Effects of a New Assessment Model on Female and Under-Represented Minority Students

Dr. Geoffrey Recktenwald, Michigan State University

Geoff Recktenwald is a member of the teaching faculty in the Department of Mechanical Engineering at Michigan State University. Geoff holds a PhD in Theoretical and Applied Mechanics from Cornell University and Bachelor degrees in Mechanical Engineering and Physics from Cedarville University. His research interests are focused on best practices for student learning and student success. He is currently developing and researching SMART assessment, a modified mastery learning pedagogy for problem based courses. He created and co-teaches a multi-year integrated system design (ISD) project for mechanical engineering students. He is a mentor to mechanical engineering graduate teaching fellows and actively champions the adoption and use of teaching technologies.

Dr. Michele J. Grimm, Michigan State University

Michele J. Grimm is the Wielenga Creative Engineering Endowed Professor of Mechanical Engineering. In addition to her scientific research, Dr. Grimm has spent a large part of her career focused on curriculum development and enhancement of student learning in engineering. She served on the faculty of Wayne State University for 25 years, where she developed and implemented both undergraduate and graduate programs in biomedical engineering and helped to establish a department of biomedical engineering. Her endowed professorship at MSU focuses on research to increase the success of students in engineering through creative pedagogical techniques.

Dr. Grimm completed her B.S. in Biomedical Engineering and Engineering Mechanics at The Johns Hopkins University in 1990 and her Ph.D. in Bioengineering at the University of Pennsylvania in 1994.

She has just finished a 3-year rotation as a program director for three BME-related programs at the National Science Foundation. She is also completing her 5-year appointment as a commissioner with ABET's Engineering Accreditation Commission. She is a Fellow of the American Society of Mechanical Engineers, the Biomedical Engineering Society, and the American Institute of Medical and Biological Engineering.

Dr. Ron Averill, Michigan State University

Ron Averill joined the faculty at Michigan State University in 1992. He currently serves as the Associate Chair of Undergraduate Studies in the Department of Mechanical Engineering. His research focus is on pedagogy, design optimization of large and complex systems, and design for sustainable agriculture.

Sara Roccabianca, Michigan State University

Sara Roccabianca is an Assistant Professor in the Department of Mechanical Engineering at Michigan State University (MSU). She was born and raised in Verona, Italy and received her B.S. and M.S. in Civil Engineering from the University of Trento, Italy. She received her Ph.D. in Mechanical Engineering from the University of Trento in 2011. She then was a Postdoctoral Fellow at Yale University, in the Department of Biomedical Engineering, working on cardiovascular mechanics. Sara's research at MSU focuses on urinary bladder mechanics and growth and remodeling associated with bladder outlet obstruction (e.g., posterior urethral valves in newborn boys or prostate benign hyperplasia in men over the age of 60). Her goals are to (i) develop a micro-structurally motivated mechanical model to describe the non-linear elastic behavior of the urinary bladder wall, (ii) develop a stress-mediated model of urinary bladder adaptive response, and (iii) understand the fundamental mechanisms that correlate the mechanical environment and the biological process of remodeling in the presence of an outlet obstruction.

Effects of a New Assessment Model on Female and Under-Represented Minority Students

Abstract

In 2016, Michigan State University developed a new model of classroom education and assessment in their Mechanics of Materials course. This model used a modified mastery approach that stresses formative assessment, guidance in the problem-solving process, and structured student reflection. We now refer to this new approach as SMART Assessment - short for Supported Mastery Assessment using Repeated Testing. The effects of this model have been very positive, and results on overall student success in Mechanics of Materials have been presented in full at prior ASEE conferences.

In this paper, we focus on the effects of this new assessment model on the performance of women and underrepresented minority students, while accounting for other measures such as incoming GPA and performance in the prerequisite course, Statics. The evaluation was conducted across 3.5 academic years and involved 1222 students divided among 9 experimental sections and 6 control sections.

Statistical analysis indicated that women from non-underrepresented ethnicities and men from underrepresented ethnicities were not negatively affected by the introduction of the SMART method, with both groups earning slightly higher grades than their male, non-underrepresented peers. However, female students who also were a member of an underrepresented racial or ethnic minority did earn statistically lower grades than their peers. Though from a very small group of students (n = 14), this result demonstrates a need for additional research and interventions.

Background

The SMART pedagogical method was developed at Michigan State University in 2016 [1]. The acronym SMART stands for Supported Mastery Assessment using Repeated Testing. The goal of the SMART method is to address concerning trends in student understanding and performance in STEM courses, especially those that focus on problem-solving. The method was developed in response to growing indications that students were passing classes by achieving a level of learning that is lower than what is expected for an engineering graduate. This lower level of learning is often not discerned by current assessment methods. Preliminary investigations determined that this trend was not due to deficiencies in quality of instruction or lack of student effort. Rather, students had figured out a way to subvert the dominant learning paradigm and pass classes without developing the requisite understanding of fundamental concepts or problem-solving skills [1].

This trend in lower levels of learning has coincided with a rise in student usage of online solution guides for homework and other out-of-class learning activities. This approach removes the need for meaningful practice and struggle, which are crucial to the learning process [2,3]. Then, in order to pass examinations, students have increasingly relied on memorization rather than understanding. Instructors have unwittingly facilitated this change by using generous partial credit rubrics that cannot distinguish between simple mistakes and conceptual errors. After passing several college classes by using solution guides and short-term memorization, students begin to use them as their dominant study methods, according to student interviews [9].

SMART pedagogy leverages known pedagogical best practices to create a learning environment where students are challenged and supported [1]. A key component of the SMART method is to motivate students to practice and struggle through problem-solving. This is achieved through an *irreducible* set of course components that modify the standard course structure [4]. Frequent, formative assessments are used to provide spaced repetition and feedback on skills development. An exam appeals process is used to encourage structured student reflection on progress. A mastery rubric is used to set clear expectations of successful problem-solving and to help students develop skills at assessing and troubleshooting their own work. A Compass, or guided problem-solving process, is used to help students develop systematic problem-solving [5]. These components work in tandem to guide, support, and motivate students to learn key concepts and establish problem-solving skills.

Previous work on this topic has shown that the SMART method is transformatively effective. Results from previous studies have shown that **students in the SMART method outperform traditional section students by 1 full standard deviation [1]** when assessed based on correctly answering common final exam problems in ME222: Mechanics of Deformable Solids (sometimes referred to as Mechanics of Materials). These results were independent of instructor and were validated over a span of 3 years. In unpublished recent studies, SMART assessment has been successfully applied to other courses (Dynamics and Fluid Mechanics at Michigan State University) and at other universities (Thermodynamics at the University of Maryland).

The SMART Assessment grading rubric, examination schedule, and grading strategy [1] have several potential benefits -- ones that apply to all students but that may have an even greater impact on underrepresented or at-risk students:

- The potential for bias in awarding partial credit is reduced.
- Testing anxiety may be reduced by increasing the number of lower-stakes examinations and providing multiple attempts at each one [6-8].
- Multiple chances on each examination results in students receiving direct feedback on areas that need improvement prior to the second attempt at the examination. The first attempt at each examination can serve as formative assessment. The time between examination attempts can then be used to provide additional assistance or corrective intervention aimed at improving understanding or skill related to the identified shortcomings. This process could be formalized, though this has not been done to date.

The current paper looks at 3.5 years of student data to determine how student sub-groups benefit from the SMART method. It is the authors' hypothesis that all subgroups benefit equally from the new method, and -- in this particular study -- the authors investigate the effects of gender and race/ethnicity on student success. This study looks specifically at student performance in ME222: Mechanics of Deformable Solids. As such, the purpose of the study is to investigate whether student success in SMART sections of ME222 is similar for all subgroups.

Study Methodology

The SMART method was first implemented in ME222 in the fall of 2016. In this semester, two instructors (B & C) adopted the new method and one instructor (A) maintained a traditional approach to act as a control. A common final exam was used, but course grades were determined independently by each individual section instructor based on their established assessment methods. A similar format was used in the fall of 2017. In the fall of 2018, the control instructor (A) adopted the SMART assessment method (Table 1). Concurrent with the introduction of the SMART format, four sections of ME222 were offered using the traditional assessment system without a comparative, common final exam to the SMART sections. A total of 6 instructors taught the course over this period.

Summer offerings of ME222 have been excluded in this study because of their condensed, 7-week format rather than the typical 15-week semester, making comparisons less relevant. One additional section was excluded as the professor started with the SMART format but shifted to a traditional format of grading after only 2-3 weeks of the semester. It was determined, therefore, that this section could not be classified as either "SMART" or "Traditional".

Table 1 – SMART implementation in ME222 at Michigan State University. Instructor A taught with both methods, Instructors B and C used the SMART method, and instructors D and E taught only using traditional methods. Enrollments varied from 60-130 students per section over this period.

	Fall 2016	**Spring 2016**	**Fall 2017**	**Spring 2018**	**Fall 2018**	**Spring 2019**	**Fall 2019**
Traditional Sections	A	D & E*	A	D		F**	
SMART Sections	B & C		B	C	A & B	C	B & C
Total	3 sections	3 sections	2 sections	2 sections	2 sections	2 sections	2 sections

* Instructor E had two sections this semester
** Instructor F taught in a style that cannot be characterized as traditional or SMART. This data is excluded from the study.

A total of 1222 students are included in this study. There are 9 sections taught by three instructors (A,B,& C) for the SMART method and 6 sections taught by three instructors (A,D,& E) using the traditional approach. Of this population, 253 (about 21%) are women (Table 2), and 104 (about 8.5%) are under-represented minorities (URM) (Table 3). Women who are URMs comprise just over 2% (28 students) of the study population.

All demographic data was obtained from student records. IRB approval was received from the university to allow the researchers to access this data. Students were classified as belonging to an underrepresented minority group if they identified themselves as African American, Hispanic, Native American, Pacific Islander, or multi-racial (2 or more races) within their university record.

Table 2 – Study population summary – Study population based on gender.

	Traditional Sections	SMART Sections	Total
Female	116 (9.5%)	137 (11%)	253 (21%)
Male	370 (30%)	599 (49%)	969 (79%)
Totals	486 (40%)	736 (60%)	1222 (100%)

Table 3 – Study population summary – Study population based on URM status.

	Traditional	SMART	Total
URM	35 (2.9%)	69 (5.6%)	104 (8.5%)
Not URM	451 (37%)	667 (55%)	1118 (91.5%)
Totals	486 (40%)	736 (60%)	1222 (100%)

Anonymized data on each student also included the students' grade in ME222, their grade in the pre-requisite course CE221 (Statics), and their GPA at the beginning of the semester in which they enrolled in ME222. Both of these markers were used as indicators of student preparation prior to enrollment in ME222. All grades at Michigan State University are numeric in nature (e.g. 4.0, 3.5, 2.0, on a 4.0 scale) rather than alphabetic (e.g. A, B+, C). This allowed for simple ratios to be used to calculate indexes of performance in ME222 in comparison to indicators of student preparation. The first index (prerequisite performance) divided the grade in ME222 by the grade in CE221. The second index (overall academic record) divided the grade in ME222 by the cumulative GPA through the previous semester of coursework.

Several statistical tests were used to explore the performance of the various study populations as measured through differences in ME222 grades, both based on the actual assigned grade and normalized by CE221 or GPA. A one-sample t-test was used to compare a sub population to the overall ME222 student results, while the Welch's t-test was used to understand the significance of differences between population subgroups. A significance level was set as $p < 0.05$ for all tests.

Study Results

Assessing the impact of the SMART method requires a profile of the students who are taking the courses. The students entering the traditional and SMART sections had similar pre-requisite and GPA profiles when comparing Male and Female students. Female students entering the SMART sections had higher GPAs than their male peers (3.45 for all women, 3.40 for all men), but this was not statistically significant. There were noticeable differences between URM and non-

URM students, the most notable being Female URM students. This group of students had significantly lower pre-requisite and GPA scores, as seen in Table 4.

Table 4 – Incoming student comparisons. Statistically significant differences are noted by **.

	Population	CE221 Grade Average	CE221 Grade Std. Dev	CUM GPA Average	CUM GPA Std Dev
SMART	**736**	**3.209**	**0.818**	**3.408**	**0.372**
Female – Non-URM	123	3.107	0.864	3.485	0.363
Female - URM	14	2.750**	0.691	3.123**	0.318
Male – Non-URM	544	3.237	0.813	3.405	0.374
Male - URM	55	3.269	0.751	3.336	0.339
Traditional	**486**	**3.297**	**0.747**	**3.410**	**0.382**
Female – Non-URM	102	3.282	0.733	3.509	0.361
Female - URM	14	3.036**	0.771	3.378	0.416
Male – Non-URM	349	3.319	0.744	3.387	0.378
Male - URM	21	3.175	0.863	3.323	0.462
Grand Total	**1222**	**3.244**	**0.791**	**3.409**	**0.376**

Metrics of student performance in ME222 are presented for the study groups in Table 5, both in terms of their assigned grade and in terms of the ratio of the ME222 grade with measures of student preparation. The only statistically significant differences in student performance were found for the Female URM students in the SMART classroom – when compared to Female students in the SMART sections, non-URM students in the SMART sections, and the overall student population. This is an interesting finding, since both Non-URM women and Male URM students performed better (based on ME222 grade) under the SMART model than their Male Non-URM counterparts (although not at a level that is statistically significant). The very small sample size of the Female URM in this study should be noted here.

Table 5 – Average student performance in ME222 and performance in ME222 normalized by grade in pre-requisite CE221 (preparation index 1) and incoming GPA (preparation index 2).

	Population	**ME222 Grade**	**Preparation Index 1: ME222/CE221 Grade**	**Preparation Index 2: ME222/GPA**
SMART	**736**	**2.646**	**0.843**	**0.762**
Female – Non-URM	123	2.671	0.910	0.750
Female - URM	14	1.857 **	0.622 **	0.577
Male - Non-URM	544	2.659	0.836	0.768
Male - URM	55	2.664	0.816	0.780
Traditional	**486**	**2.667**	**0.832**	**0.772**
Female – Non-URM	102	2.642	0.802	0.738
Female - URM	14	2.429	0.803	0.715
Male – Non-URM	349	2.696	0.847	0.786
Male - URM	21	2.452	0.767	0.736
Grand Total	**1222**	**2.654**	**0.839**	**0.766**

A Welch's t-test (or unpaired variance t-test) was used to assess the significance in differences seen in the performance of groups and subgroups. Welch's t-test allows for comparisons between groups with different variances and is robust against skewness. Results of the Welch's t-test are shown in Table 6. No statistical significance was found for Female/Male or URM/Non-URM comparisons in either the SMART or the Traditional classroom. The subgroup of Female URM students did show statistically significant differences in the SMART method only. These results indicate that closer monitoring and study of the Female URM students is needed.

Table 6 – Welch's t-test results for ME222 grades for the SMART and Traditional methods. The smaller F/F-URM comparison is included in this summary because that subgroup (F-URM) uniquely demonstrated statistically significant differences in measures of student performance in SMART course sections.

	SMART		Traditional	
Groups being compared.	t	p	t	p
Differences between Female and Male students	-0.623	0.534	-0.620	0.536
Differences between URM and Non-URM students	1.01	0.316	1.49	0.144
Differences between Female - Non-URM and Female - URM students	-2.14	0.0488**	-0.854	0.404

A Welch's t-test was also applied to the calculated preparation indices – the ME222 grades normalized by either the grade in CE221 (Preparation Index 1) or cumulative GPA through the prior semester (Preparation Index 2) (Table 7). There are no statistically significant results, indicating that the SMART method benefits were equally shared by students in these groups.

Table 7 – Welch's t-test results for Preparation Indices - ME222 grades normalized by CE221 grade or previous cumulative GPA.

	SMART		Traditional	
Groups being compared.	t	p	t	p
Preparation Index 1: ME222/CE221 - Male vs Female (all ethnicities)	1.21	0.229	-1.33	0.184
Preparation Index 1: ME222/CE221 - URM vs Non-URM (all genders)	1.3	0.196	1.17	0.25
Preparation Index 2: ME222/GPA - Male vs Female (all ethnicities)	-1.21	0.226	-1.77	0.0778
Preparation Index 2: ME222/GPA - URM vs Non-URM (all genders)	0.598	0.551	1.13	0.263

Conclusions

The SMART method has been shown to improve student understanding and problem-solving skills in ME222 [1]. This study confirms that these benefits are generally shared across race/ethnicity and gender. When viewed as larger groups, no statistically significant difference in course performance was seen for men compared to women or underrepresented minorities as compared to non-underrepresented ethnicities.

The data does indicate that the performance of Female URM students should be studied further. While this group had a statistically significant decrease in performance in ME222 when normalized to their performance in the pre-requisite course CE221, their scores were not statistically significant when normalized by their overall GPA prior to entering ME222 (p =

0.138). This discrepancy will be further explored; however, with only 28 students in this category (14 in Traditional sections and 14 in SMART sections), the sample size is quite small.

In conclusion, the study indicates that Male URM students and Female non-URM students perform as well as their peers in courses organized using the SMART method. Further study is warranted to determine if targeted interventions will counter the negative effect seen among Female URM students, which may be impacted by their significant degree of double underrepresentation (2% of course population).

References

1. Ronald C. Averill, Sara Roccabianca and Geoffrey Recktenwald. "A Multi-Instructor Study of Assessment Techniques in Engineering Mechanics Courses." *Conference Proceedings of ASEE Annual Conference & Exposition*. June 16-19, 2019. Tampa (FL).
2. Peter C. Brown, Henry L. Roediger III and Mark A. McDaniel, *Make It Stick: The Science of Successful Learning*, Cambridge, MA: The Belknap Press of Harvard University Press, 2014.
3. James M. Lang, *Small Teaching: Everyday Lessons from the Science of Teaching*, San Francisco, CA: Jossey-Bass, 2016.
4. Ronald C. Averill, Geoffrey Recktenwald, Sara Roccabianca and Ricardo Mejia-Alvarez. "The Need for Holistic Implementation of SMART Assessment." *Conference Proceedings of ASEE North Central Section Conference. March 27-28, 2020.* Morgantown (WV).
5. Ronald C. Averill. "The Seven C's of Solving Engineering Problems." *Conference Proceedings of ASEE Annual Conference & Exposition*. June 16-19, 2019. Tampa (FL).
6. Bangert-Drowns, R. L., Kulik, J. A., and Kulik, C.-L. C. (1991) "Effects of frequent class-room testing," *Journal of Educational Research* 85, 89-99. https://doi.org/10.1080/00220671.1991.10702818
7. Asghari, A., Kadir, R., Elias, H., and Baba, M. (2012) "Test anxiety and its related concepts: A brief review," *GESJ: Education Science and Psychology* 22, 3–8.
8. Zeidner, M. (1998) *Test Anxiety: The State of the Art*. Plenum, New York, NY.
9. Ronald C. Averill, Geoffrey Recktenwald and Sara Roccabianca, "Effect of Assessment Methods on Performance in Mechanics of Materials," *2018 ASEE Annual Conference & Exposition*, Salt Lake City, Utah, June 24-27, 2018.