

Återkoppling genom automaträttning

Fredrik Heintz och Tommy Färnqvist, Linköpings universitet

Sammanfattning—Vi har undersökt olika former av återkoppling genom automaträttning i en kurs i datastrukturer och algoritmer. 2011 undersökte vi effekterna av tävlingsliknande moment som också använder automaträttning. 2012 införde vi automaträttning av laborationerna. Vi undersökte då hur återkoppling genom automaträttning påverkar studenternas arbetssätt, prestationsgrad och relation till den examinerande personalen. Genom automaträttning får studenterna omedelbar återkoppling om deras program är tillräckligt snabbt och ger rätt svar på testdata. När programmet är korrekt och resurseffektivt kontrollerar kursassistenter att programmet även uppfyller andra krav som att vara välskrivet och välstrukturerat. Efter kursen undersökte vi studenternas inställning till och upplevelse av automaträttning genom en enkät. Resultaten är att studenterna är positiva till automaträttning (80% av alla som svarade) och att den påverkade studenternas sätt att arbeta huvudsakligen positivt. Till exempel svarade 50% att de ansträngde sig hårdare tack vare automaträttningen. Dessutom blir rättningen mer objektiv då den görs på exakt samma sätt för alla. Vår slutsats är att återkoppling genom automaträttning ger positiva effekter och upplevs som positiv av studenterna.

I. INTRODUKTION

ÅTERKOPPLING är grundläggande för lärande. Återkoppling ger information om vad man gör rätt och vad man gör fel vilket gör det möjligt att korrigera eller förbättra sitt beteende. Vårt mål är att genom automaträttning ge studenter bättre, snabbare och mer anpassad återkoppling. Automaträttning ersätter inte alla andra former av återkoppling utan är ett viktigt komplement. Några fördelar med automaträttning är att den är opartisk, konsistent, direkt, kan hantera stora grupper av studenter och kan individanpassas. Intresset för automaträttning har stadigt ökat som en effekt av bättre system för automatisk återkoppling och de allt större och populärare onlinekurserna.

Vi har undersökt olika former av automatisk återkoppling i en kurs i datastrukturer och algoritmer (DALG) vid Linköpings universitet under 2011 och 2012. DALG-kursen ges i början av det andra året på våra tre program med datavetenskaplig inriktning. Civilingenjör Datateknik (D), Civilingenjör Informationsteknologi (IT) och Kandidat Datavetenskap (C). Totalt tas kursen av ungefär 140 studenter. Kursen är organiserad i traditionell monolitisk form, med veckovisa föreläsningar, lektioner i klasser, laborationer och en skriftlig tentamen i slutet av kursen. Kursen ger 6 hp och betygskalan är U, 3, 4, 5.

Det har gjorts många försök inom högre utbildning runt automatisk återkoppling och examination. Att använda automatisk rättning av programmeringsuppgifter faller sig naturligt och har gjorts i mer än 40 år. Det sparar värdefull lärartid samt ger opartisk och direkt återkoppling. Några exempel är Enström et al. som beskriver hur de använder automatisk rättning av laborationer [1]. Guerreiro och Georgouli som dessutom använde automatisk rättning för självvärdering [2], [3]. García-Mateos och Fernández-Alemán

försökte ersätta den avslutande examinationen med en serie av uppgifter som automaträttades via ett webbaserat system [4].

II. DALG 2011

Programmering är ett hantverk. Akademiska kurser i programmering lär ut grunderna och de viktiga koncepten, men för att bli en riktigt bra programmerare krävs det av studenten att denne investerar betydande egen tid på kvalitativ träning. Det finns studier som indikerar att det tar ungefär 10 år att gå från novis till expert [5]. Detta stöds av Ericsson et al., vars forskning visar att det krävs ungefär 10000 timmar av genomtänkt träning för att bli en expert inom ett område [6].

För att bli en expert på något krävs det att man aktivt engagerar sig i genomtänkt träning (deliberate practise) [6]. Aktiviteterna ska tänja individens förmåga bortom dagens nivå, ge omedelbar återkoppling, repeteras flera gånger och kräva betydande insats och full koncentration. En teori som stödjer skapandet av genomtänkta övningar är Dreyfus och Dreyfus modell för färdighetsträning [7]. Enligt modellen går man normalt igenom fem stadier, från novis till kompetent, kunnig och därefter expert för att slutligen bli mästare. I de tidigare stadierna krävs detaljerade instruktioner medan i senare stadier har en tyst förståelse för hur man uppnår önskat resultat även i nya situationer utvecklats. Detta betyder att typen av aktiviteter och återkoppling ändras beroende på vilket stadie en student befinner sig i.

Vår erfarenhet från programmeringstävlingar är att dessa stimulerar och inspirerar studenter att lösa programmeringsuppgifter på egen hand. Detta ökar studenternas programmerings- och problemlösningsfärdigheter vilket signifikant ökar deras anställningsbarhet – ett faktum som är tydligt för oss, då flera företag vill synas i samband med våra tävlingar och träffa studenterna. En förutsättning för dessa tävlingar är automaträttning, både för att kunna hantera många samtidigt deltagare och för att vara opartisk.

I DALG-kursen 2011 undersökte vi två olika sätt att använda tävlingsliknande moment för att stödja studenternas genomtänkta träning inom programmering. Den frivilliga tävlingen var helt beroende av automatisk återkopplingen medan laborationstävlingen i huvudsak rättades för hand, men som i grunden lämpar sig för automaträttning. Med start 2012 har vi även infört automaträttning av laborationerna.

A. Laborationstävling

För att undersöka olika format för laborationstävlingen delade vi in studenterna i fyra grupper. Den första gruppen tävlade baserat på hur snabbt (antal dagar från starten av kursen) och korrekt (+3 extra dagar för varje felaktig inskickning) de löste laborationerna. Den andra gruppen tävlade baserat på kvalitet (cyklomatisk komplexitet och antal instruktioner) och effektivitet (körtid och minnesanvändning).

Den tredje och fjärde gruppen var kontrollgrupper och tävlade inte. Tabell I ger bakgrundsinformation om de olika grupperna vid kursstart. Vi har gjort utförliga statistiska tester och den enda signifikanta skillnaden (på 5%-nivån) mellan grupperna gäller antalet matematikpoäng för Grupp 1 och Grupp 3, där både medelvärdet och fördelningen signifikant avviker. Grupp 1 och Grupp 2 består av studenter från D-programmet, Grupp 3 av studenter från både C- och D-Programmen och Grupp 4 består av studenter från IT-programmet. IT-studenterna har ganska annorlunda kurser jämfört med C och D och pedagogiken på programmet bygger på problembaserat lärande. Detta gör att det inte är meningsfullt att ta upp deras resultat från första året i den här jämförelsen.

TABELL I
GENOMSNIITTLIGA POÄNG OCH BETYG FÖR STUDENTERNA I
PROGRAMMERING, MATEMATIK OCH DATAVETENSKAP.

	Alla	Grupp 1	Grupp 2	Grupp 3
Antal studenter	140	32	33	35
Totalt antal poäng	42.4	43.6	43.4	40.0
Programmeringspoäng	11.0	10.9	11.1	11.0
Programmeringsbetyg	3.73	3.75	3.84	3.58
Matematikpoäng	15.0	16.2	16.1	12.4
Matematikbetyg	3.61	3.57	3.65	3.60
Datavetenskapspoäng	15.1	14.6	14.8	16.0
Datavetenskapsbetyg	3.78	3.75	3.84	3.74

Alla grupperna skickade in sina laborationer genom ett webbaserat system så fort de trodde de var klara. Vi fann att båda grupperna som tävlade påverkades av tävlingarna. Den starkaste effekten är i mönstret för när laborationerna skickades in, vilket skiljer sig märkbart mellan grupperna. Detta syns tydligt i Fig 1, där Grupp 1 arbetade väldigt snabbt och Grupp 2 något snabbare än kontrollgrupperna (Grupp 3 och Grupp 4). Ett tydligt exempel är att föreläsningarna gick igenom materialet för den första laborationen först andra veckan, så mönstret för Grupp 4 är vad vi kan förvänta oss utan tävlingar. Då Grupp 3 bestod av studenter från samma program som de som tävlade tror vi att de också drogs med i tävlingsandan. Det var även flera av dem som uttryckte en besvikelse över att inte få tävla. Det fanns även en stor variation i kvaliteten hos koden från de olika grupperna.

B. Frivilligtävling

För att motivera till ytterligare träning införde vi en frivillig tävling med uppgifter från Universidad de Valladolids automatiska domarsystemet [8]. Efter varje föreläsning fick studenterna som anmält sig till tävlingen en uppgift att lösa. Uppgifterna valdes antingen för att förstärka eller repetera det som togs upp på föreläsningen eller för att utmana studenten genom att kräva lite djupare kunskaper eller insikter. Den automatiska domaren gav omedelbar återkoppling och efter-som tävlingen krävde att studenterna skulle lösa uppgifterna så fort som möjligt så blev det en intensiv upplevelse. Även tävlingen uppfyller kraven på genomtänkt träning.

Tabell II visar att 30 studenter anmälde sig till den frivilliga tävlingen och att 15 löste minst en uppgift. Tendensen att det

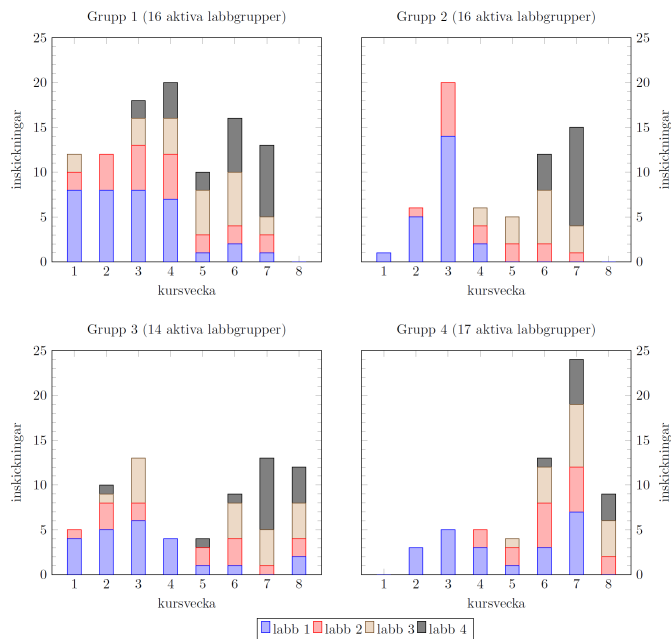


Fig. 1. Inskickningsaktivitet på laborationerna för de fyra grupperna.

TABELL II
RESULTAT FÖR OLIKA GRUPPER AV STUDENTER.

Studentgrupp	Totalt antal	Skrev tentan	Klarade tentan	Medelbetyg
Alla	140	118	95	3.36
Besvarade enkäten	79	74	62	3.47
Slutförde labbserien	76	73	62	3.47
Anmälda till frivilligtävling	30	29	27	3.56
Löste uppgifter i tävlingen	15	15	15	3.6

genomsnittliga betyget stiger med ökande aktivitet i kursen är inte tillräckligt stark för att vara signifikant om vi bara tittar på medelvärdet. Däremot är fördelningen av betyg för studenter som antingen var med i tävlingen och/eller löste minst en uppgift signifikant bättre (på 10%-nivån) jämfört med betygsfördelningen för alla studenter.

TABELL III
GENOMSNIITTLIGA POÄNG OCH BETYG FÖR STUDENTERNA I
PROGRAMMERING, MATEMATIK OCH DATAVETENSKAP.

	Alla	Svarade på enkät	Godkänd på labbar	Anmäld tävling	Löst minst en uppgift
Antal studenter	140	79	76	30	15
Totalt antal hp	42.4	45.7	46.2	45.4	50.0
Progr.-poäng	11.0	11.4	12.1	12.6	13.5
Progr.-betyg	3.73	4.00	3.83	4.00	4.27
Matematikpoäng	15.0	16.5	16.1	15.8	16.9
Matematikbetyg	3.61	3.68	3.70	3.68	3.89
Datavet.-poäng	15.1	15.6	16.4	16.9	18.4
Datavet.-betyg	3.78	4.06	3.89	4.04	4.26

En möjlig förklaring till att studenter som är med i den frivilliga tävlingen får bättre betyg skulle kunna vara att de är duktigare studenter, baserat på deras tidigare meriter. Enligt

Tabell III är det genomsnittliga antalet poäng som dessa studenter tagit signifikant högre (på 5%-nivån) än det genomsnittliga antalet poäng som tagits av alla studenter. Det samma gäller för programmeringspoäng för studenter som anmält sig och/eller löst minst en uppgift och för poäng i datavetenskap för studenter som löst minst en uppgift. Det genomsnittliga antalet programmeringspoäng för studenter som löst minst en uppgift är signifikant högre än det genomsnittliga antalet poäng för alla studenter (på 10%-nivån). Inget annat sådant förklarande signifikant samband kunde hittas. Det verkar därför som, statistiskt sett, att inte hela effekten av det bättre resultatet på tentan för de studenter som deltagit i den frivilliga tävlingen kan förklaras av deras starkare bakgrund.

C. Enkät

Efter kursen gjorde vi en enkätstudie med undersökande och fördjupande frågor kring studenternas aktivitet i kursen. 79 av 140 studenter svarade. Den enda signifikanta skillnaden med avseende på bakgrund mellan de som svarade på enkäten och hela populationen är det medelvärde antalet datavetenskapspoäng (på 10%-nivån).

Vi ställde både flervalsfrågor och fritextfrågor. Här tar vi bara upp några av de viktigaste frågorna. På frågan "Vad är din allmänna inställning till DALG-tävlingarna?" svarade 22% mycket positiv, 35% ganska positiv, 30% neutral, 11% ganska negativ och 0% mycket negativ, vilket förstärker vår och kursassistenternas uppfattning att tävlingsmomenten mottogs väl. Det kan tyckas märkligt att så många studenter är positiva trots att relativt få deltog aktivt. Förklaringen kan till stor del hittas i fritextsvaren där studenterna säger att de ville vara med men inte hade tid eller att de valde att inte vara med då det inte gav några kurspoäng eller poäng på tentan. Enligt den centrala kursvärderingen var studenterna mycket nöjda med kursen.

III. DALG 2012

Till DALG-kursen 2012 gjorde vi två viktiga förändringar. För det första så införde vi automaträttade laborationer. Det betyder att vi använder återkoppling genom automaträttning vid examination av kursens lärandemål. För det andra så gav den frivilliga tävlingen extra poäng på tentan till de studenter som löst tillräckligt många uppgifter. 2012 var 138 studenter aktiva på kursen.

A. Laborationer

Kursen har fyra laborationer där studenterna fick omedelbar återkoppling om deras program var korrekt genom automaträttning. För att bli accepterat måste deras program vara tillräckligt snabbt och ge rätt svar på all testdata. När programmet var korrekt och resurseffektivt kontrollerade assistenterna att programmet uppfyllde andra krav, som att vara välstrukturerat och välstrukturerat. Samma laborationer som 2011 användes, vilket gör att vi kan jämföra resultaten mot föregående års. För automaträttning användes Kattis, utvecklat vid KTH [1].

För att testa olika återkopplingsmodeller delades studenterna in i fyra olika grupper. Den första gruppen fick enbart den automatiska domarens normala återkoppling, vilket innebär att

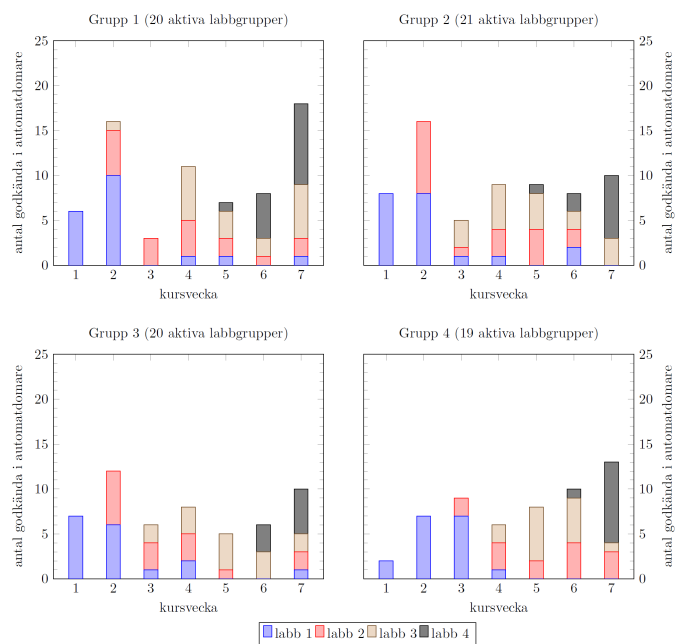


Fig. 2. Antalet slutförda labbar i Kattis per kursvecka.

de får reda på om programmet kraschar, om det tar för lång tid att exekvera, om det ger fel svar eller om det ger rätt svar. De tre övriga fick viss information om det fallande testfallets art och också tillgång till all utdata från det inskickade programmet. Tabell IV ger diverse bakgrundsstatistik för grupperna.

TABELL IV
GENOMSNIITTLIGA POÄNG OCH BETYG FÖR STUDENTERNA I
PROGRAMMERING, MATEMATIK OCH DATAVETENSKAP.

	Grupp 1	Grupp 2	Grupp 3
Totalt antal poäng	46.3	46.2	46.0
Programmeringspoäng	13.6	12.8	13.3
Programmeringsbetyg	4.3	4.0	4.2
Matematikpoäng	15.2	14.9	11.9
Matematikbetyg	3.6	3.5	3.5
Datavetenskapspoäng	17.5	16.3	19.5
Datavetenskapsbetyg	4.3	4.0	4.0

Vi har gjort utförlig statistisk testning och funnit att: antalet programmeringspoäng för Grupp 2 skiljer sig signifikant (på 10%-nivån) från Grupp 1 när det gäller medelvärdet (men inte fördelningen), antalet matematikpoäng för Grupp 3 skiljer sig signifikant (på 5%-nivån) från Grupp 1 och 2 både när det gäller medelvärde och fördelning, antalet datavetenskapspoäng skiljer sig signifikant (på 1%-nivån) från Grupp 1 och 2 när det gäller medelbetyg och signifikant (på 5%-nivån) från Grupp 1 och 2 när det gäller fördelningen. Grupp 1 och 2 hade enbart studenter från D-programmet, medan Grupp 3 bestod av studenter från både C- och D-programmen. Den fjärde gruppen bestod enbart av IT-studenter, vars programplan för första årets skiljer sig relativt mycket från de för C- och D-programmen, varför jämförelser av den här typen inte bedöms som meningsfulla.

I Fig 2 ser vi hur många labbgrupper från respektive grupp

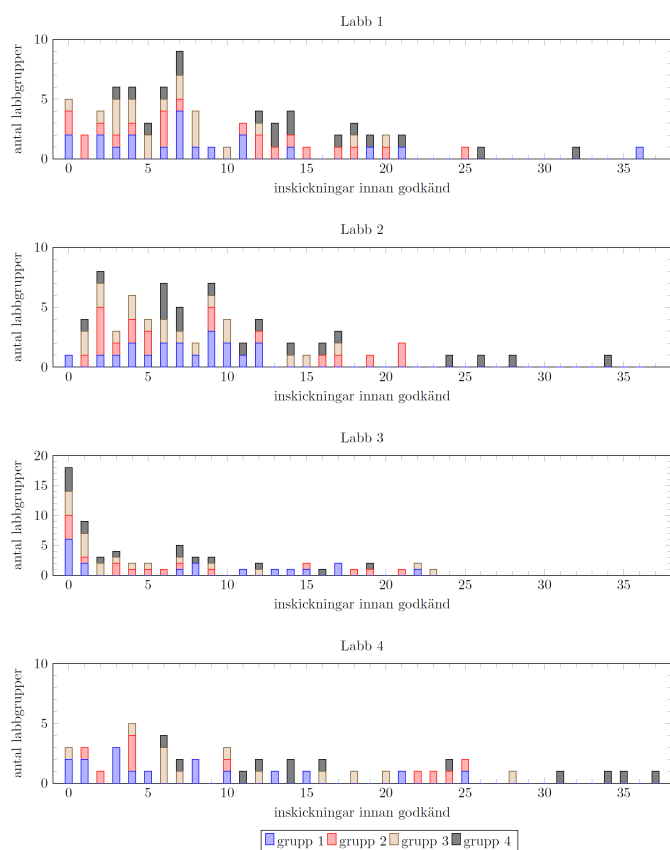


Fig. 3. Antalet inskickningar innan AC per labb och grupp.

som slutförde de olika labbarna per kursvecka. Den enda signifikanta skillnaden är att fördelningen för Grupp 4 skiljer sig för några av labbarna jämfört med övriga grupper. Samma sak gäller för datat i Fig 3, där vi kan se antalet inskickningar innan Kattis godkände labben för de olika grupperna per labb¹.

I Fig 4 kan vi se hur många godkända inskickningar labbgrupperna har gjort per labb och grupp. Ett av den automatiska domarens kriterier för att godkänna en labb är att koden är tillräckligt effektiv, så Fig 4 ger ett mått på hur länge olika labbgrupper valt att fortsätta att arbeta med att optimera exekveringstiden hos sin kod efter att de redan är godkända. Av figuren att döma skulle man förvänta sig att labb 3 och labb 4, men även labb 1, har optimeringskaraktär medan labb 2 troligtvis är en labb som har tyngdpunkten på funktionalitet snarare än effektivitet. Detta är också fallet — labb 3 går ut på att implementera en viss variant av quicksort, medan uppgiften i labb 4 är att snabba upp ett givet program så att det går från att ta dagar till någon tiondels sekund att köra.

Sammanfattningsvis ser vi att trots att Grupp 1 bara fick Kattis normala återkoppling, medan övriga grupper fick tips om typ av testfall samt allt utdata från programmet så kan vi inte observera någon skillnad i när studenterna blir färdiga

¹Fig 3 visar inte att: en labbgrupp ur Grupp 4 använde 103 inskickningar för att bli godkända på labb 1, en labbgrupp ur Grupp 2 använde 47 inskickningar för att bli godkända på labb 3, en labbgrupp ur Grupp 2 använde 61 inskickningar för att bli godkända på labb 4, en labbgrupp ur Grupp 2 använde 68 inskickningar för att bli godkända på labb 4, och att en labbgrupp ur Grupp 4 använde 93 inskickningar för att bli godkända på labb 4.

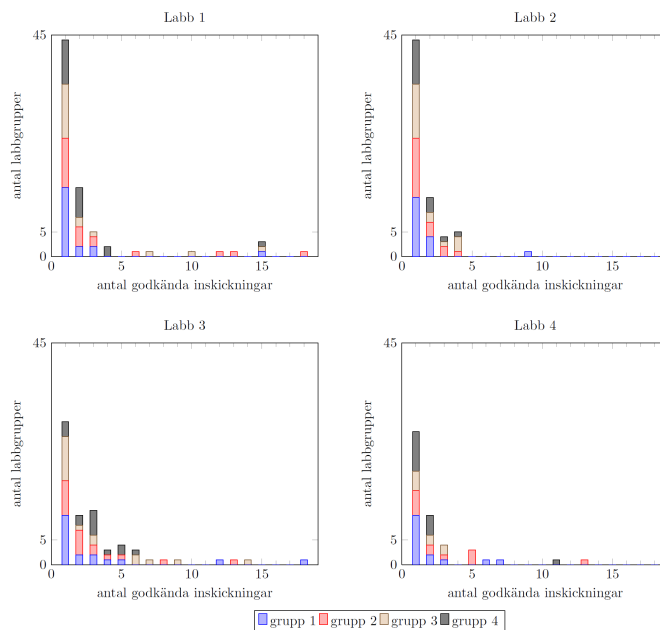


Fig. 4. Antalet inskickningar med AC per labb och grupp.

med laborationsuppgifterna. Vi vet från enkätsvaren att enstaka studenter använde att de hade allt utdata för att skriva program som bara var tabeller för testfallen i Kattis, men i de flesta fall kan vi inte se någon skillnad mellan studenter som fick normal återkoppling och de som fick utökad återkoppling. Man kan spekulera i att antagligen så skrev inga av de här studenterna någon egen testkod. Hade de som fick utökad återkoppling använt informationen de fick ut kunde de ha skrivit egen testkod och på så sätt behövt färre inskickningar i Kattis eftersom de då hade behövt fundera mycket mer på vad koden ska göra i stället för att göra inkrementella ändringar.

B. Frivilligtävling

Även 2012 arrangerades en frivillig tävling, men nu helt baserad på rättning i automatdomaren Kattis. De flesta av problemen var samma båda åren, den stora skillnaden var att 2012 delade vi inte ut några (penning)priser vilket vi gjorde 2011. Däremot kunde de studenter som löste tillräckligt många problem få bonuspoäng på tentan.

TABELL V
RESULTAT FÖR OLIKA GRUPPER AV STUDENTER.

Studentgrupp	Totalt antal	Skrev tentan	Klarade tentan	Medelbetyg
Alla	138	128	51	3.27
Besvarade enkäten	57	51	27	3.37
Slutförde labbserien	68	64	35	3.34
Anmälda till frivilligtävling	65	63	36	3.31
Löste uppgifter i tävlingen	50	49	28	3.36
Fick bonuspoäng på tentan	5	5	4	4.00

Till att börja med vill vi kommentera att tentaresultatet för alla studenter var exceptionellt dåligt. Antalet underkända brukar ligga runt 20% till 30%. Vi noterar dock att enkäten

säger att 90% av studenterna var ganska nöjda eller mycket nöjda med kursen. Hur som helst var den normala ordningen återställd efter första omtentatillfället.

Vi ser i Tabell V att 65 studenter var aktiva i tävlingen och att 50 löste minst en uppgift. (Att jämföra med 2011, då 30 studenter var aktiva i tävlingen och 15 löste minst en uppgift.) Avsaknaden av penningpriser hade alltså ingen negativ effekt på intresset för tävlingen. Däremot var det för svårt att tjäna ihop till bonuspoäng. Detta har åtgärdats till 2013 års kursomgång, där en bonuspoängstege finns implementerad.

Tendensen att ökande engagemang i aktiviteter i kursen ger högre betyg på tentan är inte tillräckligt stark för att vara signifikant förutom om vi jämför medelbetyget för alla studenter med de som tjänade ihop till bonuspoäng på tentan — då skiljer sig både medel och fördelning åt på ett signifikant vis (på 10%-nivån resp. 5%-nivån). Detta har dock sin förklaring i att de fyra studenterna med bonuspoäng är starka överlag.

C. Enkät

Efter kursen bad vi studenterna att fylla i en enkät med frågor relaterade till deras engagemang i olika aktiviteter i kursen. 57 av 138 studenter svarade på frågorna.

På frågan om assistenternas tillgänglighet under labbtillfällena svarade 15% Mycket bra, 33% Ganska bra och 28% Ok. På frågan om studenterna tycker att de kunde jobba mer självständigt jämfört med andra laborationskurser (tack vare automaträttningen) svarade 36% Ja, mycket och 50% Ja, lite. Detta indikerar att studenterna både tycker att de kan jobba mer självständigt och att assistenternas tillgänglighet är bra. Tidigare år har det funnits klagomål på att rättning har tagit för lång tid och att assistenterna inte har hunnit med att svara på alla frågor under laborationerna.

På frågan om studenterna tyckte att assistentens roll skiljer sig från hur den är i andra laborationskurser svarar knappt hälften Ja. Bland de som svarade Ja finns fritextkommentarer som säger att assistentens examinerande roll vid labbarna har kommit att handla mindre om redovisning av kod och mer om själva problemlösandet — helt i linje med Enström et. al. [1].

Vi frågade också om vad studenterna skulle vilja ändra i Kattis återkoppling. Bland svaren kan vi se att från Grupp 1 önskar man att få lite mer information om vad som gick fel med inskickningen, t.ex. genom att berätta vilken typ av testfall som gått snett — alltså precis den information övriga grupper fick. Samtidigt verkar det inte som om studenterna som får information om vilken typ av testfall de misslyckas på använder informationen på ett systematiskt sätt.

På frågan "Vad är din allmänna inställning till automaträttning?" svarade 33% Mycket positiv och 51% Ganska positiv. På frågan "Har automaträttningen haft någon medveten påverkan på din insats i kursen?" svarade 9% Ja, mycket positiv, 47% Ja, ganska positiv och 29% Nej, ingen. På frågan "Har du lagt mer tid på DALGen p.g.a. automaträttningen än du tror att du skulle ha gjort annars?" svarade 11% Ja, mycket mer, 56% Ja, lite mer och 29% Nej, lika mycket. Sammantaget indikerar detta att studenterna är positiva till automaträttning och att mer än hälften anser både att det har en positiv påverkan och att de jobbar mer. Förhoppningsvis

jobbar de även mer effektivt då de inte behöver vänta på återkoppling från assistenten utan de kan direkt jobba vidare.

Vi frågade också om studenterna trodde att rättning blev mer rättvis med automaträttning. På den frågan svarade 27% Ja, mycket och 46% Ja, lite. Det hade varit intressant att undersöka den här frågan mer, vad menar studenterna när de säger att det blir lite mer rättvist? Betyder det att de upplever rättningen i allmänhet som rättvis, vilket vore väldigt positivt, eller upplever de att automaträttningen inte gör så stor skillnad utan att det fortfarande är ganska orättvis rättning?

50% av studenterna trodde även att de fått anstränga sig mer för att bli godkända jämfört med om det hade varit handrättade labbar. Då vi har tillgång till laborationerna från 2011 i digitalt format är vår plan att jämföra rättningskvaliteten med och utan automaträttning. Det skulle ge intressant information om sanningshalten i detta påstående.

På frågan om studenterna föredrar automaträttning framför handrättning svarar 80% Ja vilket visar att studenterna, precis som vi, tycker att automaträttning är bra.

IV. SLUTSATSER

Vårt mål är att genom automaträttning ge studenter bättre, snabbare och mer anpassad återkoppling. Automaträttning ersätter inte alla andra former av återkoppling utan är ett viktigt komplement. Några viktiga fördelar med automaträttning är att den är opartisk, konsistent, direkt, kan hantera stora grupper av studenter och kan individanpassas. Våra experiment och undersökningar i en kurs i datastrukturer och algoritmer under 2011 och 2012 indikerar att studenterna är positiva till automaträttning och att den påverkar studenternas sätt att arbeta huvudsakligen positivt. Till exempel svarade 50% 2012 att de ansträngde sig hårdare tack vare automaträttningen. Dessutom blir rättningen mer objektiv då den görs på exakt samma sätt för alla. Att rättningen upplevs som mer rättvis är också tydligt i enkäten från 2012 där 73% av studenterna svarade att de tyckte att automaträttning var mer rättvist.

Vår slutsats är att återkoppling genom automaträttning ger de positiva effekter vi hoppats på och att det upplevs som positivt av studenterna. Detta inspirerar oss att fortsätta att utveckla den automatiska återkopplingen. Vi har även fått förfrågningar från andra lärare på institutionen om de kan använda automaträttning i sina kurser. Vår förhoppning är att det sprider sig som ringar på vattnet.

REFERENSER

- [1] E. Enström, G. Kreitz, F. Niemelä, P. Söderman, and V. Kann, "Five years with Kattis - using an automated assessment system in teaching," in *Proc. of IEEE Frontiers in Education Conference (FIE)*, 2011.
- [2] P. Guerreiro and K. Georgouli, "Combating anonymity in populous CS1 and CS2 courses," in *Proc. of ITiCSE*, 2006.
- [3] —, "Enhancing elementary programming courses using e-learning with a competitive attitude," *Int. Journal of Internet Education*, 2008.
- [4] G. García-Mateos and J. Fernández-Alemán, "A course on algorithms and data structures using on-line judging," in *Proc. of ITiCSE*, 2009.
- [5] L. E. Winslow, "Programming pedagogy - a psychological overview," *ACM SIGCSE Bulletin*, vol. 28, no. 3, pp. 17–22, 1996.
- [6] K. A. Ericsson, K. Nandagopal, and R. W. Roring, "Toward a science of exceptional achievement: Attaining superior performance through deliberate practice," *Annals of New York Academy of Science*, 2009.
- [7] S. E. Dreyfus and H. L. Dreyfus, "A five-stage model of the mental activities involved in directed skill acquisition," *Tech. Rep.*, 1980.
- [8] U. O. Judge, 2012. [Online]. Available: <http://uva.onlinejudge.org/>