Student attitude towards automatic and manual exercise and evaluation systems

Teemu Tokola, Rauli Puuperä, Kreetta Askola, Teiju Hietikko, Juha Röning

University of Oulu, Department of Electrical and Information Engineering
Oulu University Secure Programming Group, OUSPG
PO BOX 4500
90014 University of Oulu
Finland

Abstract. Automated teaching aids have become widespread during the last few years also in the universities in Finland. Studies have been made on the impact of these systems on the teaching results, and results have generally been encouraging. In this research the general attitudes of students towards automated learning environments and automated assessment are evaluated by surveys made to the participants of two courses in which automated systems have been recently adopted.
The research shows that automated systems are in general trusted and preferred by the students. Independent of their own success, students consider automated systems to be more professional and to treat the students equally. However, the students did not endorse the adoption of automated systems without reservations: developers and teachers need to remember the importance of human interaction and to spend appropriate effort in evaluating the need for and the quality of automated systems instead of adopting them without reservations.

1 Introduction

Automatic systems for teaching, grading and assessment have become more and more widespread. In Finnish universities, a variety of different systems have been employed in instruction and grading [1–3]. This paper is concerned with two automated systems employed in the University of Oulu.

The benefits of automated systems seem obvious at least to their developers – researchers are spared from grading endless piles of papers, students can learn in an assisted environment at any time of the day and automated systems keep track of all relevant data for later reference. Additionally, the automated systems employed have typically manifested at least an apparent increase in learning results.

Previous research has been mostly interested about the actual teaching results [4, 5]. While the impact of these systems is perhaps the most important incentive to adopting them, the attitude of students towards their use may be another strong argument. Many teachers still prefer traditional approaches to teaching and are hard to convince otherwise. The widespread adoption of automated methods in teaching requires a strong case that combines favourable teaching results with data that shows that the students approve and endorse the approach. Studying student attitude is important also because,

as related research suggests, a negative student attitude towards the method of learning adversely effects the student performance and vice versa [6].

This work aims to assess what is the current state of the student attitude towards systems such as the ones employed in the University of Oulu, using two courses as case studies. A survey was made to assess the attitudes of students and their perceptions of the difference between automated and manual systems.

2 Background

Automated systems have not previously seen wide adoption in courses taught in the Department of Electrical and Information Engineering in the University of Oulu. Part of the reason may be relatively small course sizes compared to universities in which such systems are widely adopted. Recently in the computer engineering[7] and the programming exercise[8] courses, automated systems have been adopted in an effort to increase quality of teaching, achieve better teaching results and to decrease the work-load on course assistants.

2.1 Programming Exercise

In the programming exercise course the students write a short exercise program in C or in Java. The students are expected to complete on their own a program that meets the specifications during the span of a few months. After completing the program, each student has to attend a private review session where the course assistant goes through the source code and documentation and does some test runs.

Before autumn 2006 almost all of the test-runs were done by hand. The assistant made up some test cases to see if the program crashes for the most typical cases and malformed inputs. In an effort to ensure equal treatment of students and to speed up the testing process while allowing a more complete test-run, Python scripts were adopted for testing in Autumn 2006. The assistant uses the script to test the program, and the student makes remarks in case improvements were requested. The teaching personnel consider the automatic testing a considerable improvement compared to the old, manual approach. Automated testing made the review sessions proceed faster, improved the test coverage and ensured all students had to pass the same collection of tests.

2.2 Computer Engineering

The computer engineering course concentrates on low-level computer operation. In addition to the lectures and exam, the course contains a laboratory exercise during which students are expected to produce a functional program made in the x86 assembly language. Before 2005 the students were expected to complete a set of pre-exercises prior to attending the lab and return them on paper or email to the course assistant. Despite the fact that the pre-exercises were ambitious and attempted to guide the students to attain a suitable skill-level, the starting level of students remained poor. A good example was the fact that although students were expected to compile and link code in the pre-exercises, most of them did not know how to do that when the laboratory session

started! Additionally, given the large volume of students, the assistant did not have time to check through the answers and therefore could not request corrections.

As a response to the low skill-level exhibited by the students, an automated pre-exercise system was devised in Spring 2005. This system, later named as Remote Automat (sic) In Programming Pre-exercise Assignments or RAIPPA, presented the student with series of questions, and finally allowed the student to register to the laboratory session [9]. Adoption of this system has lead to a significant increase in the starting level of students and has virtually eliminated cases that were completely at loss at the start of the exercise. This has allowed the adoption of a significantly more demanding and rewarding laboratory exercise.

The pre-laboratory system has now seen two iterations, the 2005 version and the 2006-2007-version. The latter one included a much more sophisticated guidance system and user interface compared to the 2005 prototype version.

3 Survey

Ages

Course	20	21	22	23	24	25	26	27	28	29	30	Not specified
Computer Engineering	2	13	21	28	22	11	4	1	0	0	0	3
Programming exercise	10	9	7	13	18	14	4	3	1	1	1	1

Major

	Computer Science	Electrical Engineering	Electronics	Telecommunications	Information Networks	Industrial Engineering	Physics	Mechanical engineering	Biophysics	Not specified
Computer Engineering	32	31	28	8	9	11	2	1	1	0
Programming Exercise	25	18	20	5	7	4	0	1	0	2

Starting year

Course	95	96	97	98	99	00	01	02	03	04	05	06	Not specified
Computer Engineering	1	0	0	0	0	2	8	9	35	26	19	5	0
Programming exercise	0	0	2	1	2	6	13	14	13	9	8	12	2

Year of attending the course

Course	2005	2006	2007	Not specified
Computer Engineering	42	21	38	4

Table 1: Demographic of the survey

To assess student attitudes towards automated systems in comparison to manual systems, a survey was performed in April 2007 to students that had attended either the

programming exercise during the last year or the computer engineering course during the last three years.

3.1 Survey statistics

For the programming exercise the survey was sent to 198 students that attended review sessions during Autumn 2006 and Spring 2007. Due to technical reasons, the invitation to the survey was also sent to 130 students that had been in correspondence with the course email address but had not attended the review. Of those who received the invitation to the survey, 82 students answered, over 41% of those that attended the review.

For the computer engineering course the survey was sent to the 455 students that had signed into the prelab system during the years 2005, 2006 and 2007. Of these, 105 students answered the survey. This is approximately 23% of the total population.

The invitations to attend the survey, that was hosted at the department web server, were sent via email. The survey was then held open for a period of one week. The demographics are listed in table 1.

The survey for the programming exercise contained 16 questions [10]. Of these, 14 were claims to which the students answered on a 1-5 scale, in which 1 stood for completely agree, 2 for partially agree, 3 for uncertain, 4 for partially disagree and 5 for completely disagree. Two remaining questions asked the students to select adjectives which described manual and automatic testing. In the computer engineering survey there were 20 questions, 18 of which were answered on the 1-5 scale and 2 were adjective choice questions [11].

4 Analysis

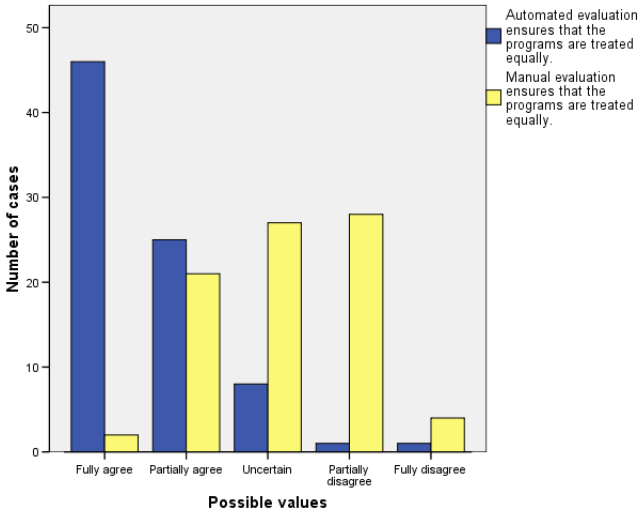
4.1 Automated systems promote equality

		Mean	std. error	95% conf.	t	df	p
Q1	Computer Engineering			[-1.179, -0.668]	-7.168	104	< 0.000
	Automated	2.05	1.101				
	Manual	2.97	0.093				
Q2	Programming Exercise			[-1.850, -1.236]	-10.007	80	< 0.000
	Automated	1.59	0.091				
	Manual	3.14	0.105				

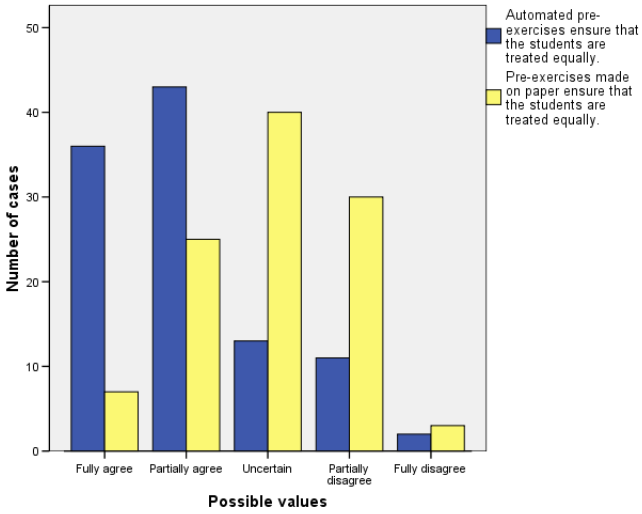
Q1 Do the automatic and manual pre-exercises ensure that students are treated equally?
Q2 Do the automatic and manual evaluation methods ensure that students are treated equally?

Table 2: Equal treatment of students in manual and automatic systems

As visible in figure 1 and shown in table 4.1, students are likely to believe that factors not related to their work affect the assessment they receive. When posed with



(a) Programming exercise



(b) Computer engineering

Fig. 1: Equal treatment of students

two questions, whether automatic assessment ensures equality of students and whether manual assessment ensures equality of students, the students were uncertain of the objectivity of manual assessment. The results for the automatic assessment and automated

pre-exercises, however, indicate that the students consider automated systems to be very objective.

4.2 Automatic pre-exercises are a comfortable way to work on prelab questions

		Mean	std. error	95% conf.	t	df	p
Q1	Programming Exercise			[-1.184, -0.501]	0.921	81	0.360
	Automated	2.33	0.123				
	Manual	2.17	0.122				
Q2	Computer Engineering			[-1.466, -0.725]	-5.863	104	< 0.000
	Automated	2.17	0.118				
	Manual	3.27	0.108				

Q1 Are the automatic and manual pre-exercise systems comfortable ways to do pre-exercises?
Q2 Are automated and manual evaluations comfortable ways of checking your work?

Table 3: The comfortability of automatic and manual systems

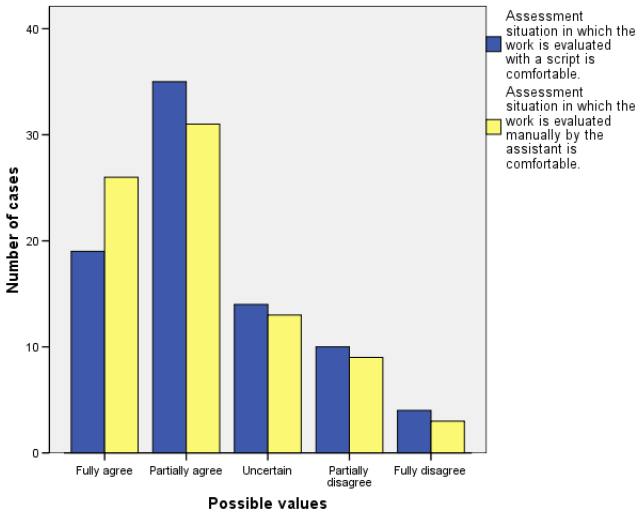
When students were enquired how comfortable they felt in the assessment situation, the programming exercise showed no significant difference between automated and manual evaluations as evident from table 3. However, the computer engineering pre-exercises again showed a significant difference as seen in table 3. The results are clearly visible also in the figure 2.

Clearly, whether the assistant checks the work manually or by using automated scripts, the students feel the situation to be equally comfortable; the work on the students part is ready, and the comfort level of the situation does not depend on the actual method of evaluating the program. In the pre-exercises, however, the situation is drastically different: in paper version the student has no immediately available support and doesn't know if the answers he already made are correct, whereas the automatic system supports the students and tells whether the answers are correct. This makes the automated pre-exercises more comfortable.

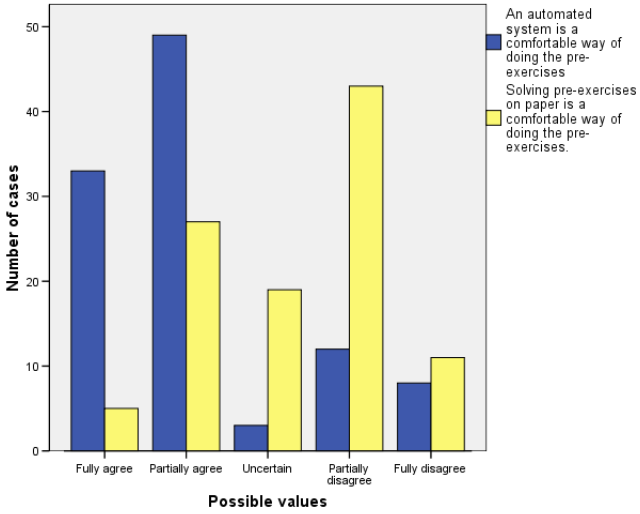
4.3 Automated systems are a more professional, but not necessarily an automated choice

As shown in table 4, students consider automated pre-exercises to be a professional choice. A more reserved opinion was expressed to the question on whether automatic pre-exercises should be adopted for all exercises where it is feasible - the results show no significant difference from uncertain (3) as shown in table 4. Additionally, there was no significant difference in opinion for either question between the students that had made the exercise in years 2005 and 2007, as shown in table 5. This indicates that the quality of the prelab system the students have been using has no bearing on this opinion.

[ht]



(a) Programming exercise



(b) Computer engineering

Fig. 2: Comfortability of automated and manual systems

The clear reservation expressed in these results towards automatic assessment is not in line with the positive results gathered in the rest of the survey. It is clear that the students are wary of giving their automatic acceptance for any kind of automated

Q1 Computer Engineering	Mean	std. error	t	df	p
Comparison value 3 (uncertain)			-5.407	104	< 0.000
Automated	2.46	0.100			
Q2 Computer Engineering					
Comparison value 3 (uncertain)			-0.783	103	0.435
Automated	2.92	0.098			
Q3 Computer Engineering					
Comparison value 3 (uncertain)			-3.157	81	0.002
Automated	2.61	0.124			
Q4 Programming Exercise					
Comparison value 3 (uncertain)			-3.208	81	0.002
Automated	2.65	0.110			

Q1 Automated systems are a professional method of arranging the pre-exercises
Q2 Pre-exercise shoud always be automated
Q3 Automated systems are a professional method of arranging the review sessions
Q4 Reviews shoud always be automated

Table 4: The significance of survey statistics on professionality and adoption

	t	p
Professional	-1.212	0.229
Should be used	-0.760	0.449

Table 5: Significance of difference of opinion between years 2005 and 2007

system despite of the fact that they consider them professional. This shows that despite the apparent strong trust of technology, students still retain a healthy dose of criticism towards automated systems.

In the script-based automatic assessment in the programming exercise course, similar reservations did not arise from the survey. As shown in table 4, students think that automatic assessment is both more professional and that it generally should be employed.

The difference between the two courses is obvious: whereas automatic pre-exercises are an individual exercise where the student is left alone with the system, in the programming exercise the evaluation script is a tool used by the assistant in a personal meeting with the students. As the quality of the tools has no direct impact on how the students do the exercise and they promote equality in the assessment, this fact leads the students to more readily recommend automated systems when employed by the teaching personnel.

4.4 Change requests are well founded

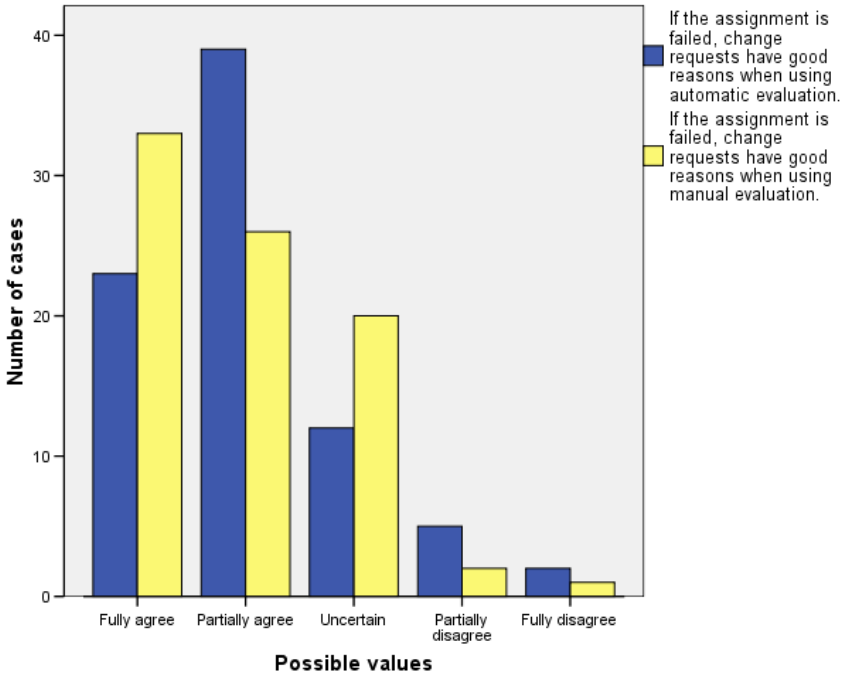


Fig. 3: Change requests are considered reasonable

When asked whether change requests for the failed programs were well justified, the students expressed a strong opinion that the change requests both in manual and automatic assessment in the programming exercise course are well founded as illustrated in figure 3. As visible in table 6, no significant difference exists between the two. This promotes the idea that the students are ready to accept the opinions of automated scripts as equal to the evaluation of the teaching assistants. Additionally this shows that automated systems are not considered to be unjustifiably strict as is sometimes suggested.

4.5 The script does not haggle

A common phenomenon in programming exercise evaluation is the one in which students try to convince the assistant to accept a work that would otherwise be failed. The department statistics show that a considerable amount of students pass the programming exercise as their last course and these are typically the ones with little or no programming skills. The students were asked in the survey whether they thought that a program

	Mean	std. error	95% conf.	t	df	p
Q1 Programming Exercise			[-1.159, 0.406]	0.869	80	0.387
Automated	2.06	0.106				
Manual	1.94	0.103				

Q1 Change requests are well justified

Table 6: The change requests are considered justified in both systems

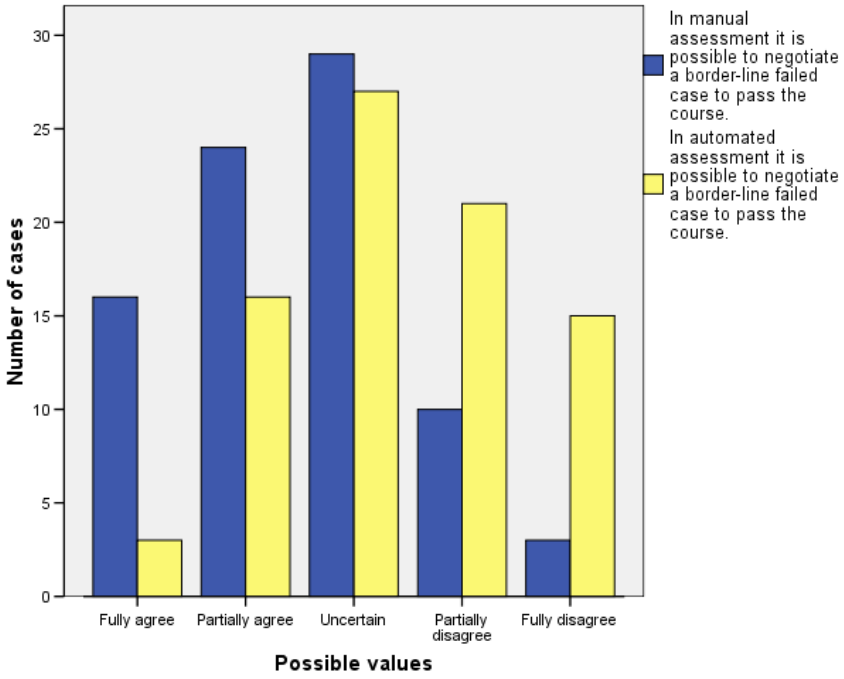


Fig. 4: Is it possible to negotiate a failed program to pass the course?

that was a borderline failure could be negotiated to pass the exercise in both manual and automatic assessment systems.

As shown in the figure 4 and table 7, students clearly express that assessment based on automatic scripts leaves very little room for trying to convince the assistant to accept the assignment. Manual testing showed a clear difference, indicating that the students felt that without the script making the checking it would be significantly easier to get a deficient program to be accepted.

		Mean	std. error	95% conf.	t	df	p
Q1	Comparison			[-1.149, -0.534]	-5.440	81	< 0.000
	Manual	2.51	0.117				
	Automated	3.35	0.122				

Q1 Is it possible to haggle a failed program to pass the course?

Table 7: A significant difference exists between the opinions on automated and manual systems

4.6 Student success has no bearing on opinion on automated systems

	r	p
Computer Engineering		
Q1	0.109	0.272
Q2	0.072	0.463
Programming Exercise		
Q3	-0.100	0.372
Q4	-0.027	0.810

Q1 Correlation of success to opinion that automatic pre-exercise systems should be adopted
Q2 Correlation of success to opinion that automatic pre-exercise systems are more professional
Q3 Correlation of success to opinion that automatic testing scripts should be adopted
Q4 Correlation of success to opinion that automatic testing scripts are more professional

Table 8: Correlation between ease of passing the courses and opinion on automated systems

Further we studied the correlation between the favourable student opinion and the success of the student in the course. The students were asked whether they passed the exercises easily, and this was correlated with whether they thought automated systems should always be adopted and whether automated systems are a professional choice.

The results summarised in table 8 show that there is no correlation to either way in either course. As mentioned before (table 5), whether the student attended the prelab exercises using the inferior 2005 implementation or the much more user-friendly 2006-2007 implementation had no bearing on opinion either. This strongly suggests that the student opinion on automated systems is not dependent on individual experiences of using such systems or their own skill level.

This result indicates that the results of this survey with regards to the adoption of automated systems may well be valid also outside the scope of these two case studies.

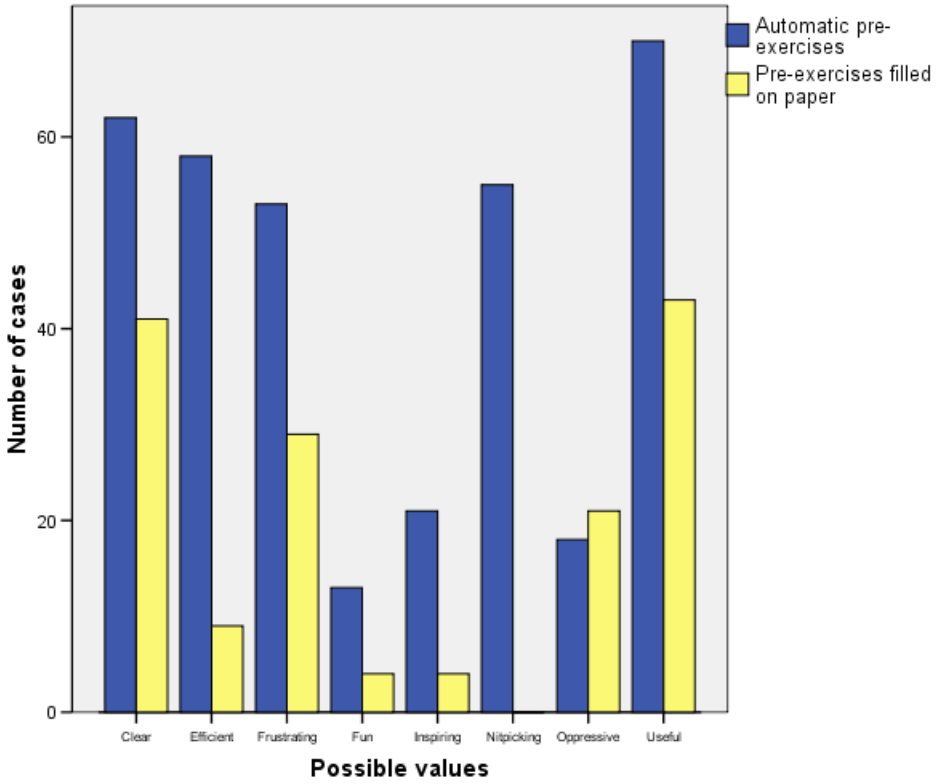


Fig. 5: Adjectives associated with pre-exercises

4.7 Student perceptions of automatic and manual systems

When studying the adjectives that the students associate with the automatic and manual versions in each course, the students had much more to say of the automatic systems, as is visible in the results shown in figures 5 and 6.

The automatic pre-exercises gathered both positive and negative evaluations, receiving a high amount of mentions of being clear, efficient, useful and inspiring, but also for being frustrating and nitpicking. Manual pre-exercises were mostly described as being clear, useful, frustrating and oppressive. Automatic evaluation was described as being clear, efficient and official in contrast to the manual evaluation which was considered clear, educational, friendly and relaxed.

5 Conclusion

The results of the survey were to large extent very positive and they provide a very good justification for the continued use of automated systems.

The students feel that the automated systems treat the students equally, are equally ready to accept feedback from automated and manual systems and feel that when assessed using automated tools, students are not allowed to talk their way into passing exercises when they do not deserve it. Additionally, the students consider these systems to be professional and endorse the adoption of automated tools for student work evaluation.

Students also thought adopting automatic tools for evaluation was a good choice. In contrast they did not express similar opinion on automated pre-exercises. This opinion reflects the students worry that automated systems might become too adopted and that the amount of human contact in teaching might be minimised in the name of progress,

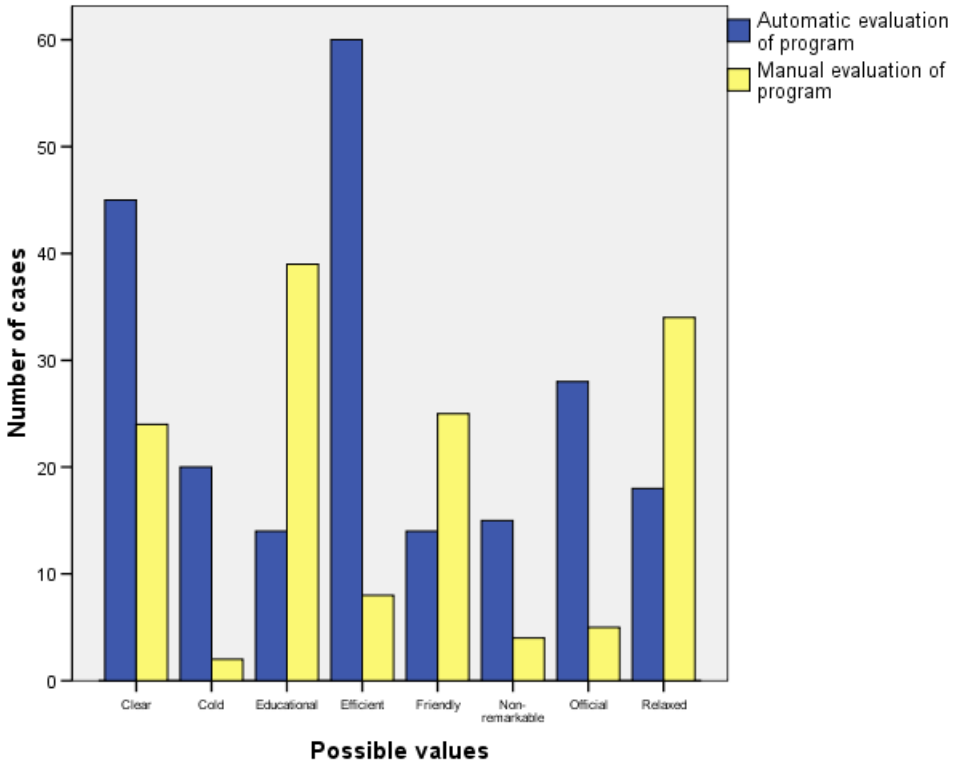


Fig. 6: Adjectives associated with evaluation situations

incidentally allowing teaching personnel to do more research and write papers instead of helping students.

6 Discussion

Related research and the past experiences in these courses clearly show that automated systems have undeniable advantages in achieving good teaching results [4, 5]. The results of this survey are a strong, additional justification for the continued development and deployment of automated systems both in these and other courses. Students want to be treated equally and consider automated systems to do that better than manual approaches.

The reservations students expressed towards the adoption of automated teaching systems is a clear reminder to teaching personnel that develop or deploy automated systems: automated systems should not be an automated choice, their quality and results should be carefully monitored and students must not be deprived completely of human contact. Especially one-on-one contact with teaching personnel is clearly valued by the students, who had no reservations of automated systems when used by teaching personnel in their presence.

It is obvious that in the future teaching will be more and more centred around automated learning environments in the web. Additionally, evaluation and assessment of student work will be more and more dependent on automatic tools such as [2] - the last bastion of manual evaluation will probably be essays in non-technical subjects.

The authors believe that light-weight solutions that can be tailored to the needs of individual courses are going to represent the main-stream of adopted solutions. Although institutions have a tendency to favour heavy, cumbersome and commercial solutions, individual course teachers and assistans are likely to prefer customisable solutions built around freely available open-source implementations. This is also the aspiration for the future development of the RAIPPA system.

While such systems are deployed, further studies both on the impact on the learning of students and on the development of the attitude of students towards such systems is important to assess both the development of the quality of teaching systems and the impact they have on student motivation and attitude.

References

1. `http://trakla.cs.hut.fi`
2. `http://www.dcs.warwick.ac.uk/boss/`
3. `http://goblin.tkk.fi/goblin /`
4. Mikko-Jussi Laakso and Tapio Salakoski The feasibility of automatic assessment and feedback University of Turku, Department of Information Technology
5. Ari Korhonen, Lauri Malmi, Pertti Myllyselkä, Patrik Scheinin, Does it make a difference if students exercise on the web or in the classroom?, Department of Computer science and engineering, Helsinki University of Technology, Department of Education, University of Helsinki
6. Siva R. Sankaran, Dalila Sankaran and Tung X. Bui: Effect of Student Attitude to Course Format on Learning Performance: An Empirical Study in Web vs. Lecture Instruction

7. http://www.ee.oulu.fi/opiskelu/kurssit/521267A.shtml
8. http://www.ee.oulu.fi/opiskelu/kurssit/521482A%20.shtml
9. http://www.ee.oulu.fi/research/tklab/courses/521415A/prelab/RAIPPA.pdf
10. http://www.ee.oulu.fi/research/tklab/courses/521415A/prelab/kysely.php?stats=1
11. http://www.ee.oulu.fi/research/tklab/courses/521415A/prelab/kysely.php?stats=2