

Kontinuerlig examination via automatrerade webbquizzar

Frank Wikström

Matematikcentrum, Lunds Universitet, Box 118, 221 00 Lund; frank.wikstrom@math.lth.se

Sammanfattning — Vi beskriver ett experiment att på en kurs i matematik införa veckovisa webbaserade och automatrerade färdighetsprov i syfte att på ett kostnadseffektivt sätt introducera moment av löpande formativ bedömning i undervisningen och därigenom få studenterna att ligga bättre i fas och att hålla en jämnare arbetsbelastning under hela läsperioden.

Index Terms — computer aided assessment, continuous assessment, formative assessment, automatic grading

I. INLEDNING

KURSEN *Funktionsteori*, FMAF01, läses under årskurs 2 av studenter på teknisk fysisk och teknisk matematik (HT) samt industriell ekonomi och elektroteknik (VT).

Många studenter uppfattar kursen som svår men intressant. Sedan länge har det funnits två obligatoriska skriftliga inlämningsuppgifter på kursen. Syftet med inlämningsuppgifterna är dels att träna den skriftliga presentationen av matematiska resonemang, dels att se till att studenterna arbetar löpande med materialet under kursens gång utan att halka efter alltför mycket.

I huvudsak har inlämningsuppgifterna fungerat bra och den skriftliga återkoppling vi lämnar är mycket uppskattad, men det finns ett antal upplevda problem:

- Studenterna tycker att inlämningsuppgifterna tar mycket tid, och att uppgifterna är för svåra. Detta beror åtminstone delvis på att syftet att studenterna inte ska halka efter inte uppnås. Många studenter försöker göra inlämningsuppgifterna utan att ha räknat tillräckligt många övningsuppgifter på det aktuella materialet, och inlämningsuppgifterna blir därför mycket svårare än vad de är tänkta att vara.
- Rättningen av inlämningsuppgifterna tar väldigt mycket lärartid. Vi har haft som tradition att ”i uppfostringsyfte” rätta uppgifterna hårdare än hur vi normalt rättar skriftliga tentamina. Lösningar som inte är felfria måste korrigeras och lämnas in igen för en ny bedömning tills de är korrekta. Detta innebär att vissa studenter behöver lämna in sina uppgifter tre eller fyra gånger innan de blir godkända.
- Vi uppmanar studenterna att samarbeta, i första hand två och två, men vi har ibland haft problem med att samarbetet närmar sig gränsen för plagiat. Några av uppgifterna brukar dessutom återkomma varje eller nästan varje kursomgång, och vi är medvetna om att ”mönsterlösningar” från tidigare år cirkulerar bland studenterna. Vid några tillfällen har institutionen varit tvungen att anmäla studenter till disciplinnämnden för plagiat.

För att komma tillrätta med dessa problem har vi under höstterminen 2018 testat att ersätta de två omfattande inlämningsuppgifterna med fem kortare webbaserade och automatrerade färdighetsprov, ungefär ett per vecka. Tanken är att studenterna i ännu högre grad ska uppmuntras att kontinuerligt arbeta med kursmaterialet, men att den totala arbetsbördan för dem (liksom för de rättande lärarna) ändå sjunker. Automatrerade tester kan inte helt ersätta den återkoppling vi tidigare har lämnat på de skriftliga lösningarna, även om systemet vi använder har stöd för individualiserade kommentarer, så vi har också behållit en, nedbantad, skriftlig inlämningsuppgift.

II. DATORSTÖDD BEDÖMNING

Datorstödd bedömning (*Computer aided assessment*, CAA) har använts nästan lika länge som det funnits datorer. Redan 1959 användes automatiserad rättning av studentuppgifter på en kurs i programmering vid Rochester Polytechnic Institute [1]. Datorstödd bedömning i matematik har inte en riktigt lika lång historia, men är en i dag en välbeprövad teknik som använts framgångsrikt i många olika sammanhang. Det finns ett antal system för automatiserade tester på marknaden med olika styrkor och svagheter, såväl kommersiella som fritt tillgängliga.

Datorstödd bedömning är tilltalande av flera skäl: webbaserade tester är tillgängliga för studenterna när och var som helst – det enda som krävs är en fungerande internetuppkoppling, och det är fullt möjligt att göra tester på en telefon eller surfplatta; rättningen sker omedelbart – så studenterna får direkt feedback och information om vilka områden de måste träna mera på; studenterna kan göra tester upprepade gånger tills de känner sig säkra på materialet; frågorna kan varieras via slumpmässigt valda parametrar, vilket dels gör att varje fråga kan återanvändas av samma student utan att bli alltför repetitiv, dels gör att risken för plagiat minskar. Även om två studenter gör samma test samtidigt kommer de att få olika frågor.

Dessutom är tekniken mycket kostnadseffektiv. När man väl byggt upp en bank av välfungerande frågor är arbetsinsatsen som krävs från lärarhåll försumbar.

Automatiserade tester kan med fördel ges löpande under kursens gång och fungerar då som formativ, snarare än summativ, bedömning av studenternas arbetsinsats. Formativ bedömning har visat sig ge positiva effekter på inläringen [2, 3]. En studie av formativ datorassisterad bedömning på tidiga matematikkurser för ingenjörstudenter på Nya Zeeland [4] visade också på positiv korrelation mellan resultatet på de automatrerade testerna och resultatet på

kursens avslutande tentamen. Denna studie visade dock även på en liten negativ korrelation mellan total tid som spenderades på de automaträttade testerna och tentamensresultatet, vilket antyder att vissa studenter tenderar att lägga alltför mycket tid på de automatiska testerna jämfört med andra inlärningsaktiviteter.

Systemet som används på Matematikcentrum kan ge individualiserade kommentarer som beror av hur studenten besvarat frågan och det går därför att fånga upp vissa typer av vanliga fel eller missuppfattningar och ge studenten personliga och relevanta kommentarer med till exempel hänvisningar till lämpliga avsnitt i läroboken eller andra tips på hur man kommer vidare. Denna typ av individualiserad återkoppling har också påvisats förbättra resultaten vid summativ bedömning som görs vid kursens slut, [5]. Systemet som vi använder kan även ge information till läraren om vilka typer av felaktiga svar som studenterna lämnar. Det går därför att gå in och lägga till anpassade kommentarer till de vanligaste felen, antingen direkt medan testet är öppet eller i efterhand, så att studenter som gör testerna kommande kurstillfällen kan dra nytta av tidigare års resultat. På det sättet uppnås en slags dubbelriktad återkoppling som är svår eller omöjlig att åstadkomma på traditionella manuellt rättade tester.

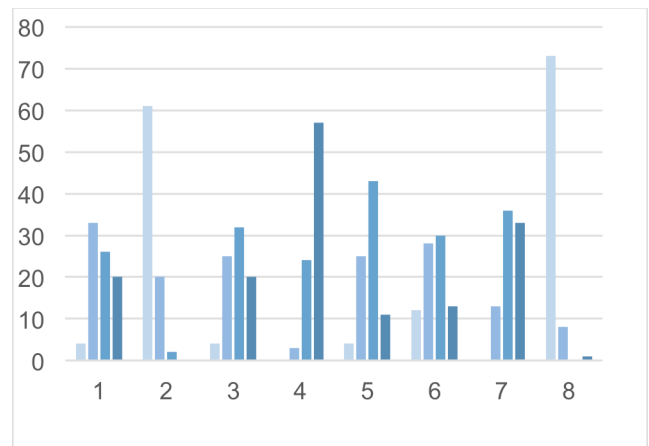
En nackdel med automaträttade frågor och tester är dock att tekniken sätter vissa begränsningar på vilka typer av frågor som är möjliga att ställa. Frågor som är mer komplicerade och mäter djupare förståelse av kursmaterialet tenderar att vara svåra att automaträtta.

Eftersom det är svar, snarare än lösningar, som bedöms, finns det också risk för missvisande rättning. Det kan hända att studenten kommer fram till rätt svar på en fråga, fast på helt felaktiga grunder: ”Är denna serie konvergent?” är därför en fråga som inte fungerar särskilt bra på ett automaträttat test, även om den fungerar väl på en manuellt rättad tentamen, eftersom det är själva motiveringen, snarare än svaret som ska bedömas. Omvänt, en principiellt korrekt lösning som innehåller något trivialt räknepel kommer att automaträttas som felaktig, även om den skulle ge full eller nästan full poäng vid manuell rättning. Se [6] för ytterligare exempel på fallgropar vid konstruktion av uppgifter som ska automaträttas. Det krävs således en hel del eftertanke vid uppgiftskonstruktionen för att säkerställa att det lärandemål som frågan verkligen kontrollerar är det avsedda lärandemålet.

III. GENOMFÖRANDE

Matematikcentrum använder sedan 2016 det kommersiella systemet *Mozquizto* för automatiska tester och instuderingsfrågor. Systemet är utvecklat i nära samarbete med Matematikcentrum och har använts för fler än 40 000 tester, framför allt de obligatoriska färdighetsprov som görs av nybörjarstudenter vid LTH under läsperiod 1 på kursen i endimensionell analys.

Vid årets kursomgång i funktionsteori fick studenterna göra fem obligatoriska korta tester, vardera omfattande två till tre frågor som behandlade det grundläggande innehållet från föregående läsveckas föreläsningar. För att bli godkänd krävdes att alla frågor skulle besvaras korrekt. Testerna var öppna från tisdag till torsdag under läsvecka 2, 3, 4, 5 respektive 7 och studenterna kunde göra ett obegränsat antal



Figur 1. Enkät svar, antalet svar på y-axeln. Staplarna visar antalet svar av typen ”instämmer inte alls”, ”instämmer delvis”, ”instämmer till stor del” respektive ”instämmer helt”.

försök under de tre dagar som testet var tillgängligt. Den nedbantade skriftliga inlämningsuppgiften låg läsvecka 6. Dessutom öppnades de automaträttade testerna igen under läsvecka 8 för att ge studenter som av någon anledning missade det ordinarie tillfället en extra chans att bli godkända. Godkänt resultat på alla fem tester och på den skriftliga inlämningsuppgiften krävs för att vara behörig att tentera.

Totalt konstruerades ett sextiototal frågor, där de allra flesta har dussintals (ibland tusentals) olika variationer på grund av slumpmässigt valda parametrar.

IV. STUDENTERNAS UPPFATTNING AV EXPERIMENTET

I slutet av läsvecka 6 genomfördes en enkät med studenterna om hur de upplevt årets obligatoriska moment. Totalt inkom 84 enkät svar. Överlag var studenterna mycket nöjda med upplägget, vilket underströks av fritextsvar.

Enkätfrågorna som ställdes var:

1. Quizzarna har fått mig att ligga mer i fas än vad jag skulle ha gjort utan dem.
2. Quizzarna har tagit för mycket tid i förhållande till vad jag har fått ut av dem.
3. På en del av quizfrågorna har systemet lämnat kommentarer på dina svar. Dessa kommentarer har varit värdefulla.
4. Systemet vi använder (Mozquizto) har varit enkelt att använda.
5. Det har varit roligt att göra quizzarna.
6. Jag hade gjort quizzarna även om de hade varit frivilliga.
7. Vi borde använda liknande quizzar på fler kurser i matematik.
8. Jag hade hellre haft två omfattande skriftliga inlämningsuppgifter (som det varit tidigare) än att göra som i år.

med svarsalternativen ”instämmer inte alls”, ”instämmer delvis”, ”instämmer till stor del” och ”instämmer helt”. Resultaten på enkätfrågorna syns i figur 1.

Studenterna upplever alltså att testerna haft viss positiv effekt på att de ligger bättre i fas, att systemet varit enkelt att använda, att testerna inte tagit för mycket tid och att vi borde använda liknande upplägg på fler kurser. Det är anmärkningsvärt att bara en enda student sade sig föredra det gamla upplägget på kursen.

Det var också ett överraskande positivt svar på frågan om det har varit roligt att göra testerna. Detta resultat kan möjligen ses som en bekräftelse av observationen att ”spelifiering” av läxor och andra typer av uppgifter som

görs utanför klassrummet gör dem mer attraktiva för studenterna [7], även om spelinslagen denna gång varit mycket begränsade.

Vi ställde dessutom tre enkätfrågor om antalet webb-quizzar och om svårighetsgraden på frågorna på quizzarna respektive på inlämningsuppgiften. Praktiskt taget alla svar (>90%) sa att det hade varit lagom många tester och lagom svåra frågor.

V. LÄRARNAS UPPFATTNING AV EXPERIMENTET

Totalt har tre lärare varit inblandade på kursen. Författaren har gett föreläsningarna och haft en av tre övningsgrupper, och dessutom varit ansvarig för frågekonstruktion och genomförandet av de automaträttade testerna. Två andra seniora lärare vid Matematikcentrum har haft de båda övriga övningsgrupperna. Alla tre har hjälpt till att rätta den kvarvarande skriftliga inlämningsuppgiften, och samtliga tre lärare har varit inblandade i kursen vid flera tidigare kurstillfällen.

Lärarnas subjektiva uppfattning är att de studenter som regelbundet har varit närvarande på övningarna har legat bättre i fas än vanligt, men detta är inget vi kunnat mäta objektivt, eftersom vi inte har tillgång till den typen av data från tidigare år.

Inlämningsuppgiften har gått ungefär som den brukar, men eftersom den var mindre omfattande än tidigare år, har lärarna haft lite mer tid och därmed större möjlighet att lämna ännu mer utförliga kommentarer på studenternas skriftliga lösningar än vanligt. En nackdel med årets upplägg är dock att vi inte har lika bra kontroll av att studenterna kan lösa lite mer komplicerade uppgifter inom de områden som bara behandlades på de automaträttade testerna. Å andra sidan har vi detta år också kunnat ställa (automatrerade) frågor under kursens gång på de sista kapitlen i kursboken, något vi inte kunde göra tidigare, eftersom den manuella rättningen tar ganska lång tid och inlämningsuppgifterna kunde därför inte täcka det material som går igenom under kursens sista två veckor.

I skrivande stund har kursens sluttentamen ännu inte gått av stapeln, varför det är omöjligt att säga något om experimentets eventuella effekter på tentamensresultatet.

VI. SLUTSATSER

Experimentet med att införa automatrerade tester för att uppnå inslag av formativ bedömning och kontinuerlig examination har slagit väl ut.

Både studenter och lärare upplever, åtminstone subjektivt, att målet att få studenterna att ligga bättre i fas under hela kursen har uppnåtts. Det verkar vidare som att studenterna har ägnat totalt sett mindre tid åt årets obligatoriska moment än vad de gjort tidigare år, fast jämnare utspritt över hela läsperioden, vilket också var ett av målen med experimentet. De inblandade lärarna har framhållit att rättningen av inlämningsuppgifterna har varit mindre betungande än vanligt, och att de därför kunnat fokusera mer på den skriftliga återkopplingen.

Vi planerar att genomföra intervjuer med ett litet antal studenter för att få en mer detaljerad förståelse för hur de arbetat med de automatrerade testerna och vilka detaljer som skulle kunna förbättras till nästa gång.

Experimentet kommer att upprepas vid nästa kurstillfälle (VT 2019) och om även detta slår väl ut, talar det mesta för

att kursupplägget ändras permanent och att vi fortsätter med årets modell.

Vi planerar även att undersöka möjligheterna att införa liknande automatrerade tester och/eller frågelistor på fler kurser i matematik.

REFERENSER

- [1] Hollingsworth, J., "Automatic graders for programming classes," *Comm. ACM*, 3 (10), 1960, pp. 528–529.
- [2] Bell, B. and Cowie, B., "The characteristics of formative assessment in science education." *Formative Assessment in Science Education* 85 (5), 2001, pp. 536–553.
- [3] Black, P. and Wiliam, D., "Assessment and classroom learning. Assessment in Education: Principles," *Policy & Practice* 5 (1) 1998, pp. 7–68.
- [4] Hannah, J., James, A. and Williams, P., "Does computer-aided formative assessment improve learning outcomes?," *Int. J. of Math. Education in Science and Technology*, 45 (2), 2014, pp. 269–281.
- [5] Gill, M. and Greenhow, M., "How effective is feedback in Computer Aided Assessments?," *Learning, Media and Technology* 33 (3), 2008, pp. 207–220.
- [6] Lawson, D., "Computer-aided assessment in mathematics: Panacea or propaganda?," *United Kingdom's LTSN Mathematics, Statistics and Operational Research*, 9, 2001.
- [7] Goehle, G., "Gamification and web-based homework," *Problems, Resources, and Issues in Mathematics Undergraduate Education* 23 (3), 2013.