørsmålet ble et realt "bomskudd".

Avisbudoppgaven ble ikke valgt ut til å være med i PISA's pilotutprøving. Jeg har likevel lyst til å peke på et par faktorer som viser hvilke intensjoner jeg selv hadde med oppgaven, og hvordan jeg ønsket å knytte den til "*mathematical literacy*".

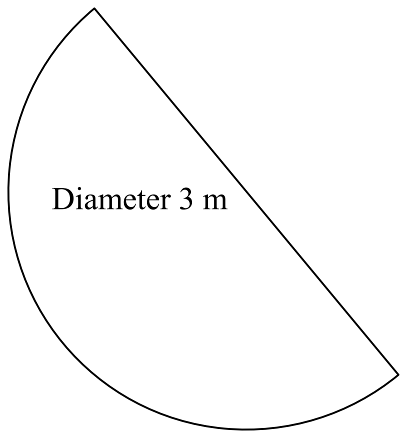
I kapittel 7.3.1. argumenterte jeg for at det å kunne tolke diagrammer, hente ut noe av den informasjonen som ligger der, for deretter å kunne formulere den verbalt, utgjør en svært viktig del av det å være en "*mathematical literate*" person. Ideen med denne oppgaven var at elevene skulle få mulighet til nettopp å dokumentere sin forståelse av sammenhengen mellom tidsbruk og timelønn, så vel verbalt som grafisk, gjennom å gi sin tilslutning til de forelagte alternativene. Oppgaven faller således først og fremst inn under den sentrale ideen *forandring og sammenheng.* Slik sett mener jeg at også avisbudoppgaven kan knyttes til konstruktet "*mathematical literacy*". Det er imidlertid åpenbart at jeg ikke helt har truffet med spørsmålene, særlig har altså spørsmål 21 falt dårlig ut.

Selv om det altså kan argumenteres for at spørsmålene er relevante i forhold til det konstruktet man ønsker å måle, og at situasjonen / konteksten er nært knyttet til elevenes hverdag, ser vi altså at enkeltspørsmål i denne oppgaven ut fra aktuelle statistiske data, egner seg dårlig til å måle konstruktet "*mathematical literacy*" i mitt utvalg. (Med et annet utvalg, for eksempel elever fra videregående skole, kunne selvsagt oppgaven fungert langt bedre.)

## 8.7 Oppgave 6

Her skal elevene tenke seg at de plasserer steiner med en bestemt radius langs omkretsen av et halvsirkelformet bed med oppgitt diameter. Oppgaven krever at elevene kan turnere begrepene areal, omkrets, diameter og radius på en adekvat måte.

# BLOMSTERBED




Formelen for arealet av en hel sirkel er : $A = \pi r^2$

Formelen for omkretsen av en hel sirkel er : $O = 2\pi r$

Cathrine har et halvsirkelformet blomsterbed i hagen, som vist på figuren. I boksen ovenfor står noen formler som du kan få bruk for.

**Spørsmål 22:** BLOMSTERBED

Hva er arealet av bedet ? Vis utregning

**KODER FOR SPØRSMÅL 22 (PISA GENERALPRØVEN I PARENTES)**

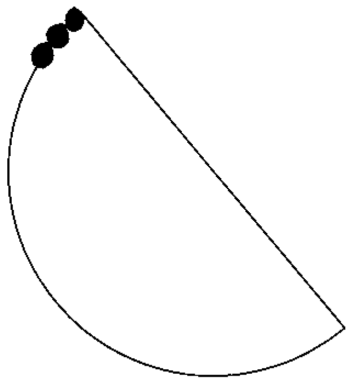
| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **57 (70)** | **37,3 (30,7)** | |
| 20 | 3,5m$^2$ (ingen trekk for feil benevning) | 57 (70) | 37,3 (30,7) | 0,71 (0,76) |
| **Delvis riktig** | | **26 (22)** | **17,0 (9,6)** | |
| 10 | 7m$^2$ | 23 (22) | 15,0 (9,6) | 0,28 (0,11) |
| 11 | Desimalfeil i utregning | 1 | 0,7 | -0,68 |
| 12 | 14m$^2$/28m$^2$ (r=3m - dividert/ikke dividert med 2) | 2 | 1,3 | 1,02 |
| **Ikke riktig** | | **36 (56)** | **23,5 (24,6)** | |
| 70 | Brukt formelen for omkretsen, dividert/ikke dividert med 2 | 10 | 6,5 | -0,57 |
| 79 | Øvrige svar | 26 (56) | 17,0 (24,6) | -0,61 (-0,40) |
| **Ikke svart** | | **34 (80)** | **22,2 (35,1)** | |
| 99 | Blankt | 34 (80) | 22,2 (35,1) | -0,78 (-0,42) |

Av tabellene med de statistiske verdiene ser vi at dette er en oppgave som fungerer bra psykometrisk. Den diskriminerer godt mellom de ulike svarkategoriene, og det er høy korrelasjon mellom elevenes score på denne oppgaven og deres totalscore. P-verdien viser at litt under halvparten av elevene i mitt utvalg svarer fullstendig, eller delvis riktig på dette spørsmålet.

Igjen er andelen elever som ikke har besvart spørsmålet en god del høyere i PISA G. Det er imidlertid relativ god overensstemmelse mellom z-verdiene for de ulike kategoriene. Vi ser for eksempel at Z(2) verdiene er nærmest identiske. Det er altså åpenbart de sterke elevene som lykkes i å besvare oppgaven fullstendig.

| Sp.M. 22 | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Min pil. | 0,46 | 0,64 | 22,2 | -0,60 | 0,30 | 0,71 | -0,78 |
| **PISA G**: | **0,36** | | **35,1** | **-0,40** | **0,11** | **0,76** | **-0,42** |

**Spørsmål 23:** BLOMSTERBED



Langs buen vil hun plassere runde steiner. De har alle en radius på ca. 5cm. Hvor mange hele steiner får hun plass til dersom de skal ligge tett inntil hverandre (slik figuren til venstre viser)? Vis hvordan du regner.

**KODER FOR OPPGAVE 23 (PISA GENERALPRØVEN I PARENTES)**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **23<br>(25)** | **15,0<br>(11,0)** | |
| 20 | 47 steiner | 23<br>(25) | 15,0<br>(11,0) | 1,21<br>(1,01) |
| **Delvis riktig** | | **18<br>(16)** | **11,8<br>(7,0)** | |
| 10 | 4,7/9,4/92/94/98 steiner | 18<br>(16) | 11,8<br>(7,0) | 0,83<br>(0,55) |
| **Ikke riktig** | | **53<br>(76)** | **34,7<br>(33,3)** | |
| 79 | Øvrige svar | 53<br>(76) | 34,6<br>(33,3) | -0,17<br>(-0,18) |
| **Ikke svart** | | **59<br>(111)** | **38,6<br>(48,7)** | |
| 99 | Blankt | 59<br>(111) | 38,6<br>(48,7) | -0,57<br>(-0,19) |

Vi observerer av de statistiske verdiene i tabellene at også dette spørsmålet fungerer greit psykometrisk. P-verdien er imidlertid såpass lav, at oppgaven testteoretisk må regnes for å være en ”dyr” oppgave, som hovedsakelig diskriminerer mellom det sterkere segmentet av elever. Man ser videre at "ikke svart"- prosenten her er forholdsvis høy. Det er rimelig å anta at dette hovedsakelig har to årsaker. For det første ble nok dette spørsmålet muligens oppfattet som noe vanskelig. I tillegg befant det seg nær slutten av heftet. Sett i lys av den noe varierende motivasjonen blant elevene, er derfor ikke svarprosenten lavere enn man vel kunne forvente.

Som man ser av kodeskjemaet har jeg her valgt å benytte få koder. I utgangspunktet hadde jeg langt flere, men disse ble senere slått sammen, slik at jeg endte opp med disse fire kodene. Det som kjennetegner svarene som er samlet under kode 10, er at elevene her har benyttet riktig algoritme, men gjort ulike småfeil under utregningen, slik at de tilslutt likevel ikke har kommet fram til det korrekte svaret.

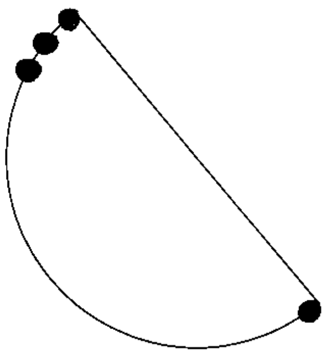
Bortsett fra at andelen elever som ikke har besvart spørsmålet, igjen er en god del høyere i PISA G, og Z(9) følgelig noe lavere, ser vi at de øvrige verdiene er relativt like.

Alle oppgavene som var med i generalprøven i PISA, ble plassert på ulike steder i de 10 delheftene man opererte med. Hver oppgave var stort sett med i tre forskjellige hefter. Dette ble gjort for å utligne eventuelle ulemper knyttet til en spesiell plassering i heftet. I min egen pilotutprøving hadde jeg kun ett hefte. Gjennom verdiene hentet fra PISA G hadde jeg håpet å få indikasjoner på at plasseringen av oppgavene i mitt hefte hadde betydning for svarprosenten. Som vi kan observere, har jeg ikke fått noe klart svar på dette, ettersom det generelt er langt flere som ikke besvarer oppgavene i PISA G, enn i min pilotundersøkelse.

| Sp.m.23 | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Min pil. | 0,21 | 0,63 | 38,6 | -0,17 | 0,83 | 1,21 | -0,57 |
| **PISA G.** | **0,15** | | **48,7** | **-0,18** | **0,55** | **1,01** | **-0,19** |

**Spørsmål 24:** BLOMSTERBED

Cathrine har 30 slike steiner. Hun legger først to steiner ved bedets endepunkter. De resterende ønsker hun å plassere med lik avstand dem imellom rundt hele bedet (se figuren til venstre).



Hva blir da avstanden mellom steinene ? Oppgi svaret i cm med en desimal. Ha med utregningene dine.

**KODER FOR OPPGAVE 24**

| Koder | Svaralternativer | Frekvens | Prosent | Z - verdi |
|---|---|---|---|---|
| **Fullstendig** | | **5** | **3,3** | |
| 20 | 5,9cm | 5 | 3,3 | 1,54 |
| **Delvis riktig** | | **27** | **17,7** | |
| 10 | 5,7cm (dividert med 30 istedenfor 29) | 7 | 4,6 | 1,73 |
| 11 | 6,0cm/6,1cm | 5 | 3,3 | 1,13 |
| 12 | 11cm/10,7cm | 3 | 2,0 | 1,66 |
| 13 | 15,7cm/16,4cm | 8 | 5,2 | 0,70 |
| 14 | Riktig satt opp, men småfeil i utregning | 4 | 2,6 | 0,48 |
| **Ikke riktig** | | **35** | **22,9** | |
| 79 | Øvrige svar | 24 | 15,7 | 0,22 |
| 97 | Vet ikke eller spørsmålstegn | 6 | 3,9 | -0,45 |
| 98 | Tull og tøysekommentarer evt. tegninger | 5 | 3,3 | -0,74 |
| **Ikke svart** | | **86** | **56,2** | |
| 99 | Blankt | 86 | 56,2 | -0,43 |

Her har jeg, i motsetning til hva jeg gjorde ved forrige spørsmål, valgt å beholde et elaborert og omfattende kodeskjema.

Jeg har vært temmelig liberal med 10-er kodene, det vil si at alle elever som har brukt tilnærmet riktig algoritme, har havnet under denne koden, og altså blitt belønnet med poeng. Ut fra z-verdiene ser vi at jeg psykometrisk sett har hatt belegg for å gjøre et slikt valg, ettersom Z(1) er vesentlig høyere enn Z(0). I tillegg vil jeg igjen understreke, at disse elevene altså har gjort noe riktig under utregningen.

Interessant er det å legge merke til at de som har svart feil, men som likevel har forsøkt å gi et fornuftig svar, det vil si kode 79, også har en klart positiv z-verdi, mens de som overhodet ikke har forsøkt å finne noen løsning, åpenbart er svakere elever.

Jeg registrerer at "ikke svart"- prosenten her er svært høy, og vil i den forbindelse henvise leseren til de refleksjonene jeg gjorde meg til det samme fenomenet under kommentarene til det foregående spørsmålet. Disse vil også være relevante her.

Det man ellers kan observere, er at det er de desidert flinkeste elevene som har svart riktig på denne oppgaven. Verdiene til så vel Z(1) som Z(2) er svært høye, og skiller seg markert fra de andre z-verdiene. At Z(9) er noe høyere enn vanlig, skyldes åpenbart at en så stor del av utvalget her havner i denne kategorien.

Som nevnt i kap.8.1 opererte jeg i utgangspunktet med kodene 97 og 98 i samtlige kodeskjemaer, men da det i de fleste tilfeller viste seg at kun to til tre elever havnet under disse kategoriene, valgte jeg i stedet å inkludere dem under kode 79. For de fire gjenstående spørsmålene i ”testen”, har jeg gjort et unntak, og latt dem stå som separate koder. Vi ser av z-verdiene at det, ikke overraskende, først og fremst er relativt svake elever som morer seg med tull og tøys og som svarer ”vet ikke”. Jeg vil likevel poengtere at det altså er svært lite av den slags, besvarelsene sett under ett. Grunnen til at det ble noe mer mot slutten av oppgaveheftet, er vel rett og slett at motivasjonen hadde en fallende kuve, etter som friminuttet nærmet seg.

| Oppgave | P - verdi | Diskriminering | Ikke svart | Z(0) | Z(1) | Z(2) | Z(9) |
|---|---|---|---|---|---|---|---|
| Spørsmål 24 | 0,12 | 0,60 | 56,2 | -0,03 | 1,12 | 1,54 | -0,43 |

### *8.7.1 Blomsterbedoppgaven, PISA og "mathematical literacy"*

Dette var en av de oppgavene som altså ble utvalgt til å være med i generalprøven i PISA. Den besto da kun av spørsmålene 22 og 23, spørsmål 24 var ikke med. Ut fra den gjennomgangen av elevsvarene vi nå har foretatt, kan dette synes fornuftig. Vi har jo sett at "ikke svart"- prosenten på spørsmål 24 var svært høy.

Selv om denne oppgaven kanskje ikke er så nært knyttet til elevenes liv som enkelte av de øvrige (det er nok et solid mindretall av 15-åringer som tilbringer fritiden sin med hagestell), er ferdighetene som man trenger for å løse den, åpenbart, av den nyttige typen. Å kunne regne ut areal og omkrets, og forholde seg til disse begrepene, er noe man absolutt bør mestre for å kunne anses som en "*mathematical literate*" person. Spørsmål 23 og 24 er knyttet til en praktisk anvendelse av slike kunnskaper, i en autentisk kontekst. Vi husker fra kapittel fire at det nettopp er matematiske kunnskaper og ferdigheter "*put into functional use*" (OECD PISA 2001, s.5), som man i PISA ønsker å fokusere på. Blomsterbedoppgaven vil derfor, så vidt jeg kan vurdere, passe godt til å måle aspekter ved konstruktet "*mathematical literacy*". Det er altså den ***konstruktrelaterte validiteten*** til oppgaven jeg her argumenterer for. At denne oppgaven som nevnt også er med i PISA's egen pilotutprøving, anser jeg som en støtte for min validitetsargumentasjon.

Blomsterbedoppgaven er av PISA plassert inn under den sentrale ideen "*rom og form*". Et viktig aspekt ved denne ideen var å kunne forstå "*the properties of objects, and the relative positions of objects*" (OECD PISA 2001a, s.15). Dette er evner man får bruk for ved løsning av spørsmålene i denne oppgaven. I rammeverket i PISA blir det, som omtalt i kapittel 6, understreket at grensene mellom de fire sentrale ideene er flytende, og at de på sett og vis glir over i hverandre. Som et eksempel på dette vil jeg påpeke at Blomsterbedoppgaven også kan knyttes til den sentrale ideen "*kvantitativt resonnement*". Det vil jo være nødvendig å ha "*tallforståelse*" og "*å forstå meningen med operasjoner på tallstørrelse*r", som var to kjennetegn ved denne sentrale ideen (se kap.5.3.1), for å kunne løse oppgaven.

Et karakteristisk trekk ved den sentrale ideen "*forandring og sammenheng*", var at oppgaver tilhørende denne kategorien, bør utfordre elevenes evner til å oversette det matematiske innholdet fra en representasjon til en annen. I Blomsterbedoppgaven illustreres spørsmålene av figurer. Elevene må kunne se sammenhengen mellom disse illustrasjonene, og de beregninger de blir bedt om å gjøre. Slik kan det synliggjøres at Blomsterbedoppgaven også berører denne sentrale ideen.

Vi ser altså at denne oppgaven kan knyttes til hele tre av de sentrale ideene i PISA's rammeverk, og at det følgelig kan argumenteres for dens ***innholdsrelaterte validitet*** i forhold til dette rammeverket er relativ høy.

## 8.8 Oppgave 7 er sperret

### *8.8.1 Spørsmål 25 er sperret.*

### *8.8.2 Spørsmål 26 er sperret.*

### *8.8.3 Spørsmål 27 er sperret.*

### *8.8.4 Kommentarene til oppgave 7 er sperret.*

# 9. Reliabilitets- og validitetsbetraktninger

## 9.1 ”Testens” reliabilitet

Som jeg allerede har understreket, og som jeg nå igjen vil poengtere, er ikke hensikten med mine oppgaver å komponere en fullverdig matematikktest, men å prøve ut enkeltoppgaver. For å finne ut hvorvidt oppgavene måler noen av de samme trekkene, vil jeg likevel for et øyeblikk se på oppgavene som en helhetlig ”test”, slik at jeg skal kunne undersøke ”testens” reliabilitet.

Korrelasjonskoeffisientene (diskriminering) har jeg allerede presentert i tabeller i etterkant av spørsmålene. I vedlegg 1 har jeg lagt fram en samlet, mer utførlig tabell over dette. Av denne kan man se at korrelasjonen mellom elevenes score på hvert spørsmål og deres totalscore, er signifikant på 0,01 nivå for alle spørsmålene, bortsett fra spørsmål 21. Dette indikerer at ”testen” har høy reliabilitet, det vil si at oppgavene i stor grad måler noen av de samme trekkene. For ytterligere å få bekreftet dette, gjennomfører jeg en reliabilitetstest, og får da følgende tabell:

```
  R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (A L P H A)

        N of Cases =       153

Item-total Statistics
```

| | Scale Mean if Item Deleted | Scale Variance if Item Deleted | Corrected Item-Total Correlation | Alpha if Item Deleted |
|---|---|---|---|---|
| R1 | 14,1830 | 68,2952 | ,6062 | ,8937 |
| R3 | 14,3660 | 70,7599 | ,5204 | ,8956 |
| R4 | 14,8497 | 75,0628 | ,4190 | ,8979 |
| R5 | 13,9281 | 68,7514 | ,5571 | ,8952 |
| R6 | 14,5425 | 73,4472 | ,4939 | ,8963 |
| R7 | 14,7386 | 73,8786 | ,4977 | ,8965 |
| R8 | 14,7124 | 73,2852 | ,5604 | ,8955 |
| R9 | 14,1961 | 70,7771 | ,4878 | ,8966 |

| | | | | |
|---|---|---|---|---|
| R10 | 14,4379 | 72,6425 | ,5990 | ,8946 |
| R11 | 14,6667 | 72,8684 | ,5915 | ,8949 |
| R12 | 14,4902 | 74,7647 | ,3377 | ,8988 |
| R13 | 14,7059 | 75,2353 | ,3081 | ,8970 |
| R14 | 14,8170 | 73,9794 | ,5475 | ,8961 |
| R16 | 14,4837 | 70,5672 | ,5385 | ,8951 |
| R17 | 14,0719 | 69,2645 | ,5139 | ,8965 |
| R18 | 13,2484 | 74,2011 | ,3333 | ,8991 |
| R19A | 14,2680 | 75,3422 | ,3203 | ,8990 |
| R19B | 14,3791 | 74,7238 | ,3586 | ,8993 |
| R20 | 14,5882 | 69,3622 | ,6020 | ,8936 |
| R21 | 14,9673 | 77,5713 | ,0811 | ,9012 |
| R22 | 14,1046 | 68,5021 | ,5715 | ,8948 |
| R23 | 14,6013 | 70,3071 | ,5710 | ,8943 |
| R24 | 14,7778 | 72,8845 | ,5625 | ,8952 |
| R25 | 14,8170 | 71,8216 | ,6043 | ,8941 |
| R26 | 14,8824 | 73,3150 | ,5424 | ,8957 |
| R27 | 14,9542 | 75,9255 | ,3229 | ,8991 |

R E L I A B I L I T Y   A N A L Y S I S   -   S C A L E   (A L P H A)

Reliability Coefficients    28 items

Alpha =   ,8999           Standardized item alpha =   ,9010

Cronbachs alfa, eller bare alfa som det oftest omtales som, er som beskrevet i kapittel 6.2 gjennomsnittet av alle mulige korrelasjonskoeffisienter mellom to "halvdeler" av en test, og er et viktig mål for en tests indre konsistens.

En alfa på 0,9, slik som jeg her har fått, sier at kun 10 % av det min ”test” måler er irrelevant ”støy”. Dette støtter opp under påstanden om at ”testen” har høy reliabilitet.

I kolonnen, ”Alpha if Item Deleted”, ser jeg at spørsmål 21 med fordel kan utelates fra ”testen”. Da vil jeg nemlig kunne få en høyere alfa - verdi. Dette betyr altså at denne oppgaven i liten grad måler noen av de samme trekkene som de øvrige. Jeg har dermed fått ytterligere en indikasjon på at dette spørsmålet fungerer dårlig i denne ”testen”.

For samtlige av de andre spørsmålene vil alfaverdien derimot reduseres, dersom noen av dem utelates.

Jeg redegjorde i kapittel 7.2 for en del faktorer man bør være oppmerksom på ved tolkning av Cronbachs alfa. Blant annet vil det at mange svake elever ikke besvarte de siste spørsmålene i min pilotundersøkelse, og at deres ”prestasjoner” på disse oppgavene da vil være helt konsistente, medføre at jeg får en noe høyere alfaverdi totalt sett enn jeg ellers ville ha fått.

I tillegg så vi at jeg utfra en høy alfaverdi ikke direkte kan slutte at testen måler mitt utvalgte konstrukt, ettersom korrelasjonen mellom oppgavene kan skyldes mer enn en faktor. Jeg har imidlertid nå grunnlag for å si at reabilitetsberegningene indikerer at oppgavene mine (bortsett fra spørsmål 21) i høy grad måler noen av de samme trekkene. Det innebærer imidlertid altså ikke at jeg kan konkludere med at ”de samme trekkene” er mitt utvalgte konstrukt ”*mathematical literacy*”. For å kunne si noe mer om dette, er det nødvendig å foreta en analyse og vurdering av oppgavenes validitet.

I neste delkapittel, gjennom de validitetsbetraktninger jeg der foretar, vil jeg søke å sannsynliggjøre at oppgavene måler ulike aspekter ved ”*mathematical literacy*”, altså at dette er et fellestrekk ved oppgavene.

### 9.1.1 Oppgavenes validitet

Jeg har under drøftingen av oppgavene forsøkt å argumentere for deres ***konstruktrelaterte validitet*** i forhold til ”*mathematical literacy*”, og deres ***innholdsrelaterte validitet*** i forhold til de sentrale ideer i PISA’s rammeverk.

Når jeg nå har presentert samtlige oppgaver, finner jeg det interessant også å undersøke hvorvidt det kan argumenteres for deres ***samtidige kriterierelaterte validitet*** (se kap.7.2.2).

Dersom oppgavene virkelig måler ”*mathematical literacy*”, bør det være en høy korrelasjon mellom samlet score på mine oppgaver, og ferdigheter/kunnskaper i matematikk. De sistnevnte blir i skolen vurdert gjennom karakterer i dette faget. Elevene i mitt utvalg har tilkjennegitt sin siste terminkarakter i matematikk, og det er derfor mulig å finne korrelasjonen mellom disse og totalscore på min ”test”.

**Correlations**

| | | KAR.M | SCORE |
|---|---|---|---|
| KAR.M | Pearson Correlation | 1,000 | ,790** |
| | Sig. (2-tailed) | , | ,000 |
| | N | 150 | 150 |
| SCORE | Pearson Correlation | ,790** | 1,000 |
| | Sig. (2-tailed) | ,000 | , |
| | N | 150 | 153 |

**. Correlation is significant at the 0.01 level

Jeg ser her at korrelasjonen mellom samlet poengsum og terminkarakter i matematikk er 0,79 (signifikant på 0,01 nivå), dvs.at mine oppgaver altså har høy korrelasjon med matematiske ferdigheter slik de blir evaluert gjennom skolekarakterer. Korrelasjonen er imidlertid ikke så høy at oppgavene kun måler ferdigheter i skolematematikk, ettersom det er en viss forskjell på alphaverdien på 0,9 som jeg fant i reliabilitetstesten, og korrelasjonskoeffisienten på 0,79 som jeg fikk her. Dette betyr at mine oppgaver også måler noe annet, hvilket jo også er ønskelig, ettersom "*mathematical literacy*" er ment å omfatte noe mer enn *utelukkende* matematiske ferdigheter. Dessuten skal fagkarakterer i skolen ikke bare være knyttet til elevenes faglige prestasjoner, men også avspeile elevenes innsats og arbeid med stoffet i timene, eventuelt hjemme, og i enkelte tilfeller også deres faglige utvikling. Kort sagt er det derfor ikke ønskelig med en for høy korrelasjon.

Måler man kun rene matematiske ferdigheter, har man ofte et ønske om at språklig forståelse ikke skal komme inn som et forstyrrende element, slik at ikke elever som er språklig svake får en lavere poengsum enn de fortjener. Da vil deres **sanne score**[10], være vesentlig høyere enn den vi måler. Ettersom jeg i min test ønsker å måle "*mathematical literacy*", og dette innebærer et element av språklig forståelse, ønsker jeg derimot at det er en viss korrelasjon mellom samlet poengsum og språklig forståelse. Elevene i mitt utvalg har også gitt meg kjennskap til deres terminkarakterer i norsk. Jeg finner korrelasjonen mellom totalscore på mine oppgaver og disse karakterene, og får:

**Correlations**

| | | KAR.N | SCORE |
|---|---|---|---|
| KAR.N | Pearson Correlation | 1 | ,416** |
| | Sig. (2-tailed) | , | ,000 |
| | N | 149 | 149 |
| SCORE | Pearson Correlation | ,416** | 1 |
| | Sig. (2-tailed) | ,000 | , |
| | N | 149 | 153 |

**. Correlation is significant at the 0.01 level

Tabellen viser at jeg får en korrelasjon på 0,42, som er signifikant på 0,01 nivå, mellom terminkarakterer i norsk og samlet poengsum. Den er altså vesentlig lavere enn korrelasjonen mellom karakterer i matematikk og totalscore. Dette er for så vidt også ønskelig, ettersom konstruktet ”*mathematical literacy*", bør inneholde en større porsjon av ferdigheter i matematikk enn i norsk. Hadde den sistnevnte korrelasjon vært for høy, ville ”testen” i for stor grad målt språklige ferdigheter.

Jeg har nå, gjennom å undersøke korrelasjonen mellom terminkarakterer i fagene matematikk og norsk og samlet score på mine oppgaver, sett at ”testen” åpenbart måler aspekter ved konstruktet ”*mathematical literacy*”, som er knyttet til kunnskaper og ferdigheter i matematikk og språk. Slik har jeg forsøkt å finne støtte for påberopelse av oppgavenes ***samtidige kriterierelaterte validitet***. Som en foreløpig konklusjon vil jeg derfor hevde at alt tyder på at ”testen” har en høy reliabilitet. Jeg har også vist at det kan argumenteres for at oppgavenes ***konstruktrelaterte, innholdsrelaterte og kriterierelaterte validitet*** er rimelig høy, selv om dette synes langt vanskeligere definitivt å avgjøre.

10. **Sann score :** enhver observert testscore kan vi tenke oss er satt sammen av to komponenter, nemlig en sann score og en tilfeldig feilkomponent. Dette kan uttrykkes slik:

$X = T + E$

Her er X den observerte score, T den sanne score og E den tilfeldige feilkomponenten. I alle undersøkelser ønsker man at feilkomponenten skal være så ubetydelig som mulig, slik at det er høy grad av overensstemmelse mellom den observerte score og den sanne score. (Angell 1996)

# 10. Konklusjon.

Utgangspunktet for denne avhandlingen er min utvikling av matematikkoppgaver til PISA 2003. Disse oppgavene skulle lages utfra bestemte beskrivelser av det man ønsker å måle, slik dette blir redegjort for i rammeverket for PISA-undersøkelsen. Her går man i detalj inn på intensjonene og målene for hele prosjektet, og hvilke konkrete kriterier oppgavene må tilfredsstille for å kunne bli benyttet som testinstrumenter.

Oppgavene, som jeg altså forsøkte å utvikle i samsvar med disse kravene, ble først gjort til gjenstand for kritisk behandling i PISA - gruppa ved instituttet, og jeg mottok i den forbindelse mange konstruktive kommentarer. På bakgrunn av disse, ble kontekster og spørsmålsstillinger bearbeidet og omformulert. Enkelte oppgaver ble allerede på dette trinn forkastet. De gjenstående oppgaver ble så pilotert på 153 elever ved 4 østlandsskoler. Resultatene fra denne piloteringen utgjør en viktig del av grunnlaget for den drøftingen jeg har foretatt av mine matematikkoppgaver. Jeg har blant annet gjennom denne drøftingen, forsøkt å belyse noen av de forskningsspørsmål jeg innledningsvis stilte.

Første delen av min hovedproblemstilling var: **"Hva er "mathematical literacy"?** Den har jeg forsøkt å besvare på følgende måte: I kapittel 2 redegjorde jeg for tidligere internasjonale komparative undersøkelser knyttet til matematikkfaget. Jeg refererte i den sammenheng ulike typer kritikk som har framkommet i kjølvannet av disse, og hvordan man for første gang i TIMSS 95, delvis for å imøtekomme noe av denne kritikken, gjennomførte en undersøkelse med utgangspunkt i konstruktene "*science literacy*" og "*mathematical literacy*". Man ønsket derigjennom å finne ut hvilke naturfaglige- og matematiske kunnskaper elevene satt igjen med etter 12 års skolegang. I de undersøkelsene som blir gjennomført i regi av PISA, har disse konstruktene, i tillegg til "*reading literacy*", fått en helt sentral posisjon.

For å utdype innholdet i "*mathematical literacy*", som jeg er opptatt av, har jeg videre foretatt en summarisk gjennomgang av hvordan matematikkfaget er blitt legitimert opp gjennom historien. Jeg har, gjennom referanser til ulike matematikkdidaktikere, hevdet at denne legitimeringen for det meste har vekslet mellom nytte- og dannelsesargumenter. I forlengelsen av dette har jeg påpekt at den definisjonen av "*mathematical literacy*" man opererer med i PISA, på sett og vis kan betraktes som et forsøk på en syntetisering av disse to retningene. Man ønsker altså å fremme en matematikkundervisning som legger vekt på så vel de nyttige som de dannende aspekter ved faget.

Jeg har i tillegg til dette ønsket å trekke linjer til "*literacy*"- begrepet, slik dette utvikles av Freire innenfor hans kritiske pedagogiske teori. Freire understreket at det å lære og lese og skrive, og derigjennom tilegne seg kunnskaper, var en nødvendighet for å kunne delta aktivt i utformingen av samfunnet. På lignende måte hevder framtredende matematikkdidaktikere, at det å være en "*mathematical literate*" person, er en forutsetning for at man i tilstrekkelig grad skal kunne forstå sosiale prosesser og derigjennom framstå som demokratisk deltakende og kritiske aktører. Det er kanskje særlig i forhold til dette kritiske aspektet at disse tankene avviker fra det vi finner i rammeverket i PISA. Også i PISA er man opptatt av viktigheten ved at medlemslandenes borgere er "*mathematical literate*", men begrunnelsen er her noe annerledes. I tillegg til å trekke fram de individuelle fordelene dette innebærer, knytter man det til samfunnets stadig økende behov for kompetent arbeidskraft. Man er imidlertid forsiktig med å trekke politisk ladede konklusjoner, og unngår bruk av betegnelsen ”kritisk”.

I kapittel 4 har jeg referert noe av den kritikken som av framtredende matematikkdidaktikere er blitt reist mot avgrensningen og bruken av "*mathematical literacy*" konstruktet i PISA, og presentert deres alternative måter å utlegge og forstå det på.

Slik har jeg altså ønsket å belyse den første delen av min hovedproblemstilling.

Grunnen til at jeg har valgt denne vinklingen, jeg tenker da særlig på tilknytningen til Freire og de refererte matematikkdidaktikere, er at jeg personlig synes et slikt perspektiv på matematikk og matematikkundervisning er svært interessant og utfordrende, og at det gir spennende perspektiver for den framtidige utviklingen av faget. Det kan innvendes at mine egne oppgaver ikke gjenspeiler, og er konsistente med, de tanker og ideer jeg her har referert, og på sett og vis gjort meg til talsmann for. Til dette vil jeg svare at jeg på grunn av korte tidsfrister, nærmest måtte kaste meg rett ut i oppgaveproduksjon ved starten av mitt hovedfagsstudium, og at jeg da var temmelig blank når det gjaldt kunnskap om, og innsikt i, matematikkdidaktiske tanker og teorier. Den prosessen jeg senere har vært igjennom, har vært utrolig lærerik og utviklende. Utfra de forutsetningene jeg i utgangspunktet hadde, er jeg derfor stort sett godt fornøyd med de oppgavene jeg fikk være med på å utvikle til PISA. I dag, etter å ha fått grundigere kjennskap til matematikkdidaktikk som fag, ville jeg nok likevel forsøkt å vinkle noen oppgaver mer i tråd med de synspunkter jeg ut fra de ovenfornevnte teorier har tilkjennegitt.

For å besvare den siste delen av min hovedproblemstilling, ”..**og hvordan konstruere matematikkoppgaver som kan benyttes til å teste ”*mathematical literacy*” på en adekvat måte?**”, har jeg i kapittel 6 redegjort for de kravene man i PISA stiller til matematikk - oppgaver som skal brukes som instrumenter for å måle ”*mathematical literacy*”.

Mine to siste problemstillinger, som henger nøye sammen, har jeg forsøkt å besvare på følgende måte:

I avslutningsdelen av drøftingen av hver oppgave, har jeg foretatt en vurdering av i hvilken grad den aktuelle oppgaven er i samsvar med PISA's ulike krav, med et særlig blikk på "*mathematical literacy*".

Gjennom statistiske analyser av elevresultater, har jeg argumentert for at majoriteten av oppgavene, psykometrisk sett, er relativt velegnet til å benyttes i en test som har til formål å rangere elevene, ut fra i hvilken grad de kan sies å være "*mathematical literate*". For å understøtte dette, har jeg pekt på at reliabilitetsanalysen viser at oppgavene mine i stor grad måler noen av de samme trekkene, og jeg har argumentert for at "de samme" refererer til konstruktet "*mathematical literacy*". Selv om dette er svært vanskelig, for ikke å si umulig å fastslå med sikkerhet, har jeg likevel ment å ha funnet belegg for å kunne komme med en slik påstand.

Dette har jeg forsøkt å begrunne ved å vise hvordan oppgavene kan sies å måle visse sider ved konstruktet "*mathematical literacy*", som definert i PISA's rammeverk, og at de er nært knyttet til de sentrale ideene som omtales og avgrenses i det samme dokumentet. Det faktum at flere av oppgavene er blitt valgt ut til generalprøven i PISA, har jeg ansett som en solid støtte for denne påberopelsen av ***konstruktrelatert-*** og ***innholdsrelatert validitet***.

I tillegg har jeg gjennomført korrelasjonsanalyser mellom elevenes testscore på mine oppgaver, og deres terminkarakterer i fagene matematikk og norsk. Dette har jeg brukt til å argumentere for oppgavenes ***kriterierelaterte validitet.***

Min endelige konklusjon blir derfor at jeg, utfra en vinkling jeg personlig har funnet svært interessant, mener å ha redegjort relativt grundig for utviklingen av konstruktet ”*mathematical literacy*”. I tråd med den ovenfornevnte argumentasjonen, vil jeg også hevde at jeg langt på vei har lykkes i å utvikle matematikkoppgaver som kan benyttes som instrumenter til å teste elever i forhold til dette konstruktet, og som er i overensstemmelse med kravene som kommer til uttrykk i PISA's rammeverk. Jeg har argumentert for at oppgavene i relativt høy grad er valide i forhold til dette rammeverket, og i forhold til konstruktet "*mathematical literacy*". De fleste av dem bør derfor med hell kunne brukes som instrumenter for å rangere 15-årige elever utfra kriteriet ”*mathematical literacy*”.

Jeg har også tilkjennegitt en del synspunkter på framtidige utfordringer knyttet til PISA. Gjennom fokuseringen på ”*mathematical literacy*”, må det kunne sies at man i PISA uttrykker et ønske om at matematikkundervisningen struktureres slik at kommende elevkull i stadig sterkere grad framstår som ”*reflective citizens*”. Med støtte i teoretiske arbeider av flere anerkjente fagdidaktikere, har jeg da hevdet at det bør legges vekt på utvikling av mer strukturelt samfunnsrelaterte matematikkoppgaver. Å lykkes med å integrere den type oppgaver i kommende PISA-undersøkelser, synes å være en formidabel utfordring for alle engasjerte matematikk-didaktikere, involvert i oppgaveutvikling for PISA.

# Referanser/Litteraturliste


Abraham, J. & Bibby, N.(1988). Mathematics and society: ethnomatematics and a public educator curriculum. *For the Learning of Mathematics* **8**(2), pp.2-11.

Angell, C.(1996). *Elevers fysikkforståelse. En studie basert på utvalgte fysikkoppgaver i TIMSS.* Dr.scient-avhandling ved UIO, Oslo, Norge.

Angell, C., Kjærnsli, M. & Lie, S.(2000). Exploring Students Responses on Free-Response Science Items in TIMSS. I Shorrocs-Taylor, D. & Jenkins, E.W. (Eds.), *Learning from Others.* Nederland: Kluwer Academic Publishers.

Brekke, G., Kobberstad, T., Lie, S. & Turmo, A.(1999). *Hva i all verden kan elevene i matematikk. Oppgaver med resultater og kommentarer.* Oslo, Norge: Universitetsforlaget.

Brown,C.(1987). Literacy in 30 Hours: Paulo Freire's Process in Northeast Brazil. I Ira Shor (Eds.), *Freire for the Classroom*. London, UK: Meinemann, pp.215-231.

Crocker, L & Algina, J.(1986). *Introduction to Classical and Modern Test Theory*. Harcourt Brace Jovanovich College Publishers.

D'Ambrosio, U.(1985). Mathematics Education in a Cultural Setting. *International Journal of Mathematical Education in Science and Technology, Vol.16*, pp.469-477.

Dørfler,W & McLone, R.R.(1986). Mathematics as a School Subject. I Christiansen,B., Howson,A.G.& Otte,M.(Eds.), *Perspectives on Mathematics Education*. Holland: D.Reidel Publishing Company, (pp.49-97).

Ernest, P.(2000). Why teach mathematics? I White,J.&Bramall,S.(Eds.), *Why Learn Maths.* London, UK:London University Institute of Education.

Ebel, R.L. & Frisbie, D.A.(1991). *Essentials of Educational Measurement*. Cliffs, New Jersey, U.S.A: Prentice Hall, Englewood.

Freire, P.(1987). Letter to North-American Teachers. I Ira Shor (Eds.), *Freire for the Classroom.* London, UK: Meinemann (pp.211-214).

Gable, R.K. & Wolf, M.B.(2000). *Instrument Development in the Affective Domain*.(2.nd.ed.) London, UK: Kluwer Academic Publishers.

Gisselberg,K., Kjærnsli, M.,Lie, S.& Weng, P.(1996).*Preliminary notes from a Nordic study on item formats in TIMSS.* UiO, Oslo, Norge. (Upublisert materiale).

Gjone, G.(1994). *Matematikkundervisningen i etterkrigstidens enhetsskole*. UiO, Oslo, Norge. (Foredrag holdt på Nordisk Forskersymposium, Island)

Gjone, G.(2001). Matematikkdidaktikk som vitenskap – nasjonal utvikling og internasjonal organisering. I Elstad, E. (Eds.), *Fagdidaktikkens identitet og utfordringer*. Acta Didactica #5/2001 Oslo, ILS, UiO. (pp.81-105).

Gronlund, N.E.(1998). *Assessment of Student Achievement.* MA 02194 USA: Allyn & Bacon Needham Heights.

Grønmo, L.S.(1991). *Det lykkelige valg – matematikk i videregående skole ?* Hovedfagsoppgave i matematikkdidaktikk UIO, Oslo, Norge.

Howson, A.G. & Mellin Olsen, S.(1986a ). Social Norms and External Evaluation. I Christiansen,B., Howson,A.G.& Otte,M.(Eds.), *Perspectives on Mathematics Education*. Holland: D.Reidel Publishing Company, (pp.1-47).

Howson, A.G. & Mellin Olsen, S.(1986 b). Mathematics as a School Subject. I Christiansen,B., Howson, A.G.& Otte, M.(Eds.), *Perspectives on Mathematics Education*. Holland: D.Reidel Publishing Company, (pp.49-97).

Keitel, C. & Kilpatrick, J.(1999). The Rationality and Irrationality of International Comparative Studies. I Kaiser, G., Luna, E. & Huntley, I.(Eds.), *International Comparisons in Mathematics*. Educatio, London: Falmer Press.

Kind, P.M., Kjærnsli, M., Lie, S. & Turmo, A.(1999). *Hva i all verden gjør elevene i realfag? Praktiske oppgaver i matematikk og naturfag*. ILS, Universitetet i Oslo, Norge.

Kleve, B.(1994). *En testteoretisk og diagnostisk analyse av flervalgsoppgaver i matematikk fra TIMSS' pilottes*t. Hovedfagsoppgave i realfagsdidaktikk. ILS, Universitetet i Oslo, Norge.

Kobberstad,T.(1991). *SIMS-ala-Bærum. Hovedfagsoppgave i matematikkdidaktikk*. ILS, Universitetet i Oslo, Norge.

KUD(1987). *M87: Mønsterplan for grunnskolen.* Oslo, Norge: KUD (Dep.).

KUF (1996). *Læreplanverket for den 10 årige grunnskolen*. Oslo, Norge: KUF (Dep.).

Kuhn,T.S.(1970). *The Structure of Scientific Revolutions*, 2nd.edn.Chicago, USA: University of Chicago Press.

Lakatos, I.(1979). *Proofs and Refutations. The Logic of Mathematical Discovery*. UK: Cambridge University Press.

Lie, S., Kjærnsli, M. & Brekke, G.(1997). *Hva i all verden skjer i realfagene. Internasjonalt lys på trettenåringers kunnskaper, holdninger og undervisning i norsk skole.* ILS, Universitetet i Oslo, Norge.

Lie, S., Kjærnsli, M., Roe A. & Turmo, A.(2001). *Godt rustet for framtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv*. ILS, Universitetet i Oslo, Norge.

Lindenskov, L & Wedege,T.(2000). *Numeralitet til Hverdag og Test*. Roskilde, Danmark: Center for forskning i Matematiklæring.

Messick, S. (1989).Validity. I R.L.Linn(Eds): *Educational measurement* (3rd ed). New York, USA: Macmillian Publishing Company, pp.13-104.

Moore, D.S. & McCabe ,G.P.(1999): *Introduction to the Practice of Statistics*. New York., USA: . W.H.Freeman and Company

Ness, A.(1978). *Filosofiens historie. Fra oldtid til renessanse*. Oslo, Norge, Universitetsforlaget.

Niss, M.(1987). Applications and Modelling in the Mathematics Curriculum – State and Trends. *International Journal of Mathematical Education in Science and Technology, Vol.18*, pp. 487-505.

Niss, M.(1994). Mathematics in Society. I Biehler, R: Scholz, R.W., Straesser, R. & Winkelmann, B. (Eds.) *The Didactics of Mathematics as a Scientific Discipline*. Dordrecht.Kluwer Academic, pp.367-378.

Niss, M.(1996). Goals of Mathematics Teaching. I Bishop, A.J.(Eds) *The International Handbook of Mathematics Education., Volume 1.* Dordrecht: Kluwer Academic, pp.11-47.

Niss, M.(1998). *Aspects of the Nature and State of Research in Mathematics Education*. Tekster fra IMFUFA/RUC (nr.351) pp.1-24.

OECD PISA(2000). *Mathematics Item Development for PISA 2003 and Item Submission Guidelines.* Dokument til møte i "Mathematics Forum", Berlin, Tyskland des.2000.

OECD PISA (2001a). *Draft Mathematics Framework for OECD/PISA 2003.* Notat til møte i National Project Managers Group, Nijmegen, Nederland, september 2001.

OECD PISA (2001b). *Mathematics Item Development and Selection*. Notat til møte i "Mathematics Experty Forum, Nijmegen, Nederland sept. 2001.

OECD PISA (2002): *Mathematics Marking Guide, PISA 2003 Items*.

Olsen,R.V.(2000*). Fra tall til ord. Å tilordne meningsfulle verbale beskrivelser til en måleskala*. ILS, Universutetet i Oslo.Norge (Foredrag på Notoddenkonferansen for naturfagdidaktikk, 17.okt.2000.)

Olsen, R.V.,Turmo, A. & Lie, S.(2001). Learning about students' knowledge and thinking in science through large-scale quantitative studies. *European Journal of Psychologyo of Education, Volume 16, nr.3.*

Orpwood,G. & Garden, R.A.(1998). *Assessing Mathematics and Science Literacy*. Vancouver, Canada: Pacific Educational Press.

Popper, K.(1959). *The Logic of Scientific Discovery*. New York, USA: Basic Books

Rand, G.(1965). *Elementær Testteori*, Universitetsforlaget, Oslo, Norge.

Resnick, L. & Resnick, D.(1992). Assessing the Thinking Curriculum: New Tools for Educational Reform. I Gifford, B.& O'Connor, M (Eds.), *Changing Assessments; Alternative Views of Aptitude, Achievement and Instruction.* National Commision on Testing and Public Policy.

Robitaille, D.F. & Garden, R.A.(1996). *Research Questions & Study Design*. TIMSS monograph No.2. Vancouver, Canada: Pacific Educational Press.

Satterly, D.(1989). *Assessment in Schools.* Oxford,UK: Basil Blackwell Ltd.

Sjøberg,S.(1998). *Naturfag som almenndannelse. En kritisk fagdidaktikk*. Oslo, Norge: Ad notam Gyldendal.

Skovsmose, O.(1992). Democratic Competence and Reflective Knowing. *For the Learning of Mathematics **12**(2),* pp.2-11.

Skovsmose, O.(1994). Towards a Critical Mathematics Education. *Educational Studies in Mathematics, 27*, pp..35-57.

Skovsmose, O.(2002): Matematikken er hverken god eller dårlig – og da slet ikke neutral. *Tangenten 3/2002*, Bergen, Norge: Caspar Forlag, pp.22-26.

# VEDLEGG 1 (to sider)

Tabellen viser korrelasjonen mellom elevenes score på hver enkeltoppgave og deres totalscore.

**Correlations**

| | | SCORE |
|---|---|---|
| R1 | Pearson Correlation | ,669** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R3 | Pearson Correlation | ,578** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R4 | Pearson Correlation | ,454** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R5 | Pearson Correlation | ,625** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R6 | Pearson Correlation | ,536** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R7 | Pearson Correlation | ,536** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R8 | Pearson Correlation | ,595** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R9 | Pearson Correlation | ,558** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R10 | Pearson Correlation | ,635** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R11 | Pearson Correlation | ,626** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R12 | Pearson Correlation | ,389** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R13 | Pearson Correlation | ,354** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R14 | Pearson Correlation | ,581** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |

**. Correlation is significant at the 0.01 level

**Correlations**

| | | SCORE |
|---|---|---|
| R15 | Pearson Correlation | ,496** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R16 | Pearson Correlation | ,598** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R17 | Pearson Correlation | ,587** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R18 | Pearson Correlation | ,396** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19A | Pearson Correlation | ,362** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19B | Pearson Correlation | ,406** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19C | Pearson Correlation | ,345** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R20 | Pearson Correlation | ,659** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R21 | Pearson Correlation | ,107 |
| | Sig. (2-tailed) | ,187 |
| | N | 153 |
| R22 | Pearson Correlation | ,639** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R23 | Pearson Correlation | ,627** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R24 | Pearson Correlation | ,601** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R25 | Pearson Correlation | ,645** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R26 | Pearson Correlation | ,580** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R27 | Pearson Correlation | ,357** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |

**. Correlation is significant at the 0.01 level

**Correlations**

| | | SCORE |
|---|---|---|
| R15 | Pearson Correlation | ,496** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R16 | Pearson Correlation | ,598** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R17 | Pearson Correlation | ,587** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R18 | Pearson Correlation | ,396** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19A | Pearson Correlation | ,362** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19B | Pearson Correlation | ,406** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R19C | Pearson Correlation | ,345** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R20 | Pearson Correlation | ,659** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R21 | Pearson Correlation | ,107 |
| | Sig. (2-tailed) | ,187 |
| | N | 153 |
| R22 | Pearson Correlation | ,639** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R23 | Pearson Correlation | ,627** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R24 | Pearson Correlation | ,601** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R25 | Pearson Correlation | ,645** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R26 | Pearson Correlation | ,580** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |
| R27 | Pearson Correlation | ,357** |
| | Sig. (2-tailed) | ,000 |
| | N | 153 |

**. Correlation is significant at the 0.01 level