

Automatiserte tester

Svend Andreas Horgen
Avdeling for informatikk og e-l ring, H gskolen i S r-Tr ndelag
November 2007
L restoffet er utviklet for kurset Pedagogisk bruk av Moodle

Resym : I denne leksjonen skal introdusere litt teori knyttet til automatiserte tester (flervalgstester). Vi ser p  motivasjonen for   bruke slike, og noen tema knyttet til det   lage gode sp rsm l. Vi kommer ogs  inn p  Blooms taksonomi og introduserer en helhetlig referansemodell som kan brukes for   lykkes med   lage gode tester.

1.1	HVORFOR AUTOMATISERTE TESTER?.....	2
1.1.1	Begrepsavklaring: automatiserte tester, vurdering og evaluering.....	2
1.1.2	Utbredelse og bruk.....	3
1.1.3	Formativ og summativ vurdering.....	3
1.1.4	Motivasjonen for � bruke automatiserte tester	4
1.1.5	Gammel skepsis? Blooms taksonomi	5
1.2	HVORDAN LYKKES MED AUTOMATISERTE TESTER – FIRE VIKTIGE FASER.....	7
1.2.1	Utformingsfasen (1).....	8
1.2.2	Distribusjonsfasen (2).....	8
1.2.3	Vurderingsfasen (3).....	9
1.2.4	Forbedringsfasen (4).....	9
1.3	LITTERATURLISTE OG REFERANSER.....	10

1.1 Hvorfor automatiserte tester?

Flervalgstester (engelsk: multiple choice) er ikke noe nytt. De fleste kjenner til testformen fra førerprøven, hvor en må ta stilling til en rekke utsagn og velge det svar en mener er riktig. Slike tester har vært mye brukt i USA, og gir gjerne negativt ladede assosiasjoner: effektivitet, drilling av faktakunnskap, ”papegøyevurdering” [Olsen 2007].

Automatiserte tester – brukt på en god måte og med gjennomtenkte spørsmål – vil utgjøre en god og svært effektiv måte å teste studentenes kunnskap på. Det er mulig å benytte tester til å beskrive kunnskap, forståelse og evner på en god måte.

1.1.1 Begrepsavklaring: automatiserte tester, vurdering og evaluering

Hva legger vi så i begrepet *automatiserte tester*? Vi tenker på en test som har spørsmål og noen veldefinerte alternativer. En skal ta stilling til hvilke(t) alternativ(er) som er riktige, og så krysse av eller på annen måte angi/velge sitt svar. Digitale testverktøy i dag tillater bruk av flere typer spørsmål. Uavhengig av hvilke spørsmålstyper som inngår i en slik test, vil spørsmålene ha veldefinerte svar som kan rettes ved å anvende en nøkkel (fasit). Retting går fort fordi nøkkelen kan sammenliknes med testkandidatens svar. Rettingen kan skje automatisk ved bruk av digitale verktøy. Da blir det også mulig med automatisert tilbakemelding og automatisk oppsummering av resultatet. Med automatisert test, mener vi derfor en digital test som har spørsmål med veldefinerte svar som kan rettes automatisk. Vi snakker stort sett om automatiserte tester, men du vil også kunne finne ordene flervalgsprøve, flervalgstest eller multiple choice test i litteraturen. Tenk på det som ulike ord på samme begrep.

En test lages av noen, og tas av noen andre. Vi omtaler ofte den som lager en test som ”læreren”, men det trenger selvsagt ikke bare være lærere som lager tester. Tilsvarende omtaler vi ofte den som tar en test som ”student” (noe som skyldes at forfatterne underviser studenter til daglig i høyere utdanning). Vi kunne like gjerne benyttet ”elev” eller kanskje heller ”den lærende” eller ”test-subjektet” eller ”test-kandidaten”.

Læring oppstår blant annet i det studenten bearbeider informasjon og omformer denne til kunnskap. For å påvise læring er det nødvendig å måle kunnskapsendringen. Vi må vurdere studenten. Noen kaller dette for evaluering, men vi velger her å skille begrepene fra hverandre og omtaler *vurdering* som det å vurdere kunnskap, mens *evaluering* blir det å undersøke noe (for eksempel å lage et spørreskjema for å finne ut hvor fornøyd noen er med noe). Vurdering og evaluering er for øvrig nært beslektet.

Vurdering av studentenes kunnskap er viktig i alle læreprosesser. Underveis (som motivasjon og rettesnor for studentene og for at læreren kan følge utviklingen), og ved avsluttet kurs (for at læreren kan gi en sluttvurdering av hva studentene har oppnådd).

Det er mange metoder som kan brukes for vurdering, og det er fullt mulig å benytte tradisjonelle metoder også når læringstilbudet administreres via et LMS (for eksempel Moodle). I vår sammenheng har vi spesiell interesse for de nye mulighetene og utfordringene som åpner seg når vi tar i bruk nye verktøy på Internett.

De fleste LMS-systemer støtter ulike vurderingsformer. De vanligste er antageligvis oppgaveinnlevering, flervalgstester og mappevurdering. Når det gjelder flervalgstester (automatiserte tester), åpner dette for en rekke muligheter som berører mange kjerneområder innen IKT-støttet læring: umiddelbar tilbakemelding, god oppfølging av den enkelte student,

effektivitet, gjenbruk, individtilpasning og kompetansekartlegging, for å nevne noe. Samtidig vil slike tester utfordre læreren på mange plan. Dette kommer vi tilbake til snart.

1.1.2 Utbredelse og bruk

Tester på Internett kan gjøres varierte, spennende og utfordrende. De kan gjøre mer lik ”games and puzzles” enn nerveslitende, tradisjonelle tester [Horton 2000]. Tester kan virke mer motiverende og mindre stressende for studentene.

Den mest åpenbare bruken av tester (for mange) er den *summative*: for å vurdere kunnskap og sette en karakter. Det er unektelig mange utfordringer ved bruk av slike tester i summativ vurdering. Hvordan lage gode spørsmål som tester utover ren faktakunnskap, er et av mange ikke-trivielle spørsmål (som dette kurset skal forsøke å gi innsikt i). Det er også fordeler. Effektiv, automatisert retting er en åpenbar gevinst. Helt objektiv retting, hvor verken studentens personlighet eller håndskrift virker inn på resultatet, er en annen, og vi kunne nevne mye mer. Den mest iøynefallende fordel sett fra læreren side, er at slike tester vesentlig reduserer den tiden læreren bruker til å rette besvarelsene. Jo flere studenter som tar testen, jo mer dramatisk blir denne effekten. Læreren får dermed friggitt tid som kan brukes til å støtte studentene på andre måter gjennom læringsprosessen. Dette er også til fordel for studentene.

Stadig flere lærere oppdager at automatiserte tester kan være et flott og smågenialt pedagogisk verktøy som det er verdt å ha i sin verktøykasse. Automatiserte tester innehar nemlig ikke bare et stort potensiale for å realisere mer effektiv vurdering, men også en mulighet for å tilby *bedre læring*. Bruk av automatiserte tester underveis i læringsprosessen åpner for mange gode pedagogiske aktiviteter. Selvtester med umiddelbar tilbakemelding kan bekrefte kunnskap og forståelse, og hjelpe den som skal lære til å fokusere på de riktige tingene og/eller til å identifisere sine kunnskapshull. Tenk på en test som en mulighet for å oppnå automatisert veiledning, og kanskje også til en viss grad individtilpasning. Slik bruk omtales gjerne som *formativ* bruk av tester, og gjennomgås i detalj senere i kurset. I tillegg vil vi se nærmere på ulike kreative opplegg der tester inngår, for eksempel bruk av en test før en læringsøkt (time, forelesning, selvstudium) for å øke motivasjonen og stimulere til læring.

Automatiserte tester har for mange lærere blitt ekstra relevante de siste årene. Internett og moderne IKT-teknologier åpner for nye muligheter for læring. I dag benytter de fleste utdanningsinstitusjoner en eller annen form for læringsplattform (LMS). I Norge er plattformer som Fronter, it's learning, Blackboard, Moodle og egenutviklede systemer i bruk. Disse har (til dels gode) verktøy for å lage, gjennomføre og analysere tester på en enkel måte. Automatiserte tester kan rettes automatisk av systemet, og dermed tilby umiddelbar tilbakemelding til den som tar testen. Testene er gjerne lette å opprette for læreren. Et system kan (i teorien) også støtte gjenbruk av spørsmål på en enkel måte, og hjelpe til med analyse og bearbeiding av resultater i ettertid. Vi skal se nærmere på testverktøyet i Moodle i en egen leksjon.

1.1.3 Formativ og summativ vurdering

Tradisjonelt sett har sluttvurdering med karaktersetning hatt en sentral rolle i norsk utdanning. Studenter grupperes sammen for å lære et emne eller innøve en ferdighet. En eksamen gjennomføres etter en viss tid, vurderes av en eller flere sensorer, og resultatet dokumenteres. Studentene karaktersettes, og fortsetter videre med nye emner.

Summativ vurdering innebærer som regel en konstatering eller anerkjennelse av kunnskapsnivå på et gitt tidspunkt, ofte etter endt læringsforløp. For denne type vurdering

settes dokumentasjonen i sentrum, og dialogen er tradisjonelt sett tilnærmet enveis i sin kommunikative form: Studenten tar en eksamen (leverer et produkt) og får en karakter eller enkel vurdering tilbake. Summativ vurdering er altså vurdering av læring. Legg merke til at veien til kunnskap som regel krever stor grad av læring i forkant fra studentens side, men ikke nødvendigvis. Det kan være at studenten kan stoffet fra før og bare skal vurderes for å bli sertifisert/påvise kunnskap.

Formativ vurdering skal være til hjelp for læringen, og dialogen mellom student og den som vurderer står i sentrum. Den vil være veiledende i formen, og er som regel noe som skjer underveis i selve læringsprosessen [wikipedia]. Formativ vurdering kan gi grunnlag for karaktersetning, men ikke nødvendigvis. Den formative vurderingen kan anses å være et viktig element i læringsprosessen [Lauvås 2004]. Dette gjelder også i stor grad for e-læring.

Kvalitetsreformen satte fokus på bruk av underveisvurdering for å bedre læringen, øke gjennomstrømningen av studenter og effektivisere ressursbruken i norsk utdanning [KR 2001]. For å lykkes i gjennomføringen av Kvalitetsreformen er det viktig å lykkes med formativ vurdering. Dette er en utfordring for mange lærere, både tidsmessig og pedagogisk.

Legg merke til at formativ vurdering underveis i læringsprosessen gir et større grunnlag for å identifisere kunnskapshull, forutsatt at læringsmålene er kjent. Via et LMS er det lettere å effektivisere den formative vurderingen, samtidig som studenten gjerne er mer "alene" om sin læring. Det er dermed nærliggende å påstå at formativ vurdering er en viktig suksessfaktor i e-læring hva læringskvalitet angår. Slik vurdering blir både et nyttig redskap for studenten (å bli bevisst på egen læring) og for læreren (å kvalitetssikre og følge opp læring).

Oppsummert kan vi si at summativ vurdering tar utgangspunkt i et endelig produkt, mens formativ vurdering er opptatt av prosessen.

1.1.4 Motivasjonen for å bruke automatiserte tester

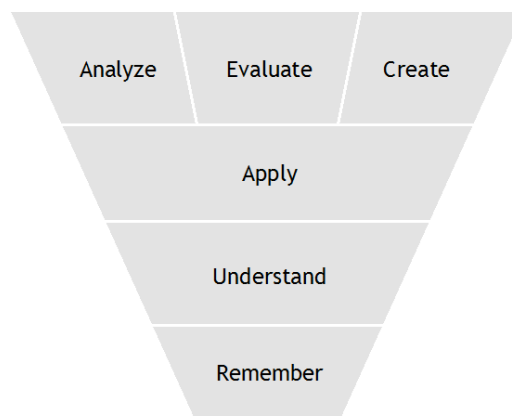
Automatiserte tester passer godt inn i en kontekst hvor læringstilbudet støttes av en læringsplattform som Moodle, og kan brukes i både formativ og summativ vurdering. Hva er motivasjonen for å bruke automatiserte tester? Hva er målet med testing generelt? Her er noen viktige punkter:

- Tester gir mulighet for *umiddelbar* tilbakemelding av studentens kunnskap, forståelse og evner. Dette kan stimulere til selvstudium, et jevnere læringsforløp og motivere studenten.
- Selvtesting kan brukes for å bekrefte hva en (ikke) kan og veilede videre i læringsprosessen. En digital test åpner for fleksibel bruk, der studenten potensielt sett selv kan teste seg når det måtte passe. Tester kan dermed også gjøre selvstudium mer attraktivt og kan bidra til et jevnere læringsløp.
- Tester kan dekke deler av pensum som en ellers ikke har tid/anledning til å kartlegge (bredde i pensum) på en avsluttende eksamen, hvor man gjerne må nøye seg med å plukke ut noen sentrale emner.
- Repetisjon av tidligere tema kan effektivt gjennomføres slik at videre læring blir best mulig.
- Vurderingen er i utgangspunktet 100% objektiv. Resultatet påvirkes ikke av studentens håndskrift og evne til å uttrykke seg muntlig eller skriftlig. Spesielt for skrivesvake studenter kan dette være en stor fordel. Det er klart at noen fag egner seg bedre enn andre, men også i for eksempel norskfaget kan tester ha mye for seg som supplement til for eksempel stil- og drøftingsoppgaver.

- Spørsmål kan utformes til å vurdere kunnskap, evner, ferdigheter og forståelse i henhold til Bloom's taksonomi (utdypes senere i denne leksjonen).
- Det å måtte velge blant flere alternativer, betyr ikke nødvendigvis at en kan gjette seg til riktig svar, og heller ikke at det er umulig å teste delvis kunnskap eller prosessen underveis. Det kan være utfordrende å lage spørsmål som møter slike krav, men ikke umulig.
- En *kreativ* bruk av tester, kan supplere øvrig pedagogikk og gi interessante bieffekter.
- Effektivitet kan oppnås som følge av automatisert retting og veiledning, både for lærer og student.
- Mulighet for gjenbruk av tester og enkelt-spørsmål kan sikre både effektivitet og kvalitet. Statistikk over tidligere resultater kan avsløre at spørsmål er gode eller dårlige.
- Individualisering kan oppnås slik at den lærende kan bli tilbudt et skreddersydd læringsforløp. Dette er et utfordrende, men spennende tema som vi trolig vil se mye mer av i fremtidens skole.

1.1.5 Gammel skepsis? Blooms taksonomi

På tross av mange fordeler, er det nok mange som har en viss skepsis til ”multiple choice tests” som et onde fra et USA. Hva testes og hvor beskrivende er et resultat på den faktiske kunnskap og forståelse? Ikke alle er enige i at automatiserte tester kan tilby en tilfredsstillende måte å teste studentenes ferdigheter på utover ren faktakunnskap. En ser for seg at effektivisering av eksaminasjon står i motsetning til å teste studentenes evne til analyse og anvende kunnskap. Mange er derimot uenige, og mener at *riktig utformede* tester kan teste ferdigheter på de fleste trinnene i Blooms ferdighetstrapp [Bloom 1956].



Figur 1: Blooms taksonomi

Noen stiller seg også tvilende til om slike tester virkelig er ressursbesparende for læreren. Det er ingen tvil om at det gir en stor besvarelse når testene skal rettes og studentene skal ha tilbakemelding. Men til gjengjeld krever det mye kompetanse, omtanke og tid av læreren når testene og spørsmålene skal utformes. Svaret på om flervalgstestene er ressursbesparende, blir derfor et regnestykke hvor antall studenter og mulig gjenbruk av testen blir viktige faktorer. I dette regnestykket bør også gevinster knyttet til læring, motivasjon og umiddelbar, skreddersydd veiledning, tas med.

Blooms taksonomi av pedagogiske mål kategoriserer evner som kan vurderes i ulike nivåer, der det er mer krevende å nå opp mot høyere nivåer. Tabellen viser hvilke ferdigheter som kan vurderes for ulike kognitive kompetansenivåer.

Blooms taksonomi for læring [Bloom 1956]
<p>1. Kunnskap: Evne til å huske informasjon og gjengi faktakunnskap. Eksempel på spørsmålsord: Nevn, list opp, definer, beskriv, ...</p>
<p>2. Forståelse: Evne til å tolke mening og reformulere kunnskap med egne ord. Eksempel på spørsmålsord: tolk, diskuter, oppsummer, klassifiser, sammenlikn, oversett...</p>
<p>3. Anvendelse: Evne til å bruke kunnskap, teknikker og metoder i nye situasjoner for å løse problemer. Eksempel på spørsmålsord: bruk, demonstrer, vis, relater, ...</p>
<p>4. Analyse: Evne til å bryte ned (dekomponere) informasjon, for så å identifisere og gjenkjenne mønstre, komponenter og sammenhenger. Eksempel på spørsmålsord: analyser, ordne, forklar, utled, konkluder, trekke slutning, sammenlikn, kategoriser, ...</p>
<p>5. Syntese: Evne til å generalisere (kombinere) gammel informasjon til ny kunnskap eller foreslå alternative løsninger. Kunne forutsi og dra konklusjoner. Eksempel på spørsmålsord: samordne, modifiser, finn opp, modeller, utform, lag, planlett, formuler, arranger, ...</p>
<p>6. Evalering: Evne til å presentere og forsvare meninger ved å bedømme informasjon og idéer. Sammenlikne og avgjøre forskjell mellom fenomener. Vurdere. Eksempel på spørsmålsord: bedøm, evaluer, vurder, forsvar, ranger, konkluder, ...</p>

Normalt anser en de tre siste nivåene for å være ”læring i dybden”. Spørsmål av type “Hva er ...?” kjennetegner ofte de tre første trinn i modellen. Spørsmål som “Hvorfor er det slik?” beveger seg oppover mot analyse og syntese-nivå, der innsikt og forståelse skal avdekkes.

Det er viktig å påpeke at modellen er relativ, avhengig av kontekst og ferdigheter. Det som krever dyp innsikt og oppleves som utfordrende for en nybegynner, kan være ren rutinegjenkjenning for en ekspert [Brown 2001].

I Moodle er det spesielt enkelt å vurdere kandidaten på ulike nivåer av Blooms taksonomi. Det er ikke noe vanskeligere å realisere Blooms nivåer i et digitalt miljø enn i et ikke-digitalt, men det kan være en utfordring å gjøre det mer effektivt. Bruk av såkalte ”stimuli” som videosnutter, lyd, artikler på Internett og liknende, er lett å innlemme i spørsmålene i digitale tester. Poenget for læreren er å tenke gjennom på hvilket nivå en ønsker å vurdere, og så velge en egnet metode.

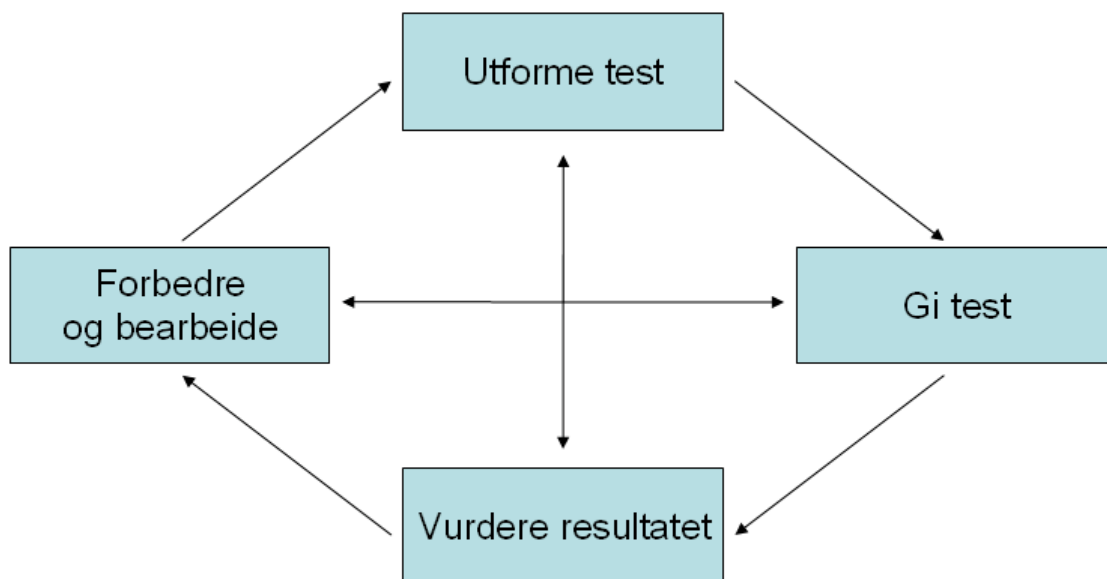
Det er vanligvis et mål å teste studentenes kognitive ferdigheter gjennom hele stigen i Blooms taksonomi (gjengi kunnskap med egne ord, anvende tillærte kunnskaper, analysere, syntetisere og vurdere) og ikke bare teste evnen til å gjengi fakta (det nederste trinnet). Dette er fullt mulig i flervalgstester, men mye mer krevende enn å oppnå en ren faktatesting.

Det er alltid et mål å teste studentenes ferdigheter i et fagfelt, og ikke deres evne til å ta flervalgstester. Her er det lett å feile. Hvis vi ikke passer på, kan vi ved vår måte å utforme testen på enten gjøre det lett for kandidaten å gjette svaret, eller vi kan lure dem med mangelfull informasjon og snedige spørsmål til å svare feil.

1.2 Hvordan lykkes med automatiserte tester – fire viktige faser

Som vi har vært inne på kan vi oppnå en rekke fordeler ved å benytte automatisert testing i undervisningen. Men det er mange ting å tenke over hvis vi skal lykkes med å sikre høy kvalitet i alle ledd i forbindelse med testing. Vi tar her opp noen overordnede, initielle utfordringer og oppgaver som bør tenkes gjennom og planlegges ved bruk av automatisert testing.

Ved å se nærmere på hva arbeidet med testen består i, kan en dele det inn i ulike faser. Vi kan dele inn i fire ulike faser, illustrert i Figur 2. Ved å identifisere kritiske spørsmål og rette fokus mot ulike oppgaver innenfor hver fase, kan kvaliteten totalt sett økes og dermed også mulighetene til å lykkes. Det er også mulig å fordele arbeidet i de ulike fasene på flere personer eller systemer.



Figur 2: Fire iterative faser for å lage gode tester.

De fire fasene er:

1. Utforme testen. I denne fasen bestemmer vi oss for hva slags test vi skal lage og vi utformer både selve testen (antall spørsmål, hvordan den skal brukes, hva slags tilbakemeldinger vi skal gi osv.) og vi utformer hver av spørsmålene.
2. Gi test. I denne fasen informerer vi om testen, vi distribuerer vi testen til studentene og vi tar imot svarene.

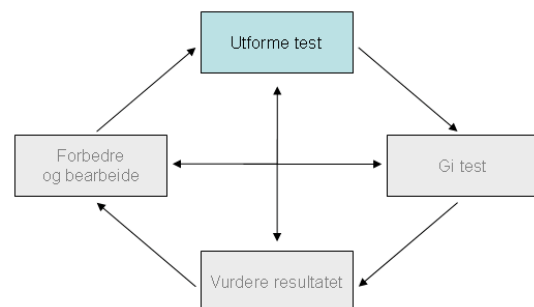
3. Vurder resultatet. I denne fasen skal vi tolke resultatet av testen og vi skal omsette oppnådd resultat til et sluttresultat for studenten.
4. Forbedre testen. I denne fasen skal vi vurdere godheten av testen, basert på oppnådd resultat, tilbakemeldinger fra studentene og egne analyser, dette er grunnlag for å forbedre testen.

Legg merke til at prosessen mot å utvikle gode tester er iterativ. Etter avsluttet syklus bør fasene gjentas, med det formål å stadig forbedre testen.

1.2.1 Utformingsfasen (1)

Testen må først og fremst utformes basert på mål, studentforutsetninger og fagdidaktikk. Basert på valgt strategi, handler utformingsfasen om å:

- Angi egenskaper ved testen.
- Lage gode spørsmål. Dette er en kompleks affære som vi ikke kan gå så nøye inn på i dette kurset. Se www.auto.hist.no for et annet kurs som går i dybden på flere aspekter knyttet til testing.
- Ha en gjennomtenkt strategi for informasjon og tilbakemelding.
- Bestemme grad av gjenbruk: Helt, delvis, eller ren nyskriving?
- Oppnå effektivitet.
- Foreta tilstrekkelig kvalitetssikring.



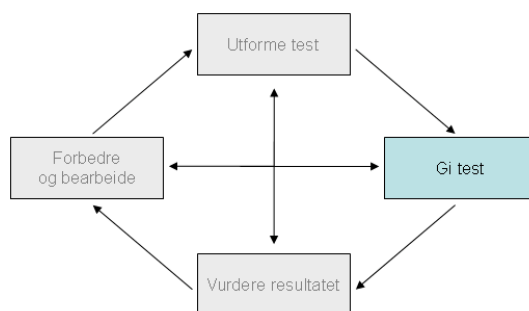
Målet er som regel å teste kunnskap, evner og forståelse knyttet til fagfeltet. En må unngå at utfallet av testen avhenger av hvor dyktig en student er til å besvare slike tester, bruke logikk eller andre teknikker for å eliminere gale alternativer. Måten testen er utformet på, spørsmålsformuleringene, mangelfull informasjon og liknende, er faktorer som kan bidra til at testen oppleves urettferdig. Disse temaene vil dukke opp senere i dette kapittelet.

Hvordan kan studenten prioritere spørsmål i tilfelle tidsnød oppstår? Blant tiltak for å gjøre det lettere å velge, er å ha en oversikt over alle spørsmål som inkluderer informasjon om maksimal uttelling for hvert spørsmål, gruppere spørsmålene på tema eller etter vanskelighetsgrad, eller vise stikkord foran spørsmålene.

1.2.2 Distribusjonsfasen (2)

Når en test er laget, må den distribueres slik at mottakeren kan ta/gjennomføre testen. I distribusjonsfasen er det noen overordnede hensyn å ta:

- Hvordan skal testen initieres, og av hvem?
- Hvor mange ganger og hvor ofte kan testen tas?
- Hvordan kan juks forhindres (dersom det er et tema)?

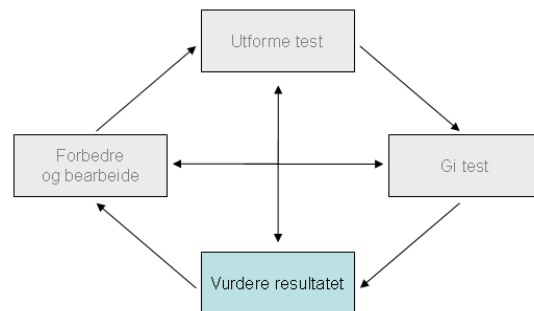


- Hvem gir informasjon, og hva må en informere om?

Noen studenter vil prøve jukse om de har anledning, og vi vil gjerne forhindre det. Vi kan jo si at juks er verst for studenten, fordi det egentlige målet for studenten burde være læring og ikke et kunstig godt resultat. Men så enkelt er det ikke, og det er derfor allikevel nødvendig å vurdere tiltak for forhindre juks.

1.2.3 Vurderingsfasen (3)

Det er ikke trivielt å vurdere et testresultat. Basert på en poengstrategi skal svarene belønnes. Deretter må poeng tolkes og omsettes til et resultat som eventuelt rapporteres til studenten. Hva sier resultatet om studentens ferdigheter? Skal resultatet foreligge i form av en karakter, en poengsum eller godkjent/ikke godkjent? Når skal resultatet foreligge?



Det fins ingen absolutte svar på disse spørsmålene. Alt avhenger av de mål og egenskaper for testen som er bestemt i de tidligere fasene.

Vi har tidligere sett på viktigheten av å informere i forkant av testen, slik at studenten kan ta testen med en best mulig forståelse for hvordan de faktiske forhold fungerer. Det er også viktig å tenke gjennom hvilken informasjon som skal vises til studenten under og etter fullført test.

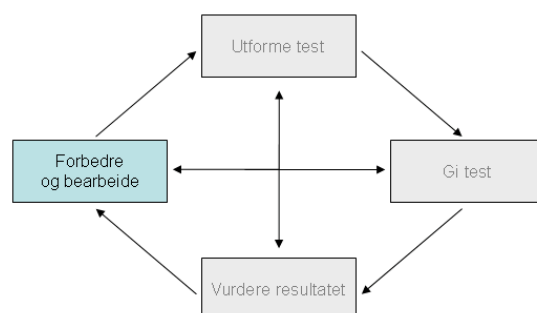
Tilbakemelding om oppnådd resultat er viktig for motivasjon og læring, men når og i hvilken form skal tilbakemelding forekomme? Det er ikke gitt at umiddelbar tilbakemelding er det beste i enhver situasjon. Et testverktøy kan typisk settes opp til å gi tilbakemelding (poengsum/karakter/resultat) umiddelbart, etter en viss tid, når en bestemt poengsum er oppnådd, eller liknende. De fleste verktøy gir også mulighet for kommentarer basert på rett/galt svar. Disse kan typisk vises etter hvert spørsmål eller i etterkant av fullført test. Det kan også være ønskelig å ikke gi tilbakemelding i det hele tatt.

Til slutt: Det er ikke alltid alle spørsmål i en test har samme poengsum til utdeling. Noen ganger kan det være riktig å ta spesielle hensyn slik at de som svarer riktig/galt på vanskelige spørsmål, på tross av gitt poengsum, kan justeres manuelt opp eller ned.

1.2.4 Forbedringsfasen (4)

Den siste fasen i modellen vår, er interessant og ikke minst veldig viktig. Bearbeiding og analyse av resultater fra en utført test gir et godt grunnlag for forbedring i neste runde.

Forbedringspotensialet kan avdekkes ved hjelp av en objektiv analyse basert på resultatene fra enkeltstudenter eller større grupper av studenter. Det er derimot krevende, både teknisk og tidsmessig, å foreta gode analyser og ikke minst å tolke resultatene i ettertids. Samtidig er denne fasen viktig for å kunne oppnå god kvalitet. Det er derfor essensielt å ha test-verktøy som støtter opp om



forbedringsfasen. Vi skal se på hvordan Moodles testverktøy kan brukes til dette i en annen leksjon.

1.3 Litteraturliste og referanser

Denne leksjonen har tatt opp noen aspekter ved automatiserte tester. Det er utviklet et kurs (tilsvarende med dette) som heter ”Pedagogisk bruk av automatiserte tester”. Se mer om dette kurset dersom du er interessert i å lære mer, på nettstedet www.auto.hist.no.

Her er en referanseliste som har gitt grunnlag for denne leksjonen:

- [Olsen 2007]: Flervalgsoppgaver – en papegøye fra Amerika? Rolf Vegar Olsen, Institutt for lærerutdanning og skoleutvikling, UiO. Artikkel tilgjengelig fra Internett: http://www.utdanning.ws/templates/udf_15085.aspx.
- [Horton 2000]: Horton, William: *Designing Web-based Training* kap. 7. New York, Wiley&sons 2000. Denne boken går svært grundig gjennom ulike spørsmålstyper og når de egner seg, diskuterer formulering av spørsmål og alternativer, viser oppbygging av informasjon til studentene etc.
- [Brown 2001]: “Assessment: A Guide for Lecturers”, George Brown, 2001, http://www.heacademy.ac.uk/resources.asp?process=full_record§ion=generic&id=3
- [wikipedia]: <http://en.wikipedia.org/wiki/Assessment>
- [Lauvås 2004]: “Mappevurdering – av og for læring. Med eksempler fra helse- og sosialfag”. Taasen, I., Havnes, A. og Lauvås, P. Gyldendal, Oslo 2004.
- [KR 2001]: “St.meld. nr. 27: Gjør din plikt - krev din rett”, 2000-2001, <http://odin.dep.no/kd/norsk/dok/regpubl/stmeld/014001-040004/hov005-bn.html>
- [online-tutoring]: Online Tutoring e-Book, chapter 4 New Assessment Strategies.
- [Bloom 1956]: “Bloom’s Taxonomy of Educational Objectives”, Bloom, 1956.
- [Race 2001]: “Using feedback to help students learn”, Phil Race, 2001, http://www.heacademy.ac.uk/resources.asp?process=full_record§ion=generic&id=432