Measuring Differences in Performance by Varying Formative Assessment Construction Guided by Learning Style Preferences

Dr. Shanon Marie Reckinger, Montana State University

Shanon Reckinger joined the department of Mechanical and Industrial Engineering at Montana State University in Fall 2015. She received her PhD in Mechanical Engineering at the University of Colorado Boulder in August of 2011. Before her position at MSU, she was a Clare Boothe Luce Professor at Fairfield University in the department of Mechanical Engineering for four years. Her research interests include ocean modeling, computational fluid dynamics, fluid dynamics, and numerical methods. Shanon has taught courses in thermodynamics, numerical methods (graduate), fluid dynamics, gas dynamics (graduate), computational fluid dynamics (undergrad/graduate), fundamentals of engineering, and introduction to programming in MATLAB.

Dr. Bryce E. Hughes, Montana State University

Bryce E. Hughes is an Assistant Teaching Professor in Adult and Higher Education at Montana State University, and holds a Ph.D. in Higher Education and Organizational Change from the University of California, Los Angeles, as well as an M.A. in Student Development Administration from Seattle University and a B.S. in General Engineering from Gonzaga University. His research interests include teaching and learning in engineering, STEM education policy, and diversity and equity in STEM.

Measuring differences in performance by varying formative assessment construction guided by learning style preferences

Abstract

In this evidence-based practice paper, the relationship between assessment design guided by learning style preferences and student performance in a programming course is investigated. One of the National Academy of Engineering's 14 Grand Challenges for Engineering is to tailor and differentiate instruction to improve the reliability of learning. A manner in which this differentiation may be accomplished is through attention to the various preferences and styles by which students learn. As such, the purpose of this paper is to present evidence on the effect of formative assessment design on student performance, and whether this effect varies by student learning style. The results from this study can be used by engineering educators to either diversify or personalize their assessment style.

This work is grounded in the Felder-Soloman learning style model, a model that was developed within engineering education and has been validated and widely used within the field. This model categorizes learning styles along four distinct dimensions: perception (sensing versus intuitive), input (visual versus verbal), processing (active versus reflective), and understanding (sequential versus global). Along each of these dimensions, students are categorized as having a mild, moderate, or strong preference in each of these four learning style scales.

This study takes place in a mid-size, public university in the western United States. The sample for this study includes mechanical engineering undergraduate students across four sections of a required programming course in MATLAB, taught by the same instructor. These students were provided the Index of Learning Styles at the beginning of the semester. Students were administered a weekly quiz to assess their ability to write code, but construction of this assessment varies by section to favor different preferences of one of the four Felder-Soloman learning style dimensions. Performance on these quizzes is objectively scored using a standardized rubric. General linear modeling is used to determine if quiz scores differ by quiz construction condition, and if learning style preference interacts with quiz condition to predict performance on each assessment. Findings portray a complex relationship between quiz construction, learning style preference, and assessment performance.

Introduction / Statement of Problem

Colleges and schools of engineering award approximately 22,000 mechanical engineering bachelor's degrees each year, yet only 12% of these degrees are awarded to women (NSF 2015). This percentage increased significantly between 1970 and the mid-1980s, but has remained stagnant since then. Racial diversity also remains a concern; the National Science Foundation reports that, in 2013, 4% of all engineering bachelor's degrees awarded to Black and African American students, and 9% of all engineering bachelor's degrees awarded to Latina/o students (Landivar 2013), demonstrating the persistent underrepresentation of people from these racial and ethnic groups in the field. Very little research has examined the experience of being LGBTQIQ in STEM, though one analysis demonstrated that lesbian, gay, bisexual, and queer students are less likely to be retained in STEM undergraduate programs after four years than their heterosexual peers (Hughes 2016). In summary, engineering, and particularly mechanical engineering, does not reflect the diversity of the nation: the demographic composition of mechanical engineers should resemble the diversity of the population these engineers serve. We will do that best when we are properly representing a broad perspective, and gender, racial, gender identity, and sexual identity are only a few ways to define and assess our representativeness.

One of the challenges facing engineering educators aiming to diversify the field is the pervasiveness of teaching practices that place students from diverse backgrounds at a disadvantage. Faculty in the sciences and engineering who teach introductory courses specifically, but to some extent upper-level disciplinary courses as well, frequently replicate the pedagogies used when they were students, practices that emphasize passive learning (lecture) and competition among students for course grades (norm-reference grading) aimed at "weeding out" underperformers as opposed to cultivating the talents of all students who aspire to engineering careers (Gasiewski, Eagan et al. 2012). The National Academy of Engineering recognized this issue as part of its 14 Grand Challenges for Engineering with the inclusion of tailoring and differentiating instruction to improve the reliability of learning (National Academy of Engineering 2017).

The purpose of this study was to examine the effect of varying the design of formative assessments in a programming course on student academic performance, using the Felder-Soloman learning styles model (Felder and Soloman n.d.) as a heuristic guiding assessment construction. The lead author had observed in her courses that students prefer to learn using a diversity of learning styles, but that assessments were traditionally constructed in a manner that favored a narrow range of styles. Our conjecture is that varying the manner in which these formative assessments were constructed could affect student learning in beneficial ways. Therefore, this led to the following research question: how might the design of formative assessments in a programming class affect student performance? Specifically, does varying assessment design along each dimension of the Felder-Soloman learning style model affect student performance on those assessments? Second, does controlling for individual students' learning style preferences account for this effect, suggesting that students learn better when assessments align with their style preferences? We hypothesize that altering the construction of formative assessments will affect student learning, as measured through performance on the assessment, but that the interaction between individual style preference and assessment design will not be significant given the lack of evidence supporting the relationship between learning style preference and learning.

Background

Several different learning style models have been developed to describe differences among students in terms of effective learning strategies (Dunn and Dunn 1975, Gregorc 1979, Fleming 2001, Kolb and Kolb 2005) in additional to personality type models and cognitive style models. These models have been fairly widely applied in practice, but often demonstrate inconsistent results when tested in research. This work is grounded in the Felder-Soloman learning style model (Felder and Silverman 1988, Felder and Brent 2005, Felder and Soloman n.d.), a model that was developed within engineering education and has been validated and widely used within the field (Van Zwanenberg, Wilkinson et al. 2000, Zywno 2003, Litzinger, Lee et al. 2005, Hawk and Shah 2007, Litzinger, Lee et al. 2007). The Felder and Soloman (n.d.) Index of Learning Styles (ILS) instrument was also chosen because it is web-based and convenient for the students to use. This model categorizes learning styles along four distinct dimensions: perception (sensing versus intuitive), input (visual versus verbal), processing (active versus reflective), and understanding (sequential versus global). Along each of these dimensions, students are categorized as having strong or neutral preference in each of these four learning style scales.

That said, the evidence supporting the use of learning style models to improve student learning through alignment of teaching methods with learning style preferences is inconsistent at best (Pashler, McDaniel et al. 2008). Pashler et al. found in their review that across studies that used an experimental design, most showed no relationship between student performance and learning style-based instruction when controlling for learning styles (Constantinidou and Baker 2002, Massa and Mayer 2006, Cook, Thompson et al. 2009). They concluded that the use of these assessments in practice is likely a waste of resources, though solely based on the "meshing," or alignment of teaching with learning style preference, hypothesis at the center of their review. However, a reason these learning style models are popular among faculty is they provide a heuristic basis for varying teaching methods, such as in-class activities, assignments, and assessments. No research has been conducted to demonstrate increased student performance as a result of application of the Felder and Soloman (n.d.) learning styles model to teaching, but other models have demonstrated evidence of improved performance (Kolb 1984, Brokaw and Merz 2000). One study in particular examined the effect of designing formative assessments to favor different learning styles (Wang, Wang et al. 2006). Both learning style and formative assessment strategy significantly affected student achievement, though, consistent with Pashler et al.'s conclusion, the interaction between these factors was not significant. Additionally, as this study was performed with a web-based middle-school biology course, a gap remains with regard to undergraduate engineering education.

As the primary motivation for this study is increasing the diversity of the engineering graduates that colleges and universities prepare for the workforce, some evidence demonstrates that varying teaching approaches to favor a multitude of learning styles may aid in achieving that particular end. A validation study of the Felder and Soloman (n.d.) ILS found that female engineering students tended to be more sequential, more sensing, and less visual than male engineering students (Litzinger, Lee et al. 2005). Little research has been performed to determine whether learning style preferences differ for college students by race (Claxton and Murrell 1987), although one study demonstrated Native American students preferred visual (vs. verbal) and written (vs. oral) compared to nonnative students (Haukoos and Satterfield 1986). Research at the K-12 level shows that minority populations definitely do have different learning style patterns than the dominant White, American culture (Dunn, Gemake et al. 1990). The number of students of color and women in this study was too small to test these differences directly, and represents an important future direction for this work.

Sample Description

This study takes place in a mid-size, public university in the western United States. The sample for this study includes approximately 123 mechanical engineering undergraduate students across four sections of a required programming course in MATLAB, taught by the same instructor. The effectiveness of the overall course design is discussed in previous work (Reckinger 2014, Reckinger 2016). At its highest, enrollment was at 169 students, however, over the course of the semester 46 students dropped the class. Of all the students who withdrew, 19 of the students withdrew before the first day of class, and the remaining withdrew at a constant rate until the drop deadline. Most students who withdrew did not participate in any or very few in-class activities. Of the 123 students who were enrolled in the class when final grades were posted, an additional five students were removed from the study due to low participation in-class activities. All five of the removed students had a final score in the class of <35%. The full demographics of the class can be found, broken down by section, in Appendix 1. The gender diversity in this course is slightly better than the national average of female's receiving a bachelor's degree in Mechanical Engineering. There are approximately 19.5% females who completed the class and were included in the study, the national average of females who receive bachelor's degrees in Mechanical Engineering is approximately 12% (NSF 2015).

Methods

The research question to answer is: how might the design of formative assessments in a programming class affect student performance? Specifically, does varying assessment design along each dimensions of Felder and Soloman (n.d.) learning style model affect student performance on those assessments? Second, does controlling for individual students' learning style preferences account for this effect, suggesting that students learn better when assessments align with their style preferences?

When interpreting student's learning style preference scores, they were grouped into three categories along each dimension. On each dimension, learners can receive a score of -11, -9, -7, -5, -3, -1, 1, 3, 5, 7, 9, or 11. Range (1) was designated as a strong preference for scores of +11, +9, +7, or +5. Range (3) was designated as a strong preference for scores of -11, -9, -7, or -5. Range (2) was designated as a neutral learner for scores -3, -1, 1, or 3. There were 31 active (1), 69 neutral (2), and 14 reflective (3) learners. There were 79 visual (1), 33 neutral (2), and 2 verbal (3) learners. There were 42 sequential (1), 60 neutral (2), and 12 global (3) learners. For each of the three dimensions, there were three different quiz styles constructed that correspond with the same three Styles (1, 2, and 3). In favor of brevity, this paper will not include the methods of how quizzes were constructed. In summary, Visual/Verbal and Sequential/Global quizzes were written differently. For the Active/Reflective quizzes, students took quizzes with partners under two different conditions and independently. They were not partnered randomly, they were partnered by course grade. They were also partnered in a way that resulted in a good distribution of mixed and matched learning style partner pairs.

This study used a quasi-experimental design, illustrated in Table 1, for several reasons. For each quiz, the quiz construction varied across sections. Additionally, each section was administered all different quiz styles over the course of nine quizzes. According to Pashler, McDaniel et al. (2008), four criteria need to be met in order for the study to be able to show style-based instruction is a significant factor in affecting student performance: (1) Learners must be divided into two or more groups; (2) Subjects within each learning style group must be randomly assigned to one of at least two different learning methods; (3) All subjects must be given the same test of achievement; (4) The results need to show that the learning method optimizes test performance of one learning style group is different than the learning method that optimizes the test performance of a second learning style group. The proposed method varies from this recommendation in three ways. First, students are not completely randomly assigned due to practical limitations of this classroom study. Second, engineering students do not have a huge diversity of learning styles, so some dimensions studied do not have adequate numbers of subjects to test. Third, the study is, by definition, studying the test of achievement. Therefore, all students are taking different formative assessment styles.

Quiz	**Sec. 2 (n=31)**	**Sec. 3 (n=37)**	**Sec. 5 (n=30)**	**Sec. 4 (n=20)**
Quiz 1-Basics	GLO (1G)	SEQ (1S)	GLO(1G2)	GLO+SEQ(1GS)
Quiz 2-Arrays	GLO+SEQ(2GS)	GLO(2G)	SEQ(2S)	SEQ(2S2)
Quiz 3-Plots	**None (V1)-3I**	**None (V2)-3I2**	**None (V3)-3E**	**None (V4)-3E2**
Quiz 4-Logic	VER (4E)	VIS+VER (4IE)	VIS (4I)	VIS (4I2)
Quiz 5-Loops	VIS (5I)	VER (5E)	VIS+VER (5IE)	VIS+VER(5IE2)
Quiz 6-Func	VIS+VER (6IE)	VIS (6I)	VER (6E)	VER (6E2)
Quiz 7-Data	**None (V1)**	**None (V2)**	**None (V3)**	**None (V4)**
Quiz 8-Stats	SEQ (8S)	GLO+SEQ(8GS)	GLO+SEQ(8GS2)	GLO(8G)
Quiz 9-Interp	REF (9R)	ACT (9A)	REF+ACT (9RA)	REF+ACT(9RA2)
Quiz 10-Int/Diff	REF+ACT(10RA)	REF (10R)	ACT (10A)	ACT (10A2)
Quiz 11-LinAlg	ACT (11A)	REF+ACT(11RA)	REF (11R)	REF (11R2)
Quiz 12-Optimiz	**None (V1)**	**None (V2)**	**None (V3)**	**None (V4)**

Table 1 - The experimental design. It is important to point out that the Section 4 lab is a bit of an outlier in all results presented here. This is due to the timing of that section (Friday, 8 am), lower mean GPA compared to other sections, small sample (only 20 students), and it is comprised mostly sophomore students. The demographic table in the Appendix shows this in more detail.

Data Analysis

Statistical analysis was performed using IBM SPSS analysis software. Each set of learning style dimensions was tested separately through a series of analysis that controlled for increasing numbers of covariates to ensure robustness. After computing means for each section on each quiz, the differences among these means were tested using one-way and two-way analyses of variance (ANOVA) to determine if quiz scores varied significantly by the condition under which each quiz was administered. The one-way ANOVA tested the bivariate relationship between quiz scores and condition, to determine if they varied significantly by condition. The two-way ANOVA then used students' learning style preference, categorized into three groups (one style, middle, other style), as a second grouping variable, testing the hypothesis that learning style preference might explain variation in quiz scores by condition.

One limitation to running separate analyses for each quiz, however, is the fact that three quizzes were used to test each learning style dimension. As an alternative to running separate models and trying to account for confounding factors, three repeated-measures ANOVAs were also tested for each of the three learning style conditions utilized for this study. Repeated-measures ANOVA provides a more robust analysis in accounting for dependence among individual student scores across each condition, and accounts for differences in student performance among the three conditions. The limitation to running a repeated-measures ANOVA is that students took different quizzes under different conditions. To overcome this limitation, course section was included as a factor to account for differences among the sections, such as differences in quiz content as well as the inability to randomly assign students to sections. Scores were then treated as repeated measures, and quiz condition as well as learning style group were tested as factors.

ANOVAs were checked for violations of Levene's test for equality of variance, and the appropriate results are reported depending on whether variance was homoscedastic or heteroscedastic. The repeated-measures ANOVAs were checked for sphericity violations, and the appropriate results are reported depending on whether this assumption was violated. F-values, or the test statistic used to determine whether effects in ANOVA models are significant, are reported, along with the associated p-value. The α-level set for significant effects is 0.05, meaning a p-value of 0.05 or less is considered significant, a standard level used within social science research. In a few instances, marginal effects are reported that are close to significance ($p<0.10$). For significant effects, effect sizes (ω^2) were computed for significant effects and are reported; 0.01, 0.06, and 0.14 indicate small, medium, and large effects (Kirk 1996).

Results and Discussion
Visual Verbal Results

Quiz 4. Quiz scores for quiz 4 did not differ by quiz style, $F(2, 106) = 3.029$, $p>0.05$, though the effect was marginally significant ($p=0.053$). No effects were significant when learning style group was included in a two-way ANOVA: for quiz style, $F(2, 103) = 0.249$, $p>0.05$; for learning style group, $F(1, 103) = 0.539$, $p>0.05$; and for the interaction between style and preference, $F(2, 103) = 0.915$, $p>0.05$. This visual quiz construction incorporated a visual description of what the quiz task was, as well as using plots to visualize the results of the task. The verbal style quiz used neither of these visual techniques. One reason the quiz scores on the visual style quiz are not higher may be due to the fact that Sec. 4 took the visual style quiz and this section is an outlier.

Quiz 5. Quiz scores for quiz 5 also did not differ by style, $F(2, 105) = 0.133$, $p>0.05$. However, when learning style group was included in the two-way ANOVA, the interaction between learning style group and quiz style was significant, $F(2, 102) = 5.440$, $p<0.01$, $\omega^2 = 0.08$, and this difference is depicted in the plot below. The effects for quiz style, $F(2, 102) = 0.511$, $p>0.05$, and learning style group, $F(1, 102) = 1.297$, $p>0.05$, were not significant in the second analysis.

Quiz 6. The effect of quiz style on quiz 6 scores was also not significant, $F(2, 105) = 0.271$, $p>0.05$. No effects were significant in the two-way ANOVA also accounting for learning style group: quiz style, $F(2, 102) = 0.306$, $p>0.05$; learning style group, $F(1, 102) = 0.010$, $p>0.05$; and interaction between style and preference, $F(2, 102) = 0.186$, $p>0.05$.

Repeated-measures ANOVA. The results of the repeated-measures ANOVA did not demonstrate a significant effect for quiz condition on quiz score, $F(2, 190) = 2.604$, $p>0.05$. However, both the interaction between section and quiz condition, $F(6, 190) = 4.153$, $p<0.01$, $\omega^2 = 0.02$ (see Figure 1, left), and between learning style preference and quiz condition, $F(2, 190) = 4.170$, $p<0.05$, $\omega^2 = 0.02$ (see Figure 1, right), were significant. One note on Sec. 3 results appearing quite different than the other sections in Figure 1, right. Sec. 3 took the visual style quiz on the most difficult topic of the semester (Quiz 6), it has the lowest mean score in all sections except

Section 2. This likely explains the low performance on the visual quiz for that section. The result in Figure 1 (left) further verifies what was seen in Quiz 5.

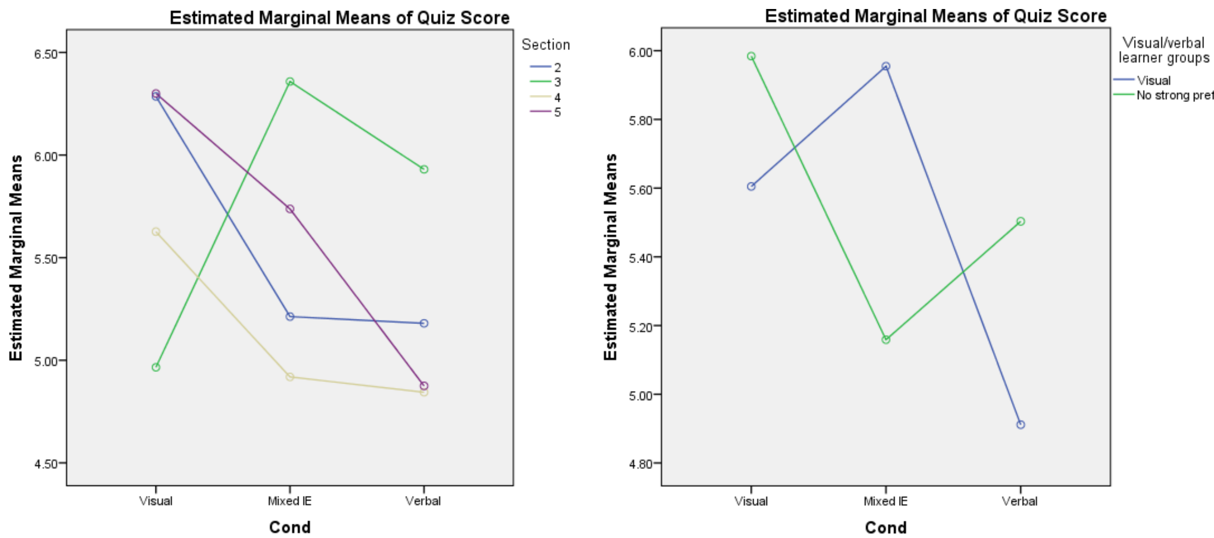

Figure 1- Estimated marginal means for all three quizzes grouped by quiz styles compared by section (left). - Estimated marginal means for all three quizzes grouped by quiz style compared by learning style preference (right).

Although the evidence is fairly inconsistent, students appear to score lower under the verbal condition than the other two conditions, and especially students who had a visual learning style preference. In other words, faculty might consider it worth adding visual items to assessments, as wholly verbal assessments seem to work to visual style students' detriment to some extent. The results do show, though, that the type and number of visual elements do appear to affect student performance. Since this was tested only over the course of three quizzes, another study will need to be conducted to further understand these effects. That said, without adequate numbers of verbal learners, we aren't able to assess if visual elements might hinder verbal student performance. The two verbal learners scored very high on all three quizzes, much higher than the mean (i.e. they scored 8.5 & 10 on visual, 8.5 & 9.5 on mixed, and 10 & 9 on verbal quizzes). This is nice to know that the visual elements on the quizzes are not obviously negatively impacting the most dominantly verbal learners. Other work (Thomas, Ratcliffe et al. 2002) in using the Felder Learning Style model in a programming class categorized students as either visual or verbal learners. They presented results on which type of learner performed better in the class and on the final exam. There results showed that verbal learners scored better. Another study (Allert 2004) on learning style preferences and performance in a computer science class found the same result. The results from this study indicate that perhaps we should stop focusing on what type of students are performing best in our standard way of assessing and start focusing more on constructing our assessment so students can perform best.

Active/Reflective Results

Quiz 9. Quiz scores differed significantly by learning condition, $F(2, 115) = 6.096$, $p<0.01$, $\omega^2 = 0.08$. A Bonferroni post hoc test revealed this difference was due to a significant difference between the mixed active/reflective condition ($M = 6.93$, $SD = 2.33$) and the reflective condition ($M = 5.242$, $SD = 2.08$), $p<0.01$. Students who took quiz 9 under the reflective condition performed worse than students who took the quiz under the mixed condition. Accounting for students' learning style preference, a two-way ANOVA demonstrated that quiz scores no longer differed by learning condition when learning style preference group was included in the analysis, $F(2, 104) = 1.661$, $p>0.05$. The main effect for learning style preference group, $F(2, 104) = 0.476$, $p>0.05$, and interaction between preference and condition, $F(4, 104) = 0.791$, $p>0.05$, were also not significant. When both preference and condition are included in the model, the effect is no longer present, which may suggest that preference helps explain the difference observed for condition.

Quiz 10. Scores on quiz 10 also differed significantly by learning condition, $F(2, 115) = 7.646$, $p<0.01$, $\omega^2 = 0.10$. A Bonferroni post hoc test revealed this difference was due to a significant difference between the reflective condition ($M = 3.45$, $SD = 1.75$) and both the active ($M = 4.55$, $SD = 1.52$), $p<0.05$, and mixed active/reflective conditions ($M = 5.19$, $SD = 2.49$), $p<0.01$. These results were consistent with quiz 9, that students scored lower under the reflective learning condition. Results from a two-way ANOVA demonstrated quiz scores differed significantly by learning condition, $F(2, 100) = 6.197$, $p<0.01$, $\omega^2 = 0.08$, and by learning style preference group, $F(2, 100) = 3.820$, $p<0.05$, $\omega^2 = 0.05$. The interaction effect between condition and preference was not significant, however, $F(4, 100) = 0.931$, $p>0.05$. Bonferroni posthoc tests reveal that students performed significantly different under each of the three conditions. Students performed the highest under the mixed condition, lowest under the reflective condition, and in-between on the active condition. Additionally, reflective learners ($M = 5.58$, $SD = 1.41$) performed higher than active learners ($M = 4.19$, $SD = 1.23$), $p<0.05$, on quiz 10.

Quiz 11. The score on Quiz 11 also differed significantly by learning condition, $F(2, 115) = 3.987$, $p<0.05$, $\omega^2 = 0.05$. A Bonferroni post hoc test revealed this difference was attributable to a significant difference between the active ($M = 7.26$, $SD = 2.59$) and reflective conditions ($M = 5.76$, $SD = 2.36$), $p<0.05$. Although the comparison group was different, students who took the quiz under the reflective condition scored lower than at least one other group for all three quizzes.

A two-way ANOVA revealed quiz scores continued to differ significantly by learning condition, $F(2, 100) = 7.171$, $p<0.01$, $\omega^2 = 0.10$, with learning style preference included. The main effect for learning style preference, $F(2, 100) = 0.925$, $p>0.05$, and the interaction effect between preference and condition, $F(4, 100) = 1.140$, $p>0.05$, were not significant. A Bonferroni posthoc test confirmed the one-way ANOVA results: students who took the quiz under the reflective condition scored lower than students who took the quiz under the active condition, $p<0.001$. The posthoc test also revealed that, accounting for learning style preference group, students who took the quiz under the reflective condition also scored lower than students who took the quiz under the mixed condition ($M = 7.12$, $SD = 1.75$), $p<0.05$.

Repeated-measures ANOVA. The repeated-measures ANOVA then demonstrated two significant effects, including the main effect for learning condition, $F(2, 184) = 10.957$, $p<0.001$, $\omega^2 = 0.02$, and the interaction effect between learning condition and course section, $F(6, 184) = 15.553$, $p<0.001$, $\omega^2 = 0.03$. A Bonferroni posthoc test confirmed what had been observed in the previous individual quiz analyses; students performed lower under the reflective condition ($M = 5.26$, $SD = 0.27$) than under either the active ($M = 6.17$, $SD = 0.22$), $p<0.01$, or the mixed condition ($M = 6.42$, $SD = 0.28$), $p<0.001$. Sections scored highest under the mixed condition with the exception of section 2. Section 2 scored lowest under the mixed condition and highest under the active condition. Section 5 also scored unusually low on the active quiz. That quiz had an unintentionally tricky part in the quiz that seemed to affect partners more than individuals.

Sequential/Global Results

Quiz 1. According to the results of the one-way ANOVA, scores on quiz 1 did not differ by quiz style, $F(2, 115) = 2.870$, $p>0.05$, though the effect was marginally significant ($p=0.61$). None of the effects were significant in the two-way ANOVA between quiz style and learning style group either: quiz style, $F(2, 105) = 1.241$, $p>0.05$; style group, $F(2, 105) = 0.868$, $p>0.05$; and the interaction between style and group, $F(4, 105) = 0.447$, $p>0.05$. It must be noted that the two-way ANOVA violated the assumption of homogeneity of variance, $F(8, 105) = 4.446$, $p<0.001$, but given the lack of significance of any effects in the model, the more robust tests recommended for this assumption violation would not have provided different results. Since this was the very first quiz of the semester, with a very high mean across all sections, this is likely the cause of no significance. The level of difficult on this quiz may have precluded any observational difference by learning style.

Quiz 2. The one-way ANOVA demonstrated scores on quiz 2 did not differ by quiz style, $F(2, 115) = 0.968$, $p>0.05$. No significant effects were determined in the two-way ANOVA either: quiz style, $F(2, 105) = 0.694$, $p>0.05$; learning style group, $F(2, 105) = 1.063$, $p>0.05$; and the interaction between style and preference, $F(4, 105) = 1.835$, $p>0.05$.

Quiz 8. Results from the one-way ANOVA showed scores on quiz 8 did not differ by quiz style, $F(2, 115) = 0.263$, $p>0.05$. Results from the two-way ANOVA demonstrated one significant effect, however: quiz scores differed by learning style group, $F(2, 100) = 3.171$, $p<0.05$, $\omega^2 = 0.04$. Results from a Bonferroni posthoc test revealed that this difference is attributed to a difference between sequential ($M = 7.03$, $SD = 1.51$) and neutral learners ($M = 6.33$, $SD = 1.71$). No effect was found for quiz style, $F(2, 100) = 0.159$, $p>0.05$, or the interaction between style and preference, $F(4, 100) = 1.875$, $p>0.05$.

Repeated-measures ANOVA. The results of the repeated-measures ANOVA testing student scores across demonstrated one significant effect. The interaction between condition and section was significant, $F(6, 190) = 6.256$, $p<0.001$, omega. The effects for condition, $F(2, 190) = 1.625$, $p>0.05$, the interaction between condition and learning style group, $F(4, 190) = 0.863$, $p>0.05$, and the three-way interaction among condition, group, and section, $F(12, 190) = 1.677$, $p>0.05$, were not significant.

Conclusions

The dominant learning style in the visual-verbal dimension among undergraduate engineering students is the visual dimension. Programming, in general, but programming assessment, in particular, is strongly verbally focused. The results in this work show that student performance on assessment will likely significantly improve if quiz construction is more visually focused. It is recommended that assessment design works to achieve visual descriptions or explanations of what is being asked for, but also guides students to solve the problem by requiring them to visualize their results. In programming, this can be most easily achieved by incorporating a plot into the solution of the problem, even when it does not appear necessary to understand the problem or algorithm. The results from this study indicate that it is possible to assess students on exactly the same concepts, but construct the assessment to better suit visual learners. This result likely extends to assessment throughout other engineering courses.

The dominant learning style in the active-reflective dimension among engineering students is the active dimension. In general, assessment is rarely active in almost any discipline, but particularly in engineering. This study shows that all students perform better on assessment when they are allowed to collaborate and actively discuss during the assessment period. While student performance on the assessment is highest when the students are given to freedom to choose to collaborate or not, observationally, student learning appears to be highest when students are required to collaborate on the assessment. It is important to point out that partner pairing certainly cannot be random. And further work is being done to understand if successful partner pairing can be predicted by properly mixing and matching learning style preferences among the pair. From the instructor's experience, pairing must always at least be based on GPA or course grade, however, this has not been quantitatively tested.

The dominant learning style in the sequential-global dimension among engineering students is the sequential dimension. For this course, the assessment has been traditionally written globally. Gradually, the instructor started designing the assessment in a more sequential manner and observationally it appeared to improve student performance and students indicated that they found it clearer and preferred it. This study was designed to quantitatively determine if this design did in fact improve student performance. The results were mostly inconclusive due to data limitations. The three quizzes that were selected to vary the sequential-global style were not consistent enough in level of difficulty. Survey results indicate that students prefer the sequential style and find it most clear. However, there were many interesting student comments that indicated that students found it easier to score higher on the sequential style quiz, but thought they learned more by struggling through the global quiz. Therefore, it could be that global style quizzes might be recommended for higher performing students and better learners, and sequential style quizzes might be recommended for lower performing students and novice learners.

Previous work has focused on categorizing learners who are most typically successful in an existing learning environments. In particular, there has been evidence that reflective and verbal learners are particularly successful academically. This work shows that those learners are still successful, but by adjusting assessment in lieu of the opposite dimensions (e.g. active and visual learners), all student performance improves. A final evaluation of all students' performance in this class showed that there was no significant difference in this class among any single learning style preference. In other words, reflective learners did not perform significantly better in the final course grade than active learners (same for the other two dimensions, as well).

Acknowledgements

First, thank you to the seven undergraduate teaching assistants who helped run this semester's course: Daniel Moore, Seth Whiteside, Lincoln DiLorenzo, Jacob Perry, Jaclyn Wing, Connor

Klein, and Dayton Cooney. We would like to acknowledge their enormous efforts in this study. These TAs took all the quizzes prior to giving them out in lab, and provided their feedback on overall clarity, consistency, and level of difficulty. They also helped proctor all the quizzes and did all the grading based on the standardized rubric. We would like to thank the TAs for all the data entry and record keeping that allowed us to analyze this work on so many dimensions. Secondly, we would like to thank the 123 EMEC 203 students who gave their time and effort to this class. They went above and beyond to provide additional feedback through the many surveys, outside of the class requirements. Lastly, we would like to thank Montana State University and the department of Mechanical and Industrial Engineering for providing the resources needed to complete this study.

Appendix

APPENDIX 1 – Demographic data from course by section.

Characteristics	All (%) (n=118)	Sec. 2 (%) (n=31)	Sec. 3 (%) (n=37)	Sec. 4 (%) (n=20)	Sec. 5 (%) (n=30)
Gender					
Male	80.5	80.6	78.4	85.0	80.0
Female	19.5	19.4	21.6	15.0	20.0
Race					
Caucasian	83.9	90.3	73.0	85.0	90.0
Asian	6.8	0	10.8	10.0	6.7
American Indian	3.4	3.2	2.7	5.0	3.3
Hispanic	2.5	0	8.1	0	0
Black/African Amer.	0	0	0	0	0
Other/No Response	3.4	6.5	5.4	0	0
Class					
Sophomore	34.7	12.9	16.2	80.0	50.0
Junior	46.6	71.0	54.1	20.0	30.0
Senior	16.9	9.7	29.7	0	20.0
Cumulative GPA (4.00)					
Mean	3.25	3.34	3.21	3.12	3.30
Residency					
Montana	39.8	25.8	48.6	40.0	43.3
Nearby States	38.1	35.5	32.4	20.0	40.0
Beyond	22.0	38.7	18.9	20.0	16.7
Age					
18-22	91.5	87.1	89.2	100.0	93.3
23-28	5.9	9.7	8.1	0	3.3
>28	2.5	3.2	2.7	0	3.3
Previous programming experience					
None	62.7	64.5	62.2	75.0	53.3
High School Class	12.7	19.4	8.1	15.0	10.0
College Class	9.3	3.2	10.8	0	20.0
College MATLAB	0.8	0	2.7	0	0
Significant Exp.	1.7	3.2	2.7	0	0
Overall Affect towards programming					
Negative	5.1	3.2	8.1	10.0	0
Nervous but hopeful	22.9	22.5	29.7	5.0	26.7
Neutral	7.6	6.5	5.4	15.0	6.7
Practical skill	38.1	41.9	32.4	55.0	30.0
Excited	16.9	19.4	13.5	10.0	23.3
Likeliness to go into computational field/specialty					
Never	8.5	9.7	5.4	15.0	6.7
Probably not	24.6	3.5	21.6	30.0	13.3
Don’t know	28.8	22.6	27.0	35.0	33.3
Maybe	24.6	22.6	32.4	10.0	26.7
Definitely	2.5	3.2	2.7	0	3.3

References

Allert, J. (2004). Learning style and factors contributing to success in an introductory computer science course. Advanced Learning Technologies, 2004. Proceedings. IEEE International Conference on, IEEE.

Brokaw, A. J. and T. E. Merz (2000). "The effects of student behavior and preferred learning style on performance." Journal of Business Education **1**(2): 44-53.

Claxton, C. S. and P. H. Murrell (1987). Learning Styles: Implications for Improving Educational Practices. ASHE-ERIC Higher Education Report No. 4, 1987, ERIC.

Constantinidou, F. and S. Baker (2002). "Stimulus modality and verbal learning performance in normal aging." Brain and language **82**(3): 296-311.

Cook, D. A., et al. (2009). "Lack of interaction between sensing–intuitive learning styles and problem-first versus information-first instruction: A randomized crossover trial." Advances in Health Sciences Education **14**(1): 79-90.

Dunn, R. and K. Dunn (1975). "Learning styles, teaching styles." NASSP Bulletin **59**(393): 37-49.

Dunn, R., et al. (1990). "Cross ‐ Cultural Differences in Learning Styles of Elementary ‐ Age Students From Four Ethnic Backgrounds." Journal of Multicultural Counseling and Development **18**(2): 68-93.

Felder, R. M. and R. Brent (2005). "Understanding student differences." Journal of engineering education **94**(1): 57-72.

Felder, R. M. and L. K. Silverman (1988). "Learning and teaching styles in engineering education." Engineering education **78**(7): 674-681.

Felder, R. M. and B. A. Soloman (n.d.). "Index of Learning Styles." Retrieved from http://www.ncsu.edu/felder-public/ILSpage.html.

Felder, R. M. and B. A. Soloman (n.d.). "Learning styles and strategies." Retrieved from http://www.ncsu.edu/felder-public/ILSdir/styles.pdf.

Fleming, N. D. (2001). Teaching and learning styles: VARK strategies, IGI Global.

Gasiewski, J. A., et al. (2012). "From Gatekeeping to Engagement: A Multicontextual, Mixed Method Study of Student Academic Engagement in Introductory STEM Courses." Research in Higher Education **53**(2): 229-261.

Gregorc, A. F. (1979). "Learning/teaching styles: Their nature and effects." Student learning styles: Diagnosing and prescribing programs: 19-26.

Haukoos, G. D. and R. Satterfield (1986). "Learning styles of minority students (Native Americans) and their application in developing a culturally sensitive science classroom." Community/Junior College Quarterly **10**(3): 193-201.

Hawk, T. F. and A. J. Shah (2007). "Using learning style instruments to enhance student learning." Decision Sciences Journal of Innovative Education **5**(1): 1-19.

Hughes, B. E. (2016). Coming Out in STEM: Factors Affecting Retention of Sexual Minority STEM Students. Association for the Study of Higher Education. Columbus, OH.

Kirk, R. E. (1996). "Practical significance: A concept whose time has come." Educational and psychological measurement **56**(5): 746-759.

Kolb, A. Y. and D. A. Kolb (2005). "Learning styles and learning spaces: Enhancing experiential learning in higher education." Academy of management learning & education **4**(2): 193-212.

Kolb, D. A. (1984). Experiential learning: Experience as the source of learningand development, Englewood Cliffs, NJ: Prentice-Hall.

Landivar, L. C. (2013). "Disparities in STEM employment by sex, race, and Hispanic origin." Education Review **29**(6): 911-922.

Litzinger, T. A., et al. (2005). A study of the reliability and validity of the Felder-Soloman Index of Learning Styles. Proceedings of the 2005 American Society for Education Annual Conference & Exposition.

Litzinger, T. A., et al. (2007). "A psychometric study of the index of learning styles©." Journal of engineering education **96**(4): 309.

Massa, L. J. and R. E. Mayer (2006). "Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style?" Learning and Individual Differences **16**(4): 321-335.

National Academy of Engineering, N. (2017). "Grand Challenges in Engineering."

NSF (2015). Science and Engineering Degrees: 1966–2012. Detailed Statistical Tables NSF 15-326. National Center for Science and Engineering Statistics. Arlington, VA.

Pashler, H., et al. (2008). "Learning styles concepts and evidence." Psychological science in the public interest **9**(3): 105-119.

Reckinger, S. M. (2016). Implementation and Evaluation of Different Types of Peer Learning Instruction in a MATLAB Programming Course. 2016 ASEE Annual Conference & Exposition, June 26-29, 2016, New Orleans, LA.

Reckinger, S. M., Reckinger, Scott James (2014). An Interactive Programming Course Model for Mechanical Engineering Students. American Society of Engineering Education Annual Conference, June 15-18, 2014, Indianapolis, IN.

Thomas, L., et al. (2002). Learning styles and performance in the introductory programming sequence. ACM SIGCSE Bulletin, ACM.

Van Zwanenberg, N., et al. (2000). "Felder and Silverman's Index of Learning Styles and Honey and Mumford's Learning Styles Questionnaire: how do they compare and do they predict academic performance?" Educational Psychology **20**(3): 365-380.

Wang, K. H., et al. (2006). "Learning styles and formative assessment strategy: enhancing student achievement in Web‐based learning." Journal of Computer Assisted Learning **22**(3): 207-217.

Zywno, M. S. (2003). A contribution to validation of score meaning for Felder-Soloman's index of learning styles. Proceedings of the 2003 American Society for Engineering Education annual conference & exposition, American Society for Engineering Education Washington, DC.