

Uniped, årg. 38, nr. 4-2015, s. 336-344
ISSN online: 1893-8981

FAGFELLEVDERT ARTIKKEL

Kombinert summativ og formativ vurdering i matematikk

Knut Bjørkli
Høgskolelektor
Høgskolen i Sør-Trøndelag
knut.bjorkli@hist.no

Ketil Arnesen
Høgskolelektor
Høgskolen i Sør-Trøndelag
ketil.arnesen@hist.no

SAMMENDRAG

Artikkelen framlegger resultater fra et undervisningsopplegg i matematikk for 1. års ingeniørstudenter ved HiST høsten 2013, hvor korte, digitale flervalgstester inngikk som en del av den summative vurderingen i matematikkfaget. Testene ble gjennomført på de ordinære forelesningene, hvor studentene brukte det HiST-utviklede testverktøyet Peer Learning Assessment System (PeLe) til å besvare testene med sine egne mobile enheter. Undervisningsopplegget foregikk over 2 x 45 minutter, og i den første, summative halvdel ble studentene den digitale testen individuelt. Testverktøyet gjorde resultatene umiddelbart tilgjengelig for læreren, som kunne bruke svarfordelingene som utgangspunkt for læringsaktiviteter i oppleggets andre, formative del. Læringsaktivitetene la vekt på samarbeid mellom studentene og ga studentene muligheten til å vise at de hadde lært av egne, og hverandres, feil. Læringseffekten av undervisningsopplegget ble målt i form av resultater på en ordinær, skriftlig matematikkeksamen, hvor kontrollgruppa besto av studenter som fulgte et identisk matematikkpensum, men som ikke gjennomførte digitale underveistester. Testgruppa fikk signifikant bedre resultater enn kontrollgruppa på en tradisjonell skriftlig matematikkeksamen etter å ha gjennomført undervisningsopplegget med kombinerte summative og formative tester.

Nøkkelord

formativ vurdering, summativ vurdering, mobillæring, ikt-støttet læring, læringseffekt.

ABSTRACT

We present results from a study on an introductory mathematics course for first-year engineering students at HiST in the autumn of 2013, in which digital multiple-choice tests were used as a part of the summative assessment in this


UNIVERSITETSFORLAGET

 idunn.no
Nordiske tidsskrifter på nett

This article is downloaded from www.idunn.no. © 2015 Knut Bjørkli og Ketil Arnesen. This is an Open Access article distributed under the terms of the Creative Commons CC-BY 4.0 License (<http://creativecommons.org/licenses/by/4.0/>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

subject. Six lecture sessions throughout the term were used for these tests, which were run using a HiST-developed tool called Peer Learning Assessment System (PeLe). Tests comprised 2 x 45 minute-sessions: in the first, summative part; students answered a digital test individually, the result of which would count towards their final grade. The PeLe tool immediately made the results available to the teacher, who would then use the data to plan the formative review phase in the second half. During the review, students engaged in peer learning activities, and the PeLe tool enabled students to show their ability to learn from their mistakes. Learning outcomes were measured in the form of an ordinary, constructed response exam at the end of the term. The control group was made up of students who followed an identical curriculum, but who did not participate in digital tests. The test group performed significantly better on the final exam than did the control group, after having taken part in combined summative and formative tests throughout the term.

Keywords

formative assessment, summative assessment, mobile learning, ict enhanced learning, learning outcomes.

INNLEDNING

Begrepene formativ og summativ vurdering

Et klart skille mellom formativ og summativ vurdering ble formulert på 1960-tallet. Summativ vurdering kvantiserer graden av læring eller måloppnåelse i form av en karakter eller en poengsum, og utgjøres i praksis av standpunktprøver, midtsemesterprøver, eksamen, karaktergivende innleveringer osv.

Det har vært endel diskusjon om hva begrepet formativ vurdering skal innebære (Bennett, 2011; Cech, 2007). Enkelte definerer formativ vurdering som et sett med tester som tas underveis i semesteret, og som har til hensikt å avdekke lærestoff elevene har problemer med. Testresultatene brukes av læreren til å forme videre undervisning, for eksempel ved at elevenes problemområder vies ekstra oppmerksomhet.

Som en nyanseforskjell toner andre ned betydningen av testing i forbindelse med formativ vurdering, og anser formativ vurdering som en prosess med fortløpende tilbakemelding mellom lærer og elev der grunnlaget for tilbakemeldingen ikke nødvendigvis er en test. Det prosessorienterte synet deler imidlertid oppfatningen om at hensikten med formativ vurdering er å forme både elevenes læring og lærerens undervisning, for at elevene i best mulig grad skal nå læringsmålene.

I denne artikkelen bruker vi begrepet formativ vurdering om diagnostiske tester som har til hensikt å kartlegge hva slags lærestoff som flertallet av studentene har problemer med, og som brukes av læreren til å gi elevene umiddelbar tilbakemelding på sin læringsprosess. I tillegg brukes testene til å forme fram-

tidig undervisning, for eksempel ved at testresultatene gjør læreren i stand til å finne andre, mer hensiktsmessige måter å presentere vanskelig lærestoff på.

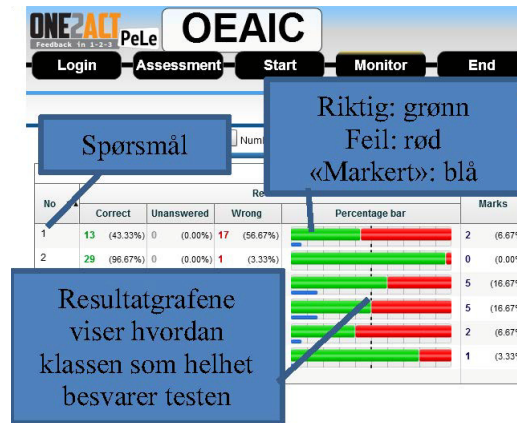
Forskningslitteraturen er sprikende når det gjelder læringseffekten av formativ vurdering – dvs. hvorvidt formativ vurdering gir forbedrede resultater på standardiserte tester og på eksamen. En metastudie av Kluger og DeNisi (1996) fant ingen entydig sammenheng mellom formativ vurdering og økt læring, mens en tilsvarende studie av Black og William (1998) hevdet at formativ vurdering faktisk kan ha en signifikant, positiv effekt på læring.

Testverktøyet Peer Learning Assessment System (PeLe)

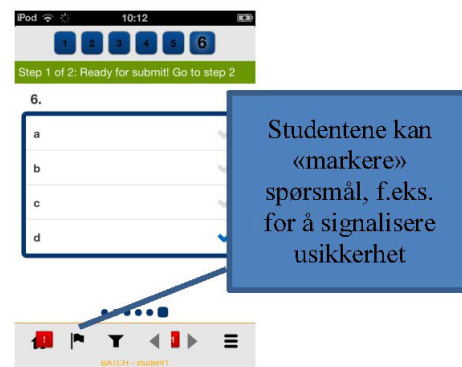
HiST har siden 2009 gjennomført flere internasjonale prosjekter om bruk av mobilteknologi i læring, og forsknings- og utviklingsarbeidet har så langt resultert i en portefølje bestående av tre ulike læringssystemer som anvender mobilteknologi: studentresponssystemet SRS; evalueringsverktøyet Eval, samt vurderingsverktøyet Peer Learning Assessment System eller PeLe, alle tre tilgjengelige fra One2act (2015). Vurderingsverktøyet PeLe er tema for denne artikkelen.

Det eksisterer et utall verktøy som gjør det mulig å gjennomføre kombinerte formative og summative tester – for eksempel har de fleste læringsplattformer, som it's learning, Moodle og Fronter, slike testverktøy. De fleste av disse verktøyene er imidlertid tilrettelagt for asynkrone, elevstyrte vurderinger, i motsetning til lærerstyrte tester som foregår i klasserommet. I tillegg er mange eksisterende testverktøy tilpasset bruk på datamaskiner, og kan være tungvinte å navigere på mobile enheter.

PeLe er et testverktøy for å kjøre flervalgstester over internett på studentenes mobile enheter, slik som bærbar PC; smarttelefon; nettbrett etc. PeLe består av to deler: 1) Lærerklienten, som er installert på lærerens PC. Den brukes i klasserommet til å kjøre og overvåke tester, og hjelper læreren å strukturere læringsaktivitetene i etterkant av selve testen; 2) Studentklienten, som kjører på studentenes mobile enheter i en nettleser eller i en egen app, og som studentene bruker til å besvare testen. Bilder av de to komponentene er vist i *Figur 1* og *Figur 2* under.



Figur 1. Lærerklienten i PeLe som viser hvordan klassen besvarer testen.



Figur 2. Studentklienten i PeLe, som studentene bruker til å besvare testen

Læreren klargjør testen på forhånd, ved å angi antall spørsmål; antall alternativer for hvert spørsmål, samt markere hvilke(t) alternativ som er riktige på hvert spørsmål. PeLe retter automatisk testen, men resultatene blir ikke gjort tilgjengelige for studentene før læreren eksplisitt åpner for dette (vanligvis etter at læringsaktivitetene i etterkant av testen er ferdige).

I løpet av en kort pause lager læreren en skisse for gjennomgangen i etterkant, med utgangspunkt i testresultatene. For eksempel: de testspørsmålene som ble besvart riktig av en viss prosentandel av studentene, er godt egnet for samarbeid og samhandling i gjennomgangsfasen. Testspørsmål som et fåtall av studentene klarte å besvare riktig, kan kreve mer direkte inn gripen fra læreren, i form av en grundig gjennomgang av oppgaven på tavla.

PeLe gjør det mulig for studentene å vise at de har lært av de feilene de gjorde under selve testen, ved at læreren kan la studentene svare på nytt i gjennomgangsfasen, i kjølvannet av læringsaktiviteter hvor studentene får anledning til å diskutere og dele løsningsstrategier med hverandre. En student som svarte feil på testen, men svarer riktig i gjennomgangsfasen, kan få delvis poenguttelling.

HENSIKT

HiST-studenter har et matematikkurs i grunnleggende analyse i første semester, og dette kurset ble valgt for gjennomføringen av et undervisningsopplegg med kombinerte summative og formative tester. Den individuelle effekten av hhv. summativ og formativ vurdering på læringsutbytte har blitt inngående studert i litteraturen, mens effekten av kombinerte formative og summative tester typisk har blitt studert i forbindelse med storskala nasjonale prøver – se for eksempel Looney (2011).

Problemstillingen for studien var som følger: Hvordan påvirker bruken av kombinerte formative og summative tester, som besvares ved hjelp av mobilteknologi, læringsutbyttet? Spesifikt: Svarer studenter som følger et slikt undervisningsopplegg signifikant bedre på en tradisjonell skriftlig matematikkksamensammenliknet med studenter som ikke følger et slikt opplegg?

Ettersom tre store klasser deltok ($n > 100$ i hver klasse), var det mulig å gjennomføre studien på en slik måte at hver student var enten i testgruppen som gjennomførte undervisningsopplegget med kombinerte formative og summative tester, eller i kontrollgruppen. På denne måten kunne man eliminere smitteeffekter som oppstår når en og samme student deltar i både test- og kontrollgruppe.

PROSEDYRE

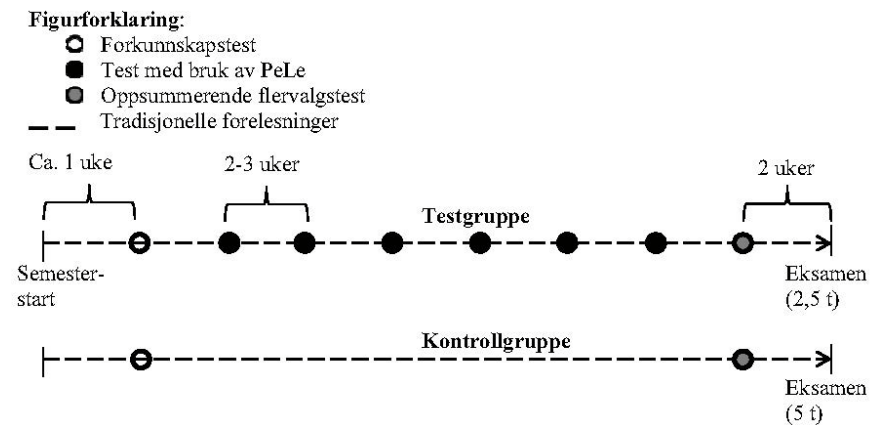
Oppbygningen av studien

Studien ble gjennomført høsten 2013 i et ett-semesterers innføringskurs i matematikk på 10 studiepoeng for 1. års logistikk-, kjemi- og materialteknologistudenter ved HiST ($n = 113$). Det ble definert to kontrollgrupper bestående av 1. års byggingeniørstudenter ($n = 112$) og maskiningeniørstudenter ($n = 96$), som fulgte et identisk pensum, men som ikke brukte noen form for underveisvurdering.

Alle tre gruppene hadde tre 90-minutters forelesninger per uke, men for testgruppa ble seks forelesninger, jevnt fordelt over semesteret, brukt til kombinert summativ og formativ vurdering med PeLe. Så i praksis hadde testgruppa 12 timer mindre forelesning enn kontrollgruppa.

For å avdekke hvorvidt det var signifikante forskjeller i faglig nivå mellom de tre gruppene i utgangspunktet, ble det helt i starten på semesteret gjennomført en standardisert test utformet av Norsk Matematikkråd. Denne testen måler grunnleggende matematikkferdigheter som aritmetikk, algebra, prosentregning og geometri, og ble besvart på papir uten bruk av mobilteknologi.

En tidslinje for studien er gjengitt i *Figur 3* under.



Figur 3: Tidslinje for studien

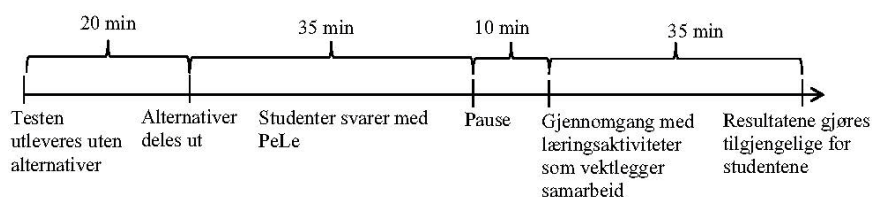
Den oppsummerende flervalgstesten inneholdt spørsmål fra hele pensum, og både test- og kontrollgruppa besvarte denne ved hjelp av PeLe på mobile enheter.

For alle gruppene var den avsluttende eksamen en tradisjonell skriftlig matematikkeksamen som hverken inneholdt flervalgsspørsmål eller tok i bruk mobilteknologi. For kontrollgruppa hadde eksamen en varighet på fem timer, bestående av 24 deloppgaver, som telte 100 % for sluttkarakteren i faget. For testgruppa telte de seks digitale testene underveis i semesteret 60 % av totalkarakteren, og varigheten av den avsluttende eksamen, som telte de resterende 40 %, ble derfor redusert til 2,5 timer. Testgruppas eksamen var derfor ikke identisk med kontrollgruppas, og besto av omtrent halvparten av oppgavene fra kontrollgruppas eksamen.

Gjennomføringen av kombinerte summative og formative tester med PeLe for testgruppa

Testene som ble brukt i forbindelse med PeLe, besto av 5–6 spørsmål, bestående av en blanding av konseptuelle oppgaver og tekniske regneoppgaver. Testene ble gjennomført i løpet av 2 x 45 minutter, med en kort pause imellom. Studentene fikk utdelt oppgavene på papir, og spørsmålene var i utgangspunktet formulert som åpne tekstoppgaver uten alternativer, som krevde utregning på tradisjonelt vis. Først etter en stund fikk studentene utdelt svaralternativene, og de besvarte så testen digitalt med PeLe.

Tidsbruken i løpet av en test med bruk av PeLe er illustrert i *Figur 4* under.



Figur 4: Tidsbruk under en kombinert summativ og formativ test med PeLe

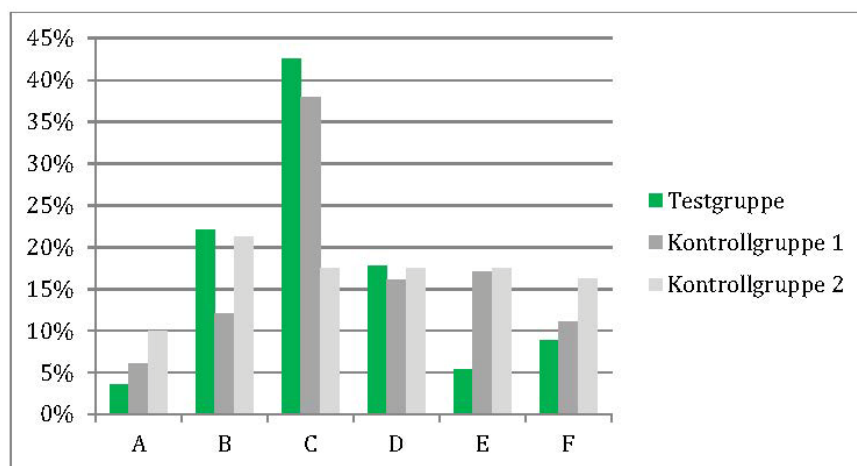
RESULTATER

Tabell 1 viser resultatene på hhv. forkunnskapstest, oppsummerende flervalgstest og avsluttende eksamen for testgruppen og de to kontrollgruppene. Her angir *n* antall studenter i gruppa; *M* er gjennomsnittlig prosentvis poengsum, og *SD* er standardavviket av (prosentvis) poengsum.

Tabell 1. Resultater fra de tre testene som ble gjennomført gjennom semesteret i test- og kontrollgruppe.

Gruppe	Forkunnskapstest			Oppsummerende flervalgstest			Avsluttende eksamen		
	n	M	SD	N	M	SD	n	M	SD
Testgruppe	100	51 %	18 %	59	64 %	15 %	113	64 %	18 %
Kontrollgruppe 1	94	54 %	19 %	61	43 %	17 %	100	54 %	23 %
Kontrollgruppe 2	78	54 %	19 %				80	52 %	24 %

Figur 5 viser karakterfordelingen på den avsluttende, skriftlige matematikk-eksamen for test- og kontrollgruppene:



Figur 5. Karakterfordelinger på avsluttende skriftlig matematikk-eksamen.

DISKUSJON OG KONKLUSJON

Begrensninger ved studien

Når datamaterialet i studien analyseres, må det tas høyde for enkelte metodiske svakheter med studien:

- Forskjellig eksamenslengde og -oppgaver: Testgruppa hadde halvert eksamenslengde sammenliknet med kontrollgruppene, og oppgavene var en delmengde av oppgavene fra kontrollgruppenes 5-timers eksamen. Eksamensresultatene det refereres til i Tabell 1, gjelder de respektive oppgavsettene for hhv. test- og kontrollgruppen, og altså ikke delmengden oppgaver som var felles. En slik sammenlikning var uhensiktsmessig ettersom delmengden med felles oppgaver hadde noe ulik vekt i test- og kontrollgruppene.
- Separering av teknologi og metodikk: Slik studien ble gjennomført, er det ikke mulig å separere effekten av bruk av mobilteknologi fra effekten av metodikken med umiddelbar tilbakemelding. Med så store klasser ville det imidlertid være umulig å hente inn data fra studentene som grunnlag for umiddelbar tilbakemelding, uten bruk av teknologi.

Diskusjon av resultater

En «Student t-test» anvendt på datamaterialet i *Tabell 1* viser, med et signifikansnivå på 0,05, at forskjellen mellom test- og referansegruppene ikke er signifikant på forkunnskapstesten, dvs. gruppene er å regne som faglig likeverdige. På den oppsummerende flervalgstesten, samt den avsluttende eksamen, var testgruppas resultater signifikant bedre en kontrollgruppas.

Karakterfordelingene i *Figur 4* gir et inntrykk av hvilke studenter som ser ut til å ha en effekt av formative og summative tester, slik de ble gjennomført i denne studien: det er de svake studentene. Som figuren viser, er både strykprosent og andelen med karakteren E markant lavere i testgruppa.

I fokusgruppeintervjuer ble studentene spurt om hvilket aspekt av testene – summativt versus formativt – de anså som viktigst for sin egen læring. At testene var karaktergivende ble framhevet som viktig av studentene: Selv om studentene vurderte læringseffekten av de formative læringsaktivitetene i etterkant av testen som betydelig, ga det summative aspektet en ytre motivasjon til å jobbe ekstra med matematikkfaget. Umiddelbar tilbakemelding gjennom semesteret ble også trukket fram av studentene som sentralt for læringsprosessen.

KONKLUSJON

Studien viser at mobilteknologi kan brukes til å gjennomføre kombinerte formative og summative tester i matematikk på en slik måte at studentene oppnår signifikant bedre resultater på en tradisjonell, skriftlig matematikkeksamen.

LITTERATUR

- Bennett, R. E. (2011). Formative assessment: a critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25.
- Black, P., & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139–48.
- Cech, S. J. (2007). Test industry split over «formative» assessment. *Edweek* 28(4: 1), 15. Hentet 10.02.2015 fra http://www.edweek.org/ew/articles/2008/09/17/04formative_ep.h28.html
- Kluger, A. N., & A. DeNisi (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin* 119(2), 254–84.
- Looney, J. W. (2011). *Integrating Formative and Summative Assessment: Progress Toward a Seamless System?*. Paris: OECD Publishing.
- One2act. Hentet 10.02.2015 fra <http://one2act.no>.