

Danske 4. klasseelever i TIMSS 2007

Peter Allerup

Forord

TIMSS 2007 projektet blev startede i 2003, kort efter, at den sidste undersøgelse, TIMSS 2003 var afsluttet. TIMSS dækker over titlen: *Trends In International Mathematics and Science Study* og er en international komparativ sammenligning mellem elevpræstationer, og deres baggrund, i fagene matematik og naturfag. Danmark deltog ikke i TIMSS 2003, bl.a. fordi det ikke var muligt at skaffe den fornødne finansiering. Danmark deltog senest i TIMSS 1995 (Weng, 1998; Beaton et al, 1997). TIMSS undersøgelserne har været gennemført regelmæssigt siden 1960'erne, og har alle årene været planlagt og styret af IEA, *International Association for the Evaluation of Educational Achievement*, som er en sammenslutning af forskningsinstitutioner verden over.

Rektoratet for DPU besluttede efter sonderinger om eventuel medfinansiering af TIMSS 2007 selv at afholde udgifterne over de interne forskningsbudgetter. Dansk deltagelse blev vurderet som et betydningsfuldt alternativ til de igangværende runder med PISA-undersøgelser, med muligheder for at opnå grundlæggende forskningsresultater på 4. klassetrin. På dette klassetrin dannes vigtige faglige og holdningsmæssige sider af elevernes forhold til fagene matematik og naturfag, som har betydning for dem senere i skoleforløbet.

På DPU er TIMSS 2007 blevet gennemført med hjælp fra en central stab af medarbejdere og medhjælpere. Hanne Ankerdal og André Torre har fra starten af projektet medvirket til, at samtlige praktiske og forskningsmæssige sider af processen er forløbet problemløst. Peter Allerup har været den interne leder af projektet og har ansvaret for afviklingen af TIMSS 2007.

Fra projektets start har Erika Thuebæk, Jane Frøhling og Pernille Thorborg deltaget som medhjælp i alle faser af projektets fremdrift, planlægning og praktisk deltagelse i afholdelse af pilottest, indsamling af informationer fra skolerne og kontakt til skoler og lærere. Alle har deltaget i det omfattende arbejde med at kode elevsvarene på de åbne spørgsmål i TIMSS, hvor svarene skal sammenholdes med manualer for scoring. Skrivning af den internationale Encyklopædi er gennemført med Jane Frøhlings hjælp, og i den fase af projektet, hvor kontakten til lærere og skoler, som skal modtage og benytte resultaterne af TIMSS 2007, vil deres medhjælp fortsat være værdsat. Erika Thuebæk har foretaget de sidste korrekturer af manuskriptet.

Fra DPU har Helene Sørensen og Anne-Marie Møller Andersen desuden medvirket i de faser, hvor opgaverne skulle oversættes fra engelsk til dansk. Peter Weng har deltaget i både oversættelser af hovedmaterialet i TIMSS opgaverne og planlægning af indholdet af den 'ekstra' TIMSS, som består i, at samtlige deltagende skoler fik tilbud om at deltage i en ekstra test med vægt på at foretage grundige, formative evalueringer af elevsvarene.

Indledning og baggrund

Hvad er TIMSS?

TIMSS 2007 eller *Trends In Mathematics and Science Study* er betegnelsen for en international sammenlignende undersøgelse, hvor færdigheder og kundskaber samt den undervisningsmæssige baggrund er i fokus. TIMSS 2007 er det sidste skud på en stamme bestående af mange TIMSS undersøgelser, der oprindeligt startede i 1960'erne. Danmark har tidligere deltaget i 1995 undersøgelsen, der som nu omfattede fagene matematik og naturfag (biologi, fysik, kemi og geografi). I denne undersøgelse deltager Danmark med elever fra 4. klasse. TIMSS 2007 udføres tillige i en del lande på 8. klasseniveau.

Forventninger om at komme nærmere til en forståelse af, hvorfor mange elever betragter matematik og naturfag som uinteressante i deres valg af uddannelse, er én af årsagerne til, at DPU valgte at deltage i TIMSS 2007. Samtidig kan man i de højere klassetrin se, at mange elever har svært ved at klare de faglige udfordringer, som de to fag stiller. Ligger noget af forklaringen i nogle forhold, som kan ses allerede i 4. klasse?

TIMSS gentages

Som navnet 'Trends' i TIMSS antyder, gentages TIMSS undersøgelserne i øjeblikket med ca. 4 års intervaller, således at den næste er planlagt til 2011. Det er et internationalt formuleret formål med TIMSS undersøgelserne (IEA, 2007), at de kan bidrage til en faglig forståelse af, hvorfor landets elever klarer sig 'godt' eller 'skidt', og at successive TIMSS målinger kan bidrage til at evaluere en udvikling – hvad enten den er igangsat af tidligere TIMSS målinger, eller den er skabt på baggrund af andre pædagogiske interventioner eller reformer af skolesystemet.

Indhold af TIMSS besluttet

TIMSS undersøgelsesernes faglige indhold og planlægning besluttet på generalforsamlinger af medlemmerne af IEA¹. De spørgsmål, der undersøges via TIMSS undersøgelserne, er skabt ud fra ønsket om at kunne svare på spørgsmål om, *hvorfor* nogle elever er bedre eller evt. dårlige end andre elever, og *hvordan* eleven er nået frem til det svar, som er afgivet. Begrundelser og beskrivelser som er fastlagt ud fra faglige (didaktiske) kriterier inden for matematik og naturfag.

¹ IEA: *The International Association for the Evaluation of Educational Achievement* er et konsortium af forskningscentre inden for uddannelsesområdet. Det omfatter medlemmer fra hele verden og etablerer internationale studier inden for uddannelsesområdet gennem et samarbejde mellem medlemmerne fra de forskellige lande. Danmark har via DPU (og tidligere DPI) været medlem af IEA siden 1986.

Man kan kort opsummere de spørgsmål, som TIMSS forsøger at sammenholde:

- hvilke faglige forventninger kommer til udtryk via de formål med undervisningen, der angives i læseplanerne (den *planlagte* læseplan, trinmålene)
- hvordan søges den planlagte læseplan *gennemført* på skolerne og i klasseværelset?
- hvilket *udbytte* har eleverne fået af undervisningen (den præsterede læseplan)

Der er andre studier end TIMSS

TIMSS er defineret som et internationalt projekt, fordi man ønsker at se på elevfærdigheder på så forskellig 'baggrund' for elevsvarene som muligt. Både i forhold til den kognitive og fysiske baggrund. Det ville fx næppe være særligt givende at benytte rent danske undersøgelser som grundlag for besvarelse af 'hvorfor' og 'hvordan', fordi det danske skolesystem er meget homogent, og samtlige elever har dermed været 'udsat' for nogenlunde den samme type af undervisning. For at kunne besvare 'hvorfor'- og 'hvordan'- spørgsmålene på en mere dækkende facon, kræves det, at man via undersøgelsen har mulighed for at foretage *formative* evalueringer af de svar eleverne giver. Det kræver svar, hvor strukturen over én eller flere opgavesvar kan lægges til grund for konklusioner og beslutninger i forbindelse med konkrete anbefalinger rettet mod den videre undervisning og læring.

Til forskel fra TIMSS, stiller andre internationale undersøgelser spørgsmål af mere *normativ* karakter. De OECD - initierede PISA undersøgelser og de årlige NAEP – undersøgelser i USA (National Assessment of Educational Progress) er begge eksempler på store undersøgelser, hvis egentlige formål er at undersøge, om *samfundet får nok ud af de investerede penge i uddannelsessystemet?* 'Nok' defineres og evalueres i denne forbindelse hovedsageligt ud fra konkrete sammenligninger med andre landes præstationer og deres omkostninger i uddannelsessystemet. Det er *ikke* de enkelte elevers præstationer, der er i centrum; det er større grupper af elever, ultimativt 'alle' testede elever, set i et nationalt perspektiv som har interesse.

Præstationer og holdninger

I TIMSS 1995 var det fx studiet af kønsforskelle og lærer-elev interaktioner, der var i fokus, da IEA og de deltagende lande formidlede resultaterne i videnskabelige artikler. Modsat disse individorienterede analyser gennemførte det engelske økonomiske fagblad The Economist en simpel 'normativ' sammenligning mellem TIMSS' placeringer af de gennemsnitlige præstationer af landene med samme landes bruttonationalprodukt pr. indbygger. Derved ankom The Economist til den umiddelbart noget nedslående konklusion, at Danmark var et af de lande, der fik

mindst ud af de investerede penge pr. elev i form af lave TIMSS scores, mens fx Japan for relativt få penge investeret pr. elev opnåede et af de *højeste* TIMSS gennemsnit. The Economist's gennemregning havde ikke den store effekt på danske forhold, den danske Folkeskole indførte ikke japanske undervisningsmetoder eller lignende som konsekvens af disse nye sammenligninger.

**37 lande med
185000 elever
deltager**

Samtidig med at elevfærdigheder undersøges i TIMSS 2007, kortlægges de forskellige landes undervisningssystemer, læseplaner og undervisningspraksis, så den nationale undervisning kan studeres på denne baggrund og i et internationalt perspektiv.

Der deltog i TIMSS 2007 i alt ca. 185.000 elever på 4. klassetrin fra 37 forskellige lande sammen med deres lærere (se Appendiks xxx).

Undersøgelsen er planlagt og styret på det internationale niveau af IEA, *The International Association for the Evaluation of Educational Achievement*, som i de sidste 50 år har gennemført internationale sammenlignende undersøgelser inden for uddannelsesområdet.

**Danmark
deltager med 4.
klasser**

I Danmark har Danmarks Pædagogiske Universitetsskole under Århus Universitet (DPU) stået for undersøgelsen, og har samtidigt påtaget sig den fulde finansiering inden for universitetets forskningsrammer.

TIMSS 2007 inddrager på verdensplan elever fra 4. klasse og 8. klasse. Heri deltager Danmark med ca. 3.500 eleverne fra 4. klasse fordelt repræsentativt på knap 137 skoler landet over. Når DPU ikke medtager eleverne fra 8. klasse i undersøgelsen skyldes det flere ting. Dels vil undervisningsmæssige konsekvenser, som evt. kan udledes fra undersøgelsen have den største mening og effekt, når det drejer sig om elever i starten af deres skolegang, dels ville omkostningerne til afholdelse af TIMSS 2007 på 8. klasse overstige, hvad universitetet kunne afsætte af forskningsmidler.

Offentliggørelsen af TIMSS 2007 resultaterne falder tidsmæssigt nogenlunde sammen med offentliggørelsen af resultaterne fra PISA 2006 undersøgelsen. Denne internationale OECD undersøgelse har i 2006 naturfag i fokus. Det vil derfor være muligt at diskutere og sammenholde naturfagsfærdigheder for elever i TIMSS' 4. klasse med PISA's elever fra 9. klasse. De to undersøgelser kan på en måde siges at dække hver sin ende af skolesystemet.

TIMSS skemaer

Informationerne fra TIMSS 2007 undersøgelsen kommer fra:

1. Spørgeskemaer til lærere og skoleledere
2. Opgavehefter i matematik og naturfag til eleverne
3. Spørgeskemaer til eleverne

Eleverne benyttede to timer i april-maj 2007 til at besvare opgavehefterne og én time til at udfylde spørgeskemaet. Der er 14 forskellige opgavehefter i gang, og hvert hefte er delt mellem opgaver i matematik og naturfag. Ved hjælp af nogle opgaver, der går igen fra hefte til hefte kan man 'binde hefterne sammen', så det fra TIMSS' side ser ud, som om, at man har fået besvarelser ind på ca. 300 opgaver i alt, fordelt på de to fagområder. Det giver god spredning inden for hvert af de to fag; TIMSS dækker de trin- eller læringsmål, der ligger til grund for undervisningen i de to fag (Encyclopedia, IEA, 2007).

Fortrolige oplysninger

Opgavebesvarelserne er blevet behandlet fortroligt. Når det umiddelbart fra hefter og spørgeskemaer ikke ser sådan ud, idet der er tydelige navne på skemaerne, skyldes det, at analyserne på DPU skal kunne sammenholde besvarelser af opgavehefter med spørgeskemaer fra samme elev, og forbindelsen skal være klar til læreren (lærerspørgeskemaet) og skolen (skolelederspørgeskemaet). Endelig var det vigtigt at man undervejs kunne fastholde en slags identitet af eleven, fordi deltagelse i TIMSS 2007 af IEA kræver, at to uafhængige hold personer 'scorer'/koder de opgaver, hvor eleverne skriver forslag til løsninger med almindelige ord og sætninger (såkaldt åbne opgaver).

TIMSS' Instrumenter

Opgaver der besvares rigtigt eller forkert

TIMSS' instrumenter består af nogle opgavehefter til eleverne og spørgeskemaer til elever og lærere. Opgaverne er konstrueret ud fra idéer til evalueringsinstrumenter, der nok i sin form ser 'gammeldags' ud, men som samtidigt lever op til moderne krav om testning. Det 'gammeldags' kommer ud på, at evalueringen, og dermed sammenligningen mellem forskellige elever, er baseret på elevers svar til opgaver, der er stillet på forhånd. Evaluering af eleven sker på grundlag af, at eleven svarer på nogle stillede spørgsmål eller opgaver af mere eller mindre *sværhedsgrad*. Svarene vurderes 'snævert' som rigtige eller forkerte og fx ikke som 'morsomme' eller 'fantasifulde'. Svarene fører i sidste ende frem til en beregning af elevens *dygtighed*.

I andre fag eller i andre testsituationer kan opgaverne være ret abstrakte og svarene udformet som 'fri' stil eller i en essayform. Der er i sådanne former, som bl.a. er kendt fra folkeskolens afgangsprøver ikke nødvendigvis noget, der er rigtigt eller forkert. Evalueringen af eleven spiller på andre sider af besvarelsen end om det er 'rigtigt' eller 'forkert'.

Åbne opgaver scores

Opgaver i TIMSS skal altså i hovedsagen besvares 'rigtigt' eller 'forkert'. Også selv om opgaven rent faktisk består i, at eleven skal svare med en længere skriftlig begrundelse (åbne opgaver, se Bilag xxx). I sådanne tilfælde 'scores' svaret efter fælles vurdering i en gruppe af personer, som 'retter' opgaven, resultatet bliver 'rigtigt' eller 'forkert'. Ved besvarelsen af andre typer af opgaver (Multiple Choice, MC, se appendiks xx) sætter eleven ét kryds i en boks ud fra nogle fortrykte forslag til svar, og da der kun findes én boks med det rigtige svar bliver svaret også her 'rigtigt' eller 'forkert'. Afgørelsen af om svaret er rigtigt eller forkert, sker ved at man (elektronisk) ser i hvilken boks eleven har sat sit kryds, dvs. uden at det er nødvendigt at trække på en arbejdsindsats fra personer, der først 'scorer' svarene.

Pilottestning af opgaver

En række internationale aktiviteter er gået forud for fastlæggelsen af det endelige sæt opgaver i TIMSS 2007. Aktiviteter, som sikrer, at danske elever ikke kommer ud for opgaver der 'rammer helt ved siden af'. Dels er der afholdt møder, hvorunder opgaverne har været drøftet og eventuelt revideret, set fra faglige, nationale synsvinkler på læseplaner (curriculum), dels blev der i foråret 2006 i alle deltagende lande afholdt en omfattende pilottest af TIMSS-opgaverne. Ved den lejlighed blev en del opgaver yderligere kasseret på grund af, at de ikke fungerede på en ensartet måde inden for alle deltagende lande². Det er indfrielsen af disse krav, der kvalificerer øjensynligt 'gammeldags' udformede opgaver til at leve op til moderne psykometriske krav.

opgavetyper

Der er tre typer af opgaver i TIMSS:

- flervalgsopgaver, det er opgaver, hvor eleven skal vælge mellem 4 til 5 mulige svar (Multiple Choice, MC)
- opgaver hvor eleverne skal give et kort svar
- opgaver, hvor eleverne skal gøre rede for, hvordan de er nået frem til løsningen

² Det er et omfattende statistisk analysearbejde at sikre, at opgaverne tilfredsstiller psykometriske krav, fx om at opgavernes *relative sværhedsgrad* er den samme for alle elever – dvs. er en bestemt opgave dobbelt så svær som en anden opgave, så skal dette være tilfældet for alle elever, uanset køn, generel dygtighed, nationalitet osv. En anden version af de psykometriske krav går ud på, at elevers 'dygtighed' skal kunne beregnes uanset hvilke opgaver, der lægges til grund for beregningen – en nyttig og nødvendig egenskab, når en elev kun besvarer opgaver fra ét hefte ud af de 14 hefter totalt. Analyserne af svarene fra pilottesten foregår ud fra såkaldte Rasch Modeller (IRT modeller).

Det er de sidste to typer af opgaver, som skal 'scores' på DPU mht. til om de er rigtigt besvaret.

**Opgaver fra
fælles
pensum**

Opgaverne er udvalgt på grundlag af den omtalte pilotundersøgelse, og er desuden taget fra tidligere runder af TIMSS, fx TIMSS 2003, hvor Danmark som nævnt ikke deltog. Endvidere har eksperter fra de deltagende lande bedømt opgavernes egnethed, og det har fx været nødvendigt hurtigt at få afklaret, om opgaverne ligger inden for den læseplan (trinmål), som er gældende for danske elever. Da det ikke er realistisk, at samtlige 37 deltagende landes læseplaner falder inden for TIMSS's opgaver, vil danske elever, såvel som elever fra andre lande, nogle få gange møde opgaver, der ligger uden for den læseplan, der gælder for 4.klasse. Antallet af den slags opgaver er få, og de er af tekniske grunde spredt over de forskellige opgavehefter, så i praksis møder eleverne kun den slags opgaver ganske få gange. I øvrigt bliver der taget højde for dette i de analyser af TIMSS, hvor elevernes 'dygtighed' beregnes, idet en opgave som falder helt uden for læseplanen, tages ud ved beregningen af elevens 'dygtighed'.

Opgaverne er konstrueret efter to kriterier, med et fagligt – og et kognitivt specificeret indhold. Inden for matematik er definitionerne på det *faglige område* (Content domain): 'Tal' (teori), 'geometri' og 'data analyse'. Inden for naturfag er det tilsvarende: 'Biovidenskab', 'naturvidenskab' og 'geovidenskab' (earth science). Det *kognitive domæne* er for både matematik og naturfag beskrevet som kategorier, der beskriver elevens 'viden', muligheder for 'anvendelse' og 'ræsonnering' (reasoning) – det sidste i forbindelse med opgaver, hvor man vurderer de *overvejelser*, der har ført eleven frem til sit svar.

**Det kognitive
domæne i
matematik**

De kognitive områder i *matematik* drejer sig mere detaljeret om (se i øvrigt TIMSS Assessments Framework, IEA, 2007)

Viden

- at genkalde definitioner, terminologi, talegenskaber, geometriske egenskaber
- at kunne genkende matematiske genstande, former, tal og udtryk
- at kunne gennemføre udregninger ud fra de simple regneoperationer
- at kunne uddrage information fra grafer, tabeller, mv.
- at kunne benytte måleredskaber som en lineal
- at kunne klassificere genstande efter numeriske eller geometriske egenskaber

Anvendelse

- at kunne vælge relevante metoder eller strategi til problemløsning
- at kunne repræsentere matematisk information i grafer og tabeller
- at kunne generalisere, dvs. finde en passende model for relationer
- at kunne udføre 'ordrer' om at opstille tabeller og lave grafer
- kunne løse problemer ud fra numeriske eller geometriske egenskaber

Ræsonnering

- at kunne bestemme og benytte numeriske og geometriske relationer, mv.
- at kunne generalisere en løsning fra en konkret opgave
- at kunne kombinere viden fra forskellige kilder
- at kunne afgøre sandheden af matematiske udsagn
- at kunne løse opgaver, som kræver mere end simpel rutine

**TIMSS
opgaver
konstrueret
efter to
principper**

Fordelingen af TIMSS' opgaver inden for disse to kriterier er sket med respekt for, hvor meget de pågældende *faglige* elementer 'fylder' i en almindelig undervisningstime, og tilsvarende, hvor megen tid, der i det *kognitive* område anvendes på at appellere til elevernes 'paratviden', evner til 'anvendelse' og 'klargøring af de tankeprocesser', der er forbundet med løsning af opgaver. Der blev udarbejdet i alt 214 opgaver eller *items* i naturfag og 215 items i matematik. De fordeler sig på følgende måde, vist i tabel 1.

**et faglige
indhold i
naturfag**

De faglige områder i *naturfag* drejer sig mere detaljeret om (se i øvrigt TIMSS Assessments Framework, IEA, 2007)

Biovidenskab

- livsprocesser for levende genstande
- reproduktion og arvelighed
- interaktion med miljø
- økosystemer
- mennesker og helbred

Naturvidenskab

- klassifikation og egenskaber ved genstande
- fysiske egenskaber og forandringer
- energikilder, varme, temperatur
- lys og lyd
- elektricitet og magnetisme

- kræfter og bevægelse

Geovidenskab

- Jordens struktur og fysiske egenskaber og resurser
- Jordens udviklingshistorie
- Jorden i solsystemet

Et fagligt princip

I matematik dækker 'tal' (teori) over beregninger med hele tal og brøker, evnen til udføre simple beregninger og at kunne gennemføre mere komplekse beregninger, der kræver, at eleven kan overskue flere skridt i beregningerne. Dertil kommer evnen til at 'se' relationer mellem tal og formulere en 'regel', der forbinder ét sæt tal med et andet. 'Geometri' dækker over kendskab til to- og tredimensionale figurer i forskellige opsætninger. 'Data' repræsenterer det matematiske felt at kunne organisere informationer, repræsentere grupper af tal numerisk og geometrisk, således at 'data'-fremstillingen støtter en given problemløsning.

Inden for naturfag benyttes de faglige kategorier 'biovidenskab', 'naturvidenskab' og 'geovidenskab' som faglige betegnelser for det, der kort beskrives som 'naturfag', 'fysik' og 'geologi/geografi' i andre sammenhænge. 'Biovidenskab' dækker over elevens evne til at forstå og beskrive livsprocesser i mennesker, dyr og planter, samt nogle vinkler på sundhed og sygdom. 'Naturvidenskab' dækker over kendskab til fysiske processer og materialer samt nogle praktiske vinkler på elektricitet, sollys og energi. Endelig dækker 'geovidenskab' bredt over kendskabet til Jorden som bosted for mennesker og dyr samt dens placering i solsystemet. Dertil kommer fysiske egenskaber ved Jorden og forståelse af, hvorledes man evaluerer, fortolker og drager konklusioner af resultater af forsøg inden for området. Elevernes evne til at argumentere er belyst i dette område.

Et kognitivt princip

Fælles for de to faglige områder er, at opgaverne i TIMSS er underkastet en *kognitiv klassifikation*, som kort går ud på at beskrive, hvad det er for en type 'tænkning' som kræves ved løsningen af en opgave. 'Viden' dækker i den korte udgave over 'paratviden' og refererer inden for både matematik og naturfag til, at fagene på 4. klassetrin har store mængder af 'viden', som kun i mindre omfang er resultatet af, at eleverne har opbygget et 'lager' af viden ud fra selve undervisningen. Eleverne forventes fx på dette klassetrin at kende til elementære sider af fotosyntesen i forbindelse med bladenes skiftende farve over året, selv om flertallet af eleverne

naturligvis ikke kender den grundlæggende kemiske proces der ligger bagom.

'Anvendelse' dækker, som navnet antyder over elevernes evne til at anvende den viden, de har til løsning af problemer. Der findes i TIMSS en række opgaver, som kan pege på, om eleven har den fornødne viden på området men fx ikke magter at anvende den fornuftigt i forbindelse med løsning af konkrete problemer. På 4. klassetrin findes der med denne analytiske indfaldsvinkel adgang til at se på mere eller mindre 'hensigtsmæssig anvendelse' inden for matematik, som er relativt mere regelbundet, og dermed vidensbelagt, sammenlignet med forholdene i naturfag. Elevernes præstationer inden for 'ræsonnering' drejer sig om at evaluere evnen til at foretage overvejelser og beslutninger, som ikke er en direkte følge af de eksisterende 'regler' og kendsgerninger på området. Det er en udfordring at fremstille opgaver inden for dette område, når der i princippet kun kan besvares 'rigtigt' eller 'forkert' på opgaven. Som det fremgår, omfatter evalueringen af et svar fra en elev flere facetter end 'bare' at se efter, om svaret består af et 'rigtigt' eller 'forkert' ræsonnement.

Tabel 1 TIMSS' opgaver i matematik og naturfag.

**Opgaver i
Matematik**

	Domæne	Antal MC items	Antal åbne items	Totalt antal items
Fagligt	tal (teori)	50	43	93
	geometri	32	28	60
	data	14	12	26
Kognitivt	viden	45	24	69
	anvendelse	37	33	70
	ræsonnering	14	26	40

**Opgaver i
naturfag**

	Domæne	Antal MC items	Antal åbne items	Totalt antal items
Fagligt	biovidsk.	36	40	76
	naturvds.	21	21	42
	geovds.	31	24	55
	viden	54	27	81

Kognitivt	anvendelse	48	40	88
	ræsonnering	15	31	46

Ekstra opgaver ud over TIMSS' egne

Det besluttes at rette henvendelse til de deltagende lærere i TIMSS i de to fag med et spørgsmål, om de også kunne tænke sig at være med til en 'ekstra TIMSS' med opgaver, som er specielt rettet mod det danske undervisnings- og evalueringssystem. I matematik sagde ca. 80% af TIMSS lærerne ja, således at der fra disse lærere og klasser dels eksisterer TIMSS 2007 - data og dels besvarelser fra en række opgaver fra 'ekstra TIMSS'. I denne supplerende undersøgelse er opgaverne betydeligt mere 'diagnosticerende' - end i TIMSS opgaverne. Opgaverne er konstrueret med specielt blik for at søge indsigt *i den måde*, eleven overvejer spørgsmålet og kommer frem til et svar; altså selve løsningsprocessen. Det forventes, at resultaterne fra denne del af den samlede undersøgelse er klar i det tidlige forår 2009. Formidlingen af resultaterne kommer bl.a. til at foregå ved afholdelsen af en række workshops, hvortil lærere og interesserede fra CVU'erne inviteres.

TIMSS Spørgeskema til elever

I tillæg til de 14 opgavehefter består TIMSS' instrumenter af et spørgeskema til eleverne og et spørgeskema til lærerne. Spørgeskemaet til eleverne omfatter 16 spørgsmål, der afdækker forhold i hjemmet, forhold vedrørende forberedelse og arbejde med fagene, samt elevernes holdning til at arbejde med problemer inden for de to fag. Nogle af variablene benyttes til et skøn over elevernes socioøkonomiske baggrund.

-- og til lærerne

Lærerspørgeskemaet indeholder tilsvarende nogle faktuelle oplysninger om lærernes baggrund som lærere (linjefag, nuværende undervisningsplan, mv.) samt deres oplevelser af at være lærere netop på den skole, hvor de er nu. Fx muligheder for at gennemføre undervisningen på præmisser, som passer dem og hvis ikke, hvad der eventuelt lægger hindringer i vejen for at kunne gøre det, dvs. gennemføre 'en god undervisning'. Relationen mellem lærer og elev belyses gennem beskrivelser af, hvad læreren lægger vægt på, jf. den karakteristik, der er vist vedrørende det faglige og kognitive indhold bag opgaverne, over brug af hjemmearbejde til anvendelsen af tekniske hjælpemidler i undervisningen. Desuden svarer lærerne på spørgsmål vedrørende deres egen opfattelse af at 'være klædt på' til at undervise i fagene.

Kan man stole på TIMSS resultaterne – og hvad kan de bruges til?

3.500 elever

Det siger sig selv, at TIMSS' stikprøve på ca. 3.500 elever kun er et billede af, hvad

**tegner et
billede af
65.000 elever**

samtligte ca. 65.000 elever i danske 4. klasser ville have svaret. På den anden side foregår indsamlingen af data via statistiske metode, der sikrer, at de resultater, man beregner ud fra de 3.500 elever er 'dækkende'³ for, hvad man ville have fået, hvis man havde involveret samtlige 65.000 elever i TIMSS. Metoden til udvælgelse af de 3.500 elever har samtidigt sikret, at de forhold, der beregnes, forgår med en acceptabel sikkerhed. Der er således ikke anledning til at tro, at TIMSS resultater, der måtte afvige noget fra forventningerne, som årsag kan have at 'det jo kun er få elever, der er testet' eller 'at eleverne nok er skævt udvalgte'. Begge dele er der taget højde for.

**TIMSS'
målinger skal
ikke være
skæve**

TIMSS' resultater udgøres hovedsageligt af 'målinger' af forskellig art. Spørgsmålene i spørgeskemaet 'måler' fx i hvor høj grad eleverne 'kan lide at lære om matematik'. Undertiden går flere spørgsmål sammen, og der konstrueres et indeks til 'måling' fx af 'selvtillid'. I opgavehefterne sammenfattes i princippet 60 svar på delopgaver i geometri – 32 i formatet multiple choice (MC) og 28 'åbne' – til ét tal som 'mål' for elevens dygtighed i geometri. Den proces der 'komprimerer' 60 enkelt svar til ét tal, forgår via statistiske (Rasch) analyser. Herunder afgøres, om opgaverne psykometrisk bør indgå i optællingen af *antal rigtige* besvarelser. Hvis dette er tilfældet indgår opgaven i den fase, hvor beregningen af et mål (indeksværdi) for elevens 'dygtighed' finder sted. Kun hvis opgaverne har én fælles (latent) egenskab, som fx kan kaldes 'geometrisværdhed', kan det lade sig gøre at beregne ét tal pr. elev som udtryk for 'geometridygtighed'.

**Inhomogene
opgaver skal
tages ud**

Analyserne af om dette kan lade sig gøre, gennemføres med kritisk blik for at opdage, om *alle* opgaverne i en serie af 'geometriopgaver' rent faktisk refererer til 'geometrisværdhed', eller om opgaverne viser sig at være i besiddelse af andre typer af 'sværdhed' end 'geometrisværdhed' – og derfor skal elimineres. Der kan desuden være tale om sværdheder, der er mindre for drenge sammenlignet med piger på en sådan måde, at selv om en dreng og en pige på samtlige andre opgaver 'er lige', så henter drengen en slags 'gratispoint' fra en sådan opgave. Også af denne grund skal opgaven elimineres. Man kalder opgaven for *inhomogen*.

At det ikke altid kan lade sig gøre at tilfredsstille disse krav, er set flere gange i forbindelse med de internationale undersøgelser (Allerup, 1997, 2002, 2007).

³ Den stikprøve der er anvendt er *simpel tilfældig udvælgelse* blandt samtlige 65000 elever. Antallet N=3500 elever i stikprøven er bestemt ud fra ønsket om, at den statistiske usikkerhed på gennemsnit over *alle* involverede elever (altså landsgennemsnit af forskellig art) skal være 10% af statistiske den usikkerhed som manifesterer sig ved 'målinger' af én elev.

Undlader man at eliminere inhomogene opgaver, kan resultater fx vedrørende kønsforskelle eller vedrørende forskelle mellem danske og ikke-danske elever blive misvisende.

I TIMSS 2007 blev et stort antal opgaver introduceret i en pilotafprøvning i 2006. Herunder skulle det bl.a. afgøres, om opgaverne havde de omtalte homogenitetsegenskaber. Det viste sig at ca. 50% af opgaverne måtte udgå og til selve hovedstudiet i 2007, det indsamlede data i april-maj 2007, blev kun de opgaver, der passerede samtlige statistiske psykometriske krav benyttet (Technical Report, IEA).

Et problem om fortolkning af resultater fra en måleskala

*Procent
rigtige og
TIMSS' point
skala*

TIMSS' opgaver besvares, som omtalt, næsten alle med udfaldet 'rigtigt' eller 'forkert'⁴ og da opgaverne er blevet 'godkendt' af de statistiske (Rasch) analyser, dvs. er *homogene*, kan man for hver elev beregne et antal rigtige besvarelser som grundlag for beregningen af elevens 'dygtighed'. Nogle elever har besvaret opgaver fra ét opgavehefte, mens andre elever har besvaret opgaver fra et helt andet opgavehefte, evt. ét, der ingen overlap har med det første opgavehefte. Det er en del af egenskaben for opgaverne 'at være homogene', at man kan placere nogle af dem i ét hefte og resten i et andet hefte og alligevel være i stand til at sammenligne elever som har besvaret to forskellige hefter på én, fælles skala.

TIMSS har valgt at anvende en skala, som også benyttes i fx PISA for denne sammenligning. En skala, der har 500 *TIMSS point* som internationalt midtpunkt og som spreder eleverne fra ca. 200 til ca. 800 point. I statistiske vendinger benyttes, at samtlige knap 200.000 elevers 'dygtigheder' i de 37 lande ses som normalfordelt med en middelværdi på 500 og en spredning på 100.

En elev med 492 point er derfor bedre end en elev med 480. Indregner man den såkaldte 'error of measurement', kan det godt være, at der *ikke*, statistisk set, er signifikant forskel på de to præstationer – de afviger alene med et beløb, som kan klassificeres som 'tilfældige afvigelser'.

*12 point
signalerer
forskelle
afhængig af
antallet
bagved*

Hvis en *klasse* har 492 i gennemsnit til sammenligning med en anden klasse med 480 point i gennemsnit, er der større håb om at der er statistisk signifikant forskel på niveauerne for de to klasser. Hvis et *land* afviger med beløbet $492 - 480 = 12$ skala point fra et andet land i gennemsnit, er der (se appendiks xx) meget stor chance for, at de to

⁴ Nogle få opgaver benytter flere grader (op til 3) af 'rigtighed'

lande vurderes som signifikant forskellige. Afhængig af hvor mange elever der indgår i en bestemt vurdering, kan afgørelsen om, at de to grupper er statistisk set signifikant forskellige altså afhænge af, hvor mange elever der indgår i grupperne. Disse betragtninger følger almindelig statistisk praksis og er allerede del af forsøgsplanen, der indhenter data til TIMSS, netop ved at sætte grænser for, hvor stor den statistiske usikkerhed skal være på beregningen af et landsgennemsnit.

***Hvad betyder
så 12 points
forskell?***

Et mere presserende problem består i at fortolke, hvad det betyder, at der evt. *er* en forskel på 12 point. Hvad er den pædagogiske konsekvens af denne forskel? Uanset om forskellen statistisk set vurderes signifikant eller ej, og afhængig af hvor mange elever, der ligger bag ved forskelsberegningen, så kan problemet 'hvad betyder 12 point på skalaen' godt stå alene. Dvs. et problem man skal kunne give et pædagogisk bud på. Problemet kan i virkeligheden føres tilbage til at skulle fortolke, om en elev – eller et gennemsnit over et antal elever – der har præsteret 492 point i fx geometri, har præsteret noget godt eller noget skidt?

Ser man på det internationale gennemsnit på 500, er resultatet 492 jo lidt under, måske ikke signifikant, når det kommer til analyse ud fra de givne data – men, mere interessant, hvordan ser man og forstår man, fagligt set, en forskel på 12 point ud fra en *formativ* vinkel, kunne man spørge?.

***En formativ
evaluering***

Det vil umiddelbart, for denne diskussion, være nemmest at flytte tallene fra TIMSS' pointskala tilbage til den skala, hvorpå man tæller antallet af rigtigt løste opgaver.

Denne tilbageregning sker éntydigt, dvs. én TIMSS pointværdi på 500-skalaen svarer til ét antal rigtigt besvarede opgaver – eller én procent rigtig værdi, hvis man foretrækker at vurdere rigtigheden ud fra procent. Hvis 492 TIMSS point svarer til en procent-rigtig værdi på 52%, betyder det altså, at eleven (eller gruppen af elevernes gennemsnit) ligger på en værdi, der svarer til, at lidt mere end halvdelen af opgaverne er besvaret rigtigt. Er det nu godt eller skidt?

Svaret afhænger af om de opgaver, eleverne har fået, er svære eller lette opgaver.

Altså om opgaverne har lave eller høje værdier af 'opgavesværhed'. Hvis de er lette, dvs. har lave værdier af opgavesværhed, må man forvente, at mange elever kan løse opgaverne rigtigt – og så er det nok 'skidt', at eleven, eller den gennemsnitlige elev, ikke kan løse mere end godt halvdelen af opgaverne rigtigt.

Man kan øge forvirringen omkring disse forhold ved at minde om, at de fleste opgaver uden store problemer kan konstrueres, så det faglige indhold er bevaret (fx 'vinkelrethed' i geometri), mens 'indpakningen' gør opgaven mere eller mindre svær.

***Svære eller
lette opgaver
bestemmer
man selv***

Groft sagt, er det således problemløst at ændre på opgavers sværhedsgrad uden at ændre på det faglige indhold. Man kan dermed 'styre' den samlede rigtighedsprocent hen hvor man vil, fx lægge den på et niveau, der pr. tommelfingerregel fungerer psykometrisk godt, nemlig på ca. 65%⁵.

***Hvornår er
noget godt
eller skidt?***

At tage stilling til om 492 TIMSS point er udtryk for et 'godt' eller 'skidt' resultat, er derfor ubestemmeligt i forhold til det faktum, at opgavernes sværhedsgrader kan justeres efter behag. Man fristes til at drage den konklusion, at det eneste man derefter kan uddrage af information fra tallet 492 TIMSS point er nogle rent *normative* betragtninger vedrørende eleverne, fx at elever der scorer mere end 492, er bedre, og elever der scorer mindre, er svagere. Betragtninger, der gerne kan suppleres med tal (hyppigheder) over hvor mange elever, der ligger over eller under 492 (dvs. bestemmelse af 492 som fraktilværdi). Dette faktum kan ikke ændre på den grundlæggende mulighed for at ændre på opgavernes sværhedsgrader, og udsagn som 'at det er skidt at ligge under 492, fordi kun 10% af eleverne ligger under 492' er ikke med til at løse problemet. Tværtimod. En sådan *rent normativ* udmelding vil kunne placere en elev i gruppen af elever, der klarer sig 'skidt' *uanset* om eleven løser mere eller mindre rigtigt af de stillede opgaver – bare der hele tiden er 90% af eleverne, som kan gøre det bedre. Denne type fejltagelse vedrørende fortolkningen af en præstation udtrykt ved point blev fx begået i officielle fortolkninger af PISA 2003's resultater i læsning, der førte til en karakteristik af elever under 417 PISA point som værende "funktionelle analfabeter" (Allerup, 2005).

***De svage
elever***

Et nærmere kik på procent-rigtige bag ved PISA's pointtal viste, at forskellen mellem den 'gennemsnitlige præstation' (500 på PISA skalaen) og præstationen svarende til den funktionelle analfabet (417 point) modsvarede en forskel på procentskalaen på mellem ca. 60% og 42% rigtige besvarelser. Derfor kan selve tallet '42%' ikke kan retfærdiggøre karakteristikken 'analfabet'.

***Spredning på
TIMSS
skalaen
uanset***

En sidste komplikation i relationen mellem opgavesværheder og målte dygtigheder på TIMSS skalaen er, at uanset om opgavekonstruktørerne fremstiller opgaver, der fungerer sådan, at præcis 52% af eleverne kan løse *hver af* opgaverne rigtigt, eller om konstruktørerne spreder opgaverne på en sådan måde, at der er svære og lette opgaver blandet imellem hinanden, så fastlægger TIMSS pointskalaen med 500 som internationalt midtpunkt og 100 som spredning. Variationsbredden af procent-rigtige

⁵ Så bliver svage elever ikke kede af det og dygtige elever synes, at der stadigvæk er lidt udfordring i opgaven og undgår – som det vil være tilfældet med værdier på fx 80% - at begå mystiske fejl, fordi sådanne elever ikke forstå så let et spørgsmål ("der må stikke noget under—")

opgaverne

bliver meget forskellig i de to situationer, forskelle som ikke reflekteres i TIMSS pointskalaen.

**Curriculum
versus
literacy**

Mulighederne for tildeling af pædagogisk mening til tallet 492 som et fortolkeligt mål for dygtighed af eleven, kan derfor synes svær at få øje på. Men helt sådan er det ikke. I TIMSS er opgaverne fremstillet ved at se på de lærings- eller trinmål, der ligger til grund for undervisningen i faget. Opgaverne udformes derfor efter et *curriculum-orienteret* princip, der forsøger at forholde sig ret tæt til de læseplaner, som karakteriserer faget. Det giver, i hvert fald i Danmark, håb om at kunne vurdere, om en elev har indfriet de krav, der er beskrevet i læringsmålet og kan derfor bane vejen for at forholde sig konstruktivt og fortolkende til den præstation, hvor eleven klarer at løse 52% af de stillede udfordringer.

Inden for andre typer af undersøgelser, fx PISA, benyttes en såkaldt *literacy orienteret* vinkel på konstruktionen af opgaverne. Inden for læsning kommer forskellen mellem et curriculum orienteret princip og et literacy orienteret princip ud på, at den curriculum orienterede opgave *eksplicit* refererer til, og medtager kendte komponenter i teoretiske læsemodeller. Dvs. komponenter som omfatter begreberne 'fonem', 'ortografi' og 'grammatik', mens den tilsvarende udfordring ved den literacy orienterede opgave går ud på at læse 'mellem linjerne', eller 'bag om linjerne' for at kunne svare på de stillede spørgsmål.

Inden for matematik refererer en curriculum-orienteret opgave til færdigheder, der har med specifikke matematiske løsningsprocesser at gøre, som evt. beskrevet i læseplanerne, mens den literacy orienterede opgave i højere grad beder eleven om at drage konklusioner og konsekvenser af nogle præmisser, der opstilles i opgaveteksten.

**Curriculum
orienterede
opgaver gør
det nemmere
at vurdere om
noget er godt?**

Det er en påstand, som mere detaljerede studier af TIMSS og PISA kan afgøre, at det er 'nemmere' at forstå og fortolke graden af rigtighed, når opgaven er udsprunget fra et curriculumorienteret sæt læringsmål i stedet for at skulle vurdere, om 'mere eller mindre' grad af en kompetence eller literacy er til stede i besvarelsen af nogle opgaver. I TIMSS fra 1995 overlod man til eksperter og passende repræsentanter for samfundet at vurdere, om opgaverne var passende. Ydermere bad man samme personer om at komme med et bud på, hvor stor en andel af en klasse, *burde* kunne svare rigtigt på opgaverne, vurderet én for én. Denne teknik anvendes ikke længere. På den anden side har årene, med stadig klarere (curriculumorienterede) mål beskrevet for undervisningen i højere og højere grad, gjort det muligt at tage stilling til om 492, og

dermed 52% rigtigt besvaret er i orden og tilhører 'godt' kategorien eller bør klassificeres som ikke tilfredsstillende og dermed placeres i 'skidt' kategorien.

Danske elever i internationalt perspektiv

De danske 4. klasseelevers præstationer på TIMSS' matematik- og naturfagstest gengives i tabel 2 ved hjælp af de omtalte TIMSS scores⁶ og som 'procent rigtige'. Oversigtsformen fra tabel 1 benyttes med samme faglige og kognitive kategorier.

Tabel 2 Danske elevers præstationer på TIMSS' opgaver i matematik og naturfag.

	Domæne	TIMSS gennemsnit	Pct rigtig gennemsnit	Alle domæner gennemsnit
Matematik				
	tal (teori)	509	51%	523 57%
	geometri	544	60%	
	data	529	68%	
	viden	528	59%	523 57%
	anvendelse	513	60%	
	ræsonnering	524	47%	

	Domæne	TIMSS gennemsnit	Pct rigtig gennemsnit	Alle domæner gennemsnit
Naturfag				
	biovidsk.	527	56%	517 54%
	naturvds.	502	52%	
	geovds.	522	53%	
	viden	515	60%	517 54%
	anvendelse	516	50%	
	ræsonnering	525	49%	

Tabel 2 viser, at danske elever både i matematik og i naturfag klarer sig bedre end den gennemsnitlige internationale elev, som er sat til værdien 500.

**Danske elever
klarer sig
godt i**

Danske elever præsterer på et niveau, så der 'kun' er 14 lande blandt de 37 lande, der

⁶ International middelværdi på 500 med standardafvigelse på 100, dvs. 95% af tallene fra ca. 300 til ca. 700

**internationalt
perspektiv**

er effektivt bedre i naturfag og 'kun' 10 lande, som er effektivt bedre i matematik. De elever, som internationalt i naturfag præsterer bedre end danske elever, stammer fra Singapore, Kina, Hongkong, Japan, Rusland, Letland, England, USA, Ungarn, Italien, Kazakhstan, Tyskland, Australien og Sverige. I matematik drejer det sig om landene Hongkong, Singapore, Kina, Japan, Kazakhstan, Rusland, England, Letland, Holland og Litauen.

**Nogle elever
præsterer
bedre end
danske elever**

Der er ikke noget præcist mønster i den overordnede undervisningsstruktur, som præger de lande, hvis elever gennemsnitligt præsterer bedre end danske elever. Dog er flere østeuropæiske og asiatiske lande repræsenteret, og der er øjensynligt ingen lande fra det arabiske område (bilag).

**Danske
elevers
relative
styrker og
svagheder**

Det ses af tabel 2, at danske elever i matematik er relativt bedre til geometri end til tal (teori) – herunder regulære 'regnestykker' med de fire regningsarter, mens præstationerne i data, der handler om at præsentere tal, mv. i tabeller og figurer, ligger på gennemsnittet. Det passer godt med forestillingen om, at undervisningen i 4. klassetrin ikke fokuserer så meget på de automatiske færdigheder, der fx er forbundet med hurtigt at kunne gange 376 med 47 – en udfordring som nogen kalder 'talterperi'. Ud fra en kort vurdering af resultaterne fra det kognitive domæne, hvor 'anvendelse' bl.a. står for at kunne *anvende* 'regneregler' for at gange to tal sammen, passer den viste gennemsnitlige profil med relativt lave præstationer inden for 'anvendelse' i forhold til 'viden' og 'ræsonnering' godt med betragtningerne over, hvor danske elevers faglige styrke befinder sig.

De to nordiske lande Norge og Sverige placerede sig på følgende måde, set fra de overordnede gennemsnitlige værdier på TIMSS skalaen:

Tabel 3 De tre nordiske lande

Land	Matematik	Naturfag
Sverige	503	525
Norge	473	477
Danmark	523	517

**Danske elever
i forhold til
elever i Norge
og Sverige**

Man ser, at for naturfags vedkommende er Svenske elever en smule bedre end de danske elever, $525 \Rightarrow 517$, dvs. 8 TIMSS point; forskellen er statistisk set *ikke* signifikant.

Opgaverne i TIMSS er sorteret efter fag og indpasset i 14 forskellige opgavehefter, således at ca. halvdelen af opgaverne i ét hefte, ca. 25 opgaver, er matematikopgaver og den anden halvdel, ca. 25 opgaver er naturfagsopgaver. Nedenfor er nogle opgaver gengivet, som er 'fritstillede til offentliggørelse', fordi de ikke kommer til at indgå i den næste udgave af TIMSS i 2011. Andre opgaver ønsker man med begrundelse i fremtidige sammenligningsmuligheder mellem 2007 og 2011 ikke at offentliggøre nu. Opgaverne repræsenterer et spektrum af sværheder inden for hvert af de faglige områder, som er nævnt i tabel 2. En svær opgave forventes alene løst korrekt af de dygtigste elever.

Et eksempel på *en svær naturfagsopgave* inden for det faglige område 'naturvidenskab' er nedenstående opgave. Opgaven besvares korrekt af 50% af de danske elever, hvor det internationale gennemsnit ligger på 39%; af de bedste elever i verden svarer 60% korrekt.

11 Susie har tre isterninger i forskellig størrelse. Hun kommer isterningerne i tre ens bægre, der indeholder samme rumfang af vand, som vist på tegningen. 502_11

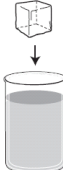
Terning 1



Terning 2



Terning 3



Hvad sker der med isterningerne, når de kommer i vandet?

- (A) Terning 1, 2 og 3 vil synke.
- (B) Terning 1, 2 og 3 vil flyde.
- (C) Terning 1 vil flyde og terning 2 og 3 vil synke.
- (D) Terning 1 og 2 vil flyde og terning 3 vil synke.

**Eksempler på
svære og
mindre svære
opgaver**

En *middelsvær naturfagsopgave* inden for det faglige område *biovidenskab* er gengivet herunder. Opgaven besvares korrekt af 45% af de danske elever, det internationale gennemsnit ligger på 33%, og af de bedste elever i verden svarer 93% korrekt på spørgsmålet.

3 Tegningen nedenfor viser et møls livscyklus.
Skriv navnet på hvert stadium i boksene.
Et stadium er udfyldt for dig.

502_03

voksent møl

504018

Endelig kan følgende opgave illustrere en opgave i naturfag inden for området 'biovidenskab' *som er let*, og som derfor forventes løst af de fleste elever. 84% af de danske elever svarede korrekt mod et internationalt gennemsnit på 68%. Af eleverne fra gruppen af de bedste elever, svarede 92% korrekt.

Opgaven med bjørnen er et eksempel på, at selv om elever fra nogle lande via personlige erfaringer har kendskab til, *hvad* en ørken er og *hvad* den indeholder for liv, mens andre elever ikke selv har oplevet en ørken, så har opgaven under afprøvningen vist sig egnet til at blive inkluderet i TIMSS med en fælles, relativt lille sværhedsgrad over for samtlige elever i verden. Faktisk kunne det måske i det lys undre, at kun 28% af eleverne i Yemen kender det rigtige svar. Blandt de danske elever mener kun ca. 1% af eleverne at det er bjørnen, som ret sikkert lever i ørkenen, mens resten, ca. 15% fordeler deres svar ligeligt på en tiger og en krabbe. Selvom mange danske elever sikkert kender til den 'isørken' der findes på Grønland, så undgår de (is)bjørnen som svar.

17 Hvilket af disse dyr lever ret sikkert i en ørken?

507_04

A bjørn

B krabbe

C firben

D tiger

5031283

Inden for matematik er følgende opgave et eksempel på *en svær opgave* fra området 'tal'(teori). Opgaven besvares rigtigt af 17% af de danske elever over for et internationalt gennemsnit på 15%. Danske elever ligger ret præcist på det gennemsnitlige her. I gruppen af de bedste elever svarer 39% rigtigt på det stillede spørgsmål. At danske elever ikke er stærke i udfordringer, der kræver hurtig talbehandling, indikeres her.

4

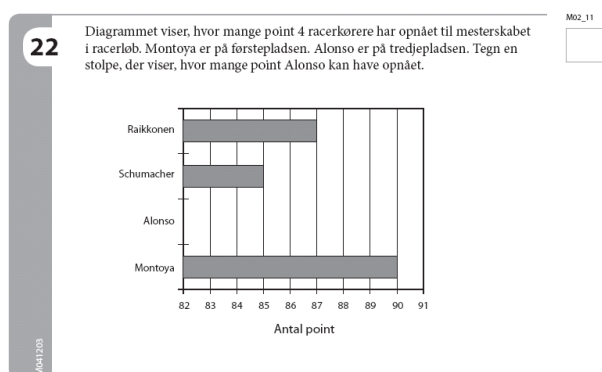
M01_1227

Søren brugte samme regneregul for at få tallet i fra tallet i . Beskriv regnereglen?

Svar: _____

M01_04

En *middelsvær matematikopgave* er følgende, som tilhører den faglige gruppe 'data', hvor opgaven består i en grafisk illustration af nogle talrelationer. Denne opgave besvares korrekt af 48% af de danske elever, mens det internationale gennemsnit ligger på 38%. Blandt de bedste løser 77% af eleverne opgaven korrekt.



Endelige illustrerer den følgende opgave en *let matematikopgave* inden for området 'tal'(teori). 75% af de danske elever løste opgaven korrekt over for det internationale gennemsnit på 60%. Blandt de bedste elever i verden klarede 95% af eleverne at svare korrekt.

**Der
fastlægges
flere niveauer
af dygtig-
heden**

MO313D1

8

Allan ville vide, hvor meget hans kat vejede. Han vejede først sig selv og vægten viste 57 kg. Derefter vejede han sig selv sammen med katten. Vægten viste nu 62 kg.

Hvor mange kg vejede katten?

Svar: _____ kilogram

M01_08

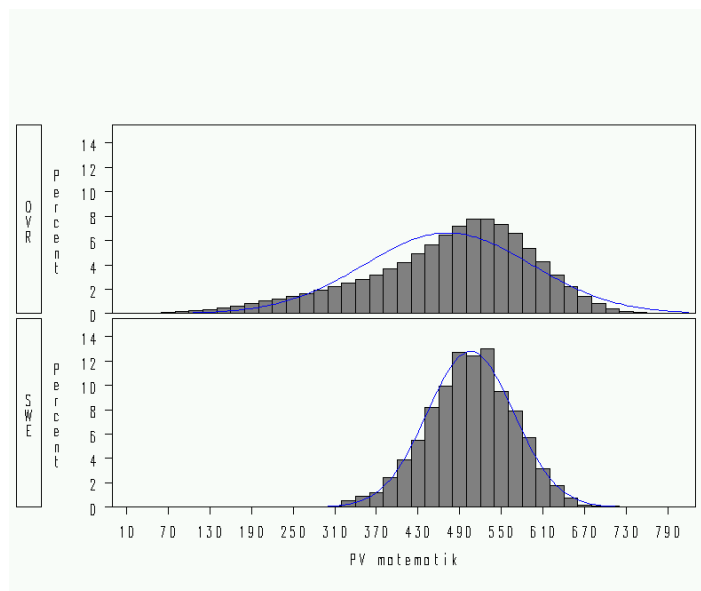
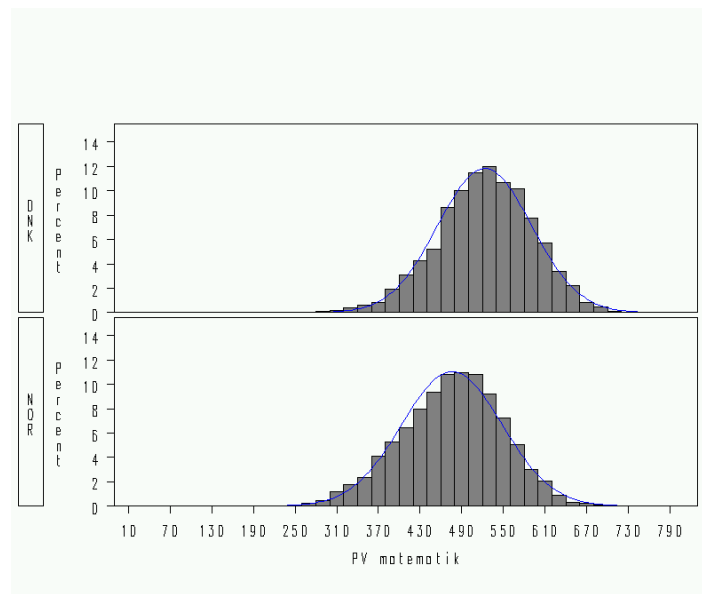
Som omtalt er det af flere grunde svært at angive meningsfulde grænser for TIMSS scorerne i retning af at pege på præstationer, der står for et 'godt' resultat eller et 'skidt' resultat. TIMSS' officielle internationale rapport (IEA, 2008) definerer sådanne grænser alligevel, men undgår en diskussion af rimeligheden i deres valg. Nærværende nationale rapport gør ikke brug af detaljerne bag ved disse grænser, men for at belyse en vigtig side af fordelingen af elevpræstationerne skal to af de internationalt benyttede grænser fremdrages.

Samtlige 200.000 TIMSS scores fra eleverne fra alle lande standardiseres, så gennemsnittet er 500 med en statistisk spredning på 100. Ser man på fordelingerne over TIMSS scorerne i Figur 1, specielt for 'OVR', som dækker over de 175.000 elever, der *ikke* går i danske, svenske eller norske skoler, fremgår det ret tydeligt, at normalfordelingen er uegnet som teoretiske statistisk model. Fordelingerne er både i matematik og naturfag ret skæve, hvilket indikerer, at TIMSS scorerne ikke lader sig simpelt beskrive ved hjælp af middelværdier og spredningsmål, noget som ellers karakteriserer anvendelsen af normalfordelingen. Det har indflydelse på udpegningen af 'svage' og 'dygtige' elever, dvs. udpegning af grænseværdier for 'godt' eller 'skidt'. Andre metoder er derfor bragt i anvendelse i TIMSS' internationale rapport, hvor værdierne 625 som grænseværdien for 'godt' og 400 for 'skidt' er benyttet (TIMSS 2007 International Report IEA, 2008). Resultaterne er opsummeret i tabel 4.

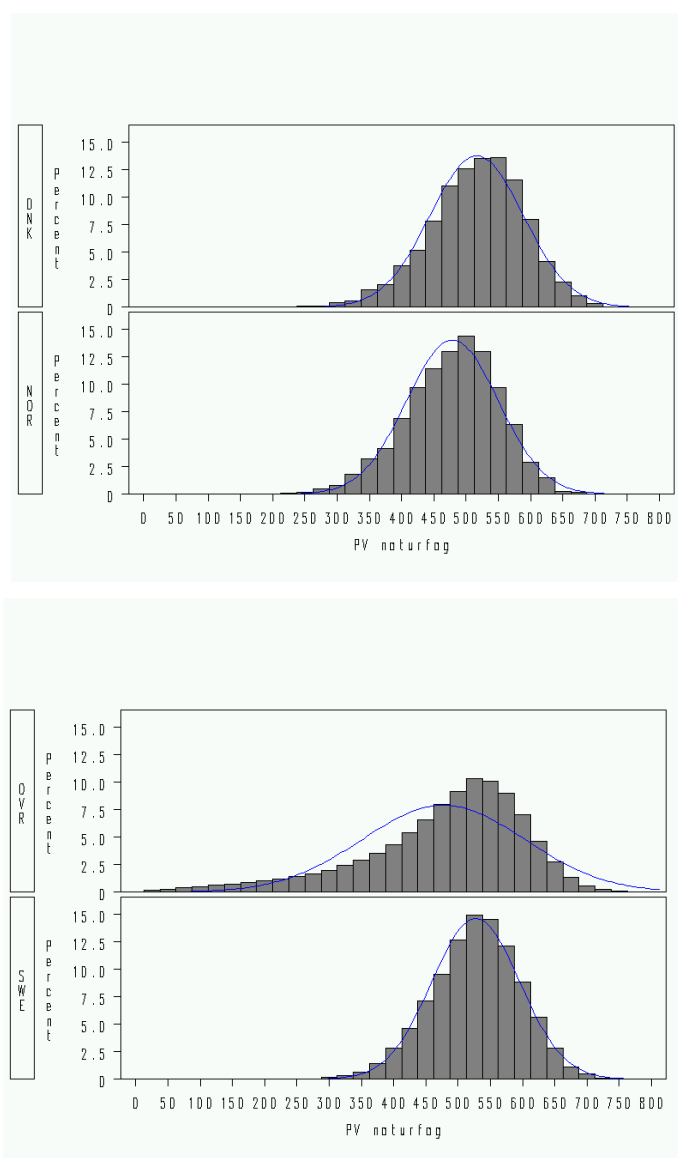
**Fordeling af
TIMSS
præstationer
i de nordiske
lande**

Figur 1 Fordelinger af TIMSS scores i de nordiske lande Norge (NOR), Sverige (SWE) og Danmark (DNK) samt i lande uden for de nordiske lande (OVR).

Matematik



Naturfag



***Svage og
dygtige elever***

Tabel 4 procentdel af elever, som ligger over den internationalt fastsatte grænse for 'godt' og under den internationalt fastsatte grænse for 'skidt'.

Matematik	Procentdel elever over 625 ('godt')	Procentdel elever under 400 ('skidt')
Internationalt	5%	10%
Danmark	7%	5%
Norge	2%	17%
Sverige	3%	7%

***Spredning af
TIMSS
scorerne***

Naturfag	Procentdel elever over 625 ('godt')	Procentdel elever under 400 ('skidt')
Internationalt	7%	7%
Danmark	7%	7%
Norge	1%	16%
Sverige	8%	5%

Fra tabel 4 ses det, at danske elever bidrager pænt med 7% til gruppen af elever i kategorien 'godt' inden for *matematik*. Det er hyppigere, end det sker internationalt, og samtidigt lander halvt så mange danske elever i kategorien 'skidt', 5% sammenlignet med det internationale niveau på 10%. Sammenlignet med Norge og Sverige ser fordelingen i de to kategorier også ud til at falde ud til danske elevers fordel.

***Profil af de
dygtige elever***

Når man ser på naturfag, ligger danske 4. klasseelever præcis på det internationale gennemsnit mht. til at levere elever i de to ekstreme dygtighedskategorier. Kun i kategorien 'godt' har Sverige en procent flere elever end i Danmark

En del af forklaringen på de forskelle mellem de nordiske lande, som ses i tabel 4 er, at de er et resultat af *de generelle* forskelle i TIMSS scorerne mellem landene, som er skildret i tabel 3. Ser man desuden på fordelingerne af TIMSS scorerne i figur 1, fremgår det, at scorerne i *matematik* spreder sig mest, internationalt set, næstmest i Norge, så Danmark og til sidste Sverige. For *naturfags vedkommende* er rækkefølgen internationalt set, Danmark, så Norge og til sidst Sverige. Dette giver ikke noget klart billede for Danmarks vedkommende, ud over at svenske elevers præstationer ligger 'tættere' end danske elevers, måske fordi individualiseret undervisning er slået igennem på forskellig vis i de to lande? Det skal i øvrigt bemærkes, at spredningen ved de danske resultater er størst for drengenes vedkommende.

Man kan ved hjælp af opgørelser over elevernes 'dygtigheder' og svar fra elevernes spørgeskemaer stykke et billede sammen af 'de dygtige' elever, fx. elever som har TIMSS scores på mere end de 625, som er sat som øvre grænse i tabel 4. Inden for *matematik* fører dette til følgende profil af de dygtige elever. Der er tale om:

**Man måler
socioøkonomi
på mange
måder**

- yngre drenge
- der kommer fra hjem med høj socioøkonomisk baggrund
- som er jævnt enige i at de gerne vil lære mere matematik
- som synes, at man hurtigt lærer noget i matematik
- som er *enige* i, at de gerne vil lære noget mere i matematik (til forskel fra den svage gruppe som er *meget uenig* i dette)
- som er *enige* i, at de kan lide at gå i skole (til forskel fra de svage som er *meget enige* i, at de ikke kan lide det)
- som har forældre, der er født i Danmark – og som selv er født i Danmark

**Graden af
sammenhæng
bag den
negative
sociale arv**

Billedet er nogenlunde det samme for gruppen af dygtige elever i naturfag, dog med en svagere kønsbunden forskel, idet der er en lille overvægt af drenge, 59% i gruppen elever, der klarer sig 'godt' mod kun 50% i gruppen af elever, der klarer sig 'skidt'.

Strukturelle faktorer bag ved et TIMSS resultat

– den socioøkonomiske baggrund

Det er siden PISA 2000 blevet almindeligt ved analyser af data, fra de internationale undersøgelser, at inkludere et mål for hjemmets *socioøkonomiske niveau*. Formålet har angiveligt været ønsket om at justere sammenligninger af præstationer mellem grupper af elever med forskellig socioøkonomisk baggrund. Hvis den socioøkonomisk baggrund faktisk *har* sammenhæng med værdien af elevens TIMSS score er argumentationen den, at denne sammenhæng 'ligger bagom', og styrer enhver beregning og vurdering af TIMSS scorerne. Derfor tilsiger samme argumentation, at når elever sammenlignes, fx fra skole til skole, så skal elevernes TIMSS præstationer først 'renses' for den indflydelse, som socioøkonomisk baggrund har på TIMSS scorerne. Det er ikke alle nærværende analyser, som er blevet 'renset' for elevernes socioøkonomiske baggrund.

**Antallet af
bøger
benyttes**

I PISA undersøgelserne, som er rettet mod 15-årige elever, måles den socioøkonomisk baggrund via oplysninger om forældrenes uddannelse, erhverv og indkomst. Hvis man har adgang til registerdata (Danmarks Statistik) bliver indkomsten præcist fastsat, ellers 'regner' man den ud fra elevens oplysninger i spørgeskemaet om forældrenes beskæftigelse.

**Høj
korrelation er
en
nødvendighed**

I sin enkleste udgave med baggrund i de tre forældrefaktorer uddannelse, erhverv og indkomst, viser PISA's relation mellem socioøkonomisk baggrund og PISA scoren en lineær sammenhæng, som kaldes 'den negative sociale arv'. Sammenhængen er ikke

imponerende stærk, idet den såkaldte R^2 – værdi (for den statistiske regressionsanalyse) ligger på ca. 0.15 i Danmark, svarende til, at 15% af variationen i præstationsscoren 'forklares' ud fra den socioøkonomiske baggrund.

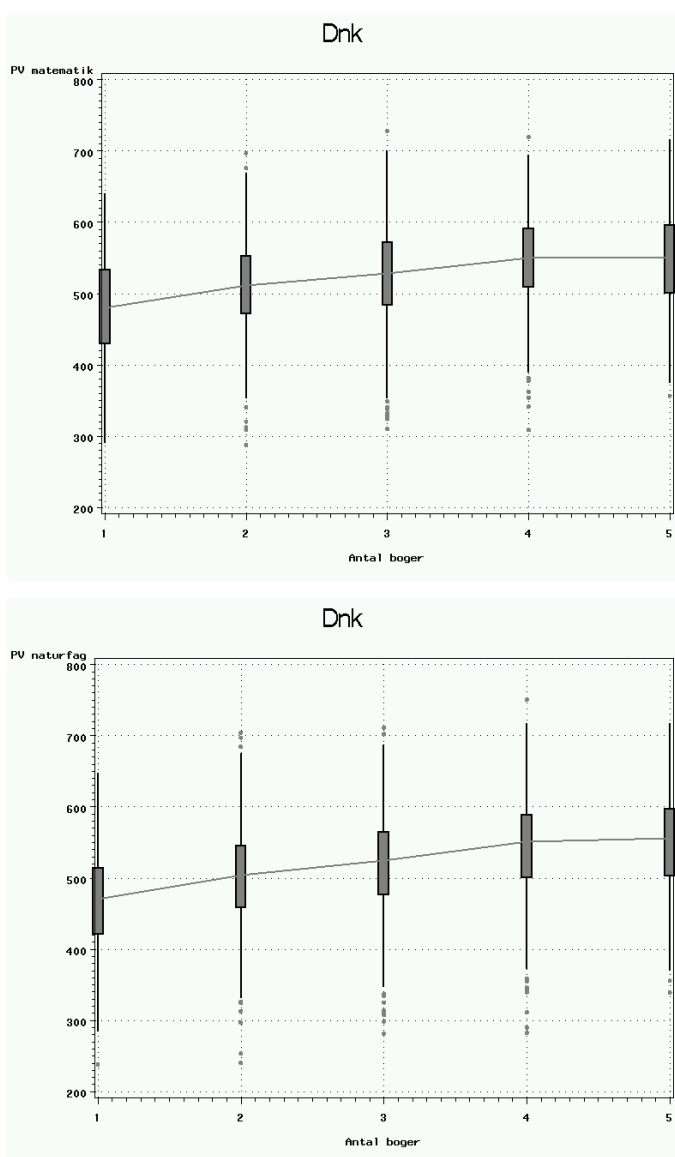
I tilfælde af TIMSS, og specielt i forbindelse med dataindsamlingen på 4. klassetrin, kan man ikke forvente at få pålidelige svar angående de tre nævnte hjemlige forhold, og man må derfor benytte andre fremgangsmåder til belysning af den socioøkonomiske baggrund. Tidligere TIMSS undersøgelser (Beaton et al, 1996) – og med den en lang række internationale undersøgelser - har med succes anvendt 'antallet af bøger' i hjemmet. Det er et mål, som via en passende grafisk illustration i elevspørgeheftet med boghylder, hvor spørgsmålet om bøger kan besvares relativt sikkert af eleverne, og som repræsenterer en blanding af 'økonomi' og 'kultur'. Kulturelle kapitaler er i øvrigt via Bourdieu's definitioner forsøgt indført i bredere sammenhæng i senere udgaver af de internationale undersøgelser uden stort held.

***Relation
mellem
TIMSS scores
og antallet af
bøger***

Det er klart, at der ud over en diskussion af de faglige komponenter, der ligger bag ved mål for socioøkonomi, også skal indgå en mere neutral vurdering af den sammenhæng med TIMSS præstationerne, som hele argumentationen bag det 'at justere for socioøkonomisk baggrund' er bygget på. Under alle omstændigheder skal der være en vis statistisk korrelation mellem præstationsscorerne og målet for socioøkonomisk baggrund. Korrelationen bør med respekt for den teknik, der efterfølgende benyttes til at justere præstationsscorerne, være lineær.

I figur 2 er der i lighed med grundrelationerne bag 'den negative sociale arv' i PISA illustreret, hvorledes antallet af bøger korrelerer med de to TIMSS scores. Antallet af bøger er via elevernes spørgeskemaer kodet som grupperede værdier i et antal intervaller.

Figur 2 Relation mellem TIMSS scores i matematik (PV matematik), og naturfag (PV naturfag) og antallet af bøger i hjemmet.



Den negative sociale arv er ikke stærkere i Danmark end i de andre nordiske lande

Der er i begge figurere en tydelig, næsten lineær tendens til at TIMSS scorerne vokser med antallet af bøger i hjemmet. Dermed kan man acceptere, at antallet af bøger kan benyttes ved statistiske justeringer (covariansanalyser), hvor TIMSS scores sammenlignes over forskellige elevgrupper med forskellig hjemlig baggrund.

De numeriske udtryk for sammenhæng (R^2 –værdier i regressionsanalyse) er af samme størrelsesorden som dem, man fx finder i PISA's angivelse af 'den negative sociale arv'. I tabel 5 er målene for de nordiske lande opgjort.

Tabel 5 Sammenhæng (R^2 –værdier) mellem TIMSS scores og antallet af bøger som mål for den socioøkonomiske baggrund, matematik og naturfag.

*Der er
signifikant
forskul på
niveauerne
fra skole til
skole*

Land	Matematik	Naturfag
Danmark	0.08	0.11
Norge	0.07	0.08
Sverige	0.10	0.14

I Danmark 'forklares' dermed ca. 10% af den generelle variation i TIMSS scorerne ved hjælp af antallet af bøger, i Norge ligger forklaringsgraden på nogenlunde samme niveau, mens Sverige har lidt større værdier af korrelationen.

Man kan derfor med en vis ret sige, at den negative sociale arv, som i PISA er fundet til at være 'stærk' i Danmark (dvs. høj R^2 værdi) sammenlignet med de øvrige nordiske lande, i dette studie *ikke* markerer Danmark som særligt 'tyngt' af den negative sociale arv.

– fire baggrundsvARIABLE styrer en del af TIMSS scorerne

Formålet med statistiske analyser af TIMSS' data er bl.a., at kunne 'forklare' variationen i TIMSS' præstationer mellem forskellige elever og grupper af elever. Det ligger i den måde, at man har fastlagt forsøgsplanen for TIMSS på (Frame Work, se xx), at de variable, der kan komme på tale som 'prædiktorer' af en elevs TIMSS score, må belyses gennem spørgeskemaer til eleverne, til lærerne og skolelederne. Dertil kommer de 'strukturelle' variable, der også indsamles viden om. Til den sidste kategori hører oplysninger om hvilken skole, eleven tilhører, hvilken klasse og fx skolestørrelsen. Når forsøgsplanen i PISA snævert definerer deltagende elever i undersøgelsen, som elever med en præcis alder, er det fx ikke muligt at studere effekten af, at eleverne går i en bestemt klasse. 'Klassen' falder væk som mulig 'årsag' til, at nogle elever klarer sig bedre end andre.

I TIMSS kan elevernes svar sammenholdes med skolens svar (skolelederskema), lærerens svar (lærerskema) og man ved, hvilke andre elever, der går i samme klasse. Derfor er det muligt og relevant at benytte sådanne oplysninger som

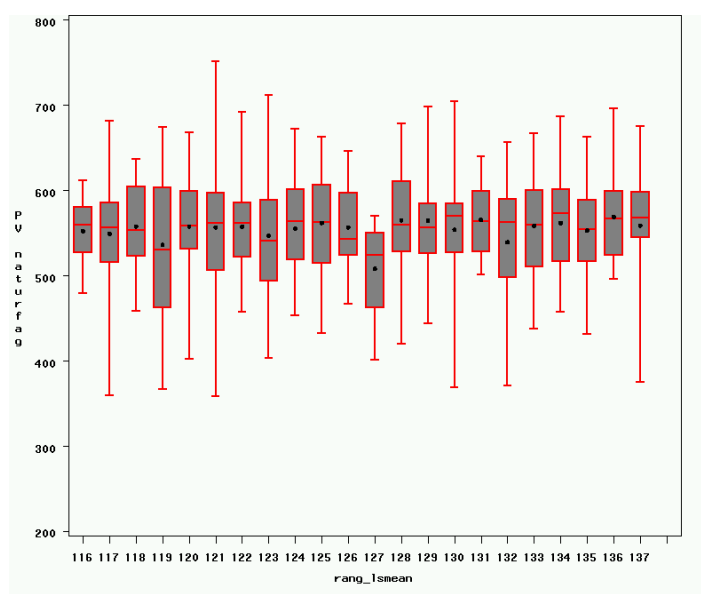
baggrundsfaktorer, når man analyser elevernes præstationer, dvs. som mulige 'forklarende' variable for TIMSS-forskellene.

En fornemmelse af hvilken betydning skolen har, kan illustreres ved at se på TIMSS scorerne fra skole til skole. En mindre del af alle 137 TIMSS skoler indgår i figur 3. Figuren illustrerer inden for naturfag de 22 *gennemsnitligt bedst* scorende skoler og de 23 *gennemsnitligt set lavest scorende skoler*. Figuren (et Boxplot) består af en kasse hvis top og bund markerer de 25% højest og 25% lavest scorende elever. Derudover er der en pind opad, som ender i de 10% højest scorende elever og en pind nedad, der markerer de 10% lavest scorende elever. Prikken i kassens midte er gennemsnittet, mens strengen repræsenterer medianen (50% af eleverne over og 50% under). Skolen der står yderst til venstre har fx et gennemsnit på ca. 350. De svageste 10% ligger på ca. 250, mens de 10% dygtigste lige klarer at passere det internationale gennemsnit på 500.

Der er statistisk set forskel på skolerne, især hvis man foretager en sammenligning fra de 23 lavest scorende til de 22 højest scorende skoler i de to delfigurer i figur 3.

Resultatet er næppe overraskende og spørgsmålet er snarere om man kan 'forklare' hvad der skiller skolerne.

Figur 3 De 22 gennemsnitligt højest scorende skoler og 23 gennemsnitligt lavest scorende skoler i naturfag blandt deltagende 137 skoler i TIMSS. Gennemsnittet for eleverne på en skole er angivet med en prik i kasserne.

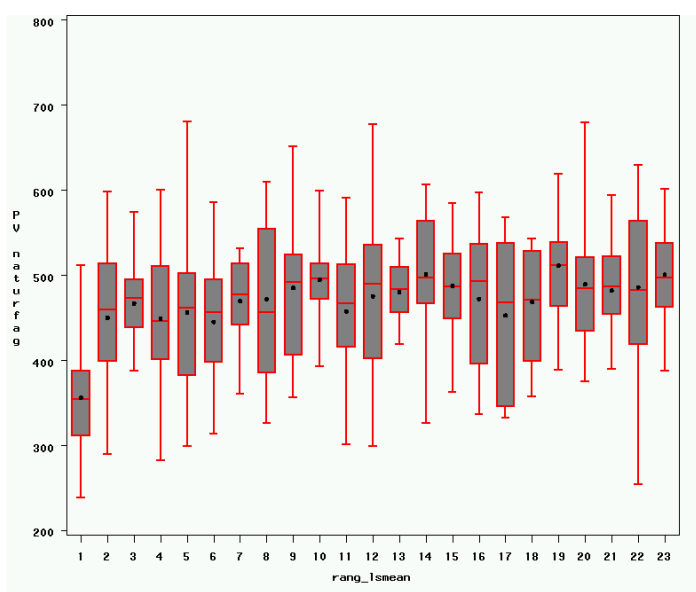


*Man kan
forklare ca.
25-28% af*

**variationen i
TIMSS
scorerne**

**Variation
mellem skoler
reduceres
med 14-16%**

**Køn,
tosprogethed,
etnicitet og
antallet af
bøger
benyttes som
prædiktorer**



En mulig 'forklarende' faktor, der som omtalt ofte bringes i anvendelse, når forskelle mellem grupper af elever skal analyseres, er målet for den socioøkonomiske baggrund – her antallet af bøger. Sammenligningen mellem skolerne kan justeres for forskelle mellem elevernes socioøkonomiske baggrund, således at det er skolernes forventede præstationsniveauer, der sammenlignes under forudsætning af, at de 'stilles lige' mht. til socioøkonomiske baggrund.

**Andre
analyser har
benyttet flere
faktorer**

Justeringen finder sted i en statistisk model, der er fundet velegnet til at beskrive (forudsige) TIMSS scorerne i matematik og naturfag. I nærværende sammenhæng sker det ved, at man ved siden af 'antallet af bøger' inddrager baggrundsfaktorerne 'køn', 'tosprogethed' og 'etnicitet'. Tilsammen udgør disse variable et sæt prædiktorer af de overordnede TIMSS præstationer i matematik og naturfag, som har en forklaringsgrad på 25% i matematik og 28% i naturfag. Altså et effektivt løft i forhold til de 'forklaringsgrader' der i tabel 5 er angivet med antallet af bøger som eneste prædiktør. Den samlede reduktion af variationen fra skole til skole kan udregnes til ca. 14% i faget matematik og til ca. 16% i naturfag. Den oprindelige 'rå' variation mellem skolernes gennemsnit bliver altså reduceret med henholdsvis 14% og 16%, når man justerer (eller kontrollerer) for den socioøkonomiske baggrund, målt ved antallet af bøger. Det betyder, at mange af de signifikante forskelle, der eksisterer mellem de oprindelige skoleniveauer bliver reduceret til at være insignifikante. At der stadigvæk er nogle tilbage, som repræsenterer signifikante skoleforskelle ses af, at ud af de 9316 parvise skolesammenligninger er ca. 27% af dem signifikante, altså et godt stykke over de 5% som ville være et forudsigeligt resultat alene pga. 'statistisk tilfældighed'. I disse vurderinger er variationen mellem skolerne forsøgt beskrevet statistisk set ved

hjælp af baggrundsfaktorerne 'antal bøger', 'køn', 'tosprogethed' og 'etnicitet', altså fire faktorer.

En omfattende analyse af gennemsnitskarakterer (Rangvid, 2008) benytter en længere række af 'hjemmefaktorer' som baggrundsfaktorer. Listen omfatter følgende variable. Flere af variablene er af natur således, at det er nødvendigt at indhente oplysninger via Danmarks Statistik. Elever i 4. klasse vil ikke kunne oplyse om værdier på disse variable, ligesom der vil være problemer mht. til oplysning selv i 9. klasse, hvor fx PISA afholdes.

4.3 Social baggrund: de forklarende variable

Som forklarende variable benyttes en række indikatorer for elevens socioøkonomiske baggrund fra Danmarks Statistiks administrative registre¹². Følgende variable står til rådighed i vores data:

- Elevens køn
- Elevens etnicitet (dansker, indvandrerbaggrund)
- Antal søskende (ingen, én, to, flere end to)
- Plads i søskenderækken (approksimeret ved antallet af yngre søskende)
- Familiestruktur (bor med: begge forældre, enlig far/mor, far/mor i nyt par)
- Forældrenes alder ved barnets fødsel: mor (under 20 år, 20-40 år, over 40 år), far (under 20 år, 20-45 år, over 45 år)
- Fars og mors højeste uddannelse (ufaglært, faglært, studentereksamen, kort videregående uddannelse (KVU), mellemlang videregående uddannelse (MVU), lang videregående uddannelse (LVU))
- Fars og mors erhvervsstatus (lønmodtager, selvstændig/medhjælpende ægtefælle, under uddannelse, på kontanthjælp (ej tilmeldt AF), uden for arbejdsmarkedet (orlov, pension, hjemmegående))
- Fars og mors årsledighedsgrad
- Fars og mors lønindkomst (deflateret med industriens lønindeks)
- Boligens størrelse (kvadratmeter pr. person)

*Skolerne er
ens efter
korrektions
med mange
faktorer*

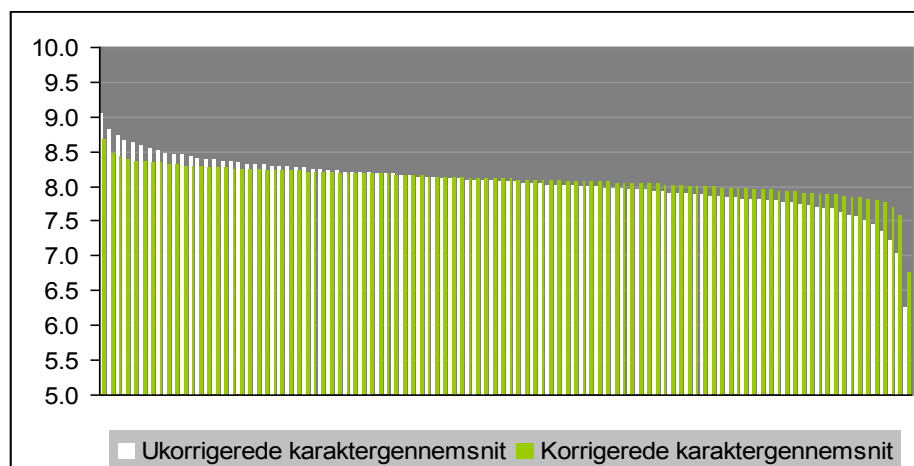
Citat: Beatrice Rangvid, "Skolegennemsnit af karakterer ved folkeskolens afgangsprøver - Korrektions for social baggrund" (Rangvid, 2008).

Det er vigtigt for forsøgene på, at konstruere præcise statistiske modeller, at variationen i den afhængige variabel, der søges 'modelleret', bliver beskrevet så detaljeret som muligt, dvs. man opsøger så præcise prædiktorer som muligt, som opfanger variationen i den afhængige variabel. Øvelsen går bl.a. ud på, at man reducerer uforklaret variation – den såkaldte 'residuale' variation. I disse analyser går det ud på at beskrive variationen af TIMSS scorerne. I den citerede undersøgelse af Rangvid drejer det sig om variationen fra skole til skole, målt ved gennemsnitlige karakterer ved folkeskolens afgangsprøve.

Med anvendelse af de i boksen nævnte variable opnås, at næsten al variation mellem skolernes gennemsnitlige karakterer kan beskrives ud fra den statistiske model. Det

fremgår af figur 4 (Rangvid, 2008), hvor de enkelte skoler er repræsenteret ved deres gennemsnit.

Figur 4 Ukorrigerede og korrigerede gennemsnitlige karakterer ved folkeskolens afgangsprøve.



Skolerne er ikke ens når man korrigerer med få faktorer

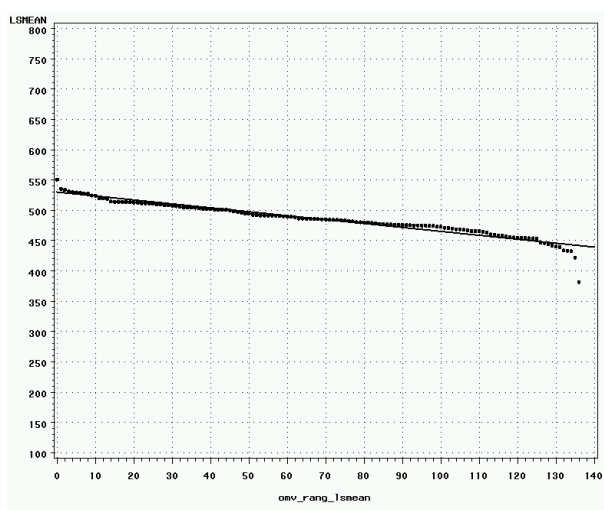
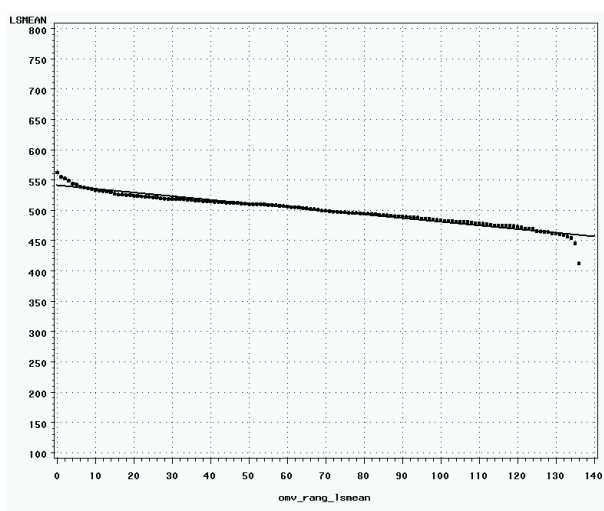
I figuren kan man iagttage variationen mellem skolerne før og efter justering af de gennemsnitlige karakterer. Det er karakteristisk, at S-formen, er meget udtalt for de ukorrigerede gennemsnit, 'flades' ud til en næsten vandret linje, når der korrigeres. Det betyder, at der 'næsten' ikke er forskel på skoler til venstre i billedet sammenlignet med skoler til højre. Beskrivelsen af 'vandrethed' holder for ca. 80% af de skoler, som ligger i midten af figuren. 80% af skolernes indbyrdes variation er således væk, når man justerer gennemsnitskaraktererne mht. til de variable, som listen oven for indeholder. De sidste 20% af skolerne ligger med ca. 10% i hver sin ende og adskiller sig også efter justeringen fra de 80% ens skoler i midten.

I TIMSS analyserne af skoleforskelle er, som nævnt, baggrundsfaktorerne 'køn', 'tosprogethed', 'etnicitet', 'socioøkonomiske baggrund' benyttet sammen med en ikke specificeret skolefaktor, dvs. fem faktorer. Anvendes disse ved justeringen af skolegennemsnittet af TIMSS scorerne, bliver den grafiske analog til figur 4 det, som vises i figur 5.

Figur 5 Fordeling af korrigerede TIMSS scores for matematik (øverst) og naturfag (nederst). Justering mht. til 'køn', 'tosprogethed', 'etnicitet' og 'socioøkonomisk baggrund'.

*Skolerne
bytter plads i
rangnumre*

*Bytte af plads
i rangfølgen
har sine
årsager*



Det faktum, at de indlagte rette linjer ikke er parallelle med X-aksen indikerer, at der stadigvæk er en vis variation mellem de justerede TIMSS-gennemsnit for skolerne, i god overensstemmelse med, at den oprindelige variation mellem skolerne kun blev reduceret med 14-16% .

Det er en følge af justeringerne, at den rangorden, som skolerne placeres i før og efter justeringen, skifter noget, idet det fremgår, at ca. 20% af skolerne skifter pladsnummer. Det betyder en ændring på 15 numre eller flere. Et eksempel på skoler, som foretager store skift mht. rangnummer før og efter justeringen er skole nr. 81 og

skole nr. 69. Skole 81 har gennemsnitlige TIMSS præstationer *før* justering på ca. 550 og 560 for henholdsvis matematik og naturfag, mens de tilsvarende tal for skole 69 er 520 og 510. Altså to skoler, der er lidt forskellige *før* justeringen, hvad rangnumrene 122 (skole 69) og 67 (skole 81) viser. *Efter* justering har de samme skoler fået nye rangnumre: nr. 120 for skole 81 og nr. 80 for skole 69, som dermed har byttet plads med et anseeligt antal numre.

Ser man på baggrundsvariablene for skolerne som er del af forklaringen på, at de to skoler flytter så meget, falder det i øjnene:

- at det gennemsnitlige antal bøger varierer fra 1.78 på skole 69 til 4.40 på skole 81; skolen 81 har således en markant højere socioøkonomisk baggrund
- at 57% taler dansk hjemme i skole 69 mens det er 77% i skole 81

- skolestørrelsens betydning for TIMSS scorerne

Ved flere lejligheder (Allerup, 2006,2007) har *skolestørrelsens* betydning for elevernes præstationer været diskuteret. I TIMSS analyserne er relationen mellem skolestørrelse og elevpræstationerne angivet i figur 6. Punkterne repræsenterer justerede gennemsnitlige værdier af TIMSS scores for de elever, der deltager fra en given skole. Der er taget højde for 'køn', 'tosprogethed', 'etnicitet' og 'socioøkonomisk baggrund' ved gengivelsen af de justerede værdier af elevernes TIMSS scores. Figuren indeholder 137 punkter svarende til 137 skoler. Den indlagte rette linje repræsenterer den bedst tilpassede lineære approksimation til punkterne, og af den medfølgende regressionsanalyse kan man aflæse hældninger på ca. 0.002 i begge delfigurer. Hældninger, som ikke adskiller sig signifikant fra værdien nul, dvs. linjer der er parallelle med X-aksen, og som derfor signalerer, at den justerede TIMSS scorer ikke afhænger af skolestørrelsen.

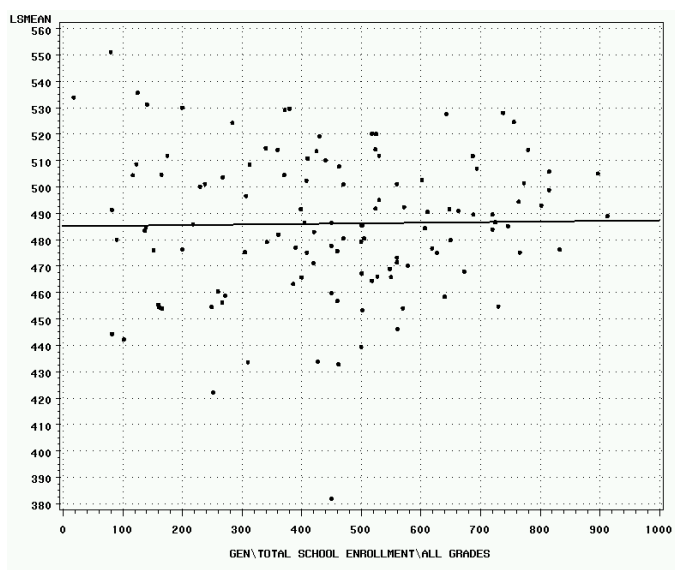
Skole- og klasse størrelse har ingen betydning for elevernes præstationer

Det blev i forbindelse med den tidligere diskussion (Allerup, 2007) fremført, at skoleledelsen har *større* frihedsgrader til at sætte relevante faglærere til undervisningen i de to fag, på store skoler, sammenlignet med de noget strammere betingelser på de små skoler. PISA's tal viste klart, at skoleledere *ikke* føler større allokeringproblemer på små skoler sammenlignet med store skoler. Samme resultat findes ved analyserne af TIMSS' skolelederspørgeskemaer, hvor et spørgsmål gående på, om 'skoleledelse føler, at der er problemer med at skaffe relevante og uddannede lærere i fagene på 4. klassetrin', besvares på helt samme måde for store og for små skoler.

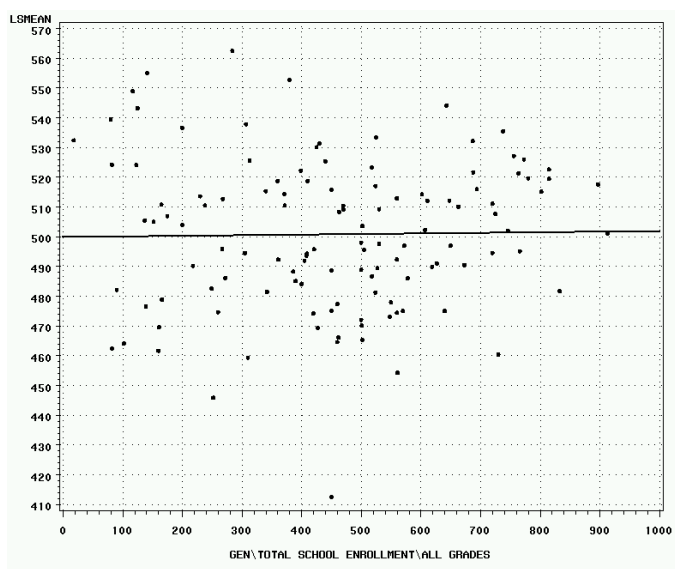
Går man ned på klassestørrelse (Bilag xx), viser beregningerne, at den gennemsnitlige elevpræstation for elever, der går i klasser med 1-19 elever er 529 (matematik) og 521 (naturfag), mens den gennemsnitlige elevpræstation for elever, der går i klasser med 20-32 elever er 521 (matematik) og 517 (naturfag). Den første elevkategori dækker 34% af eleverne, resten udgør 66%. Forskellen er ikke signifikant, og størrelsen af klassen påvirker derfor ikke det gennemsnitlige TIMSS resultat.

Det er derfor tydeligt, at nærværende TIMSS-analyser støtter det tidligere fremførte synspunkt, *at skolestørrelsen ikke har sammenhæng med elevernes præstationer. Dertil er kommet en analyse, der viser at klassestørrelsen heller ikke påvirker det forventede TIMSS resultat.*

Figur 6 Relation mellem gennemsnitlige elevpræstationer (TIMSS scores) pr. skole og skolens størrelse, oplyst fra spørgeskema til skoleleder.



***En skala
måler noget
bestemt***



- andre faktorer, der påvirker TIMSS resultatet

Det er en klassisk vinkel på data fra TIMSS og andre internationale undersøgelser, at man ser på eventuelle kønsforskelle. Der synes at herske nogen forvirring mht. til omfanget og stabiliteten af sådanne forskelle, når man samler resultaterne fra nationale og internationale undersøgelser. Det er derfor ikke altid, man finder bekræftelse af den klassiske hypotese, der postulerer at 'drengene er bedre til matematik og naturfag' i de yngre klasser, mens 'pigerne er bedre til at læse'.

***Det er
nødvendigt, at
opgaverne er
homogene***

At 'være bedre til –' er under alle omstændigheder et udtryk, der bør præciseres i lyset af, at opgaverne er resultatet af TIMSS' homogenitets procedurer. Derved er der fastlagt bestemte statistiske skala-egenskaber ved opgaverne. Ved skalaer benyttes *grupper af opgaver* til at 'måle' med. Hermed menes, at det er den *samlede* besvarelse af hele rækken af opgaver, der fører frem til et mål for elevens dygtighed. Ikke den enkelte opgave. Herved kommer rækken af opgaver til at fungere som en slags målestok, hvorpå de enkelte opgaver kan siges at være delepunkter. Som følge af de omtalte homogenitetsegenskaber er det altid *antallet af rigtigt besvarede opgaver*, der *alene* fører frem til beregningen af elevens *dygtighed*. Man ser ikke på, hvordan de enkelte opgaver er løst.

Hver opgave bliver repræsenteret ved sin 'sværhedsgrad', og hvis samtlige opgaver i samme skala er konstrueret som 'biovidenskabs' opgaver, forsøger skalaen i et sådant tilfælde at måle 'dygtighed' hos eleverne mht. til 'biovidenskab'.

TIMSS' opgaver dækker, jf. tabel 1 tre faglige og tre kognitive områder inden for

***Hvad sker
der, hvis
opgaverne
ikke er
homogene?***

hvert fag. Tilsammen bliver TIMSS således dækket af 12 skalaer for hver elev.

At 'være bedre til – ' betyder derfor, set fra TIMSS' side, at 'have størst TIMSS score uden at se på andet end scoren'. Det er forudsætningerne om homogenitet, der sikrer, at dette kan lade sig gøre, egenskaber, som TIMSS opgaverne alle har været undersøgt for.

Medtager man inhomogene opgaver i sine skalaer, risikerer man at kreere forskelle mellem grupper af elever, som ikke er rimelige. Fx blev der i TIMSS 1995 påvist omfattende kønsforskelle ved besvarelsen af matematikopgaverne. Dvs. forskelle som fremkom, når målingerne blev udført med de skalaer, som var konstrueret i TIMSS 1995. En nærmere analyse (Allerup, 2002) viste, at forskellen skyldtes én enkelt opgave, som uheldigvis blev benyttet som 'sammenbindingsopgave' mellem hefterne. En opgave som selv dygtige piger svarede forkert på. Selv om en dreng og en pige på andre opgaver svarede helt ens – og derfor måtte siges at være lige dygtige – begik pigerne flere fejl på den nævnte opgave. Den var relativt sværere for piger end for drenge, drenge får en slags 'gratispoint' forærende ved sådan en opgave. I sidste ende førte dette til de udmeldinger om kønsforskelle, som ret få fagfolk med rod i nationale erfaringer kunne genkende.

Det er tilsvarende vigtigt, at TIMSS' opgaver bliver undersøgt for homogenitet mht. til etnicitet og tosprogethed. At det i lighed med 'fejlen' i TIMSS 1995 ikke altid er muligt, er (Allerup, 2007) vist tidligere hvor specifikke opgaver i PISA kunne vises at give danske elever nogle 'gratispoint' sammenlignet med elever, som har ikke-dansk baggrund.

I TIMSS 2007 er der som omtalt i alt 12 skalaer inden for de to fag. Målinger og sammenligninger mellem drenge og piger er vist i tabel 6.

**Kønsforskelle
i TIMSS**

Tabel 6 Gennemsnitlige TIMSS scores for piger og drenge inden for matematik og naturfag.

	Domæne	TIMSS scores Piger	TIMSS scores Drenge
Matematik			
	tal (teori)	503	514
	geometri	546	540
	data	527	531
Kognitivt	viden	524	531
	anvendelse	509	516
	ræsonnering	522	527

***Drenge er lige
så dygtige
som piger***

	Domæne	TIMSS scores Piger	TIMSS scores Drenge
Naturfag			
	biovidsk.	527	527
	naturvdk.	500	505
	geovdk.	515	529
Kognitivt	viden	512	519
	anvendelse	511	521
	ræsonnering	528	523

Selv om tabellen antyder en positiv forskel til drengenes fordel i matematik - med geometri som eneste undtagelse – så kan det vises, at kun forskellen 514-503 i 'tal' (teori) er statistisk signifikant. De øvrige markerede forskelle kan derfor betragtes til at være udslag af statistiske tilfældigheder.

***Analyser af
enkeltopgaver
mht. til
kønsforskelle***

I naturfag er det umiddelbare indtryk, at piger og drenge ligger på samme niveau. Statistiske tests bekræfter dette med den undtagelse, at forskellen 529-515 i 'geovidenskab' er eneste signifikante forskel, til drengenes fordel. *En overordnet konklusion er altså, at der er ingen eller få kønsforskelle i TIMSS.*

Der er ikke umiddelbart simple forklaringer på de to markerede signifikante forskelle. Analyser, der går i detaljer for at finde 'årsagerne' til disse forskelle vil bliver gennemført senere. Der er imidlertid visse grænser for mængden af 'ny viden' produceret af sådanne fremtidige analyser. Da TIMSS opgaverne er blevet undersøgt

**Item Maps i
den
internationale
litteratur**

for homogenitet - med et succesfuldt resultat - betyder det, at når man måler en forskel mellem en dygtigere elev og en svagere elev, så skyldes forskellen *alene*, at den dygtigere elev vil kunne svare rigtigt på nogle (relativt sværere) opgaver, som den svagere elev svarer forkert på. Både den dygtigere og den svagere elev vil forventeligt kunne svare rigtigt på opgaver, som, rent sværhedsmæssigt, ligger 'under begge' deres niveauer. Her vil der ikke kunne måles forskel. De afgørende 'forskelle' kommer alene frem i de opgaver, som den dygtigere elev kan svare rigtigt på, men hvor den svagere elev laver fejl.

Ifølge den type psykometri, som opgaverne i TIMSS har været underkastet (Item Response Models, Rasch Modeller), giver det ikke anledning til andre kommentarer end, at 'spændet' mellem opgaverne, som den dygtigere elev kan klare, og opgaverne som den svagere elev kan klare, matcher 'spændet' mellem de to elevers 'dygtigheder'. Betragtninger og fortolkninger af denne forskel tager ikke hensyn til, at der er et bestemt indhold i disse forskels-opgaver, kun at opgaverne har nogle bestemte værdier af sværhedsgrader.

Når en dygtig elev fx kan svare rigtigt på de to opgaver, som er gengivet herunder i figur 7, svarer vedkommende rigtigt på to svære opgaver. Disse opgaver vil forventeligt ikke kunne besvares rigtigt af en elev med et middelmådigt niveau af 'dygtighed'.

Med denne forskel for øje, vil det være en oplagt fejl for fortolkningen - men ikke desto mindre én, der ofte mødes – at sige, 'at den dygtige elev ved meget om skildpadder og deres formeringsbetingelser' og 'meget om smørs evne til at klistre bønner under varmpåvirkning'. Altså et forsøg, om end overdrevet, på at dreje fortolkningen af forskellen ind på selve indholdet bag ved opgaven, dvs. på *mere* end bare sværhedsgraden.

I de internationale rapporter manifesterer denne slags konklusioner sig i de såkaldte 'Item Maps', hvor en graf med elevernes dygtigheder gengives på den ene side af en lodret akse, og opgavernes sværheder på den anden side. Således forsøger Item Maps at 'konkretisere', hvad det er for opgaver *med beskrivelse af deres indhold* som eleverne kan klare eller ikke klare ud fra elevernes 'dygtigheder'.

**Muligheder
for at foretage
formative
evalueringer**

Figur 7 Eksempler på to svære opgaver i naturfag.

15 Der er en kæmpe-skildpadde, der lever på ø. Han er den eneste skildpadde, der er tilbage af en særlig slags kæmpe-skildpadde.
Kan han formere sig, så denne slags skildpadde ikke uddør?

(Sæt kryds i én boks.)

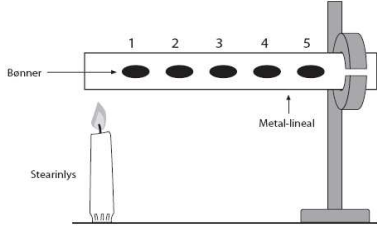
☐ Ja

☐ Nej

Giv en begrundelse for dit svar.

S07_02

19



Bønner er sat fast på en metal-lineal med smør, som vist på figuren ovenfor. Linealen opvarmes i den ene ende. I hvilken rækkefølge vil bønnerne falde af?

(A) 1, 2, 3, 4, 5
(B) 5, 4, 3, 2, 1
(C) 1, 3, 5, 4, 2
(D) Alle på samme tid

S03_06

**Når elever
begår fejl er
der adgang til
god
information**

**Der findes
forskellige
opgavetyper i
TIMSS**

Der ligger i TIMSS' kodesystem af svarene, nogle gode muligheder for at foretage *formative* eller 'diagnostiske' evalueringer af eleverne. Under kodningen er der mulighed for ikke alene at kode information om rigtigheden af svaret. Den grundlæggende kodning indeholder også information om det detaljerede svar. Fx vedrørende *hvilken* svarmulighed, eleven har valgt, hvis der er svaret forkert. Ligesom der er muligheder for at score bestemte åbne svar. Dermed kan man 'forstå' eller fortolke en registreret kønsforskel på anden måde end blot det simple *normative* udsagn, som bæres frem af sammenligningen mellem TIMSS scorerne.

TIMSS opgaverne falder som omtalt i to, måske tre grupper: 'Multiple Choice' (MC) -

**Formativ
evaluering ud
fra analyser
af profiler**

er et opgaveformat, hvor svaret samt nogle alternativer på forhånd er angivet, fx med bogstaver som i det sidste eksempel i figur 7. Ud over det rigtige svar gives der her mulighed for at vælge blandt tre andre svarmuligheder, tre distraktorer, som alle repræsenterer 'fejl svar'. Det ligger i forlængelse af bemærkningerne oven for, at sådanne distraktorer ikke kan vælges helt tilfældigt, de skal være psykometrisk set i orden i forhold til at svare på den rigtige kategori. Der ligger, som nævnt undertiden god information i at konstatere, at en given elev faktisk har valgt en bestemt distraktor. Måske er det oven i købet muligt at se på flere opgaver under ét, så man kan samle indtrykket af, om der er et *mønster i den måde*, eleven vælger distraktorerne på – altså at svare 'forkert' på en bestemt, organiseret måde. Det er før lykkedes at foretage sådanne analyse af distraktorerne med henblik på at lave 'elevprofiler' defineret ud fra fejlsvarene (Allerup et al.2006).

De to andre typer af opgaveformater er begge 'åbne opgaver' og er, rent formatmæssigt varianter af den første opgave i figur 7. I denne opgaves sidste del bedes eleven redegøre for sit svar. Ca. halvdelen af TIMSS' opgaver er konstrueret som åbne opgaver, hvor der skal gives skriftlige begrundelser eller redegørelser i stedet for at svare i et MC-format. Svarene på disse åbne opgaver, mere end 250.000 i alt, blev alle kodet efter et fælles kodesystem. Rigtigheden af svaret blev afgjort, og der var mulighed for at registrere bestemte rigtigheds- og fejlvarianter, som kan tænkes at have didaktisk interesse i den formative evaluering. Det anvendte kodesystem var gældende for alle deltagende lande, men det var muligt at tilføje nationale varianter af 'interessante fejlsvare' som man måtte ønske at analysere også.

Denne type formativ evaluering af elevsvarene vil stå i centrum i de kommende mere detaljerede analyser, som TIMSS data bliver underkastet i fremtiden. Gennem møder med lærere i fagene vil disse analyser forhåbentlig bidrage til brugbar information i den daglige undervisning. Den formative evaluering båret af elevprofiler er et eksempel på anvendelse af elevsvarene på en anden måde end 'blot' at beregne antallet af rigtige svar. Desværre sker det ofte, at ønsket om at kunne tegne faglige profiler via de faktisk afgivne svar møder det problem, at opgaver, som man ønsker skal indgå som 'basis' i profilen, ikke kan findes i samme opgavehefte. En forudsætning for at kunne tegne profiler er, at man har adgang til samme elevs svar på en række af opgaver.

Nogle få eksempler illustrerer baggrunden for at tegne profiler:

**Elevernes
etniske og
sproglige
baggrund**

- i en matematikopgave skal eleverne finde resultatet af regnestykket: **5.3-3.8**.
Det viser sig at, 65% af pigerne og 51% af drengene finder det rigtige svar. To helt specifikke 'måder' at regne forkert på er kodet som 70 og 71, mens standardkoden for 'andre fejltyper' er 79. Her er svarfordelingerne for piger 17%, 5%, 13% og tilsvarende for drenge 25%, 3%, 21% - dvs. markante forskelle både mht. til at finde det rigtige svar og mht. at begå særligt udvalgte fejltyper.
- I en geometriopgave skal eleven tegne en 4-sidet figur. Det lykkes 78% for pigerne og for 67% af drengene. To specielt udpegede måder at lave fejl er kodet 70 og 71, og igen er 79 koden for 'alle andre fejl'. Fordelingen på disse kategorier viser sig at være 1%, 0%, 21% og for drenge 1%, 5%, 27%. Fejltypen '71' begås derfor næsten udelukkende af drenge.
- I matematikopgaven nedenfor viser det sig, at det rigtige svar (C) besvares af 45% piger og 45% drenge. De øvrige svarmuligheder, distraktorerne besvares af pigerne med hyppighederne 22%, 25% og 8% for pigerne og 12%, 32% og 11% for drengenes vedkommende. Igen en ret stor forskel i valget af distraktor, når eleven ikke vælger det rigtige svar, som er (C). Pigerne 'afkoder' fx indholdet af (A) hyppigere end drengene til deres valg. Der er jo forskel på, om man *først* lægger 1 til og så ganger med 2 *eller omvendt*!

2, 5, 11, 23, ...

Det viste talmønster starter med 2. Hvilken af de følgende regneregler giver tallene i mønsteret?

- (A) Læg 1 til og gang med 2 for at få det næste tal.
- (B) Gang med 3 og træk 1 fra for at få det næste tal.
- (C) Gang med 2 og læg 1 til for at få det næste tal.
- (D) Træk 1 fra og gang med 3 for at få det næste tal.

- Elevernes etniske og sproglige baggrund som faktorer, der påvirker TIMSS resultatet

Eleverne har i spørgeskemaet svaret på nogle spørgsmål, der giver oplysninger om etnisk baggrund og det sprog, som tales i hjemmet. Etniciteten kan i lighed med andre undersøgelsers definitioner fastlægges ud fra forældres og eget fødselssted. I denne

analyse fastlægges etniciteten ud fra eget fødselssted – er det uden for Danmark, er man indvandrer. De sproglige aspekter belyses ud fra elevens vurdering af, hvor hyppigt dansk anvendes hjemme: 'altid', 'næsten altid', 'en gang i mellem' og 'aldrig'.

Disse definitioner betyder, at der i TIMSS materialet er 10% indvandrere, og 25% af eleverne, som ikke altid taler dansk i hjemmet (dvs. svar i kategorierne 'næsten altid', 'en gang i mellem' og 'aldrig'). Hyppighederne passer ret godt med tal fra Danmarks Statistik og definitionerne af de to begreber. I oversigtsformen er hyppighederne gengivet i tabel 6.

Tabel 6 Hyppigheder af elever med dansk og ikke-dansk baggrund.

gruppe	Dansk	Ikke dansk	Sprog dansk næsten altid	Sprog dansk engang imellem	Sprog dansk aldrig
etnicitet	90%	10%			
Sprog	75%	25%	18%	6%	1%

Med disse definitioner på etnicitet og tosprogethed kan man beregne de gennemsnitlige TIMSS præstationer. Tabel 7 gengiver resultaterne for matematik og naturfag.

Tabel 7 Oversigt over præstationer af elever med dansk og ikke-dansk baggrund.

*Når TIMSS
scorer
justeres for
elevernes
etniske og
sproglige
baggrund*

	Matematik				
gruppe	Dansk	Ikke dansk	Sprog dansk næsten altid	Sprog dansk engang imellem	Sprog dansk aldrig
etnicitet	528	486			
Sprog	528	511	525	471	493

	Naturfag				
gruppe	Dansk	Ikke dansk	Sprog dansk næsten altid	Sprog dansk engang imellem	Sprog dansk aldrig
etnicitet	522	479			
Sprog	523	500	519	447	461

Tabel 7 angiver de 'rå' direkte beregnede gennemsnit. Det er klart fra tabellerne, at

elever med ikke-dansk baggrund gennemsnitligt set har lavere TIMSS præstationer end danske elever. Forskellen er statistisk signifikant, og de markerede forskelle ud fra TIMSS scorerne kan oversættes til:

**Elevernes
sproglige
baggrund**

- at '*den sproglige effekt*' dansk - ikke dansk, svarer til en forskel i rigtighedsprocent på 5%-7%
- at '*den etniske effekt*' dansk - ikke dansk, svarer til en forskel i rigtighedsprocent på 11%-13%

Man kan imidlertid anvende den tidligere statistiske model, der benytter skolen, køn, etnicitet, sprog og antal bøger som prædiktorer af TIMSS resultatet til at justere disse forskelle i forhold til elevernes socioøkonomiske baggrund (antallet af bøger bliver benyttet som covariat). Som omtalt er den statistiske tilpasning af de to modeller acceptabel med forklaringsgrader på 25% henholdsvis 28% af variationen i TIMSS scorerne ($R^2=0.25$ og $R^2=0.28$) for henholdsvis matematik og naturfag. De justerede eller korrigerede TIMSS præstationer er opført i tabel 8.

Tabel 8 Oversigt over korrigerede præstationer af elever med dansk og ikke-dansk baggrund.

**Elevernes
holdninger til
matematik og
naturfag**

	Matematik				
gruppe	Dansk	Ikke dansk	Sprog dansk	Sprog dansk	Sprog dansk
etnicitet	515	498	næsten altid	engang imellem	aldrig
Sprog	516	505	513	485	488

	Naturfag				
gruppe	Dansk	Ikke dansk	Sprog dansk	Sprog dansk	Sprog dansk
etnicitet	498	474	næsten altid	engang imellem	aldrig
Sprog	513	496	508	466	456

Bortset fra en lille 'udjævning' af tallene, der går ud på, at der ikke er signifikant forskel på niveauerne svarende til 'dansk (altid)' vs. 'dansk næsten altid' og 'engang imellem' vs. 'aldrig', så er de signifikante forskelle, som de ses i tabel 7 opretholdt i tabel 8. Det er derfor *ikke* sådan, at elevens socioøkonomiske baggrund kan 'forklare'

**PAT indekset
måler graden
af positiv
indstilling til
matematik og
naturfag**

de i tabel 7 fundne forskelle – forskellene *er* der, også når man justerer tallene på en måde, så sammenligningerne gennemføres under forudsætning af, at eleverne har samme socioøkonomiske baggrund.

Danske elevers forhold til matematik og naturfag

Det er velkendt at eleverne, i de højere klassetrin af folkeskolen, ikke har naturfag som 'bedste og sjoveste' fag. Faktisk er ét af problemerne, at så *få* elever synes, at fagene på 8. til 9. klasseniveau, og det gymnasiale niveau, er spændende og udfordrende.

I 4. klasse har eleverne ikke færdige forestillinger om, hvad man kan bruge fagene til senere hen, og man kan måske derfor antage, at når eleverne svarer på spørgsmål som 'jeg kan godt lide at lære om matematik', 'matematik er kedeligt', 'jeg kan godt lide matematik', så vil svarene i høj grad være udtryk for en umiddelbar fornemmelse, skabt af undervisningen i klassen. Svarene på de tre spørgsmål er passende omkodet og bragt på en form, så man via et indeks kan måle elevens 'enighed' i om fagene er spændende, og noget de kan lide at lære mere om. Den målte 'enighed' går på en skala fra 'meget uenig' til 'meget enig'.

Tabel 9 Hyppighed af elever med Højt (Høj), Medium (Med) og Lavt (Lav) værdi af på PAT-indekset til registrering af *elevernes positive holdning til fagene* (Positive Affect Towards), og deres gennemsnitlige TIMSS scores i matematik fordelt Høj PAT-værdi svarer til mest positiv holdning (PATM i matematik og PATS i naturfag).

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DK	55%	526	24%	521	21%	523	3 point
	Int.	72%	488	14%	457	14%	454	29 point
Nat	DK	59%	525	20%	510	21%	507	18 point
	Int.	77%	485	13%	456	11%	452	33 point

**Danske elever
elsker ikke
fagene
matematik
eller naturfag
men klarer
sig godt
alligevel**

Det fremgår af tabellen, at Danmark med tallene 55% og 59% mønstrer færre elever, der har en positiv holdning til fagene end man ser internationalt, hvor tallene er 72% og 77%. Faktisk ligger de 55% danske elever i matematik næstnederst blandt landene hvad angår 'høj' værdi af 'positiv holdning' – kun i Taiwan er der med 50% færre elever, som har en 'høj' positiv holdning. Men ligesom i Taiwan præsterer de 'få' danske elever i denne gruppe alligevel høje TIMSS scores, gennemsnitligt 526 for matematik og 525 for naturfag.

***Elevernes
selvtillid i de
to fag måles***

Ser man på forskellen (vandret) over de tre grupper af elever, mødes man internationalt af en ændring på 29 – og 33 TIMSS score point henholdsvis for naturfag og matematik. Forskellene er signifikante og går i en retning, der harmonerer med den opfattelse, at jo mere positiv holdning man har, jo bedre klarer man sig.

Det samme kan ikke siges, at være gældende blandt danske elever. Dels ligger matematikpræstationerne 526, 521, 523 på konstant niveau tæt på det generelle gennemsnit for alle elever (523). Dels flyttes tallene op og ned for naturfags vedkommende. Det betyder, at selv om forskellen mellem ydergrupperne i dette tilfælde er 18 point, dækker ændringen ikke, som det ses internationalt, over en *systematisk* falden af præstationerne i takt med at den positive holdning indskrænkes.

Det bliver derfor en konklusion, at danske 4. klasseelever i gennemsnit relativt set har en udpræget ikke-positiv holdning til matematik og naturfag, men at disse holdninger ikke har indflydelse på deres evne til at frembringe gode TIMSS resultater.

Det er til diskussion, om gruppen af 'meget positive' elever på 55% og 59% er et bekymrende lille tal. Set fra den tidligere omtalte forventningsnegative synsvinkel på elevernes holdning til de to fag, er tallene alligevel udtryk for mere positiv holdning end den, eleverne vokser ind i med alderen, og som de giver udtryk for i elevspørgeskemaer på 8. og 9. klasseniveau.

***Danske
elevers
selvtillid i top***

Ved siden af elevernes holdninger til fagene belyst gennem de to indekssværdier PATM og PATS er der via elevernes spørgeskema adgang til at vurdere elevernes grad af *selvtillid*. Det sker ud fra spørgsmålene (her matematik) 'jeg er normalt god til matematik', 'matematik er sværere for mig end for mine kammerater', 'jeg er bare ikke god til matematik' og 'jeg lærer hurtigt tingene i matematik'. Beregningerne er samlet i de to indekssværdier SCM og SCS for henholdsvis matematik og naturfag.

Tabel 10 Hyppighed af elever med Højt (Høj), Medium(Med) og Lavt (Lav) værdi af elevernes indeks SCS (naturfag) og SCM (matematik) til registrering af *elevernes selvtillid* samt gennemsnitlige TIMSS scores. Høje værdier af SCS og SCM svarer til høj grad af selvtillid.

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DK	70%	540	23%	493	7%	469	71 point
	Int.	57%	500	32%	449	11%	429	71 point
Nat	DK	68%	531	26%	494	7%	485	46 point
	Int.	61%	497	30%	453	8%	437	60 point

Drenge er mere selvsikre end piger

En umiddelbar sammenligning mellem Danmark og det internationale gennemsnit viser en overvægt af danske 4. klasseelever med 'høj' selvtillid. Både i matematik og i naturfag. I matematik overgås mængden af elever med 'høj' selvtillid kun af tre lande, herunder Sverige, hvor 77% af eleverne lander i denne kategori. I naturfag er 68% af de danske elever udstyret med 'høj' grad af selvtillid, hvilket er mange. Her finder man 9 lande (ikke Sverige) med relativt flere elever i denne gruppe end i Danmark. *Billedet af danske elever som meget 'selvsikre' elever fremgår således af de mange elever i kategorien 'Høj' og kompletteres af en lille hyppighed af elever i den modsatte ende af SCS- og SCM skalaen med 'lav' grad af selvtillid.* Som det fremgår af tabellen er der, især i matematik (7%), færre elever i denne kategori i Danmark sammenlignet med andre lande.

Tabel 11 Hyppighed af piger (P) og drenge (D) med Høj (Høj), Medium (Med) og Lav (Lav) værdi af elevernes indeks SCS (naturfag) og SCM (matematik) til registrering af *elevernes selvtillid*. Høje værdier af SCS og SCM svarer til høj grad af selvtillid.

Fag		Høj		Med		Lav	
		%elever		%elever		%elever	
Mat	P	66%		25		9	
	D	73%		21		6	
Nat	P	66%		26		8	
	D	69%		25		5	

Danske elever har også hjemme arbejde

Tabel 10 ser på eleverne under ét. I tabel 11 er eleverne delt op efter køn. Det strider

ikke mod forventningerne at se, at andelen af 'selvsikre' elever især findes inden for drengene: 73% ⇔ 66% i matematik og 69% ⇔ 66% i naturfag, selvom forskellen i naturfag ikke er stor.

Når danske elever arbejder hjemme

Brug af hjemmearbejde veksler fra lærer til lærer, og er ikke omfattet af fælles regler. Der er desuden vekslende syn på de pædagogiske fordele ved at benytte hjemmearbejde. Hvem skal fx have fordel af det?⁷ I TIMSS måles graden af hjemmearbejde gennem en kortlægning af anvendt tid og hyppighed. Disse to aspekter er samarbejdet i et indeks for den enkelte elev, som måler 'hvor meget' hjemmearbejde eleven har anvendt. Passende grupperet i tre kategorier 'høj', 'medium' og 'lav' er eleverne kategoriseret i tabel 12 sammen med deres gennemsnitlige TIMSS scores.

Tabel 12 Hyppighed af elever med Højt (Høj), Medium (Med) og Lavt (Lav) forbrug af hjemmearbejde og deres gennemsnitlige TIMSS scores i matematik og naturfag.

Dygtige elever i Danmark laver ikke hjemmearbejde

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DK	23%	514	52%	524	25%	538	24 point
	Int.	21%	469	58%	479	21%	468	1 point
Nat	DK	2%	-	13%	501	85%	522	- point
	Int.	9%	446	35%	474	57%	488	42 point

Det ses fra tabel 12, at udbredelsen af hjemmearbejde (23%, 52% og 25%) inden for faget matematik næsten er, som det gennemsnitligt finder sted i andre lande (21%, 58% og 21%). Det ses i form af den hyppighed hvormed hjemmearbejde anvendes. Hvis 'effekten' af hjemmearbejdet i matematik gøres op i form af indflydelse på TIMSS scoren, ses der ikke store variationer internationalt. Mens man fristes til den konklusion i Danmark, at det er de svage elever, der i 'høj' grad udsættes for hjemmearbejde, idet de 23% af eleverne som hyppigt eller i 'høj' grad udsættes for hjemmearbejde, gennemfører TIMSS med et gennemsnit på 514. Elever med 'lavt' forbrug af denne aktivitet gennemfører TIMSS med et gennemsnit på 538, noget over

⁷ Seneste debat fandt sted i oktober 2007, hvor undervisningsminister Bertel Haarder opfordrer til at fastholde hjemmearbejdet som en værdifuld aktivitet

det danske gennemsnit for elevgruppen som helhed på 523.

Det internationale konstante niveau af TIMSS scorerne kunne forlede én til at tro, at der i alle lande mangler sammenhæng mellem brug af hjemmearbejde og niveauet af TIMSS præstationerne. Altså at der internationalt set, er påvist manglende 'effekt' af hjemmearbejdet. Det vil muligvis være en fejlslutning, baseret på et – ikke på nuværende kendt resultat – at det påpegede konstante niveau i virkeligheden dækker over stærkt varierende systematikker i de forskellige lande. Systematikker, hvor nogle lande øger TIMSS score niveauet i takt med mere brug af hjemmearbejde og nogle lande, hvor det modsatte gør sig gældende.

Fagligt indhold i timerne – hvad fylder det?

I naturfag benyttes hjemmearbejde næsten ikke i Danmark i 'høj' grad (2%). Også internationalt er man på dette klassetrin tilbageholdende med 'høj' grad af hjemmearbejde (9%). Næsten alle elever (85%) i Danmark har intet eller i ringe omfang hjemmearbejde. Det er mere, end man ser internationalt, hvor der 'kun' findes 57% i denne kategori. *Både internationalt og i de tal, der er tilgængelige for Danmark ses samme tendens, som fandtes i Danmark for matematiks vedkommende: Lav udnyttelse af hjemmearbejde er kombineret med dygtigere elever.*

Undervisningen – tid og indhold

Den faktisk gennemførte undervisning i matematik og naturfag kan anskues fx ud fra, hvor 'meget' de enkelte dele af faget fylder i den daglige undervisning. Altså en relativ betragtning, hvoraf man kan se, hvordan de tre hovedområder, som TIMSS har valgt, er vægtet i forhold til hinanden af skolen og lærerne. Et andet aspekt af tid og indhold er den samlede årlige undervisningstid i de to fag. Oplysninger over begge forhold findes i tabel 13.

Betyder antallet af timer i faget noget?

Tabel 13 Procentvis fordeling af tid på de tre grundfaglige områder i matematik og naturfag – og samlet gennemsnitlig tid (timer pr. år).

Fag		tal	geometri	Data	Andet	Samlet tid (årligt)
Mat	DK	49%	26%	18%	7%	125
	Int.	50%	24%	16%	4%	144
		Biov.	Naturv.	Geov.		
Nat	DK	37%	26%	33%	4%	59
	Int.	40%	25%	24%	11%	67

Der er ingen sammenhæng mellem TID og præstationer internationalt

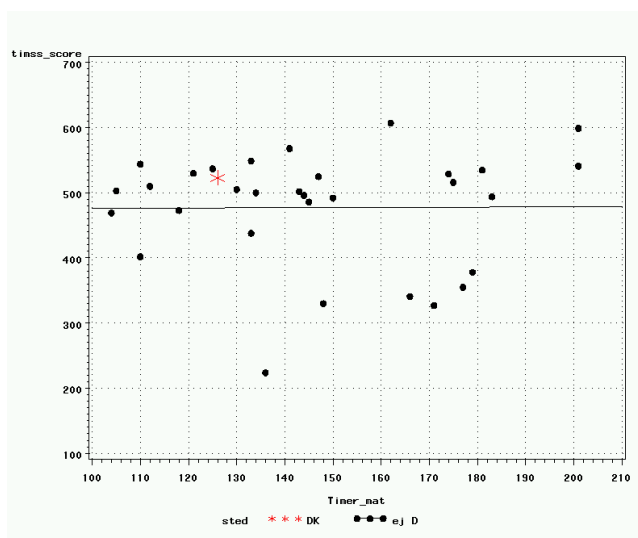
Det fremgår af tabellen, at den interne fordeling på tre delområder, som hvert af

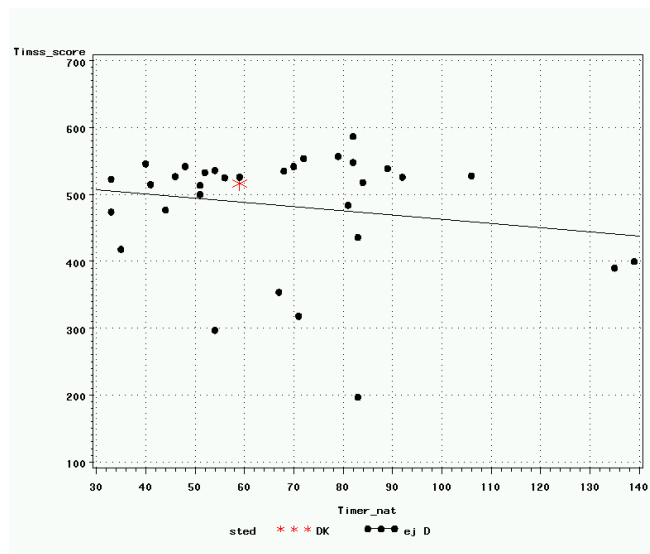
fagene er bygget op af, vægtes nogenlunde, sådan som det sker gennemsnitligt uden for Danmark. Dog ser det ud til at 'geovidenskab' fylder noget mere i danske læseplaner på 4. klassetrin sammenlignet med andre lande. Også det samlede antal undervisningstimer, der årligt er afsat til de to fag, er sammenlignelige med lande uden for Danmark – dog med tendens til at der bruges lidt færre timer, svarende til ca. 15%, til fagene i Danmark sammenlignet med andre lande.

Men betyder antallet af timer i faget noget for det præstationsniveau som testes ved TIMSS? Sammen med spørgsmålet om størrelsen af klassen og/eller skolen er dette spørgsmål om tid et tilbagevendende tema ved undersøgelser som TIMSS – internationale eller nationale. I de følgende figurer 8 og 9 belyses spørgsmålet internationalt og nationalt.

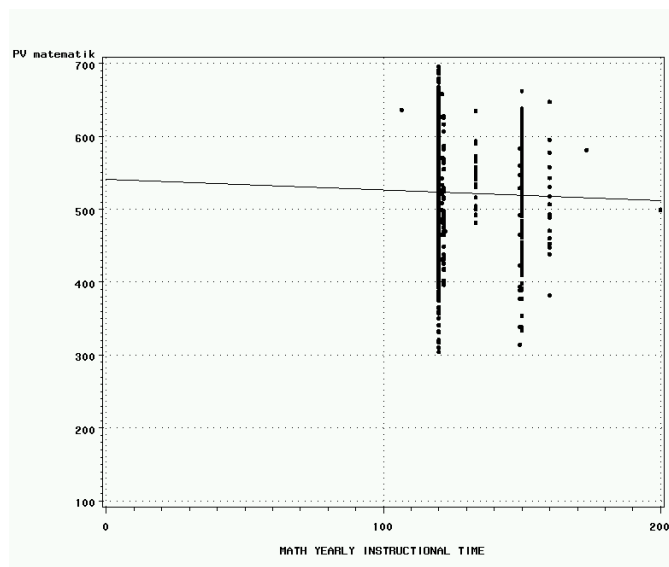
Figur 8 Relation mellem TIMSS score i matematik og naturfag og det årlige antal timer på undervisningsplanen. *Alle 37 deltagende lande.* Danmark er markeret med en stjerne (naturfag er første deltegning, derefter matematik).

**Der er ingen
sammenhæng
mellem TID
og
præstationer i
Danmark**



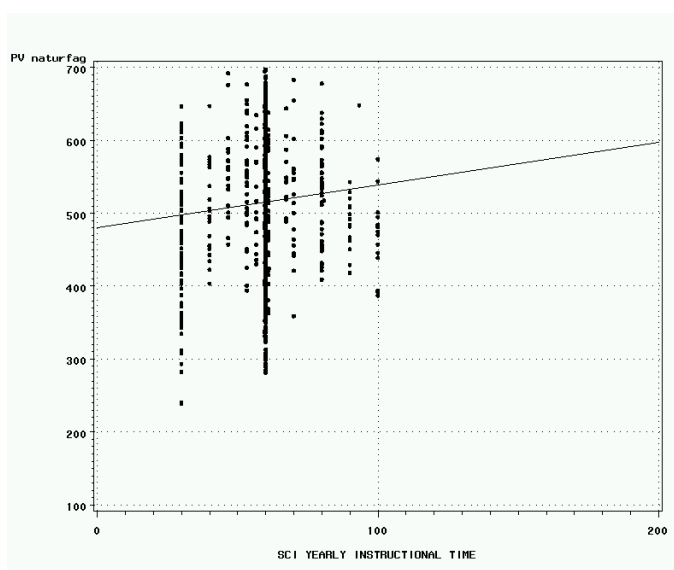


Figur 9 Relation mellem TIMSS scorer i matematik og naturfag og det årlige antal timer på undervisningsplanen *alene for deltagende skoler i Danmark*. Hvert punkt repræsenterer en elev.



*Føler lærerne
at de er
ordentligt
'klædt på'*

*Danske
naturfags
lærere er ikke
helt enige*



I figur 8 belyses dette internationalt ved at gengive gennemsnitstal for 37 deltagende lande. For naturfags vedkommende ses ingen sammenhæng. Den indlagte regressionslinje har hældningen nul, mens samme linje for matematiks vedkommende er negativ. Dette er et numerisk udtryk for, at den gennemsnitlige TIMSS præstation falder med et øgende antal timer på læseplanen. Imidlertid kan det ses, at det er ganske få afvigende lande, der er ansvarlige for dette indtryk, og matematikrelationen kommer til at ligne relationen i naturfag, hvis man ser væk fra disse punkter.

Den internationale konklusion er derfor, som fundet mange gange tidligere, at der ingen sammenhæng er mellem antallet af timer på læseplanen og præstationsniveauet.

*Er lærernes
arbejds-
betingelser
passende?*

I figur 9 undersøges spørgsmålet nationalt, idet punkterne alle stammer fra de deltagende danske skoler. Hvert punkt refererer til én elev. Mens matematik ser ud til at præsentere en faldende TIMSS præstation med stigende årlige timer på læseplanen, gør det modsatte sig gældende for naturfags vedkommende. Under alle omstændigheder er relationerne ikke tydelige, og de indlagte linjer repræsenterer ikke et system i punkterne, som alle er fælles om. Specielt ser punkterne i naturfagsrelationen ud til at falde med helt store værdier af antallet af årlige timer. Også i dette tilfælde med data fra de danske elever, bliver den korte konklusion, at antallet af timer på læseplanen ikke har sammenhæng med TIMSS' præstationsniveauer for eleverne.

Ikke helt passende siger både matematik og naturfagslærere

Forberedelse til undervisningen – elever og lærere

Blandt de faktorer, som har betydning for at eleverne opnår gode resultater af undervisningen, er den 'lærerfaktor', der har med lærerens 'parathed til undervisningen' at gøre. I tabel 14 gengives resultaterne af lærernes svar på spørgsmålet om at føle sig 'godt klædt på til undervisningen'.

Tabel 14 Hyppighed af elever, hvis lærere føler at *de er velforberedte til at undervise i de to fags hovedområder.*

Fag		Tal	geometri	Data
Mat	DK	94%	92%	91%
	Int.	77%	68%	71%
		Biov.	Naturv.	Geov.
Nat	DK	57%	50%	60%
	Int.	59%	46%	56%

Det fremgår af tabel 14, at inden for *faget matematik mødes danske elever i meget høj grad af lærere, som er 'parate' til at udføre undervisning i faget*. Omkring 90% af eleverne befinder sig i denne situation, hvilket er betydeligt flere, end man finder i lande uden for Danmark, hvor kun omkring 72% af eleverne har lærere, som føler, at de er parate.

Situationen er lidt anderledes i faget *naturfag, hvor kun omkring 56% af lærerne føler, at de er parate til at undervise i faget*, med 50% i delområdet 'naturvidenskab' som det laveste. De danske tal adskiller sig ikke væsentligt fra det internationale gennemsnit, som for alle delområderne set under ét ligger omkring 54%.

Det kan bemærkes, at både nationalt og internationalt gælder det, at lærere føler sig mindre parate til at undervise i naturfag sammenlignet med matematik.

Lærernes arbejdsbetingelser

TIMSS har stillet en række spørgsmål vedrørende lærernes arbejdsbetingelser i lærerspørgeskemaet. Spørgsmålene berører lærernes vurderinger af vedligeholdelsesstand af skolebygninger, overfyldte klasselokaler, adgang til lokaler uden for klasseværelset, og i tilfælde af naturfag, adgang til materialer til brug for eksperimenter. Svarene på disse spørgsmål er omsat til beregning af indekssværdierne TAWS og TAWM, i henholdsvis naturfag og i matematik.

Behov for resurser til

**under-
visning**

I tabel 16 er værdierne af TAWS og TAWM grupperet i tre grupper svarende til høje, mellem og lave værdier.

Tabel 16 Hyppighed af elever med Høje (Høj), Medium (Med) og Lave (Lav) værdier af skolernes adgang til at give ansatte lærere tilfredsstillende arbejdsvilkår (TAWS og TAWM, henholdsvis naturfag og matematik). Høje værdier af TAWS og TAWM svarer til at skolen har relativt få problemer med at give lærerne tilfredsstillende vilkår.

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DK	7%	545	58%	522	35%	523	
	Int.	15%	477	56%	475	29%	472	
Nat	DK	6%	527	61%	521	33%	514	
	Int.	8%	490	56%	477	36%	475	

Man ser af tabellen, at danske elever har lærere, der *i matematik* føler, at de har lidt mindre passende arbejdsbetingelser end man finder internationalt. Det ligger dels i halveringen 7% i Danmark, i forhold til de 15% som findes i gruppen internationalt af lærere, der mener, at arbejdsbetingelserne er helt passende. Dels i det faktum, at 35% af eleverne i Danmark mod 29% internationalt har lærere, der mener, at arbejdsbetingelserne bestemt ikke er passende.

Inden for *naturfag* er forskellen mellem udsagnene fra danske lærere og lærere internationalt betydeligt mindre end i matematik. Det bør på trods af dette bemærkes, at man, som i matematik finder, at en tredjedel (33%) af lærerne mener, at deres arbejdsbetingelser *ikke* er passende, og kun 6% mener, at arbejdsbetingelserne 'i høj grad' er passende.

Det er ikke klart fra de internationale opgørelser, hvilket mønster der tegner sig i sammenhængen mellem læreres opfattelse af egne arbejdsbetingelser og elevernes TIMSS præstationer. For både matematik og naturfag er der tale om et numerisk fald i de gennemsnitlige TIMSS præstationsniveauer med faldende grad af passende arbejdsbetingelser. Men i matematik er faldet kun 5 score point, og det er ikke signifikant, mens de 15 points forskel i naturfag er signifikant.

**I Danmark
dækkes
behovene**

For danske elever gælder en lidt modsat tendens end den internationale. *Mens der i matematik er tydelig sammenhæng mellem læreres oplevelse af passende arbejdsbetingelser – jo bedre betingelser, jo højere TIMSS præstationer – antyder de*

13 points forskel (527 $\Rightarrow$ 514) mellem 'høj' og 'lav' en mindre forskel for naturfags vedkommende.

Skolen og resurser til undervisningen

Skoleklimaet betyder noget

Skolen udgør den fysiske ramme for undervisningen. Lærebøger og andre undervisningsmidler, penge til vedligeholdelse af skolebygninger, varme og lys, samt antal klasseværelser er faktorer, der indgår i dette kompleks af rammebetingelser.

Lærere og skoleledere er i TIMSS spørgeskemaerne blevet bedt om, at vurdere om der er passende adgang til at få dækket behov for resurser på de nævnte områder. Svarene er gengivet i tabel 15, hvor tallene er 'set' gennem elevernes øjne.

Forskel mellem skoleledere og læreres vurderinger

Tabel 15 Hyppighed af elever hvor skolen kategoriseres ud fra Høje (Høj), Medium (Med) og Lave (Lav) værdier af skolernes adgang til at få dækket behov for resurser til skolen (ASRMI og ASRSI, oversat fra 'Availability of School Resources for Mathematics/Science') i forbindelse med afviklingen af undervisningen. Høje værdier af ASRMI og ASRSI svarer til, at skolen har relativ let adgang til resurser.

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DK	50%	530	49%	518	1%	-	-
	Int.	36%	480	55%	472	9%	429	51 point
Nat	DK	47%	523	52%	512	1%	-	-
	Int.	31%	483	59%	477	10%	442	41 point

Det ses af tabel 15, at ca. halvdelen (50% i matematik og 47% i naturfag) af eleverne

går på skoler, hvor der skønnes at være god adgang til at få dækket behovet inden for de nævnte områder. Omvendt er gruppen af elever, hvor behovene kun i ringe udstrækning (Lav) dækkes, meget lille (1% i begge faggrupper). *Danmark ligger bedre, end man ser det internationalt, både hvad angår informationerne vedrørende 'høj' grad og 'lav' grad af muligheder for at få dækket specifikke behov for resurser i forbindelse med undervisningen i de to fag*. Også sammenlignet med forholdene i Sverige (Høj matematik = 37%, Høj naturfag = 33%) og i Norge (Høj matematik = 23 Høj naturfag = 27%) er forholdene bedre i Danmark.

Skolen og det interne klima

Skoleledere og lærere har svaret på en række spørgsmål vedrørende deres oplevelser af 'skoleklimaet' på den skole, de arbejder på. Dette 'klima' beregnes ud fra svar på 8 spørgsmål om 'læreres arbejdsglæde', 'læreres forståelse af skolens læseplaner', 'læreres succes med at implementere den planlagte læseplan', 'lærerforventninger mht. til elevers præstationer', 'forældreengagement i elevernes undervisning', 'skole-hjem samarbejde', 'elevernes opmærksomhed på skolens ejendom' og 'elevers ønsker om at få gode resultater i skolen'.

Målingerne samles i et indeks som kaldes PPSC, når det er skolelederen, der har svaret og TPSC, hvis det er lærernes svar, der ligger til grund for beregningen.

**Jo bedre
skoleklima, jo
bedre
elevresultater**

Tabel 17 Hyppighed af elever med Høje (Høj), Medium (Med) og Lave (Lav) værdier af 'skoleklimaet' belyst gennem 8 spørgsmål til skolelederen (PPSC, Index of Principals' Perception of School Climate (PPSC)), og lærere (TPSC, Index of Teachers' Perception of School Climate). 'S' står for skoleleder, 'L' for lærer, 'DK' for værdier i Danmark og 'Int' for internationalt. 'Høje' værdier af PPSC og TPSC svarer til at skolen har et godt 'klima'.

Fag		Høj	Høj	Med	Med	Lav	Lav	Forskel
		%elever	score	%elever	score	%elever	score	Høj-Lav
Mat	DKL	19%	537	68%	526	13%	497	40 point
	DKS	26%	538	69%	519	5%	509	29 point
	IntL	17%	488	67%	473	16%	453	35 point
	IntS	22%	487	68%	471	10%	441	46 point
Nat	DKL	17%	538	69%	520	14%	495	43 point

	DKS	26%	533	69%	514	5%	485	48 point
	IntL.	17%	494	66%	477	17%	454	40 point
	IntS	22%	491	68%	474	10%	444	47 point

Den korte aflæsning af skoleklimaet foregår ved at læse første (Høj) og sidste (Lav) søjle med procentværdier. Hermed ser man, at der er relativt få elever, som oplever et 'højt' skoleklima – et hurtigt gennemsnit over tallene i første kolonne fortæller, at det drejer sig om ca. 25% af eleverne, i begge fag. Det er tilsvarende påfaldende, at så mange som ca. 10% af eleverne går på skoler med et 'lavt' eller dårligt skoleklima – som det ses ved hurtigt at sammenfatte tallene i sidste kolonne.

Det er derimod ikke overraskende, at der både internationalt og nationalt observeres, at skolelederne internationalt (IntS) og danske skoleledere (DKS) ser mere positivt på skoleklimaet end lærerne, både internationalt set (IntL) og for de danske lærere (DKL). Procenttallene for skolelederne er systematisk højere ved kategorien 'høj' og lavere ved kategorien 'lav'. Midterkategorien medium, 'med' holder sig relativt konstant for 2/3 af eleverne for både skoleledere og lærere – og inden for begge fag.

Der hersker ikke tvivl om, at den gennemsnitlige TIMSS score varierer systematisk med skoleklimaet – ved at læse vandret mellem 'høj' og 'lav' fås et billede af den forskel, der findes mellem en skole med 'højt' skoleklima og en skole med et 'lavt' klima. Forskellene ligger fra ca. 30 point til ca. 50 point, afhængig af fag og hvilke elevgrupper, der afgrænses ved grupperingen i de tre grupper 'Høj', 'Med' og 'Lav'. *I alle tilfælde er forskellene signifikante. Forskellene peger éntydigt i retning af, at jo bedre skoleklimaet er, jo højere er TIMSS scoren.*

Referencer

Allerup, P. 2002: *“Gender Differences in Mathematics Achievement”*. I: al, S. N. e., Measurement and Multivariate Analysis, Tokyo: Springer Verlag.

Allerup, P., 2002: *“TIMSS Results in Denmark – On Gender Differences in*

Science Achievement". Proc. AERA meeting, Chicago, marts 1997.

Allerup,P:2007 "*PISA 2000's læseskala, Vurdering af psykometriske egenskaber*"; Rockwool fonden, Syddansk Universitet,Odense 2007

Allerup,P, 2006: "*Det kræver kørekort at bruge P-værdier – en replik til en analyse af skolestørrelser*", Psykologisk Pædagogisk Rådgivning nr.6, 2006

Allerup, P, 2007 "*Om skolestørrelser og statistik - et svar fra professor til professor*" www.folkeskolen.dk 2007,

Allerup,P, Lindenskov,L.,Weng,P. 2006 '*growing up*' –*The story behind two items in PISA 2003*. Nordic Light, Nordisk Råd 2006.

Allerup,P:2005 "*PISA præstationer – målinger med skæve målestokke?*" Dansk Pæd. Tidsskr. maj 2005

Allerup, P. (2005). "*PISA præstationer* ", 2005 (nr. 1), s. 13.

Allerup, P.: *De statistiske F2000 analyser – er det alt?* Dansk Pædagogisk Tidsskrift / 2 / maj 2001.

Allerup P., "*Identification of group differences Using PISA scales*" in PISA according to PISA, LIT, Vienna, 2007

Beaton, A et al, 1997 "*Mathematics Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*". November 1997.

Beaton, A et al, 1997 "*Science Achievement in the Middle School Years: IEA's Third International Mathematics and Science Study*". November 1997.

IEA, 2007 "*TIMSS 2007 International Report*" "(i tryk, offentliggøres december 2008)).

IEA, 2007: TIMSS 2007 *Assessment Frameworks*, IEA(International Association for the Evaluation of Educational Achievement), Lynch School of Education, Boston College

IEA, 2007: TIMSS 2007 Encyclopedia: "*A guide to Science Education around the world*", vol 1 and 2 Edited by Mullis, I.V.S.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (2008). Chestnut Hill, MA: TIMSS &

IEA, 2007: TIMSS 2007 Encyclopedia: "*A guide to Science Education around the world*", vol 1 and 2 Edited by Mullis, I.V.S.S., Martin, M.O., Olson, J.F., Berger, D.R., Milne, D., & Stanco, G.M. (2008). Chestnut Hill, MA: TIMSS &

Rangvid, B "*Skolegennemsnit af karakterer ved folkeskolens afgangsprøver korrektion for social baggrund*" Working paper, februar 2008AKF Working paper, 2008(1), AKF

Weng,P, 1998: "*Matematik og Naturvidenskab i Ungdomsuddannelser – en international undersøgelse*"; Danmarks Pædagogiske Institut 1998.04; 1998

Weng,P, 1996: "*Matematik og Naturvidenskab i Folkeskolen – en international undersøgelse*"; Danmarks Pædagogiske Institut 1996.35; 996.35

