

Erfaringsrapport IKT-baserte nasjonale prøver i engelsk

BITE^{IT}

Bergen Interaktive tester i engelsk

Juli 2004

InterMedia 31. juli 2004

Skrevet av

Mona Elise Arntzen (*testutvikler*)
Steinar Dragsnes (*teknisk kontaktperson*)
Cecilie Hansen (*administrativ skolekontakt*)
Olve Hansen (*systemutvikler/arkitekt*)
Konrad Morgan (*stedfordredende prosjektleder*)
Henrik Schlanbusch (*teknisk prosjektleder*)
Lars Schlanbusch (*testutvikler*)

Prosjektgruppens medlemmer

Mona Elise Arntzen (*testutvikler*)
Jarle Bareksten (*testutvikler*)
Steinar Dragsnes (*teknisk kontaktperson*)
Cecilie Hansen (*administrativ skolekontakt*)
Olve Hansen (*systemutvikler/arkitekt*)
Øyvind Meistad (*systemutvikler*)
Konrad Morgan (*stedfordredende prosjektleder*)
Henrik Schlanbusch (*teknisk prosjektleder*)
Lars Schlanbusch (*testutvikler*)
Barbara Wasson (*prosjektleder*)

<http://www.intermedia.uib.no/projects/bite-it>
<http://info.nationaltests.no/>

Innholdsfortegnelse

1	SAMMENDRAG	4
2	PROSJEKTGRUNNLAG FOR BITE^{IT}	5
3	OPPSUMMERING AV PROSJEKTET.....	6
4	TALLENES TALE.....	9
4.1	VURDERING AV ELEVENE	9
4.2	KARAKTERER OG RESULTATER	10
4.2.1	<i>Resultater fordelt på oppgavesett</i>	<i>11</i>
4.2.2	<i>Hendelser bak tallene</i>	<i>11</i>
4.3	JUKS	12
4.4	ARKITEKTUR	12
5	DESIGN OG UTVIKLING	14
5.1	DESIGNVALG FOR ET BRUKERVENNLIG GRENSESNI TT	14
5.2	FASER I UTVIKLINGSARBEIDET	15
5.3	SAMARBEID MELLOM ULIKE FAGGRUPPER	16
5.4	HVA ER ET OPPGAVEFORMAT?	17
5.5	HOVEDFORMATER	17
5.6	PILOTUTPRØVINGER - FORMATIV EVALUERING	21
5.7	GJENNOMFØRING	21
5.8	ENDRINGER SOM RESULTAT AV UTPRØVINGENE.....	22
6	WEBSYSTEMENE	25
6.1	ELEVSYSTEM.....	25
6.2	OPPGAVEBANK.....	26
6.3	EVALUERINGSSYSTEM.....	29
6.4	SKOLEKONTAKTSYSTEM	29
6.5	ADMINISTRASJONSSYSTEM	31
7	KONTAKT MED SKOLENE	33
7.1	INFORMASJON PR BREV	33
7.2	ELEVREGISTRERING	34
7.3	KONTAKTPERSONER VED SKOLENE	34
7.4	ORGANISERING OG SKOLENS INFRASTRUKTUR.....	35
7.5	OMREGISTRERING AV ELEVER	35
7.6	UTHENTING AV KARAKTERER	36

7.7	LEVERING AV KARAKTERER TIL SKOLEPORTEN	36
7.8	ELEVENS GJENNOMFØRING	36
7.9	ELEVENS REAKSJONER	37
7.10	NEGATIVE REAKSJONER FRA TESTPERIODEN	37
7.11	POSITIVE REAKSJONER FRA TESTPERIODEN	38
7.12	ALDERSAVHENGIGE REAKSJONER FRA TESTPERIODEN	38
7.13	SKOLEREAKSJONER.....	39
7.14	REGLER OG RETNINGSLINJER - JUKSING PÅ NASJONALE PRØVER	40
7.15	IKT OG NYE MEDIER - BRO ELLER HINDER.....	40
7.16	ANBEFALINGER	41
8	TEKNISKE ERFARINGER VED SKOLENE.....	42
8.1	DEN TEKNOLOGISKE KLIENTLØSNINGEN FOR SKOLENE	42
8.2	GENERELT INNTRYKK AV UTSTYR VED SKOLENE	43
8.3	INNTRYKK AV TEKNISK KOMPETANSE VED SKOLENE	44
8.4	OPERATIVSYSTEMER I SKOLENE.....	45
8.4.1	<i>Nettlesere generelt.....</i>	<i>46</i>
8.5	OPPSUMMERING AV KONFIGURASJONER.....	47
8.6	ORGANISERING AV BRUKERSTØTTEGRUPPEN	48
8.7	DRIFTING AV DEN SENTRALE BITE ^{IT} -SERVEREN	50
9	KLIENTARKITEKTUR	51
9.1	TEKNISK LØSNING	51
9.1.1	<i>Formatutvikling.....</i>	<i>51</i>
9.1.2	<i>Kommunikasjon med tjener</i>	<i>52</i>
9.1.3	<i>Produksjon.....</i>	<i>52</i>
9.2	SIKKERHET.....	53
9.3	CD-ROM.....	53
9.4	ANBEFALINGER	53
10	GLOSSARY	54
11	APPENDIX	57

1 SAMMENDRAG

På oppdrag fra Undervisnings og forskningsdepartementet (UFD) og Læringscenteret ble de elektronisk baserte nasjonale prøvene i engelsk lesing utviklet ved Universitetet i Bergen. Prøvene ble utviklet som webbaserte prøver der elevene logget seg på fra datamaskiner på skolene. Elevene gjennomgikk to prøver – en forprøve og en hovedprøve – og fikk til slutt et resultat som var målt opp mot de europeiske målene for språkbeherskelse – CEF (Common European Framework)¹.

Prøvene var en stor utfordring for prosjektet og hver enkelt skole. Lignende prøver har aldri vært gjennomført nasjonalt i Norge. Flere skoler fikk problemer av teknisk/administrativ art som gjorde at gjennomføringen av prøvene ikke gikk helt feilfritt for seg. Andre skoler fikk problemer på grunn av feil på den sentrale serveren prøvene ble avholdt fra. Totalt logget 62137 elever seg på i en periode på 8 dager. Av disse var det 43412 som fullførte feilfritt (uten tekniske/administrative problemer).

Tilbakemeldingene før og etter gjennomføringen av de nasjonale prøvene i engelsk lesing var at gjennomføringen av prøvene var positiv i den forstand at den ikke krevde ekstra innsats fra skolene i form av retting eller administrasjon av resultater. På den negative siden var det enkelte skoler som opplevde store tekniske problemer. Selv om brukerstøtten løste de fleste av disse problemene, var det ikke alle skoler som fikk sine problemer løst og dermed gjennomført prøvene på en tilfredsstillende måte.

Denne rapporten beskriver i detalj det arbeidet som ble gjort i forbindelse med forberedelse og gjennomføring av de nasjonale prøvene i engelsk lesing. I tillegg er det kapitler som beskriver de tilbakemeldingene som ble gitt fra skolene og en egen seksjon som omhandler IKT-ferdighetene ved skolene.

¹ Mer informasjon om CEF: <http://www.alte.org/index.cfm>, se også kapittel 10 Glossary.

2 PROSJEKTGRUNNLAG FOR BITE^{IT}

Universitetet i Bergen, InterMedia fikk i oppdrag å utvikle de IKT-baserte nasjonale prøvene i engelsk i lesing 2004. Oppdragsgiver var UFD (Utdannings- og forskningsdepartementet), hvis rolle senere ble overtatt av Læringssenteret. Prosjektet ble kalt BITE^{IT}, som er et akronym for Bergen Interaktive Tester i Engelsk, Informasjonsteknologi. Prøvene skulle gjennomføres nasjonalt, ved at elever besvarte oppgaver på datamaskin tilkoblet internett og ble evaluert maskinelt. Resultatene på prøvene skulle måles opp mot felles europeiske mål, operasjonalisert i CEF (Common European Framework).

De nasjonale prøvene er et ledd i kvalitetsreformen i skolen. Prøveresultatene skal være et middel for elevene til å plassere sine språklige ferdigheter i forhold til de europeiske målene. Skoler, foreldre og myndigheter skal få oversikt over hvordan deres elever og skoler relaterer seg til de europeiske nivåene og andre norske skoler.

Nasjonale prøver avholdes på alle skoler i Norge over et relativt kort tidsrom, slik at alle elever får den samme prøven. På denne måten skal det i prinsippet la seg gjøre å sammenligne resultater mellom ulike skoler og regioner. Det korte tidsrommet prøvene avholdes i hindrer at oppgavene blir allment kjent og at skoler som gjennomfører prøvene sent vil ha fordeler.

Tradisjonelle prøver er ofte dannet for middeleven. Det vil si at dyktige elever får for lette oppgaver mens svake elever får vanskelige oppgaver. Dette kan være demotiverende for begge grupper. De nasjonale prøvene skal være *adaptive*, hvilket vil si at prøveinnholdet skal tilpasse seg ferdighetsnivået til den enkelte elev. En dyktig elev skal basert på denne modellen få mer utfordrende oppgaver, mens en svakere elev vil få mindre vanskelige oppgaver.

To grupper ved Universitetet i Bergen har deltatt i arbeidet med prøvene. En gruppe språk- og testekspert har jobbet med det faglige innholdet i prøvene. En annen gruppe med IKT-eksperter – testutviklere – har jobbet med selve utformingen av prøvene og et system for administrasjon av disse. IKT-delen av prosjektet ble kalt BITE^{IT}. Denne rapporten beskriver arbeidet som er gjort i BITE^{IT}-prosjektet og de erfaringene som er gjort under dette arbeidet.

BITE^{IT}-prosjektet ønsker å takke alle som har bidratt i prosessen med å utvikle de nasjonale leseprøvene i engelsk. En spesiell takk går til:

Utdannings- og forskningdepartementet

Læringssenteret

IT-avdelingen ved Universitetet i Bergen

Alle skoler, lærere og IKT-ansvarlige som har deltatt, spesielt skolene i Bergensområdet

SkoleLinux-prosjektet ved Frode Stiansen

InterMedia, Aksis

Dr Glenn Fulcher

Dr Felianka Kaftandjjeva

3 OPPSUMMERING AV PROSJEKTET

BITE^{IT}-prosjektets oppdrag var å utvikle IKT-baserte nasjonale prøver som skulle kunne gjennomføres over internett på skolene for alle elever i 10. klassetrinn (ca 65.000). Prøvene skulle tilpasse seg hver enkelt elev, ved at sterke elever fikk mer utfordrende oppgaver mens svake elever fikk enklere oppgaver. Elevenes resultater skulle gis i form av et ferdighetsnivå relatert til CEF-skalaen. BITE^{IT}-prosjektet ble inndelt i tre hovedfaser for å løse de utfordringene som oppdraget gav. Dette kapitlet beskriver disse utfordringene og de fasene som ble gjennomført for å løse dem. Senere kapitler vil gå mer i detalj om hvordan arbeidet ble utført i de ulike fasene.

IKT-baserte nasjonale prøver har aldri vært gjennomført før i Norge, noe som har stilt store krav til arbeidsprosessen for å utvikle prøver som har latt seg gjennomføre på skolene. Prosjektet har stått ovenfor en rekke utfordringer av IKT-messig art for å oppnå oppdragets mål:

- Norges grunnskoles maskinpark er ikke homogen, hvilket vil si at en opererer med flere forskjellige maskinkonfigurasjoner, operativsystem og nettlesere. Denne utfordringen setter store krav til utformingen av prøver hvor resultatene ideelt sett skal være sammenlignbare mellom skolene.
- Når prøvene skulle holdes ville ca 65.000 elever aksessere prøvene i en kort periode. Dette stilte strenge krav til hvordan leveringssystemet skulle oppføre seg under høyt press fra mange samtidige brukere. Dersom systemet ville reagere sent – eller i verste fall gå ned - ville det kunne påvirke resultatene hos de elevene dette rammet.
- Elever og lærere har hatt krav på god oppfølging når prøvene skulle gjennomføres. En har derfor måttet bygge opp en kompetanse på hvilke problemer som kunne oppstå på skolene og løse disse i forkant. De problemene en ikke har kunnet løse på forhånd måtte en kunne løse ved hjelp av en effektiv brukerstøtte.
- IKT-ferdighetene hos lærere og administrativ stab ved skolene varierer betydelig. Gjennomføringen av nettbaserte prøver har vært en uvant situasjon for mange og tvunget en del til å operere systemer de ellers sjelden forholder seg til.

Prosjektet ble tidsinndelt og oppgaveinndelt for å løse disse utfordringene. Utviklingen av prøvene kan grovt deles inn i tre hovedfaser:

1. *Prøveutforming.* Ved starten av prosjektet var det stort fokus på selve prøveutformingen – hvordan prøvene skulle utformes for å redusere den mulige forstyrrende effekten av at de måtte løses på maskin.
2. *Plattformtilpasning.* Etter hvert gikk fokuset over til hvordan prøvene fungerte på en heterogen maskinpark.
3. *Skalering og administrasjon.* Siste steg var å fininnstille leveringssystemet slik at det kunne håndtere de store mengdene samtidige brukere og å bygge et administrasjonssystem som ville fungere som støtteverktøy for brukerstøtten under gjennomføringen av de nasjonale prøvene.

En utfordring ved IKT-baserte prøver er å skape et grensesnitt som lett lar seg forstå av brukerne, slik at en reduserer IKT-effekten mest mulig. En IKT-effekt oppstår når elever besvarer oppgaver annerledes på datamaskin enn de ville gjort i et annet medium, for eksempel på papir. Dersom effekten er negativ fordi mediet er nytt og annerledes må det tas hensyn til. For å undersøke dette ble det gjennomført en rekke pilotutprøvinger i InterMedias datalaber. I denne prosessen ble skoleklasser invitert til å besøke datalabene og gjennomføre pilotutgaver av prøvene. Elevene ble observert under oppgaveløsingen, og prøveformatene ble finpusset under denne prosessen. Prosessen ble gjennomført sen vår og tidlig høst 2003 og omfattet observasjon av ca 300 elever.

Etter at prøveformatene var etablert og utprøvd på elever i datalab, var neste steg å gjennomføre pilotprøver på skolene. Det var hovedsaklig tre grunner til å gjøre dette.

- Nettbaserte prøver vil oppføre seg noe ulikt på ulike system og konfigurasjoner. Dette måtte testes grundig, slik at forskjellen i oppførsel mellom skolene ville være minimal.
- Videre var det et mål å undersøke prøveutdelingssystemets evne til å skalere over et større antall samtidige brukere.
- Endelig ønsket prosjektet å skaffe seg erfaringer i hvilke tekniske og infrastrukturelle problemer som ville oppstå på skolene når de skulle kjøre prøvene, slik at hjelpemateriale og brukerstøtte kunne forberedes i god tid før gjennomføringen av de nasjonale prøvene.

Tre til fire klasser ble invitert til å delta fra et tilsvarende antall skoler ved hver utprøving i denne fasen. Elevene logget seg på til samme tid for å oppnå flest mulig samtidige brukere. Det ble gjort en del erfaringer som var nyttige for senere brukerstøtteaktivitet. Dette var erfaringer med hvordan skolenes maskiner håndterte programvaren og hva nettverkshastigheten, maskinkonfigurasjon, nettleser og operativsystem hadde å si på den tekniske gjennomføringen ved skolene.

Et mål med gjennomføringen av de nasjonale prøvene var å kunne gi elevene oppgaver som var tilpasset den enkelte elevens nivå. I tillegg skulle prøvene nivåplassere elevenes ferdigheter i henhold til CEF-skalaen. Denne skalaen består av kvalitative beskrivelser av hva en person skal kunne bruke det engelske språket til. Nivåene er inndelt i tre hovedgrupper:

A: Beginner

B: Intermediate

C: Native speaker

For å kunne bestemme hvilket CEF-nivå hver enkelt oppgave målte elevene til, ble pilotprøveresultater fra en større mengde elever beregnet og kalibrert opp mot CEF-skalaen. Dette ble gjennomført med to store datainnsamlinger – kalibreringsprøver, én før jul i 2003 og en rett over nyttår 2004. Under disse kalibreringsprøvene deltok ca 11.000 elever. Man testet her utelukkende om oppgavene hadde riktig vanskelighetsgrad i forhold til CEF-nivåene.

Etter at kalibreringsprøvene var gjennomført gikk 12 språkesperter gjennom alle prøvene og registrerte et CEF-nivå på hver oppgave/hvert spørsmål. Ekspertene var lærere, språkforskere, språkstere og personer med bakgrunn i andrespråklæring. Disse estimerte

hver oppgaves vanskelighetsgrad gjennom kvalitative vurderinger. Basert på resultatene fra kalibreringstestene og evalueringene av oppgavene, var det mulig å foreta statistiske beregninger som valgte ut hvilke oppgaver som skulle inngå i prøvesettene som senere skulle presenteres for elevene i de nasjonale prøvene. Dette resulterte i tre forprøve-sett og tre hovedprøve-sett for 10. klassetrinn.

I uke 16 og 17 ble de nasjonale prøvene i engelsk lesing gjennomført på landsbasis. Elever fra hele landet logget seg på og gjennomførte de nettbaserte prøvene. Etter at prøven var gjennomført fikk hver elev tilbakemelding om hvilket CEF-nivå de befant seg på.

Under gjennomføringen av de nasjonale prøvene ble det samlet inn resultater fra over 60.000 elever.

Noen uheldige erfaringer ble gjort når det gjelder utdelingen av riktig antall oppgaver til hver elev. En kombinasjon av høy belastning på server og et høyt antall ulike konfigurasjoner av maskinpark på elevsiden førte til at noen elever fikk tildelt uriktig antall og/eller “hoppet over” noen oppgaver. Dette er utelukkende et teknisk problem, som vil la seg løse for fremtidige gjennomføringer av IKT-baserte nasjonale prøver.

Hovedinntrykket etter gjennomførte nasjonale prøver var at det teknisk sett var mulig å gjennomføre disse prøvene på stort sett alle maskiner i den norske skolen med dagens standard. Det store mangfoldet av konfigurasjoner og de mange eldre maskinene førte imidlertid til uforutsette problemer som hadde vært lettere å kontrollere i en nyere og mer homogen maskinpark. IKT-ferdighetene på skolene var høyst varierende. Bedring av den generelle IKT-forståelsen hos skolens personale vil være viktig for suksessen av fremtidige gjennomføringer av tilsvarende nasjonale IKT-baserte prosjekt i skolen.

De neste kapitlene vil ta for seg i nærmere detalj hva som ble gjort for å utforme prøvene, hvilke resultater som ble registrert i de nasjonale prøvene og hvilke reaksjoner en har hatt fra skolene og elevene.

4 TALLENES TALE

4.1 Vurdering av elevene

Engelskkunnskapene til elevene som gjennomførte de nasjonale prøvene ble elektronisk evaluert. Det vil si at systemet de løste oppgavene i automatisk vurderte elevenes ferdigheter i engelsk lesing². Umiddelbart etter at en elev hadde gjennomgått en full prøve, ville dens resultater bli presentert på skjermen. Samtidig ble resultatene tilgjengelig for skolens kontaktperson.

De nasjonale prøvene skulle implementeres slik at de oppførte seg adaptivt. Hver elev skulle få tildelt oppgaver som var tilpasset sitt nivå. For å gjennomføre dette ble det utviklet en algoritme som ville nivåplassere eleven og basert på dette tildele oppgaver som var elevtilpasset. Modellen som ble valgt bestod av følgende elementer:

- Forprøve – et oppgavesett bestående av ca 20 korte oppgaver som målte elevens nivå for å finne riktig nivå på hovedprøven
- Hovedprøve – et oppgavesett som inneholdt ca 35 lengre oppgaver som skulle måle elevens faktiske nivå relatert til CEF
- Utvelgelsesalgoritme – algoritme som ville velge ut en hovedprøve på riktig nivå basert på elevens besvarelser i forprøven.

Vurderingsmodellen krever at hver elev først gjennomfører en forprøve. Hvert oppgavesvar i denne forprøven blir vektet etter en på forhånd fastsatt vektskala. Basert på summen av de vektete resultatene fra forprøven, blir en hovedprøve valgt. For de ulike elevene kunne hovedprøven være av ulik vanskelighetsgrad. Dersom eleven hadde svake besvarelser på forprøven, ville en enklere hovedprøve blitt valgt enn dersom eleven hadde gjort det sterkt. Hovedprøvene var inndelt i tre ulike vanskelighetsgrader. Når eleven hadde besvart alle oppgavene i hovedprøven, ble hans/hennes karakternivå bestemt basert på en vektning av alle svarene fra hovedprøven. Summen av disse vektete resultatene bestemte hvilket karakternivå eleven fikk. Elevens karakternivå ble så rapportert til eleven og skolens kontaktperson som et nivå i CEF-skalaen.

For å bestemme hvilken vekt de enkelte oppgavene i for- og hovedprøven skulle ha var det på forhånd utført en større kalibrering av oppgavene. Her ble samtlige oppgaver utprøvd av ca 11.000 elever fra et representativt utvalg av Norges skoler. Videre hadde hvert enkelt spørsmål i oppgavene blitt kvalitativt vurdert av et ekspertpanel bestående av lærere og spåk-/testeksperter. Dette panelet relaterte hvert spørsmål til det CEF-nivået de mente var passende. Ved å sammenkoble resultatene fra kalibreringstestene med de kvalitative vurderingene fra ekspertpanelet var det mulig å bestemme hver enkelt oppgaves vekt og hvilket klassetrinn de var egnet for.

² (jfr Personopplysningsloven §22)

4.2 Karakterer og resultater

Til leseprøvene i Engelsk 2004 ble det registrert 62189 elever, og av dem var det kun 52 som ikke logget seg inn i det hele tatt. Tar en bort de 52, sitter en igjen med 62137 elever som logget seg inn under utprøvingen. Frafallet fra forprøve til hovedprøve kom totalt på 9823 elever. Årsaker som kan føre til et slikt frafall kan være mange, de mest sannsynlige skisseres opp under.

- Dersom prøvene ble boikottet av elever på forskjellige skoler, vil disse komme inn under gruppen av de som ikke kom videre til hovedprøven. Disse har mest sannsynlig logget seg inn og ikke gjort noe mer – eventuelt gjort kun få oppgaver fra forprøven.
- Tekniske problemer hos skolene kan forårsake at elever ikke kommer gjennom oppgavesettet på grunn av årsaker som trege linjer, linjebrydd og liknende. Mer om dette i kapittelet som omhandler skolenes datautstyr.
- Tekniske problemer ved BITE-tjeneren. I noen tilfeller erfarte vi for eksempel at enkelte oppgaver ble hoppet over. Feil som oppstår på tjeneren som får konsekvenser for elevene ble heller ikke tilfredsstillende rapportert til brukerens grensesnitt.
- I mange tilfeller vil problemet ligge i en kombinasjon av punkt to og tre, ved at tjeneren ikke reagerer tilfredsstillende på ulike feil og unntakssituasjoner. I tilfeller med trege linjer på skolene vet vi at elever i noen tilfeller har svart på for få prøver og fått for dårlig karakter. Disse er rettet opp ved at de med feil under gjennomføring ikke får utskrift av sin karakter, men en beskjed om at karakteren ikke er satt pga feil under gjennomføring.

Antall elever registrert i elevsystemet: 62189 (100 %)

Av disse er det kun 52 (0,8 %) som ikke har vært innlogget.

Antall elever fullført: 43412 (70 %)

Antall elever påbegynt hovedprøve: 52314 (84 %)

Antall som ikke fullførte hovedprøven: 8902 (14 % av alle, 17 % av antall på hovedprøve)

Disse fordeler seg slik over de tre hovedprøvene:

<i>Oppgavesett</i>	<i>%</i>	<i>Sum elever</i>
Sett 1	2%	1077
Sett 2	94%	40666
Sett 3	4%	1669
Sum	100%	43412

Tabell 1. Sett 1 er det enkleste oppgavesettet, sett 2 er middels vanskelig og sett 3 det vanskeligste.

Sum av antall som ikke fullførte: 18777

Antall som begynte på prøven uten å komme til hovedprøvene: 9875

Antall som begynte på hovedprøvene uten å fullføre: 8902

Av de som startet, men som ikke fullførte hovedprøven er fordelingen som følger:

<i>Oppgavesett</i>	<i>%</i>	<i>Sum elever</i>
Sett 1	5%	459
Sett 2	93%	8285
Sett 3	2%	158
Sum	100%	8902

Tabell 2. Fordeling av elever som startet, men ikke fullførte prøven.

Dette er gruppen av elever som har besvart ett eller flere spørsmål fra hovedprøven, men ikke alle. Av de som ikke har fullført er det 8479 som er markert som fullført i systemet, dvs at det finnes en mulighet for at så mange har fått dårligere karakter enn fortjent. Av de 43412 som har fullført er det 22 som er markert som fullført, men som ikke har fått tildelt karakter av systemet.

4.2.1 Resultater fordelt på oppgavesett

Karakterfordeling alle oppgavesettene under ett og fordelt på de individuelle.

<i>Karakter</i>	<i>Antall</i>	<i>%</i>	<i>Sett 1</i>	<i>Sett 2</i>	<i>Sett 3</i>
A2	81	0,19%	81		
A2/B1	378	0,87%	320	58	
B1	17420	40,15%	648	16772	
B1/B2	12589	29,02%	22	12565	2
B2	8819	20,33%	5	8787	27
B2/C1	1765	4,07%		1481	284
C1	2334	5,38%		978	1356
Sum	43386	100%	1076	40641	1669

Tabell 3. Karakterfordeling.

I tillegg er det 26 elever som pga teknisk feil ikke har fått tildelt karakter selv om de har gjennomført prøven tilfredsstillende. Disse er fordelt med én person på test 1, og 25 på test 2.

4.2.2 Hendelser bak tallene

Dette prosjektet har nå i første fase vært et pilotprosjekt, noe som innebærer at ting er under utprøving og at feil nødvendigvis vil skje i et så omfattende system med så mange brukere.

Hendelser som har innvirkning på den tekniske gjennomføringen kan være mange. Noen vet vi om, mens andre hendelser kan være av skjult natur, slik at sammenhengen ikke er opplagt.

Belastning av testtjeneren

Under testperioden var tjeneren under stor belastning. Dette er en mulig grunn til at så mange som 1/5 av elevene som begynte på hovedprøvene ikke fikk fullført. Dette problemet er minst tosidig. Dersom tjeneren ikke gir svar innen et visst tidspunkt, sender klienten/elevmaskinen (flashmodulen) svaret på nytt. Dette har i noen tilfeller blitt tolket av tjeneren som to svar, slik at den neste oppgaven i sekvensen har blitt hoppet over. Derfor såpass mange besvarelser med for få svar.

Denne oppførselen ble observert under perioder med høy belastning. Dersom noen fra BITE^{IT} prøvde systemet under denne belastningen fant vi ingen problemer. Rapporter fra IT-avdelingen sentralt på universitetet indikerte heller ingen feil eller for stor belastning på nettet lokalt. Dette kan indikere at skoler kan ha problemer med sin linjehastighet, ved at responstiden blir for lang pga stor trafikk lokalt hos de enkelte skolene.

Designproblemer

Et av problemene her er bruken av Macromedia Flash[®] til å lage oppgavene. Denne teknologien håndterer ikke avbrudd fra tjeneren like bra som for eksempel vanlig html i en nettleser. Det er enklere å håndtere unntakssituasjoner ved bruk av html i en nettleser enn ved bruk av Flash[®]. For eksempel må en funksjonalitet som 'Trykk ok for å prøve igjen' dersom en tilkobling ikke svarer, lages fra grunnen av i Flash[®], mens alle normale nettlesere støtter dette uten noe ekstraprogrammering.

Planlegging av gjennomføringen

Planlegging av gjennomføringen ble pålagt BITE^{IT} kort tid før selve prøven skulle startes. I ettertid ser en at grupper av skoler burde blitt pålagt å ta testene innenfor et på forhånd angitt tidsrom på for eksempel 3 dager. På denne måten kunne en kontrollert at ikke alle skolene logget seg på samtidig, og dermed unngå store toppe i belastningen på tjeneren. Trafikken fra forskjellige skoler kunne da fordeles jevnt utover en gitt periode.

Gjennomføringen foregikk på en slik måte at veldig mange skoler startet på samme dag, til nesten samme tid – noe som førte til (for) høy belastning på tjeneren.

4.3 Juks

Muligheter for juks er alltid til stede. I tjenerdelen av dette systemet vil juks være begrenset til tilfeller av uautorisert tilgang til systemet. I et pilotprosjekt er systemet hele tiden i endring, og det vil være vanskelig å garantere at det er umulig å få uautorisert tilgang til systemet. Uansett er systemet sikret med passord og roller, slik at en med studentrolle ikke kan få tilgang til menyer for lærere. Sikkerhetsevaluering av systemet er absolutt et viktig punkt for neste gjennomføring, men vanskelig å gjennomføre før systemets arkitektur nærmer seg en stabil (med hensyn på grad av endring) versjon.

I forhold til klientdelen oppstod det problemer med nysgjerrige lærere. I testperioden hendte det ofte at lærere ville forsøke seg på prøvene. Dette førte til en del merarbeid, da resultatene slike utprøvinger genererte ikke kunne være med i graderingen av de ulike testenes vanskelighetsgrad. Tilbakemeldinger viste at digitale prøver kanskje er vel så interessant og spennende for lærerne som for elevene. Problemet med lærere som genererte ut brukernavn og passord til seg selv fortsatte i de "ekte" prøvene også.

4.4 Arkitektur

Systemet for gjennomføring av nasjonale prøver i engelsk er et websystem bygd opp rundt åpne standarder. Systemet består av en tjener som leverer html-kodede dokumenter, generert utifra hvilken situasjon eleven/læreren er i.

Systemet er basert rundt flere åpne/semi-åpne standarder:

- Macromedia Flash[®] er benyttet for å tegne opp brukergrensesnittet til oppgavesettene. Flash[®] er en åpen standard utviklet av Macromedia for å lage interaktive tegninger/animasjoner.
- Kommunikasjonen fra Flash-animasjonen tilbake til tjeneren som leverer disse foregår i XML. XML i seg selv er en standard for å lage dokumenter som lett kan utveksles mellom forskjellige systemer. BITE^{IT} har utviklet sin egen dialekt tilpasset prosjektet for denne kommunikasjonen.
- Tjeneren er utviklet i programmeringsspråket Java. Java er ikke en åpen standard, men allikevel regnet som industristandard for mange typer webløsninger. Videre er tjeneren utviklet etter spesifikasjonene for Enterprise Java Beans (EJB) som er en standard for å lage tjenerkomponenter
- Alt datamateriale generert fra dette systemet lagres på en MySQL database. Denne databasen er et åpen-kildekode prosjekt fra det svenske firmaet MySQL AB.
- Selve tjenermaskinen er plassert på UiBs IT-avdeling, på deres tjenerrom. De er ansvarlige for daglig drift og backup av systemet, og har gjort en god jobb på den fronten.

5 DESIGN OG UTVIKLING

5.1 Designvalg for et brukervennlig grensesnitt

Aldersspennet i brukergruppen for de nasjonale prøvene er fra 9-åringer i 4. klasse i grunnskolen, til 17-åringer i 1. klasse på videregående skole. Ved utforming av brukergrensesnittet for de nasjonale prøvene i engelsk, etterstrebet man å arbeide i tråd med kjente HCI (Human Computer Interaction)-prinsipper, for utvikling av brukervennlige og enkle løsninger som skulle være intuitive og lette å forstå for elevene på alle de aktuelle klassetrinn.

HCI, eller menneske-maskin-samhandling, handler om å utforme datasystemer eller programmer på en slik måte at de oppleves som en støtte for mennesker som bruker dem. Dette innebærer at de som utformer programvare bør ha god menneskekunnskap generelt, og kunnskap om hvert programs målgruppe spesielt. Med kunnskap om mennesker menes for eksempel at en bør kjenne til hvordan mennesker bearbeider informasjon, og hvordan de bruker språk og kommuniserer. Kunnskap om målgruppen innebærer for eksempel å vite hvilke arbeidsoppgaver brukerne skal anvende programmet til og hvilke behov og ønsker de har i den anledning, slik at man kan utnytte tilgjengelig teknologi til å lage mest mulig brukervennlige programmer.

Eksempler på elementer i et grensesnitt som kan påvirke brukervennligheten, er brukerkontroll, navigasjon og gjennomsiktighet (*interface transparency*) (Tessmer, 1995)³. Brukergrensesnittets gjennomsiktighet dreier seg om hvor intuitivt grensesnittet virker for brukeren, og hvor godt han eller hun kan bruke programmet uten hjelp. Jo mer transparent selve grensesnittet er for brukeren, desto mer kan denne konsentrere seg om *innholdet* i programmet, og slippe å tenke så mye på hvordan programmet skal brukes.

Graden av brukerkontroll i et program dreier seg om i hvor stor grad faktorer som sekvens (rekkefølge), innhold, hastighet og vanskelighetsgrad skal kunne kontrolleres av brukeren. Dette er faktorer som kan være fastlagt på forhånd; bestemt av programmet, eller rettere sagt av programutviklerne. Hvor stor grad av kontroll brukerne skal ha, bør avgjøres med bakgrunn i brukerkarakteristika. Her vil både kompetansenivå, ambisjoner og alder kunne spille en rolle. Ofte vil programutviklerne måtte gjøre kompromisser mellom bruker- og programkontroll, for å tilfredsstille ulike brukergruppers behov.

Målgruppen for de nasjonale prøvene er som nevnt en gruppe med stor aldersspredning. Det er også en gruppe der man kan forvente stor variasjon når det gjelder generell datakompetanse. Noen elever vil ha brukt datamaskiner i mange år, mens andre har liten eller ingen slik erfaring. Denne variasjonen kan blant annet skyldes at noen elever har tilgang til eller egen datamaskin hjemme, mens andre ikke har det. Det kan også skyldes at datamaskinressursene i norske skoler varierer mye. Videre vil ulik brukererfaring naturlig nok variere i takt med at noen elever er svært interessert i data, mens andre ikke er det. Her kan også kjønnsmessige variasjoner spille inn. Det er en vanlig antagelse at gutter generelt er mer opptatt av datamaskiner og dataspill enn jenter er. Blant de yngste elevene kan man risikere at enkelte har så liten dataerfaring at de har problemer med å bruke musepeker

³ Tessmer, Martin (1995): Formative Multimedia Evaluation. *Training Research Journal*, volume 1, s. 127-149.

effektivt. Andre elever på samme alder kan imidlertid være storbrukere av datamaskiner og dataspill, og har dermed den nødvendige finmotorikken, slik at det ikke er problematisk for dem å for eksempel bruke musepeker med nødvendig presisjon. I leseprøvene er det imidlertid ingen oppgaver som krever at brukerne må skrive, det er kun lesingen som skal prøves. Erfaring i tastaturbruk var dermed noe man ikke trengte å ta hensyn til ved utforming av prøvene.

En så uensartet målgruppe krevde at man vurderte hvor stor grad av brukerkontroll som skulle implementeres i leseprøvene. Dette temaet ble diskutert med språkforskerne på forhånd. Prøvene ble lagt opp slik at brukerne ikke har noen direkte innflytelse på programsekvens og innhold, bortsett fra at elevens prestasjoner på *forprøven* bestemmer vanskelighetsgraden på oppgavene eleven får i *hovedprøven*. Programmet registrerer elevens tidsbruk, men det forelå ingen tidsfrist eller tilmålt tid for gjennomføringen av årets nasjonale leseprøver i engelsk. I det ferdige brukergrensesnittet har man lagt vekt på et "gjennomsiktig" grensesnitt med enkel navigasjon. I forprøven har hver oppgave bare ett spørsmål eller én påstand som eleven må ta stilling til. Oppgavene i hovedprøven har imidlertid ofte flere spørsmål eller deloppgaver. Brukerne kan da navigere fram og tilbake med Next- og Previous-knapper (innenfor en oppgave), og har dermed en viss mulighet til å gå tilbake for å rette opp feil. Et eget tekstfelt holder oversikt over hvor langt eleven er kommet, dvs hvilket spørsmål hun/han holder på med i forhold til totalt antall spørsmål i den aktuelle oppgaven. Men når eleven har klikket på OK-knappen og gått videre til neste oppgave, kan han/hun ikke gå tilbake til den forrige. Da er svarene avgitt og sendt til server, og kan ikke endres.

Videre har man vektlagt en mest mulig lik og gjenkjennbar design på elementer som går igjen utover i oppgavene (for eksempel fonter, ok-knapp, knapper for å navigere fram og tilbake, tekstfelt og svarfelt). Det ble tidlig i arbeidet utviklet en stilmanual for å sikre et mest mulig ensartet utseende eller brukergrensesnitt fra oppgave til oppgave. Senere ble dette ivaretatt gjennom faste stiler i et eget script. Dette øker brukervennligheten fordi det bidrar til at brukerne kun behøver å konsentrere seg om nytt oppgaveinnhold for hver gang de trykker Next-knappen, de trenger ikke bruke tid til å venne seg til et helt nytt brukergrensesnitt samtidig.

For å gjøre menneske-maskin-samhandlingen enda enklere for elevene under selve prøvene, ble skolene oppfordret til å la elevene prøve seg på demonstrasjonstester som lå tilgjengelig på internett, før de nasjonale prøvene tok til. På denne måten ville elevene kunne gjøre seg kjent med oppgaveformatene og oppgavetyperne på forhånd. Det ses på som en fordel, men ikke en nødvendighet, at elevene har fått prøve ut grensesnittet på forhånd. Også i tilknytning til framtidige leseprøver vil skolene få mulighet til å la elevene prøve et demonstrasjonsoppgavesett på internett.

5.2 Faser i utviklingsarbeidet

Arbeidet med å utvikle et brukergrensesnitt for leseprøven, samt utvikle oppgaveformatene, foregikk over en periode på ca 12 måneder (april-desember 03, pluss jan-mars 04). De ulike fasene er kort oppsummert nedenfor.

- Utvikling av brukergrensesnitt og oppgaveformater, først grafiske utkast, deretter formater med full funksjonalitet
- Produksjon av oppgaver, basert på oppgaveformatene
- Pilottesting av prøvesett med ca 300 elever, InterMedias og UiBs lokaler

- Evaluering av pilottesting og endring av formater i tråd med evalueringsresultater
- Pilottesting av prøvesett ute i skolene, Bergen kommune
- Endring av formater i tråd med evalueringsresultater
- Kalibreringstest med ca 11.000 elever
- Ny evaluering og endring av formater foretas
- Vurdering av kalibreringsresultater, samt avgrensning/utvelgelse av oppgaver i samarbeid med anerkjent statistiker
- Korrekturlesing av utvalgte oppgaver, grundig feilsøking og -feiloppretting

Oppgave --- Periode	2003										2004			
	Apr	Mai	Jun	Jul	Aug	Sep	Okt	Nov	Des	Jan	Feb	Mar	Apr	
Utvikle brukergrensesnitt	■	■												
Utvikle oppgaveformater, produksjon av oppg.	■	■	■	■	■	■	■	■	■	■	■	■	■	
Pilotteste prøvesett med elever i datalab		■	■	■										
Endre formater i tråd m. evalueringsresultater		■	■	■	■									
Pilotteste prøvesett i skolene, Bergen komm.					■	■	■	■	■					
Endre formater i tråd m. evalueringsresultater					■	■	■	■	■					
Gjennomføre kalibreringstest									■	■				
Ny evaluering og endring av formater								■	■	■	■	■	■	
Utvelgelse av oppg. basert på kalibr.res.											■	■	■	
Korrekturlesing, feilsøking											■	■	■	
Nasjonal prøve avholdes													■	

Figur 1. Gantt-diagram, design og utvikling.

5.3 Samarbeid mellom ulike faggrupper

Utviklingen av formater og oppgaver foregikk som et samarbeid mellom språkforskere ved Norsk Språktest i Bergen (etter hvert kalt BITE – Bergen interaktive tester i engelsk) og testutviklere ved InterMedia (BITE^{IT}), Universitetsforskning i Bergen (UNIFOB). Tekstbaserte utkast til oppgaver ble levert til testutviklerne på papir eller over e-post fra språkforskerne. I samråd diskuterte språkforskere og testutviklere seg deretter fram til mulige måter å "løse" oppgavetyperne på for å overføre dem fra papir til data. Språkforskerne forklarte hvilke språkferdigheter hver oppgave var ment å teste og hvilken funksjonalitet de ønsket, slik at de kunne være sikre på at intensjonen med oppgaven ble ivaretatt gjennom utviklings- og implementasjonsfasen. Testutviklerne kunne så vurdere ulike måter å implementere hver oppgavetype på, og komme med konkrete forslag til språkforskerne.

Til å begynne med ble det laget grafiske utkast uten reell interaktivitet eller funksjonalitet for å gi et inntrykk av hvordan oppgaven kunne komme til å se ut, og for å gi et diskusjonsgrunnlag mellom IT- og språkfolk. Etter hvert gikk man over til å utvikle oppgaveformatene direkte i programmet Macromedia Flash[®] og gjøre dem tilgjengelige for språkforskerne på et eget nettsted. Der kunne de enkelt gå inn og vurdere de ferdige formatene, godkjenne dem eller eventuelt foreslå endringer. Utvikling av nye og videreutvikling av eksisterende oppgaveformater pågikk gjennom hele prosjektperioden. Underveis ble det lagt vekt på å utvikle et begrenset antall oppgaveformater, som kunne brukes som basis og utgangspunkt for en større mengde oppgaver slik at disse lettere kunne masseproduseres. Slik bygget man gradvis opp en innholdsrik *oppgavebank* basert på de

ulike hovedformatene, som var navngitt etter funksjon eller type interaktivitet. For hvert format ble det laget en rekke oppgaver der bare instruksjoner, spørsmål og evt. tegninger var byttet ut, men der grensesnittet for øvrig var det samme.

Utviklingen av grensesnitt for formatene, som foregikk i Flash[®], var til å begynne med hovedsaklig grafisk basert, det vil si at de ble designet (dels tegnet, dels programmert) i programmet Flash[®] sitt grafiske brukergrensesnitt. For å forenkle den ønskede og nødvendige masseproduksjonen av oppgaver gikk man imidlertid over til bare å programmere en stadig større del av oppgavene i Flash[®] ActionScript[®], som er Flashs eget programmeringsspråk.

I tillegg til språkforskere og testutviklere har også en illustratør deltatt i arbeidet med å utvikle oppgaveformater. Illustratøren produserte alle tegningene som er brukt i oppgavene, på oppdrag fra og i samarbeid med språkforskerne som utarbeidet oppgavetekstene. Han samarbeidet også en del med testutviklerne for å tilpasse tegningenes format og fargebruk m.m. slik at de kunne brukes i et webbasert grensesnitt. Illustratøren var også en viktig medspiller for testutviklerne under utviklingen av det endelige grensesnittet for Nasjonale prøver i engelsk.

5.4 Hva er et oppgaveformat?

Et oppgaveformat er en mal for en oppgavetype. Et typisk oppgaveformat består ofte av fire elementer:

- En instruksjon for oppgavetypen, som forklarer hva man skal gjøre for å løse oppgaven.
- En oppgavetekst/problembeskrivelse, for eksempel en tekst som forteller en liten historie. For lavere vanskelighetsgrader kan oppgaveteksten være byttet ut med et bilde som skal formidle et budskap.
- En påstand eller et spørsmål, relatert enten til oppgaveteksten eller bildet. Ofte må man lese oppgaveteksten/eller studere bildet én gang til etter at man har lest selve spørsmålsformuleringen.
- Et felt for oppgavebesvarelse, der eleven besvarer spørsmålet eller tar stilling til påstanden. Besvarelsen av en flervalgsoppgave skjer ved at eleven klikker i en radioknapp eller en rute for å svare på et spørsmål eller avgjøre om en påstand er rett eller gal. Andre formater har ikke et slikt felt fordi besvarelsen gjøres på annet vis, for eksempel ved å klikke direkte på et bilde eller en tekst.

Arbeidsmåten for å løse de fleste oppgavetyperne er å lese instruksjonen og oppgaveteksten for så å svare på spørsmålet. Når eleven har avgitt sitt svar vil en 'ok'-knapp dukke opp nederst til høyre i skjermvinduet. Trykker man på denne vil elevens besvarelse bli sendt til systemets server, og neste oppgave kommer opp på skjermen.

5.5 Hovedformater

Formatene kan deles inn i seks hovedformater. Under noen av disse tilkommer formater som er ulike varianter av hovedformatene.

Multiple Choice/Flervalg: Et format der eleven har flere ulike og gjensidig utelukkende svaralternativer. Eleven svarer ved å klikke på en såkalt radioknapp foran det valgte

svaralternativet, se illustrasjon under. I andre varianter avgis svar ved å klikke på en av flere tekstbokser eller ett av flere bilder.

Nasjonale prøver i engelsk

Read the text. Answer the questions.

Anna

Anna Lundy, aged 12, lives with her parents, her older sister, Catherine, and her younger sister, Sarah. Catherine is 14 while Anna has just finished kindergarten and will start school in a few months.

Anna's parents both work at the local school. She Lundy teaches maths while Mr Lundy is an English teacher. He has not been working for the last year, because of a back injury.

Question 2 of 4 Previous Next

Anna loves...

- reading and singing
- reading and reading
- singing and reading

Bilde 1. Eksempel på en flervalgsoppgave. Oppgaveteksten er duset ned for ikke å kompromittere innholdet i oppgaven.

Highlight/Marker ord: I dette formatet svarer eleven ved å markere et ord, deler av eller en hel setning i den oppgitte teksten. Dette gjøres ved å plassere markøren foran ordet eller setningen og dra musepekeren over det området eleven mener bør inngå i svaret. Å markere hele teksten er mulig, men vil ikke gi riktig svar.

Nasjonale prøver i engelsk

Mark the right word.

Anna

Anna Lundy, aged 12, lives with her parents, her older sister, Catherine, and her younger sister, Sarah. Catherine is 14 while Anna has just finished kindergarten and will start school in a few months.

Anna's parents both work at the local school. She Lundy teaches maths while Mr Lundy is an English teacher. He has not been working for the last year, because of a back injury.

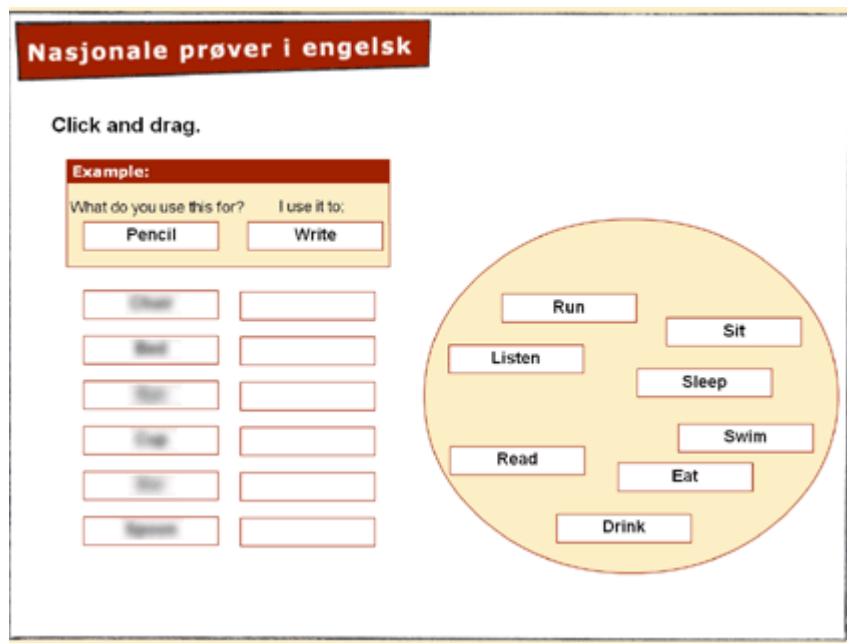
Task:

Mark the word roughly meaning ...

Your answer:

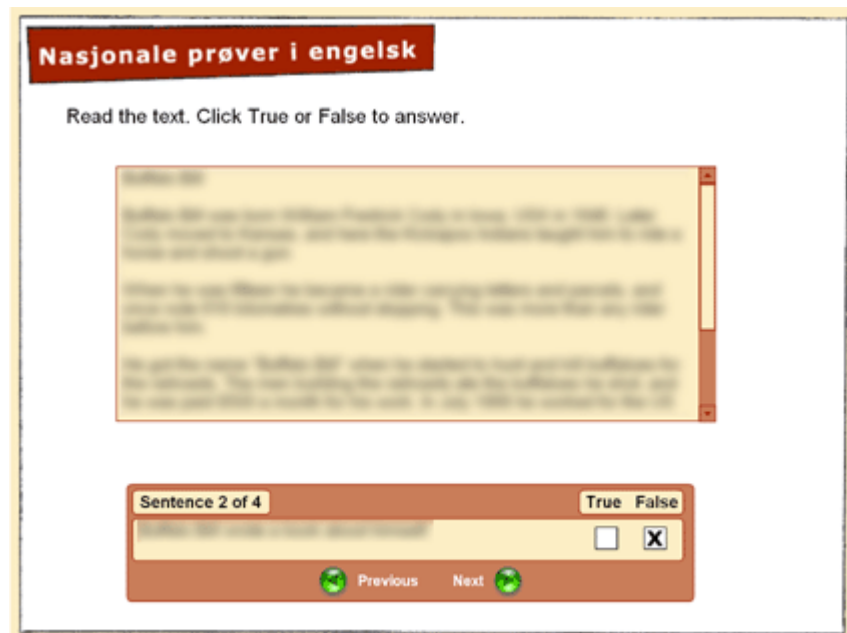
Bilde 2. Eksempel på en "marker ord"-oppgave.

Drag and drop/Dra og slipp: Her avgis svar ved å klikke og dra et avsnitt, en setning eller et ord/tekstfelt til riktig plass. Det kan for eksempel dreie seg om å matche ord med tilhørende ord, ord med tilhørende bilder, eller å sortere setninger eller avsnitt i riktig rekkefølge.



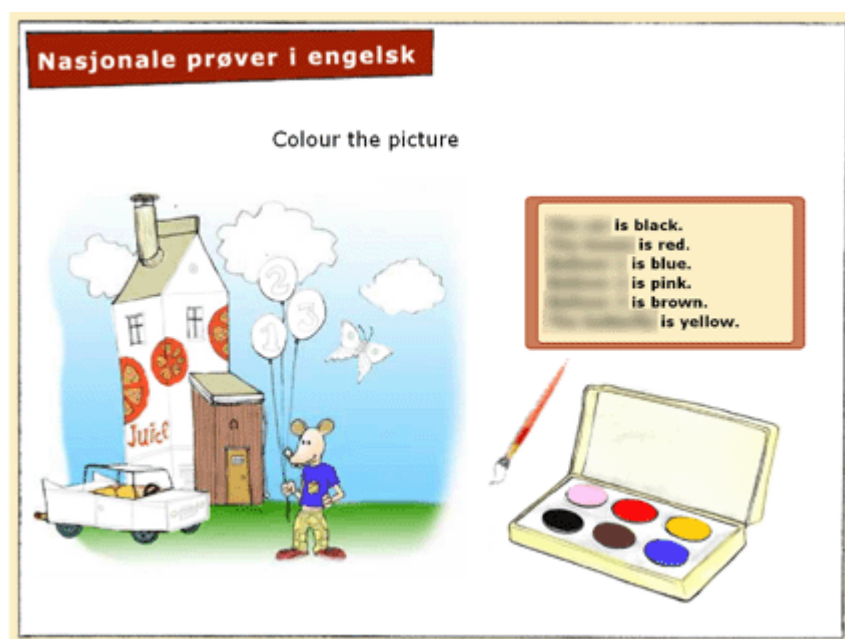
Bilde 3. Eksempel på en "dra og slipp"-oppgave. Oppgaveteksten er duset ned for ikke å kompromittere innholdet i oppgaven.

True-False/Rett-Galt: Et format der eleven får oppgitt en eller flere påstander som må tas stilling til ved å klikke i en rute for True/Rett eller False/Galt, som så markeres med et kryss. Påstandene er knyttet til en tekst som må leses først, eller til et bilde.



Bilde 4. Eksempel på en "rett-galt"-oppgave.

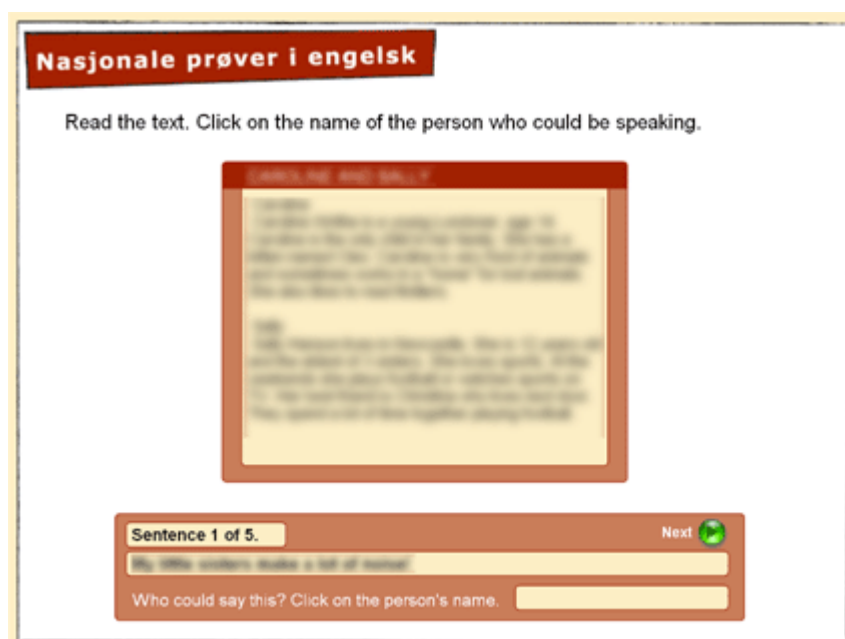
Colour/Fargelegg: Dette formatet ble kun brukt på fjerde og syvende klassetrinn.



Bilde 5. Eksempel på en fargeleggingsoppgave. Oppgaveteksten er duset ned for ikke å kompromittere innholdet i oppgaven.

Elevene møter her en tegning med noen hvite felter. Disse skal fargelegges i tråd med oppgitte instruksjoner. Et eksempel på en slik instruksjon kan være "Colour the house red." Fargeleggingen gjøres ved at eleven klikker med musepekeren, for anledningen omgjort til en malepensel, for å velge farge i et maleskrin. Deretter klikker eleven på det feltet han/hun mener er det som skal ha den aktuelle fargen.

Who could say/Klikk Navn: Dette formatet består av en tekst som omhandler en familie



Bilde 6. Eksempel på en "klikk navn"-oppgave. Oppgaveteksten er duset ned for ikke å kompromittere innholdet i oppgaven.

eller annen gruppe personer, noen ganger supplert med en tegning. Eleven må etter å ha lest teksten ta stilling til en rekke utsagn, og finne ut med bakgrunn i teksten hvem som kunne ha kommet med disse utsagnene. Svar avgis ved at eleven deretter klikker på navnet til den riktige personen i teksten. Navnet framkommer da i et eget felt som avgitt svar. Denne oppgaven krever at eleven klarer å trekke slutninger ut fra innholdet, fordi svaret ikke nødvendigvis står i klartekst. Dette kan være en uvant problemstilling for noen.

5.6 Pilotutprøvinger - formativ evaluering

I løpet av design- og utviklingsfasen av de nasjonale prøvene ble det jevnlig avholdt pilotutprøvinger med skoleelever fra Bergensområdet. Et systemutviklingsprosjekt har vanligvis stramme tids- og økonomiske rammer, og derfor er det viktig å gjennomføre evalueringer av produktet man utvikler, mens det fremdeles er mulig å foreta endringer dersom dette viser seg å være nødvendig. Slike utprøvinger underveis kalles gjerne *formativ* evaluering, i motsetning til *summativ* evaluering, som utføres etter at et produkt er ferdig utviklet. I skolesammenheng kan man for eksempel bruke en summativ evaluering for å fastsette elevenes karakterer etter endt semester, mens man kan bruke en formativ evaluering dersom man ønsker å fokusere på elevenes læringsprosess mens den enda pågår, for eventuelt å gjøre endringer eller korrigeringer underveis i et semester.

At en evaluering er *formativ*, betyr at informasjonen samles inn mens programmet enda er under utforming, og produktet fortsatt kan formes i en retning som best mulig ivaretar programmets målsetning. Resultatene fra en formativ evaluering bruker man fortløpende i den videre programutviklingen. Nedenfor følger en beskrivelse av hvordan pilotutprøvingene i regi av BITE^{IT} foregikk.

5.7 Gjennomføring

De første pilottestene ble holdt i InterMedias og Universitetet i Bergens lokaler, hvor skoleklasser ble invitert på besøk for å prøve ut det testmaterialet som til da var ferdig utviklet. Tilstede under utprøvingene var både testutviklerne, representanter for språkforskerne og til dels teknisk ansvarlige ved BITE^{IT}. Ved noen anledninger var det også representanter tilstede for andre prosjektgrupper, som jobbet med nasjonale prøver i norsk og matematikk, samt ansatte ved Læringscenteret som fulgte arbeidet med leseprøvene.

Elevene ble delt i grupper etter størrelse på de tilgjengelige datalabene, og observatørene kunne fritt gå rundt blant elevene mens de jobbet med testen. Før elevene begynte, ble de ønsket velkommen og ble forklart litt om bakgrunnen for at de var invitert til disse utprøvingene. For å fjerne litt av presset elevene kunne komme til å føle ved å skulle gjøre en engelskprøve på data mens fremmede voksne sto rundt dem og så på, ble de forklart hva observatørene var mest interessert i å finne ut av.

Testutviklerne ønsket å få inntrykk av om elevene forsto selve oppgaveformatene, det vil si om de forsto hva de skulle gjøre i de ulike oppgavetyperne. Det var også svært viktig for testutviklerne å få inntrykk av hvilken datakompetanse skoleelever på ulike klassetrinn hadde. For eksempel trengte man å finne ut om 4. klassinger uten problemer kunne bruke musepekeren til å klikke og dra objekter med og om de forsto ”dra-og-slipp”-konseptet som brukes i mange oppgaver på dette trinnet. Det var også interessant å observere om elever på de ulike klassetrinnene forsto at de i enkelte oppgaver måtte bruke rullefelt/scrollte teksten for å få lest hele, på grunn av begrenset plass i skjermbildet. Disse observasjonene var svært viktige for å få bekreftet eller avkreftet om de oppfatninger man på forhånd hadde om

målgruppens datakompetanse, stemte. Dersom det hadde vist seg at mange 4. klassinger hadde vanskeligheter med presis bruk av musepeker for å dra og flytte på objekter, kunne man ikke benyttet dette formatet i leseprøvene, i hvert fall ikke til den aldersgruppen.

Også for språkforskerne var det nyttig og viktig å få et inntrykk av om elevene forsto oppgaveformatene, men det var også interessant å få en vurdering av om oppgavetekstene holdt en passende vanskelighetsgrad i forhold til aldersgruppen. Imidlertid ble det overfor elevene lagt vekt på at observatørene på dette tidspunktet ikke var ute etter å vurdere hvor flink den enkelte elev var i engelsk.

Mens elevene jobbet seg gjennom prøvene, gikk observatørene rundt og noterte seg eventuelle åpenbare problemer, på skjemaer som var utarbeidet for dette formålet. I tillegg ble elevene etter hver utprøving oppfordret til å si sin mening om testene de nettopp hadde gjennomgått. Etter at utprøvingen var ferdig, samlet observatørene seg for å gå gjennom notater og inntrykk. På disse møtene kom man raskt fram til konkrete endringer som måtte gjøres i oppgaveformatene, og det framkom forslag til nye oppgaveformater som kunne utvikles og prøves ut.

5.8 Endringer som resultat av utprøvingene

Eget instruksjonsskjerm bilde

Mye nyttig informasjon kom fram gjennom disse utprøvingene, og observasjoner som ble gjort underveis førte også til endringer i designet. Det ble blant annet klart at elevene relativt ofte hoppet over overskrifter og instruksjoner, og gikk rett i gang med oppgaven. Dersom de ikke klarte å ”knekke koden” og løse oppgaven basert på egne erfaringer og forståelse, rakte noen opp handa og ba om hjelp, mens andre bare klikket seg videre i oppgaven uten å avgi svar. Dette skjedde til tross for at de før de begynte på prøvene fikk beskjed om å lese instruksjonene nøye.

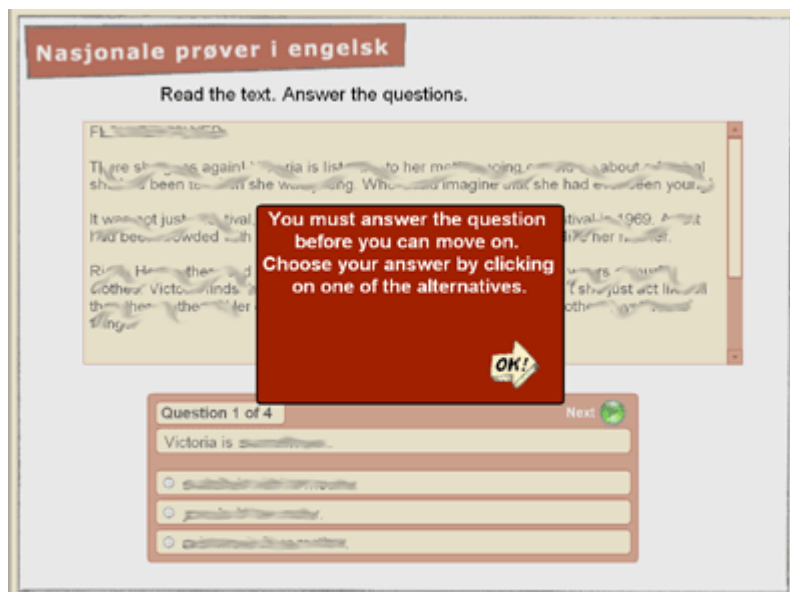
Ettersom elevene ikke skal få hjelp fra lærerne når de nasjonale prøvene avvikles, forsøkte man å finne en løsning som kunne øke sannsynligheten for at elevene leser instruksjonene før de prøver å løse oppgavene. Løsningen ble å innføre et ekstra skjerm bilde før hver enkelt oppgave i testen, som kun inneholder selve instruksjonen til oppgaven, for eksempel ”Read the text. Click True or False to answer.”. Skjerm bildet har en noe transparent farge, slik at selve oppgaven vises svakt gjennom fargen. Etter tre sekunder kommer en ok-knapp opp nederst til høyre i skjerm bildet, og elevene klikker på denne når de er klar til å gå i gang med oppgaven. Se illustrasjon nedenfor.



Bilde 7. Skjerm bilde med instruksjon.

”No Answer check”

I løpet av pilottestene ble det også klart at en del elever klikket seg videre gjennom oppgaver uten å svare på alle spørsmålene, enten fordi de glemte det, var usikre på hva de skulle gjøre i det aktuelle formatet eller fordi de ikke visste svaret. For å sikre at alle spørsmål ble besvart, ble det derfor lagt inn et nytt skjerm bilde i de formatene der dette var aktuelt, det vil si oppgaver med mer enn ett spørsmål. I oppgaver bestående av kun ett spørsmål var dette ikke aktuelt, fordi ok-knappen ikke dukker opp på skjermen før spørsmålet er besvart. Dette skjerm bildet (se illustrasjon nedenfor) dukker kun opp dersom eleven klikker seg videre i oppgaven (ved hjelp av Next-knappen) uten å besvare det til enhver tid aktuelle spørsmålet.



Bilde 8. Skjerm bildet elevene får opp dersom de ikke besvarer et spørsmål.

Samtidig som det opprinnelige skjerm bildet duses litt ned, kommer det opp en melding i skjerm bildet som forklarer at det ikke er mulig å gå videre til neste spørsmål før dette er besvart. Dette forhindrer at eleven taper poeng ved en forglemmelse. Men samtidig tvinger

det eleven til å gjette, dersom han eller hun ikke vet svaret, og det ble diskutert hvorvidt dette var riktig å gjøre. I noen av testene ble dette løst ved at man inkluderte en "Vet ikke"- eller "I don't know"-knapp i svarfeltet. Imidlertid ble det litt senere i testperioden besluttet at denne skulle fjernes, ut fra en antagelse om at det er mest rettferdig at alle elever som ikke vet svaret umiddelbart, tvinges til å svare/gjette. Andre endringer som ble gjort underveis, gikk blant annet på tekststørrelse, fonter og fargevalg.

Tekststørrelse

Det kan være vanskelig å forutsi hvordan skjermbildet av leseprøvene vil se ut for ulike brukere, ettersom det finnes så mange ulike operativsystemer og konfigurasjoner på skolene. Tidlig i prosjektperioden ble oppgavene laget med tekst i størrelse 11 til 14 punkter (overskriftene var større). Tilbakemeldinger fra utprøvingene på skolene viste imidlertid at dette var i minste laget. De fleste tekstene i oppgavene har derfor nå en skriftstørrelse på 16 punkter, noe som normalt bør være godt lesbart. Likevel var det en god del skoler som meldte om at de hadde problemer med å lese tekstene. For de fleste løste dette problemet seg da de fikk beskjed om å gjøre oppgavene i fullskjermmodus (klikke på F11-tasten i nettleseren). Imidlertid er dette et godt eksempel på at brukerne ikke nødvendigvis har samme opplevelse av et produkt som utviklerne har. Det understreker bare hvor viktig det er å foreta evalueringer underveis.

Fonter

Macromedia Flash[®] bruker klientens (elevmaskinens) fonter som standard på all tekst. Det betyr at en oppgave som bruker en font som ikke er installert på systemet til klienten, ikke vil kunne vise denne teksten hos klienten. Det er mulig å legge ved fontene som brukes, men dette vil gjøre filene som sendes fra tjener til klient mye større, og dermed øke kravet til båndbredde.

I utgangspunktet valgte testutviklerne å bruke fonten Verdana. Denne fonten ble valgt fordi den er spesielt utviklet for bruk på internett, og skal være lett å lese på skjermen. Under pilotprøvene ble det imidlertid oppdaget at denne fonten ikke er tilgjengelig for alle operativsystemer. Dette førte til at man måtte skifte fonttype for samtlige oppgaver. Etter hvert falt valget på "sans"-familien. Ved å bruke en *fontfamilie* i stedet for en spesifikk font, vil klientmaskinen selv velge en font fra denne familien. Testutviklerne hadde dermed ikke full kontroll på hvordan tekstene ville se ut rundt om på skolene, men sans-fontene er stort sett enkle å lese på skjerm.

Fargevalg

Når det gjelder fargevalg er det forsøkt tatt hensyn til fargeblinde, ettersom så mange som 8 % av den mannlige og 0,5 % av den kvinnelige befolkningen er fargeblinde. Man har derfor etterstrebet å unngå bruk av fargene rød og grønn samtidig fordi rød-grønn-fargeblinde har vanskeligheter med å skille de to fargene fra hverandre. Videre er bakgrunnsfargene stort sett lyse, og teksten mørk. De såkalte True/False-oppgavene hadde i et tidlig utkast teksten "True" i grønt og teksten "False" i rødt over avkrysningsboksene. Dette gikk man bort fra, og all tekst er nå enten svart eller hvit. Man har også forsøkt å ta dette med i betraktning ved fargelegging av tegningene til oppgavene, der fargene har betydning for oppgaveløsningen. For eksempel benyttes ikke fargene rød og grønn i samme malerskrin i Colour-oppgavene.

6 WEBSYSTEMENE

Ved utvikling av nasjonale prøver er det mange ulike interessenter og/eller roller tilknyttet prosessen, og disse har ofte behov for ulike typer informasjon. Det kan for eksempel dreie seg om språkekspert, testutviklere, prøveevalueringspanel, prøveadministratorer, elever og skolekontakter. For å tilfredsstille alle disse ulike interessentene valgte en å utvikle webbaserte applikasjoner. Et webbasert system har den fordel at det samler alle data på et sted og kan gjøre disse dataene tilgjengelig for et stort antall brukere uavhengig av tid, sted og rolle. I prosjektet ble det utviklet fem ulike webbaserte systemer: Design-, implementasjons-, evaluerings-, elev- og skolekontaktsystem. Dette kapitlet beskriver hvert av disse systemene.

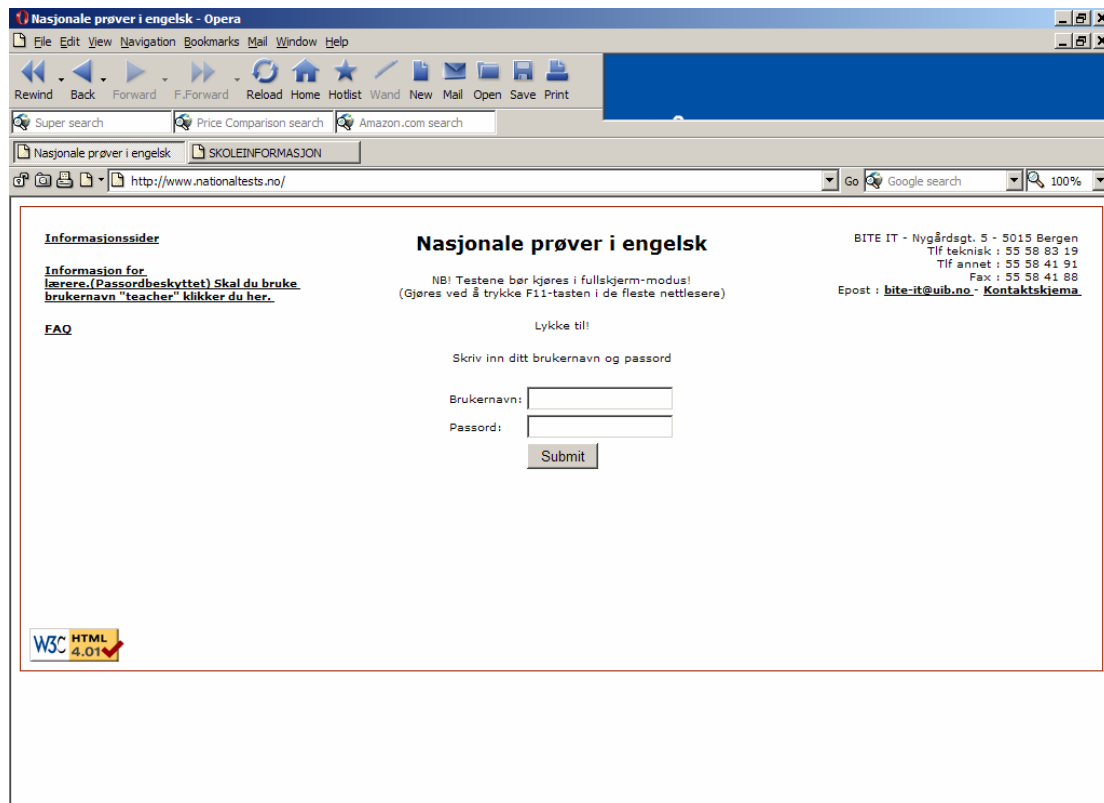
6.1 *Elevsystem*

Elevsystemet er det systemet elevene brukte da de skulle gjennomføre de nasjonale prøvene i engelsk. Systemet håndterte pålogging, utdeling av riktige oppgaver, registrering av svar, beregning av resultater og presentasjon av karakternivå for eleven.

Elevsystemet ble bygget med tanke på at det skulle være minst mulig forstyrrende for elevene under besvarelsen av oppgavene. Systemet var derfor ikke synlig for elevene på annen måte enn at de måtte skrive inn sitt brukernavn og passord for å logge seg på systemet. Se illustrasjon nedenfor – bilde 9.

I det eleven logget seg på ved hjelp av brukernavn og passord, ville systemet velge ut hvilket oppgavesett som skulle besvares basert på elevens ferdigheter. Når alle oppgaver i forprøve- og hovedprøvesettet var besvart, ble elevens nivå presentert umiddelbart og en ble informert om at prøven var over.

Dersom noe gikk galt under gjennomføringen av prøven, ville eleven bli styrt ut til innloggingssiden og måtte logge seg på på nytt. Hvis noen av oppgavene var besvart før, ville en ikke måtte besvare disse på nytt.



Bilde 9. Viser elevenes påloggingsside. Når eleven hadde skrevet inn brukernavn og passord ble riktig oppgave valgt for eleven og prøven ble startet. Dersom noe galt skjedde under prøven, ville eleven automatisk bli styrt tilbake til denne påloggingssiden for ny pålogging.

Utfordringen for elevsystemet var todelt. For det første skulle responstiden være lav slik at ventetiden mellom hver oppgave skulle oppleves som akseptabel for elevene. Videre skulle systemet i minst mulig grad påvirke elevens gjennomføring av oppgavene.

Under gjennomføringen av de nasjonale prøvene var det i travle tidsrom over 200 treff per sekund mot serveren fra elevsystemet. Hovedsaklig ble det ikke registrert responstid fra server på over 2 sekunder, de fleste responsene var på noen få millisekunder. Elever ved skoler med eldre utstyr kunne likevel ofte oppleve systemet som tregt fordi skolenettverk og maskiner ikke håndterte prøvene like godt.

6.2 Oppgavebank

Oppgavebanken er prøvesystemets kjerne. Den består av store mengder *oppgaver*, som senere ble satt sammen i ulike *oppgavesett* som skulle presenteres for elevene. En oppgave består av ett eller flere spørsmål/påstander. Et oppgavesett er en samling av oppgaver.

Filosofien bak en oppgavebank er at dens verdi vil bli større jo større banken blir. Ved et tilstrekkelig stort antall oppgaver i en oppgavebank vil det være mulig å gjenbruke store deler av banken hvert år ved gjennomføringen av de nasjonale prøvene, uten at bankens innhold blir vesentlig kompromittert. Ulike elever vil få ulike oppgaver med individtilpasset vanskelighetsgrad. Det vil derfor ikke være så lett å avsløre prøvens innhold i etterkant. På denne måten vil oppgaver som har blitt brukt ved tidligere års nasjonale prøver kunne brukes i nye nasjonale prøver senere.

Utviklingen av oppgavebanken ble foretatt av et team språkekspertter og testutviklere. Språkekspertene var ansvarlig for innhold, og var i tillegg ansvarlig for presentasjonsformatene i samarbeid med testutviklerne. Et designsystem ble utviklet for å støtte samarbeidsprosessen mellom disse to gruppene. Systemet ble utviklet som et websystem, der språkforskerne kunne legge inn tekster/bilder og spørsmål til disse i et webgrensesnitt. Basert på denne databasen over tekster og oppgaver, kunne testutviklerne implementere ferdige prøver. Systemet var utviklet med tanke på at man skulle få en god oversikt over de oppgavene som var blitt utviklet, og at en lettere skulle kunne formidle status på oppgaver i oppgavebanken mellom de ulike gruppene.

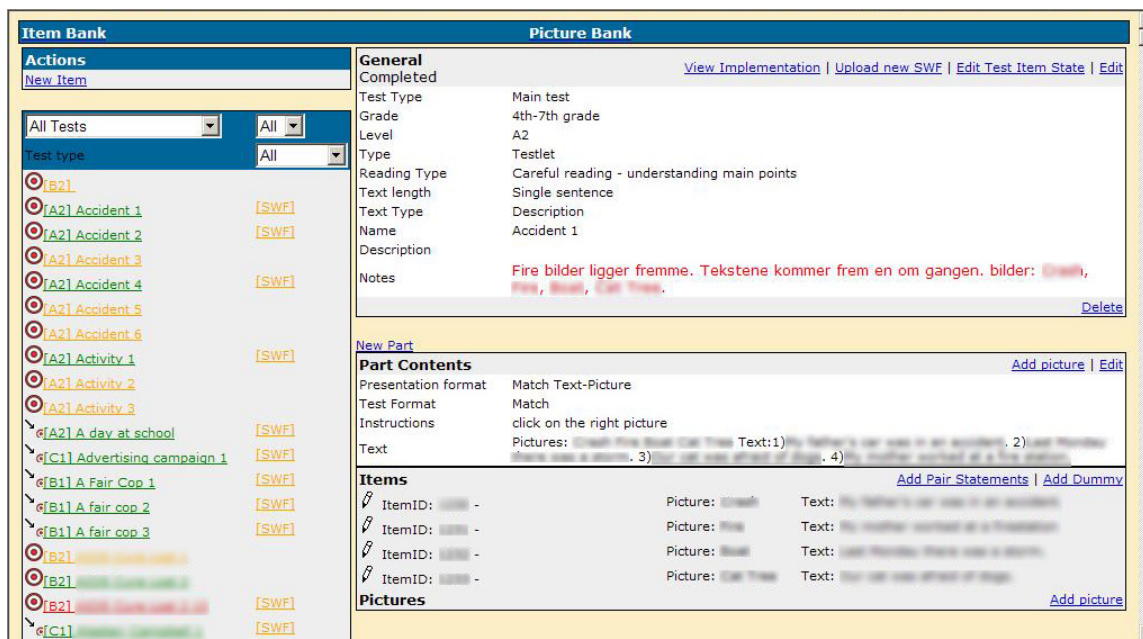
Designsystemet ble utviklet basert på de krav språkekspertene hadde til metadata og prøveformat.

Testutviklerne implementerte de fysiske oppgavene i Macromedia Flash[®]. De ble utformet basert på de spesifikasjonene språkekspertene hadde lagt inn i oppgavebanken. Når en oppgave var ferdigutviklet, ble den eksekverbare prøven lastet opp i oppgavebanken og gjort tilgjengelig for språkforskernes kommentarer.

For å klassifisere de ulike oppgavene tilbød websystemet muligheten til å legge inn metadata til hver oppgave. Her kunne en legge inn data om testens

- navn – navn som beskrev oppgavens innhold
- beskrivelse – beskrivelser av hva oppgaven omhandlet
- type – om oppgaven skulle delta i en hovedprøve eller om den var tenkt som forprøve
- klasstrinn – hvilket klasstrinn oppgaven var tiltenkt
- nivå – hvilket ferdighetsnivå oppgaven var antatt å ha relatert til CEF-nivåene
- oppgavefokus – hvilken type leseferdighet oppgaven skulle teste
- tekstlengde – hvor mange ord oppgaven bestod av
- teksttype – hvilken type tekst oppgaven skulle inneholde for eksempel faktaartikkel, reklametekst, brev osv
- notater – notater til forfatteren av oppgaven
- status – hvilken status oppgaven hadde i testutviklingen

Se illustrasjon nedenfor; bilde 10.



Bilde 10. Viser hvordan oppgavebanken var organisert. Venstre side lister alle oppgaver som er definert, og inneholder verktøy for å filtrere listen på ulike måter. De ulike fargene indikerer ulik status på oppgavene. Høyre side viser metadata og spesifisering av en enkelt oppgave. Øverste seksjon viser metadataene, mens nederste seksjon viser spesifisering av det faktiske prøveinnholdet. Testutviklerne baserte sine implementasjoner på slike spesifikasjoner.

Basert på disse metadataene kunne en sette sammen oppgavesett som ville prøve elevene i alle de ulike leseferdighetene det var relevant å teste, og i tillegg kunne en lettere kontrollere at oppgavesett inneholdt god variasjon i teksttyper og -lengder.

Når en oppgave ble utviklet gikk den gjennom en firefasert prosess. Under vises en idealrekkefølge for hvordan fasene ble gjennomgått, men de ble ikke alltid foretatt i denne rekkefølgen. En kunne bevege seg fritt mellom dem dersom ønskelig.

- Design – språkeksperter la inn oppgavespesifikasjoner, tekster og spørsmål til disse i databasen
- Implementasjon – testutviklerne utviklet en ferdig, kjørbart prøve av spesifikasjonene
- Korrekturlesing – språkvasking av de ferdig utviklede prøvene
- Ferdigstilt – prøven klar for oppgavebank

Når en prøve hadde status ferdig var den klar for oppgavebanken. Det vil si at den kunne inngå i et oppgavesett som skulle presenteres for elever. Dersom prøven av en eller annen grunn fikk en ny status (ikke ferdig), ville den ikke lenger kunne delta i et oppgavesett.

For å kunne skape adaptive prøver måtte det utvikles store mengder oppgaver, hvorav mange ville bli forkastet etter kalibreringstestene⁴. Dette stilte krav til at en hele tiden måtte ha oversikt over hvor mange oppgaver en hadde designet og implementert for hvert klassetrinn til hvert CEF-nivå. I tillegg til detaljinformasjonen om hver oppgave, inneholder

⁴ En del oppgaver ble forkastet etter kalibreringstestene, fordi de ble vurdert som for dårlige av ulike grunner – eksempelvis at de ikke målte det de var ment å måle, at teksten var for vanskelig eller at den kunne feiltolkes eller misforstås.

oppgavebanken derfor oppsummerende informasjon der antall tester og hvilke nivå de tilhører blir presentert.

Pre-test	4	4-7	7	7-10	10-11
C2	0	0	0	0	0
C1	0	1	7	23	132
B2	0	10	68	285	147
B1	54	54	300	228	98
A2	163	414	155	60	28
A1	144	0	0	0	26
Total	0	0	0	0	0

Bilde 11. Viser status på oppgavebanken før de nasjonale prøvene ble gjennomført. Oversikten viser hvor mange oppgaver som var blitt utviklet for de ulike klassetrinn og til hvilke nivå oppgavene var definert.

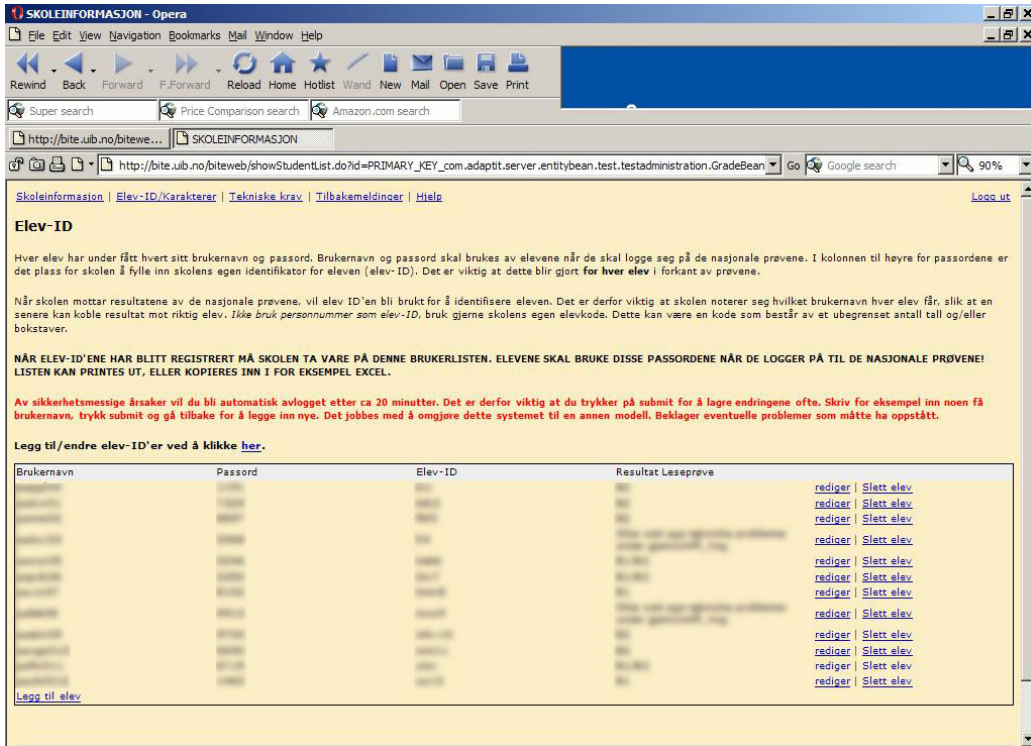
6.3 Evalueringssystem

For at språkprøvene skulle ligge så tett opp til de kvalitative målene i CEF-rammeverket som mulig, måtte hver oppgave evalueres og nivåbestemmes av språkekspertene. Da språkekspertene befant seg overalt i landet, ble det utviklet et websystem som kunne brukes til denne registreringen. Når språkekspertene logget seg på, fikk de opp en liste over hvilke oppgaver som skulle evalueres og hvilke evalueringer som hadde blitt gjort før. Listen over oppgaver som skulle evalueres kom direkte fra den ferdigstilte oppgavebanken. Språkekspertene måtte i denne prosessen se på hver individuelle oppgave og markere hvilket CEF-nivå det befant seg på. Videre hadde de muligheten til å legge inn kommentarer om de enkelte oppgavene.

6.4 Skolekontaktsystem

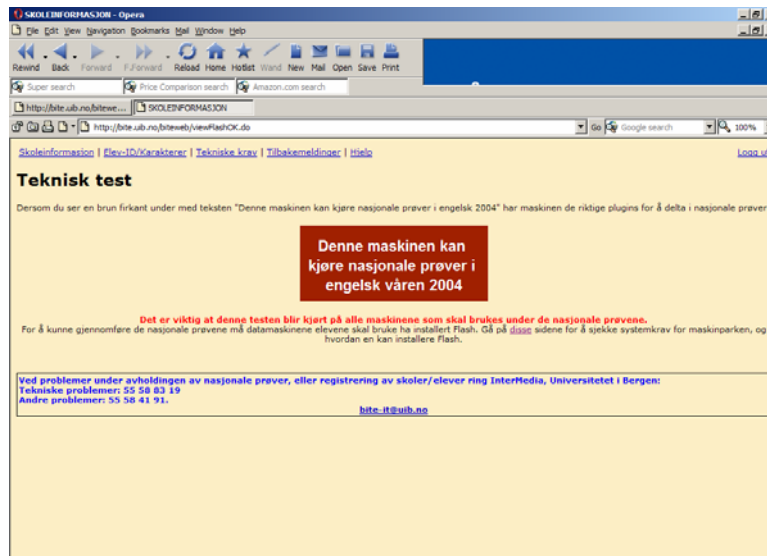
For at elevene skulle få registrert sine resultater, måtte hver elev registreres i brukerdatatabasen. Siden det var ca 65.000 elever i 10. klassekullet, ble elevregistreringen overlatt til skolene. Hver skole fikk tilsendt et brev med brukernavn og passord til det webbaserte brukerhåndteringssystemet.

Når en skoles kontaktperson logget seg på systemet kunne han registrere antall elever i hver klasse som skulle delta. Når dette ble registrert i systemet, ville det automatisk bli opprettet en bruker pr elev, og gjøre brukernavn og passord til disse tilgjengelige for skolekontakten. Denne kunne videre registrere skolens egen elev-id på hver elev, slik at de visste hvilken elev som hadde hvilket brukernavn.



Bilde 12. Viser elevregistreringssystemet. Skolens kontaktperson registrerte antall elever i hver klasse. Systemet ville da generere brukernavn og passord til hver elev i denne klassen. Skolekontakten kunne også registrere sin egen ID på hver elev dersom det var ønskelig. Kolonnen helt til høyre viser resultatene til elevene etter gjennomført prøve.

Skolekontaktsystemet kunne også brukes til å vurdere skolens maskinparkes tekniske standard. Når skolekontakten brukte skolekontaktsystemet ville han få tilbakemelding på om maskinen han brukte kunne kjøre de nasjonale prøvene eller ei. Hvis en maskin var kompatibel med de tekniske kravene ville en se en melding om dette på elevmaskinen, og dersom den ikke var kompatibel ble man instruert om hvordan en skulle gjøre den kompatibel, ved å installere de plugins som krevdes. Meldinger om elevmaskinenes tekniske standard ble registrert i databasen, slik at det ville være lett for brukerstøtten å sjekke for eventuelle tekniske problemer når en ble oppringt av de enkelte skolene.



Bilde 13. Skolekontaktsystemet kunne brukes til å sjekke om skolens tekniske utstyret tilfredstilte kravene til de nasjonale prøvene. Bildet over viser at maskinen tilfredstiller kravene.

Etter gjennomføringen av de nasjonale prøvene kunne skolens kontaktperson også hente ut resultatene for skolens elever gjennom skolekontaktsystemet.

6.5 Administrasjonssystem

For å kunne monitorere bruken av systemet for feilsøk eller uthenting av resultater fra de nasjonale prøvene, ble et administrasjonssystem utviklet. Dette systemet administrerte data for elever, skoler og prøver. Her kunne en følge med på hva som skjedde, hvilken progresjon elevene hadde ved de ulike skolene, hvilket teknisk utstyr de brukte under gjennomføringen og hvilke prøver de var blitt tildelt. Denne informasjonen var helt nødvendig for å kunne utføre en tilfredsstillende brukerstøtte til brukerne.

I administrasjonssystemet ble prøvene satt sammen og registrert opp mot riktige klassetrinn. Dette ble gjort ved at de enkelte oppgavene fra oppgavebanken ble valgt og lenket opp mot et oppgavesett. Det ble opprettet tre hovedprøvesett for hvert klassetrinn, med ulik vanskelighetsgrad. Videre ble det opprettet tre forprøvesett med lik vanskelighetsgrad for hvert klassetrinn.

Planned Tests	Active
260204Pre [481]	No
290104Main [39311]	No
290104Pre [5821]	No
aa [01]	No
bb [01]	No
Demo gammel [1401]	No
Demosett [01]	No
DO NOT USE Testsett 23 [24529]	No
NP04 10.KLASSE MAINTEST SETT 1 [55271]	No
NP04 10.KLASSE MAINTEST SETT 2 [185409]	No
NP04 10.KLASSE MAINTEST SETT 3 [76381]	No
NP04 10.KLASSE PRETEST SETT 1 [17287]	No
NP04 10.KLASSE PRETEST SETT 2 [17286]	No
NP04 10.KLASSE PRETEST SETT 3 [17108]	No
NP04 4.KLASSE MAINTEST SETT 1 [01]	No
NP04 4.KLASSE MAINTEST SETT 2 [01]	No
NP04 4.KLASSE MAINTEST SETT 3 [01]	No
NP04 4.KLASSE PRETEST SETT 1 [01]	No
NP04 4.KLASSE PRETEST SETT 2 [01]	No
NP04 4.KLASSE PRETEST SETT 3 [01]	No
NP04 7.KLASSE MAINTEST SETT 1 [211]	No

Bilde 14. Viser administrasjon av oppgavesett. Til venstre er listen over hvilke oppgavesett som er definert. Til høyre er et eksempel på ett oppgavesett som består av 10 oppgaver for 10. klassetrinn. Oppgavesettet er ikke publisert, hvilket vil si at det ikke er mulig å aksessere før det blir publisert.

Når et oppgavesett var ferdig definert, kunne en gjennom administrasjonspanelet publisere og avpublisere disse online. Hvis elever logget seg på elevsystemet og ingen oppgaver for elevens klassetrinn var publisert, ville han/hun få beskjed om at ingen prøver var tilgjengelig. Dersom et eller flere oppgavesett var publisert for det aktuelle klassetrinnet, ville eleven få tildelt et tilfeldig oppgavesett for riktig klassetrinn. Under de nasjonale prøvene var det opprettet tre ulike forprøvesett, hvilket vil si at når en elev logget seg på ville han/hun få tildelt ett av disse forprøvesettene.

7 KONTAKT MED SKOLENE

7.1 Informasjon pr brev

Nasjonale prøver i engelsk ble utført i tidsperioden 13.-23. april 2004 for 10. klasse ved landets grunnskoler. De nasjonale prøvene i matematikk, norsk og engelsk skriving skulle gjennomføres i løpet av en dag hver på landsbasis. På grunn av den teknologiske infrastrukturen ved landets skoler ble det bestemt at dette ikke var gjennomførbart for de nettbaserte leseprøvene i engelsk. Det mest hensiktsmessige ble da at skolene fikk en tidsperiode på ti dager hvor prøvene skulle gjennomføres.

Etter avtale med Læringscenteret ble det bestemt at informasjon om leseprøven til skolene skulle sendes fra InterMedia i brevform. Man mente at det ikke ville fungere å bruke e-postsystemet til dette da Læringscenteret mangler oppdaterte og fullstendige e-postadresselister til skolene, samt at e-post som kommer til skolene kan bli oppfattet som useriøse henvendelser.

Ettersom hver skole skulle tildeles et unikt brukernavn og passord måtte skolene få tilsendt et unikt brev hver. Dette ble løst ved at InterMedia fikk tilsendt en adresseliste over landets ungdomskoler fra Læringscenteret. Denne listen med brevet (se vedlegg) ble sendt til Posten Norge AS, Kommunikasjon/Mediahuset. Her ble brevet flettet med adresselisten, klargjort for utsendelse og utsendt til skolene.

I brevet ble det beskrevet hva den enkelte skole skulle gjøre for å få registrert sine elever for prøven. Dette skulle helst gjøres noen dager før selve prøven slik at passord og brukernavn til hver enkelt elev skulle bli å finne i datasystemet definert med egen ID. Passord og brukernavn måtte så deles ut til elevene rett før prøven skulle gjennomføres.

Til tross for denne informasjonen viste det seg at omtrent 300 skoler ennå ikke hadde registrert sine elever dagen før første prøvedag. Vi fikk også flere telefoner fra rektorer som ikke hadde fått brevet med skolens unike brukernavn og passord. Vi besluttet derfor å ringe til de skoler som til da ikke hadde registret elever i systemet. Det viste seg da at de fleste av disse skolene ikke kunne se å ha mottatt brevet.

Vi tok kontakt med Posten Norge AS. Produksjonsavdelingen sjekket sitt produksjonssystem og kunne finne at den 4. mars var det loggført 1208 brev for konvoluttering, ferdig kl.13.05. Dette stemte overens med det antallet skoler vi hadde på adresselisten fra Læringscenteret. Det er derfor uklart hvor feilen ligger i forhold til dette. Riktignok var det noen av skolene som stod oppført med feil adresse i Læringscenterets adresseliste, men det forklarer ikke alle de 300 skolene som ikke hadde mottatt brevet. Ved noen skoler kan brevet ha blitt arkivert uten at rektor fikk sett det. Det kan også være at brevet ikke har blitt satt i sammenheng med "Nasjonale prøver", ettersom all tidligere korrespondanse som vedrørte de nasjonale prøvene var gått gjennom Læringscenteret.

Et siste aspekt med utsendelse av brevet var i forbindelse med spesialskoler og andre fritatte skoler. Listen fra Læringscenteret var en komplett liste over ungdomsskoler. Passord og brukernavn ble generert til samtlige av disse skolene hvorpå brevet ble sendt ut. I denne listen var også de skoler som var fritatt fra å ta prøven. Under dette punktet kommer både

spesialskolene - f.eks. skoler med elever med adferdsvansker og privatskolene med annet pedagogisk innhold enn i den offentlige skolen - f. eks. steinerskolene. Andre feil i listen var institusjoner som viste seg å ikke være skoler - f.eks. Karasjok Skolestyre.

7.2 *Elevere registrering*

For å kunne gjennomføre de databaserte nasjonale prøvene i engelsk måtte hver enkelt elev få utdelt hvert sitt brukernavn og passord til å logge seg på systemet med. Dette måtte knyttes videre til den enkelte elev gjennom en unik ID slik at det ble mulig å identifisere hvilken elev som hadde mottatt hvilket brukernavn/passord.

Ettersom vi ikke hadde oversikt over hvor mange elever som skulle ta prøven på hver enkelt skole ble det bestemt at en kontaktperson ved skolene skulle ta seg av å registrere antall elever inn i datasystemet, som så ville generere ut riktig antall brukernavn og passord. Dette var også mest hensiktsmessig i forhold til identifiseringen av elevene. Av sikkerhetsmessige grunner skulle det ikke være mulig å kunne lete seg frem til brukernavn og passord i systemet gjennom elevenes for- eller etternavn. Skolene ble derfor bedt om å lage elev-ID'er som ikke tok utgangspunkt i elevenes navn i sin helhet.

For at dette skulle fungere mest mulig smertefritt fikk hver enkelt skole sitt eget unike passord og brukernavn som skulle tas i bruk av skolens kontaktperson når vedkommende registrerte elevene. I forhold til mulige komplikasjoner eller spørsmål som kunne oppstå ble det opprettet to telefonnumre ved InterMedia slik at det var mulig for skolene å ta kontakt for teknisk assistanse og problemer av administrativ art. Til tider var det stor pågang på disse telefonene, til tross for at 6 personer satt klare til å svare på spørsmål.

Det var også mulig å ta kontakt via det administrative datasystemet hvor skolekontakten skulle registrere sine elever. Slike beskjeder ble distribuert gjennom systemet til de samme 6 personene og besvart fortløpende.

7.3 *Kontaktpersoner ved skolene*

Informasjonsbrevene med brukernavn og passord til en kontaktperson var adressert til rektor ved hver skole. Imidlertid valgte skolene ulike løsninger for hvem som skulle være kontaktperson og ta seg av elevregistrering. Mange steder tok rektor på seg denne rollen selv, mens andre rektorer delegerte kontaktpersonansvaret videre til IKT-ansvarlige, klasseforstandere eller faglærere i engelsk. Ofte hadde informasjonen i brevet heller ikke nådd fram til alle som burde fått den. Postrutinene fungerer ulikt ved forskjellige skoler, og der noen rektorer sørget for at brevet ble gjort tilgjengelig for alle lærere, har andre valgt å kun informere muntlig etter hvert som spørsmålene meldte seg. Dette førte til en del tilfeller der ikke alle impliserte parter ved skolene var klar over hvem som hadde kontaktpersonansvar eller hva dette innebar, noe som skapte en del forvirring. Det kom flere telefoner til brukerstøtten med spørsmål om brukernavn og passord fra skoler der elevene viste seg å være registrert allerede. Andre steder har faglærere gått ut fra at de kun hadde ansvaret for gjennomføring av selve prøven, mens rektor har trodd at lærerne også tok seg av elevregistreringen.

Denne uklarheten i forhold til ansvarsforholdene førte også til noe forvirring i forhold til uthenting av resultater etter prøvene. Leseprøvene i engelsk er databaserte, og prøvene blir rettet automatisk. Elevenes resultater kunne hentes ut av datasystemet ved å logge seg inn med kontaktpersonens brukernavn og passord. Dette var ikke alle klar over, og disse fikk

problemer med å hente ut resultatene i tilfeller der kontaktpersonen ikke var tilstede. Dette løste seg vanligvis da de aktuelle skolene kontaktet brukerstøttetelefonen og fikk brukernavnet og passordet sitt oppgitt derfra.

Ved framtidige prøver anbefales det at rektor er kontaktperson på hver skole, da dette har vist seg å være den løsningen som har fungert best. Det viktigste er imidlertid klare ansvarsforhold og at alle impliserte parter får den nødvendige informasjonen.

7.4 Organisering og skolens infrastruktur

Svært få skoler har et helt klassesett med datamaskiner, derfor kunne sjelden en hel klasse ta prøven samtidig. Dette førte til stor kreativitet på skolene. Under pilotutprøvingen fikk noen skoler låne lokaler med maskiner hos nabobedrifter mens andre hadde lærere som tok med seg elevene hjem eller benyttet maskinene på lærerrommene. Ved enkelt skoler var gjennomføringen altså avhenging av lærernes kreativitet og oppfinnsomhet for å tilfredsstille den infrastrukturen som er nødvendig for denne type prøver. De fleste skolene har løst dette med å ta elevene puljevis med til datarommet. Prisgitt den teknologiske infrastrukturen og kunnskapen på dette området kunne prøven ta alt fra 30 - 120 minutter. Det var nødvendig at en lærer med IKT-kunnskaper var til stede store deler av tiden.

Andre steder har en ikke sett mulighet til å få gjennomført prøvene pga. sene reparatører, tyveri, ombygging, feil i strømleveranse, ekskursjoner, manglende teknologisk kompetanse ved skolene eller pga ulike tekniske problemer (se kapitlet om tekniske erfaringer ved skolene).

7.5 Omregistrering av elever

Elevenes passord og brukernavn er i datasystemet tilknyttet den skolen eleven tilhører. Det er imidlertid mulig å legge til og/eller fjerne enkeltelever fortløpende. Til tider er dette nødvendig da elever overføres/flyttes og bytter skole i løpet av den perioden prøven gjennomføres.

Situasjoner med elever som tilhører en "hjemmeskole", men av ulike grunner (læringsproblemer og lignende) er plassert på en annen/alternativ skole, skapte en del problemer. Skolen eleven opprinnelig tilhørte lot i mange tilfeller være å registrere eleven, fordi de regnet med at han/hun ble registrert på den skolen der eleven til daglig var elev. Skolen som eleven var plassert på lot også å være å registrere eleven, fordi de gikk ut fra at han/hun ble registrert på sin "hjemmeskole", og ventet på å få tilsendt brukernavn og passord til eleven derfra. Det bør derfor klargjøres overfor skolene hva som er riktig framgangsmåte i slike tilfeller.

Det oppstod også problemer med nysgjerrige lærere. I testperioden hendte det ofte at lærere ville "teste ut" prøvene. Dette førte til en del merarbeid, da resultatene slike utprøvinger genererte ikke kunne være med i graderingen av de ulike testenes vanskelighetsgrad. Tilbakemeldinger viste at digitale prøver naturlig nok er vel så interessant og spennende for lærerne som for elevene. Problemet med lærere som genererte ut brukernavn og passord til seg selv fortsatte i de "ekte" prøvene også.

7.6 Uthenting av karakterer

Skolens kontaktperson eller personer med tilgang til skolens brukernavn og passord hadde i det administrative datasystemet mulighet til å hente ut elevenes karakterer. På samme nettside som skolens kontaktperson genererte ut brukernavn og passord til de enkelte elevene, kom karakterlistene frem. Dette har fungert meget tilfredsstillende. De tilbakemeldingene som har kommet frem har vært entydig positive. Tilbakemeldinger viste at flere lærere fryktet merarbeidet med å notere, kopiere og sende inn karakterene som de enkelte elevene fikk. Reaksjonene fra lærerne var svært positive da karakterene automatisk kom frem i det administrative datasystemet, ettersom dette ikke medførte merarbeid for lærerne.

7.7 Levering av karakterer til Skoleporten

Tatt i betraktning at årets nasjonale prøver i engelsk var et pilotprosjekt, anbefales det ikke at resultatene legges ut til allmenn beskuelse på Skoleporten. Medienes sammenlikninger av skolens karakterer, som sannsynligvis blir et resultat av en slik offentliggjøring, vil ikke være basert på rettferdig sammenlikningsgrunnlag.

Årsaken til dette er de feil som har blitt oppdaget i systemet underveis, samt problemene som skolens infrastruktur har påført elevene ved gjennomføringen. Disse problemene førte til at noen elever ikke fikk riktig antall oppgaver, og karakteren ble dermed basert på feil grunnlag. Dette kan ha resultert i karakterer som ikke er reelle i forhold til elevenes faktiske kunnskapsnivå. Dette er noe som vil bli tatt høyde for i det administrative systemet. Til neste gjennomføring av nasjonale prøver vil det bli lagt vekt på å få til en rapporteringsrubrikk linket til de ulike karakterene. Det vil si at lærerne vil kunne gå inn og se om det har vært ureglementerte feil i systemet i forbindelse med den enkelte elevs karakter.

7.8 Elevenes gjennomføring

Elevene gjennomførte prøven i løpet av uke 16 og 17. Prøvene var tydelig ikke samkjørt med mange andre oppgaver som også skulle gjennomføres i de samme ukene. Det gjaldt elevkursjoner, tentamener, nasjonale prøver i matematikk, etc. Dette medførte et stort press på elevene og skolene som hadde mye å gjennomføre i løpet av to korte uker. Det ble etter stor etterspørsel fra flere skoler bestemt at uke 19 skulle bli siste uke for gjennomføring av prøvene. Etter dette ble systemet lukket - til tross for at det enda var skoler som av ulike grunner ikke hadde gjennomført prøven.

Når det gjelder selve utviklingen av prøvene har dette vært et kontinuerlig samarbeid med ulike skoler i Norge med stor vekt på de tilbakemeldinger som har kommet fra lærere og elever.

Som eksempel kan en nevne elevenes ønske om å vite hvor mange oppgaver som var igjen av prøven. En prøve som utføres på papir vil være oversiktlig i forhold til arbeidsmengden som er utført og hvor mye som er igjen. Papiret visualiserer dette tydelig for eleven. Dette er ikke minst viktig i forhold til å kunne disponere tiden en bruker på de enkelte oppgavene. Denne digitale prøven hadde ingen tidsbegrensning, men elevene uttrykte dette behovet likevel. Det ble derfor lagt inn en "teller" som viste hvor mange oppgaver som var gjennomført og hvor mange oppgaver som var igjen.

7.9 Elevenes reaksjoner

Digital kultur og teknologiske verktøy og løsninger er ikke fremmed for elevene i dagens skolenorge. En kan i mange sammenhenger se at den teknologiske kompetansen kan være bedre hos elevene enn hos mange av lærerne. Den teknologiske erfaringen som hver enkelt elev har var også viktig å bygge på i forhold til disse prøvene. Det er likevel ulike aspekt ved papirbaserte prøver som også vil gjelde digitale prøver – digitale prøver er fremdeles prøver og vil alltid møte på samme type problematikk som papirbaserte prøver. Eksempler på dette kan være at elever som ikke kan stoffet kan bli ukonsentrerte, frustrerte, kjede seg og gjøre prøven fort unna for å bli ferdige – som på en papirbasert prøve.

Nyhetens interesse ved digitale prøver vil etter hvert forsvinne. En ser at mange av tilbakemeldingene som har kommet i forhold til de digitale prøvene er de samme som en kunne forvente fra elever som tar papirbaserte prøver.

7.10 Negative reaksjoner fra testperioden

Før selve prøven ble gjennomført av 10. klassingene ble det kjørt jevnlig tester av system og oppgavesett for å få en oversikt over elevens reaksjoner på denne type prøver. Dette var pilotutprøvingene som senere i denne rapporten er omtalt som *formativ evaluering*, og som ble gjennomført med elever fra både 4.-5.-7.-8.- og 10.klassetrinn. På bakgrunn av de tilbakemeldingene som kom kunne enkelte problemer lukes vekk. Tilbakemeldingene viste seg å være svært ulike. Noen elever opplevde prøvene som langtekkelige fordi det ble for mye å lese. Det var vanskelig å finne ord i teksten når man skulle markere ordene, det ble kjedelig, lite variert og vanskelig. Svake elever fikk lite til og prøven ble derfor en negativ opplevelse for disse. Andre gav tilbakemeldinger om at prøven ikke var morsom nok eller at det var vanskelig å forstå hva en egentlig skulle gjøre. Noen elever sa at de ble så lei at de ikke løste oppgavene på beste måte, men bare klikket seg videre for å bli ferdige. Dette gjaldt både de enkelte oppgavesettene og hvordan de var satt sammen.

En annen ting er det fysiske forholdet mellom datamaskin og elev. Noen elever klaget over vondt i hodet fordi de måtte sitte og se lenge på skjermen, andre mente det var vanskelig å få oversikten når en måtte rulle teksten opp og ned. Noen elever ga tilbakemelding om at de likte den papirbaserte "gamlemåten" best, da det kunne ta lang tid å laste ned en ny oppgave. Mange av disse faktorene førte til at elevene mistet konsentrasjonen og derfor ble mindre motivert til å gjennomføre prøven.

De fysiske problemene som elevene gav tilbakemelding om er noe som er mulig å rette opp i ved bedre skjermkvalitet – dette er et problem som vil bli diskutert mer utførlig i et senere avsnitt (se avsnitt om Aldersavhengige reaksjoner). Problemer rundt kjedsomhet er noe som er vanskeligere å rette opp i. Noe av dette kan handle om elevenes vanlige opplevelse av å jobbe med datamaskinene. Datamaskinen oppleves som og forbindes med et underholdningsprodukt og utnyttes på denne måten av den generasjonen som ikke bruker datamaskinen i jobbsituasjonen. Dermed kan opplevelsen bli skuffende i og med at prøven ikke er like "fartsfylt" som et dataspill. Ved videre utvikling bør en kanskje se etter oppgaveløsninger som baserer seg mer på barne- og ungdomskulturen for å dra nytte av den kunnskapen som interesserer de ulike klassetrinnene. For de minste elevene kan dette være nødvendig og viktig, men eldre elever bør være vant med at prøvesituasjonen ikke er lek og underholdning.

7.11 Positive reaksjoner fra testperioden

Tilbakemeldingene har også vært av en slik art at det ikke ville spilt noen rolle om prøven ble tatt på en datamaskin eller ikke. Tilbakemeldinger som at tegningene var tøffe kunne like gjerne gjelde papirbaserte prøver.

Det at elevene virker interesserte og liker prøveformen generelt handler derimot om opplevelsen av prøvens digitale form – dette gjelder tilbakemeldinger om at databaserte prøver oppleves som bedre enn å skrive på ark og at opplevelsen av interaktivitet er positiv. Et av de viktigste aspekter i forhold til opplevelsen av nettbaserte prøver som positive er at oppgavene ikke må føles vanskeligere å håndtere fordi en sitter og lur på "hva en skal gjøre". Dette gjelder i stor grad elevenes kompetanse i forhold til IKT og nye digitale medier. Dette spørsmålet vil bli behandlet i avsnittet om Aldersavhengige reaksjoner – men i korthet kan en si at så lenge manglende datakompetanse ikke var et problem var tilbakemeldingene at elevene opplevde oppgavene, bildene og prøven som veldig bra. Lærerne uttrykker at databaserte oppgaver er noe elevene liker å jobbe med, og det blir enkelt for dem å vite hva de skal gjøre for å løse oppgavene.

De aller fleste ser ut til å like seg framfor dataskjermen, som er et medium som "tilhører" elevenes generasjon. Flere av tilbakemeldingene gikk på at elevene forstod oppgavene og ikke hadde problemer med å skjønne hva de skulle gjøre. En av de viktigste faktorene var at elevene var positive til å ta prøven på data. Dermed jobbet elevene konsentrert og noen av elevene hadde lyst til å ta prøven flere ganger.

Tilbakemeldingene er med andre ord svært varierte og dikotomien i tilbakemeldingene viser dette; morsomt; kjedelig; spennende; statisk etc. De erfaringene vi kan trekke av dette er at det er vanskelig å måle kunnskaper om ikke bestemte faktorer ligger til rette for dette. I denne prøveformen er det viktigste at elevene har gjort seg kjent med de ulike oppgaveformatene før prøven begynner. Dette vil bli tatt opp i avsnittet om IKT og nye medier - bro eller hinder.

7.12 Aldersavhengige reaksjoner fra testperioden

De nevnte problemene og opplevelsene var stort sett like for de ulike klassetrinnene. Alle klassetrinnene viste at de hadde god erfaring med bruk av IKT og mestringen av det og likte måten å løse oppgaver på.

Lærere for 4. klasse ga imidlertid tilbakemelding om at så mye tekst ikke hørte hjemme på dette klassetrinnet, da det her fokuseres mest på det muntlige; å snakke og synge. Flere lærere på dette trinnet ønsket derfor audiovisuelle prøver. Et annet aspekt ved 4. klassetrinn var behovet for å få instruksjonene på norsk. Svake elever vil ikke ha mulighet for å skjønne hva de skal gjøre om det ikke er instruksjoner på norsk. Andre tilbakemeldinger når det gjaldt 4. klasse var at elevene ikke var komfortable med å sitte og jobbe så intenst ved datamaskinen. De ble derfor utålmodige fordi prøven krevde mer konsentrasjon enn de var vant med fra tidligere. Imidlertid fant en også tilbakemeldinger fra 4. trinn om at elevene fant at dette var gøy og syntes "prøve på data" var spennende og motiverende.

Lærerne i 7. og 8. klasse ga tilbakemeldinger om at disse trinnene hadde god evne til å kunne konsentrere seg over lengre tid foran en dataskjerm. Elevene her var vant med å sitte foran en dataskjerm i lengre tid. Her må det imidlertid skilles mellom ulike måter å jobbe foran en datamaskin på. I dagens Norge er barn etter hvert blitt svært vant til å sitte lenge konsentrert foran en dataskjerm. Likevel er aktivitetene som unge vanligvis driver med på datamaskinen

svært forskjellig fra det å lese lange tekster på skjermen. F.eks. krever spilling en annen form for konsentrasjon og blikkaktivitet enn det å lese tekster. Det kan antas at mange av skolenes skjermer skiller seg vesentlig fra de skjermene som barna bruker hjemme. Skolenes skjermkvalitet kan være en undervurdert faktor når elever skal jobbe foran en datamaskin. Faktorer som lys, farge og flimring kan ha mye å si for elevenes resultater og bør vurderes nøye.

For 10. klasse skal ikke oppgavene bare være morsomme. Norske 10. klassinger krever gjerne at arbeidsoppgavene de skal jobbe med også skal "fenge" dem - temaene må være interessante for dem. Én tilbakemelding var at tematikken i tekstene var "underlig fjernt fra en femtenårings verdensbilde", noe vedkommende (som var lærer) mente gjorde prøven demotiverende å gjennomføre. I noen tilfeller førte dette til at elever "klikket seg gjennom" prøven uten å ta den seriøst. Dette var selvfølgelig uheldig for en test som skulle fungere som en korreksjon før den endelige prøven, da det ikke ga en reell tilbakemelding på vanskelighetsgrad, tidsbruk og liknende.

7.13 Skolereaksjoner

"Vi tror nasjonale prøver på data kan bli en god og effektiv prøveform. Enkelte oppgaver skilte godt mellom dei flinke og dei litt svakare elevar" (ref.e-post)

"Som lærer var dette en spennende måte å utføre nasjonale tester på! Elevene navigerer seg lett fra side til side og kommer etter ca 45 minutter til slutten og prøven er unnagjort." (ref.e-post)

Tilbakemeldingene har vært mange og konstruktive. Det er ikke bare elevenes tilbakemeldinger som har vært sprikende og varierte, det har også lærernes og skoleledelsens tilbakemeldinger vært. Mange av reaksjonene er avhengig av noen bestemte faktorer. Dette gjelder om skolens tekniske infrastruktur har fungert bra eller ei, og det gjelder skoleledelsens og lærernes holdninger til bruk av IKT og nye medier i skolen. Skolens tilbakemeldinger har variert fra at de ikke kan forstå meningen med prøvene til at det har vært en spennende utfordring hvor en ser positivt på å bruke denne typen prøver.

Det er viktig å berømme skolens innsatsvilje til å gjennomføre et prosjekt som dette - ikke minst de skolene som har vært med på testing av systemet under selve utviklingen. På tross av at digital kultur og bruken av den er voksende, ikke bare i det norske samfunnet generelt men også i ulike institusjoner som skolen, krever det stor innsats og ikke minst vilje og kunnskap. Dette kan være krevende i en allerede travel skoleadministrasjon og lærerhverdag.

Pedagogiske og fagdidaktiske spørsmål er også viktige i denne situasjonen. Det er viktig at vanskelighetsgraden samsvarer med læreplanen i faget og det uttrykkes bekymring for at prøvene som blir tatt i bruk vil styre undervisningen, slik at det som måles i slike tester blir vektlagt sterkere enn de områdene som er vanskelig å teste med en slik prøveform. Dette er viktig i en vurdering av hvordan nasjonale prøver skal brukes, som kvalitetssikringsverktøy.

Lærere poengterte at andre faktorer ble testet ut før selve engelskkunnskapene. Det gjaldt som nevnt datakunnskapene, men også leseferdigheter, tålmodighet og logisk tenkning. Når det endelig var engelskkunnskapene som telte var det ikke mye igjen av utholdenheten på slutten av prøven.

Flere lærere har gitt uttrykk for at de opplever det som en ryddig og oversiktlig prøveform forutsatt at infrastrukturen på skolens datautstyr kan bringes til den standard som er

nødvendig. Om det tekniske fungerer er det veldig fort gjort å gjennomføre en slik prøve, spesielt siden rettingen foregår elektronisk.

Den digitale prøveformen blir av flere lærere altså sett på som en praktisk måte å gjennomføre en prøve på. Men lærere som hadde elever med andre behov i klassen påpekte problemer som bør undersøkes nærmere. Spesielt gjelder dette lesesvake elever, som fort får problemer med å klare denne form for prøve, siden den inneholder mye tekst. Disse elevene er ikke nødvendigvis svake i engelsk, men blir på grunn av lesesvakhet ikke testet i sine faktiske kunnskaper fordi de er prisgitt det medium som benyttes til testingen.

I regjeringens Program for digital kompetanse 2004-2008 (utgitt av Utdannings og forskningsdepartementet) legges det vekt på oppbygging av IKT-ferdighetene. Enkelte av lærerne har gitt tilbakemelding om at testen inneholder mye positiv "datalæring". Denne prøveformen er dermed også med på å øke den digitale kompetansen og er ikke kun en test av en fagkunnskap.

7.14 Regler og retningslinjer - juksing på nasjonale prøver

En del praktiske løsninger var ikke påtenkt før problemene og spørsmålene oppstod. Som eksempel kan nevnes hvorvidt maskinene bør stå lenger fra hverandre for å unngå juks. I utgangspunktet kan det virke som om adaptive prøver – at elevene får en prøve tilpasset deres kunnskapsnivå – i noen grad vil forhindre at elevene kan se og kopiere hverandres svar.

Etter en forprøve vil datasystemet finne frem til det av tre ulike prøvesett som passer best til eleven. Ettersom elevene som sitter ved siden av hverandre ikke nødvendigvis vil få den samme prøven vil muligheten for juks bli mindre. Men om man tar utgangspunkt i at gjennomsnittlig flest elever vil havne på den nest vanskeligste prøven vil det likevel være store sjanser for at elever som sitter ved siden av hverandre får like prøver. Ikke bare det – om en elev kopierer sin sidemanns svar fra starten av prøven vil jo disse to elevene få samme hovedprøve og kan fortsette "samarbeidet". Det bør derfor utarbeides et regelverk som bl.a. kan regulere avstanden mellom maskinene, eller ses på muligheter for skillevegger.

Et annet problem er at det var mulig for elevene å logge inn og gjøre prøven ferdig hjemme. Elevene er avhengig av å kunne komme tilbake til den oppgaven som var den siste de gjorde dersom datamaskinen låser seg. Det samme gjelder om andre tekniske problemer ved systemet oppstår slik at de ikke får gjennomført prøven. Dette kan gjøre det mulig å jukse på prøven ved at eleven ikke gjør ferdig prøven på skolen, husker passord og brukernavn og logger på hjemme for å gjøre ferdig prøven der. Systemet registrerte tilfeller av at elever hadde ferdigstilt prøver klokken 23 om kvelden, med beste karakter som resultat. Det er altså helt nødvendig at elevene ferdigstiller prøven på skolen under oppsyn av lærer. Eventuelt bør det være mulig for lærer å hente ut karakterene sammen med tidspunkt for ferdigstillelse.

7.15 IKT og nye medier - bro eller hinder

Før prøven klarer å måle det kunnskapsnivået eleven har i engelsk blir elevens IKT-kunnskaper også satt på prøve. En slik test må så langt som mulig skrelle vekk andre ting enn det som ønskes målt. Dette er et svært viktig poeng i denne sammenhengen - men det er ikke et fenomen som kun gjelder databaserte prøver.

Under papirbaserte prøver vil også resultatet være avhengig av leseferdigheter, tålmodighet og logisk tenkning. Prøvesituasjonen er en konstruert situasjon som viser elevenes ferdigheter

i den gitte situasjon – men kan aldri helt og fullt måle hvordan elevenes kunnskaper vil uttrykkes i en faktisk situasjon.

Det er viktig at elevene på forhånd har prøvd seg på "prøve"-testen for å trene opp sine IKT-ferdigheter i denne situasjonen. Dersom gjennomgangen av de ulike oppgavene er tatt med elevene på forhånd er det få som trenger forklaring på hvordan de skal gjøre dette. I og med at mange av oppgavene bygger på prinsipper som elevene kjenner fra før gjennom dataspill, som for eksempel *klikk og dra*-prinsippet, ble tilbakemeldingen ofte at gjennomføringen gikk veldig greit og at denne måten å teste på virket positivt på elevene.

En person som aldri tidligere har sett en bok vil før instruksjonen også være ukjent ved dette mediet. Dette er en faktor som alltid vil spille inn og som vil være vanskelig å fjerne. Det er derfor viktig at prøvene bygger på kunnskap som allerede er kjent gjennom tidligere erfaringer med datamaskinen som et medium. Elevene bør derfor bli gjort kjent med de ulike oppgavetyperne før den "ekte" prøven gjennomføres. Dette har mye å si for om prøven kan oppleves positivt eller negativt av elevene.

7.16 anbefalinger

1. All korrespondanse til skolene bør skje gjennom Læringscenteret. En tredjepart kan føre til kommunikasjonsproblemer eller at informasjonen "forsvinner" og ikke settes i forbindelse med den riktige saken.
2. Det må utarbeides oppdaterte lister på hvilke skoler som skal ta prøvene.
3. Det må sendes ut tydelig instruksjon til de skoler som har utplasserte elever fra andre "hjemmeskoler" om hvordan de skal forholde seg til registrering av disse elevene.
4. Skolene må få retningslinjer for hvem som skal være skolens kontaktperson i forhold til brukernavn og passord.
5. Læringscenteret må informere om at den enkelte skole ved hjelp av skolens passord og brukernavn kan gå inn og se om det har skjedd uregelmessigheter i systemet i forhold til enkeltelevers karakternivå.
6. Det bør utvikle løsninger for elever med spesielle behov i prøvesituasjonen.
7. Det må utvikles klarere instruksjoner på hvordan prøven praktisk skal gjennomføres – for eksempel når det gjelder avstanden mellom datamaskinene og ferdigstilling av prøven på skolen.
8. Det er viktig at IKT-kunnskaper inngår som en del av fagkretsen.
9. Læringscenteret bør ha oversikt over oppdaterte e-postlister til alle skolene

8 TEKNISKE ERFARINGER VED SKOLENE

Tekniske forberedelser for å lykkes med gjennomføringen av de nasjonale prøvene startet i september 2003. Dette arbeidet bestod av:

- å teste ut prosjektets valg av teknologisk løsning på klientsiden (altså skolene)
- å videreutvikle leveringssystemet og velge serverløsning
- å organisere og gjennomføre landsdekkende pilottester for å teste trafikkpress på server og leveringssystem, samt fortsette med pilottesting av klientprogrammet på skolene
- å analysere data fra testene for å evaluere hvilke oppgaver som kvalifiserte til å bli presentert i de endelige prøvene
- å lage et brukerhåndteringssystem for generering av brukernavn og passord
- å organisere en brukerstøttegruppe for raskt å yte støtte til skoler som kontakter BITE^{IT} angående gjennomføring av prøvene.

Alle disse faktorene var kritiske for å kunne lykkes med gjennomføringen av nettbaserte leseprøver. I dette kapitlet fokuseres det på erfaringer fra det tekniske forarbeidet mot skolene gjennom uttesting, samt hvordan brukerstøttegruppen ble organisert. Til slutt i kapitlet vil det bli en kort presentasjon av hvordan hovedserveren er satt opp, hvordan den blir driftet og hvordan sikkerheten blir ivarettatt.

8.1 Den teknologiske klientløsningen for skolene

Prosjektets tekniske kontaktperson fikk hovedansvaret for å reise rundt til skoler i Bergensområdet og prøve ut hvordan ulike prototypversjoner fungerte på typiske skolemaskiner. Dette var nødvendig for å kartlegge hvilke systemer skolene har og hvilke begrensninger disse systemene setter for prosjektets muligheter. Ikke minst var det viktig å få dannet et inntrykk av den tekniske kompetansen skolenes IKT-ansvarlige besitter.

Ved prosjektstart ble det vedtatt at klientprogrammene skulle være nettbaserte og at de skulle benytte seg av Macromedia Flashteknologi. Macromedia Flash[®] (se nettside www.macromedia.com) er et skriptprogrammeringsspråk fra Macromedia som er blitt konstruert med tanke på å lage innhold til nettsider og interaktive programmer som leveres over internett. Flash[®] benytter i hovedsak vektorgrafikk⁵ istedenfor punktgrafikk (bitmap) som ellers er vanlig for bilder og animasjoner. Vektorgrafikk gjør det mulig å vise avanserte

⁵ Punkt- eller bitmapgrafikk beskriver innholdet i hvert eneste punkt i bildet. Derfor er det grenser for hvor mye bildet kan forstørres før det går synlig ut over kvaliteten. Ved å forstørre et bilde tilfører man det ikke flere detaljer, men man fordeler bildets informasjon over et større antall piksler. Alle digitale fotografier er bitmaps. Noen av de vanligste filformatene i denne kategorien er jpg, gif og tif, og disse blir anbefalt dersom man skal vise detaljerte motiver med urene fargeflater (for eksempel naturlige landskapsbilder, fjellsider osv).

Vektorgrafikk derimot beskriver bildet ved hjelp av geometriske verdier. Av den grunn kan man forstørre denne typen grafikk uendelig antall ganger uten at bildet mister kvalitet. Vektorgrafikk egner seg spesielt godt der det er relativt få detaljer og rene flater. Typiske filformater med vektorgrafikk er swf (Flash), ai (Illustrator) og svg.

bilder, animasjoner og andre interaktive løsninger på en smalere båndbredde enn dersom man bruker andre typer grafikk. Dette frigir båndbredde til andre aktiviteter som for eksempel overføring av levende lyd og bilder i såkalte strømmer ("streams"), noe som Flash[®] også støtter. I tillegg til at Flash[®] kan brukes til å lage komplette internettsider, kan programmet også lage innhold som plasseres i vanlige sider skrevet i det tradisjonelle vevspråket HTML.

En av forutsetningene for å kunne bruke Flash[®] som teknologi for klientsiden er at det finnes en Flash-avspiller, som er et gratis tilgjengelig program som "plugges inn" i nettleseren (lastes ned fra Macromedias hjemmeside). Det skal være forholdsvis enkelt for skoler å installere denne tilleggsprogramvaren siden det ikke krever at andre deler av operativsystemet endres. Dette høres enkelt og greit ut, men et av hovedproblemene ved dette teknologivalget var å få Flash-avspilleren "plugges inn" i skolenes nettlesere og få dette til å fungere på en tilfredsstillende måte. Dette problemet vil bli utdypet i punktet om brukerstøttegruppen.

8.2 Generelt inntrykk av utstyr ved skolene

For å få et førsteinntrykk av hvilke systemløsninger som finnes i den norske grunnskolen, reiste prosjektets tekniske personell rundt til flere skoler i Bergen kommune og nabokommunene Fjell og Øygarden. På grunn av variasjon i kommunenes økonomi regnet man med å få tilgang til utstyr og nettverk av ulik kvalitet, og således få et inntrykk av hva gjennomsnittlig utstyr på landsbasis kunne være.

Elevmaskiner ble hovedsaklig testet etter følgende mønster:

1. Finn den svakeste maskinen og test ut prototypen på denne først.
2. Finn den beste maskinen og gjør en identisk test.
3. Test så mange maskiner som mulig innenfor den tilgjengelige tidsperioden.

Dette testmønsteret ble valgt fordi prosjektet var interessert i å kartlegge det svakeste utstyret skolene hadde. Hypotesen var: "Dersom de svakeste maskinene kan kjøre prototypen mens en sterkere maskin ikke klarer det, skyldes det ikke maskinvaren, men er et programvareproblem som kan løses ved rekonfigurering." Å kjøre tester på de beste maskinene ble gjort for å bekrefte denne arbeidshypotesen.

Ved hjelp av denne framgangsmåten erfarte prosjektet følgende:

- Det ble ikke funnet noen maskiner som var for svake og ikke kunne kjøre klientløsningen. Den svakeste maskinen som ble funnet var en Pentium 60 (MHz) med 16 MB RAM (internminne). Dette er en maskinspesifikasjon som var vanlig i 1995-96. Riktignok tok det lang tid å laste inn prøvene, men maskinen klarte å kjøre gjennom flere tester. Dette var et svært positivt funn, da de fleste skoler har nyere og kraftigere maskiner enn dette.
- Videre fant man at maskinvarespesifikasjonen Pentium 166 (MHz) med 64 eller 128 MB RAM gikk igjen i veldig mange skoler. Dette er typiske maskiner som mange skoler kjøpte inn rundt 1998. Basert på denne forundersøkelsen kom man fram til at dette var det som i verste fall kunne forventes å være gjennomsnittlig maskinvareprofil i skolene.

- I tillegg erfarte man i løpet av forundersøkelsen at kombinasjonen av svake maskiner (som ligger rundt minimumskravet) og ISDN-nettverk er uheldig. Uttesting av ISDN-linjer og belastning på denne type nettilgang gav svært dårlige resultater ved flere enn 4 synkrone tilkoblinger. Prosjektet anbefaler derfor at skoler med ISDN-linjer ikke bør tillate mer enn fire samtidige tilkoblinger, og dersom dette ikke fungerer skal man redusere antall tilkoblinger. For skoler med svært mange elever kan ISDN-nett by på problemer, særlig for barneskoler, der oppgavene vil inneholde mer grafikk og dermed være tyngre å overføre.
- Skjermstørrelsen på skolenes elevmaskiner er ofte liten, med en god del 14"- og 15"-skjermer. Dette fører til at testene som utvikles må kunne vises på en svært liten flate. På grunn av disse begrensningene valgte man å lage testene med en oppløsning på 800 x 600 punkter. De skal da kunne sees i sin helhet på 15"-skjermer, dersom nettleseren er i fullskjermmodus (oppnås ved å trykke på F11-tasten). Informasjon om dette ligger ute på nettsidene til prosjektet. Imidlertid viste det seg at få skoler hadde fått med seg denne tekniske opplysningen, og noen skoler med små skjermer fikk problemer fordi de ikke fikk se hele testen på ett og samme skjerm bilde. Brukerstøttegruppen fikk mange henvendelser om dette under gjennomføring/testing.

Basert på disse erfaringene spesifiserte vi følgende maskinkrav:

Prosesor: Pentium 166 MHZ

Minne: 64 MB RAM

Grafikkort: 2 MB

Skjermstørrelse: Minimum 15"

Skjermoppløsning: 800x600 piksler

Antall farger: Minimum 8bit (eller mer enn 256 farger).

Erfaringer fra testene viste at kravene i denne maskinvarespesifikasjonen var omtrent så lave som de kunne være, dersom man samtidig skulle ha klienter som kjørte forholdsvis raskt, med lav ventetid i innlasting og utføring av instruksjoner i oppgavene. I tillegg ville det være mulig for marginalt svakere spesifikasjoner å kjøre prøvene på tross av noe lengre responstid. Innlastningstiden (ventetid for å laste inn en ny oppgave) avhenger sterkt av nettilgangen, men ved bredbånd var innlastningstiden sjelden mer enn 15 sekunder; noe mer dersom det var stor trafikk på nettet generelt og vår sentrale server spesielt. Rapporter fra skoler med ISDN viser til flere ganger så lang innlastningstid.

8.3 Inntrykk av teknisk kompetanse ved skolene

Samtidig med at gjennomføringen av disse utstyrstestene pågikk i skolene, fikk man dannet seg et inntrykk av den tekniske kompetansen som IKT-ansvarlige på skolene har. Kompetansen varierte veldig fra skole til skole. Noen steder hadde skoler som rutine å søke etter ferdigutdannede dataingeniører som ville avtjene sivilteneste og samtidig ha relevant arbeid. Ved andre skoler var IKT-ansvaret fordelt mellom tre eller fire datainteresserte lærere der det ofte var uklart hvem som hadde ansvar for hva. Det ble avdekket mange tilfeller der tidligere IKT-ansvarlige ble pålagt en større arbeidsmengde i vedlikehold av systemer enn

hva oppgitt ansvarsområde og stillingsprosent tilsa. I noen situasjoner førte dette til mindre samarbeidsvillige lærere. Ellers opplevde prosjektet svært hjelpeløse lærere og er takknemlige for alt ekstraarbeidet de har nedlagt.

Et generelt inntrykk er at IKT-kompetansen blant lærere og IKT-ansvarlige bør økes. Brukerstøttegruppen erfarte svært ofte at lærere som ringte inn og spurte om teknisk assistanse ikke visste hvilken nettleser og/eller hvilket operativsystem som ble brukt på maskiner hvor problem hadde oppstått. I tillegg hadde flere lærere problemer med å forklare konkret hva problemet var; hva som gikk galt, og hvor og når dette skjedde. Dette vanskeliggjorde hjelpeprosessen, og brukerstøttegruppen måtte svært ofte "famle seg fram i blinde" i slike situasjoner. Økning av teknisk kompetanse til lærere og særlig IKT-ansvarlige bør bli et satsningsområde dersom man tar sikte på å gjøre dagens datarom om til eksamenslokaler. Dersom elever skal avlegge nettbaserte nasjonale prøver (og eksamen) ved hjelp av datamaskiner, stiller dette et større krav til teknisk utstyr, vedlikehold og kompetanse enn det som svært mange skoler kan vise til i 2004.

8.4 Operativsystemer i skolene

Prosjektet tar mål av seg til å være plattform- og systemuavhengig. Skoler skal fritt kunne velge hvilke systemer de vil kjøre. Så lenge man installerer Flash[®] 6.0 eller nyere skal dette være mulig. Dessverre har disse målsettingene vært vanskelige å nå. Flere skoler har rapportert om problemer ved installasjon av Flash-avspilleren og andre tilsynelatende uforklarlige problemer under gjennomføring av prøvene. Imidlertid er det nok en sammenheng mellom IKT-kompetansen til den som erfarer et problem, og dennes evne til å forstå og forklare brukerstøttegruppen hva som er problemet, noe som får betydning for om problemet kan løses eller ei. Prosjektet har blant annet erfart at kombinasjonen av Windows 98 med nettleseren Internet Explorer (hvor Flash[®] blir plugget inn) har vært problematisk. Ellers har det blitt rapportert at mange skoler har gamle operativsystemer som Windows 95, Windows 98, Windows ME og NT 4.0. En generell anbefaling her er at alle skoler bør oppdatere systemene sine til minimum *et system som det kommer sikkerhetsoppdateringer til*. Nettverk med gamle Microsoft operativsystemer er potensielle sikkerhetshull der både virus og ormer lettere får spillerom. Det kan dermed være et interessant regnestykke å sammenligne hva som er billigst i lengden av å drifte gamle sårbare systemer eller å investere i nyere systemer. Skoler som velger Microsofts løsninger, bør ha et nyere operativsystem, som Windows 2000 eller Windows XP som det finnes sikkerhetsoppdateringer for.

Selv om storparten av skolene kjører Microsoft-baserte operativsystemer, er det også et voksende antall skoler som tar i bruk Skolelinux. Prosjektet har ikke opplevd noen store problemer med Skolelinux hittil. Et lite problem var at noen fonter ikke er støttet av "gratisprogramvareverdenen" på grunn av lisensieringspolitikk, men dette problemet ble løst ved å skifte font i oppgaveformatene. Et annet problem som Skolelinux møter er at Flash[®] ikke er forhåndsinstallert og integrert i distribusjonsplaten. Dette problemet ble løst ved at skolene tok kontakt med teknisk personell ved InterMedia som veiledet dem i denne type installasjon. I fremtiden kan dette problemet løses ved at distributørene av Skolelinux på forhånd lager oppsett som integrerer Flash[®] i nettleserne.

Det finnes sikkert en god del skoler som også kjører Mac-maskiner med operativsystemet OSX installert. Brukerstøttegruppen hørte svært lite fra disse miljøene og man regner derfor med at disse systemene klarte seg bra.

Noen skoler kjører på klient-terminalløsningen Citrix Metaframe. Denne type teknisk infrastruktur har mange drifts- og økonomifordeler, men en av ulempene er at prosessorkraften på den sentrale tjeneren må deles mellom alle klientene. Derfor tilrås sjelden teknisk personell at Flash[®] programvare blir installert og tilgjengeliggjort, da noen typer Flash-animasjoner stjeler svært mye prosessorkraft. IKT-ansvarlige for slike løsninger rapporterte at 4-5 klienter kunne tvinge ned hele skolens nettverk. Etter nærmere undersøkelse viste det seg at disse skolene hadde kjørt Flash-animasjoner funnet på Internett, og at de ikke hadde testet ut de ulike demoversjonene av leseprøvene i engelsk. Prosjektet har alltid hatt stor fokus på å lage så små og effektive Flash-animasjoner som mulig. Mengden av prosessorkraft som disse krever, er mye mindre enn andre typer animasjoner som man vanligvis finner på nettet. Prosjektet opplevde imidlertid at disse skolene ikke ville teste ut systemet fordi det brukte Flash-teknologi. Det ble foreslått en løsning for disse skolene der Flash-teknologi kun blir tilgjengelig i ett av flere sett med profiler på maskinene. Dette innebærer at elevene deres logger seg på skolens system med et annet brukernavn og passord enn de normalt bruker, den dagen de skal avlegge nasjonale leseprøver i engelsk. Vi vet med sikkerhet at mange skoler som kjører Citrix Metaframe har gjennomført leseprøvene med sine 10.-klasseelever, men det har ikke kommet tilbakemelding om hvorvidt disse fulgte prosjektets løsningsforslag.

8.4.1 Nettlesere generelt

Generelt for alle plattformer (operativsystemer) er det blitt rapportert om problemer ved oppgaver med mye interaktivitet. Dette gjelder spesielt maleoppgavene (se bilde 5, kap. 5). I flere typer nettlesere, på tvers av plattformer, forsvinner malepenselen (musepekeren) som brukes i denne oppgavetypen. Dette kan bli et problem for fag som krever mer dynamikk og interaksjon. Kanskje er ikke skolene godt nok teknisk utrustet for denne type nettbaserte oppgaver enda? Mer systematisk testing og datainnsamling bør gjøres før prosjektet kan uttale seg sikkert på dette området. Ellers har prosjektet hatt gode erfaringer med de nyeste nettleserne, og særlig med den norske nettleseren Opera. Denne nettleseren er gratis, liten, krever lite minne og er lett for svake maskiner å kjøre. Den kan være en god løsning for ressursvake skoler. For gamle nettlesere som Netscape 4-serien, Opera 6, Internet Explorer 5.4 og eldre versjoner, har prosjektet hatt dårlige erfaringer og anbefaler å oppgradere til nyere versjoner. Det er viktig å påpeke at svært mange skoler fremdeles bruker disse gamle utdaterte nettleserne som ikke (automatisk) støtter dagens nettstandarder.

8.5 Oppsummering av konfigurasjoner

En oppsummering av innrapporterte og tilbakevendende problemer blir presentert i tabellen under:

Enhet	Problem
Windows 95, Windows 98, Windows Millenium	Fungerer dersom konfigurasjonen er helt riktig. Men dette kan være vanskelig å få til og det fungerer ofte ikke sammen med Internet Explorer.
Windows NT	Fungerer med et nytt og rent oppsett. Fungerer dårlig i profilstyrte systemer mot Netscape 4.x familien. Generelt dårlige erfaringer med kombinasjonen av NT og Netscape 4.x.
Windows 2000, Windows XP	Fungerer med riktig og rent oppsett, best å installere alle oppdateringer for å få Flash [®] til å fungere riktig. Fungerte mot alle nye nettlesere på dette operativsystemet.
Mac OSX	Fungerte bra med Flash [®] plagget inn i nettleser, særlig bra mot Internet Explorer på OSX.
Skolelinux	Når Flash [®] blir riktig installert ser det ut til å fungere greit. I noen tilfeller hadde man problemer med fontene. Dersom prøvene mangler tekst kan dette ordnes ved å oppdatere fontene.
Internet Explorer 3.x og 4.x	Har aldri fått oppgavene/prototypene til å kjøre mot denne type konfigurasjon.
Internet Explorer 5.5 og nyere.	Med standard innstillinger fungerer denne nettleseren fint mot Flash [®] og nyere operativsystemer. Problemer kommer ofte av enten eldre operativsystemer eller lokale innstillinger i registeret. En negativ side med Internet Explorer er at den er så tett integrert i de Windows-baserte operativsystemene. Dette gjør at noen systeminnstillinger definert av andre programmer (i registeret) kan påvirke nettleseren og føre til at Flash [®] ikke kjører som forventet. Slike innstillinger er umulig for brukerstøttegruppen å finne over telefon.
Netscape 4.x	Det er mulig å installere Flash [®] for Netscape 4.x-nettleserene, men skolene har ikke klart å få til dette. Hovedsakelig skyldes dette at skoler som bruker en så gammel nettleser ofte har et gammelt operativsystem. Prosjektet anbefaler ikke at skoler bruker nettlesere av Netscape 4.x-familien lengre.
Netscape 6.0 og nyere.	Netscape 6.0 og nyere kjører prøvene veldig godt. Nettleseren er ikke så tett integrert mot operativsystem så det er lettere å styre konfigurasjonen for eventuelle tilleggsprogrammer som plugges inn. Et minus med denne nettleseren er at penselen i maleoppgavene av og til forsvant. Ved å høyreklikke i prøven fikk man fram den vanlige musepekeren og kunne fortsette som normalt. Denne løsningen ble oppdaget ved en tilfeldighet og det var ikke lett å opplyse skolene som opplevde dette problemet om hva de kunne gjøre for å fortsette som normalt.

Mozilla-familien.	Mozilla er beslektet med Netscape, men er hovedsakelig en forbedring av denne nettleseren. Mozilla-klienter kjører også bra mot de fleste operativsystemer og mot Flash [®] . Ellers har Mozilla også samme uheldige feil i maleoppgavene som erfart med slektningen Netscape 6.0 (se ovenfor).
Opera 7.x	Opera 7.23 har fungert veldig bra mot alle operativsystemer. Flash [®] blir også automatisk installert med Opera, så skoler som bruker Opera slipper å tenke på installering av Flash [®] . Det største minuset med Opera har vært at malepenselen ofte forsvinner i maleoppgavene, men i tillegg kommer ikke musepekeren fram igjen. Dette er kun ut mot Windows operativsystemer.
Nautilus, Konqueror og Epiphany	Dette er nettlesere som man omtrent bare finner i Linux-distribusjoner. Disse nettleserne kjørte Flash [®] og prøvene helt fint.
14" Skjermer	For liten fysisk skjermflate til å vise hele oppgavevinduet på skjermen uten å måtte rulle ned eller til sidene.
15" Skjermer	Problemer med liten skrift dersom punktoppløsningen ble satt til 1024x768. Ellers fungerte det fint med 800x600 i oppløsning og fullskjerm modus.
17" Skjermer og større	Ingen registrerte problemer.
ISDN-nettilgang	Fungerte dårlig dersom det var stor trafikk mot prosjektets sentrale server. Anbefalt antall samtidige oppkoblinger er fire eller færre.
Bredbånd, minimum 700kilobits/sekund	Med bredbånd ble overføringstiden for å hente ned og laste opp svar liten og klientprogrammet fungerte fint. I noen tilfeller delte hele kommunen samme bredbånd (oftest med kapasitet på 2mb/s) og det kunne da være perioder på dagen med stor regional nettrafikk som økte kommunikasjonstiden mellom klient og tjener.

8.6 Organisering av brukerstøttegruppen

Under de landsomfattende pilot-/kalibreringstestene høsten 2003 fikk BITE^{IT} svært mange henvendelser fra skoler som skulle delta. Skolene ringte, sendte e-post og fakser om alle mulige spørsmål de måtte ha, om alt fra organisatoriske til tekniske problemer. Av de tekniske henvendelsene dreide det seg hovedsakelig om hjelp til å installere Flash-avspilleren i skolens nettlesere. Før den nasjonale avviklingen av prøvene for 10.-klassinger skulle gjennomføres, ble det opprettet en brukerstøttegruppe. Det ble satt opp et gruppenummer som fordelte innkommende telefonsamtaler etter prioritet (da noen personer i prosjektet jobbet mer mot skolene enn andre). Denne organiseringen gjorde det mulig å besvare opptil 6 henvendelser samtidig. Brukerstøtteordningen så ut til å fungere bra og vi fikk hjulpet svært mange skoler med deres problemer. Nedenfor følger en oversikt over de hyppigste tekniske henvendelsene.

Uklar ansvarsfordeling mellom IKT-ansvarlig (ofte flere) og engelsklærere gjorde at teknisk klargjøring ikke ble gjort eller ble utsatt. Ofte ringte engelsklæreren og fortalte at ingenting fungerte selv om skolens IKT-ansvarlig hadde klargjort datamaskinene dagen før.

Klargjøring og testing av skolens maskiner ble ikke alltid gjort på en tilfredsstillende måte. Ofte sjekket lærerne om klientprogrammet kjørte på lærernes kontormaskiner. Kontormaskinene er konfigurert på en helt annen måte med fulle eller utvidete tilgangsrettigheter til operativsystemet. I tillegg er lærermaskinene ofte nyere med oppdatert programvare. Lærere som prøvde demoversjonen av leseprøvene på sine kontormaskiner fikk det ofte til der og trodde dermed at alt også ville fungere på elevmaskinene, enda det ble klart spesifisert i brev at all testing og klargjøring måtte gjøres på de maskinene som elevene skulle bruke.

Andre typer problemer gikk ut på at IKT-ansvarlig eller lærer bare hadde testet to-tre maskiner. Det skulle da vise seg at Flash[®] allerede var installert i nettleserne på de maskinene som ble testet ut. Neste dag ringte lærer under gjennomføring av testen og lurte på hvorfor det ikke fungerte på de andre maskinene, og etter litt testing kunne brukerstøtten fastslå at Flash[®] manglet på disse maskinene.

Ved henvendelser hvor det kunne fastslås at Flash[®] manglet på elevmaskinene, var svært få villige til å gå tilbake til instruksjonssidene og lese disse for å få en forklaring på hvordan Flash[®] installeres. Ofte oppdaget brukerstøtten at lærerne ikke engang var klar over at informasjonssidene fantes. Dette mener BITE^{IT} skyldes for dårlige forberedelser, og at de ansvarlige ikke tar seg tid til/har tid til å lese den informasjonen som de blir henvist til. Dette fenomenet gjelder særlig for elektroniske brev. Det virker som om denne informasjonskanalen blir tatt mindre seriøst enn tradisjonelle papirbaserte brev. I tillegg kan det hende at informasjon fra InterMedia blir tatt mindre seriøst enn om denne kom direkte fra Læringscenteret.

I en del tilfeller lot Flash[®] seg ikke installere på elevmaskinene. Dette ble registrert ved samtlige Windows operativsystemer sammen med nyere Internet Explorer nettlesere (versjon 5.0 og høyere). Mer presist, Flash[®] ble tilsynelatende installert uten feilmeldinger, men da nettleseren skulle kjøre klientprogrammet var det ingenting som fungerte. Dersom det kom en helt hvit, tom side opp, var dette et symptom på at Flash[®] manglet eller ikke var riktig installert/fungerte sikkert. For disse skolene løste brukerstøtten problemet ved å hjelpe dem med å installere nettleseren Opera. I Opera er Flash-avspilleren integrert, samt at Opera er løsere koblet mot operativsystemet. Dette gjør det mulig å ha større kontroll over hvilke variabler som påvirker nettleseren og dermed uttale seg med større sikkerhet om hva som går galt.

Mange henvendelser dreide seg også om liten skrift og/eller at skrift forsvinner i deler av oppgavene. Dette problemet ble identifisert for 15" skjermer som ikke kjørte prøvene i fullskjerm modus, eller 17" skjermer med en oppløsning lik 800x600 og som heller ikke kjørte i fullskjerm modus. Løsningen i slike tilfeller er å velge riktig oppløsning i forhold til skjermens fysiske størrelse og eventuelt blåse prøvene opp ved å velge fullskjerm modus (trykke på F-11 tasten når man er i nettleseren).

I tillegg fikk brukerstøttegruppen en del henvendelser angående registrering av elever og hvordan de skulle få brukernavn og passord. Dette var også beskrevet på våre informasjonssider, i tillegg til at informasjonen var sendt ut pr brev til alle skolene. En del lærere hadde vansker med å forstå hvordan registreringssystemet (hvor man registrerte skolens elever og laget elevidentiteter) skulle brukes. Her bør prosjektet ta selvkritikk, da dette systemet ikke fungerte helt feilfritt før rundt 10. mars. I tillegg burde det ha vært en brukermanual for systemet slik at lærerne kunne ha funnet frem til den informasjonen de trengte selv (om de leser manualer eller instruksjonssider).

8.7 Drifning av den sentrale BITE^{IT}-serveren

Fra et teknisk perspektiv kan det være problematisk å gjennomføre en så omfattende prøve som de nasjonale leseprøvene er. De største serverne ved Universitetet i Bergen har en trafikktopp på 30 spørringer i minuttet. Vi opplevde til tider å ha 200 spørringer i sekundet. For å kunne håndtere en så stor belastning må prosjektet investere i en sikker server som tåler denne høye belastningen over korte tidsperioder. Det ble dermed anskaffet en server hvor kravspesifikasjonen ble utarbeidet i samråd med serverdriftsavdelingen ved IT-avdelingen, Universitetet i Bergen. IT-avdelingen ved UiB har 30 års erfaring med serverdrift, og vi har derfor mye å lære av dem. I tillegg til å måtte takle høy belastning, ble fokus satt på at serveren måtte være robust. Serveren er blitt installert av IT-avdelingen med avdelingens RedHat Server-versjon som de har på flere servere. IT-avdelingen sørger også for at det blir tatt sikkerhetskopier av denne serveren hver natt.

9 KLIENTARKITEKTUR

9.1 Teknisk løsning

For at oppgavene skal fungere må funksjonaliteten i oppgaveformatene programmeres, slik at grensesnittet faktisk oppfører seg slik som forventet. Hver enkelt oppgave skal etter elevens besvarelse rapportere svaret til tjener. Teknisk oppbygning av brukergrensesnittet og kommunikasjon med tjener beskrives i dette kapitlet.

Brukergransnittets tekniske oppbygning kan deles i tre deler:

1. Formatutvikling
2. Kommunikasjon med tjener
3. Produksjon

9.1.1 Formatutvikling

Oppgaveformatene er utviklet i Macromedia Flash[®]. Funksjonaliteten i formatene, eller måten brukergrensesnittet oppfører seg på i forhold til elevens valg, er programmert ved hjelp av programmeringsspråket Flash[®] ActionScript[®]. Når kildekoden kompiles/slås sammen med flashfilen (en flashfil har endelsen .fla), blir det generert en enkelt fil (.swf), som er den faktiske oppgaven. Hver oppgave består, før kompilering, av totalt seks moduler med forskjellig innhold. Noen av disse modulene er felles for samtlige oppgaver, noen er felles for hvert enkelt oppgaveformat, mens noen er helt unike for hver enkelt oppgave. Årsaken til at oppgavene ble produsert på denne måten var å kunne gjøre felles endringer for samtlige oppgaver. Dermed kunne en endring gjøres i én modul, før alle oppgavene ble automatisk rekompilert.

Oppgaveformatene er bygget av følgende moduler:

Den grafiske modulen kompiles til det som blir selve oppgaven som elevene ser. Denne modulen inneholder alt av illustrasjoner og andre grafiske elementer.

Rammeverkmodulen inneholder all kode som er felles for samtlige oppgaveformat. Alt som angår plasseringen av grafiske elementer i rammeverket, elementenes ”oppførsel” eller funksjonalitet samt de forskjellige skriftformatene er spesifisert i denne modulen.

Instruksjoner og oppgavetekster som hører til den enkelte oppgave ligger i en egen modul. I denne modulen ligger også koden som sjekker om eleven svarer rett eller galt på oppgavene. Alle svar som blir gitt får en oppgave-id, som er unik for hvert enkelt spørsmål/oppgave. Disse oppgave-id'ene blir opprettet allerede under utarbeidelse av oppgavene. Oppgave-id'ene brukes for å kunne registrere hvilket spørsmål/oppgave eleven har svart på.

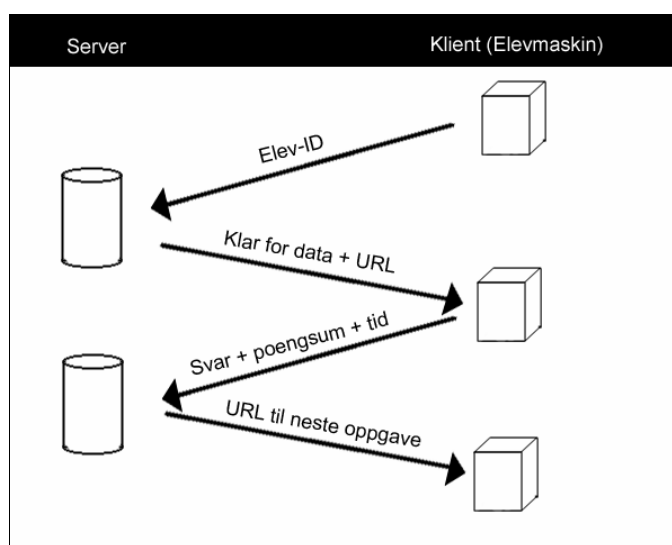
All funksjonalitet som brukes i de forskjellige formatene ligger i en egen modul. Denne modulen er lik for alle oppgaver av samme format.

Alt som har med kommunikasjon med tjener å gjøre ligger for seg, i en egen modul. Her ligger funksjonen som kaller opp tjener og sender elevens besvarelser og poengsum til

databasen. Etter at svarene er sendt får funksjonen beskjed fra tjener om at svar er mottatt. Neste oppgave blir dermed lastet ned til klienten.

9.1.2 Kommunikasjon med tjener

Når en elev har besvart en oppgave, dukker det opp en "ok-knapp" nederst til høyre i brukergrensesnittet. Denne må eleven klikke på for å bekrefte svarene som er gitt. Idet eleven trykker på ok-knappen sender klienten en forespørsel (et kall) til tjener. Dermed sendes elev-id'en. Tjener svarer med beskjed om at den er klar for å ta imot besvarelsen og gir beskjed om hvor data skal sendes (URL). Klienten sender så svarene, poengsum og tiden eleven har brukt på denne oppgaven. Tjeneren mottar disse data og sender neste oppgaveformats URL til klienten som svarer med å laste ned denne.



Figur 2. Kommunikasjon mellom tjener/server og klient/elevmaskin.

Muligheten for små avbrudd i forbindelsen mellom elev-pc/klient og tjener gjorde at det ble lagt inn en repetisjon av kallene til tjener. Slike avbrudd kan komme av lav båndbredde ute på skolene og/eller press på egen tjener. Hvis et kall mislykkes med å nå tjener, eller hvis klienten ikke får svar fra tjener, vil det samme kallet sendes på nytt. Denne prosessen gjentas tre ganger. I utgangspunktet var intervallene mellom kallene satt til 3 sekunder. Dette viste seg å være for lite under gjennomføringen av de nasjonale prøvene. Problemer med å sende og motta raskt nok fra tjener var en feil som ofte oppsto på enkelte skoler der flere elever forsøkte å løse oppgavene over én felles ISDN-linjetjener. Etter tre forsøk på å sende poeng tolket klienten dette som mangel på kontakt med tjener. Etter at intervallene ble økt til 5 sekunder ble frekvensen av feil på grunn av manglende kontakt med tjener betraktelig redusert.

9.1.3 Produksjon

Når alle oppgaveformat og kommunikasjonen med tjener var som den skulle, ble produksjonen av oppgaver intensivert. Det ble produsert ca 500 oppgaver der formatene ble gjenbrukt, mens innholdet ble endret. Hver enkelt oppgave ble fortløpende lastet opp i oppgavebanken slik at de var tilgjengelige for korrekturlesing fra språkforskerne. Testutviklerne hadde også et webbasert system der alle oppgavene var listet. I dette systemet noterte utviklerne produksjonsstatus for hver enkelt oppgave, så alle til enhver tid hadde oversikt over hvor langt i produksjonsprosessen en hadde kommet.

9.2 Sikkerhet

Elevenes besvarelse kontrolleres som nevnt mot fasit i en av modulene som et oppgaveformat består av. Det faktum at dette gjøres på klientmaskinen er en svakhet, og bør i fremtiden gjøres på tjenersiden. Årsaken til dette er muligheten for å dekompile oppgavene og hente ut fasit for hver enkelt oppgave for elever med høy nok datakompetanse. Det er også mulig å finne måter å manipulere svar og poengsum på idet de sendes til tjener. Disse tingene krever imidlertid høy datakompetanse og vil være så ressurskrevende for elevene, at sjansen for at dette skal skje er regnet som svært liten. Oppgavene er dessuten ikke tilgjengelig mer enn en viss periode på dagen i det tidsrommet elevene skal ta prøven, noe som også minsker mulighetene for juks. Eleven har med andre ord begrenset tid på seg til å bryte seg inn i systemet på denne måten.

9.3 CD-ROM

Det har blitt produsert CD-ROM for presentasjon av oppgaveformatene. I utgangspunktet skulle det produseres CD-ROM for utsendelse til skolene slik at elevene skulle kunne øve seg på denne typen prøve. Øvingsoppgavene ble imidlertid lagt ut på egne nettsider i stedet, da dette var mindre ressurskrevende. CD-ROM som er blitt laget har blitt brukt av språkforskerne når de har reist rundt og presentert prosjektet på konferanser både i innland og utland.

9.4 Anbefalinger

Det anbefales at all kontroll av elevens svar mot fasit gjøres på tjener, og ikke på klienten, som det gjøres i dag. Dette på grunn av faren for juks.

Det bør også utvikles flere demosett som elevene kan øve seg på. Pr dags dato finnes der bare ett slikt sett som er tilgjengelig for elevene. Det bør utvikles minst ett sett for hvert nivå.

10 GLOSSARY

BITE – Bergen Interaktive Tester i Engelsk. Prosjektgruppen som utvikler nasjonale leseprøver i engelsk. Språkeksperter.

BITE^{IT} – Den delen av prosjektet som arbeider med system- og testutvikling; IKT-delen av de nasjonale leseprøvene i engelsk.

Brukergrensesnitt – ”møtepunktet” mellom menneske og maskin, som muliggjør interaksjon og gir brukeren mulighet til å utnytte maskinens funksjoner gjennom å styre dens operativsystem.

CEF – Common European Framework. Et resultat av mange års forskning på språklæring, -undervisning og -vurdering i regi av Europarådet. Skalaen er utarbeidet for skrive-, lese-, tale- og lytteferdigheter i et fremmedspråk, og den beskriver hva elevene språklig sett er i stand til å gjøre på tre hovednivåer: Beginner, Intermediate og Native, med to undernivåer for hvert hovednivå: A1, A2, B1, B2, C1 og C2. A1 er nivået for en nybegynner i et språk, mens C2 er så avansert at det er uopnåelig for de fleste på et andrespråk. Mer informasjon her:

http://www.alte.org/can_do/cef.cfm

<http://www.cambridge-efl.org/exam/5level.cfm>

<http://publishing.cambridge.org/ge/elt/booksforteachers/26833/32278/>

<http://culture2.coe.int/portfolio/documents/0521803136txt.pdf>

<http://info.nationaltests.no/cefr.html>

Forprøve – en forprøve som består av ca 20 oppgaver. Forprøvens funksjon er å måle elevens nivå slik at eleven får riktig vanskelighetsgrad på hovedprøven.

HCI – Human Computer Interaction, Et forskningsområde som dreier seg om menneske-maskin-samhandling, eller å utforme datasystemer på en slik måte at de oppleves som støttende og gode hjelpemidler for menneskene som skal bruke dem.

Hovedprøve – etterfølger forprøven, og består av ca 35 spørsmål, påstander eller instruksjoner fordelt på/ i tilknytning til ulike oppgaver.

Kalibreringstesten – En større prøvegjennomkjøring i november 2003, med ca 11.000 elever. Formålet var å beregne vanskelighetsgrad for hver enkelt oppgave, samt å få en teknisk gjennomgang i stor skala.

Klient: en applikasjon/et program som kjører på en datamaskin eller arbeidsstasjon, og som kommuniserer med en tjener/server for å få utført visse operasjoner

Oppgave – en oppgave i forprøven består normalt av bare ett spørsmål, én påstand eller én instruksjon. En oppgave i hovedprøven består vanligvis av flere spørsmål, påstander eller instruksjoner.

Oppgavebank – alle oppgaver som er produsert i forbindelse med leseprøvene. Ut fra denne oppgavesamlingen velges oppgaver til oppgavesettene.

Oppgaveformat – en mal for en oppgavetype, for eksempel et True/False-format, som viser det utseende og den funksjonalitet som er vanlig i den aktuelle oppgavetypen.

Oppgavesett – et oppgavesett består av et bestemt antall Forprøveoppgaver eller et bestemt antall Hovedprøveoppgaver. Oppgavesettene har ulik vanskelighetsgrad.

Pilottester: Utprøving av et ikke helt ferdig produkt/program

Prototyp: Uferdig produkt/program

Tjener/server: en datamaskin i et nettverk som styrer flere andre ressurser/datamaskiner

URL: Uniform Resource Locator, den globale adressen til nettsider (dokumenter og andre ressurser) på internett/World Wide Web.

Bergen, 31. juli 2004

Prof. Konrad Morgan
Stedfortredende prosjektleder

11 APPENDIX

Vedlegg: Brev til skolene v. rektor



InterMedia

NYE MEDIER • NETTBASERT LÆRING

(skolenavn)
2004
(adresse)

(poststed)

Bergen 17. august

Saksnr.200400111/1/IM/CJH

Informasjon til rektor/kontaktperson for Nasjonale prøver i engelsk, 10. klasse

Det skal i år arrangeres Nasjonale prøver ved alle landets skoler. Nasjonale prøver i engelsk avholdes i tidsrommet 13. – 23. april 2004, mellom kl 08.00 og kl 16.00 (norsk tid).

Vedlagt følger brukernavn og passord for skolens kontaktperson. Disse gir tilgang til nettsidene <http://www.nationaltests.no>. Her skal dere registrere klassene og antall elever i hver klasse som skal delta i de nasjonale prøvene i engelsk. Deretter genereres det brukernavn og passord til hver elev. Her må skolen også registrere sin egen identifikator for hver elev. Dette er nødvendig for å kunne koble resultatene fra prøvene tilbake til den enkelte elev. NB! Dere kan ikke benytte elevenes personnummer til dette.

Det forutsettes at skolens kontaktperson logger seg på og foretar disse registreringene snarest mulig etter mottagelsen av dette brevet.

På de samme nettsidene er det mulig å teste om skolens datautstyr er godt nok til å gjennomføre de nasjonale prøvene i engelsk. Det er svært viktig at dette blir gjort, da leseprøvene skal foregå over Internett. Skolen må sørge for at alle datamaskinene som skal brukes av elever under prøvene er testet på forhånd. Når dette er gjort kan vi hjelpe til med å rette opp eventuelle problemer dere måtte oppdage, i god tid før selve gjennomføringen av de nasjonale prøvene.

Vær oppmerksom på at skolens kontaktperson har et eget brukernavn og passord, som følger nedenfor. Kontaktpersonene har tilgang til informasjon som kun skal være tilgjengelig for den respektive skole; informasjon som er av mer sensitiv karakter. De tildeles derfor hver sitt brukernavn og passord.

Brukernavn for kontaktperson: (flettefelt for br.navn settes inn her)
Passord for kontaktperson: (flettefelt for passord)

På nevnte nettsted <http://info.nationaltests.no> finner dere mer informasjon om Nasjonale prøver i engelsk. Blant annet finnes det egne sider med informasjon til lærerne. Disse sidene er passordbeskyttet, men alle lærere benytter samme brukernavn og passord (se nedenfor).

Brukernavn for lærere: teacher
Passord for lærere: pokemon

Dersom dere har spørsmål vedrørende prøvene, finnes kontaktinformasjon på nettsidene, <http://info.nationaltests.no>. Vi ønsker dere lykke til med gjennomføringen av Nasjonale prøver i engelsk.

Med vennlig hilsen,



Konrad Morgan
Forskningsleder Prof.