

Improvement of Spreadsheet Quality through Reduction of End-User Overconfidence: Case Study

Lazar Raković¹, Marton Sakal¹, Vuk Vuković^{1*}

¹ Department of Business Informatics and Quantitative Methods, Faculty of Economics in Subotica, University of Novi Sad, 24000 Subotica, Segedinski put 9-11, Serbia

* Corresponding author, e-mail: vuk@ef.uns.ac.rs

Received: 12 April 2018, Accepted: 26 October 2018, Published online: 28 January 2019

Abstract

This paper is prompted by and based on earlier research into developers' overconfidence as one of the main causes of spreadsheet errors. Similar to related research, the aim of the paper was to ascertain the existence of overconfidence, and then examine the possibility of its reduction by means of experimental treatment designed for the needs of the research. A quasi-experiment was conducted to this end, in which 62 students of the Faculty of Economics of the University of Novi Sad participated, divided into the experimental and control group. Participants of both groups developed domain free spreadsheets in two iterations each. After the first iterations, students in the experimental group were subjected to experimental treatment: they attended lectures on spreadsheet errors taxonomies supported by real-life examples, and about spreadsheet best practices in the area of spreadsheet error prevention. Results showed that spreadsheet developers who were informed about spreadsheet error taxonomies and spreadsheet best practices create more accurate spreadsheets and are less self-confident in terms of accuracy of their spreadsheets.

Keywords

spreadsheets, spreadsheet development, end user development, spreadsheet errors, spreadsheet taxonomies, spreadsheet best practices

1 Introduction

Shadow IT, manifested in the form of applications and processes developed and/or acquired without the knowledge and/or support of the IT department, exists in organisations in parallel with the official IT infrastructure (Chua et al., 2014; King and Azad, 2014; Rentrop and Zimmermann, 2012; Sakal et al., 2017a; 2017b; Zimmermann et al., 2016). In the broad spectrum of its manifestation forms, ranging from smart phones to complex on-demand cloud services (Kopper and Westner, 2016), a prominent place is taken up by spreadsheets, developed by end users of non-IT domicile, for the needs of performing their own or their colleagues' tasks (Bellino et al., 2010; Hill and Barnes, 2011). The frequency of use of spreadsheets as shadow IT artefacts is not surprising, given that they are one of the first end-user computer tools to achieve widespread popularity (Kruck et al., 2003), owing to their usability and flexibility (Baker et al., 2006; Panko and Aurigemma, 2010; Pemberton and Robson, 2000). Their range of use is broad and diverse, practically universal, ranging from

keeping simple records to supporting key business processes (Panko, 2013; Panko and Sprague, 1999).

Spreadsheets are often associated with the following concepts: End User Development (EUD), End User Programming (EUP); End User Tailoring (EUT); End User Software Engineering (EUSE). (More information on individual terms is given in the paper by Sakal et al., 2016b). End users developing applications for themselves are often called user developers or end user programmers, while those who only use such applications are referred to end user consumers or just end users (Govindarajulu, 2003; Ko et al., 2011). As regards spreadsheet development, they are mostly created and used by the same persons. This paper will use the term spreadsheet developer, i.e. term denoting the end user who creates and uses spreadsheets.

In addition to benefits such as analyze and visualize data sets in accessible ways, spreadsheets often result in errors as well, with devastating financial and reputation-related consequences (EuSpRIG). The examples of spreadsheet errors from practice are justifiably described by the epithet

"horror stories" (Aurigemma and Panko, 2014; Caulkins et al., 2007; EuSpRIG). Panko states that a large portion of spreadsheet errors are never detected, and therefore calls spreadsheets "The Dark Matter (and Dark Energy) of Corporate IT" (Panko, 2013).

In addition to the lack of adequate spreadsheet techniques related knowledge (Ferreira and Visser, 2012; Kulesz and Zitzelsberger, 2012), self-learning and unawareness of the dangers and risks of the use of spreadsheets (Lawson et al., 2009; Powell et al., 2009a; 2009b), the occurrence of spreadsheet errors is also greatly contributed to by developers' overconfidence (Panko, 2007; 2009; Raković, 2014). Similar to related research, the aim of the paper was to examine whether overconfidence, and consequently number of spreadsheet errors, can be reduced by education on error taxonomies, providing examples of the consequences of spreadsheet errors from practice, and pointing to the best practices of spreadsheet error prevention.

In addition to the introduction, the paper is structured into five other sections. Section 2 presents the results of related research that were used to define the research hypotheses and set up the research design. After the research design, described in Section 3, research results are presented in Section 4, and discussed in Section 5. Finally, Section 6 presents conclusive considerations, limitations of the conducted research and proposals for future research. In addition to the list of used sources, the paper also contains four appendices.

2 Related research

Using the domain-free (no need for particular domain knowledge) Wall problem (the task of creating a bid for wall construction, see Appendix 1), Panko (1996) conducted a study in which 72 students participated. Students created a spreadsheet for the Wall problem at home. In addition to developing their spreadsheets, the students filled in a questionnaire to assess their own spreadsheet knowledge, adequacy of their knowledge for resolving tasks, etc. The author included a control group as well, who created a spreadsheet for the same task, but in the classroom. Research results found that the first group of students (the experimental group who created spreadsheets at home) created 38 % incorrect spreadsheets and the other (control group) 30 %. Later, Panko confirmed his research results with his co-author Sprague (Panko and Sprague, 1999), conducting an experiment with a bigger number of participants (152 students who created spreadsheets based on the Wall problem). In this research as

well, the percentage of incorrect spreadsheets was higher than one-third (35 % incorrect spreadsheets). Irons (2003) conducted a study in which, in addition to Panko's Wall problem, he also used the Ball task (a task requiring more advanced knowledge of algebra, and therefore was not domain free). Having built two spreadsheet models, the participants were interviewed. The study only included 11 participants. The participants were students and academic staff (fellows). 18.18 % of spreadsheets built for the Wall problem contained errors, whereas the percentage of incorrect spreadsheets for Ball problem amounted to 71.42 %. Researching the types of errors (quantitative and qualitative errors), which can occur even within the domain-free spreadsheet problems, Teo and Tan (1999) used the modified Wall problem. The research included a total of 176 undergraduate business students, who attended a course in information systems. In the first part of the experiment, the students created spreadsheets at home, like in Panko's study (1996). The second part of the experiment involved creating spreadsheets in the classroom, in order to establish whether there are differences between completing the task at home and in the classroom. The study included 168 developed spreadsheets for both parts of the experiment. In the first part, 41.7 % spreadsheets contained an error, whereas this number was somewhat higher in the second (50 % incorrect spreadsheets).

In addition to studying spreadsheet errors, the researchers also dealt with spreadsheet developers' overconfidence. The influence of awareness of the errors is prominent as one of the ways to influence the reduction in the number of errors. Benham and Giullian (2005) proved that the number of incorrect spreadsheets can be reduced if the spreadsheet developers are aware of the existence of spreadsheet errors. Panko (2007; 2009) points out that overconfidence is high among spreadsheet developers. In the study he conducted, he first proved spreadsheet developers' overconfidence, and then the fact that it can be reduced by warning of spreadsheet errors. Spreadsheet developers' overconfidence was also demonstrated in Raković (2014), which was reached by analyzing the spreadsheets submitted by spreadsheet developers who participated in the study. In most cases, spreadsheet developers underestimated the number of errors, or were completely unaware of the existence of errors in their spreadsheets. Purser and Chadwick (2006) found that the ability to identify spreadsheet errors is directly correlated to spreadsheet experience and awareness of the existence of particular types of errors.

Although human errors are impossible to eliminate (Panko, 2008; Powell et al., 2008; 2009b), influencing the spreadsheet developers' awareness of their existence, as well as the use of certain design guidelines by spreadsheet developers may influence the reduction of the number of errors, with simultaneous improvement of the quality of spreadsheets (Bewig, 2005; Raković, 2014). Kulesz and Zitzelsberger (2012) point out that applying certain good practices and guidelines may improve the quality of spreadsheets, especially after modification of spreadsheets during their use. Some of the guidelines and good practices include: separated spreadsheets inputs and outputs (positions of cells containing inputs separated from the positions of cells containing formulas or functions into two wholes) (Bewig, 2005; Dunn, 2010; Raković, 2014; 2019); recommendation not to use constants (hard-coding, numbers in formulas) in formulas (isolating them into separate cells) (Dunn, 2010; Grossman and Özlük, 2004); set validation on cells for input (e.g. a cell can receive only integers which are in a certain range or only elements belonging to a list) (Bewig, 2005; Raković, 2014; 2019; Read and Batson, 1999); using as simple a formula as possible (separating complex formulas into several cells; the formula should contain as few operands as possible) (Colver, 2004; O'Beirne, 2010) and different formatting spreadsheet inputs and outputs (using different colors for background of cells containing the user's inputs and cells containing formulas or functions) (Raffensperger, 2003; Read and Batson, 1999).

The above described studies and conclusions of individual authors served as the starting point for problem definition and research design. Most related studies used students as participants, which is a good basis for comparing research results. Students who participated in the study were divided into two groups. Both groups created the same spreadsheet for a problem that did not require additional knowledge except the use of spreadsheet programs. After developing the spreadsheets, only the first group of students attended lectures on spreadsheet errors and selected good practices of spreadsheet development. Then both groups once again created spreadsheets for the same domain free problem. The aim was to ascertain whether the lecture that the students attended influenced the number of errors and the quality of the developed spreadsheet. The purpose of the other group of students was to eliminate the possible influence of external variables on the results. A detailed description of research design is presented in the following section.

3 Research design

The following research hypotheses were defined based on the set research objective:

- H1 – If spreadsheet developers are informed about spreadsheet errors and taxonomies and good practices guidelines, they create spreadsheets containing a smaller number of errors.
- H2 – If spreadsheet developers are informed about spreadsheet errors, they will be less overconfident about the accuracy of their spreadsheets.

To examine the set research hypotheses, a quasi-experiment was conducted. The participants were third-year students of undergraduate studies of the Faculty of Economics of the University of Novi Sad. Before the experiment, within the course in Information Technologies, the students completed the elementary and advanced spreadsheet course.

The research design is schematically presented in Fig. 1 (the design was created based on Ristić, 2011, p. 107). To control a possible influence of numerous external variables, two groups of students were used: experimental and control. The groups were not formed by randomizing; the existing groups of students were taken. As the Faculty of Economics of the University of Novi Sad teaches courses in two locations (the cities of Novi Sad and Subotica), students from Novi Sad were used as the experimental group, and students from Subotica were the control group. The students of both groups did not know each other, had a similar previous knowledge (they attended the same courses taught by the same lecturers in each location) and had a similar grade point average. They did not have experience in the use of spreadsheets except for what they acquired in the Information Technologies course. Both groups of students developed a spreadsheet for the Wall problem (based on Panko, 1996, described in Appendix 1) in two iterations. This is a domain free task, which does not require domain-specific knowledge from a specific area. In the

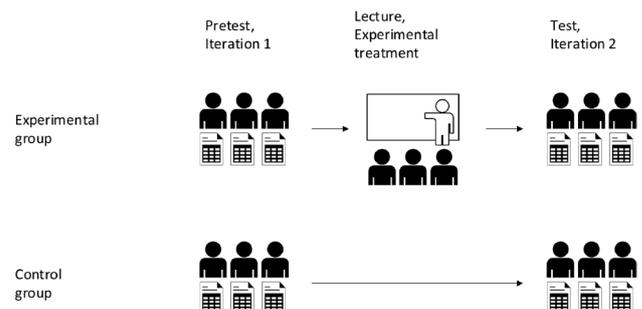


Fig. 1 Schematic presentation of the research design

specific case, the students were presented with a task to draft a calculation for a bid for building a wall of stone or brick. They were given only initial parameters, that are conditions regarding the number of workers, number of man hours required for the construction of the wall, the size of the wall, the workers' wages, cost price of brick and stone, and regarding the amount of the desired profit. Appendix 2 presents a possible solution to the problem.

Having created the spreadsheet, the students filled in the questionnaire shown in Appendix 3, based on Panko (1996) up on completion of the task. After the first iteration, only the experimental group of students were subjected to experimental treatment. The experimental treatment comprised a three-part lecture. In the first part, the students were informed about numerous examples of spreadsheet errors, the so-called "horror stories" (EuSpRIG; Raković, 2014), for instance, an error caused by reformatting of a spreadsheet built in MS Excel resulted in the largest bankruptcy lawsuit in the history of the USA (Havenstein, 2008). A minor administrative error (cut and paste) cost the Canadian company Trans Alta 24 million dollars (French, 2003), etc. The students were advised that, although such stories appear, they are considered to be a mere tip of an iceberg, as many errors are never publicized do to fear of reputation damage, and many errors are not even detected. The second part of the lecture addressed the existing taxonomies of spreadsheet errors (Raković, 2014; Sakal et al., 2015; 2016a). Four taxonomies were presented, which were elaborated in Sakal et al., 2016a; 2016b. For example, Powel et al. (2008; 2009a; 2009b) divide errors into: logic errors, reference errors, hard-coding errors, copy / paste errors, data input errors, and omission errors. The third part of the lecture elaborating on good practice guidelines for creating spreadsheets (Raković et al., 2015; Raković, 2014; 2019). At the end of the experiment, within briefing about research results, students of both control and experimental group, as potential influencers of business life and corporate social responsibility (Binzberger, 2006; Tokarčíková et al., 2015), were also informed about the significance of responsible creation and use of spreadsheets.

A total of 62 students participated in the study. The pretest (the first iteration of the quasi-experiment) was taken by 37 and test iteration 2 (the second iteration of the quasi experiment, conducted after experimental treatment) by 35 students from the experimental group. Within the control group, pretest was taken by 25, and posttest by 24 participants. The responses of participants (students) who did not participate in the posttest as well

were ignored. As the size of the sample was not sufficient, Bootstrap method was applied (bootstrap results are mostly based on 1000 bootstrap samples).

4 Research results

The quasi-experiment was conducted in two iterations. Each iteration consisted of two parts: developing the spreadsheet and filling in the questionnaire. The following section first presents the research results based on the analysis of created spreadsheets, and then from the questionnaire.

4.1 Spreadsheets evaluation results

Based on the literature overview, eight categories were defined according to which spreadsheets created by students were evaluated:

1. accurate output results;
2. errors in model setup;
3. hardcoding errors;
4. input errors;
5. the use of validation;
6. separate inputs and outputs;
7. differently formatted inputs and outputs;
8. use of complex formulas.

Categories number 1., 2. and 4. point to the existence of errors, whereas other categories do not imply quantitative errors. Categories 3., 5., 6. and 7. refer to what are called qualitative errors (Leon et al., 2015; Powell et al., 2008; Przasnyski et al., 2011), which can negatively influence the use of spreadsheets and may result in quantitative errors. Results show that more than two-thirds of spreadsheets within the pretest contained an error (68.57 % experimental group, 66.67 % control group). The results changed within the posttest, so that more than four-fifths of students from the experimental group created an accurate spreadsheet model, compared to somewhat more than one-third of students from the control group. Table 1 shows the distribution of errors belonging to the eight categories.

Appendix 4 gives a table that presents the results of bootstrap analysis (column sig (2-tailed)). There is no statistically significant difference between the control and experimental group in the pretest. The table clearly shows that the experimental treatment has the effect on the dependent variable ($p < 0.05$), which specifically means:

- Students who were subjected to experimental treatment create more accurate spreadsheet tables ($p = 0.002$).
- Students who were subjected to experimental treatment make a smaller number of hardcoding errors ($p = 0.003$).

Table 1 Results of the analysis of the created spreadsheets

Spreadsheet assessment category	Pretest				Posttest			
	Experimental group		Control group		Experimental group		Control group	
	YES %	NO %	YES %	NO %	YES %	NO %	YES %	NO %
E1 – Accurate output results	31.43	68.57	33.33	66.67	82.86	17.14	37.50	62.50
E2 – Errors in model setup	51.43	48.57	41.67	58.33	14.29	85.71	50.00	50.00
E3 – Hardcoding errors	85.71	14.29	87.50	12.50	11.43	88.57	79.17	20.83
E4 – Input errors	57.14	42.86	50.00	50.00	8.57	91.43	45.83	54.17
E5 – Use of validation	0	100	0	100	28.57	71.43	0	100
E6 – Separated inputs and outputs	5.71	94.29	4.17	95.83	57.14	42.86	4.17	95.83
E7 – Differently formatted inputs and outputs	0	100	0	100	28.57	71.43	0	100
E8 – Use of complex formulas	51.43	48.57	66.67	33.33	8.57	91.43	70.83	29.17

- Students who were subjected to experimental treatment make fewer input errors ($p = 0.007$).
- Students who were subjected to experimental treatment use data validation to a greater extent ($p = 0.006$).
- Students who were subjected to experimental treatment separate inputs and outputs from the spreadsheet to a greater extent ($p = 0.001$).
- Students who were subjected to experimental treatment format model inputs and outputs differently ($p = 0.008$).
- Students who were subjected to experimental treatment use complex formulas to a smaller extent ($p = 0.001$).

4.2 Questionnaire results

Within the pretest, the average task completion time in the experimental group was 18.4, and in the control group 18.8 minutes. Both the experimental and the control group assessed the task of spreadsheet development as medium difficult. Individual average values and standard deviations are presented in Table 2.

Application of T-test for individual questions (questions 1, 2, 3, 4 and 6) in the questionnaire did not result in establishing statistically significant differences between responses in the pretest, which implies that the participants' knowledge of stories about errors, spreadsheet error taxonomies and good

Table 2 Mean value and standard deviation of assessments for individual questions

Question	Pretest				Posttest			
	Experimental group		Control group		Experimental group		Control group	
	Average	StDev	Average	StDev	Average	StDev	Average	StDev
Q1 – Time required for task completion (min)	18.40	9.07	18.83	6.83	21.22	5.80	19.33	6.37
Q2 – Task difficulty (assessments from 1 to 7; 1 – easy, 7 – difficult)	3.47	1.33	3.25	0.98	2.85	1.06	2.70	0.80
Q3 – My previous knowledge was (assessments from 1 to 7; 1 – sufficient, 7 – insufficient)	2.78	1.80	2.58	1.44	2.89	1.69	2.95	1.19
Q4 – Assess your experience in work with Excel (assessments from 1 to 7; 1 – no experience, 7 – expert experience)	3.49	1.09	3.50	1.06	3.48	1.01	3.45	1.14
Q6 – Assess the accuracy of your application from 0 to 100 % (0 % -completely inaccurate, 100 % -completely accurate)	78.85 %	16.28 %	79.28 %	16.32 %	74.77 %	20.17 %	87.50 %	16.49 %
Q7 – Do you believe that errors in Excel workbooks can affect decision making in an organisation? (assessments from 1 to 7; 1 – cannot affect decision making, 7 – errors may affect decision making extremely negatively)	4.97	1.12	4.66	1.04	6.02	0.89	4.54	1.38

practices guidelines for creating spreadsheet models does not affect the responses and attitudes of users surveyed by means of the questionnaire. For this reason, Mann-Whitney test was also applied after the T-test. The results of the conducted Man-Whitney test point that, if pretest and posttest are considered, only question 7 contains statistically significant differences ($p < 0.05$) relating to the participant’s opinion on the influence of spreadsheet errors on decision making in organizations (Table 3).

When observing question five (Questionnaire, Appendix 2), regarding the participant's assessment whether their created spreadsheet contains errors, we find that most students, from both groups, believe that their spreadsheets contain errors. Based on hypothesis 2, which reads if spreadsheet developers are informed about spreadsheet errors, they will be less overconfident about the accuracy of their spreadsheets, two statistical hypotheses are set here:

- H20: Selfconfpre = Selfconfpost
- H21: Selfconfpre > Selfconfpost.

The result of the application of McNemar test for the significance of changes on the experimental group ($p = 0.008$) shows that self-confidence about the accuracy of one's own spreadsheet models is higher before learning about

spreadsheet error than after learning about spreadsheet errors (Tables 4 and 5). Therefore, statistical hypothesis H20 is rejected, and research hypothesis H2 is confirmed.

5 Discussion

In comparison with related researches (Benham and Giullian, 2005; Panko, 2007; 2009), students who participated in the quasi-experiment showed approximately the same knowledge (experience) and self-confidence about their own spreadsheets. Research results show that before the experimental treatment, more than two-thirds of participants (68.57 %) created spreadsheets that contained errors. The errors were mostly in model setup, use of constants in the formulas, or in input. This is a very high level of incorrect spreadsheets if compared with related research (presented in Section 2): Panko (1996) – 38 % incorrect spreadsheets; Irons (2003) 18.18 % incorrect spreadsheets; Panko and Sprague (1999) – 35 % incorrect spreadsheets, and Teo and Tan (1999) with 41.7 % and 50 % incorrect spreadsheets. After the experimental treatment, more than four-fifths (82.86 %) of participants in the experimental group created accurate spreadsheets. The accuracy of spreadsheets in the control group did not differ much in the pretest and posttest: 33.33 % and 37.5 % accurate spreadsheets

Table 3 Test Statistics^a

	P7 – pretest	P7 – posttest
Mann-Whitney U	383.500	165.500
Wilcoxon W	708.500	465.500
Z	-1.178	-4.061
Asymp. Sig. (2-tailed)	0.239	0.000
Monte Carlo Sig. (2-tailed) Sig.	0.243 ^b	0.000 ^b
99 % Confidence Interval		
	Lower Bound	0.232
	Upper Bound	0.254
Monte Carlo Sig. (1-tailed) Sig.	0.121 ^b	0.000 ^b
99 % Confidence Interval		
	Lower Bound	0.113
	Upper Bound	0.130

^a Grouping Variable: Group

^b Based on 10000 sampled tables with starting seed 1573343031

Table 4 Q7 pre * Q7 post

		Q7 post		Total
		YES	NO	
Q7 pre	YES	25	0	25
	NO	8	2	10
Total		33	2	35

Table 5 McNemar Test

	Value	Exact Sig. (2-sided)
McNemar Test		0.008(a)
N of Valid Cases	35	

^a Binomial distribution used.

respectively. Spreadsheet developers subjected to experimental treatment create spreadsheets containing a lower number of errors in the setup. The use of constants in formulas in the beginning does not result in quantitative errors; however, during the use and modifications, they can often turn into a quantitative error. When the spreadsheet developers are informed about this type of errors, the use of constants in the formulas declines drastically (from 87.51 % to 11.43 %). Awareness of the errors also affects positively the quality of input, i.e. reduction in input errors. Data validation, setting certain limitations in input etc. can prevent incorrect inputs to a certain extent, and therefore, the use of data validation may affect the quality of spreadsheets positively. After the experimental treatment, the participants started using data validation. Separating cells containing inputs from wholes containing outputs (formulas or functions), as well as their different formatting may also affect spreadsheet usability positively. After the experimental treatment, the users started separating inputs and outputs more frequently and formatting them differently.

Getting to understand spreadsheet and its modification may be difficult if the spreadsheet contains complex formulas. Experimental treatment significantly affected the reduction in the number of spreadsheets containing complex formulas. Therefore, if the spreadsheet developers are aware of the specific negative consequences of spreadsheet errors, types of spreadsheet errors and good practices in creating and using spreadsheets, they will create more accurate, i.e. higher-quality spreadsheets.

Creating a spreadsheet for the Wall problem presented a medium difficult problem for participants, and their previous knowledge was sufficient for creating a bid in a spreadsheet. The participants assessed their own experience in working with Excel as very good. More than three-quarters of participants in the pretest believed that their spreadsheets did not contain errors, compared to somewhat fewer than three-quarters of participants in the posttest. The application of McNemar test for the significance of changes showed that users are more self-confident about the accuracy of their spreadsheet models before experimental treatment, which included hearing stories about specific spreadsheet errors, existing spreadsheet error taxonomies and good practices guidelines for creating spreadsheets. The experimental treatment affected spreadsheet developers' self-confidence (or overconfidence as Panko would term it). Therefore, as proved by Benham and Gullian (2005); Panko (2007; 2009); Raković (2014) and Purser and Chadwick (2006), the study confirmed

that spreadsheet developers aware of stories of spreadsheet errors, their taxonomies and good practices guidelines have a lower self-confidence regarding the accuracy of their spreadsheet models. In addition to self-confidence, the experimental treatment affected the opinion that the participants had about the claim that spreadsheet errors can be the cause of wrong business decisions. After the experimental treatment, spreadsheet developers believe to a greater extent that errors can affect decision making in organizations extremely negatively.

Based on the above discussion on the research results, it can be concluded that both research hypotheses are accepted, that is, spreadsheet developers aware of spreadsheet errors and taxonomies and good practices of spreadsheet development will create more accurate and higher-quality spreadsheets, and be less self-confident about the accuracy of spreadsheets.

6 Conclusion

Spreadsheets are mostly created by end users who lack adequate IT knowledge, and very often do not know most spreadsheet possibilities either. In addition, spreadsheet developers are overconfident and unaware of risks entailed by the use of these applications. If it is (almost) impossible to eliminate the creation and use of spreadsheets (user-developed applications), then they need to be placed under control in a way. If spreadsheet developers are well-informed about spreadsheet errors (specific examples and taxonomies), and good practices of spreadsheet creation, they will create higher-quality spreadsheets. Therefore, spreadsheet based courses also need to include topics about spreadsheet errors and good practices. Organizations basing their key business processes on user developed applications in spreadsheet programs should organize courses dealing with these subjects and create a strategy of managing this part of Shadow IT. The very awareness of the risks of spreadsheet errors is bound to affect positively the quality of developed spreadsheets. It is also important to note that participation in the quasi-experiment enabled students to acquire valuable experiences that will facilitate handling spreadsheets in their professional careers.

6.1 Limitations of the study and proposals for further research

Research design belongs to the group of quasi-experimental designs, because rather than dividing students into groups by random choice, existing groups were taken, which implicates a more lenient control of external

variables. In addition to an insufficient number of participants, which was, in this research, compensated by the use of Bootstrapping, an improved design would require forming a control and experimental group by randomizing. Another deficiency that can be mentioned is participation of students instead of participants who are employed, and create and use spreadsheets in their job. So, the proposals for future research are to form two groups of users who use spreadsheets at their workplace and conduct the same

References

- Aurigemma, S., Panko, R. (2014) "Evaluating the Effectiveness of Static Analysis Programs Versus Manual Inspection in the Detection of Natural Spreadsheet Errors", *Journal of Organizational and End User Computing (JOEUC)*, 26(1), pp. 47–65.
<https://doi.org/10.4018/joeuc.2014010103>
- Baker, K., Powell, S., Lawson, B., Foster-Johnson, L. (2006) "Spreadsheet Engineering Research Project (SERP)", [online] Available at: <http://mba.tuck.dartmouth.edu/spreadsheet/index.html> [Accessed: 15 November 2017]
- Benham, H., Giullian, M. (2005) "Reducing Spreadsheet Error Rates", *Issues in Information Systems*, 6(1), pp. 28–34.
- Bellino, C. A., Ochab, D., Rowland, J. S. (2010) "Global Technology Audit Guide (GTAG®) 14 Auditing User-developed Applications", The Institute of Internal Auditors, Altamonte Springs, USA, [online] Available at: <https://chapters.theiia.org/montreal/ChapterDocuments/GTAG%2014%20-%20Auditing%20User-developed%20Applications.pdf> [Accessed: 15 January 2017]
- Binzberger, V. (2006) "Software development as social activity: distributed cognition or hermeneutic practice?", *Periodica Polytechnica Social and Management Sciences*, 14(1), pp. 19–28.
<https://doi.org/10.3311/pp.so.2006-1.03>
- Bewig, P. L. (2005) "How do you know your spreadsheet is right?", [online] Available at: <http://www.eusprig.org/hdykysir.pdf> [Accessed: 21 September 2013]
- Caulkins, J. P., Morrison, E. L., Weidemann, T. (2007) "Spreadsheet Errors and Decision Making: Evidence from Field Interviews", *Journal of Organizational and End User Computing (JOEUC)*, 19(3), pp. 1–23.
<https://doi.org/10.4018/joeuc.2007070101>
- Chua, C. E. H., Storey, V. C., Chen, L. (2014) "Central IT or Shadow IT Factors Shaping Users' Decision To Go Rogue With IT", In: 35th International Conference on Information Systems, Auckland, New Zealand, [online] Available at: <https://aisel.aisnet.org/icis2014/proceedings/ISGovernance/2/> [Accessed: 15 January 2017]
- Colver, D. (2004) "Spreadsheet good practice: is there any such thing?", In: European Spreadsheet Risk Interest Group Conference, London, England [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1001/1001.3967.pdf> [Accessed: 18 March 2017]
- Dunn, A. (2010) "Spreadsheets – the Good, the Bad and the Downright Ugly", In: European Spreadsheet Risk Interest Group Conference, London, England. [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1009/1009.5705.pdf> [Accessed: 17 February 2017]
- EuSpRIG "European Spreadsheet Risks Interest Group", [online] Available at: <http://www.eusprig.org/> [Accessed: 30 September 2017]
- experiment. In addition to the questionnaire, it would also be good to conduct a semi-standardized interview before and after control treatment, which would gather data that could be the subject of qualitative data processing.

Acknowledgement

We would like to thank Professors Živan Ristić and Paul Mireault for valuable comments and insightful advice.

- Ferreira, M. A., Visser, J. (2012) "Governance of Spreadsheets through Spreadsheet Change Reviews", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1211/1211.7100.pdf> [Accessed: 15 January 2017]
- French, C. (2003) "TransAlta Says Clerical Snafu Costs It \$ 24 Million", [online] Available at: http://www.globeinvestor.com/servlet/ArticleNews/story/ROC/20030603/2003-06-03T232028Z_01_N03354432_RTRIDST_0_BUSINESS-ENERGY-TRANSALTA-COL [Accessed: 15 January 2017]
- Govindarajulu, C. (2003) "End users: who are they?", *Communications of the ACM*, 46(9), pp. 152–159.
<https://doi.org/10.1145/903893.903931>
- Grossman, T. A., Özlük, Ö. (2004) "A Paradigm for Spreadsheet Engineering Methodologies", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/0802/0802.3919.pdf> [Accessed: 16 January 2017]
- Havenstein, H. (2008) "Excel error leaves Barclays with more Lehman assets than it bargained for", *Computerworld*, [online] Available at: http://www.computerworld.com/s/article/9117143/Excel_error_leaves_Barclays_with_more_Lehman_assets_than_it_bargained_for [Accessed: 20 February 2017]
- Hill, M. C., Barnes, W. A. (2011) "End-User Computing Applications: implications for internal auditors and managers", *The CPA Journal*, 81(7), pp. 67–71.
- Irons, R. J. (2003) "The Wall and The Ball: A Study of Domain Referent Spreadsheet Errors", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/0804/0804.0943.pdf> [Accessed: 17 January 2017]
- King, N., Azad, B. (2014) "Feral Systems as Institutional Phenomena: A Framework for Analyzing Persistent Computer Workarounds", In: Kerr, D. V., Burgess, K., Houghton, L. (eds.) *Feral Information Systems Development: Managerial Implications*, IGI Global, Hershey, USA, pp. 43–67.
<https://doi.org/10.4018/978-1-4666-5027-5.ch003>
- Ko, A. J., Abraham, R., Beckwith, L., Blackwell, A., Burnett, M., Erwig, M., Scaffidi, C., Lawrance, J., Lieberman, H., Myers, B., Rosson, M. B., Rothermel, G., Shaw, M., Wiedenbeck, S. (2011) "The state of the art in end-user software engineering", *ACM Computing Surveys*, 43(3), article number: 21, pp. 1–44.
<https://doi.org/10.1145/1922649.1922658>

- Kopper, A., Westner, M. (2016) "Towards a Taxonomy for Shadow IT", In: 22nd Americas Conference on Information Systems, San Diego, USA, pp. 690–699.
- Kruck, S. E., Maher, J. J., Barkhi, R. (2003) "Framework for Cognitive Skill Acquisition and Spreadsheet Training", *Journal of Organizational and End User Computing (JOEUC)*, 15(1), pp. 20–37.
<https://doi.org/10.4018/joeuc.2003010102>
- Kulesz, D., Zitzelsberger, S. (2012) "Investigating Effects of Common Spreadsheet Design Practices on Correctness and Maintainability", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1211/1211.7104.pdf> [Accessed: 17 January 2017]
- Lawson, B. R., Baker, K. R., Powell, S. G., Foster-Johnson, L. (2009) "A comparison of spreadsheet users with different levels of experience", *Omega*, 37(3), pp. 579–590.
<https://doi.org/10.1016/j.omega.2007.12.004>
- Leon, L., Przasnyski, Z. H., Seal, K. C. (2015) "Introducing a Taxonomy for Classifying Qualitative Spreadsheet Errors", *Journal of Organizational and End User Computing (JOEUC)*, 27(1), pp. 33–56.
<https://doi.org/10.4018/joeuc.2015010102>
- O'Beirne, P. (2010) "Spreadsheet Refactoring", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1009/1009.1412.pdf> [Accessed: 17 February 2017]
- Panko, R. R. (1996) "Hitting the wall: Errors in Developing and Debugging a "Simple" Spreadsheet Model", In: Proceedings of HICSS-29: 29th Hawaii International Conference on System Sciences, Wailea, USA, pp. 356–363.
<https://doi.org/10.1109/HICSS.1996.495419>
- Panko, R. R. (2007) "Two Experiments in Reducing Overconfidence in Spreadsheet Development", *Journal of Organizational and End User Computing (JOEUC)*, 19(1), pp. 1–23.
<https://doi.org/10.4018/joeuc.2007010101>
- Panko, R. R. (2008) "What We Know About Spreadsheet Errors", (In 2008, the article from 1998 was revised.) [online] Available at: <http://panko.shidler.hawaii.edu/ssr/Mypapers/whatknow.htm> [Accessed: 15 October 2015]
- Panko, R. R. (2009) "Two Experiments in Reducing Overconfidence in Spreadsheet Development", In: Clarke, S. (ed.) *Evolutionary Concepts in End User Productivity and Performance: Applications for Organizational*, IGI Global, Hershey, USA, pp. 131–149.
<https://doi.org/10.4018/978-1-60566-136-0.ch009>
- Panko, R. R. (2013) "End User Computing: The Dark Matter (and Dark Energy) of Corporate IT", *Journal of Organizational and End User Computing (JOEUC)*, 25(3), pp. 1–19.
<https://doi.org/10.4018/joeuc.2013070101>
- Panko, R. R., Aurigemma, S. (2010) "Revising the Panko–Halverson taxonomy of spreadsheet errors", *Decision Support Systems*, 49(2), pp. 235–244.
<https://doi.org/10.1016/j.dss.2010.02.009>
- Panko, R. R., Sprague Jr., R. H. (1999) "Hitting the Wall: Errors in Developing and Code-Inspecting a 'Simple' Spreadsheet Model", *Decision Support Systems*, 22(4), pp. 337–353.
[https://doi.org/10.1016/S0167-9236\(97\)00038-9](https://doi.org/10.1016/S0167-9236(97)00038-9)
- Pemberton, J. D., Robson, A. J. (2000) "Spreadsheets in business", *Industrial Management & Data Systems*, 100(8), pp. 379–388.
<https://doi.org/10.1108/02635570010353938>
- Powell, S. G., Baker, K. R., Lawson, B. (2008) "A critical review of the literature on spreadsheet errors", *Decision Support Systems*, 46(1), pp. 128–138.
<https://doi.org/10.1016/j.dss.2008.06.001>
- Powell, S. G., Baker, K. R., Lawson, B. (2009a) "Errors in Operational Spreadsheets", *Journal of Organizational and End User Computing (JOEUC)*, 21(3), pp. 24–36.
<https://doi.org/10.4018/joeuc.2009070102>
- Powell, S. G., Baker, K. R., Lawson, B. (2009b) "Impact of errors in operational spreadsheets", *Decision Support Systems*, 47(2), pp. 126–132.
<https://doi.org/10.1016/j.dss.2009.02.002>
- Przasnyski, Z., Leon, L., Seal, K. C. (2011) "In Search of a Taxonomy for Classifying Qualitative Spreadsheet Errors", In: European Spreadsheet Risk Interest Group Conference, London, England, [online] Available at: <https://arxiv.org/ftp/arxiv/papers/1111/1111.6909.pdf> [Accessed: 15 January 2017]
- Purser, M., Chadwick, D. (2006) "Does an awareness of differing types of spreadsheet errors aid end-users in identifying spreadsheets errors?", In: European Spreadsheet Risk Interest Group Conference, London, England, pp. 185–204. [online] Available at: <https://arxiv.org/ftp/arxiv/papers/0803/0803.0167.pdf> [Accessed: 15 January 2017]
- Raffensperger, J. F. (2003) "New Guidelines for Spreadsheets", *International Journal of Business and Economics*, 2(2), pp. 141–154.
- Rakovic, L. (2014) "The function of spreadsheet programs in end user application development", PhD Dissertation University of Novi Sad, Subotica, Serbia.
- Raković, L. (2019) "A framework for managing spreadsheet based end user applications", *International Journal of Management and Decision Making*, 18(1), pp. 76–92.
<https://doi.org/10.1504/IJMDM.2019.10014614>
- Raković, L., Sakal, M., Tumbas, P., Matković, P., Pavličević, V. (2015) "Improvement of Spreadsheet Skills in Business-Oriented University Courses", In: 9th International Technology, Education and Development Conference, Madrid, Spain, pp. 6874–6882.
- Read, N., Batson, J. (1999) "Spreadsheet Modelling Best Practice", [online] Available at: <http://www.eusprig.org/smbp.pdf> [Accessed: 07 November 2013]
- Retrop, C., Zimmermann, S. (2012) "Shadow IT evaluation model", In: Federated Conference on Computer Science and Information Systems (FedCSIS), Wroclaw, Poland, pp. 1023–1027.
- Ristić, Ž. (2011) "Kvantitativna, kvalitativna i mešovita istraživanja – metodološki aspekti", (Quantitative, qualitative and mixed research - methodological aspects) University of Novi Sad, Novi Sad, Serbia. (in Serbian)
- Sakal, M., Raković, L., Tumbas, P., Matković, P., Pavličević, V. (2015) "Reducing Spreadsheet Users' Overconfidence Through Business Oriented University Courses", In: 7th International Conference on Education and New Learning Technologies, Barcelona, Spain, pp. 6054–6059.
<https://doi.org/10.13140/RG.2.1.5029.0407>

- Sakal, M., Raković, L., Tumbas, P. (2016a) "Taxonomies of Spreadsheet Errors as Content of Spreadsheet Related Courses", In: 8th International Conference on Education and New Learning Technologies, Barcelona, Spain, pp. 2882–2889.
<https://doi.org/10.21125/edulearn.2016.1628>
- Sakal, M., Raković, L., Vuković, V. (2016b) "End User Software Engineering", In: 8th International Conference on Education and New Learning Technologies, Barcelona, Spain, pp. 2897–2902.
<https://doi.org/10.21125/edulearn.2016.1633>
- Sakal, M., Raković, L., Matković, P. (2017a) "Taksonomija Shadow IT sistema", (Taxonomy of Shadow IT Systems) Anali Ekonomskog fakulteta u Subotici / The Annals of the Faculty of Economics in Subotica, 53(37), pp. 227–240. (in Serbian)
- Sakal, M., Raković, L., Vuković, V. (2017b) "A Survey of the Use and Purpose of Spreadsheets in SMEs in Serbia", *Ekonomika preduzeća* (Journal of Business Economics and Management), 65(3-4), pp. 294–305.
<https://doi.org/10.5937/EKOPRE1704294S>
- Teo, T. S. H., Tan, M. (1999) "Spreadsheet development and 'what-if' analysis: quantitative versus qualitative errors", *Accounting, Management and Information Technologies*, 9(3), pp. 141–160.
[https://doi.org/10.1016/S0959-8022\(99\)00006-5](https://doi.org/10.1016/S0959-8022(99)00006-5)
- Tokarčíková, E., Kucharčíková, A., Ďurišová, M. (2015) "Education of Students of the Study Program Informatics in the Field of Corporate Social Responsibility", *Periodica Polytechnica Social and Management Sciences*, 23(2), pp. 106–112.
<https://doi.org/10.3311/PPso.7473>
- Zimmermann, S., Rentrop, C., Felden, C. (2016) "Governing IT Activities in Business Workgroups—Design Principles for a Method to Control Identified Shadow IT", In: Abramowicz, W., Alt, R., Franczyk, B. (eds.) *Business Information Systems, Lecture Notes in Business Information Processing*, Springer International Publishing, Cham, Switzerland, pp. 252–264.
https://doi.org/10.1007/978-3-319-39426-8_20

Appendix 1 – Wall Problem

A calculation of bid for construction of a wall should be made in Excel, bearing in mind the following:

- Two options need to be offered: stone and brick.
- Workers will work three eight-hour days to build the wall. The number of workers is 2.
- The wall will be 6 metres long, 1.8 metres high and 0.6 metres thick.
- The workers' hourly wage is 10 \$. Another 20 % need to be added to the total of paid earnings.
- Stones cost 3 \$ per cubic metre, and bricks 2 \$ per cubic metre.
- The profit should be 30 % of the cost price.

Save the document under the name SurnameFirstname.xlsx.

Appendix 2 – A possible solution for a bid for construction of a wall (according to Panko, 1996)

	A	B	C	D	E
1	Bid for construction of a wall				
2					
3	Labour - hours				
4	Days per person	3			
5	Hours per day	8			
6	Hours per person	24	=B4*B5		
7	Number of person	2			
8	Number of person	48	=B6*B7		
9					
10	Labor cost				
11	Pay per hour	\$10,00			
12	Pay	\$480,00	=B11*B8		
13	Wage bonus – percentage	20%			
14	Dodatak na platu	\$96,00	=B12*B13		
15	Labor cost	\$576,00	=B12+B14		
16					
17	Wall				
18	Height	1,8			
19	Lenght	6			
20	Thickness	0,6			
21	Cubic metre	6,48	=B18*B19*B20		
22					
23	Material cost				
		Stones	Brick		
24	Cost per cubic metre	\$3,00	\$2,00		
25	Total material costs	\$19,44	\$12,96	=B24*B21	=C24*B21
26					
27	Total costs				
		Stone	Brick		
28	Workers plus material	\$595,44	\$588,96	=B25+B15	=C25+B15
29	Profit %	30%			
30	Profit amount	\$178,63	\$176,69	=B29*B28	=C28*B29
31	Bid	\$774,07	\$765,65	=B30+B28	=C30+C28

Appendix 3 – Questionnaire

Surname and first name:

Gender:

Year of birth:

1. Time required for completing the task (min) _____
2. Task difficulty (assessments from 1 to 7; 1 – easy, 7 – difficult) _____
3. My knowledge was (assessments from 1 to 7; 1 – sufficient, 7 – insufficient) _____
4. Assess experience in work with Excel (assessments from 1 to 7; 1 – no experience, 7 – expert level experience) _____
5. Do you believe that your application contains errors? YES NO
6. Assess the accuracy of your application from 0 % to 100 % (0 % - completely inaccurate, 100 % – completely accurate) _____
7. State the degree of agreement with the following statement: Spreadsheet errors may be a cause of wrong business decisions (assessments from 1 to 7; 1 – I do not agree, 7 – I fully agree) _____

Appendix 4

Table Bootstrap for Independent Samples Test

			Mean Difference	Bootstrap ^a			
				Bias	Std. Error	Sig. (2-tailed)	95 % Lower
E1 – Accurate output results	Pretest	Equal variances assumed	-019	0.001	0.126	0.874	-0.282
		Equal variances not assumed	-0.019	0.001	0.126	0.875	-0.282
	Posttest	Equal variances assumed	0.454	0.000	0.120	0.001	0.218
		Equal variances not assumed	0.454	0.000	0.120	0.002	0.218
E2 – Errors in model setup	Pretest	Equal variances assumed	0.098	0.007	0.133	0.455	-0.147
		Equal variances not assumed	0.098	0.007	0.133	0.456	-0.147
	Posttest	Equal variances assumed	-0.357	0.002	.117	0.003	-0.581
		Equal variances not assumed	-0.357	0.002	.117	0.004	-0.581
E3 – Hardcoding errors	Pretest	Equal variances assumed	-0.018	-0.001	0.088	0.848	-0.184
		Equal variances not assumed	-0.018	-0.001	0.088	0.847	-0.184
	Posttest	Equal variances assumed	-0.677	0.004b	0.102b	0.003b	-0.860b
		Equal variances not assumed	-0.677	0.004b	0.102b	0.003b	-0.860b
E4 – Input errors	Pretest	Equal variances assumed	0.071	0.002	0.129	0.584	-0.168
		Equal variances not assumed	0.071	0.002	0.129	0.584	-0.168
	Posttest	Equal variances assumed	-0.373	0.001	.112	0.005	-0.584
		Equal variances not assumed	-0.373	0.001	.112	0.007	-0.584
E5 – Use of validation	Pretest	Equal variances assumed					
		Equal variances not assumed					
	Posttest	Equal variances assumed	0.286	-0.003	0.072	0.009	0.143
		Equal variances not assumed	0.286	-0.003	0.072	0.006	0.143
E6 – Separated inputs and outputs	Pretest	Equal variances assumed	0.015	0.001°	0.057°	0.825°	-0.103°
		Equal variances not assumed	0.015	0.001b	0.057b	0.832b	-0.103b
	Posttest	Equal variances assumed	0.530	-0.001	0.097	0.001	0.337
		Equal variances not assumed	0.530	-0.001	0.097	0.001	0.337
E7 – Differently formatted inputs and outputs	Pretest	Equal variances assumed					
		Equal variances not assumed					
	Posttest	Equal variances assumed	0.286	-0.003	0.078	0.011	0.139
		Equal variances not assumed	0.286	-0.003	0.078	0.008	0.139
E8 – Use of complex formulas	Pretest	Equal variances assumed	-0.152	0.000	.131	0.248	-0.394
		Equal variances not assumed	-0.152	0.000	.131	0.249	-0.394
	Posttest	Equal variances assumed	-0.623	0.003	0.109	0.001	-0.820
		Equal variances not assumed	-0.623	0.003	0.109	0.001	-0.820

^a Unless otherwise noted, bootstrap results are based on 1000 bootstrap samples