ASEE'S VIRTUAL CONFERENCE
At Home with Engineering Education
JUNE 22 - 26, 2020 #ASEEVC
Paper ID #28940

Objective scoring partial credits by tracking failure cascade in mechanics problem solving

Dr. Andrew Dongjin Kim, Georgia State University

Andrew Dongjin Kim is an Assistant Professor of Computer Science and Engineering at Georgia State University.

Objective scoring partial credits by tracking failure cascade in mechanics problem solving

Abstract

This study proposes an alternative way of the ordinary open-ended problem that makes consistent and objective grading easier by tracking students' failure cascade. The proposed problem style consists of a variable-based sub-question set. By plugging student's answers from the previous steps in the equation for the current step, the grading process can detect the error propagation. The author collected multiple test results from Engineering Statics courses for the past two years and verified the grading algorithm's feasibility by running a computer program to repeat the failure tracking.

Study Background

Consistent and objective grading open-ended questions for classical mechanics problems is a challenge. Offering partial credits to failures arising from mistakes needs enormous time and effort. If students' handwriting is illegible or if the solving procedure contains skipped steps (e.g., by using a calculator), grading becomes tedious. Considering that human beings make mistakes under various circumstances; instructors commonly offer partial scores for minor errors. However, in general, no standard grading rubric for such failures exists. Even with highly prescriptive rubrics, still grading depends on each instructor as well as each question's level of difficulty and complexity [1] [2] [3] [4] [5].

While the author was using the classic open-ended questions, there were appeals and 'begging' for further partial scores (even after a detailed explanation on how the grading was made) as students considered there was a room for instructor's subjective factor. In fact, the author experienced a dramatic grade difference (standard deviations 29.8% and 53%) between two sections taught by different instructors. The author also experienced a huge grade drop (standard deviations from 44% to 19%) between two classes of similar average scores by simply itemizing the problem statement. For instance, from the statement "Determine the internal forces through member AB, BF, and FG" to "Determine the following: (1) the internal force through member AB, (2) the internal force through member BF, and (3) the internal force through member FG." With the former statement, there were distinct student groups: solved the entire procedure vs. left the answer sheet blank. On the other hand, with the latter statement, there was another group of students who put some amount of effort to solve each item, and thus obtained partial scores.

Several efforts to curtail subjective grading have put mostly by adopting statistical approaches, such as cross-checking and averaging [6] [7] . These approaches' pitfall is that we need multiple trained graders for one class (ten to twenty according to the cited works) which is still time-consuming and inefficient for small or mid-size single instructor classes without the help of a teaching assistant. The author does not have any grading assistants. This is the main reason that impelled the author to develop the proposed method.

Gross and Dinehart identified common student errors in solving fundamental mechanics problems [8]. They have categorized the errors into three parts: non-conceptual errors, minor execution errors, and major conceptual errors. Each problem was graded out of 10 points and deducted points according to the type of errors as categorized: 30% deduction for major conceptual errors; 10-20% for minor execution errors; and 10% for non-conceptual errors. From 28 different problems which cover topics of parallel forces, the centroid of composite areas, truss analysis, and distributive load analysis, they found that students lost points only 13% from major conceptual error while 49% from minor execution and 38% from non-conceptual errors indicating that students may have a better understanding of overall mechanics concepts than a complete ability to obtain the correct answers.

This study is focusing on the development of objective grading methodology for the engineering statics course which is a fundamental of mechanics courses and the test result analysis for the past two years of collection. As shown in the previous study [8], students lose a meaningful number of scores from non-conceptual and minor execution errors. Thus, the main goal of this grading tool is to detect the non-conceptual and execution errors, return the lost points, and provide more detailed feedback to students, consistently and objectively.

Methodology

Ordinary open-ended problems were redesigned in a form of a variable-based sequential question set by dividing a given problem with a specific ultimate quantity into sub-questions for variables that appear in the solving procedure. Given calculation tolerance, if a test-taker did not submit the answer or if the answer was out of tolerance for a sub-question, a full assigned point was deducted. Referring to the governing equations for each variable, test-takers' answers were reused in the calculations. The calculation results were compared with the correct reference answers and submitted answers. If the calculation result using submitted answers in the previous step is the same as the submitted answer for the current step but different from the reference answer, then it is considered that the error has been cascaded. If the calculation result is the same as the correct answer, then the error in the previous step was a one-time mistake only for that step and did not cascade. If the calculation result is different from any of the compared values, then it is considered as another individual failure. If it turned out that the incorrect answer has affected the consecutive sub-questions, then no further point deduction was made from the affected items.

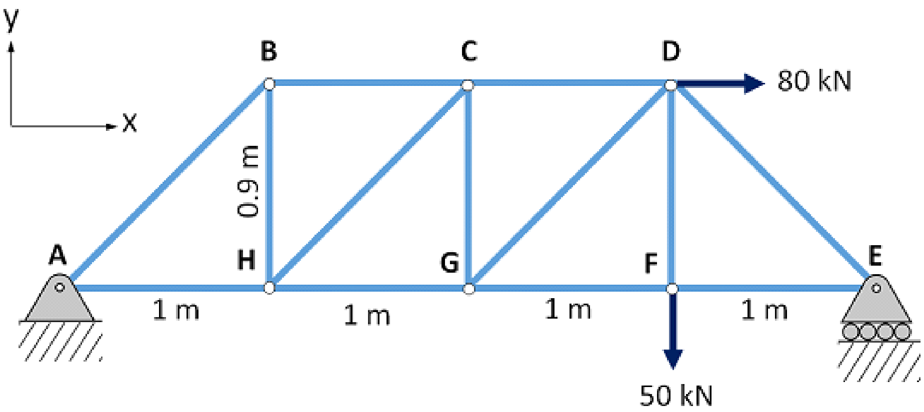
To better detect the calculation error, the concept of simplified uncertainty propagation estimation was firstly employed in determining tolerances [9]. However, estimating tolerances for each sequential question using tolerance propagation required the enormous task of complicated equation manipulations and was of little worth. Since the submitted answers were plugged in the equations for each item, the uncertainty does not accumulate but rather is reset within the standard rounding error range according to the precision (e.g., 123±0.5, 12.3±0.05). Thus, the tolerances were defined individually based on each answer's precision.

Sequential question set

In general, the open-ended mechanics problems are constructed with a narration of given conditions and values to be determined, then expect test-takers to find suitable solving methods and appropriate equations following the process of interpretation, planning, and execution (IPE)

[10] until they determine the values of the required ultimate parameter. The ordinary open-ended problems give test-takers freedom in this IPE step and thus develop their own IPE skills. However, considering that students in the low-level courses need to first build a sturdy foundation before personalizing their own methods, the author has bravely tried to reform the generic IPE steps and requested students to determine unknowns that are necessary to approach the ultimate parameter. The sequential questions are based on the equations that govern the assigned problem. Of course, there might be other approaches and equations that apply to the same problem [11]. Also, there can be a concern if the sequential question set is just exposing the solving procedure to test-takers. The author tried to compensate for this issue by generalizing the variable list. In other words, if there is a variable that is not necessary for a specific question but can be part of a governing equation, the sub-question set included the items for these variables. For instance, if there was a reaction force in one direction at a roller-support, the question asked for two or three direction force components. The author also tried to alter the sequence of the sub-questions slightly. After adopting the sequential question set, the author learned that only the students who comprehend the meaning of each variable and the solving procedure could play with the appropriate formulae and their answers were within the error-cascading range. In addition, students with test anxiety or are deficient in interpreting narrative statements could avoid turning in blank answer sheets and thus made a more precise evaluation of their ability possible.

Figure 1 shows an example of the sequential question set compared to the ordinary open-ended problem on the same topic. This problem is asking students to determine forces along truss members using either method of section or method of joints. The first step to solve the problem is identifying the support types. Once the support types were identified, the reaction forces are to be calculated by applying the equilibrium of forces and moments to the entire structure. Then, students will choose either method of section or method of joints to determine the internal forces along with the designated members. The sequential question set does not offer students the freedom of choosing methods between the two in order to limit the variation of applying equations. Instead, the problem designates the method first, then asks questions correspond to that method. The first element to determine is the reaction force at each support. Instead of asking the type of support at joint A and E, the question asks students to estimate each component of the reaction force. To determine components of reaction force in x and y direction, students first need to figure out the type of support. If the support types were improperly determined, then the reaction forces at the supports will be incorrectly calculated, and the error will be cascaded toward the next step unless it was an individual error made only for this item. To determine the internal forces, students may set up a linear algebra with a set of equilibrium equations applied to the left or right side of the imaginary section cut along with the members BC, HC, and HG. If solving a matrix for the linear algebra is not convenient, students may solve the equilibrium equations one by one in the order of given sequential questions (i.e., BC→HC→HG). Instead of requesting to indicate the direction of forces in words, the sequential questions ask students to indicate them with positive and negative values following the sign convention as noted. It is not vital to list up sub-question set in a strict solving sequence. Instead, it is recommended to make the questions variable base (i.e., numeric values of each unknown variable). The questions can be easily rearranged in sequential order when the evaluation is made.



(a)

A plane truss is loaded and supported as shown. Solve for the forces in the members BC, HC, and HG. Indicate whether each member is in tension or compression.

(b)

A plane truss is loaded and supported as shown. Solve for the following using Method Of Section:

①. x component of reaction force at A in kN
②. Moment due to 80 kN about E in kN-m
③. Moment due to 50 kN about E in kN-m
④. y component of reaction force at A in kN
⑤. x component of reaction force at E in kN
⑥. y component of reaction force at E in kN
⑦. Internal force along BC in kN
⑧. y component of internal force along HC in kN
⑨. x component of internal force along HC in kN
⑩. Internal force along HC in kN
⑪. Internal force along HG in kN

Sign convention: negative if the component of force is to negative x or negative y; negative if the internal force is in compression; negative if the moment is in clockwise.

Figure 1 – Sample truss structure problems: (a) ordinary open-ended style and (b) sequential sub-question set

Grading a sequential question set

The blank answers and the first out of tolerance answers were initially scored at zero. Then, the incorrect answer that appears after the first incorrect answer was primarily examined whether the previous error has cascaded toward this item. It was done by plugging the preceding incorrect answer in the equation and checking whether the result was equal to the answer submitted. If the result was equal to the submitted answer, then it was considered that the error has cascaded but the calculation was correct in the current step, and no point was deducted from the current question. If the result was different from the submitted answer either, then it was considered as the arbitrary answer or the result of misconception, and no point was given to the current question as well. Table 1 shows the four different sample gradings of the truss problem shown in Figure 1.

Suppose students were asked to submit answers in minimum of four significant figures if the answer has more than four figures. Then, the correct answers to the given truss problem can be obtained in order as:

$E_x = 0$

$A_x = -(80) - E_x = -80$ kN

$M_{E,80kN} = -(0.9)(80) = -72$ kN-m

$M_{E,50kN} = (1)(50) = 50$ kN-m

$A_y = (M_{E,80kN} + M_{E,50kN})/(4) = (-72+50)/4 = -5.5$ kN

$E_y = (50 - Ay) = (50 -(-5.5)) = 55.5$ kN

$F_{BC} = -A_y/(0.9) = -(-5.5)/(0.9) = 6.111$ kN

$F_{HC,y} = -A_y = -(-5.5) = 5.5$ kN

$F_{HC,x} = F_{HC,y}(1/0.9) = 6.111$ kN

$F_{HC} = F_{HC,y}(\sqrt{1 + 0.9^2}/0.9) = 5.5(1.345/0.9) = 8.219$ kN

$F_{HG} = -A_x - F_{BC} - F_{HCx} = -(-80) -(6.111) -(6.111) = 67.78$ kN

The submitted answer (1) indicates that the error due to the opposite sign (question ①) was a one-time error (might be due to missing sign when the answer was submitted) which did not affect the consecutive calculation (corresponds to the variable and the answer colored in blue). The submitted answer (2) contains two kinds of errors: incorrect calculation (question ②) and no-submission (question ⑤). For the answer not submitted, zero was simply assigned. However, the effect of the incorrect answer of item ② was tracked by plugging in the answer to the consecutive calculations (green colored items). This error was a one-time calculation error and the effect was cascaded toward the final question without any further miscalculation. Hence, no point was deducted from the subsequent items. The third submitted answer contains multiple errors propagated: the effect of a misconception of support types (questions ① and ⑤) and miscalculation (question ⑦) toward the final question ⑪. The misconception of item ⑤ resulted in an incorrect answer from item ①, and there exists no further error. Therefore, no point was additionally deducted from item ①. The incorrect answer found from question ⑪ was due to the errors occurred in both questions ⑤ and ⑦. The last case also contains multiple errors in items ⑧, ⑨, and ⑩ in a row. Therefore, the adjusted answers using the submitted answers were also updated multiple times. Here, the error occurred in item ⑧ did not cascade toward the last item ⑪ because the variable of item ⑧ is not part of the equation for item ⑪.

Table 1 – Grading sample (1) of the truss problem

Question	Correct Answer	Tolerance	Submitted Answer	Score (1)	Adjusted Answer	Score (2)	Final Score
① A_x	-80	-80.5 ~ -79.5	80	0		0	0
② $M_{E,80kN}$	-72	-72.5 ~ -71.5	-72	1	-72	1	1
③ $M_{E,50kN}$	50	49.5 ~ 50.5	50	1	50	1	1
④ A_y	-5.5	-5.55 ~ -5.45	-5.5	1	-5.5	1	1
⑤ E_x	0	0	0	1	0	1	1
⑥ E_y	55.5	55.45 ~ 55.55	55.5	1	55.5	1	1
⑦ F_{BC}	6.11	6.105 ~ 6.115	6.111	1	6.11	1	1
⑧ $F_{HC,y}$	5.5	5.45 ~ 5.55	5.5	1	5.5	1	1
⑨ $F_{HC,x}$	6.11	6.105 ~ 6.115	6.111	1	6.11	1	1
⑩ F_{HC}	8.22	8.215 ~ 8.225	8.219	1	8.22	1	1
⑪ F_{HG}	67.8	67.75 ~ 67.85	67.8	1	-92.2	0	1
			Total	10		9	**10**

Table 1 – Grading sample (2) of the truss problem

Question	Correct Answer	Tolerance	Submitted Answer	Score (1)	Adjusted Answer	Score (2)	Final Score
① A_x	-80	-80.5 ~ -79.5	-80	1	-80	1	1
② $M_{E,80kN}$	-72	-72.5 ~ -71.5	75	0		0	0
③ $M_{E,50kN}$	50	49.5 ~ 50.5	50	1	50	1	1
④ A_y	-5.5	-5.55 ~ -5.45	31.25	0	31.3	1	1
⑤ E_x	0	0	None	0	0	0	0
⑥ E_y	55.5	55.45 ~ 55.55	18.75	0	18.8	1	1
⑦ F_{BC}	6.11	6.105 ~ 6.115	-34.72	0	-34.7	1	1
⑧ $F_{HC,y}$	5.5	5.45 ~ 5.55	-31.25	0	-31.3	1	1
⑨ $F_{HC,x}$	6.11	6.105 ~ 6.115	-34.72	0	-34.7	1	1
⑩ F_{HC}	8.22	8.215 ~ 8.225	-46.71	0	-46.7	1	1
⑪ F_{HG}	67.8	67.75 ~ 67.85	149.4	0	149	1	1
			Total	2		9	**9**

Table 1 – Grading sample (3) of the truss problem

Question	Correct Answer	Tolerance	Submitted Answer	Score (1)	Adjusted Answer	Score (2)	Final Score
① A_x	-80	-80.5 ~ -79.5	-40.00	0	-40	1	1
② $M_{E,80kN}$	-72	-72.5 ~ -71.5	-72.00	1	-72	1	1
③ $M_{E,50kN}$	50	49.5 ~ 50.5	50.00	1	50	1	1
④ A_y	-5.5	-5.55 ~ -5.45	-5.500	1	-5.5	1	1
⑤ E_x	0	0	-40.00	0		0	0
⑥ E_y	55.5	55.45 ~ 55.55	55.50	1	55.5	1	1
⑦ F_{BC}	6.11	6.105 ~ 6.115	5.500	0	(6.11)	0	0
⑧ $F_{HC,y}$	5.5	5.45 ~ 5.55	5.500	1	5.5	1	1
⑨ $F_{HC,x}$	6.11	6.105 ~ 6.115	6.111	1	6.11	1	1
⑩ F_{HC}	8.22	8.215 ~ 8.225	8.222	1	8.22	1	1
⑪ F_{HG}	67.8	67.75 ~ 67.85	28.39	0	28.4	1	1
			Total	7		9	**9**

Table 1 – Grading sample (4) of the truss problem

Question	Correct Answer	Tolerance	Submitted Answer	Score (1)	Adjusted Answer	Score (2)	Final Score
① A_x	-80	-80.5 ~ -79.5	-80.00	1	-40	1	1
② $M_{E,80kN}$	-72	-72.5 ~ -71.5	-72.00	1	-72	1	1
③ $M_{E,50kN}$	50	49.5 ~ 50.5	50.00	1	50	1	1
④ A_y	-5.5	-5.55 ~ -5.45	-5.500	1	-5.5	1	1
⑤ E_x	0	0	0.000	1	0	1	1
⑥ E_y	55.5	55.45 ~ 55.55	55.50	1	55.5	1	1
⑦ F_{BC}	6.11	6.105 ~ 6.115	6.111	1	5.5	1	1
⑧ $F_{HC,y}$	5.5	5.45 ~ 5.55	10.10	0		0	0
⑨ $F_{HC,x}$	6.11	6.105 ~ 6.115	12.5	0	(11.2)	0	0
⑩ F_{HC}	8.22	8.215 ~ 8.225	9.21	0	(15.1)	0	0
⑪ F_{HG}	67.8	67.75 ~ 67.85	61.39	0	61.4	1	1
			Total	7		8	**8**

Automatic scoring tool

Digitalizing student's answers is an essential part of the error cascade detection procedure. Using the electronic files of the student's answers, an instructor can run a computing tool. For the large size classes, digitalizing students' answers might be a problem. We can consider using a separate answer list sheet beside the solution pages. Or, we may ask test-takers to turn in their answers electronically if computers are available. There can be several ways. These days, many schools are utilizing an online learning system such as D2L and blackboard. One can build a test page using the quiz module in such an online system. The author designed a webpage in the form of the paper-based-test so students can see all problems on one page and thus feel more comfortable. Next to each question item, there exists an answer box along with the required unit. Once the test taker submits the answers, the answer file saves the data in a certain form designed to the grading program. Figure 2 depicts the pseudocode of the automatic grading program that follows the scoring logic.

```
Read the correct answer and the student's submitted answer

% Evaluation of the student's submitted answers by comparing with the correct answers
For each question do
        Call:  score1 = evaluate (the correct answer, the student's answer)
End for

% Calculation using student's submitted answers
Plug in student's answers for the variables in the equations for each sub-question

% Evaluation of the student's submitted answers by comparing with the re-calculated values
For each question do
        Call:  score2 = evaluate (the re-calculated values, the student's answer)
End for

% Returning the lost scores affected by the calculation error cascade
For each question do
        new_score = MAX(score1, score2)
End for

% The user defined function that evaluates the student's answers
Function evaluate
        Pass In: the reference answer, the students' answer
        Estimate the tolerance range in relation to the reference answer
        IF the student's answer is within the tolerance range THEN
                score = 1.0
        ELSE
                score = 0.0
        ENDIF
        Pass Out: score
End function
```

Figure 2 – The pseudocode of the error tracking program

Result and Analysis

Table 2 lists the collected fifteen distinct Statics problems and their question structures. The terms ‘sequential question’ and ‘sub-question’ are used distinctly to mean that the number of sequential questions indicates the number of sub-question items that are dependent on other sub-questions’ answers so the error cascade may occur. For instance, in the sample truss problem, item ① depends on the result of item ⑤, and item ④ is affected by items ② and ③. The results from items ⑥, ⑦ and ⑧ lean on item ④. Items ⑨ and ⑩ rely on item ⑧. Item ⑪ counts on items ⑦ and ⑨. Therefore, the number of sequential questions is 8 and the portion of it out of eleven equation items is 72.7%. The x marks in Table A1 indicate the semesters in which those problems were assigned for evaluation.

Table 2 – Structure of assessment questions

Problem topics		No. of sub-questions	No. of sequential questions	Portion of sequential questions (%)
1	2D forces pulling a hook	8	2	25.0
2	2D forces supporting a cylinder weight	13	4	30.8
3	3D forces supporting a plate	34	25	73.5
4	3D force projection	14	8	57.1
5	Angle b/w two forces	18	9	50.0
6	3D force equilibrium	14	2	14.3
7	Simplification of 2D load on a beam	21	11	52.4
8	2D load on a beam w/ a two-force member	10	4	40.0
9	Simple truss w/ Method of Joints	23	21	91.3
10	Simple truss w/ Method of Section and Joints	17	11	64.7
11	Simple truss w/ Method of Joints	29	19	65.5
12	Internal load in a beam w/ a dry friction	13	11	84.6
13	Internal Load in a Beam w/ a couple-force	8	6	75.0
14	Dry friction on a belt supporting a beam	5	3	60.0
15	Dry friction on a spool pulled by a cylinder over a pulley	8	6	75.0

Table 3 shows the portion of the number of students who were in jeopardy of losing scores due to calculation error cascade and the average percentage scores related to the error cascade. The statistics reveal that an average of 42% (±26%) of students was at risk of losing points due to error cascade; an average of 16% points (±7% points) was detected as deducted points because of the failure cascade effect. These scores were returned to the overall scores along with the feedback message addressing that at which solving step the calculation failure has occurred, and up to which step the error has cascaded.

Table 3 – Assessment result summary

Problems	No. of students solved	No. of students lost scores due to error cascade	No. of students lost scores due to error cascade (%)	Average score loss due to error cascade / max. score (%)
1	58	3	5.2	16.7
2	83	22	26.5	12.1
3	83	30	36.1	14.2
4	58	10	17.2	11.1
5	83	8	9.6	9.9
6	58	34	58.6	11.1
7	91	65	71.4	12.6
8	91	67	73.6	16.0
9	19	12	63.2	17.0
10	106	42	39.6	10.1
11	67	45	67.2	12.2
12	106	32	30.2	15.0
13	67	12	17.9	20.5
14	67	22	32.8	20.9
15	17	15	88.2	38.9

The burden put on the instructors when they make test problems is weighing the points for each problem, and the weight is normally defined by the level of difficulty of the problem. However, measuring the difficulty level is apparently subjective and that can be different from the student's side. The trend lines in Figures 3 and 4 depict that the greater number of sequential questions, the higher chance to lose scores due to the calculation errors. This data is actually telling us that the longer the solution, the higher chances to lose scores due to calculation error. Therefore, the length of the solution can be a guideline in measuring the extent of the difficulty of a problem. With the proposed variable-based sub-question set, instructors can simply weigh points evenly throughout the sub-question, and the error cascade detection method will compensate the lost scores (10 to 20% points according to Figure 4) that are solely due to the calculation errors.

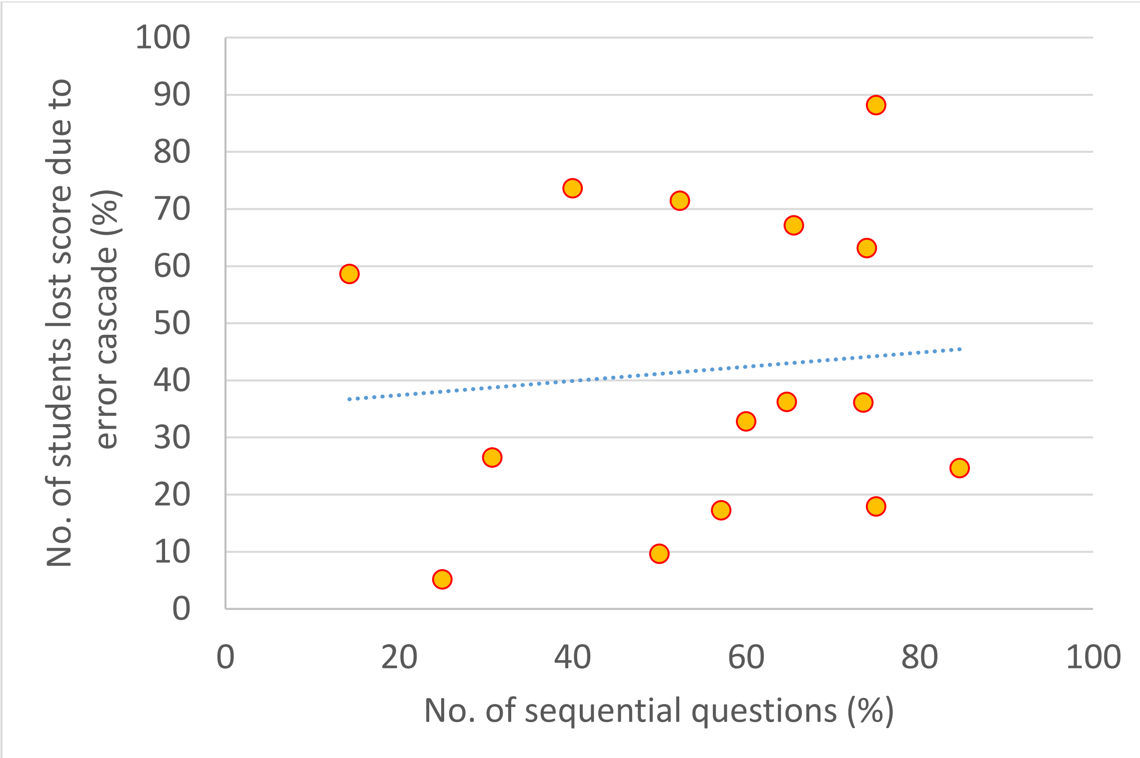


Figure 3 – Number of students who are in jeopardy of losing score due to calculation error effect cascade

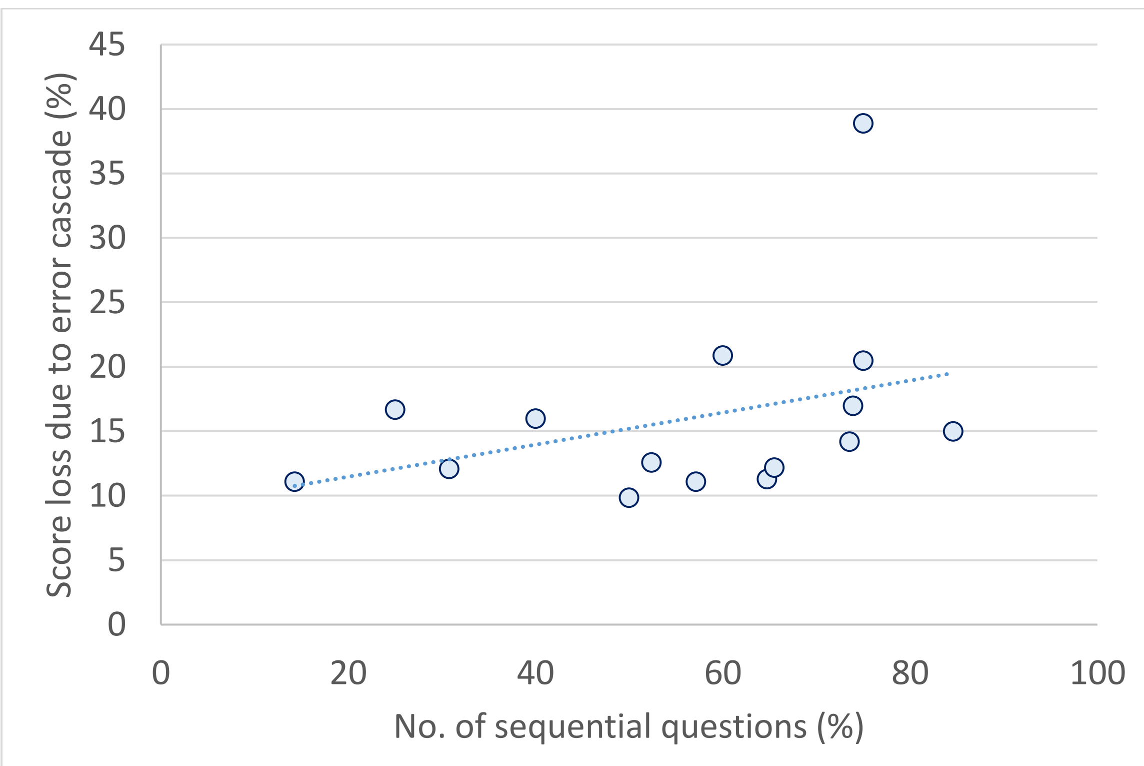


Figure 4 – Scores which were at risk of loss due to calculation error effect cascade

Table 5 shows the average scores per assessed semester for each problem. The differences of the average scores of the same problem between semesters ranges 1.6% ~ 6.6% (except the average score of the problem 12 in Spring 2019) confirming that the proposed grading method has consistency in scoring.

Table 5 – Score averages per semester for each problem

Problems	Average % scores per assessed semester			
	Summer 2018	Fall 2018	Spring 2019	Fall 2019
1	92.5	94.1		
2			57.5	60.4
3			73.7	75.4
4	73.9	75.9		
5			89.6	87.5
6	73.9	68.1		
7	70.6	72.7		70.7
8	64.4	67.6		68.2
9	64.3			
10		69.1	69.0	70.4
11			65.7	66.7
12		55.2	49.6	61.8
13			60.6	57.4
14			71.9	67.4
15	45.1			

The author has conducted a short survey to grasp students' side impression on the proposed test method. Table 6 is a result of a short survey conducted in the Spring 2020 semester with groups of students who are taking a Statics course and who already took it in the past. Comparing to the total number of data used in this study, the error margins of the survey are around 25% to 28% at a 95% confidence level. There were four questions on the favorable effect of the proposed variable-base testing and one on the adverse effect. Although it is hard to make a conclusion with this small amount of data, students who just began experiencing the new test structure show more affirmative opinions on the favorable effects and more contradicting opinion on the adverse effect. In both groups, most students think the proposed testing method provides the grading fairness, advocates comprehending physics and mathematical concepts, and mitigates the nervousness. The good thing is that more than half of the group 2 students who have taken the course in the past all the way through the final reassured that the proposed method does not clearly have an adverse effect on the improvement of problem-solving skills. In addition, the author asked students to turn in their paper solution sheets that are legible enough. The paper solution sheets are backup materials in case there are any technical issues during submitting their answer list. The author could observe from their paper solutions that students get more used to solve the problems beginning with a variable-base level until getting the computational results instead of listing equations meaninglessly.

Table 6 – Student satisfaction survey results

Question List	Group 1 (In the middle of the semester) Error Margin: 25.13% Confidence Level: 95%		Group 2 (After the semester) Error Margin: 28.14% Confidence Level: 95%	
Grading fairness	Strongly disagree	0%	Strongly disagree	0%
	Disagree	0%	Disagree	8.33%
	Neutral	0%	Neutral	0%
	Agree	20%	Agree	58.33%
	Strongly Agree	80%	Strongly Agree	33.33%
Motivation for practicing calculation skills	Strongly disagree	0%	Strongly disagree	0%
	Disagree	0%	Disagree	0%
	Neutral	0%	Neutral	8.33%
	Agree	26.67%	Agree	16.67%
	Strongly Agree	73.33%	Strongly Agree	75%
Motivation for building up concept and principle comprehension	Strongly disagree	0%	Strongly disagree	0%
	Disagree	6.67%	Disagree	8.33%
	Neutral	0%	Neutral	0%
	Agree	20%	Agree	33.33%
	Strongly Agree	73.33%	Strongly Agree	58.33%
Alleviation of test anxiety	Strongly disagree	0%	Strongly disagree	0%
	Disagree	0%	Disagree	0%
	Neutral	20%	Neutral	16.67%
	Agree	26.67%	Agree	25%
	Strongly Agree	53.33%	Strongly Agree	58.33%
Adverse impact on learning problem solving skill	Strongly disagree	0%	Strongly disagree	16.67%
	Disagree	13.33%	Disagree	8.33%
	Neutral	26.67%	Neutral	33.33%
	Agree	40%	Agree	25%
	Strongly Agree	20%	Strongly Agree	16.67%

Summary

The objective scoring method tailored to engineering statics problems was verified and examined using the variable-base sequential sub-question set. Students' comprehension of variables in terms of physical meaning and the way of estimation was tracked by plugging in the test-takers' submitted answers back for the variables in the corresponding governing equations. This calculation tracking method detected cascaded calculation errors. Once they are detected, the lost points were returned to the overall scores such that only the first execution error is reflected.

From the test results collected for the past two years, the statistics revealed that approximately 42% of students were at risk of losing scores due to calculation error and its propagation, and about 16% points were detected as the scores associated with the error cascade. The study also discovered that the number of sub-questions (or the length of solution) is proportional to the chances of error cascade. This also supports the importance of tracking calculation failure while grading.

Students’ answers were digitalized by requesting test-takers to type in their answers using computers. The error tracking and the scoring were conducted adopting a computer program such that the scoring is done quickly and objectively. Therefore, it shows that the algorithm applied in this study can be deployed to automatic grading along with providing standardized feedback messages.

The most important advantage of using the proposed grading method was that the grader and the test-takers both admitted the scoring was made objectively, and thus less complain about unfairness. Also, students recognize the importance of solving problems completely with the correct comprehension of physical and mathematical meanings of variables in the governing equations.

References

[1] M. S. a. C. C. B. Z. Dymond, "The Influence of Grading Bias on Reinforced Concrete Exam Scores at Three Different Universities," in *ASEE Annual Conference and Exposition*, Salt Lake City, UT, June 24-17, 2018.

[2] A. Karimi, "Bringing Uniformity in Topic Coverage and Grading Fairness in Multiple," in *International Mechanical Engineering Congress*, 2015.

[3] A. C. K. a. W. Sander, "Grades and Student Evaluations of Teachers," *Economics of Education Review*, pp. 59-63, 1999.

[4] C. E. Work, "Nationwide Study of the Variability of Test Scoring by Different Instructors," *Journal of Engineering Education,*, pp. 241-248, 1976.

[5] J. Malouff, "Bias in Grading," *College Teaching*, pp. 191-192, 2008.

[6] B. A. K. a. S. Miller, "Grade-a-thons and Divide-and-Conquer: Effective Assessment at Scale," in *ASEE Annual Conference & Exposition*, Columbus, OH, June 25-28, 2017.

[7] T. H. a. D. Wilkins, "A Model of a Relative Peer Grading System for MOOC's," in *The Annual ACM Southeast Conference*, Richmond, KY, March 29-31, 2018.

[8] S. P. Gross and D. W. Dinehart, "Identification of Common Student Errors in Solving Fundamental Mechanics Problems," in *ASEE Annual Conference & Exposition*, San Antonio, Texas, June 10-13, 2012.

[9] This page was last edited on 6 January 2020, at 15:34 (UTC). [Online]. Available: https://en.wikipedia.org/wiki/Propagation_of_uncertainty#cite_note-5.

[10] R. C. Hibbeler, "PowerPoints for Engineering Mechanics: Statics, 14th Edition," Pearson, 2016.

[11] J. L. Newcomer, "Inconsistencies in Students' Approaches to Solving Problems in Engineering," in *Proceedings – 40th Annual Frontiers in Education Conference, ASEE/IEEE*, 2010.

Appendix

Table A1 – Semesters that assessed the problems

Problems	Assessed semesters			
	Summer 2018	Fall 2018	Spring 2019	Fall 2019
1	x	x		
2			x	x
3			x	x
4	x	x		
5			x	x
6	x	x		
7	x	x		x
8	x	x		x
9	x			
10		x	x	x
11			x	x
12		x	x	x
13			x	x
14			x	x
15	x			