Effectiveness of an Online Automated Evaluation and Feedback System in an Introductory Computer Literacy Course

İlhan Varank, M. Fatih Erkoç, Meryem Köşkeroğlu Büyükimdat,
Mehmet Aktaş & Sabiha Yeni
Yildiz Technical University, TURKEY

Tufan Adıgüzel
Bahcesehir University, TURKEY

Zafer Cömert
Bitlis Eren University, TURKEY

Esad Esgin
Marmara University, TURKEY

The purpose of this study was to investigate the effectiveness of an online automated evaluation and feedback system that assessed students' word processing assignments prepared with Microsoft Office Word. The participants of the study were 119 undergraduate teacher education students, 86 of whom were female and 32 were male, enrolled in different sections of Computer-I course taught at one of the major public universities in Istanbul, Turkey. A total of 52 and 67 participants were assigned to the control and experimental group, respectively. No statistically significant difference was found between the experimental and control group students' post-tests performance, self-efficacy perception and technology acceptance scores after the implementation in which the experimental group students used the online automated evaluation and feedback system to get feedback on their assignments, and the control group students didn't receive any feedback. However, the interview results showed that the experimental group students had positive experiences with the system such as contributions to their learning performance, high perceptions, easy use of the system and saving time for the assignments.

Keywords: Information Technologies, Primary Education, Self-efficacy, Use of Technology

INTRODUCTION

Providing feedback to students in order to effectively support their learning processes and performances is crucial (Narciss, Körndle, Reimann, & Müler, 2004). Feedback can be defined as an informative response to a person as a result of his/her action to correct it or prevent reoccurrences of similar actions. The person receiving feedback is expected to consider his/her real performance within the framework defined by the feedback (Vasilyeva et al., 2007).

The term feedback has different interpretations in the context of education. For instance, Tosti and Jackson (1999) consider feedback as either a skill or information. As a skill, it aims to further enhance performance resulted from learning. As information, it is to provide information about impacts of skills. Feedback is also considered as (1) communication

State of the literature

- The study aims to investigate the effectiveness of an online automated evaluation and feedback system that assessed students' word processing assignments prepared with Microsoft Office Word.
- Feedback can be defined as an informative response to a person as a result of his/her action to correct it or prevent reoccurrences of similar actions.
- There are several taxonomies developed to classify feedback systems such as (1) predefined/adapted feedback developed for a specific user group, (2) compatible/adaptable feedback editable by users in a process of interaction and (3) dynamic/adaptive feedback varied according to individual characteristics and performance of users.

Contribution of this paper to the literature

- Although there are numerous other automated assessment and feedback systems to provide students feedback on their coding assignments, an extensive literature review did not yield any study investigating a similar system for word processing assignments
- Although the experimental and control group students' post-test scores did not differ significantly on all three measures — (1) learning performance and (2) self-efficacy perceptions and (3) technology acceptance — the automated assessment and feedback system was found to be an effective system according to the qualitative data analysis results

established to indicate accuracy of learner's response to an instructional question, (2) information presented to the learner for shaping his/her perception (Mory, 2003), (3) any text or displayed messages shown to learner followed by his/her answer in a technology-supported learning environment (Wager & Wager, 1985) and (4) —from a narrower perspective—, verbal responses to inform learners whether they correctly answered questions or they produced correct solutions to problems (Driscoll, 1993).

There are several taxonomies developed to classify feedback systems. For instance Vasilyeva et al. (2007) identified three types of feedback: (1) predefined/adapted feedback developed for a specific user group, (2) compatible/adaptable feedback editable by users in a process of interaction and (3) dynamic/adaptive feedback varied according to individual characteristics and performance of users. In another classification of Vasilyeva et al. (2007), feedback was categorized by its purpose (positive, negative and neutral), timing (instant, delayed and random), form of

presentation (textual, graphical, animated and with-voice), target audience (individual and group), grading system (process-driven and results-based), progress (instant, continuous and summative) and function (confirmative, informative, corrective, descriptive, evaluative, rewarding, motivating, critical and remarkable). In another classification by Çalışkan (1998), seven feedback processes were proposed: (1) students may not receive any feedback, (2) students may just give a right answer and receive an interpreted feedback for their right answer, (3) students may just give a wrong answer and receive an interpreted feedback for the wrong answer, (4) students may receive feedbacks only for truth or falsity of their answers (i.e. in the form of "your answer is correct or incorrect"), (5) students may receive a feedback about what is the right answer if their response is incorrect, (6) students may receive a feedback about why their response is incorrect and (7) students may receive a feedback about accuracy of their response or error rates without any comment.

Numerous research studies were conducted to determine the effectiveness of different feedback systems. For example, Narciss, Körndl, Reimann, and Müller (2004) found that students' success was related to number of times when Informative Tutoring Feedback (ITF) was used, a kind of feedback providing strategically useful information without giving correct answer immediately. Comparing the effectiveness of feedbacks given by teachers through handwriting and text on the computer, Russell (1992) found that feedback given to students as a text on the computer to improve their speaking skills was more effective than the one given as a handwritten text. The level of detail provided in a feedback message, called clarity, is associated with high learning performance (Annett, 1969; Schmidt, 1991). Similarly, as openness increases, feedback progressively focuses on specific behaviors and gives more information about the cause of errors. (Annett, 1969; Baron, 1988; Goldstein, Emanuel, & Howell, 1968; Payne & Hauty, 1955; Wentling, 1973).

Clarity of feedback on learning depends on size of learned skills (Goodman & Wood, 2004) and has a significant impact on learning in general, especially, when students perform unfamiliar tasks (Kopelman, 1986). Feedback given later was found to be more effective than immediate feedback in regular teaching and learning processes (Kulik & Kulik, 1988) as well as in a computer-aided instruction context (Clariana, Wagner, & Murphy, 2000). Barringer and Gholson (1979) found that symbolic and verbal feedbacks were more effective compared to rewards given in a computer-based learning environment. They also found feedbacks given for incorrect answers were more effective than the ones given for correct answers. However, the opposite situation was true for software

programs used for teaching specific skills (Dempsey & Driscoll, 1989).

On the other hand, considering the effectiveness of feedback in online learning environments gains more importance due to the fact that getting help and feedback from their teachers in such environments enables students to be aware of their teacher's existence in the social environment and helps students reduce the feeling of loneliness outside of the classroom (Henninger & Wiswanathan, 2004). Feedback is also needed in courses where intensive practices are required. For example, in a course to gain computer skills, students have to transfer their computer knowledge from theory to practice with a quite number of exercises and assignments, and get feedback from the instructors. However, some limitations such as growing number of students, restrictions of learning context and lack of classroom time, may prevent instructors from providing enough feedback for students' works. Therefore, several automated evaluation and feedback systems have been proposed to overcome this problem in the literature.

In a course, teaching programming microprocessors and cache systems, an automated feedback system was developed using C and C++. This system evaluated students immediately after they wrote the code. The evaluation was done by email until the system fully corrected students' codes. The system significantly reduced the burden of the teaching staff in the evaluation process. Another advantage of the system was that it removed human errors from the evaluation process (Chen, 2004).

A similar automated evaluation system, Scheme-Robo, was developed by Saikkonen, Malmi, and Korhonen (2001). This system provided instant feedback for online students to give them a chance to fix errors in software codes they wrote. Majority of the students found the automated evaluation system excellent and they believed it perfectly evaluated their codes and corrected the errors. Alemán (2011) also investigated the effect of an automated assessment system, Mooshak, on students' attitudes towards the system and their coding performance. He used the system in a programming course in a university. The results of the study revealed that there was a significant difference between the experimental group students who used the assessment system and the control group students who did not used the system in favor of the experimental group in terms of performance and attitude.

Two automated code evaluation systems were developed in Nottingham University, called Ceilidh and CourseMarker evolved from Ceilidh. Using an artificial intelligence method, Ceilidh gave feedbacks automatically to both students and instructors on students' programming assignments in various programming languages such as C and C++. A more advanced version of Ceilidh including marking tools such as Typographic, Dynamic, Feature, Flowchart, Object-Oriented and CircuitSim to assess different metrics of programs, CourseMarker gave instant feedback on students' programming assignments submitted online. Both systems were found to be effective to increase the students' learning performance and attitudes (Foubister, Michaelson, & Tomes, 1997; Higgins et al., 2005)

Although there are numerous other automated assessment and feedback systems, such as TRAKLA2 (Laakso et al., 2005a; Laakso et al., 2005b), Homework Project Generation and Grading (Morris, 2003), Autograder (Helmick, 2007), Online Judge (Cheang et al., 2003), AutoLEP (Wang et al., 2011) and ASSYST (Jackson & Usher, 1997; English, 2004), to provide students feedback on their coding assignments, an extensive literature review did not yield any study investigating a similar system for word processing assignments in introductory computer literacy courses.

The purpose of this study was to examine the effectiveness of an Online Automated Evaluation and Feedback System (OAEFS) that evaluated students' word processing assignments prepared with Microsoft Office Word. Specifically, the following research questions were investigated:

- *Is there a significant difference in the final learning performance of students whose word processing assignments were evaluated by the OAEFS as compared to those who did not received automated feedback?*

- *Is there a significant difference in the final self-efficacy perceptions of students whose word processing assignments were evaluated by the OAEFS as compared to those who did not received automated feedback?*

- *Is there a significant difference in the final technology acceptance of students whose word processing assignments were evaluated by the OAEFS as compared to those who did not received automated feedback?*

METHOD

This study used an explanatory mixed method that conducts a follow-up qualitative study after a quantitative one in order to (1) enrich the research quality by minimizing the possible biases which may occur due to the researchers or the nature of the research and (2) make the results more valid (Creswell, 2003). A pre-test/post-test control group design was used to evaluate the effectiveness of the OAEFS. The students in the experimental group used the OAEFS with full functions (receiving an assignment with criteria, submitting the assignment, receiving an automated feedback on the assignment, and correcting and resubmitting it) while the students in the control

group used the OAEFS for only two functions which were receiving and submitting the assignments.

Participants

A total of 119 undergraduate teacher education students enrolled in Computer-I course at one of the major public universities in Istanbul, Turkey participated in the study. Computer-I is a mandatory course for the college freshman, which is offered in fall semesters, and covers basic information technology literacy and its practices. However, students have an opportunity to be exempt from this course when they successfully pass the exemption exam administered at the beginning of the semester. Therefore, the students who did not pass the exam and participated in this study might lack of prerequisite knowledge of Computer-I course as well as basic technology skills.

A total of three sections of Computer-I course, out of five, were randomly assigned to the experimental group and two sections were assigned to the control group. The students' demographic information by groups (experimental and control) and sections/departments are summarized in Table 1.

Data collection instruments

Four data collection instruments were used in this study: (1) Word Processing Skills Self-Efficacy Perception Questionnaire (WPSSEPQ), (2) Word Processing Skills Performance Test (WPSPT), (3) Technology Acceptance Questionnaire (TAQ) and (4) an interview protocol.

Developed by the researchers and measuring to what extent participants believe they have word processing skills, WPSSEPQ included 32 five-point likert-type items (5 = Very Good, 4 = Good, 3 = Fair, 2 = Poor, 1 = Very Poor). The European Computer Driving License (ECDL) competencies were used to write the items of the questionnaire. ECDL Foundation, a non-profit organization, developed the globally recognized ECDL program in 1995 in order to certify Information and Communication Technology (ICT) and digital literacy qualifications. ECDL has a set of competencies in different areas one of which is word processing (Carpenter, Dolan, Leahy, & Sherwood-Smith, 2000). The items in the questionnaire were written in a way that each of which was matched with a word processing competency of ECDL. Then, five educational technology and one computer science experts reviewed the items in the questionnaire to ensure the face validity. Some sample items are: *I can change the font size of a selected text, I can change the color of a selected text*, and *I can change the line spacing of paragraphs*.

Also developed by the researchers based on ECDL word processing competencies, WPSPT included 20 multiple-choice questions measuring participants' performance on word processing skills. Similar to the self-efficacy perception questionnaire, it was reviewed by the same group of experts to ensure the validity. Some sample questions of WPSPT are: *Which of the following steps should be followed to add numbering or bullets to the text?, which of the following buttons is used to align a selected text to the right?* and *which of the following buttons is used to set before/after spacing and line spacing of paragraph*?

The TAQ, a modified version of the original Technology Acceptance Model (TAM) instrument, was used to determine to what extent the students accepted the OAEFS. The TAM is a well-researched instrument with a historical precedent in the validity and reliability of scores obtained from previous research studies (Teo, Ursavaş, & Bahçekapılı, 2012; Akça & Özer, 2012; Adiguzel, 2008; Usluel Koçak & Mazman, 2009; Ma, Andersson, & Streith, 2005; Hu, Chau, Liu Sheng, & Tam, 1999; Taylor & Todd, 1995; Davis, Bagozzi , & Warshaw, 1989). The TAQ was modified from several variations of the TAM and had four sub-constructs: Perceived Usefulness (PU), Perceived Ease of Use (PEU), Intention to Use (IU), and Subjective Norm (SN). The questionnaire included a total of 21 five-point likert-type items (5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, and 1 = strongly disagree) 5 of which belonged to IU, 6 belonged to PU, 7 belonged to PEU and 3 belonged to SN.

The reliability coefficients of the two questionnaires (WPSSEP and TAQ) were calculated. The alpha

Table 1. Demographic Information by Groups and Sections/Departments

Group	Sections/Departments	Gender (N) Female	Male	Total (N)
Experimental	Elementary Social Science	12	17	29
	Counseling	21	6	27
	Preschool	10	1	11
	Total	43	24	67
Control	Elementary Mathematics	24	7	31
	Elementary Science	20	1	21
	Total	44	8	52
Total		87	32	119

reliability coefficients of Skills Self-Efficacy Perception Questionnaire and Technology Acceptance Questionnaire were same, 0.96.

A total of seven following semi-structured interview questions were used to understand the experimental group students' experiences with the OAEFS: (1) Were you able to use the system easily? (2) Were you able to get help when you needed? (3) Did getting feedback on your assignment affect your next assignments in a positive way? (4) What were the strengths of the OAEFS? (5) What were the weaknesses of the OAEFS? (6) What were your other experiences while using the system? and (7) Would you like to use a similar system in other courses or in the future and why?

The online automated evaluation and feedback system (OAEFS)

The OAEFS is a web-based application developed to evaluate and give instant feedback on students' word processing assignments prepared with Microsoft Office Word based on certain criteria derived from ECDL competencies such as changing the font type, size or color of the text, underlining the text, aligning the text to the right, left, center, or justified, setting line and paragraph spacing etc.

The OAEFS in this study served to two types of users: Instructors and students. The instructors define word processing assignments for their students. Initially, the instructor created the assignment step by step. First, he/she named the assignment, set starting-ending dates to submit the assignment and determined the number of the paragraphs in the assignment. Then, formatting criteria for each paragraph were defined and saved in the system. There were a total of 32 formatting criteria taken from ECDL competencies. Then, the system automatically created the downloadable Microsoft Word document as an assignment for students. After creating the assignment, only the list of criteria, not the created Microsoft Word document as an assignment, was sent to the students' accounts by the system so that they could see it under their assignment section to complete.

After receiving the assignment via the system, the students created a Microsoft Word file based on the criteria given and uploaded it to the system. Then, the system provided an immediate feedback to the students through a pop-up window showing what criteria were met in the assignment. For each criterion that was not met, the system created a link to a related instructional video. The students watched the video to be able to learn how to correct their mistakes in the assignment. The system also allowed the students to re-upload their assignments as many times as they wanted until they corrected all the mistakes.

Data collection and analysis

In order to address the research questions of this study, quantitative and qualitative data were collected. The quantitative data was gathered from the participants' test and questionnaire scores, and qualitative data was gathered from the focus group interview with the experimental group students. The quantitative data was collected from the participants before and after and the qualitative data was collected after the implementation process that lasted 10 weeks in the 2013 fall semester.

Prior to the implementation, several meetings were held with the two course instructors to prepare (1) a common course syllabus to ensure that the sections of the course were equivalent and (2) five assignments for students to complete during the implementation. Both experimental and control group students were provided face to face orientations and printed instructions on the use of the system. Both groups were told that they would submit their five assignments via the system. However, only the experimental group students were told their assignments would be graded and they would get immediate feedbacks on the assignments. Although the study was initiated with 119 students the same number was not retained because of attendance problems and invalid responses to data collection instruments.

At the beginning of the study, both experimental and control group students completed the word processing skills performance test in the classroom ($n = 110$), and the word processing skills self-efficacy perception questionnaire ($n = 103$) and the technology acceptance questionnaire online ($n = 104$) as pre-tests. Then, the instructors gave the both groups of the students the regular instruction on word processing using Microsoft Word and assigned them tasks periodically to complete and submit via the system. While the experimental group students received immediate feedback and corrected their errors in the assignment, the control group students received no feedback. Following the implementation, all students completed the same scales as post-test. Additionally, the researchers conducted two focus group interviews with a total of nine experimental group students after the implementation.

Mann-Whitney U-test was used to determine whether significant differences occurred between the experimental and control group students' WPSSEP, WPSPT and TAQ scores since the scores were not normally distributed (as given in the results section). Qualitative analysis was also used for interview data. In the qualitative analysis, the students' verbal responses given in the interviews were analyzed to find out whether significant and/or non-significant statistical results were validated. The qualitative and quantitative

Table 2. Mann-Whitney U Test Results of Students' Learning Performance Post-test Scores

	Group	N	Mean	SD	Mean Rank	Sum of Ranks	U	p
Performance	Experimental	57	14.77	2.28	52.27	2979.50	1326.50	.340
	Control	52	15.13	2.34	57.99	3015.50		

Table 3. Mann-Whitney U Test Results of Students' Self-Efficacy Post-test Scores

	Group	N	Mean	SD	Mean Rank	Sum of Ranks	U	p
Self-Efficacy	Experimental	47	109.02	18.32	46.81	2171.00	1072.00	.945
	Control	46	111.70	13.78	47.20	2200.00		

Table 4. *Mann-Whitney U Test Results of Students' Technology Acceptance Post-test Scores*

	Group	N	Mean	SD	Mean Rank	Sum of Ranks	U	p
Technology Acceptance	Experimental	49	76.41	20.46	42.62	2088.50	863.50	.070
	Control	45	83.89	15.80	52.81	2376.50		

analyses were performed separately but the results were combined in the interpretation.

RESULTS

The data from the instruments were analyzed to determine whether there were significant differences between the experimental and control group students on scores of learning performance, self-efficacy perception and technology acceptance. The normality analyses revealed that pre- and post-test results were not normally distributed ($p < .05$). Therefore, non-parametric Mann-Whitney U-test was used in the analyses to investigate the differences between the control and experimental groups.

Mann-Whitney U-test results showed that there was no statistically significant difference between the experimental and control group students' pre-tests performance scores (Mann-Whitney U = 1373.00, $p > .05$), self-efficacy scores (Mann-Whitney U = 1041.00, $p > .05$) and technology acceptance scores (Mann-Whitney U = 1196.50, $p > .05$) indicating that both groups were equivalent at the beginning of the study.

The descriptive and inferential statistical results were reported below according to the research questions based on the three dependent variables: (1) learning performance, (2) self-efficacy perceptions and (3) technology acceptance.

Learning performance

As can be seen in Table 2, statistically significant difference was not found between the experimental and control group students' post-test learning performance scores (Mann-Whitney U = 1326.50, $p > .05$).

The descriptive statistics of the performance instrument items indicated that participants in both

experimental ($n = 57$) and control groups ($n = 52$), except one in the control group, had generally answered more than half of the questions correctly. The post-test performance scores ranged from 8 to 19 in the control group, while they ranged from 10 to 20 in the experimental group.

Using the interview results, the attributes of learning performance on the word processing skills were also assessed. When asked about their experience with the OAEFS, most of the interviewees focused on how the system positively contributed to their performances. For example, one typical comment was, "I learned a lot. To illustrate, I was able to correct my mistakes [by the help of automated feedback]." Possible contributions to the other courses were also mentioned: "I have not known about [Microsoft Word] terminology too much such as line space. The system helped me learn and this helped me use such things in the assignments of the other courses. I couldn't have done [such assignments] easily and quickly." Another participant commented, "I did not know about [Microsoft] Word too much. For example, I did not know about sub-script, super-script and character count. I have learned them and even been helping my brother, who is also taking computer course [at another university], do his assignments when we talk on the phone. . ."

Self-efficacy perceptions

As summarized in Table 3, similar to the learning performance results, no statistically significant difference between the experimental and control group students' self-efficacy post-test scores was found (Mann-Whitney U = 1072.00, $p > .05$).

Students' self-efficacy perception post-test data were analyzed to test the difference between the experimental and control group students' mean scores. As in learning

performance results, there was no statistically significant difference between experimental and control group students' post-test self-efficacy scores (Mann-Whitney U = 1072.00, $p > .05$). With the exception of two items (of the control group data), the descriptive statistics of the self-efficacy instrument items indicated that participants held generally high (mean scores greater than four) self-efficacy perceptions towards the word processing skills. The mean scores ranged from 3.76 to 4.85 in the control group ($n = 46$), while such scores ranged from 4.00 to 4.66 in the experimental group ($n = 47$). The items with the highest and lowest mean scores were the same in both groups.

Self-efficacy perceptions towards the word processing skills were also drawn from participant interviews. The results of the qualitative analysis substantiated students' high self-efficacy perceptions found in the quantitative analysis. Some participants reported: "We learned how to use [Microsoft Word] and where to do something on it. Namely, [the OAEFS] really helped us develop [our word processing skills]." Another one pointed out that "When you directly work on Microsoft Word yourself, you don't know you just do it. However, when you get detailed feedback from [the OAEFS], you feel confident about what to do and correct." Participants also expressed the importance of having instructional video in the system: "There was a video teaching how to perform related skills together with the feedbacks displaying your mistakes in [the OAEFS]. [Such videos] contributed me a lot not to make the same mistake again."

Technology acceptance

Table 4 summarizes the test results for students' technology acceptance. Similar to the learning performance and self-efficacy perceptions results, there was no statistically significant difference between experimental and control groups post-tests technology acceptance scores (Mann-Whitney U = 863.50, $p > .05$).

The descriptive statistics of the technology acceptance instrument items indicated that all participants held generally positive (mean scores greater than three) perceptions towards technology and the OAEFS. The mean scores ranged from 3.42 to 4.47 in the control group ($n = 45$), while such scores ranged from 3.10 to 4.06 in the experimental group ($n = 49$). When looked at the experimental and control group students' technology acceptance post-test sub-construct mean scores (PEU, PU, SN and IU), the ones for PEU and PU in both groups were higher than the other two sub-constructs.

The quantitative data analysis results showed that students' responses regarding technology acceptance were aggregated under two sub-constructs of technology acceptance: PE and PEU. The participants

stated that using the OAEFS was generally easy while the first-time use was sometimes problematic: "We couldn't send the first assignment, but I got help and then I was able to send it." Some others who didn't know too much about how to use computers mentioned the ease of its use: "It was not really hard to use the OAEFS, even easier than our student information system. Where [your assignment] was sent and who would look at it was known. Overall, everything was known." Another commenter noted, "I was able to send [my assignment]. It was easy to do it. However, you could send your assignments even for ten times back to back. This was allowed [by the system]." Regarding perceived usefulness of the OAEFS, the participants' comments cumulated under two themes: time and automation. Most of the interviewees indicated how the system helped them to save time for their assignments and other commitments: "[The system] shorten our time spent for assignment process. For example, instead of waiting for one or more days or asking to get feedback from the course instructor, we were able to get feedback on our mistakes and correct them right away before the second submission." Another one commented "You knew when to send [your assignment]. There was no way to send it later. This really helped us and [our instructor] to be more organized. Otherwise, nothing was clear about what and when to do." When automation is considered, many commenters pointed the importance of having automated feedback especially in a course with more than 50 students: "It was really hard [for the instructor] to be involved in [such number of students]. There was a need for the system. The instructor can miss something while evaluating such assignments which have too many details such as font types…" And, the final one said "When we submitted our assignments via e-mails, we were having problems while sending and there was no way to notice our mistakes. However, [with the help of the OAEFS], we were able to control and learn where we did mistakes one by one instantly and automatically."

DISCUSSION AND CONCLUSION

Although the experimental and control group students' post-test scores did not differ significantly on all three measures — (1) learning performance and (2) self-efficacy perceptions and (3) technology acceptance— the OAEFS was found to be an effective system according to the qualitative data analysis results.

The students' learning performance post-test scores were not significantly different between the experimental and control groups, but none of the experimental group students interviewed commented negatively on the use of the OAEFS for their assignment submission process. When the mean scores

of each group are considered, the control group post-test mean score (15.13) was higher compared to those (14.77) of the experimental group. Namely, the performance of the students who used the OAEFS fully was less than those who used the same system partially. One reason might be that since the students in the control group were from math and science education departments they might gain the word processing skills much quicker than the students in the experimental group who were majoring in social sciences (Can, 2010; Varank, 2007). Another possible reason is that the implementation period of the study was really short to be able to make the experimental group students more familiar with the system. Another evidence was that the performance mean score (14.79) of the students in the experimental group who submitted all the assignments through the system were higher than the ones (12.00) who submitted only one assignment. In general, it might be reasonable to say that an automated evaluation and feedback system in courses can effectively replace teachers' role to provide feedback (Chen, 2004; Barker, 2010; Laakso, Salakoski, & Korhonen, 2005b; Malmi, Korhonen, & Saikkonen, 2002).

The similar finding was found for students' self-efficacy perception. That is, although there was no significant difference between the experimental and control group students' self-efficacy perception post-test scores, students' comments in the interview pointed the positive contribution of the OAEFS to their perceptions of the word processing skills. When the self-efficacy perception mean scores were analyzed separately, the control group students' mean score (111.69) was again higher than the experimental group students' mean score (109.02). Although the interview results revealed that the system helped students feel competent on the word processing skills, having this finding in favor of the control group which is parallel to learning performance attribute might be a reason since the past research refers that self-efficacy perceptions and performance are constructed on the same kind of skills which produce similar results (Yi & Im, 2004; Brosnan, 1998; Gist, Schwoerer, & Rosen, 1989).

Furthermore, the self-efficacy perception mean score of the students who submitted all the assignments through the system in the control group (114.40) were higher than the ones (110.00) in the experimental group. Overall, this result is plausible given that the use of the OAEFS might not directly affect students' self-efficacy perceptions of the word processing skills since such skills can be easily performed by the participating students who grew up as digital natives or may easily interact with technology anywhere around them.

The groups of students were not significantly separated on technology acceptance as well. Although this finding was parallel to the ones obtained from the learning performance and self-efficacy perception data,

technology acceptance mean score difference between the experimental and control groups was higher compared to the previous two attributes. One factor that could affect this difference is that less use of the system may increase technology acceptance. The evidence was that some students interviewed from the experimental groups emphasized technical problems while using the system especially for the first assignment. Therefore, full use of the system with the problems might affect students' intentions to use the system.

On the other hand, the experimental group students' comments were parallel to the two technology acceptance sub-construct mean scores, which mean that students found the OAEFS useful and easy to use, and the technology acceptance scores revealed that perceived usefulness and perceived ease of use were the main contributors to students' technology acceptance.

Finally, this study might contribute to the technology acceptance literature that students who use the system with some technical problems in a short-time period without getting familiar and confident with it may resist accepting this technology fully when compared to the ones who used such system for a limited purpose (Adiguzel, Capraro, & Willson, 2011).

REFERENCES

Adiguzel, T. (2008). *Dependability and acceptability of handheld computers in school-based data collection.* Doctoral dissertation, College Station, TX: Texas A&M University. Retreived from http : / /hdl .handle .net /1969 .1 /ETD -TAMU -2958

Adiguzel, T., Capraro, R. M., & Willson, V. L. (2011). An examination of teacher acceptance of handheld computers. *International Journal of Special Education, 26*(3), 12-27.

Akça, Y. & Özer, G. (2012). The Use of Technology Acceptance Model in Enterprise Resource Planning Implementation. *Business and Economics Research Journal, 3*(2), 79-96.

Alemán, J. L. F. (2011). Automated assessment in a programming tools course. *IEEE Transactions on Education, 54*(4), 576-581.

Annett, J. (1969). *Feedback and human behavior.* Hammondsworth, England: Penguin.

Barker, T. (2010). An Automated Feedback System Based on Adaptive Testing: Extending the Model. *International Journal of Emerging Technologies in Learning, 5*(2), 11-15.

Baron, R. A. (1988). Negative effects of destructive criticism impact on conflict, self-efficacy, and task performance. *Journal of Applied Psychology,73*, 199–207.

Barringer, C. & Gholson, B. (1979). Effects of type and combination of feedback upon conceptual learning by children: Implications for research in academic learning. *Review of Educational Research, 49*(3), 459-478.

Brosnan, M. J. (1998). The impact of computer anxiety and self-efficacy upon performance. *Journal of computer assisted learning, 14*(3), 223-234.

Can, Ş. (2010). Comparison of the attitudes of the pre-service teachers from SESMTE and SESSTE departments against computer use (The case of Muğla University, Turkey). *Procedia-Social and Behavioral Sciences*, *2*(2), 4638-4642.

Carpenter, D., Dolan, D., Leahy, D., & Sherwood-Smith, M. (2000). ECDL/ICDL: a global computer literacy initiative. In *Proceedings of Conference on Educational Uses of Information and Communication Technologies (ICEUT), Beijing: Publishing House of Electronics Industry (PHEI)*.

Cheang, B., Kurnia, A., Lim, A., & Oon, W. C. (2003). On automated grading of programming assignments in an academic institution. *Computers & Education, 41*(2), 121-131.

Chen, P. M. (2004). An Automated Feedback System for Computer Organization Projects. *IEEE Transactions on Education , 47*(2), 232-240.

Clariana, R. B., Wagner, D., & Murphy, L. C. R. (2000). Applying a Connectionist Description of Feedback Timing. *EducationalTechnology Research and Development, 48*(3), 5-21.

Creswell, J. W. (2003). *Qualitative, quantitative, and mixed methods approaches* (2nd Ed.) Thousand Oaks, CA: Sage.

Çalışkan, H. (1998). Bilgisayar destekli öğretimde geri bildirim. *Kurgu Dergisi, 15,* 346-360.

Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science, 35*(8), 982-1003.

Dempsey, J. V., & Driscoll, M. P. (1989). The Effects of Four Methods of Immediate Corrective Feedback on Retention, Discrimination Error, and Feedback Study Time in Computer-Based Instruction. *American Educational Research Association Annual Meeting*, San Francisco, CA.

Driscoll, M. P. (1993). *Psychology of earning for instruction*. Needham Height, MA: Ally&Bacon.

English, J. (2004). Automated assessment of GUI programs using JEWL. *ACM SIGCSE Bulletin, 36*(3), 137-141.

Foubister, S. P., Michaelson, G. J., & Tomes, N. (1997). Automatic Assessment of Elementary Standard ML Programs Using Ceilidh. *Journal of Computer Assisted Learning , 13,* 99-108.

Gist, M. E., Schwoerer, C., & Rosen, B. (1989). Effects of alternative training methods on self-efficacy and performance in computer software training. *Journal of applied psychology, 74*(6), 884.

Goldstein, I. L., Emanuel, J. T., & Howell, W. C. (1968). Effect of percentage and specificity of feedback on choice behavior in a probabilistic information-processing task. *Journal of Applied Psychology, 52*, 163–168.

Goodman, J. S., & Wood, R. E. (2004). Feedback Specificity, Learning Opportunities, and Learning. *Journal of Applied Psychology, 89*(5), 809–821.

Helmick, M. T. (2007). Interface-based programming assignments and automatic grading of java programs. *ACM SIGCSE Bulletin, 39*(3), 63-67.

Henninger, M., & Viswanathan, V. (2004). Social presence in online tutoring: What we know and what we should know. In *Proceedings of the first joint meeting of the EARLI SIGs instructional design and learning and instruction with*

computers (CD-ROM). Tuebingen: Knowledge Media Research Center.

Higgins, C. A., Gray, G., Symeonidis, P., & Tsintsifas, A. (2005). Automated assessment and experiences of teaching programming. *Journal on Educational Resources in Computing (JERIC), 5*(3), 5.

Hu, P. J., Chau, P. Y., Liu Sheng, O. R., & Tam, K. Y. (1999). Examining the technology acceptance model using physician acceptance of telemedicine technology. *Journal of Management Information Systems, 16*(2), 91-112.

Jackson, D., & Usher, M. (1997). Grading student programs using ASSYST. *ACM SIGCSE Bulletin, 29*(1), 335-339.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed methods research: A research paradigm whose time has come. *Educational Researcher, 33*(7), 14-26.

Kopelman, R. E. (1986). Objective feedback. In E. A. Locke (Ed.), *Generalizing from laboratory to field settings* (pp. 119–145). Lexington, MA: D. C. Heath.

Kulik, J. A., & Kulik, C. C. (1988). Timing of feedback and verbal learning. *Review of Educational Research, 58*(1), 79-97.

Laakso, M. J., Salakoski, T., Grandell, L., Qiu, X., Korhonen, A., & Malmi, L. (2005a). Multi-perspective study of novice learners adopting the visual algorithm simulation exercise system TRAKLA2. *Informatics in Education: An International Journal, 4*(1), 49-68.

Laakso, M. J., Salakoski, T., & Korhonen, A. (2005b). The Feasibility of Automatic Assessment and Feedback. *IADIS International Conference on Cognition and Exploratory Learning in Digital Age* (pp. 113-122). Porto: IADIS.

Ma, W. W., Andersson, R., & Streith, K. O. (2005). Examining user acceptance of computer technology: An emprical study of student teachers. *Journal of Computer Assisted Learning, 21*, 387-395.

Malmi, L., Korhonen, A., & Saikkonen, R. (2002). Experiences in automatic assessment on mass courses and issues for designing virtual courses. *ACM SIGCSE Bulletin, 34*(3), 55-59.

Morris, D. S. (2003). Automatic Grading of Student's Programming Assignments: An Interactive Process and Suite of Programs. *33rd Annual Frontiers in Education. 3*, S3F1-S3F6. Westminster, USA: FIE.

Mory, E. H. (2003). Feedback research revisited. In Jonassen, J.H. (Ed.): *Handbook of Research on Educational Communications and Technology*, MacMillian Library Reference, New York, pp.745–783.

Narciss, S., Körndle, H., Reimann, G., & Müller, C. (2004). Feedback-seeking and feedback efficiency in web-based learning – How do they relate to task and learner characteristics? In Gerjets, P., Kirschner, P. A., Elen, J. & Joiner, R. (Eds.). *Proceedings of the first joint meeting of the EARLI SIGs Instructional Design and Learning and Instruction with Computers* (pp. 377-388). Tuebingen: Knowledge Media Research Center.

Payne, R. B., & Hauty, B. T. (1955). Effect of psychological feedback upon work decrement. *Journal of Experimental Psychology, 50*, 343–351.

Russell, B. W. (1992). *The effect of computer-generated instructional feedback and videotape on the speaking performance of college students in a basic speech course.* Doctoral dissertation, Pittsburgh, PA: University of Pittsburgh.

Saikkonen, R., Malmi, L., & Korhonen, A. (2001). Fully Automatic Assessment of Programming Exercises. *6th Annual Conference on Innovation and Technology in Computer Science Education* (pp. 133-136). New York: ACM.

Schmidt, R. A. (1991). Frequent augmented feedback can degrade learning: Evidence and interpretations. In J. Requin & G. E. Steimach (Eds.), *Tutorials in motor neuroscience* (pp. 59–75). London: Kluwer Academic.

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research*, *6*(1), 144-176.

Teo, T., Ursavaş, Ö. F., & Bahçekapılı, E. (2012). An assessment of pre-service teachers' technology acceptance in Turkey: A structural equation modeling approach. *The Asia-Pacific Education Researcher, 21*(1), 191-202.

Tosti, D., & Jackson, S. (1999). Feedback. In H. Stolovich & E. Keeps (ed) *Handbook of Performance Technology*, (2nd. Ed.), (pp. 395-410). San Francisco: Jossey-Bass/Pfeiffer.

Usluel Koçak, Y. & Mazman, S. G. (2009). Sosyal Ağların Benimsenmesi Ölçeği. *Eğitim Bilimleri ve Uygulama Dergisi*, *8*(15), 137-157.

Varank, I. (2007). Effectiveness of quantitative skills, qualitative skills, and gender in determining computer skills and attitudes: A causal analysis. *The Clearing House, 81*(2), 71-80.

Vasilyeva, E., Puuronen, S., Pechenizkiy, M., & Räsänen, P. (2007). Feedback adaptation in web-based learning systems. *International Journal of Continuing Engineering Education and Life-Long Learning, 17*(4/5), 337-357.

Wager, W., & Wager, S. (1985). Presenting questions, processing responses, and providing feedback in CAI. *Journal of Instructional Development, 8*(4), 2–8.

Wang, T., Su, X., Ma, P., Wang, Y., & Wang, K. (2011). Ability-training-oriented automated assessment in introductory programming course. *Computers & Education, 56*(1), 220-226.

Wentling, T. L. (1973). Mastery versus nonmastery instructions with varying test item feedback treatments. *Journal of Educational Psychology, 65*, 50–58.

Yi, M. Y., & Im, K. S. (2004). Predicting computer task performance: personal goal and self-efficacy. *Journal of Organizational and End User Computing (JOEUC), 16*(2), 20-37.

❖ ❖ ❖