

Are Turmo (f. 1973) har doktorgrad i realfagdidaktikk (dr.scient.) fra Universitetet i Oslo. Han har vært forsker ved Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo siden 1998. Svein Lie (f. 1941) har doktorgrad (dr.philos.) innen teoretisk kjernefysikk fra Universitetet i Oslo. Han har mange års erfaring som lærer i videregående skole der han særlig har undervist i matematikk og fysikk. Siden 1991 har han vært ansatt ved Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo, nå som professor. Begge forfatternes forskningsområde er internasjonale, komparative storskalastudier av skoleelevers kompetanse, særlig i matematikk og naturfag.

ARE TURMO

Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo, Norge
are.turmo@ils.uio.no

SVEIN LIE

Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo, Norge
svein.lie@ils.uio.no

Vurdering av naturfagkompetanse på PC - Norske resultater fra generalprøven i PISA CBAS

Abstract

This article reports from a computer-based assessment of scientific literacy. The study was implemented in Norway in the period from March to June 2005 by the use of 6 carry-in laptops. In total, 315 15-year old students participated. All students did a one hour paper-based test and a one-hour computer-based test. In addition to the cognitive response data, also behavioural data were recorded from the computer-based test. The results show no statistically significant gender difference for the paper-based test, while a substantial gender difference in favour of boys is established for the computer-based test. The PC test seems to favour boys because of the lower reading load. Boys also express higher motivation and larger preference for doing the test on PC compared to on paper. The behavioural data show that boys make more active use of the multimedia elements in the PC test than the girls do.

INTRODUKSJON

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? Denne artikkelen 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 Programme for International Student Assessment's (PISA) 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 (Informasjons- og kommunikasjonsteknologi).

SCIENTIFIC LITERACY I PISA

Scientific literacy er i PISA definert som følger:

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, 2003, 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.

COMPUTER-BASED ASSESSMENT OF SCIENCE (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.

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.

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 og 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.

Alle PC-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. Ingen av PC-oppgavene er frigitt og de kan derfor ikke beskrives i stor detalj her. Noen typiske trekk kan vi likevel beskrive. For mange av oppgaveenhetene blir en situasjon vist ved avspilling av en kort videosekvens som elevene selv må aktivere. Typisk kan videoen vise et eksperiment eller en dagligdags situasjon. Elevene skal så besvare ett eller flere spørsmål knyttet til videoen. Oppgavene er laget for å operasjonalisere definisjonen av *scientific literacy* slik den er videre presisert i rammeverket i PISA (OECD, 2003). Eksempler på typiske oppgaver som inngår i papirtestene i PISA, kan finnes på www.pisa.oecd.org, samt i nasjonale PISA-rapporter. Også disse oppgavene opptrer i oppgaveenheter og er knyttet til en beskrevet kontekst.

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-elevene, 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.

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, 2005). I gjennomsnitt deltok 8 elever per skole. Tabell 1 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 1. Antallet deltakende elever ved 40 skoler

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

Gjennomføring på skolene

Undersøkelsen ble gjennomført med ett sett av bærbare PCer bestående av fem 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 en 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 (Kjærnsli, Lie & Turmo, 2005). Dette prinsippet ble også fulgt i CBAS-undersøkelsen, både i testen på papir og i testen på PC. Totalt inngikk 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 tidsrammen på 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 treningssesjon med eksempeloppgaver og forklaringer på hvordan de skulle svare og manøvrere seg gjennom testen.

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 artikkelen.

CBAS generalprøve 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 artikkelen vil derfor fokuset være på norske resultater av spesiell interesse.

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 artikkelen, er det tatt hensyn til disse trekkene ved de tilgjengelige dataene.

EMPIRISKE RESULTATER

Sammenhengen mellom resultatene på de to testene

Som tidligere nevnt, deltok totalt 315 elever i CBAS-undersøkelsen. For disse elevene har vi med andre ord i utgangspunktet to skåreverdier for naturfagkompetanse. I Norge er reliabiliteten til papirtesten 0,68, mens reliabiliteten til PC-testen er 0,79. Fire av de 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 fem elever som skåret mer enn tre standardavvik under gjennomsnittet på en eller begge testene. Disse elevene regner vi som "outliers". Det er overveiende sannsynlig at disse elevene har boikottet deler av testen, og de er derfor ikke inkludert i analysene. Analysene er dermed basert på 306 elever.

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

Kjønnsforskjeller i faglige prestasjoner

Tabell 2 sammenlikner kjønnsforskjellene i prestasjoner på de to delene av CBAS-undersøkelsen. Resultatene for de to testene er skalert slik at 10 er gjennomsnittet og 2 er standardavviket. Resultatene viser en markert forskjell mellom de to delene av undersøkelsen. For begge delene er det en kjønnsforskjell i favør av guttene, men denne forskjellen er mye større for PC-testen.

For resultatene i tabell 2 blir feilmarginene betydelige som en konsekvens av det relativt lave antallet elever. De to testene er ikke normert helt på samme måten fordi også elevene i den vanlige PISA-undersøkelsen er med og bestemmer normering for papirtesten. En detaljert sammenlikning av resultatene for hvert kjønn gir derfor ikke mening. Derimot gir det mening å sammenlikne gutters og jenters resultater for hver prøve for seg. Som det framgår av tabell 2, skårer guttene signifikant bedre enn jentene på PC-testen. Forskjellen mellom kjønnene på papirtesten er derimot ikke signifikant.

Tabell 2. Kjønnsforskjeller i prestasjoner i CBAS-undersøkelsen. Kjønnsforskjellene er angitt ved *t* og *d* (effektstørrelse).

	Jenter N=157		Gutter N=149		<i>t</i> ¹⁾	<i>d</i>
	Gj.snitt.	S.D.	Gj.snitt.	S.D.		
Resultat PC-test	9,71	1,79	10,26	1,87	-3,52**	-0,38
Resultat papirtest	10,38	1,84	10,47	1,96	-0,57 ^{ns}	-0,06

¹⁾ ns: $p > 0,05$, * $p < 0,05$, ** $p < 0,01$

Selv om PC-testen og papirtesten begge søker å måle *scientific literacy*, er det ingen felles oppgaver i testene. Tidligere forskning har vist at kjønnsforskjeller i naturfagprestasjoner varierer avhengig av delområde i faget. I TIMSS-undersøkelsen i 2003 fant man 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 kjønnsforskjellen.

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 det dreier seg om.

Tabell 3 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 gjerne har vist seg å være et "guttefag", mens biologi gjerne er et "jentefag", kan forskjellene mellom kjønnene på de to testene tenkes å henge sammen med forskjeller i faglig innhold.

Tabell 3. Prosentvis fordeling av oppgaver på ulike fagområder for PC-testen og papirtesten.

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

For å teste hypotesen om at kjønnsforskjellen kan tilskrives forskjeller i faglig vektlegging, valgte det internasjonale PISA-konsortiet ut et utvalg av PC-oppgavene innen ulike fagområder ut fra samme fordeling som i papirtesten. I motsetning til på papirtesten viste det seg at guttene skårer bedre enn jentene innen alle fagområder (ACER, 2005).

Sammenhenger mellom faglige resultater og karakterer i sentrale skolefag

I den norske versjonen av elevspørreskjemaene ble det lagt inn spørsmål om hvilken karakter elevene fikk i sentrale skolefag ved siste karakteroppgjør. De utvalgte fagene var norsk, matematikk og natur- og miljøfag.

Tabell 4 viser sammenhengen mellom skåre på de to delene av CBAS-undersøkelsen og karakterer i disse tre fagene. Et påfallende trekk ved resultatene er at korrelasjonene er relativt lave, også for natur- og miljøfag. Faktisk er korrelasjonen med karakter i matematikk for PC-testen sterkere enn tilsvarende korrelasjon for natur- og miljøfag. Det er også interessant å påpeke at korrelasjonen mellom karakteren i norsk og resultater for papirtesten er sterkere enn tilsvarende korrelasjon for PC-testen. Dette resultatet kan forstås i lys av at papirtesten inneholder betydelig mer tekst som elevene må forholde seg til enn det PC-testen gjør. 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" mål for naturfagkompetanse.

Tabell 4. Korrelasjoner mellom karakterer i sentrale skolefag og skåre i CBAS-undersøkelsen. I parentes er vist korrelasjoner korrigert for tilfældige målefeil.

Skolefag	Korrelasjon mellom PC-test og karakter ¹⁾	Korrelasjon mellom papirtest og karakter ¹⁾
Norsk	0,28** (0,32)	0,40** (0,49)
Matematikk	0,46** (0,52)	0,46** (0,56)
Natur- og miljøfag	0,36** (0,41)	0,42** (0,51)

¹⁾ ns: $p > 0,05$, * $p < 0,05$, ** $p < 0,01$

Resultatene tyder videre på at CBAS-undersøkelsen i stor grad måler noe annet enn det karakteren i natur- og miljøfag er et uttrykk for. Det må 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 høyere enn 0,68 og 0,79 for henholdsvis papirtesten og PC-testen, vitner det om at vi har en betydelig målefeil for de målte skåreverdiene. Disse målefeilene bidrar til å "utvanne" korrelasjonene, siden de setter en øvre grense for hvor store korrelasjonene kan bli.

Det går an å "korrigere" for dette forholdet (ved å dividere med kvadratroten av reliabiliteten), og de korrigerte korrelasjonene er også vist i tabell 4. Hvis vi studerer tallene i parentes, blir det enda tydeligere at PC-resultatene korrelerer betydelig lavere med norsk karakteren enn det papirtesten gjør. Men vi merker oss også med noe undring at korrelasjonene for begge versjonene er høyest for matematikk-karakterens vedkommende.

IKT-ferdigheter / -bruk og prestasjoner

Elevspørreskjemaene inneholdt flere spørsmål om elevenes erfaringer med bruk av IKT samt deres egenvurdering 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 5 gir en oversikt over hva elevene ble spurt om å vurdere sine ferdigheter i forhold til.

Tabell 5. Oversikt over spørsmål om egenvurdering 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:

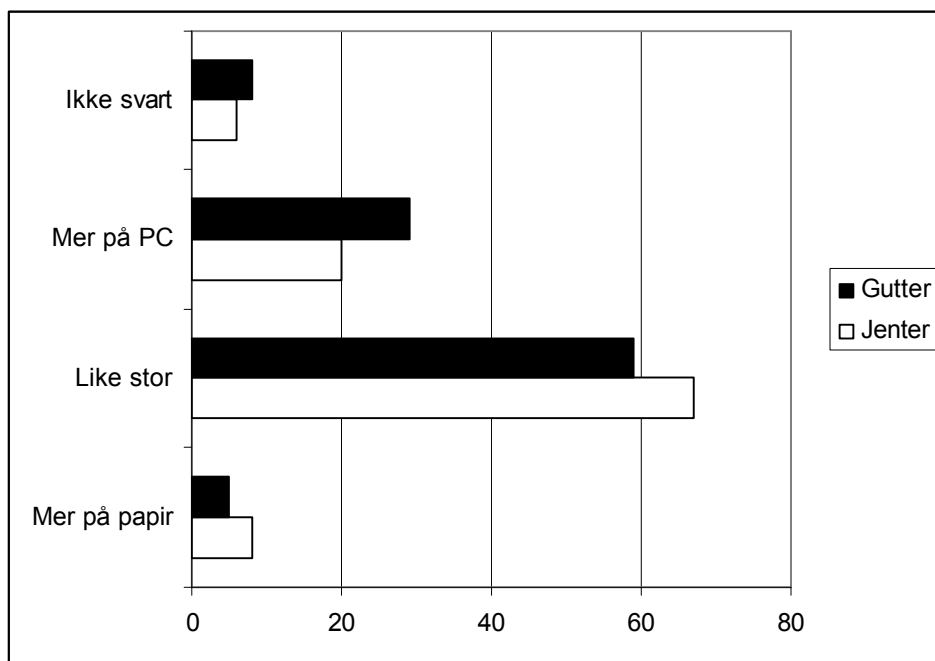
1. Dette kan jeg gjøre uten hjelp fra noen andre.
2. Jeg kan gjøre dette hvis jeg får hjelp av noen andre.
3. Jeg vet hva dette er, men vet ikke hvordan det kan gjøres.
4. Jeg vet ikke hva dette er.

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.

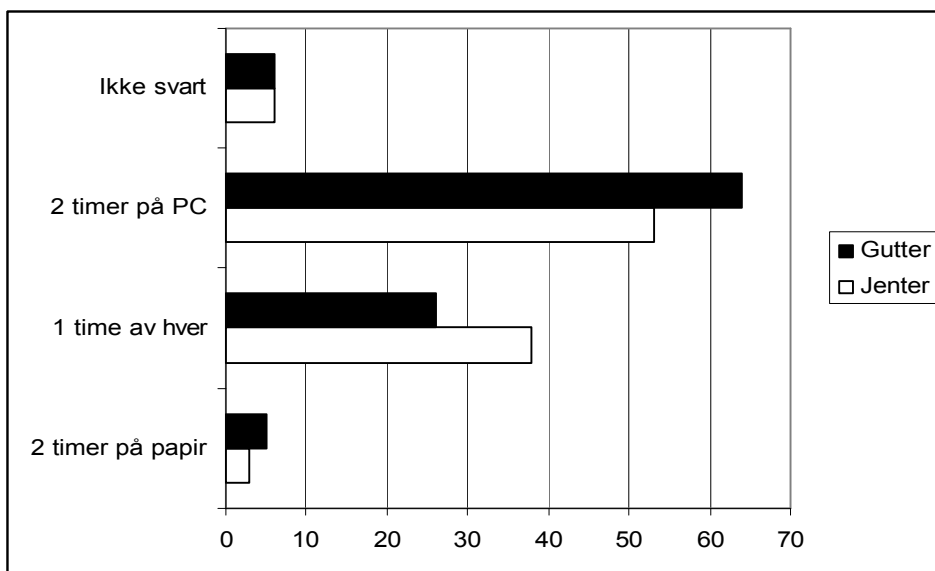
Elevenes innsats og preferanser

Figur 1 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 1: Elevenes vurdering av egen innsats på de to delene av CBAS-undersøkelsen. Prosentfordeling. (N=315)

Figur 2 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.



Figur 2: Elevenes preferanser ved en eventuell ny totimers prøve. Prosentfordeling. (N=315)

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 er 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 6 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 med PC-testen, differansen mellom PC-test og papirtest eller med begge disse målene ($p < 0,05$).

Tabell 6. 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 ¹⁾	Korrelasjon med differansen mellom de to skåreverdiene ¹⁾
Kjønn (jente=1, gutt=2)	0,20**	0,22**
Karakter i norsk	0,28**	-0,14*
Karakter i matematikk	0,46**	0,01 ^{ns}
Karakter i naturfag	0,36**	-0,05 ^{ns}
PC-kompetanse	0,07 ^{ns}	0,13*
Preferanse for PC-testen	0,09 ^{ns}	0,13*
Antall bøker hjemme ²⁾	0,41**	-0,03 ^{ns}
Gjenstander i hjemmet ²⁾	0,28**	0,03 ^{ns}

¹⁾ ns: $p > 0,05$, * $p < 0,05$, ** $p < 0,01$

²⁾ Dette brukes i PISA som indikator på sosioøkonomisk status, se Turmo (2004).

La oss kommentere noen tydelige trekk i tabell 6. 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 norskarakter. Dette er nettopp hva vi ville forvente for en vellykket IKT-basert måling av naturfagkompetanse.

Tabell 6 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 6 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 6 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.

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. 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.

I det norske utvalget er det totalt registrert i overkant av 91 000 hendelser.

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 her. 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 tilleggsinformasjon 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.

Tidsbruk og antall "media-play" på PC-testen

Tabell 7 viser gjennomsnittlig tidsbruk på PC-testen for jenter og gutter (se første linje). Tiden for avgitt svar ("item attempt") ble registrert idet eleven gikk videre til neste oppgave. Tiden for siste

Tabell 7. Gjennomsnittlig tidsbruk og antall "media-play" på PC-testen for jenter og gutter.

	Jenter N=157		Gutter N=149		t¹⁾	d
	Gj.snitt.	S.D.	Gj.snitt.	S.D.		
Tidsbruk (min) PC-test	48,9	9,0	48,7	8,9	0,17 ^{ns}	0,02
Antall "media-play"	70,1	18,9	78,7	24,2	-3,48 ^{**}	0,39

¹⁾ ns: $p > 0,05$, * $p < 0,05$, ** $p < 0,01$

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.

Resultatene i tabell 7 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 viser videre 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. Man kan studere gjennomsnittlig skåre for fire skåregrupper definert ut fra kvartilene til tidsfordelingen. Resultatene 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.

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.

Tabell 7 viser også gjennomsnittlig antall "media-play" for jenter og gutter (se andre linje). 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.

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.

Det er en signifikant positiv sammenheng antallet "media-play" og skåre på PC-testen for jentene (korrelasjonen er 0,28), mens en svak negativ og ikke signifikant sammenheng kan observeres for guttene (-0,07). 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. Forskjellen mellom kjønnene når det gjelder sammenhenger mellom antall "media play" og holdninger til og resultater på testen, er ikke enkle å fortolke. Men dette må sees i lys av at guttene generelt, også svake elever, har mye høyere interaksjon med testen enn jentene (se tabell 7).

KONKLUSJON

De to testene i CBAS-undersøkelsen tok utgangspunkt i samme definisjon av *scientific literacy*. 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. Det kan være interessant å reflektere over hva det er ved PC-oppgavene som gjør dem mer motiverende for guttene. En slik diskusjon er ikke minst interessant i lys av nyere forskning knyttet til dataspill og den motiverende effekten som kan ligge i det interaktive aspektet. Det vil imidlertid føre for langt å ta opp denne diskusjonen i detalj her.

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 *scientifically literate* (Fang, 2005, 2006; Norris & Phillips, 2003; Yore et al., 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 uproblemtisk 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.

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 anbefalte vi at Norge ikke burde delta i hovedundersøkelsen i CBAS våren 2006. Denne anbefalingen ble tatt til følge av Kunnskapsdepartementet og Utdanningsdirektoratet.

CBAS-undersøkelsen vil gjennomføres i full skala i tre land våren 2006; Island, Danmark og Korea. Vi ser spesielt med spenning fram mot de analysene som kan gjøres i Danmark og Island

basert på et større utvalg enn det vi hadde i Norge i 2005. 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. Vi vurderer denne typen løsning som vesentlig mer framtidsrettet enn det designet som ble prøvd ut våren 2005.

TAKK

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 Institutt for lærerutdanning og skoleutvikling, Universitetet i Oslo.

REFERANSER

- ACER (2005). *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.
- Fang, Z. (2005). Scientific Literacy: A Systemic Functional Linguistic Perspective. *Science Education*, 89 (2), 335-347.
- Fang, Z. (2006). The Language Demands of Science Reading in Middle School. *International Journal of Science Education*, 28 (5), 491-520.
- 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.
- 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.
- Kjærnsli, M, Lie, S. & Turmo, A. (2005). Kan elevene mindre enn før? Naturfagkompetanse i Norden i perioden 1995-2003. *NorDiNa*, 2, 51-60.
- Norris, S. P. & Phillips, L. M. (2003). How Literacy in Its Fundamental Sense Is Central to Scientific Literacy. *Science Education*, 87 (2), 224-240.
- OECD (2003). *The PISA 2003 Assessment Framework*. Paris: Organisation for Economic Co-Operation and Development.
- OECD (2005). *PISA 2003 Technical Report*. 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), 401-418.
- Roe, A. (2006): Leseopplæring og lesestrategier. Kommer i Elstad, E. & Turmo, A. (red.). *Læringsstrategier. Søkelys på lærernes praksis*. Oslo: Universitetsforlaget (august 2006).
- Turmo, A. (2004). Scientific Literacy and Socio-Economic Background among 15-year-olds - A Nordic Perspective. *Scandinavian Journal of Educational Research*, 48 (3), 287-305.
- 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), 689-725.