



Acta Didactica

2/2006

2/2006



Acta Didactica

2/2006

Acta Didactica

Are Turmo  
Svein Lie

# PISA's Computer-based Assessment of Science (CBAS):

Gjennomføring og norske resultater våren 2005

**Institutt for lærerutdanning og skoleutvikling**  
Det utdanningsvitenskapelige fakultet  
Universitetet i Oslo  
Postboks 1099 Blindern  
0317 Oslo

**Dept. of Teacher Education and School Development**  
Faculty of Education  
University of Oslo  
P.O.Box 1099 Blindern  
0317 Oslo  
Norway

[www.ils.uio.no](http://www.ils.uio.no)

ISSN: 1502-2013  
ISBN: 82-90904-83-5



9 788290 904833

unipubskriftserier

INSTITUTT FOR LÆRERUTDANNING OG SKOLEUTVIKLING  
UNIVERSITETET I OSLO

**Acta Didactica**

**2/2006**

**Are Turmo  
Svein Lie**

**PISA's Computer-based  
Assessment of Science  
(CBAS): Gjennomføring og  
norske resultater våren 2005**

© ILS og forfatterne, Oslo 2006

ISSN: 1502-2013

ISBN: 82-90904-83-5

Utgiver: Unipub AS

Trykk og innbinding: AiT e-dit AS

Rapportserien distribueres av Unipub AS

Henvendelser om denne boka kan rettes til Unipub AS:

Telefon: 22 85 33 00

Telefaks: 22 85 30 39

E-post: [post@unipub.no](mailto:post@unipub.no)

Det må ikke kopieres fra denne boka i strid med  
åndservkloven eller avtaler om kopiering inngått  
med Kopinor, interesseorgan for rettighetshavere til  
åndservk.

*Unipub AS er et heleid datterselskap av Akademika AS,  
som eies av Studentsamskipnaden i Oslo.*

## INNHold

<b>Forord</b> .....	<b>s.5</b>
<b>1. Hva er PISA?</b> .....	<b>s.7</b>
<b>2. Computer-based Assessment of Science (CBAS)</b> .....	<b>s.10</b>
2.1 Bakgrunnen for CBAS.....	s.10
2.2 Design.....	s.13
2.3 Tekniske spesifikasjoner.....	s.13
2.4 Vurdering av oppgaver.....	s.15
2.5 Oversetting av oppgaver.....	s.16
2.6 Utvalg av oppgaver.....	s.17
2.7 Gjennomføring på skolene.....	s.18
2.8 Erfaringer fra gjennomføringen.....	s.21
2.9 CBAS 2005- en ”feasibility study”.....	s.21
<b>3. Empiriske resultater</b> .....	<b>s.23</b>
3.1 Faglige prestasjoner i CBAS-undersøkelsen.....	s.23
3.2 Andeler ikke besvarte oppgaver i de to testene.....	s.27
3.3 Kjønnforskjeller i faglige prestasjoner.....	s.27
3.4 Sammenhenger mellom faglige resultater og karakterer i sentrale skolefag.....	s.29
3.5 Sammenhenger mellom sosial bakgrunn og skåre.....	s.31
3.6 IKT-ferdigheter/-bruk og prestasjoner.....	s.32
3.7 Hvor godt likte elevene CBAS-undersøkelsen?.....	s.33
3.8 Elevenes innsats og preferanser.....	s.35
3.9 Hva kjennetegner elever som skårer spesielt godt på PC-testen?.....	s.36
3.10 Introduksjon til atferdsdataene.....	s.38
3.11 Analyse av atferdsdataene.....	s.42
3.12 Tidsbruk på PC-testen.....	s.43
3.13 ”Media-play” i PC-testen.....	s.45

<b>4. Avslutning</b> .....	<b>s.51</b>
4.1 Konklusjon.....	s.51
4.2 Veien videre for CBAS.....	s.52
<b>Referanser</b> .....	<b>s.54</b>

## **Forord**

Faglige tester på PC blir stadig mer utbredt, og de kan ha mange åpenbare fordeler. Men påvirkes resultatene av at testen gjennomføres på PC istedenfor på papir? Og i tilfelle ja, hvordan? Dette er sentrale problemstillinger i denne rapporten.

Rapporten omhandler en undersøkelse av 15-åringers kompetanse i naturfag ved hjelp av en faglig test på PC. Våren 2005 deltok 315 norske 15-åringer fra 40 skoler i PISAs Computer-based Assessment of Science (CBAS). Undersøkelsen besto i at elevene gjennomførte en faglig test i naturfag på PC og en faglig test på papir. Resultatene fra de to testene kan følgelig sammenliknes på elevnivå. Begge testene varte en time. Deretter besvarte elevene et spørreskjema om egen bakgrunn, holdninger og forhold til IKT.

CBAS-undersøkelsen ble finansiert av Kunnskapsdepartementet gjennom Utdanningsdirektoratet. Vi vil rette en spesiell takk til de to testadministratorene Tristan Dobson og Jo-Rasmus Holt Zachariassen, samt til Torgeir Christiansen og Kenneth Nymoen for uvurderlig datateknisk assistanse. Takk også til Therese Nerheim Hopfenbeck, Marit Kjærnsli, Rolf Vegar Olsen og Jo-Rasmus Holt Zachariassen for kommentarer til manuset. Og ikke minst må en stor takk rettes til hele PISA-gruppa ved ILS, UiO.

Oslo, mars 2006  
Are Turmo og Svein Lie

## 1. Hva er PISA?

PISA (Programme for International Student Assessment) er et stort komparativt internasjonalt forskningsprosjekt i regi av OECD (Organisation for Economic Cooperation and Development). Prosjektet har som mål å sammenlikne 15-åringers kunnskaper og ferdigheter samt deres evne til å reflektere over egen kunnskap og erfaring i lesing, matematikk og naturfag.

Ut fra et rammeverk som er utviklet for undersøkelsen, ønsker man å belyse i hvilken grad elever som har fullført obligatorisk skolegang, har greid å tilegne seg ferdigheter og kunnskaper som regnes for å være nødvendige for å kunne delta i samfunnet på en konstruktiv måte. Prinsippet for dette rammeverket er bestemt på et politisk nivå og utviklet videre av fagpersoner fra de tre kompetanseområdene (OECD 1999, 2000, 2003a).

En omfattende orientering om PISA og de norske resultatene er gitt i de norske rapportene *Godt rustet for framtida?* (Lie, Kjærnsli, Roe & Turmo, 2001) og *Rett spor eller ville veier?* (Kjærnsli, Lie, Olsen, Roe & Turmo, 2004). OECD har publisert resultater i egne internasjonale rapporter (Artelt mfl., 2003, Kirsch mfl., 2002, OECD, 2001, 2003b, 2004a, b, c, 2005a, b, 2006, Willms, 2003). En egen rapport med sekundæranalyser av PISA-dataene i et nordisk perspektiv er også utgitt: *Northern Lights on PISA* (Lie mfl., 2003)

I PISA er policy-relevante faktorer et hovedfokus. Utdanningspolitikere vil ha svar på spørsmål som: Forbereder vårt skolesystem barna på å delta i samfunnet? Hvilken utdanningsstruktur og -praksis gir de beste mulighetene for elever som har dårlige ressurser hjemmefra? Hvor mye avhenger elevenes prestasjoner av skolens ressurser? PISA er utviklet for å bidra til en økt innsikt i blant annet disse spørsmålene. Analysene ser på

hvordan demografiske, sosiale, økonomiske og utdanningspolitiske særtrekk henger sammen med elev- og skoleresultater.

Man ønsker også å se på utviklingen over tid ved å gjennomføre en undersøkelse hvert tredje år. Alle tre kunnskapsområdene lesing, matematikk og naturfag blir dekket hver gang, men to tredeler av testtiden vil rette seg mot det som er hovedområdet. I PISA 2000 var hovedområdet lesing. I PISA 2003 var matematikk hovedområdet, og i PISA 2006 er det naturfag som skal være i fokus. I PISA 2000 deltok 32 land, 41 land deltok i PISA 2003, mens i PISA 2006 vil omlag 60 land delta.

Ved å dekke alle de tre fagområdene hvert tredje år vil det være mulig for hvert land å få en systematisk og pålitelig oversikt over endringer i elevers kompetanse innen hvert av disse sentrale kunnskapsområdene. Noen oppgaver holdes hemmelig og gjentas i nye undersøkelser slik at utvikling over tid kan studeres i absolutt forstand ved hjelp av moderne psykometriske metoder (Adams, 2002, Kjærnsli, Lie & Turmo, 2005).

I tillegg til de faglige testene inngår det i PISA et elevspørreskjema samt et skolespørreskjema som besvares av rektor (Harvey-Beavis, 2002). Elevspørreskjemaet inneholder spørsmål om blant annet elevenes selvoppfatning, læringsstrategier, motivasjon og sosiale bakgrunn. Skolespørreskjemaet har spørsmål om skolens ressurser, om lærerne og elevene ved skolen, om skolens organisering etc. (Lie mfl., 2001, Kjærnsli mfl., 2004).

PISA organiseres som et samarbeid mellom medlemslandene i OECD. Et utvalg med representanter fra hvert av deltakerlandene (PISA Governing Board, PGB) legger premissene for de politiske prioriteringene og standard for utviklingen av hva som skal undersøkes. Disse representantene kommer fra det politiske eller forvaltningsmessige nivået. Den praktiske administrasjonen og faglige ledelsen av prosjektet skjer i regi av et internasjonalt

konsortium ledet av Australian Council for Educational Research (ACER, se [www.acer.edu.au](http://www.acer.edu.au)). Konsortiet har til sammen bred erfaring når det gjelder å utforme og gjennomføre testsystemer, både i nasjonale og internasjonale sammenhenger. Mer detaljert informasjon om prosjektet finnes på den internasjonale hjemmesiden: [www.pisa.oecd.org](http://www.pisa.oecd.org).

I Norge er PISA-prosjektet finansiert av Kunnskapsdepartementet (KD) gjennom Utdanningsdirektoratet. En forskningsgruppe ved Institutt for lærerutdanning og skoleutvikling (ILS) ved Det utdanningsvitenskapelige fakultet, Universitetet i Oslo har ansvaret for gjennomføringen i Norge. Det norske prosjektets hjemmeside er: [www.pisa.no](http://www.pisa.no).

I mange tidligere internasjonale komparative undersøkelser har man konsentrert seg mer om ”skolekunnskap” ved at testen har vært basert på deltakerlandenes læreplaner. I PISA tar man ikke utgangspunkt i landenes læreplaner og skolefagenes ”pensum”, men man tar i hovedsak sikte på å måle elevenes evne til aktivt å bruke kunnskaper og erfaringer og hvordan de forholder seg til emner som trolig vil være relevante for framtiden. PISA-undersøkelsen er basert på en internasjonal konsensus på politisk nivå blant OECD-land om hva som anses å være viktig å kunne i årene framover (Wu, 2002). Til hver undersøkelse er det utviklet et rammeverk (OECD 1999, 2000, 2003a) som beskriver i detalj hva som måles innen hvert fag.

I definisjonen av de tre kunnskapsområdene er det altså lagt vekt på kunnskaper og ferdigheter som man antar blir viktig for unge mennesker for å kunne spille en konstruktiv rolle i samfunnet. Det er fokus på nyttige mer enn fagspesifikke og formelle sider ved fagene. Internasjonalt brukes begrepene *Reading Literacy*, *Mathematical Literacy* samt *Scientific Literacy*. Dette er begreper som vanskelig kan oversettes til norsk på en meningsfull måte. I denne rapporten vil vi konsentrere oss om det siste fagområdet,

naturfagkompetanse. Denne kompetansen er definert slik i PISAs rammeverk:

*Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity* (OECD, 2003a, s. 133).

Naturfagkompetanse i PISA innebærer altså å kunne bruke sentrale naturfaglige begreper for å kunne forstå og bidra til å ta avgjørelser som har med naturens verden å gjøre. Det innebærer også evne til å formulere naturvitenskapelige problemstillinger, føre beviser, trekke naturvitenskapelige konklusjoner og formidle disse videre. I dette arbeidet anvendes naturvitenskapelige begreper som har relevans for elevene, både nå og i elevenes framtid som yrkesutøvere og samfunnsborgere.

## **2. Computer-based Assessment of Science (CBAS)**

### **2.1 Bakgrunnen for CBAS**

Åpenbart er det sider ved naturfaglig kompetanse som ikke kan dekkes ved en skriftlig prøve med papir og penn, slik PISA-undersøkelsen innebærer. Det har derfor vært et ønske om å finne måter der elevene også kan engasjere seg i et dynamisk forhold til naturfaglige fenomener, enten ved ”hands on”-eksperimenter, eller ved simuleringer på datamaskin. Bruk av ”hands on”-eksperimenter ble prøvd ut i stor skala i Third International Mathematics and Science Study (TIMSS) i 1995 (Kind, 1996, Kind, Kjærnsli, Lie & Turmo, 1999). En fordel ved en IKT-løsning er at den også peker framover ved at det praktiske dataarbeidet (retting og inntasting av data) etter gjennomføringen blir sterkt forenklet. Burkhardt & Pead (2003) framhever fire mulige fordeler ved å administrere tester på PC:

1. Tradisjonelle ”papiroppgaver” kan administreres med lavere kostnader.
2. Elevsvar kan samles inn automatisk hvis de kan uttrykkes som enkel tekst, som svar på flervalgsoppgaver eller i form av informasjon uttrykt ved posisjon, for eksempel i ”drag-and-drop”-formatet.
3. Elevsvar kan kodes automatisk hvis de ikke krever menneskelig bedømmelse.
4. Nye typer oppgaver som inneholder interaktive multimedia-elementer, kan inkluderes.

I desember 2003 utlyste OECD et anbud vedrørende en PC-basert undersøkelse av elevers naturfagkompetanse innenfor rammene av PISA-prosjektet. Det ble lagt vekt på at den PC-baserte undersøkelsen skulle gi noe ekstra i forhold til den ordinære PISA-undersøkelsen i naturfag (”have added value”). I PISA-konsortiets anbud ble det lagt vekt på å skape ”added value” på to måter. For det første ønsket man å utnytte de mulighetene som IKT gir når det gjelder å presentere dynamisk stimulusmateriale, og å lage oppgaver hvor dynamisk og interaktivt stimulusmateriale er nødvendig. For det andre ville man redusere kravet til leseforståelse gjennom å presentere mer autentisk stimulusmateriale med mindre behov for forklarende tekst. Slik kan man håpe å redusere betydningen av elevenes generelle leseforståelse og derved få et ”renere” mål for kompetansen i naturfag. Til gjengjeld er det et viktig spørsmål i hvilken grad elevenes IKT-kompetanse vil påvirke resultatene. Dette er et spørsmål som blir diskutert i denne rapporten.

PISA-konsortiet innså at bruk av IKT ga en rekke nye muligheter for utforming av oppgaver, men de foreslo likevel en undersøkelse som var relativt enkel. Den skulle være mulig å gjennomføre innenfor rammene av den skisserte tidsplanen og med et relativt moderat budsjett. Det var videre et mål at studien ikke skulle stille for store krav til elevenes IKT-kompetanse. Dette medførte blant

annet at man valgte å ikke stille krav om at elevene måtte skrive tekst ved hjelp av tastaturet eller bruke hyperlenker.

Et viktig moment var i stor grad å anvende de allerede etablerte prosessene og prosedyrene i PISA-undersøkelsen også for den PC-baserte studien (heretter kalt CBAS). Konsortiet foreslo derfor at oppgavene skulle baseres på det samme rammeverket som blir brukt for å utvikle oppgaver til den ordinære PISA-undersøkelsen (OECD, 2003a). Man foreslo også å bruke de allerede etablerte ekspertgruppene, i hovedsak *Science Expert Group* (SEG) og *Technical Advisory Group* (TAG), for å gi råd om faglige og tekniske aspekter vedrørende CBAS. Konsortiet foreslo videre å bruke de samme utvalgsprosedyrene, skaleringsprosedyrene og analyseteknikkene som brukes i den ordinære PISA-undersøkelsen. Gjennom å bruke allerede etablerte prosedyrer forsøkte konsortiet å minimalisere de totale kostnadene.

PISA-konsortiet utviklet programvare for CBAS som tilfredstilte PISA-undersøkelsens spesielle behov når det gjelder oversetting og behandling av data. Programvaren som ble utviklet, er integrert med dataprogrammet *KeyQuest* slik at dette kan brukes for å trekke elever og sende inn data. Programvaren ble spesielt utviklet for å gjøre det mulig å oversette alt materiale og samtidig oppfylle kravene i PISA til kvalitetssikring (uavhengig godkjenning av alle oversettelser).

Alle oppgavene ble automatisk kodet, med andre ord var det ingen av PC-oppgavene som krevde manuell retting. Dette ble gjort for å holde undersøkelsen så enkel som mulig. De automatiske algoritmene for koding av oppgaver ble laget for å takle enkle og komplekse flervalgsoppgaver, korte numeriske svar og ”drag and drop”-svarformatet.

## **2.2 Design**

Et sentralt formål med CBAS var å sammenlikne resultater framkommet gjennom henholdsvis en papirbasert test og en test på PC. For å kunne koble CBAS til den papirbaserte testen var det derfor nødvendig at alle elevene som deltok i CBAS, også gjennomførte en test på papir. Dette medførte at de nasjonale senterne også måtte sette sammen spesifikke oppgavehefter og elevspørreskjemaer for CBAS-elevne, rette oppgaveheftene og sørge for at alle disse resultatene også ble tastet inn i dataprogrammet *KeyQuest*. Den papirbaserte testen inneholdt både flervalgsoppgaver og oppgaver hvor elevene selv måtte formulere et svar.

Alle landene som deltok i CBAS, måtte også implementere et *ICT familiarity questionnaire* for alle elevene som deltok, også for de elevene som deltok i den ordinære PISA-undersøkelsen. Elevene som deltok i CBAS, gjennomførte en en-times test på papir og en en-times test på PC. De besvarte også et elevspørreskjema på papir.

## **2.3 Tekniske spesifikasjoner**

Oppgavene i CBAS-undersøkelsen inneholder et bredt spekter av multimediaeffekter, inklusive videoer og animasjoner. Det er viktig at elevene enkelt kan navigere fra oppgave til oppgave i testen. Programvaren i CBAS ble derfor utviklet for å kunne takle store filfiler. Programvaren ble spesielt utviklet for å kunne kjøres fra en PC hvor alle filmer og animasjoner var lastet inn fra en CD eller DVD før testen startet. Dette stilte spesielle krav til maskinvaren som skulle brukes.

Det var et ønske fra noen land om å administrere testen via Internett. Dette var imidlertid ikke mulig på grunn av det store antallet multimediafiler som krevdes i testen. Administrasjon via Internett ville krevd svært stor båndbredde, og det kunne man ikke regne med at alle skoler i alle deltakerlandene hadde tilgjengelig.

Kravet om samme oppgavelayout i alle land er viktig for å sikre at dataene fra CBAS er internasjonalt sammenliknbare. PISA-konsortiet ga derfor detaljerte spesifikasjoner av de PCene som skulle brukes i CBAS. Multimediaeffekter kan oppføre seg annerledes når skjermstørrelsen blir endret. Det var derfor viktig at spesifikasjoner for skjermstørrelse og videokort ble fulgt. På samme måte var det viktig at kravet til prosessor og minne ble fulgt fordi oppgavene inneholder filer som krever et stort minne og høy prosessorhastighet. Det ble også anbefalt at spesielle PC-merker skulle brukes slik at konsortiet skulle kunne gi maksimal brukerstøtte. Minstekrav til PCene i CBAS er gitt i tabell 1.

Selv om en nettbasert løsning ikke kunne gjennomføres i 2005, så er det selvsagt at en slik løsning vil være en mer framtidsrettet versjon enn den som ble brukt. På mange måter er det derfor aktuelt å se erfaringene fra 2005 som en forberedelse av og en "feasibility"-studie for en full nettbasert gjennomføring i forbindelse med neste fase av PISA i 2009.



Tabell 1: Minimumskrav til bærbare PCer for bruk i CBAS-undersøkelsen (ACER, 2005a)

Component	Minimum required
CPU	1.6 GHz Pentium M Processor
Memory	512 MB of RAM
Hard disk	40 GB
Display	14.1" XGA (this <b>exact</b> screen size must be used)
Network	Must be equipped with wireless capability (Centrino technology) – IEEE 802.11 b/g
Optical drive	Computers must be equipped with a DVD-ROM drive for installing the software required for test administration. Also, the test administrators computer must be equipped with a CD-RW drive for writing test session result data to CDs.
Operating System	Windows XP Professional
Pointing device	Symmetrical external, optical mouse
Listening device	External stereo headphones

## 2.4 Vurdering av oppgaver

Alle oppgaver ble sendt ut til alle de deltagende landene for vurdering før det endelige utvalget av oppgaver ble gjort. Tilbakemeldingen fra landene ble tatt med i betraktningen ved det endelige utvalget. Den endelige avgjørelsen om dette ble tatt av *Science Expert Group (SEG)*. Følgende aspekter ble vurdert av det enkelte nasjonale senter for alle PC-oppgavene. Tilsvarende spørsmål blir også besvart for alle oppgavene på papir i PISA (Wu, 2002).

### 1. Relevance for 15-year-olds

- *Within curriculum?*
- *Relevant to preparedness for life?*
- *Interest and motivation level?*
- *Authentic context or application?*

### 2. Sensitivity issues

- *Cultural concerns?*
- *Comments regarding cultural, gender and/or other bias concerns?*

### 3. "Technical" characteristics

- *Translation problems?*
- *Does the computer form add value?*
- *Problems with Response Coding Guide?*

### 4. Other comments

### 5. Priority for inclusion

## 2.5 Oversetting av oppgaver

Oversetting av de PC-baserte oppgavene fulgte i hovedsak de samme prinsippene som ved oversetting av oppgaver til den ordinære PISA-undersøkelsen (Grisay, 2002, Hambleton, 2002). Tre klart spesifiserte roller måtte fylles av de nasjonale senterne; to uavhengige oversettere, samt en person til å sette de to oversettelsene sammen til én.

Alle oppgavene ble gitt i engelsk versjon, og denne versjonen ble oversatt til det enkelte lands språk av to uavhengige oversettere. En tredje person satte så disse to versjonene sammen til én versjon. Denne versjonen ble så kontrollert av en representant fra PISA-konsortiet. Denne representanten behersket selvsagt både norsk og engelsk. De enkelte nasjonale senterne gjorde deretter eventuelle

endringer i oversettelsene basert på de tilbakemeldingene som ble gitt. Til sist ble en såkalt ”endelig optisk sjekk” av oppgavene utført av en representant fra konsortiet.

En viktig nyskaping i CBAS-undersøkelsen var at alle testfiler ble lagret sentralt på PISA-konsortiets hjemmeside. Oversettelsen foregikk ved hjelp av en spesiell applikasjon som ble utviklet for CBAS. Oversetterne måtte logge seg inn på konsortiets hjemmeside for å laste ned kildeversjonene av oppgavene. Deretter ble oppgavene lastet opp igjen på siden når oversettelsen var gjort. Når begge oversetterne hadde fullført sine oversettelser, ble begge oversettelsene gjort tilgjengelige for personen som hadde som oppgave å sette versjonene sammen til én. Disse versjonene ble så igjen lastet opp på nettsiden og gjort tilgjengelige for konsortiets kontrollør. Prosessen fortsatte på tilsvarende måte helt til endelige versjon forelå. Basert på de endelige oversettelsene satte konsortiet sammen testfiler som ble sendt til det enkelte nasjonale prosjektsenter. Samtidig ble endelig versjon av CBAS-programvaren distribuert. Totalt ble 116 oppgaver oversatt til PC-testen.

## 2.6 Utvalg av elever

Skoler til den norske CBAS-undersøkelsen ble trukket blant skoler innenfor en begrenset radius rundt Oslo, noe som gjorde det mulig å reise til skolen, gjennomføre testen og returnere samme dag. Totalt deltok 40 skoler. Ved hver av disse skolene ble 10 elever født i 1989 trukket ut til å delta. Ved de fleste skolene var det noen elever som ikke deltok, enten fordi de var fraværende på testdagen, eller fordi de var ekskludert på forhånd, i tråd med PISA-undersøkelsens retningslinjer for dette (OECD, 2005a). I gjennomsnitt deltok 8 elever per skole. Tabell 2 gir en oversikt over antallet deltakende elever ved de 40 skolene. Som det går fram av tabellen, var det totalt 315 elever som deltok i CBAS-

undersøkelsen i Norge. 4 av disse elevene deltok av ulike årsaker kun på PC-testen.

Tabell 2: Antallet deltakende elever ved 40 skoler

Antall deltakende elever	Antall skoler
10	4
9	13
8	8
7	9
6	1
5	5
<i>Totalt antall elever: 315</i>	<i>40</i>

## 2.7 Gjennomføring på skolene

Undersøkelsen ble gjennomført av to testadministratorer. Begge har betydelig IKT-kompetanse og utdanning i pedagogikk på mastergradsnivå. Betydelig IKT-kompetanse var nødvendig for å kunne takle eventuelle driftsproblemer under administrering av testen.

Undersøkelsen ble gjennomført med ett sett av bærbare PCer bestående av 5 elevmaskiner og en maskin for testadministratoren. Dette medførte at ingen av skolene kunne gjennomføre testen samme dag, og det tok følgelig 40 dager å gjennomføre testen ved alle skolene. Undersøkelsen ble gjennomført i løpet av perioden fra 16. mars til 15. juni, 2005.

Ved gjennomføringen ble elevene delt i to grupper. Halvparten tok testen på PC først, mens den andre halvparten startet med testen på papir. Alle elevene satt i samme rom. Til sist svarte alle elevene på elevspørreskjemaet på papir. Hver av testene tok 1 time, mens elevene fikk 35 minutter til å svare på elevspørreskjemaet.

I PISA-undersøkelsen anvendes et rotert design, det vil si at elevene får ulike versjoner av testen som er koblet sammen

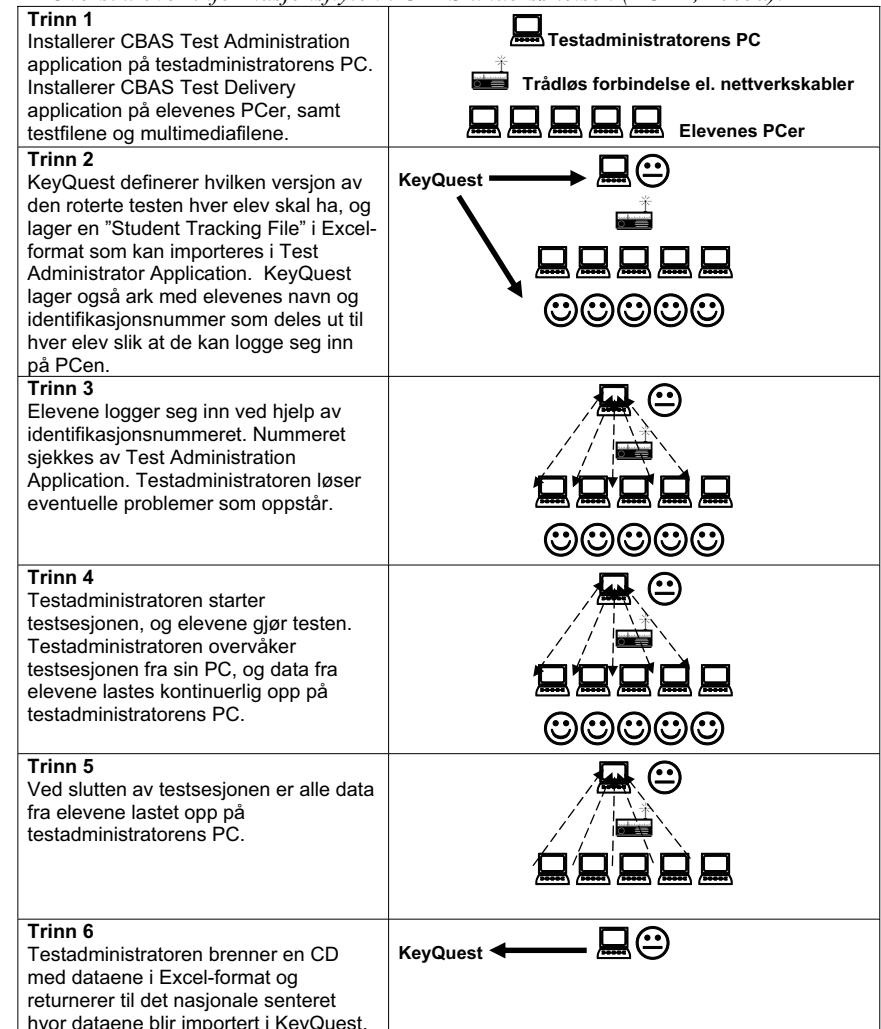
gjennom noen felles oppgaver. Basert på de felles oppgavene kan elevene plasseres på samme prestasjonsskala ved hjelp av moderne psykometriske metoder. Dette prinsippet ble også fulgt i CBAS-undersøkelsen, både i testen på papir og i testen på PC. Som tidligere nevnt, inngikk totalt 116 oppgaver i PC-testen. Hver versjon av testen inneholdt imidlertid bare 46 oppgaver. Totalt ble 10 versjoner av testen rotert blant elevene.

I og med at dette var en generalprøve, ønsket man videre å prøve ut flere spørsmål i elevspørreskjemaet enn det som kan besvares i løpet av 35 minutter. Fire forskjellige versjoner av elevspørreskjemaet ble derfor rotert blant elevene.

Under gjennomføringen av testen ble de samme instruksjonene lest høyt for elevene i alle landene som deltok. Før elevene startet med selve PC-testen gjennomgikk de en treningslesjon med eksempeloppgaver og forklaringer på hvordan de skulle svare og manøvrere seg gjennom testen. Tabell 3 viser informasjonsflyten i CBAS-undersøkelsen.

Tabell 3:

Oversikt over informasjonsflyten i CBAS-undersøkelsen (ACER, 2005a).



To typer elevdata ble registrert i PC-testen. For det første ble selvsagt elevenes svar på alle oppgavene registrert ("response data"). Dessuten ble informasjon om elevenes atferd gjennom testen lagret ("behavioral data"). Begge disse to typene av data vil bli analysert i denne rapporten.

## 2.8 Erfaringer fra gjennomføringen

Testadministratorene rapporterte at elevene gjerne var entusiastiske i forhold til den PC-baserte testen før testsesjonen startet. Noen elever ble imidlertid skuffet når de så hva testen faktisk innebar. Testadministratorenes inntrykk var for øvrig at de aller fleste elevene hadde tilstrekkelige IKT-ferdigheter til å kunne gjennomføre testen uten problemer.

Rent teknisk fungerte PC-testen stort sett som forutsatt. Ved ett tilfelle klarte imidlertid en elev å komme seg ut av testprogrammet og logge seg på skolens trådløse nettverk. Ved et annet tilfelle trakk en elev ut nettverkskabelen fra PCen. Dette førte til problemer med å overføre data fra de andre elevenes PCer. Ved noen få tilfeller sluttet musa og tastaturet å fungere under testen slik at eleven ikke fikk fullført testen. Det viste seg vanskelig å finne ut hvorfor dette skjedde.

## 2.9 CBAS 2005 – en "feasibility study"

CBAS-studien i 2005 var designet for et dobbelt formål. For det første hadde den verdi i seg selv, som en prøve for å undersøke elevenes kompetanse i naturfag på en måte som utfylte kompetansene som måles i de regulære PISA-undersøkelsene. For det andre skulle CBAS 2005 fungere som en "feasibility study" ved å finne ut hvilke tekniske og praktiske muligheter som foreligger når man skal involvere et stort antall tilfeldig uttrukne elever og skoler. Undersøkelsen ble derfor designet som en generalprøve, der både oppgavene og designet ble prøvet ut. Dette har åpenbare konsekvenser for hvilke resultater som kan rapporteres i ettertid. Blant annet er ikke dataene tilstrekkelig

kvalitetssikret til at det er meningsfullt å sammenlikne deltakerlandenes resultater. Noen av oppgavene fungerte ikke godt nok og ville ikke kunne brukes i en endelig undersøkelse. Av de 116 oppgavene var det for eksempel 17 oppgaver som hadde for lav diskriminering ( $<0,30$ ) i fem eller flere av de i alt 13 landene. I denne rapporten vil derfor fokuset være på norske resultater av spesiell interesse. De 13 deltakerlandene i CBAS-undersøkelsen våren 2005 er gitt i tabell 4.

Tabell 4: De 13 deltakerlandene i PISA CBAS våren 2005

Australia	Norge
Bulgaria	Portugal
Danmark	Russland
Irland	Slovakia
Island	Taiwan
Japan	Østerrike
Korea	

I hovedundersøkelsene i den ordinære PISA-studien våren 2006 trekkes mer enn 5000 elever ut til å delta. Dette gjør det mulig å etablere robuste funn også for flere undergrupper av populasjonen. I CBAS-undersøkelsen legger derimot det totale antallet på 315 elever begrensninger på hvilke analyser som det er hensiktsmessig å gjøre. Det relativt begrensende utvalget kan ikke splittes opp i mange underkategorier uten at feilmarginene for estimatene blir svært store, med tilhørende økt sannsynlighet for å rapportere utvalgsspesifikke funn.

Som tidligere beskrevet, var det også fire forskjellige versjoner av elevspørreskjemaet som ble rotert blant elevene. Dette medfører at ikke alle 315 elevene har svart på de samme spørsmålene i spørreskjemaene. Korrelasjoner mellom data fra spørreskjemaene og prestasjoner er derfor i mange tilfeller basert på en delmengde av utvalget. I analysene som presenteres i denne rapporten, er det tatt hensyn til disse trekkene ved de tilgjengelige dataene.

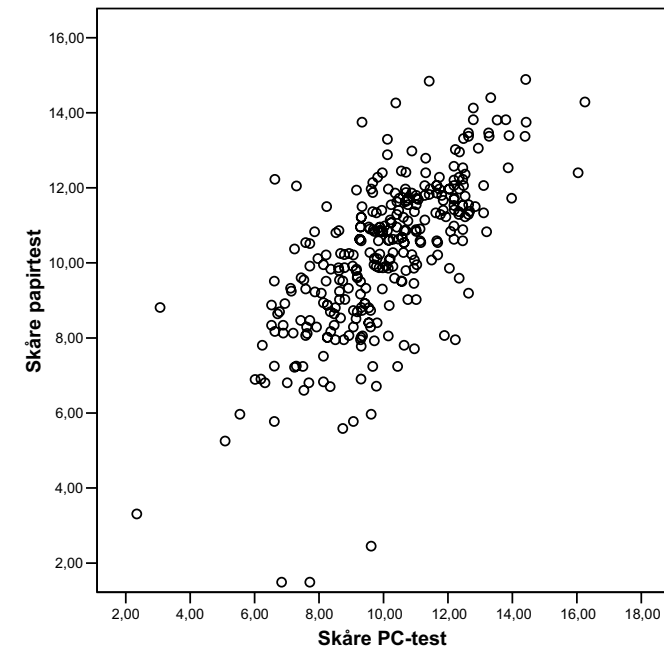
### 3. Empiriske resultater

#### 3.1 Faglige prestasjoner i CBAS-undersøkelsen

Som tidligere nevnt, deltok totalt 315 elever i CBAS-undersøkelsen. For disse elevene har vi med andre ord i utgangspunktet to skåreverdier for naturfagkompetanse. Fire av disse 315 elevene deltok imidlertid av ulike årsaker kun på PC-delen av undersøkelsen, og de er derfor ikke inkludert i analysene som nå følger. I tillegg var det 5 elever som skåret mer enn 3 standardavvik under gjennomsnittet på en eller begge testene. Som det framgår av figur 1, må disse elevene med skåre <4 regnes som "outliers" (gjennomsnittet for testene er 10, standardavviket er 2). Det er overveiende sannsynlig at disse elevene har boikottet deler av testen, og de er derfor ikke inkludert i analysene i denne rapporten. Analysene er dermed basert på 306 elever.

Når vi ser bort fra de fem elevene med ekstremt lave skåreverdier, har korrelasjonen mellom de to skåreverdiene i figur 1 verdien 0,70. Korrelasjonen er omtrent like sterk for jenter som for gutter.

Figur 1: Sammenhengen mellom skåre for PC-testen og papirtesten (N=311).

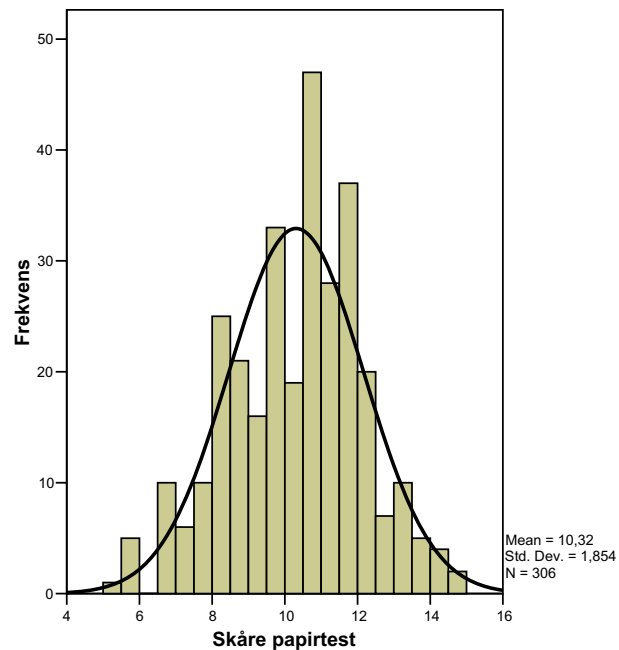


Figur 2 viser hvordan skåreverdiene for papirtesten fordeler seg for de 306 elevene. Skåreverdiene er her standardisert slik at 10 er gjennomsnittet og 2 standardavviket for alle de 1211 elevene som deltok i generalprøven i PISA våren 2005 (CBAS + ordinær PISA). De fleste av disse elevene deltok i generalprøven for den ordinære PISA-undersøkelsen. Disse elevene har fått en skåreverdi basert på en totimers papirtest. CBAS-elevne er plassert på samme prestasjonsskala, basert på papirtesten som varte i en time.

Fra resultatene på figur 2 kan vi finne ut at de 306 elevene som tok testen på papir som del av CBAS, skårer litt høyere enn

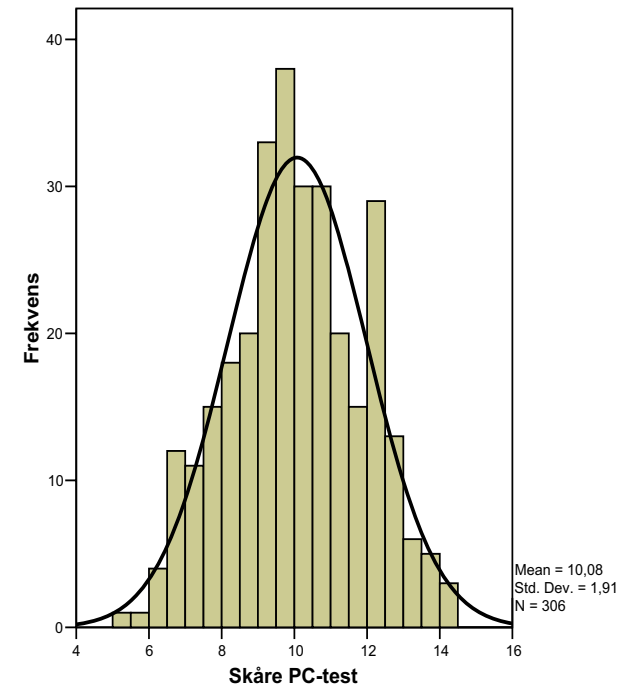
gjennomsnittet for alle PISA-elevene. Gjennomsnittet for elevene i den ordinære PISA-undersøkelsen er 9,93, mens gjennomsnittet for CBAS-elevene er 10,32. Forskjellen utgjør 20 prosent av standardavviket, men denne forskjellen er imidlertid ikke statistisk signifikant.

Figur 2: Fordeling av skåre for testen på papir for CBAS-elevene (N=306). En normalfordelingskurve er lagt til for sammenlikningens skyld.



Figur 3 viser skårefordelingen for PC-testen. At gjennomsnittet er marginalt høyere enn 10, skyldes at vi har utelatt de elevene som skårer lavere enn tre standardavvik under gjennomsnittet (se ovenfor).

Figur 3: Fordeling av skåre for testen på PC for CBAS-elevene (N=306). En normalfordelingskurve er lagt til for sammenlikningens skyld.



Som tidligere nevnt, gjennomførte omlag halvparten av elevene PC-testen først og deretter testen på papir. Den andre halvparten fikk papirtesten først. Tabell 5 viser resultatene for disse to gruppene av elever. Resultatene viser at elevene som fikk PC-testen først, skårer høyere både på PC-testen og på papirtesten. Forskjellen er imidlertid størst for PC-testen. Verdiene i tabell 5 er basert på kun i overkant av 150 elever. Dette gjør at standardfeilen blir 0,16, og differansene måtte ha vært minst 0,46 for å være

statistisk signifikante ( $p=0,05$ ). Ingen av differansene i tabell 5 er med andre ord statistisk signifikante.

Tabell 5: Resultater for elever som gjennomførte PC-testen henholdsvis først og sist ( $N=151$  og  $N=155$ )

	Resultat PC-test	Resultat papirtest
PC-testen først	10,25	10,39
PC-testen sist	9,91	10,24

### 3.2 Andeler ikke besvarte oppgaver i de to testene

Totalt sett er det en større andel ubesvarte oppgaver i testen på papir enn i testen på PC, noe som for øvrig er tilfelle i alle landene som deltok. Papirtesten har 7 prosent ikke-svart i Norge, mens i PC-testen er tilsvarende tall kun litt over 2 prosent. Hvis vi derimot utelukkende betrakter flervalgsoppgavene i papirtesten, er det kun en ubetydelig forskjell i prosent ikke-svart. Med andre ord er det de åpne oppgavene i papirtesten som trekker prosentverdien opp i papirtesten.

### 3.3 Kjønnforskjeller i faglige prestasjoner

Tabell 6 sammenlikner kjønnforskjellene i prestasjoner på de to delene av CBAS-undersøkelsen. Resultatene viser en markert forskjell mellom de to delene av undersøkelsen. For begge delene er det en kjønnforskjell i favør av guttene, men denne forskjellen er mye større for PC-testen.

Også for resultatene i tabell 6 blir feilmarginene betydelige som en konsekvens av det relativt lave antallet elever. Likevel kan vi slå fast at jentene skårer statistisk signifikant høyere på papirtesten enn på PC-testen. Og guttene skårer signifikant bedre enn jentene på PC-testen. Forskjellen mellom kjønnene på papirtesten er derimot ikke signifikant.

Tabell 6: Kjønnforskjeller i prestasjoner i CBAS-undersøkelsen

	Resultat PC-test	Resultat papirtest
Jenter (N=157)	9,71	10,26
Gutter (N=149)	10,47	10,38
Differanse i favør av gutter	0,76	0,12
Effektstørrelse (differanse målt i % av standardavviket)	38	6

Selv om PC-testen og papirtesten begge søker å måle *scientific literacy*, er det ingen felles oppgaver i testene. Tidligere forskning har vist at kjønnforskjeller i naturfagprestasjoner varierer avhengig av delområde i faget. I TIMSS-undersøkelsen i 2003 fant vi for eksempel at jenter i 8. klasse skårer noe bedre enn guttene i biologi, mens guttene skårer best i kjemi, fysikk og geofag (Grønmo, Bergem, Kjærnsli, Lie & Turmo, 2004). Eventuelle forskjeller i faglig tyngdepunkt i de to testene vil derfor kunne tenkes å påvirke størrelsen på kjønnforskjellen.

For å undersøke dette nærmere, har vi klassifisert alle oppgavene i de to testene etter hvilket fagområde de hører hjemme innenfor. Det må i denne forbindelse påpekes at man i PISA-undersøkelsen klassifiserer oppgaver etter om de krever kunnskaper *om* naturfag eller *i* naturfag. Oppgavene *om* naturfag er ikke klassifisert etter fagområde. I vår analyse har vi derimot klassifisert alle oppgavene etter det fagområdet de dreier seg om.

Tabell 7 viser prosentvis fordeling av oppgavene i PC-testen og papirtesten klassifisert ut fra fagområdene fysikk, biologi, kjemi og geofag. Resultatene viser at andelen oppgaver innen kjemi og geofag er omtrent tilsvarende i de to testene. Forholdet mellom fysikkoppgaver og biologioppgaver er imidlertid betydelig forskjellig. Omlag halvparten av oppgavene i PC-testen er fysikkoppgaver, mens omlag halvparten av oppgavene i papirtesten er biologioppgaver. Med utgangspunkt i kunnskapen om at fysikk

gjærne har vist seg å vre et ”guttefag”, mens biologi gjærne er et ”jentefag”, kan forskjellene mellom kjnnene p de to testene tenkes å henge sammen med forskjeller i faglig innhold.

Tabell 7: Prosentvis fordeling av oppgaver p ulike fagomrder for PC-testen og papirtesten.

	PC-test	Papirtest
Fysikk	47 %	26 %
Biologi	30 %	48 %
Kjemi	9 %	9 %
Geofag	15 %	17 %

For å teste hypotesen om at kjnnforskjellen kan tilskrives forskjeller i faglig vektlegging, ble et utvalg av PC-oppgavene valgt ut innen ulike fagomrder ut fra samme fordeling som i papirtesten. I motsetning til p papirtesten viste det seg at guttene skrer bedre enn jentene innen alle fagomrder (ACER, 2005b).

### 3.4 Sammenhenger mellom faglige resultater og karakterer i sentrale skolefag

I den norske versjonen av elevsprreskjemaene ble det lagt inn sprsml om hvilken karakter elevene fikk i sentrale skolefag ved siste karakteroppgjr. De utvalgte fagene var norsk, matematikk og natur- og miljfag.

Tabell 8 viser sammenhengen mellom skre p de to delene av CBAS-underskelsen og karakterer i disse tre fagene. Et pfallende trekk ved resultatene er at korrelasjonene er relativt lave, ogs for natur- og miljfag. Faktisk er korrelasjonen med karakter i matematikk for PC-testen sterkere enn tilsvarende korrelasjon for natur- og miljfag. Det er ogs interessant å ppeke at korrelasjonen mellom karakteren i norsk og resultater for papirtesten er sterkere enn tilsvarende korrelasjon for PC-testen. Dette resultatet kan forsts i lys av at papirtesten inneholder betydelig mer tekst som elevene m forholde seg til enn det PC-

testen gjr. Nettopp dette forholdet har vi tidligere beskrevet som ett av argumentene for en slik PC-test. En PC-test kan stille mindre krav til elevenes lesekompetanse og dermed gi et ”renere” ml for naturfagkompetanse.

Tabell 8: Korrelasjoner mellom karakterer i sentrale skolefag og skre i CBAS-underskelsen. I parentes er vist korrelasjoner korrigert for tilfeldige mlefeil.

	Skre p PC-test	Skre p papirtest
<b>Norsk</b>	0,28 (0,32)	0,40 (0,49)
<b>Matematikk</b>	0,46 (0,52)	0,46 (0,56)
<b>Natur- og miljfag</b>	0,36 (0,41)	0,42 (0,51)

Resultatene tyder videre p at CBAS-underskelsen i stor grad mler noe annet enn det karakteren i natur- og miljfag er et uttrykk for. Det m imidlertid understrekes at elevene selv har oppgitt karakterene, og feilrapportering kan derfor forekomme. Dette kan bidra til å svekke korrelasjonene. Og siden reliabiliteten for testene ikke er hyere enn 0,68 og 0,79 for henholdsvis papirtesten og PC-testen, vitner det om at vi har en betydelig mlefeil for de mlte skreverdiene. Disse mlefeilene bidrar til å ”utvanne” korrelasjonene, siden de setter en øvre grense for hvor store korrelasjonene kan bli.

Det gr an å ”korrigere” for dette forholdet (ved å dividere med kvadratroten av reliabiliteten), og de korrigerede korrelasjonene er ogs vist i tabell 8. Hvis vi studerer tallene i parentes, blir det enda tydeligere at PC-resultatene korrelerer betydelig lavere med norsk karakteren enn det papirtesten gjr. Men vi merker oss ogs med noe undring at korrelasjonene for begge versjonene er hyest for matematikk-karakterens vedkommende.



### 3.5 Sammenhengen mellom sosial bakgrunn og skåre

PISA-undersøkelsen inneholder flere spørsmål som kartlegger elevenes hjemmebakgrunn, noe som også har vært tilfelle i mange andre internasjonale komparative studier innen utdanning (Buchmann, 2002). Særlig i PISA 2000 var kartlegging av elevenes hjemmebakgrunn et prioritert område i elevspørreskjemaet (Turmo, 2004, 2005). Utgangspunktet er begrepene økonomisk, kulturell og sosial kapital (Bourdieu, 1984, Coleman, 1990).

Tidligere analyser har vist at spørsmålet om antall bøker i hjemmet er et meget velegnet mål på hjemmebakgrunn (Lie, Kjærnsli & Brekke, 1997; Turmo & Lie, 2004). Analyser av PISA-dataene fra 2000 og 2003 viste en korrelasjon mellom skåre i naturfag og antallet bøker på 0,32. Den var litt sterkere for jenter enn for gutter (0,36 mot 0,29). Også dataene fra generalprøven i den ordinære PISA-undersøkelsen våren 2005 viser en tilsvarende sterk sammenheng (0,33).

Analysen viser at sammenhengen med sosial bakgrunn er noe sterkere i CBAS-undersøkelsen enn det som er vanlig i PISA for øvrig. Det må imidlertid understrekes at kun 118 elever har svart på spørsmålet om antallet bøker i hjemmet som en følge av designet med roterte spørreskjemaer. Dette medfører relativt store feilmarginer, og funnet bør derfor fortolkes med varsomhet.

Analysene viser at sammenhengen med sosial bakgrunn er noe sterkere for papirtesten enn for PC-testen. Denne effekten viser seg å være stabil også i ulike tilfeldig sammensatte delmengder av utvalget.

Med forbehold om at funnene er utvalgsspesifikke, vil vi nevne at sammenhengen mellom antallet bøker i hjemmet og prestasjoner er i omtrent samme størrelsesorden for guttene som det vi typisk tidligere har funnet i PISA-undersøkelsene. Årsaken til at

korrelasjonene er sterkere i CBAS-undersøkelsen, er å finne blant jentene. Tilsvarende er det de elevene som tok PC-testen først, som i hovedsak gjør at korrelasjonen med antallet bøker i hjemmet blir sterkere i CBAS-undersøkelsen. For elevene som tok PC-testen først, er sammenhengen med sosial bakgrunn sterkere for begge testene. Sammenhengen mellom prestasjoner og sosial bakgrunn er derimot i samme størrelsesorden som det som er vanlig i PISA for de elevene som tok papirtesten først.

### 3.6 IKT-ferdigheter / -bruk og prestasjoner

Elevspørreskjemaene inneholdt flere spørsmål om elevenes erfaringer med bruk av IKT samt deres egen vurdering av brukerferdigheter. Disse spørsmålene inngikk i alle fire versjonene av elevspørreskjemaet. Elevene ble spurt om hvorvidt de noen gang har brukt en datamaskin, og om hvor ofte de bruker datamaskiner hjemme og på skolen. Tabell 9 gir en oversikt over hva elevene ble spurt om å vurdere sine ferdigheter i forhold til.

Tabell 9: Oversikt over spørsmål om egen vurdering av IKT-ferdigheter

Lage hjemmeside	Sende vedlegg til e-post
Skrive og sende e-post	Laste ned filer fra Internett
Lage multimedia-presentasjon	Søke på Internett
Laste ned musikk fra Internett	Flytte filer
Lage en presentasjon i Power Point e.l.	Kopiere data fra CD
Bruke regneark	Lage en database
Bruke tekstbehandler	Redigere digitale foto
Fjerne datavirus	Chatte

Elevene fikk her fire svaralternativer:

1. Dette kan jeg gjøre uten hjelp fra noen andre.
2. Jeg kan gjøre dette hvis jeg får hjelp av noen andre.
3. Jeg vet hva dette er, men vet ikke hvordan det kan gjøres.
4. Jeg vet ikke hva dette er.

Alle elevene i generalprøven i PISA besvarte spørsmålene om erfaringer og ferdigheter. Spørsmålene vil også inngå i hovedundersøkelsen i PISA våren 2006. En publisering av deskriptive data fra generalprøven vil derfor innebære en foregripping av PISA 2006-resultatene som skal publiseres i desember 2007. Vi vil derfor her nøye oss med å kun studere sammenhengen med prestasjoner i CBAS-undersøkelsen. Det kan nevnes at spørsmålene om ferdigheter og IKT-bruk fra PISA 2003 er analysert i detalj i en egen rapport utgitt av OECD (OECD, 2006, Schleicher, 2006).

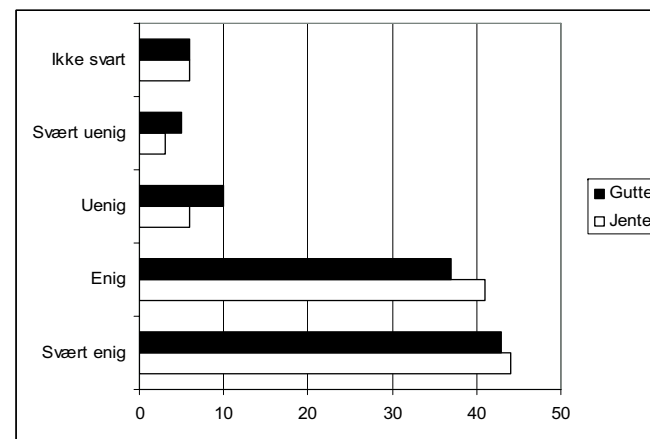
Vi finner kun svake sammenhenger mellom elevenes egenvurderinger av IKT-ferdigheter og prestasjoner i CBAS-undersøkelsen. Sammenhengene er imidlertid gjennomgående noe mer positive for PC-testen enn for papirtesten. Vi har laget en samlevariabel for alle spørsmålene om elevens egenvurdering av ferdigheter knyttet til IKT. Korrelasjonen mellom denne samlevariabelen og differansen mellom skåre på PC-testen og papirtesten er signifikant positiv (0,13). Dette indikerer at det er en svak, men positiv effekt av IKT-ferdigheter på resultatene fra PC-testen.

Analysene viser ingen sammenheng mellom hyppigheten av IKT-bruk og prestasjoner.

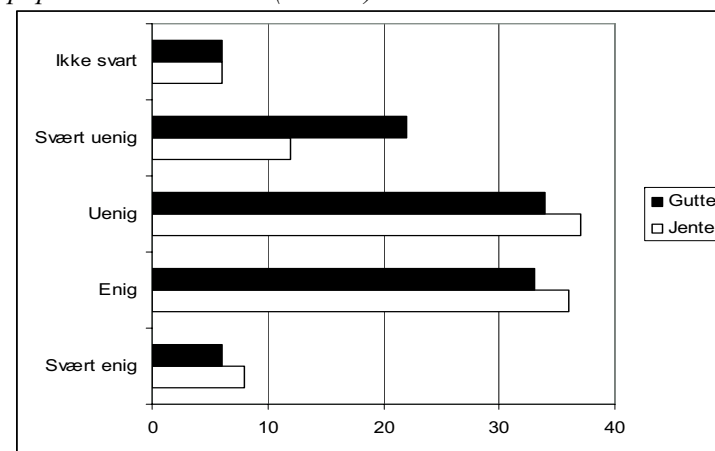
### 3.7 Hvor godt likte elevene CBAS-undersøkelsen?

Figur 4 viser resultater for spørsmålet i elevspørreskjemaene om hvorvidt elevene likte å ha prøven på datamaskin. De fleste elevene er enige i dette utsagnet, og dette gjelder både jenter og gutter. Resultatene på figur 5 viser imidlertid at betydelig færre elever likte å ha prøven på papir. Som det framgår ved å sammenlikne de to figurene 4 og 5, er det tydelig at guttene har sterkere preferanse for PC-prøven enn det jentene har.

Figur 4: Svar på spørsmålet om elevene likte å ha prøven på datamaskin: Prosentandeler (N=315)



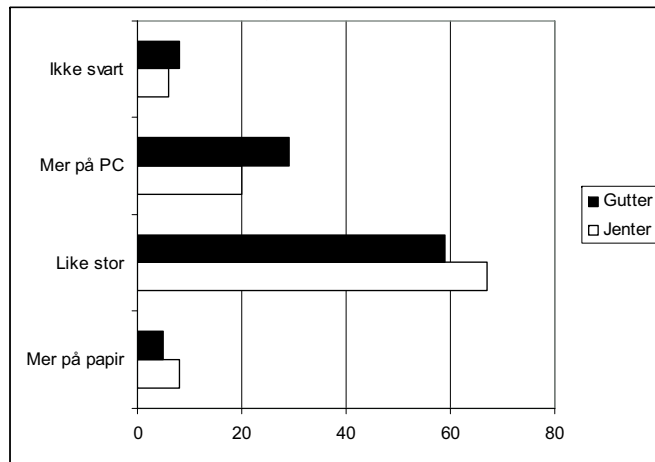
Figur 5: Svar på spørsmålet om elevene likte å ha prøven på papir: Prosentandeler (N=315)



### 3.8 Elevenes innsats og preferanser

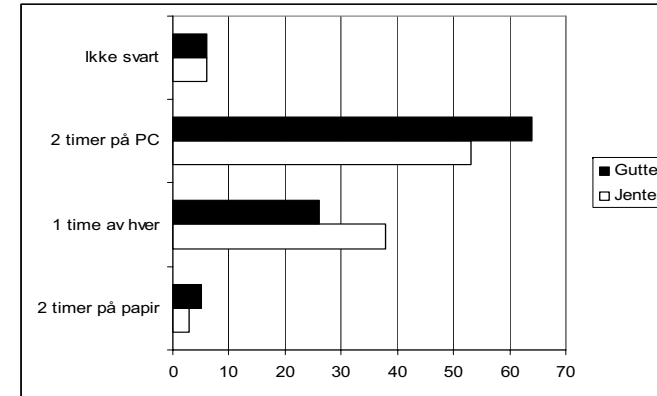
Figur 6 viser elevenes vurdering av egen innsats på de to delene av CBAS-undersøkelsen. Resultatene viser at de aller fleste av elevene oppgir at de la ned samme innsats i de to delene av undersøkelsen. Totalt sett er det imidlertid en tendens til at elevene hevder å ha lagt noe mer innsats i PC-testen enn i papirtesten. Spesielt gjelder dette guttene.

Figur 6: Elevenes vurdering av egen innsats på de to delene av CBAS-undersøkelsen (N=315)



Figur 7 viser elevenes preferanser hvis de skulle gjennomføre en ny totimers prøve. De fleste ville foretrekke en totimers prøve på PC, og dette er særlig populært blant guttene. Vi kan altså slå tydelig fast at PC-prøven har en form som slår an blant norske elever. Dette er en viktig lærdom for utvikling av databaserte prøver her i landet, et felt som er veldig aktuelt og står foran store utfordringer, ikke minst i forbindelse med de nasjonale prøvene (Lie, Hopfenbeck, Ibsen & Turmo, 2005).

Figur 7: Elevenes preferanser ved en eventuell ny totimers prøve (N=315)



### 3.9 Hva kjennetegner elever som skårer spesielt godt på PC-testen?

Med dette spørsmålet mener vi å fokusere på hvilke elever som ser ut til å ha fordel av PC-testen sammenliknet med papirversjonen. Et viktig tema har vi allerede berørt tidligere, nemlig hvorvidt PC-testen i (for) stor grad måler generell IKT-kompetanse i tillegg til det den er ment å måle. Siden databaserte prøveformer blir mer og mer utbredt, er det viktig å undersøke hvor viktig dette forholdet er sammenliknet med andre faktorer. Vi vil derfor undersøke ulike faktorer og se hvordan de kan "forklare" testresultatene.

Tabell 10 viser korrelasjonene mellom hver av de ulike faktorene og to resultatmål: skåre på PC-testen samt differansen mellom de to skåreverdiene (hvor mye bedre eller dårligere eleven skåret på PC enn på papir). Det er i denne sammenheng viktig å minne om at de to skåreverdiene er målt med samme mål (standardisert med gjennomsnitt 10 og standardavvik 2 poeng). I tabellen er det bare

tatt med faktorer som korrelerer signifikant ( $p < 0,05$ ). Ikke-signifikante verdier er vist med tegnet -.

Tabell 10: Korrelasjoner med skåre på PC-prøven samt med differansen mellom de to prøvene i PC-prøvens favør (N=306, men bare rundt 100 for de to siste pga rotasjon av hefter)

	Korrelasjon med skåre på PC-testen	Korrelasjon med differansen mellom de to skåreverdiene
Kjønn (jente=1, gutt=2)	0,20	0,22
Karakter i norsk	0,28	-0,14
Karakter i matematikk	0,46	-
Karakter i naturfag	0,36	-
PC-kompetanse	-	0,13
Preferanse for PC-testen	-	0,13
Antall bøker hjemme	0,41	-
Gjenstander i hjemmet	0,28	-

La oss kommentere noen tydelige trekk i tabell 10. At guttene skårer betydelig bedre enn jentene, har vi allerede diskutert. Det har vi også når det gjelder korrelasjoner mellom PC-skåre og karakterene, men nå kan vi se dette i en videre sammenheng. Vi ser i høyre kolonne at den spesifikke, PC-relaterte forbedringen ikke korrelerer signifikant med realfagkarakterer og korrelerer negativt med norsk karakter. Dette er nettopp hva vi ville forvente for en vellykket IKT-basert måling av naturfagkompetanse.

Tabell 10 sier oss også noe mer. Vi ser at forbedringen ikke uventet korrelerer positivt både med samleskåren for egenvurdert PC-kompetanse og med den tidligere omtalte preferanse for PC-versjonen. Her får vi et enkelt mål på hvor mye IKT-kompetanse og -motivasjon ser ut til å påvirke resultatet. Korrelasjonene i høyre kolonne, kan sammenfattes omtrent slik: Guttene har i forhold til jentene en betydelig fordel av at prøven gjennomføres på PC. Det går ut på ett om vi sier at dette delvis henger sammen med at papirversjonen av prøven favoriserer gode lesere og

skribenter, eller om vi sier at PC-versjonen favoriserer IKT-kyndighet og -motivasjon. Det ser videre ikke ut til at elevenes hjemmebakgrunn har noen spesifikk effekt overfor de to versjonene. Det er også viktig å merke seg at hyppigheten av IKT-bruk ikke forekommer i tabell 10 og altså i seg selv ikke henger sammen med prestasjonene.

En enkel regresjonsmodell kan kaste ytterligere lys over forholdet. Vi bruker den omtalte differansen av testskårene i favør av PC-versjonen som avhengig variabel og de fire faktorene fra tabell 10 som korrelerte signifikant som uavhengige variabler. Resultatene viser at de fire faktorene til sammen kan forklare (i statistisk forstand) 8 prosent av variasjonen i den nevnte differansen mellom skåreverdiene.

### 3.10 Introduksjon til atferdsdataene

Som tidligere nevnt, ble også elevenes interaksjoner med PC-testen kartlagt. I løpet av den timen testen varte, ble alle interaksjoner elevene hadde med testprogrammet loggført. Disse dataene ble eksportert og lagret som Excel-filer etter at testen var ferdig. Generelt er atferdsdataene i CBAS av svært høy kvalitet (99,9 prosent feilfrie). De få feilene i atferdsdataene skyldes tekniske forhold som nettverksproblemer.

Tabell 11 viser atferdsdata fra de fem første minuttene av testen for en tilfeldig valgt elev. Kolonnen lengst til venstre viser tiden for de ulike hendelsene. Deretter følger typen hendelse og hvilken oppgave eleven har arbeidet med. "SC848Q01" betyr spørsmål nummer 1 i oppgaveenheten SC848. Seks ulike typer hendelser er kartlagt:

1. "Item-enter": Eleven klikker seg inn på en oppgave.

2. "Item-attempt": Eleven avgir et svar. Tiden registreres idet eleven går videre til neste oppgave.
3. "Media-play": Eleven avspiller et multimedia-element.
4. "Media-stop": Eleven stopper et multimedia-element.
5. "Media-pause": Eleven setter et multimedia-element på pause.
6. "Custom-event": Spesifikk hendelse knyttet til noen av oppgavene med flash-filer. Noen flash-filer er programmert til å generere en "custom-event" hver gang en elev gjør en handling. Alle de andre hendelsene er standardisert over alle oppgavene.

Lengst til høyre er en nærmere spesifisering av hendelsen gitt, for eksempel hvilket svar eleven faktisk har gitt, eller hvilken konkret film hun har spilt av.

Fra tabell 11 kan vi se at eleven gikk rett inn på den første oppgaven idet testen startet. Etter 21 sekunder spilte eleven av en film, for deretter å avgi et svar 19 sekunder senere. Eleven brukte altså kun 40 sekunder på denne oppgaven, mens neste oppgave tok noe mer tid; 1 minutt og 4 sekunder. På denne måten kan elevens atferd gjennom hele testen studeres i stor detalj.

Tabell 11: Atferdsdata fra de fem første minuttene av testen for en tilfeldig valgt elev

Tid	Type hendelse	Oppgave	Spesifikasjon
00:00:00	item-enter	SC848Q01	null
00:00:21	media-play	SC848Q01	SC848-movie-01.avi
00:00:40	item-attempt	SC848Q01	3
00:00:40	item-enter	SC850Q01	null
00:01:14	media-play	SC850Q01	SC850-flash-01.swf
00:01:44	item-attempt	SC850Q01	2
00:01:44	item-enter	SC850Q02	null
00:02:01	media-play	SC850Q02	SC850-flash-02.swf
00:02:05	media-play	SC850Q02	SC850-flash-02.swf
00:02:07	media-play	SC850Q02	SC850-flash-03.swf
00:02:10	media-play	SC850Q02	SC850-flash-04.swf
00:02:13	media-play	SC850Q02	SC850-flash-05.swf
00:02:16	media-play	SC850Q02	SC850-flash-03.swf
00:02:20	media-play	SC850Q02	SC850-flash-02.swf
00:02:24	media-play	SC850Q02	SC850-flash-03.swf
00:02:33	media-play	SC850Q02	SC850-flash-03.swf
00:02:36	item-attempt	SC850Q02	2
00:02:36	item-enter	SC817Q01	null
00:02:58	media-pause	SC817Q01	SC817-movie-01.avi
00:03:03	media-play	SC817Q01	SC817-movie-01.avi
00:03:38	item-attempt	SC817Q01	1
00:03:38	item-enter	SC817Q02	null
00:04:28	custom-event	SC817Q02	1:dip1:8421
00:04:39	custom-event	SC817Q02	2:dip2:10055

I det norske utvalget er det totalt registrert i overkant av 91 000 hendelser. Tabell 12 viser antallet loggførte hendelser for de 10 ulike versjonene av PC-testen. Gjennomsnittlig antall hendelser per elev er også gitt for hver av testversjonene. Hendelser under den innledende treningsseksjonen som alle elevene gjennomførte, er ikke inkludert her. Disse hendelsene inkluderes heller ikke i de videre analysene som presenteres. Resultatene i tabell 12 viser at gjennomsnittlig antall hendelser varierer fra 194 i testversjon 3 til 307 i testversjon 5.

Tabell 12: Antall hendelser i de 10 ulike versjonene av testen

Versjon	Antall	Antall elever	Gjennomsnittlig antall hendelser per elev
1	9862	36	274
2	8265	31	267
3	6801	35	194
4	7674	34	226
5	8911	29	307
6	7987	32	250
7	8963	31	289
8	8767	29	302
9	8450	30	282
10	8683	33	263

Tabell 13 viser prosentvis fordeling av antall hendelser i de seks loggførte kategoriene. Resultatene viser at tre av hendelsestypene er dominerende; "Item-enter", "Item-attempt" og "Media-play". Tabell 14 viser tilsvarende prosentfordeling for hver av de 10 testversjonene. Ingen av testversjonene avviker vesentlig fra totalbildet.

Tabell 13: Prosentvis fordeling av ulike hendelser i PC-testen

Hendelse	Prosentandel
Custom-event	6 %
Item-attempt	32 %
Item-enter	32 %
Media-pause	1 %
Media-play	28 %
Media-stop	1 %

Tabell 14: Prosentvis fordeling av ulike hendelser i de 10 versjonene av PC-testen

Hendelse	Versjon av testen (prosentandeler)									
	1	2	3	4	5	6	7	8	9	10
Custom-event	10	0	2	5	11	13	1	3	12	3
Item-attempt	30	31	34	33	30	27	36	34	31	33
Item-enter	30	31	34	33	30	27	36	34	31	33
Media-pause	1	2	1	1	2	1	1	1	1	0
Media-play	28	34	28	27	26	30	24	26	23	31
Media-stop	1	2	2	1	1	1	1	2	2	1

### 3.11 Analyse av atferdsdataene

Atferdsdataene kan analyseres på ulike måter. De kan brukes til å analysere og forklare forskjeller i prestasjoner til elever med ulik bakgrunn. Hvordan oppgavene har fungert, kan også studeres i større detalj i lys av bedre informasjon om elevenes måte å løse oppgavene på. I tillegg kan man måle eksakt tiden det tar elevene å løse de enkelte oppgavene, noe som er viktig informasjon for testutviklerne når de skal vurdere testens lengde og sammensetning.

Som tidligere nevnt, inngikk totalt 116 oppgaver i PC-testen i CBAS-undersøkelsen. På grunn av det roterte designet har kun i overkant av 100 norske elever besvart hver av oppgavene. Dette legger naturlige begrensninger på hvilke analyser det er hensiktsmessig å gjøre på oppgavenivå. I tillegg kommer at oppgavene ikke er offentlige fordi mange av dem skal brukes i hovedundersøkelsen i CBAS våren 2006. Vi kan derfor ikke publisere en detaljert beskrivelse av oppgavene i denne rapporten. På bakgrunn av dette vil vi ikke presentere detaljerte analyser på oppgavenivå basert på de kognitive dataene og atferdsdataene. Likevel vil vi understreke det store potensialet som kan ligge i denne typen analyser sett i et fagdidaktisk perspektiv. Vi har tidligere argumentert for betydningen av slike analyser basert på tradisjonelle oppgaver i store internasjonale undersøkelser (Olsen,

Turmo & Lie, 2001), og atferdsdataene kan gi interessant tilleggsmateriale om elevenes oppgaveløsningsstrategier.

I det følgende vil vi derfor presentere analyser på makronivå sett i forhold til totalskåre på PC-testen. Av de seks hendelsestypene er det ”media-play” som er den mest interessante. Det er særlig denne aktiviteten som skiller PC-testen fra papirtesten. Innhold som i papirtesten må presenteres i form av tekst, kan i PC-testen illustreres ved hjelp av filmer. I analysene vil fokuset være på denne hendelsestypen. Vi vil også spesielt studere tidsbruken på PC-testen.

### 3.12 Tidsbruk på PC-testen

Tabell 15 viser gjennomsnittlig tidsbruk på PC-testen for jenter og gutter. Fordelingen for alle elevene er gitt på figur 8. Som tidligere nevnt, ble tiden for avgitt svar (”item attempt”) registrert idet eleven gikk videre til neste oppgave. Tiden for siste avgitte svar ble registrert idet testprogrammet ble avsluttet etter en time. Vi bruker derfor her tiden for nest siste avgitte svar som mål for tidsbruk, med andre ord det tidspunktet hvor eleven klikket seg inn på siste oppgave som er besvart.

Tabell 15: Gjennomsnittlig tidsbruk på PC-testen for jenter (N=157) og gutter (N=149)

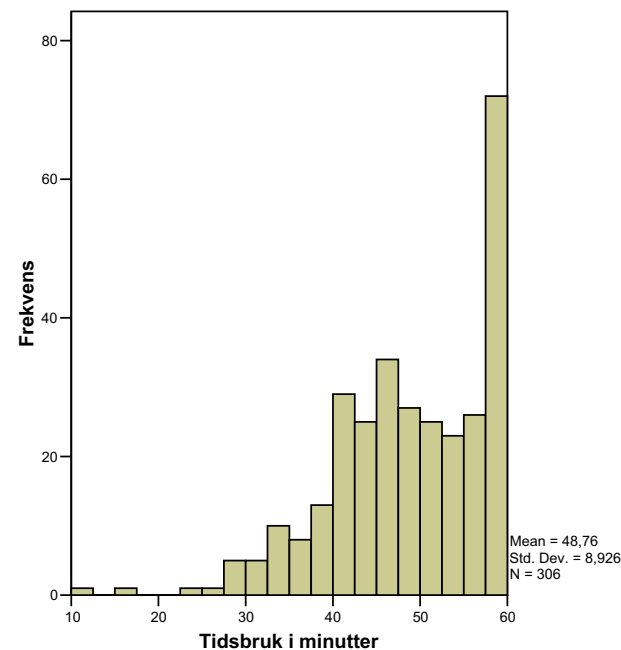
	Gjennomsnittlig tidsbruk på PC-testen (i minutter)	Spredning (SD)
Jenter	48,8	9,0
Gutter	48,7	8,9

Resultatene i tabell 15 viser bare en ubetydelig kjønnsforskjell når det gjelder gjennomsnittlig tidsbruk. Videre analyser viser også relativt liten variasjon mellom de 10 ulike versjonene av testen. Høyest gjennomsnittlig tidsbruk er 50,1 minutter (versjon 8), mens den laveste er 46,3 minutter (versjon 10). Resultatene på figur 8

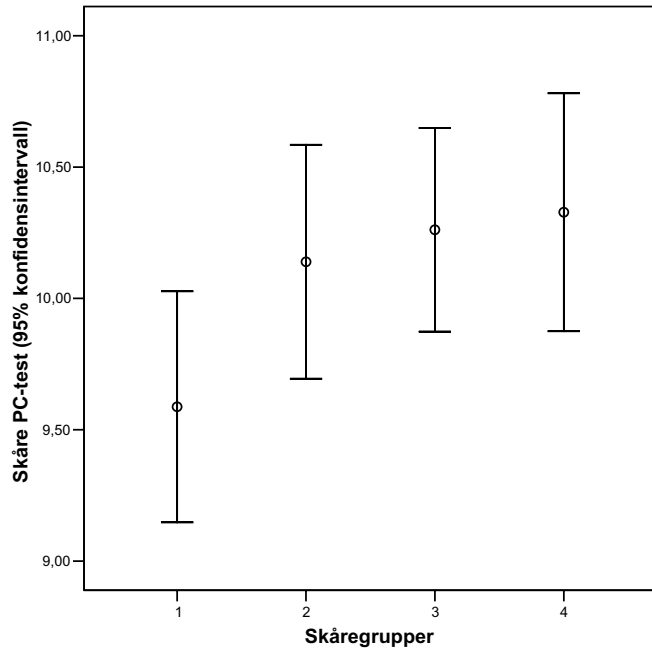
viser at en betydelig andel av elevene har brukt bortimot hele timen de hadde til rådighet; 34 prosent av elevene har brukt 55 minutter eller mer. Kun en liten andel av elevene (3 prosent) har brukt mindre enn halvparten av tilgjengelig testtid.

Korrelasjonen mellom tidsbruk og skåre på PC-testen er 0,20, og den er like sterk for jenter som for gutter. Figur 9 viser gjennomsnittlig skåre for fire skåregrupper definert ut fra kvartilene til tidsfordelingen. Figuren viser at elevene i første kvartil (anvendt tid mindre enn 43 minutter) skårer lavest. Forskjellene i gjennomsnittsskåre mellom de tre andre skåregruppene er imidlertid relativt små.

Figur 8: Tidsbruk på PC-testen (i minutter) (N=306)



Figur 9: Gjennomsnittlig skåre for fire skåregrupper definert ut fra kvartilene til tidsfordelingen (N=306)



Resultatene viser videre at det er signifikant positiv sammenheng mellom tidsbruk på PC-testen og elevenes hjemmebakgrunn målt i form av antallet bøker i elevens hjem ( $p=0,05$ ). Korrelasjonen er 0,23.

### 3.13 "Media-play" i PC-testen

Tabell 16 viser gjennomsnittlig antall "media-play" for jenter og gutter, samt for alle elevene sett under ett. Resultatene viser at guttene har flere avspillinger enn jentene, og forskjellen er klart statistisk signifikant. Spredningen blant guttene er også tydelig

større. Som tidligere nevnt, inngikk 46 oppgaver i hver av versjonene av testen, noe som gir 1,6 avspillinger i gjennomsnitt per oppgave for alle elevene sett under ett.

Figurene 10 og 11 viser fordelingen av antallet "media-play" for henholdsvis jenter og gutter. Blant jentene er det 5 prosent som har mindre enn én avspilling per oppgave i gjennomsnitt, mot bare 1 prosent av guttene. I den andre enden av fordelingen har 22 prosent av guttene to avspillinger per oppgave eller mer i gjennomsnitt, mot 10 prosent av jentene.

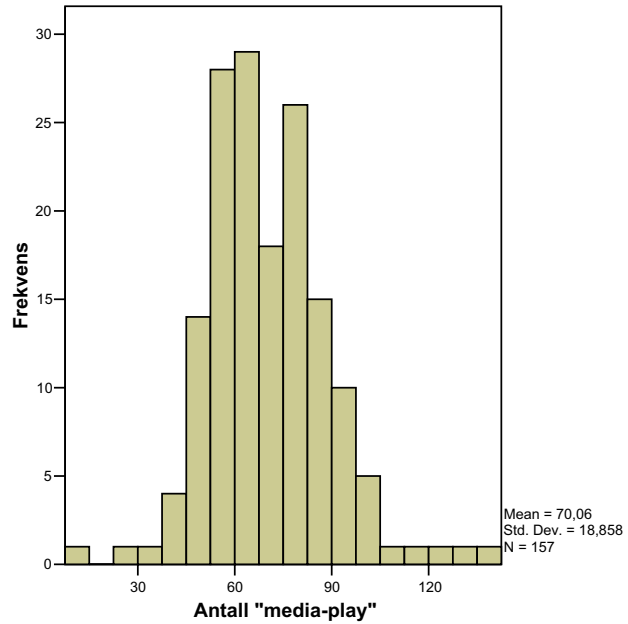
Vi ser altså her en tydelig kjønnsforskjell når det gjelder løsningsstrategier for testen som helhet. Selv om jentene og guttene i gjennomsnitt bruker like lang tid på PC-testen, anvender de tiden forskjellig. Guttene bruker mer tid til aktiv handling i forhold til testen.

Tabell 16: Gjennomsnittlig antall "media-play" for jenter (N=157) og gutter (N=149)

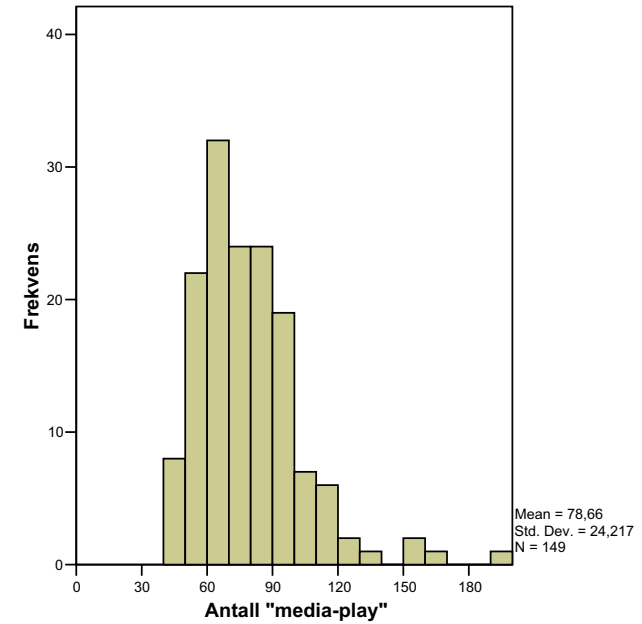
	Gjennomsnitt	Spredning (SD)
Jenter	70,1	18,9
Gutter	78,7	24,2
Alle	74,3	22,0



Figur 10: Fordelingen av "media-play" for jenter (N=157)

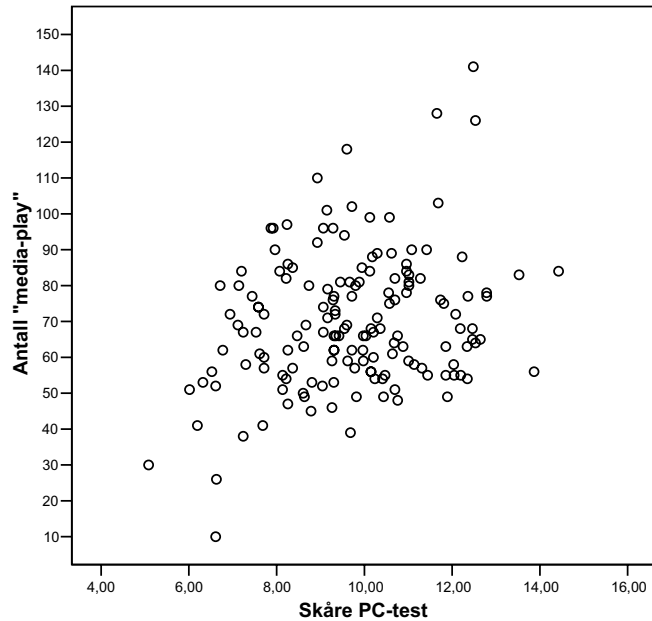


Figur 11: Fordeling av "media-play" for gutter (N=149)

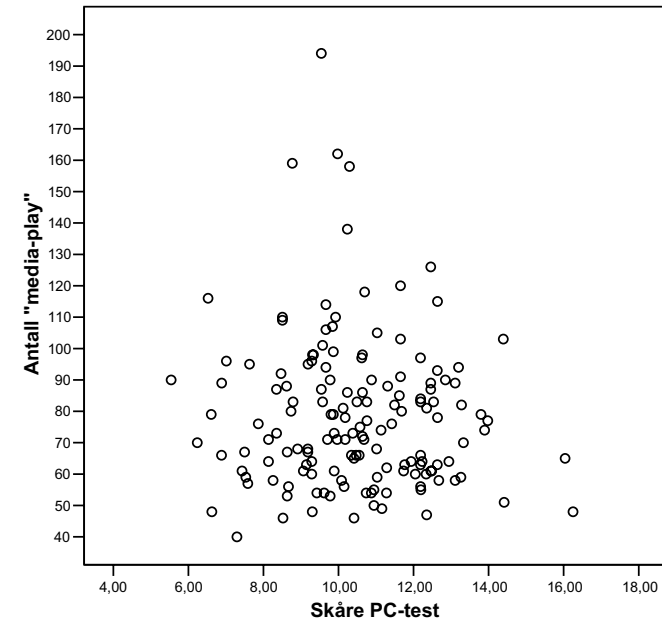


Figurene 12 og 13 viser sammenhenger mellom antallet "media-play" og skåre på PC-testen for henholdsvis jenter og gutter. Det er en signifikant positiv sammenheng for jentene, mens en svak negativ og ikke signifikant sammenheng kan observeres for guttene.

Figur 12: Skåre på PC-testen (x-aksen) og antallet "media-play" (y-aksen) for jenter (N=157). Korrelasjonen er 0,28.



Figur 13: Skåre på PC-testen (x-aksen) og antallet "media-play" (y-aksen) for gutter (N=149). Korrelasjonen er -0,07 (ikke signifikant).



Vi finner videre en svak positiv korrelasjon mellom elevenes egenvurderte PC-kompetanse og antallet "media-play" (0,11). Korrelasjonen er imidlertid ikke statistisk signifikant.

Resultatene viser også at det er en svak negativ sammenheng mellom hvor godt guttene likte PC-testen og hvor mange avspillinger de har gjort i løpet av testen. Korrelasjon er -0,17 og statistisk signifikant. For jentene finner vi imidlertid en motsatt effekt i form av en statistisk signifikant positiv korrelasjon (0,18).

Med andre ord er det en tendens til at jo flere avspillinger jentene har gjort, desto bedre likte de PC-testen.

## 4. Avslutning

### 4.1 Konklusjon

De to testene i CBAS-undersøkelsen tok utgangspunkt i samme definisjon av scientific literacy:

*Scientific literacy is the capacity to use scientific knowledge, to identify questions and to draw evidence-based conclusions in order to understand and help make decisions about the natural world and the changes made to it through human activity* (OECD, 2003a, s. 133).

Definisjonen ble likevel operasjonalisert med ulike typer oppgaver. Oppgavene i papirtesten inneholder en god del forklarende tekst, mens oppgavene i PC-testen inneholder betydelig mindre tekst, men til gjengjeld illustrasjoner i form av multimedia-elementer. PC-testen inneholder kun flervalgsoppgaver, mens papirtesten også har oppgaver hvor elevene må formulere svaret selv.

De to testene gir ganske forskjellige resultater, ikke minst sett i et kjønnsperspektiv. PC-testen favoriserer klart guttene, og hovedforklaringen synes å ligge i at kravet til lesekompetanse er mindre, og at guttene er mer motiverte for og aktive i forhold til den PC-baserte testen. Analysene av elevenes aktive bruk av multimedia-elementer ("media play") viser et interessant kjønnsmønster. Også svake elever blant guttene er aktive i forhold til mediet, mens for jentene er dette i større grad noe som kjennetegner de som skårer høyt, jfr. figurene 12 og 13.

Man kan diskutere hvilket av de to målene som gir det "sanneste" målet for *scientific literacy*. Det kan hevdes at PC-testen gir et

bedre mål enn papirtesten fordi kravet til lesekompetanse begrenses. Men man kan også hevde at lesekompetanse er en viktig og integrert del av det å være *scientific literate* (Fang, 2005, Norris & Phillips, 2003, Yore mfl., 2003). Lesekompetanse er en nødvendig forutsetning for å kunne tilegne seg kunnskaper og informasjon fra skrevne tekster, også tekster med et naturfaglig innhold. Som påpekt av blant andre Roe (2006), øker tekstmengden i samfunnet stadig, ikke minst som en følge av Internett. For å kunne tilegne seg og dra nytte av den informasjonen som er relevant og tilgjengelig, er god lesekompetanse viktigere enn noen sinne. Det er derfor sterke grunner til å regne lesekompetanse som en viktig, integrert del av begrepet *scientific literacy*. På den annen side viser resultatene fra CBAS-undersøkelsen at IKT-kompetanse og motivasjon for å ta testen på PC virker inn på resultatene på PC-testen.

En viktig lærdom fra CBAS-undersøkelsen er at det ikke er uproblematisk å flytte en test fra papir og over på PC. Selv om PC-baserte tester kan ha åpenbare fordeler, bør man ha et bevisst forhold til hvordan denne overføringen påvirker hva slags kompetanse som måles. Slike diskusjoner er ikke minst viktige i forbindelse med ønsket om å gjennomføre framtidige nasjonale prøver og eksamener i Norge på PC.

### 4.2 Veien videre for CBAS

Generalprøven i CBAS-undersøkelsen våren 2005 ble gjennomført i et tilfeldig utvalg av skoler i begrenset avstand fra Oslo. Internasjonalt var planen å gjennomføre en CBAS-undersøkelse i full skala våren 2006. Dette medfører at et tilfeldig utvalg skoler trekkes fra hele landet, og at testadministratorer reiser til den enkelte skole med alt utstyr som trengs i testen på samme måte som i generalprøven. Et slikt design ville blitt en stor logistisk utfordring i et land som Norge. Undersøkelsen ville følgelig blitt

svært kostbar. På bakgrunn av dette ble det bestemt at Norge ikke skulle delta i hovedundersøkelsen i CBAS våren 2006.

Internasjonalt viste det seg at de fleste av de 13 landene som deltok i generalprøven våren 2005, konkluderte på samme måte som Norge. Likevel vil CBAS-undersøkelsen gjennomføres i full skala i tre land våren 2006; Island, Danmark og Korea.

Fordi så mange av landene vurderte det foreslåtte opplegget som for komplisert å gjennomføre, har PISA-konsortiet lansert et forslag om å prøve ut en Internettbasert løsning våren 2007. Norge har sagt seg interessert i å delta i denne utprøvingen. Som vi tidligere har vært inne på, vurderer vi denne typen løsning som vesentlig mer framtidsrettet enn det designet som ble prøvd ut våren 2005.

### **Referanser**

ACER (2005a): *Computer-based Assessment of Science: Hardware and Software Requirements and Configuration*. Doc: CBAS (0409)4. Document to the National Project Managers' Meeting, Bratislava, Slovak Republic. Melbourne: Australian Council for Educational Research.

ACER (2005b): *CBAS 2006 Preliminary Field Trial Data Analysis*. Doc: CBAS(0510)2. Document to the National Project Managers' Meeting, Mildura, Australia. Melbourne: Australian Council for Educational Research.

Adams, R.J. (2002): Scaling PISA Cognitive Data. In Adams, R.J. & Wu, M. (Eds.): *PISA 2000 Technical Report*. Paris: Organization for Economic Co-Operation and Development.

Artelt, C., Baumert, J., Julius-McElvany, N. & Peschar, J. (2003): *Learners for Life. Student Approaches to Learning*. Paris: Organisation for Economic Co-Operation and Development.

Bourdieu, P. (1984): *Distinction: A Social Critique of The Judgement of Taste*. Harvard: Harvard University Press.

Buchmann, C. (2003): Measuring Family Background in International Studies of Education: Conceptual Issues and Methodological Challenges. In Porter, A. & Gamoran, A. (Eds.): *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington: National Academy Press.

Burkhardt, H. & Pead, D. (2003): Computer-based assessment: a platform for better tests? In Richardson, C. (Ed.): *Whither Assessment?* London: Qualifications and Curriculum Authority.

Coleman, J.S. (1990): *Foundations of a social theory*. Harvard: Harvard University Press.

Fang, Z. (2005): Scientific Literacy: A Systemic Functional Linguistic Perspective. *Science Education*, 89 (2), pp. 335-347.

Grisay, A. (2002): Translation and Cultural Appropriateness of the Test and Survey Material. In Adams, R.J. & Wu, M. (Eds.): *PISA 2000 Technical Report*. Paris: Organization for Economic Co-Operation and Development.

Grønmo, L.S., Bergem, O.K., Kjærnsli, M., Lie, S. & Turmo, A. (2004): *Hva i all verden har skjedd i realfagene? Norske elevers prestasjoner i matematikk og naturfag i TIMSS 2003*. Acta Didactica 5/2004. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Hambleton, R.K. (2002): Adapting Achievement Tests into Multiple Languages for International Assessments. In Porter, A.C. & Gamoran, A. (Eds.): *Methodological Advances in Cross-National Surveys of Educational Achievement*. Washington: National Academy Press.

Harvey-Beavis, A. (2002): Student and School Questionnaire Development. In Adams, R.J. & Wu, M. (Eds.): *PISA 2000 Technical Report*. Paris: Organization for Economic Co-Operation and Development.

Kind, P.M. (1996): *Exploring Performance Assessment in Science*. Oslo: Dr.scient.-avhandling, Universitetet i Oslo.

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

Kirsch, I., de Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J. & Monseur, C. (2002): *Reading for Change. Performance and Engagement Across Countries*. Paris: Organisation for Economic Co-Operation and Development.

Kjærnsli, M., Lie, S., Olsen, R.V., Roe, A. & Turmo, A. (2004): *Rett spor eller ville veier? Norske elevers prestasjoner i matematikk, naturfag og lesing i PISA 2003*. Oslo: Universitetsforlaget.

Kjærnsli, M., Lie, S. & Turmo, A. (2005): Kan elevene mindre enn før? Naturfagkompetanse i Norden i perioden 1995-2003. *Nordisk Didaktikk i Naturfag*, 1 (2), ss. 51-60

Lie, S., Hopfenbeck, T.N., Ibsen, E. & Turmo, A. (2005): *Nasjonale prøver på ny prøve. Rapport fra en utvalgsundersøkelse for å analysere og vurdere kvaliteten på oppgaver og resultater til nasjonale prøver våren 2005*. Acta Didactica 1, 2005. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Lie, S., Kjærnsli, M. & Brekke, G. (1997): *Hva i all verden skjer i realfagene? Internasjonalt lys på trettenåringers kunnskaper, holdninger og undervisning i norsk skole*. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Lie, S., Kjærnsli, M., Roe, A. & Turmo, A. (2001): *Godt rustet for framtida? Norske 15-åringers kompetanse i lesing og realfag i et internasjonalt perspektiv*. Oslo: Acta Didactica 4/2001. Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Lie, S., Linnakylä, P. & Roe, A. (Eds.) (2003): *Northern Lights on PISA. Unity and diversity in the Nordic countries in PISA 2000*. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Norris, S. P. & Phillips, L. M. (2003): How Literacy in Its Fundamental Sense Is Central to Scientific Literacy. *Science Education*, 87 (2), pp. 224-240.

OECD (1999): *Measuring Student Knowledge and Skills. A New Framework for Assessment*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2000): *Measuring Student Knowledge and Skills- The PISA 2000 Assessment of Reading, Mathematical, and Scientific Literacy*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2001): *Knowledge and Skills for Life. First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2003a): *The PISA 2003 Assessment Framework*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2003b): *Literacy Skills for the World of Tomorrow. Further results from PISA 2000*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2004a): *Learning for Tomorrow's World. First Results from PISA 2003*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2004b): *Problem Solving for Tomorrow's World. First Measures of Cross-Curricular Competencies from PISA 2003*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2004c): *What makes School Systems Perform? Seeing school Systems through the Prism of PISA*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2005a): *PISA 2003 Technical Report*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2005b): *School Factors related to Quality and Equity. Results from PISA 2000*. Paris: Organisation for Economic Co-Operation and Development.

OECD (2006): *Are Students Ready for a Technology-Rich World? What PISA Studies Tell Us*. Paris: Organisation for Economic Co-Operation and Development.

Olsen, R.V., Turmo, A. & Lie, S. (2001): Learning about students' knowledge and thinking in science through large-scale quantitative studies. *European Journal of Psychology of Education*, 16 (3), pp. 401-418.

Roe, A. (2006): Leseopplæring og lesestrategier. Kommer i Elstad, E. & Turmo, A. (red.): *Læringsstrategier. Søkelys på lærerens praksis*. Universitetsforlaget (juni 2006).

Schleicher, A. (2006): *Are students ready for a technology-rich world? What PISA studies tell us*. Presentasjon i Power point som kan lastes ned fra [www.pisa.oecd.org](http://www.pisa.oecd.org). Paris: Organisation for Economic Co-Operation and Development.

Turmo, A. & Lie, S. (2004): *Hva kjennetegner norske skoler som skårer høyt i PISA 2000?* Oslo: Acta Didactica 1, 2004. Oslo: Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

Turmo, A. (2004): Scientific Literacy and Socio-Economic Background among 15-year-olds - A Nordic Perspective. *Scandinavian Journal of Educational Research*, 48 (3), pp. 287-305.

Turmo, A. (2005): The relationship between the use of learning strategies and socioeconomic background in 15-year olds. *Journal of Nordic Educational Research*, 25 (2), pp. 155-168.

Willms, J.D. (2003): *Student Engagement at School. A Sense of Belonging and Participation. Results from PISA 2000*. Paris: Organisation for Economic Co-Operation and Development.

Wu, M. (2002): Test Design and Test Development. In Adams, R.J. & Wu, M. (Eds.): *PISA 2000 Technical Report*. Paris: Organisation for Economic Co-Operation and Development.

Yore, L. D., Bisanz, G. L. & Hand, B. M. (2003): Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education*, 25 (6), pp. 689-725.