

Association for Information Systems

AIS Electronic Library (AISeL)

Proceedings of the 2018 AIS SIGED
International Conference on Information
Systems Education and Research

SIGED: IAIM Conference

2018

AUTOMATED GRADING, ANALYSIS, AND FEEDBACK IN EXCEL ASSIGNMENTS

Raymond Frost

Department of Analytics & Information Systems Ohio University, frostr@ohio.edu

Lauren Kenyo

Department of Analytics & Information Systems Ohio University, kenyo@ohio.edu

Vic Matta

Ohio University, matta@ohio.edu

Follow this and additional works at: <https://aisel.aisnet.org/siged2018>

Recommended Citation

Frost, Raymond; Kenyo, Lauren; and Matta, Vic, "AUTOMATED GRADING, ANALYSIS, AND FEEDBACK IN EXCEL ASSIGNMENTS" (2018). *Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research*. 6.

<https://aisel.aisnet.org/siged2018/6>

This material is brought to you by the SIGED: IAIM Conference at AIS Electronic Library (AISeL). It has been accepted for inclusion in Proceedings of the 2018 AIS SIGED International Conference on Information Systems Education and Research by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

AUTOMATED GRADING, ANALYSIS, AND FEEDBACK IN EXCEL ASSIGNMENTS

Raymond Frost
Department of Analytics & Information Systems
Ohio University
frostr@ohio.edu

Lauren Kenyo
Department of Analytics & Information Systems
Ohio University
kenyo@ohio.edu

Vic Matta
Department of Analytics & Information Systems
Ohio University
matta@ohio.edu

Abstract:

A system was created to automate the grading of Microsoft Excel assignments while providing analytical formative feedback to students and faculty. Although automated grading in Excel is not new, these systems have encountered several hurdles in grading, such as handling open ended questions that can have multiple correct answers. This manuscript outlines some of our grading system's key features and summarizes its benefits.

Keywords: automatic grading, Excel assignments, formative feedback

I. BACKGROUND

One of the ironies of automated grading is that it is fairly easy to grade for a prescribed answer, but very difficult to grade for reasonable answers that would normally be accepted by a conscientious grader [Stern & Solomon, 2006]. In fact some systems get around this limitation by requiring the instructor to approve each alternate solution one by one, thereby cutting into the time savings that the automation would have saved [Matthews, Janicki, He, & Patterson, 2012].

Grading systems also need to provide feedback. Ideally, that feedback would be formative and summative. To be truly formative the feedback has to be easy to comprehend. In the era of pen and paper the instructor would circle the error and provide a description of what was wrong and how to fix it. Our system similarly places formative feedback at the location of the error in Excel, instead of generating a separate feedback report [Bertheussen, 2014]. What follows is a brief description of some of the system features.

II. GRADING SYSTEM FEATURES

Working with a Delphi panel, the authors have analyzed and implemented reasonable grading rules. The system operates in batch mode grading hundreds of assignments in a fraction of the time previously required. A subset of five of these rules are featured in this abstract. We employ a narrative style to try to capture the look and feel of the program.

Handling Multiple Correct Answers

Working with the panel, the authors had to define the most inclusive, yet accurate, definition of a correct answer. Grading for the correct response would simply require checking the answer. However, checking that the total has been arrived at in valid fashion is difficult. If one considers each formula cell in an Excel spreadsheet to be a miniature program then for an answer to be valid, it should consistently transform any set of given inputs into the correct outputs [Hill, 2004].

Consider the case of calculating a grand total (a value of \$644,404) as shown in figure 1. Imagine that the system is designed to check whether the grand total resulted from a sum of the row totals. However, if a student sums the column totals instead, they would be marked wrong. However, both intermediate totals share the same root precedents—the raw numeric data shown in white. So, our system was designed to recursively work back to the root precedents for each formula cell and validate that they are indeed the necessary and sufficient set of precedents necessary to solve the problem.

	Peyton	Alex	Andy	Monthly Total Sales
January	\$ 24,066	\$ 24,042	\$ 18,875	\$ 66,983
February	\$ 21,262	\$ 21,685	\$ 19,506	\$ 62,453
March	\$ 18,880	\$ 19,109	\$ 20,382	\$ 58,371
April	\$ 12,243	\$ 21,259	\$ 21,404	\$ 54,906
May	\$ 10,784	\$ 11,167	\$ 15,541	\$ 37,492
June	\$ 20,094	\$ 18,981	\$ 23,415	\$ 62,490
July	\$ 12,739	\$ 13,590	\$ 17,412	\$ 43,741
August	\$ 11,897	\$ 12,087	\$ 16,960	\$ 40,944
September	\$ 19,631	\$ 16,123	\$ 23,189	\$ 58,943
October	\$ 15,234	\$ 19,270	\$ 14,909	\$ 49,413
November	\$ 22,176	\$ 19,725	\$ 16,947	\$ 58,848
December	\$ 20,185	\$ 15,869	\$ 13,766	\$ 49,820
Annual Total Sales	\$ 209,191	\$ 212,907	\$ 222,306	\$ 644,404

Figure 1: Two equally valid paths to the grand total—with same root precedents

If this simple example can support multiple paths to a solution, imagine how many more paths there would be in a complex problem. The system needs to be able to validate any and all of them. The system can also be customized to accept specific variations of answers. Consider logically equivalent strings that are not exact matches. It is easy to program an exact match but much more difficult to set up logical equivalence. For example, “<.001”, “< .001”, and “p < .001” should all be valid answers to a question testing p values. The system was configured to accept these equivalencies. These are shown in the accepted values list of figure 2. However, “less than .001” is not acceptable, and therefore, marked wrong. Furthermore, because the accepted list is just a column in a configuration spreadsheet, it is easily updateable to account for future equivalencies. Note that figure 2 represents a rare case of accepting a hard-coded answer. In this case the student interpreted the answer from regression output of the data analysis toolpak (not shown).

Accepted Values	Submitted Values	Final Model Summary Table					
< .001			Television Advertising	Radio Advertising	Newspaper Advertising	Magazine Advertising	
<.001		Coefficient	11475.4300	3.57220281	4.21048372	0.8736152	0.921732
p<.001		P-value	---	P <.001	<.001	less than .001	p< .001
p < .001							
p <.001							
p< .001							

Graded Values	Final Model Summary Table					
		Television Advertising	Radio Advertising	Newspaper Advertising	Magazine Advertising	
	Coefficient	11475.4300	3.57220281	4.21048372	0.8736152	0.921732
	P-value	---	P < .001	P < .001	answer -4 less than .001 P < .001	P < .001

Figure 2. Logically equivalent reporting of p values.

Encouraging Best Practice

Most Excel assignments require that the same formula be repeated across multiple rows. Best practice dictates that the formula be constructed in such a way so that the range will automatically adjust as the formula is copied [Buzzetto-More & Alade, 2006] . Therefore, our system makes the same requirement of students. Formulas lose points if they lack a proper mix of relative and absolute references that would enable copying.

	A	B	C	D	E	F	G	H	I	J
35										
36		Total Slices								
37			Number of Pizzas							
38	Size	Slices/Pizza	1	2	3	4	5	6	7	8
39	Personal	4	4	8	12	16	20	24	28	32
40	Small	6	6	12	18	24	30	36	42	48
41	Medium	8	8	16	24	32	40	48	56	64
42	Large	10	10	20	30	40	50	60	70	80
43	X-Large	12	12	24	36	48	60	72	84	96
44	Party	24	24	48	72	96	120	144	168	192

	A	B	C	D	E	F	G	H	I	J
35										
36		Total Slices								
37			Number of Pizzas							
38	Size	Slices/Pizza	1	2	3	4	5	6	7	8
39	Personal	4	=C\$38*\$B39	=D\$38*\$B39	=E\$38*\$B39	=F\$38*\$B39	=G\$38*\$B39	=H\$38*\$B39	=I\$38*\$B39	=J\$38*\$B39
40	Small	6	=C\$38*\$B40	=D\$38*\$B40	=E\$38*\$B40	=F\$38*\$B40	=G\$38*\$B40	=H\$38*\$B40	=I\$38*\$B40	=J\$38*\$B40
41	Medium	8	=C\$38*\$B41	=D\$38*\$B41	=E\$38*\$B41	=F\$38*\$B41	=G\$38*\$B41	=H\$38*\$B41	=I\$38*\$B41	=J\$38*\$B41
42	Large	10	=C\$38*\$B42	=D\$38*\$B42	=E\$38*\$B42	=F\$38*\$B42	=G\$38*\$B42	=H\$38*\$B42	=I\$38*\$B42	=J\$38*\$B42
43	X-Large	12	=C\$38*\$B43	=D\$38*\$B43	=E\$38*\$B43	=F\$38*\$B43	=G\$38*\$B43	=H\$38*\$B43	=I\$38*\$B43	=J\$38*\$B43
44	Party	24	=C\$38*\$B44	=D\$38*\$B44	=E\$38*\$B44	=F\$38*\$B44	=G\$38*\$B44	=H\$38*\$B44	=I\$38*\$B44	=J\$38*\$B44

Figure 3: Checking for best practice use of absolute and relative references

Appropriate Formatting Requirements

There is also the issue of best practice formatting. Currency and accounting are both reasonable formatting equivalents and should be graded equivalently as shown in figure 3.

Television Advertising	Radio Advertising
\$ 90,000	\$ 90,000.00
\$ 100,000	\$ 95,000.00
\$ 50,000	\$ 50,000.00
\$ 71,000	\$ 75,000.00
\$ 100,000	\$ 95,000.00
\$ 58,000	\$ 65,000.00

Figure 4: Accounting vs currency formatting

As another example, say that the requirements are to fill the column headers with a certain color. Would it be wrong if students used a slightly different color? The decision was made to check for a fill rather than for a fill of a certain color. In contrast, prior attempts at automated grading of formats [Hill, 2003] do not accommodate the kind of tolerance described here.

Formative Feedback

Working with the panel, the authors identified location and content of feedback as two issues that needed to be addressed. For feedback to be usable, the panel felt that it had to be located exactly where the error took place. This meant that the feedback had to be placed inside each cell. Secondly, the panel felt that the feedback should provide all of the following:

- i. Each type of error and its point value
- ii. The student’s answer and the formula they typed
- iii. The correct answer and the correct formula

Students can earn partial points for their work. Each cell is graded to check for the correct answer, appropriate precedents, correct absolute vs relative references, and proper formatting. The point values for each of the 4 categories is customizable, but the panel felt that hard coded values and formulas should be marked wrong, since the purpose of Excel is to reference cells.

Figure 5 shows graded answers from parts of a table. Cells without errors, as in the Vehicle Type column, are left untouched. The others are marked partially incorrect. In the Salesperson ID column the student used a formula to find the salesperson ID from another table (not shown). The student’s answer and formula are shown in red. The answer is correct “74536,” but the formula has hard-coded values and does not correctly lock the appropriate cells for copying down the column. Therefore, partial points were deducted for missing precedents and missing absolute references. The student earns credit for getting the correct answer and for correct number formatting, and thus, shows no point deduction in those categories. The correct formula is shown below the student’s submitted formula in green, so the two can be easily compared by the student. Note that the system would have accepted any valid formula as long as the student arrives at the correct answer using correct precedents.

Nested If	IF - Binary
Salesperson ID	Vehicle Type
missing precedents -4, missing absolute reference -2 74536 =IF(D31=E24,74536, IF(D31=E25,20172, IF(D31=E26,89152))) 74536 =IF(D31=E\$24,D\$24, IF(D31=E\$25,D\$25,D\$26))	1

Figure 5. Deductions and solutions are shown directly in each cell

Weighted Grading

It is common for professors to have some problems and skills weighted more heavily than others. One should be able to do the same in an automated system. When launching the program, a form allows the professor to assign point values to specific skills—such as correct precedents being worth 4 points. The professor can also assign variable weights per cell using a special multiplier embedded in a comment. That multiplier is implemented by the word “special” followed by the specific multiplier—in this case “special9.”

Answer Sheet	
< .001	Author: Special9
No	
Yes	
Yes	
81.94%	
0.75	

Figure 6: Special multiplier for high value cells

III. DISCUSSION

The system has a number of other features not documented here that are the subject of a future paper. These include allowing the student to change assumptions, analytics reporting at varying degrees of granularity, “no fly” zones for the grading program to avoid over penalizing students for the same mistake, extensive anti-cheating measures, and the ability to perform all of these functions in batch mode.

We believe that the two most important contributions of this system are the ability to validate open-ended responses and the detailed formative feedback. Future plans include developing completely unstructured assignments that could nonetheless be validated.

IV. REFERENCES

- Bertheussen, B. A. (2014). Power to business professors: Automatic grading of problem-solving tasks in a spreadsheet. *Journal of Accounting Education*, 32(1), 76-87. doi:10.1016/J.JACCEDU.2014.01.002
- Buzzetto-More, N. A., & Alade, A. J. (2006). Best Practices in e-Assessment. *Journal of Information Technology Education*, 5, 251-269. doi:<https://doi.org/10.28945/246>
- Hill, T. G. (2003). MEAGER: Microsoft Excel automated grader. *Journal of Computing Sciences in Colleges*, 18(6), 151-164.
- Hill, T. G. (2004). Excel grader and Access grader. *ACM SIGCSE Bulletin*, 36(2), 101-105.
- Matthews, K., Janicki, T., He, L., & Patterson, L. (2012). Implementation of an automated grading system with an adaptive learning component to affect student feedback and response time. *Journal of Information Systems Education*, 23(1), 71.
- Stern, L. A., & Solomon, A. J. A. w. (2006). Effective faculty feedback: The road less traveled. 11(1), 22-41.