Bias and Precision in Instructor Grading of Concept Inventories in Geotechnical Engineering Courses

Dr. Simon Thomas Ghanat P.E., The Citadel

Dr. Simon Ghanat is an Assistant Professor of Civil and Environmental Engineering at The Citadel (Charleston, S.C.). He received his Ph.D., M.S., and B.S. degrees in Civil and Environmental Engineering from Arizona State University. His research interests are in Engineering Education and Geotechnical Earthquake Engineering. He previously taught at Bucknell University and Arizona State University.

Dr. James Kaklamanos, Merrimack College

Dr. James Kaklamanos is Associate Professor of Civil Engineering and Zampell Family Faculty Fellow at Merrimack College in North Andover, Mass. Jim received his M.S. and Ph.D. in Civil and Environmental Engineering from Tufts University, along with a B.S. in Civil Engineering from the same institution. He specializes in geotechnical engineering, and his published work has included research on site response analyses, ground-motion prediction equations, uncertainty in earthquake engineering, and engineering education. At Merrimack, Jim has taught courses in geotechnical engineering, foundation engineering, earth retaining structures, earthquake engineering, engineering mechanics, and engineering probability and statistics.

Dr. Tanya Kunberger P.E., Florida Gulf Coast University

Dr. Kunberger is a Professor in the Department of Environmental and Civil Engineering in the U. A. Whitaker College of Engineering at Florida Gulf Coast University. Dr. Kunberger received her B.C.E. and certificate in Geochemistry from the Georgia Institute of Technology and her M.S. and Ph.D. in Civil Engineering with a minor in Soil Science from North Carolina State University. Her areas of specialization are geotechnical and geo-environmental engineering. Educational areas of interest are self-efficacy and persistence in engineering and development of an interest in STEM topics in K-12 students.

Dr. Corrie Walton-Macaulay Ph.D., P.E., Saint Martin's University

With degrees in both Bachelor of Science and Master of Science from the University of Arkansas, and a doctoral degree from the University of Kentucky, Dr. Corrie Walton-Macaulay is now a Geotechnical Engineering Assistant Professor in the Civil and Environmental Engineering Department at Saint Martin's University. He teaches the traditional geotechnical course of soil mechanics, but also teaches civil engineering materials, mechanics of materials and pavement design. His research area is in unsaturated soil mechanics, energy geotechnics, and transportation infrastructure resiliency.

Address: 5000 Abbey Way SE, Saint Martin's University, Lacey, WA 98503

Dr. Suresh Immanuel P.E., University of Evansville

Dr. Immanuel Selvaraj is an associate professor of civil engineering at the University of Evansville, IN. He holds a PhD degree from Auburn University and a licensed professional engineer. His research interests are in engineering education, pavement design and analysis, pavement management, and pavement instrumentation. At the University of Evansville, he teaches courses such as transportation engineering, soil mechanics, geotechnical engineering, advanced pavement design, pavement management, and surveying.

Dr. David A. Saftner, University of Minnesota Duluth

David Saftner is an Associate Professor at the University of Minnesota Duluth. He received a BS in Civil Engineering from the United States Military Academy and MS and Ph.D. in Civil Engineering from the University of Michigan.

Dr. Brock E. Barry P.E., U.S. Military Academy

Dr. Brock E. Barry, P.E. is Professor of Engineering Educaiton in the Department of Civil & Mechanical Engineering at the United States Military Academy, West Point, New York. Dr. Barry holds a Bachelor of Science degree from Rochester Institute of Technology, a Master of Science degree from University of Colorado at Boulder, and a PhD from Purdue University. Prior to pursuing a career in academics, Dr. Barry spent 10-years as a senior geotechnical engineer and project manager on projects throughout the United States. He is a licensed professional engineer in multiple states. Dr. Barry's areas of research include assessment of professional ethics, teaching and learning in engineering education, nonverbal communication in the classroom, and learning through historical engineering accomplishments. He has authored and co-authored a significant number of journal articles and book chapters on these topics.

Shawn Griffiths, University of Wyoming

Shawn Griffiths is an Assistant Professor of Civil Engineering at the University of Wyoming. Shawn holds a B.S. in Civil Engineering from Utah State University (2009), M.S. in Civil Engineering from the University of Arkansas (2011) and a Ph.D. in Geotechnical Engineering from the University of Texas at Austin. He believes attitude and hard work are the two most important ingredients in successful learning. As such, he strives to have a classroom that is filled with ideas, questions, and a positive environment. He also believe school is the place to make mistakes, and encourages students to be brave enough to "try, try again," until they succeed.

Dr. Craig M. Shillaber, Northeastern University

Craig M. Shillaber is an assistant teaching professor in the Department of Civil and Environmental Engineering at Northeastern University. He earned a Ph.D. in civil engineering from Virginia Tech in 2016, an M.S. in civil engineering from Virginia Tech in 2009, and a B.S. in civil engineering from the University of New Hampshire in 2008. His research interests include sustainability education in civil engineering, geotechnical subsurface characterization, developing and improving methods for assessing life-cycle embodied energy and carbon emissions for the geotechnical profession, and sustainable geotechnics. Dr. Shillaber is a member of the American Society of Civil Engineers (ASCE) and the Deep Foundations Institute (DFI), where he has served on the Sustainability Committee since 2013.

Dr. Chris Swan, Tufts University

Chris Swan is Dean of Undergraduate Education for the School of Engineering and an associate professor in the Civil and Environmental Engineering department at Tufts University. He has additional appointments in the Jonathan M. Tisch College of Civic Life and the Center for Engineering Education and Outreach at Tufts. His current engineering education research interests focus on community engagement, service-based projects and examining whether an entrepreneurial mindset can be used to further engineering education innovations. He also does research on the development of reuse strategies for waste materials.

Bias and Precision in Instructor Grading of Concept Inventories in Geotechnical Engineering Courses

Introduction

An assessment of bias and precision in instructor grading was undertaken at several private and public institutions having civil engineering programs. At five institutions, undergraduate civil engineering majors completed a concept inventory at the conclusion of their first course in geotechnical engineering. The ten-question instrument focused on fundamental concepts in geotechnical engineering to assess students' knowledge gained throughout the course. A random sample of ten surveys was collected from each of the five institutions, leading to a dataset of n = 50 concept inventory surveys encompassing a breadth of student populations. A team of geotechnical engineering professors from seven institutions (four institutions included in the dataset, and three institutions not included in the dataset) independently graded the 50 concept inventory surveys, using an established solution to the instrument. The end result was a distribution of seven instructor scores for each question within the dataset of student responses.

The objectives of this study are (1) to quantify instructor bias in grading concept inventory surveys by examining whether there are differences between instructors' grading of their own students and instructors' grading of anonymous surveys, and (2) to quantify instructor precision (instructor-to-instructor variability) in grading. Statistical analyses were performed on the scores of the concept inventory surveys to quantify the distributions of instructor grading within the distributions of student scores. In the context of an undergraduate geotechnical engineering course, this paper discusses the concept inventory, grading criteria, institutional contexts, results of statistical analyses, and suggestions for future research. With a better understanding of instructor grading patterns, we can work towards reducing bias and increasing precision in instructor grading in undergraduate civil engineering courses.

Background

The notion of bias in grading has been an area of research for several decades. While some earlier research focused predominantly on grading variability by instructors [1] or comparing instructor versus student scoring [2], others began to investigate the impact of student-instructor interaction on both grading and learning [3]. Broader studies [4] have considered the influence different grading systems may have on student performance in general; focusing less on specific bias and placing a greater emphasis on the impact and correlation of grades and knowledge gained. More recent research [5 - 7] has recognized and begun to address the existence of both conscious (explicit) and unconscious (implicit) bias and its impact on assessment.

Related research has sought to distinguish between and potentially minimize variations in grading by instructors - either from bias or from fundamental approach differences, particularly for different sections of the same course. Investigations have addressed these instructor influences across different sections of the same course at the same university [8], different but related courses at the same university [9], and the same course at multiple different universities

[10]. The research in this paper seeks to add to the examination of instructor bias and precision in a singular course across multiple institutions.

Assessment measure and grading criteria

A ten-question concept inventory survey was administered in class to all students present on the last day of the introductory geotechnical engineering course at five institutions. Students were given 15 minutes to complete the instrument without any notes or references. It is important to note that the concept inventory survey was a low-stakes assessment because it did not affect students' grades in the course in any manner. This end-of-the-semester concept inventory survey assesses fundamental concepts in soil mechanics and geotechnical engineering, as well as material from prerequisite courses (such as Mohr's Circle). This concept inventory survey was implemented in two previous studies [11, 12] at both the beginning and end of the semester in undergraduate geotechnical engineering courses at multiple institutions; in another study [13], a similar concept inventory was applied to upper-level geotechnical engineering courses. These studies assessed students' learning of geotechnical engineering topics as a result of various pedagogical techniques and educational factors at the institutions. In the current study, the concept inventory survey was only administered at the end of the semester. The ten-question concept inventory survey is provided in Table 1.

Table 1. Geotechnical engineering concept inventory survey used in this study.

No.	Question
Q1	What are some engineering characteristics of fine-grained soils?
Q2	What does high relative density and low void ratio indicate?
Q3	Why do we need to assess the shear strength of soil?
Q4	What is the difference between compaction and consolidation?
Q5	Why do we compact soils in earthwork?
Q6	Why is determination of water content of soil important?
Q7	What are some of the causes of settlement in soils (i.e., sources of settlement in soils)?
Q8	What is the difference between normally consolidated and over-consolidated clay?
Q9	What is difference between the drained condition and undrained condition?
Q10	The major and minor principal stresses at a certain point in the ground are 450 and 200 kPa, respectively. Determine the maximum shear stress at this point.

The concept inventory survey was administered at five institutions with the undergraduate civil engineering programs listed in Table 2; the class sizes ranged from 15 to 47. In order to reduce the potential for sampling bias (to avoid the dataset being dominated by institutions with larger class sizes), a random sample of ten concept inventory surveys was collected from each of the five institutions. The end result was a dataset of $n = 50$ concept inventory surveys encompassing a breadth of student populations.

- The Citadel: small public university in the Southern U.S.
- Merrimack College: small private university in the Northeast U.S.
- University of Evansville: small private university in the Central U.S.
- Bucknell University: small private university in the Northeast U.S.
- Florida Gulf Coast University (FGCU): mid-size public university in the Southern U.S.

A team of geotechnical engineering professors from seven institutions (four institutions included in the dataset [The Citadel, Merrimack, Evansville, Bucknell], and three institutions not included in the dataset) independently graded the 50 concept inventory surveys. For the four instructors [The Citadel, Merrimack, Evansville, Bucknell], 10 of the 50 surveys were from their own students, and 40 of the 50 surveys were from students of other institutions. The three instructors whose institutions (University of Minnesota Duluth, Northeastern University, and University of Wyoming) were not included in Table 2 graded the surveys in a fully blind manner. One instructor is a faculty member at University of Minnesota Duluth, a Midwestern, public M1 University with an enrollment of roughly 10,000 undergraduates. The second additional instructor is a faculty member at Northeastern University, a private R1 institution, with a total enrollment of 13,800 undergraduates. Finally, the third additional instructor is a faculty member of University of Wyoming, a land grant university with an enrollment of 10,000 undergraduates across seven colleges.

Table 2. Characteristics of the five institutions at which the survey was administered.

Institution	The Citadel	Merrimack	Evansville	Bucknell	FGCU
Total undergraduate enrollment	2,773	3,488	2,248	3,571	13,917
Civil engineering undergraduate enrollment	235	100	65	101	335
Freshman acceptance rate	82%	82%	71%	30%	61%
Prerequisite courses before geotechnical engineering	*Mechanics of Materials* and *Fluid Mechanics*	*Mechanics of Materials* and *Fluid Mechanics*	*Mechanics of Materials*	*Solid Mechanics I* and *Fluid Mechanics*	*Mechanics of Materials* and *Fluid Mechanics*
Number of students completing survey in course	37	22	17	15	47
Number of student surveys included in this study	10	10	10	10	10

Comparisons of course curricula

Table 3 displays a cross-comparison of the topics covered in the introductory geotechnical engineering course at the five institutions used in this study. It can be seen from Table 3 that all topics of the concept inventory survey were covered at the five institutions. Table 3 also shows that the geotechnical engineering course contents are comparable; given the similarity of the courses, students at different institutions would likely give similar answers on the concept inventory survey. However, some institution-to-institution variation is expected due to differences in the student populations and instructors' teaching methods.

Table 3. Comparisons of geotechnical engineering topics covered at the five institutions.

	Citadel	**Merrimack**	**Evansville**	**Bucknell**	**FGCU**
Course topic	*Coverage at each institution*				
Geology	✓	✓	✓	✓	✓
Index Properties and Soil Classifications (Q1)	✓	✓	✓	✓	✓
Phase Relations (Q2, Q6)	✓	✓	✓	✓	✓
Compaction (Q4, Q5)	✓	✓	✓	✓	✓
Permeability	✓	✓	✓	✓	✓
Seepage/flow nets	✓	✓	✓	✓	✓
Stresses in soils	✓	✓	✓	✓	✓
Compressibility of soils (Q4, Q7, Q8)	✓	✓	✓	✓	✓
Shear Strength of soils (Q3, Q9, Q10)	✓	✓	✓	✓	✓
Lateral Earth Pressures	✓			✓	✓
Bearing Capacity	✓			✓	

Standardized rubric and solutions to the survey questions

Each instructor used an established solution to the instrument, provided in the appendix to this article. Table 4 shows the standardized grading rubric for the concept inventory survey. Graders were instructed to score each of the ten questions using the following standardized rubric: awarding a score of zero (0) for an incorrect, off-base answer or no answer at all; awarding a score of 0.5 for a partially correct answer (as detailed in the appendix to this paper, some of the solutions explicitly describe the manner in which partial credit is to be awarded); or awarding a score of one (1.0) for correct answer. In particular, questions 1, 3, 5, 6, and 7 can be challenging to answer or grade due to their open-ended nature. For students, these open-ended questions require higher-level cognition and synthesis (necessary for success in geotechnical engineering). For graders, these questions require greater interpretation and judgment when determining acceptable solutions.

Table 4. Grading rubric for the concept inventory survey.

Points Awarded per Question	Rubric
0	No credit for incorrect, off-base answer or no answer at all
0.5	Partial credit for partially correct answer (see appendix)
1.0	Full credit for correct answer

Data analysis and statistics

Figure 1 illustrates a stacked bar chart of the proportion of each of the three points (i.e., 0, 0.5, or 1.0) awarded by each instructor. Within the group of seven instructors, the individual identities are kept anonymous; they are herein referred to as Instructors 1-7. Figure 1 shows that Instructors 1 and 4, respectively, awarded the maximum (43%) and minimum (13%) percent zeros; across all instructors, an average of 28% of scores were zero. The greatest and the smallest amount of partial credit (scores of 0.5) were awarded by Instructors 6 and 7, respectively; on average, instructors awarded partial credit 19% of the time. The highest and the lowest percentages of full credit (scores of 1.0) were awarded by Instructors 7 (70%) and 1 (33%), respectively; an average of 53% of scores awarded were equal to 1.0.

Figure 2 shows question-by-question of the proportion of each of the three points awarded by each instructor. Question 10 on Mohr's Circle (a quantitative calculation) shows remarkably similar patterns among the seven graders. Question 8 appears to have the second most agreement in the scores across all instructors; this question assesses students' understanding of the difference between normally consolidated and over-consolidated soil (a question with a clearly defined answer). Questions 3, 5, and 7 appear to show the most variability between instructors. Question 3, an open-ended question, asks, "Why do we need to assess the shear strength of soil?" Questions that are more open-ended (that have a less narrowly defined answer) tend to have a wider distribution of scores, as would be expected.

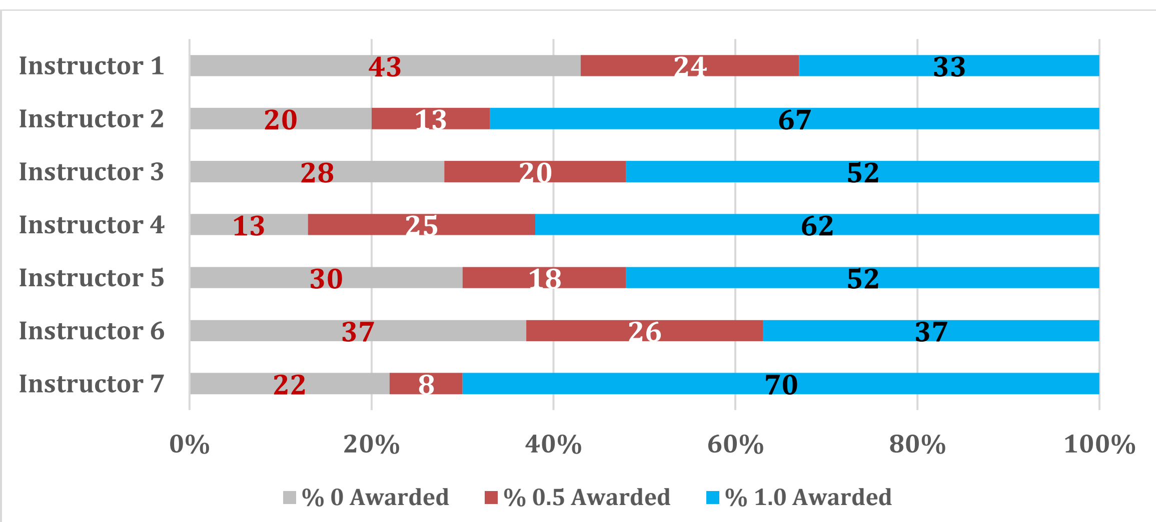

Figure 1. Percentages of each of the three scores (0, 0.5, 1) each instructor awarded on questions

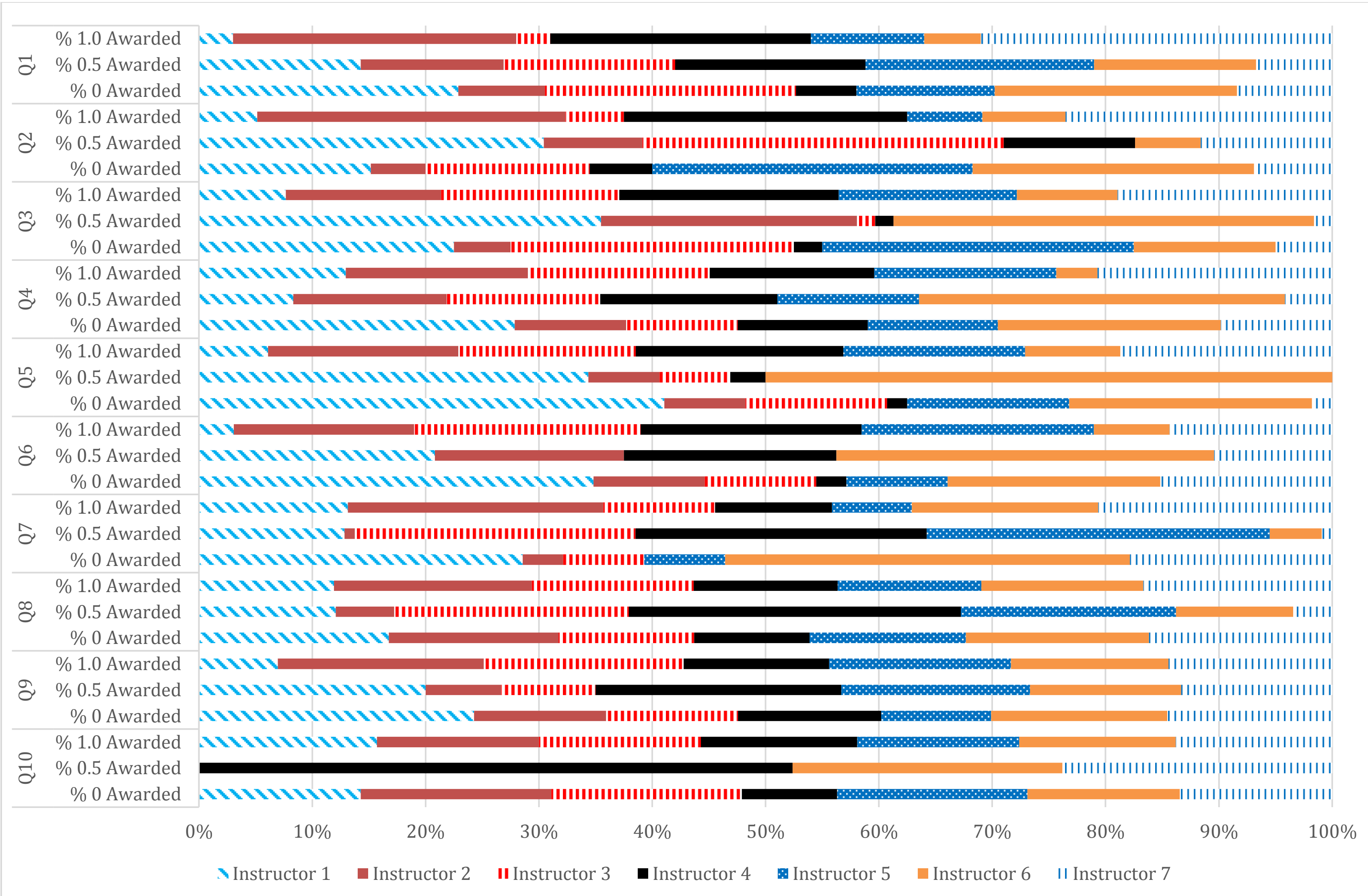

Figure 2. Question-by-question summary of the proportion of each of the three scores each instructor awarded.

Data from five institutions were examined to see whether the conditions of normality were met. Based upon inspections of the distributions of scores from each institution, we concluded that not all randomly selected samples of size n = 10 came from normally distributed populations. Therefore, nonparametric statistical tests were employed to evaluate the statistical significance of the results.

Kruskal-Wallis non-parametric tests were conducted to confirm these trends, and the results are shown in Table 5; the median scores were not statistically different for Questions 8 and 10 (arguably the two most objective questions on the survey, as previously noted), but the median scores for all other questions exhibited statistically significant differences among the instructors.

Table 5. Results of Kruskal-Wallis non-parametric test conducted for each survey question.

H_0 : $Median_1 = Median_2 = Median_3 = Median_4 = Median_5 = Median_6 = Median_7$ H_a : $Median_1 \neq Median_2 \neq Median_3 \neq Median_4 \neq Median_5 \neq Median_6 \neq Median_7$		
Question	Test statistic	P-values
Q1	48.81	<0.00001
Q2	67.82	<0.00001
Q3	48.18	<0.00001
Q4	21.57	0.0002
Q5	37.19	<0.00001
Q6	53.17	<0.00001
Q7	35.47	<0.00001
Q8	1.98	0.74
Q9	18.37	0.001
Q10	0.735	0.95

Distributions of total scores

Figure 3 illustrates box-and-whisker plots for instructors 1-7 for the total scores of the $n = 50$ surveys each instructor graded. Instructors 2, 4 and 7 have the same median score of 75%, but different distributions of scores. The box plots show a similar distribution of scores and median for instructors 3 and 5 (medians of approximately 65%). Instructors 1 and 6 provided the lowest scores in their grading, with Instructor 1 having the lowest first quartile, median, and third quartile in the dataset (median of 45%). The lowest and highest scores are associated with Instructors 1 and 7, respectively.

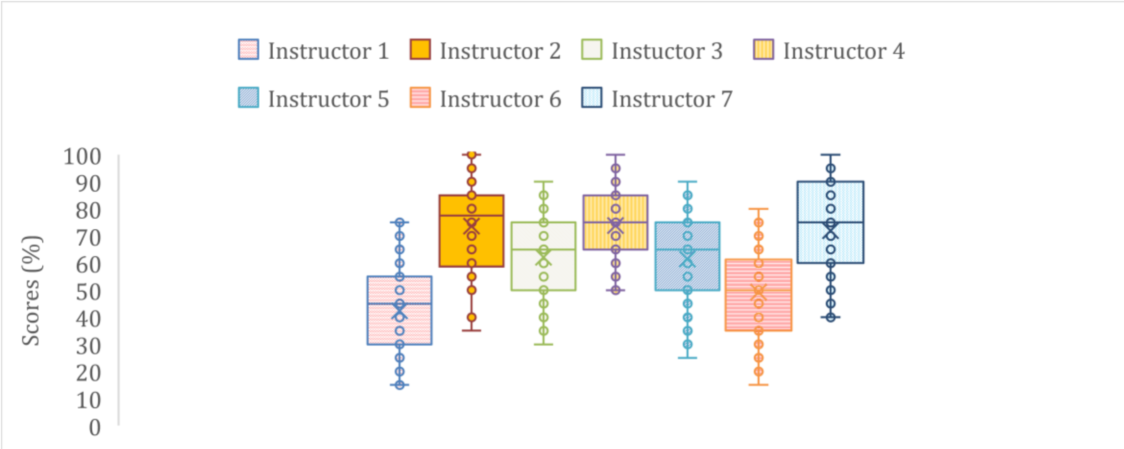

Figure 3. Box-and-whisker plot of total scores for instructors 1-7.

The Kruskal-Wallis non-parametric test was performed across the seven instructors to assess differences in the total scores each instructor provided. An overall p-value was used to determine if there were any significant differences among median scores. The p-value for the entire data set indicated there were highly significant differences within the overall result ($p < 0.00001$). A post-hoc test was conducted to determine where the significant differences were located. Table 5 shows the p-values for each instructor pair, and statistically significant differences were observed for many instructors in the dataset. Again, the post-hoc test confirmed that instructor 1 provided scores that were statistically different than the other instructors at a significance level of 5%.

Table 5. Kruskal-Wallis p-values indicating significantly different scores among the instructors.

	p-values for instructor pairings						
Instructor	vs.1	vs. 2	vs. 3	vs. 4	vs. 5	vs. 6	vs. 7
1		<0.00001	<0.00001	<0.00001	<0.00001	0.04865	<0.00001
2			0.00046	0.49	0.0003	<0.00001	0.49
3				0.0004	0.81	0.0002	0.0007
4					0.0002	<0.00001	0.76
5						0.0006	0.003
6							<0.00001
7							

Assessment of instructor grading bias

In this section, we will quantify instructor bias in grading concept inventory surveys by examining whether there are differences between instructors' grading of their own students and instructors' grading of students from other institutions. For the four instructors who graded surveys and whose students were also included in the dataset (Instructors 1 to 4), Table 6 provides two summary ratios: (1) mean from grading own students / mean across all graders for own students, and (2) mean from grading other students / mean across all graders for other students. Table 4 also shows the difference in these two ratios; as the difference increases, the potential instructor bias increases as well.

Table 6. Summary of ratios for instructors grading their own students vs. grading other students.

	Mean from grading own students / Mean across all graders for own students	Mean from grading other students / Mean across all graders for other students	Difference in ratio own students vs. ratio other students
Instructor 1	0.78	0.65	0.13
Instructor 2	1.19	1.18	0.01
Instructor 3	1.00	0.98	0.02
Instructor 4	1.22	1.19	0.03
Instructor 5	-	0.98	-
Instructor 6	-	0.81	-
Instructor 7	-	1.16	-

Table 6 shows that Instructors 3 and 5 are aligned well with other instructors, instructors 2, 4, and 7 are more generous graders than other instructors. Table also shows that instructor 1 is stricter than other instructors, and even stricter on the students of others. On average, the difference in the ratio of instructors' grading of their own students and instructors' grading of anonymous surveys is only 0.05. Individually, the difference in ratios ranges from 0.01 to 0.13. Instructor 1 shows the highest difference in ratio at 0.13, which is more than four times the next highest difference.

Figure 4 illustrates differences in how instructors graded their own students compared to the group. Note that three of the four instructors graded the students of Instructor 1 (Students #1-#10) the highest of any others, so it is possible that differences within the student population (i.e. higher scores at Institution 1) may explain the higher scores assigned by Instructor 1 to their own students.

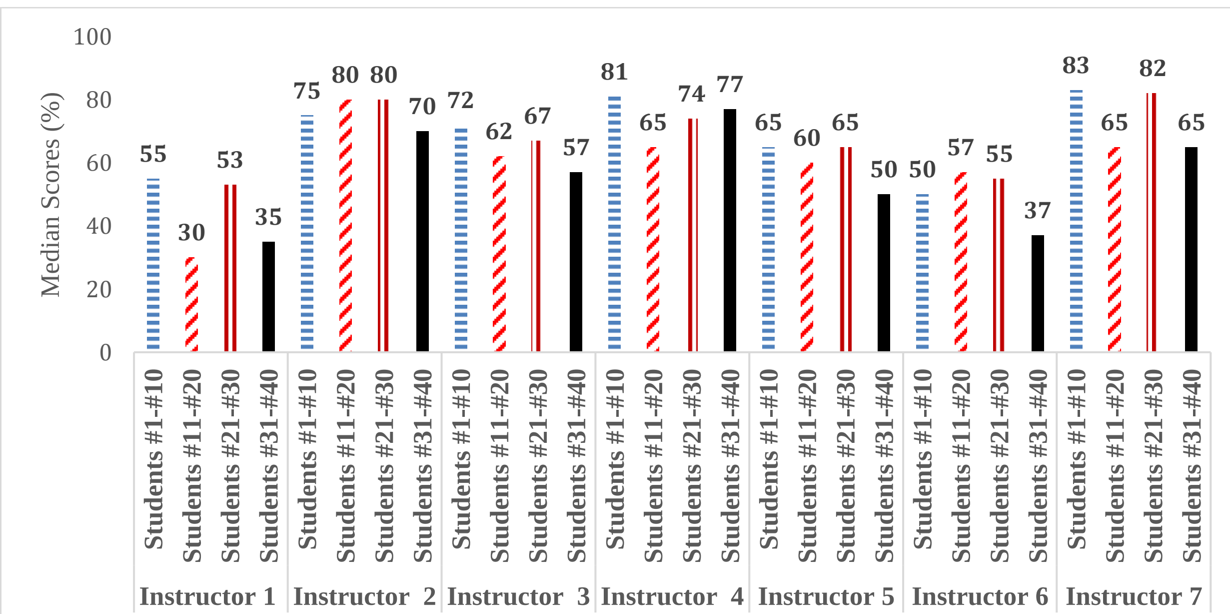

Figure 4. Instructors grading their own students vs. grading other instructors’ students.

Statistical calculations were performed to assess whether there are significant differences between instructors' grading of their own students and instructors' grading of anonymous surveys. The statistical significance for each of the four instructors was evaluated by using the Mann-Whitney test at 5% level of significance, and the results are shown in Table 7. The results showed that the difference between instructors’ grading of their own students and instructors’ grading of anonymous surveys was only significant for Instructor 1 at the 5% level of significance. However, as Figure 3 showed, it is possible that the students of Instructor 1 objectively earned higher scores than students from other institutions.

Table 7. Results of Mann-Whitney test for instructor grading bias.

Hypothesis	P-value
Median of instructor 1 grading of own students = Median of instructor 1 grading of anonymous surveys	0.005 < 0.05
Median of instructor 2 grading of own students = Median of instructor 2 grading of anonymous surveys	0.38
Median of instructor 3 grading of own students = Median of instructor 3 grading of anonymous surveys	0.50
Median of instructor 4 grading of own students = Median of instructor 4 grading of anonymous surveys	0.55

It was also hypothesized that there is no difference among the median scores of instructors’ grading for their own students and instructors’ grading of anonymous surveys from each institution. This comparison was completed using the Kruskal-Wallis non-parametric test. The results shown in Table 8 demonstrate that for Instructor 1, there was a significant difference among the median scores ($p < 0.05$).

Table 8. Results of the Kruskal-Wallis non-parametric test for instructor grading bias.

Hypothesis	Test statistic	P-value
$H_0 : Median_1 = Median_2 = Median_3 = Median_4$ $H_a : Median_1 \neq Median_2 \neq Median_3 \neq Median_4$ [Instructor 1]	17.57	0.0005 < 0.05
$H_0 : Median_1 = Median_2 = Median_3 = Median_4$ $H_a : Median_1 \neq Median_2 \neq Median_3 \neq Median_4$ [Instructor 2]	1.08	0.78
$H_0 : Median_1 = Median_2 = Median_3 = Median_4$ $H_a : Median_1 \neq Median_2 \neq Median_3 \neq Median_4$ [Instructor 3]	5.51	0.14
$H_0 : Median_1 = Median_2 = Median_3 = Median_4$ $H_a : Median_1 \neq Median_2 \neq Median_3 \neq Median_4$ [Instructor 4]	5.97	0.11

Conclusions and suggestions for future research

The objectives of this study were (1) to quantify instructor bias in grading concept inventory surveys by examining whether there are differences between instructors' grading of their own students and instructors' grading of anonymous surveys, and (2) to quantify instructor precision (instructor-to-instructor variability) in grading. We found that instructor bias is present in grading of undergraduate geotechnical engineering concept inventories (for at least one of seven instructors), but this perceived bias may ultimately be a result of differences in student populations (therefore justifying higher concept inventory scores for this instructor's students).

Despite the development of a detailed solution to the concept inventory, the distributions of scores varied from instructor to instructor, with instructor scoring patterns generally falling into categories of high (Instructors 2, 4, and 7; median ≈ 75%), medium (Instructors 3 and 5; median ≈ 65%), and low (Instructors 1 and 6; median ≈ 45-50%). The variability in instructors' scores is smallest when questions are numerical (e.g. Question 10) or have a distinct correct answer (e.g. Question 8), and greatest when questions are open-ended and require greater interpretation and judgment when analyzing students' responses.

Due to differences among instructors, student populations, and institutions, the existence of grading biases are not unexpected. Therefore, mitigation strategies are needed to reduce these biases and enhance the reliability of the instrument in measuring student differences in understanding of geotechnical engineering concepts. Possible avenues to pursue include having the responses from each institution reviewed by a group of other instructors from other institutions participating in the implementation of the instrument. This would reduce institutional bias, especially if the grading is done through a 'double-blind' process. Additionally, responses from all participating institutions can be reviewed by a single (or small group of) 'knowledge expert(s)' who are not implementing the instrument at their own institution. This latter strategy places a higher level of consistency in grading and providing increased reliability in the interpretation of results, though they may come from variable sources.

In undergraduate geotechnical engineering courses, future research may be needed on bias and precision in instructor grading on high-stakes assessments such as examinations. Ultimately, the goal is to work towards reducing bias and increasing precision in instructor grading in undergraduate civil engineering courses.

References

[1] C. Work, "A Nationwide Study of the Variability of Test Scoring by Different Instructors," *Engineering Education,* 67, 3, 241 - 248, December 1976.

[2] J. M. Moxley and D. Eubanks, "On Keeping Score: Instructors' vs. Students' Rubric Ratings of 46,689 Essays," *Writing Program Administration,* 39, 2, 53 - 80, Fall 2015.

[3] B. Harper and J. Hughey, "Effects of Communication Responsiveness Upon Instructor Judgement Grading and Student Cognitive Learning," *Communication Education*, 35, 147 - 156, April 1986.

[4] S. Kleinman, M. B. Leidman, and A. J. Longcore, "The Changing Landscape of Grading Systems in U.S. Higher Education," *Perspectives: Policy and Practice in Higher Education*, 22, 1, 26 - 33, 2018. https://doi.org/10.1080/13603108.2017.1279692.

[5] P. Steinke and P. Fitch, "Minimizing Bias When Assessing Student Work," *Research and Practice in Assessment*, 12, 87 - 95, Winter 2017.

[6] J. Malouf, "Bias in Grading," *College Teaching*, 56, 3, 191 - 192, 2008. https://doi.org/10.3200/CTCH.56.3.191-192.

[7] J. M. Malouf and E. B. Thorsteinsson, "Bias in Grading: A Meta-Analysis of Experimental Research Findings," *Australian Journal of Education*, 60, 3, 245 - 256, November 2016. https://doi.org/10.1177/0004944116664618.

[8] C. D. Wright, A. L. Huang, K. M. Cooper, and S. E. Brownell, "Exploring Differences in Decisions about Exams among Instructors of the Same Introductory Biology Course," *International Journal for the Scholarship of Teaching and Learning*, 12, 2, 14, January 2018. https://doi.org/10.20429/ijsotl.2018.120214.

[9] A. Satyanarayana, J. Kusyk, and H. Li, "Common Assessment of Two Related Courses to Reduce Grading Bias and Improve Readiness of the Students for Corporate Environments," *Proceedings of the American Society for Engineering Education 2017 Mid-Atlantic Section Conference*, Reading, Pennsylvania, 6 - 7 October, 2017.

[10] B. Z. Dymond, M. Swenty, and C. Carroll, "The Influence of Grading Bias on Reinforced Concrete Exam Scores at Three Different Universities," *Proceedings of the American Society for Engineering Education 2018 Annual Conference and Exposition*, Salt Lake City, Utah, 24 - 27 June 2018.

[11] S.T. Ghanat, J. Kaklamanos, K. Ziotopoulou, I. Selvaraj, and D. Fallon, "A multi-institutional study of pre- and post-course knowledge surveys in undergraduate geotechnical engineering courses," *Proceedings of the American Society for Engineering Education 2016 Annual Conference and Exposition*, New Orleans, Louisiana, 26–29 June 2016.

[12] S.T. Ghanat, J. Kaklamanos, C. Walton-Macaulay, S.I. Selvaraj, D.A. Saftner, C. Swan, and T. Kunberger, "Assessing the impact of educational factors on conceptual understanding of geotechnical engineering topics," *Proceedings of the American Society for Engineering Education 2018 Annual Conference and Exposition*, Salt Lake City, Utah, 24–27 June 2018.

[13] S.T. Ghanat, J. Kaklamanos, S.I. Selvaraj, C. Walton-Macaulay, and M. Sleep. "Assessment of students' prior knowledge and learning in an undergraduate foundation engineering course," *Proceedings of the American Society for Engineering Education 2017 Annual Conference and Exposition*, Columbus, Ohio, 25–28 June 2017.

Appendix: Solutions to Geotechnical Engineering Concept Inventory

1. **What are some of engineering characteristics of fine-grained soils?**

   For clays:
   Generally possess low shear strength
   Plastic and compressible
   Can lose part of shear strength upon wetting
   Can lose part of shear strength upon disturbance
   Can shrink upon drying and expand upon wetting
   Generally very poor material for backfill
   Generally poor material for embankments
   Can be practically impervious
   Clay slopes are prone to landslides

   For silts:
   Relatively low shear strength
   High capillarity and frost susceptibility
   Relatively low permeability
   Difficult to compact

   Two or three need to be stated (1 pt); if less, 0.5 pt.

2. **What does high relative density and low void ratio indicate?**

   Strong or incompressible soils

3. **Why do we need to assess the shear strength of soil?**

   Geotechnical strength parameters address the ability of soil to accept the loads imparted by the foundation without failing. The strength of the soil is governed by its capacity to sustain shear stresses, so we satisfy geotechnical strength requirements by comparing shear stresses with shear strengths and designing accordingly.

4. **What is the difference between compaction and consolidation?**

   - **Compaction** is the process of compacting soils by removing air from voids with repeated application of mechanical energy (0.5 pt)
   - **Consolidation** is settlement of soil due to the expulsion of water from the voids, as stress is transferred from the pore water to the soil skeleton (0.5 pt)

5. **Why do we compact soils in earthwork?**

   Compaction is one of the most common and cost effective means of stabilizing soils. An extremely important task of geotechnical engineers is the performance and analysis of field control tests to assure that compacted fills are meeting the prescribed design specifications. Design specifications usually state the required density (as a percentage of the "maximum" density measured in a standard laboratory test), and the water content. In general, **most engineering properties**, such as the strength, stiffness, resistance to shrinkage, and imperviousness of the soil, **will improve by increasing the soil density**.

6. **Why is determination of water content of soil important?**

   For many soils, the water content may be an extremely important index used for establishing the relationship between the way a soil behaves and its properties (hydraulic conductivity, consolidation, shear strength properties, etc.). The consistency of a fine-grained soil largely depends on its water content. The water content is also used in expressing the phase relationships of air, water, and solids in a given volume of soil.

7. **What are some of the causes of settlement in soils (i.e., sources of settlement in soils)?**

   Some acceptable answers:
   Application of structural loads on footings
   Weight of a recently placed fill
   Falling groundwater table
   Formation of sinkholes
   Underground mining or tunneling
   Secondary compression of underlying soils
   Lateral movements resulting from nearby excavations

   Two or more need to be stated (1 pt); if less, 0.5 pt.

8. **What is the difference between normally consolidated and over-consolidated clay?**

   - **Normally consolidated clay:** Vertical effective stress in the field has never been higher than its current magnitude (0.5 pt)
   - **Overconsolidated clay:** Vertical effective stress in the field once was higher than its current magnitude (0.5 pt)

9. **What is difference between the drained condition and undrained condition?**
    - **Drained condition:** a limiting condition under which there is no excess pore water pressure in the soil (0.5 pt)
    - **Undrained condition:** a limiting condition under which water is not allowed to flow into or out of the soil, leading to excess pore water pressures in response to either contraction or dilation of the soil skeleton (0.5 pt)

10. **The major and minor principal stresses at a certain point in the ground are 450 and 200 kPa, respectively. Determine the maximum shear stress at this point.**

    Maximum shear stress = (450 – 200)/ 2 = 250 / 2 = 125 kPa