The Constructs Behind the Clicks

John Whitmer, EdD, Sweet San Pedro, PhD, Ruitao Liu, PhD
Kate E. Walton, PhD, Joann L. Moore, PhD, and Alejandro Andrade Lotero, PhD

ACT Research Report | 2019-6
ABOUT THE AUTHORS

John Whitmer, EdD
John is a senior director for Data Science and Analytics at ACT. In this role, he leads a team of data scientists, psychometricians, and learning analytics researchers to conduct foundational research and to develop operational algorithms.

Maria Ofelia Z. San Pedro, PhD
Maria Ofelia “Sweet” San Pedro is a research scientist in the Learning, Assessment and Navigation Experiences group at ACT specializing in learning science and learning analytics. She researches on K-12 learning experiences, assessment and instruction within digital learning environments.

Ruitao Liu, PhD
Ruitao Liu is a senior research scientist in Research Technology, Data Science, and Analytics specializing in Bayesian statistics, predictive modeling, and game-based assessment and learning.

Kate E. Walton, PhD
Kate Walton is a principal research scientist in ACT’s Center for Social, Emotional, and Academic Learning. Prior to joining ACT in 2017, she was Associate Professor of Psychology at St. John’s University. She is interested in social and emotional skill assessment and development.

Joann L. Moore, PhD
Joann Moore is a senior research scientist in Validity and Efficacy Research specializing in prediction of secondary and postsecondary outcomes from academic and non-cognitive factors.

Alejandro Andrade Lotero, PhD
Alejandro Andrade Lotero is a research scientist II at ACTNext. Alejandro holds a PhD in Learning Sciences and Inquiry Methodology from Indiana University and a master’s in education from PUJ, Bogota, Colombia. His research has focused on modeling and supporting embodied learning using multimodal analytics as well using natural language processing tools to automatically represent the way students collaborate and learn together.

ACKNOWLEDGEMENTS
This project was facilitated through a partnership between ACT, the University of Maryland, Baltimore County, Blackboard and Vitalsource, Inc. The research analysis and findings were supported by a Research for Mission-Driven Innovation (RMDI) grant from ACT.
## Summary

Key Findings

## Introduction

Research Questions

## Study Context

Student Data Privacy and Ethical Review

## Data Sources

- Course Syllabus and Faculty Interviews
- ACT Tessera Social and Emotional Skill Assessment
- Blackboard Data
- Student Information System

## Method and Research Approach

- Computing LMS Interaction Features and Sequential Data Mining
- LMS Activity Measures
- Activity Sequence Features

## Key Findings

- Key Finding 1: Some Tessera SE skills were systematically observed in student use of the LMS
- Key Finding 2: High-level learning tactics could be extracted using sequential data mining techniques
- Key Finding 3: Predictions using LMS behavioral data were more accurate than those made using student family and educational background, providing an accurate and powerful basis for actionable interventions
- Key Finding 4: LMS behavioral data partially mediated the relationship between Grit and course grades
- Key Finding 5: No evidence of demographic bias was found in predictive models

## Implications

## Limitations and Future Work
Summary

We analyzed data from student use of the Learning Management System (LMS) in a large-enrollment undergraduate chemistry course and examined the relationships among social and emotional (SE) skills, as measured by ACT® Tessera®, course features, and course grades. Our goal was to understand how SE skills are associated with students’ online behaviors and course outcomes and to explore a principled design approach for feature extraction using individual activities and sequential data mining techniques. We found low to moderate associations (r = 0.10 - 0.34) between these features and SE skills (Grit, Teamwork, Resilience, Curiosity, and Leadership) and between these features and course grade (r = 0.18 - 0.36). These behavioral features (i.e., individual activities and learning tactics) were used to create predictive models that accurately predicted SE skills (root mean squared error [RMSE] = 0.87) and course grade (RMSE = 0.67). These models had higher predictive accuracies than baseline models created from only student background information (including college grade point average [GPA]). This finding suggests a new source of insights that can be used to create accurate and actionable predictions to improve student course grades and SE skills. Furthermore, a mediation analysis revealed a significant partial mediation effect of student online behaviors on the influence of SE skills in final course grade. This suggests that SE skills may have some influence on the daily behaviors of students, which in turn are related to class outcomes and student success. No bias based on student demographic background was observed in our predictive models.

Key Findings

1. Some Tessera SE skills were systematically observed in student use of the LMS
2. High-level learning tactics could be extracted using sequential data mining techniques
3. Predictions using LMS behavioral data were more accurate than those made using student family and educational background, providing an accurate and powerful basis for actionable interventions
4. LMS behavioral data partially mediated the relationship between Grit and course grades
5. No evidence of demographic bias was found in predictive models
Introduction

Educational technologies are becoming an increasingly important part of student learning experiences. These systems record detailed information about student interactions with learning resources and activities. As these technologies are increasingly adopted in classrooms and used for a broader variety of activities, the data collected by these systems provide an increasingly valuable potential data source for insights into student learning practices. For more than a decade, learning analytics researchers and educational technologists have demonstrated that these data can be used to make accurate predictive models of student course grade (Long & Siemens, 2011; Macfayden & Dawson, 2010; Peled & Rashty, 1999). However, the psychological constructs that underlie student online behaviors, which are needed to help faculty, advisors, and students interpret predictions and use them to improve learning outcomes, are not well understood. This limitation has been recognized in the field and has become known as the “clicks to constructs” problem that should be addressed for learning analytics to achieve the stated goal of improving student learning outcomes (Knight & Shum, 2017).

On the other hand, educational psychologists have found that self-reported SE skills have significant relationships with course grades, but there is little known about how these skills are enacted in authentic learning contexts. Several SE skills have been identified in the literature as relevant for success in school, work, and life (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). There is ample evidence, for example, that Grit-related skills are associated with academic success at all levels of education and are even on par with cognitive ability in predicting grades in secondary and tertiary education levels (Poropat, 2009). What we do not understand are the mechanisms by which positive SE skills lead to improved school outcomes. For example, we would expect that having attention to detail, being organized, and goal striving (i.e., being high on Grit) would lead to better course grades through better study habits. That is, individuals who are higher on Grit may be more likely to make an effective use of LMS features, which in turn, results in better course grades. Thus, we hypothesize that LMS feature usage mediates the relationship between Grit and course grades. This hypothesis, however, has not previously been evaluated empirically.
Research Questions

We sought to answer the following questions by employing principled design, conventional and sequential data mining, and our assessment of SE skills:

1. Can we observe SE skills from behaviors recorded in online learning environments?
2. Can high-level learning tactics be extracted using sequential data mining techniques?
3. Are predictions using LMS behavioral data more accurate than those made using student demographic data?
4. How do student online behaviors mediate the relationship between ACT Tessera SE skills and course grades?
5. Is there evidence of demographic bias in predictive models?

Study Context

We investigated the Spring 2019 second semester offering of an introduction to chemistry series at the University of Maryland, Baltimore County (UMBC). UMBC is a mid-sized, public research university accredited by the Middle States Association of Colleges and Secondary Schools and is one of the 11 campuses comprising the University System of Maryland. The fall 2012 six-year graduation rate for first-time, full-time freshmen was 68.2%, and the freshman-to-sophomore retention rate was 87.3%. The university is actively conducting research and deploying predictive analytics solutions intended to improve graduation and retention rates.

This course typically enrolls 700 students and is the second largest course on campus, second only to the first semester course in the chemistry series. This course is taught in a face-to-face modality with extensive use of online course components. Given its size and rigor, the instructor, a senior lecturer in the department of chemistry and biochemistry, has a history of pedagogical innovations designed to engage students and improve outcomes. These include using personal response devices (e.g., “clickers”) to engage students, assigning online homework, offering in-person and online office hours, and helping to establish the Chemistry Discovery Center, which is an active learning supplement to the traditional lecture recitation section. As a result of these innovations, the instructor’s students’ average use of the Blackboard Learn LMS, which UMBC has frequently found to be positively correlated with better outcomes, is among the highest on campus. This course was offered using the
Blackboard "original experience," and it was hosted on the web-based "Ultra" platform.

**Student Data Privacy and Ethical Review**

Institutional Review Board evaluation was conducted prior to beginning this study. All student data were anonymized by UMBC using an identifier generated for this study prior to providing the data to researchers. No personally identifiable information was included in the dataset.

UMBC’s data privacy policies, which are disclosed to students and posted in the student portal, allow for the analysis of learning activity data for the purpose of educational improvement. The instructor introduced the study to the class, and interested students had the option to take ACT Tessera. Participating students were given extra credit points for the course and were entered in a drawing for a gift card. As a result, all students’ recorded LMS usage data were included in the analyses, and for analysis of SE skills, a subset of students (83%) who completed ACT Tessera were included.

**Data Sources**

We used multiple data sources to inform the analyses and modeling.

**Course Syllabus and Faculty Interviews**

Initially, the course syllabus was analyzed, and relevant dates and learning activities were extracted. The faculty member was then interviewed through a set of structured questions around her use of the LMS for the course activities and her intuitions around what student behaviors lead to more or less successful outcomes. This information was used to create a preliminary map of course activity features that could be extracted from the LMS data. Once the LMS data were received, the results were discussed with the faculty member and academic technology administrator to ensure that the mappings were accurate.

**ACT Tessera Social and Emotional Skill Assessment**

ACT Tessera is a standardized assessment that measures five SE skills, which are listed and described in Table 1. Students’ skills are assessed via three methods – Likert items, forced choice items, and situational judgment tests. This multi-method approach avoids the pitfalls associated with any single item type (Kenny & Kashy, 1992), such as the ease with which test-takers can fake their responses to Likert items (Zickar, Gibby, & Robie, 2004). Responses from these three item types are then aggregated into a multidimensional IRT score (Anguiano-Carrasco, Walton, Murano, Burrus, & Way, 2018). The college version of ACT Tessera has been piloted to assess its psychometric properties. Prior validation evidence for ACT Tessera College is limited to self-report data,
including self-reported SE skills, high school GPA, and college GPA. Additional validity evidence in the form of objective data is necessary. Moreover, it would be beneficial to fully flesh out the associations among the ACT Tessera skills, concrete study behaviors, and course grades to understand the mechanisms by which ACT Tessera skills are associated with academic performance. ACT Tessera is a formative assessment system, and if we can recognize skills that require further development – either through use of the ACT Tessera assessment or though large-scale LMS data analysis – we can intervene with students and encourage them to complete the entirety, or portions, of the ACT Tessera College Playbook, which is the SE learning curriculum tied to the assessment.

Table 1. Descriptions of SE Skills Assessed

<table>
<thead>
<tr>
<th>Skill</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grit</td>
<td>Reflects the extent to which a student’s actions demonstrate persistence,</td>
</tr>
<tr>
<td></td>
<td>goal striving, reliability, dependability, and attention to detail at school.</td>
</tr>
<tr>
<td>Teamwork</td>
<td>Reflects the extent to which a student’s actions demonstrate collaboration,</td>
</tr>
<tr>
<td></td>
<td>empathy, helpfulness, trust, and trustworthiness.</td>
</tr>
<tr>
<td>Resilience</td>
<td>Reflects the extent to which a student’s actions demonstrate stress</td>
</tr>
<tr>
<td></td>
<td>management, emotional regulation, a positive response to setbacks, and</td>
</tr>
<tr>
<td></td>
<td>poise.</td>
</tr>
<tr>
<td>Curiosity</td>
<td>Reflects the extent to which a student’s actions demonstrate creativity,</td>
</tr>
<tr>
<td></td>
<td>inquisitiveness, flexibility, open-mindedness, and embracing diversity.</td>
</tr>
<tr>
<td>Leadership</td>
<td>Reflects the extent to which a student’s actions demonstrate assertiveness,</td>
</tr>
<tr>
<td></td>
<td>influence, optimism, and enthusiasm.</td>
</tr>
</tbody>
</table>

Student score distributions for the assessment and the final course grade are provided in Table 2. Ranges are slightly different between the constructs but largely fall within a -3 to 3 range. The grade distribution is smaller and is skewed toward higher grade values, typical of grading for an undergraduate course.
Table 2. Descriptive Statistics for Course Grades and SE Skills

<table>
<thead>
<tr>
<th>Variable</th>
<th>N*</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Course Grades**</td>
<td>489</td>
<td>2.78</td>
<td>3.00</td>
<td>1.07</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td><strong>SE Skill Scores</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grit</td>
<td>406</td>
<td>-0.01</td>
<td>-0.09</td>
<td>0.95</td>
<td>-2.65</td>
<td>2.58</td>
</tr>
<tr>
<td>Teamwork</td>
<td>406</td>
<td>-0.19</td>
<td>-0.19</td>
<td>0.93</td>
<td>-2.61</td>
<td>2.72</td>
</tr>
<tr>
<td>Resilience</td>
<td>406</td>
<td>-0.07</td>
<td>-0.07</td>
<td>0.88</td>
<td>-2.48</td>
<td>2.88</td>
</tr>
<tr>
<td>Curiosity</td>
<td>406</td>
<td>-0.22</td>
<td>-0.24</td>
<td>0.92</td>
<td>-2.88</td>
<td>3.45</td>
</tr>
<tr>
<td>Leadership</td>
<td>406</td>
<td>-0.18</td>
<td>-0.21</td>
<td>0.89</td>
<td>-2.64</td>
<td>3.43</td>
</tr>
</tbody>
</table>

*Difference in count due to opt-in student participation in SE skill assessment and students with missing grades

**Course grades: 0 = F, 4 = A

Blackboard Data

Our quantitative analyses were conducted on the activity log data extracted from the Blackboard Learn LMS using the native database telemetry. Events were provided at the individual student-activity level for all interactions with course materials and activities that were recorded in the database. Blackboard extracted this information from the hosted version of the LMS and provided it at the individual learner-event level, with a total of approximately 700,000 records included in the dataset.

Of note is that we only included LMS behavioral data from the activity table; in-course graded items were excluded as these items were deemed to not be independent from final course grade. As our goal was to identify predictive relationships that could be used to improve student course grade, these were eliminated as potentially misleading variables.

Student Information System

Student demographic background and prior educational experience were extracted from the Peoplesoft student information system. We included data identified as important in prior research or by UMBC. These variables included race / ethnicity, gender, first generation college student status, current college GPA, credits attempted, credits earned, ACT / SAT scores, high school GPA, and other factors. Demographic data are provided in Table 3. The sample was diverse with a substantial number of students meeting potential underserved criteria that would place them at higher risk of failing this course.
Table 3. Sample Demographics (N = 489)

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>242</td>
<td>50%</td>
</tr>
<tr>
<td>Race / Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td>141</td>
<td>29%</td>
</tr>
<tr>
<td>Black / African American</td>
<td>84</td>
<td>17%</td>
</tr>
<tr>
<td>Hispanic / Latino</td>
<td>43</td>
<td>9%</td>
</tr>
<tr>
<td>White</td>
<td>189</td>
<td>39%</td>
</tr>
<tr>
<td>Two or More</td>
<td>29</td>
<td>6%</td>
</tr>
<tr>
<td>Not specified</td>
<td>3</td>
<td>1%</td>
</tr>
<tr>
<td>First Generation College Student</td>
<td>119</td>
<td>24%</td>
</tr>
<tr>
<td>Transfer Student</td>
<td>53</td>
<td>11%</td>
</tr>
<tr>
<td>ACT Tessera Participant</td>
<td>406</td>
<td>83%</td>
</tr>
</tbody>
</table>

Method and Research Approach

Computing LMS Interaction Features and Sequential Data Mining

The online interaction features were generated from the LMS clickstream data. After manual inspection, we determined that the action field alone (e.g., “opened”) was insufficient to address our research questions and needed to be joined with the label of the item that the action was taken in reference to, which was a complex pairing. For example, course item values in the LMS data include “Week 12: Electrochemistry” or “CHEM 102 Practice Exam 4B,” which were easily interpretable from the course syllabus, while others (e.g., “2/27 CL” or “18.7 RQ”) required confirmation from the instructor. Hence, we created broader activity categories for these activity events in the LMS data using the course syllabus with confirmation from the instructor which resulted in 110 unique activity events in the LMS data that were recoded to a total of 21 activity categories as described in Table 4.
The initial design pattern was developed based on evidence-centered design principles to help us understand the LMS data and create indicators (DiCerbo, 2015) that best described the SE skills. However, the information captured in the LMS event log proved to be inadequate in generating the needed interaction features. Therefore, from the reformatted LMS data, we were able to compute student-level interaction features that described access to these activities (LMS activity features) and patterns of activity sequences indicative of a learning tactic.
(activity sequence features) within a predefined time interval (LMS data for the entire semester were used by default; predictive models were generated both using data for the entire semester as well as week by week, as discussed below). These were then merged to the individual students’ outcome information of final course grade, SE skill scores, and background information. This process required extensive manual interpretation and recoding.

**LMS Activity Measures**

From the LMS log data, relevant information about a student accessing a particular activity included the session ID, the duration, and number of instances of that activity. Therefore, for each activity category, we computed its average measures of engagement across all of a student’s recorded sessions. These calculations included three average measures:

- Average count or number of attempts of an activity in a session (activity appended with “Attempt”)
- Average time spent or total duration for an activity in a session (activity appended with “TimeSpent”)
- Average variability of time spent or total duration for an activity in a session (activity appended with “TimeSpentSD”)

We accounted for the number of sessions when looking at counts and duration of an activity, instead of count and sum across the entire semester, to have a more nuanced measure of activity engagement. For example, throughout the semester, a student accessed an assignment activity in the LMS across three login sessions, and for each session the student accessed it twice. Thus, the feature assignment_Attempt for this student would be equal to 2. If we included the number of sessions a student had as an interaction feature, there were a total of 64 unique LMS activity features across all students who took the chemistry course.

**Activity Sequence Features**

To discover learning behavior patterns, we sought to categorize student activity into frequent patterns that included all activities they conducted within a login session. We grouped sessions of activity sequences into clusters using a method based on finite mixtures of Markov transition matrices. The method assumes that each activity sequence in a session is generated from one of the K transition probability matrices (TPM). Each entry in a TPM is a probability of transitioning from one activity to another in a single step. We can generate a sequence by first randomly selecting an initial activity and a TPM according to a mixture probability distribution. Then, the rest of the activities in the sequence can be generated
iteratively using the chosen TPM. We set the number of clusters, K, equal to five, according to the Bayesian Information Criteria (BIC; Melnykov, 2016).

**Key Findings**

**Key Finding 1: Some Tessera SE skills were systematically observed in student use of the LMS**

To address our first research question, whether we can observe SE skills from behaviors recorded in online learning environments, we evaluated the correlations between activity features and final course grade and the five SE skills. Out of the 64 potentially relevant features calculated, 52 were significantly correlated with grade or one of the SE skill scores (p < .06; a significance level of .06 was used to account for a few associations that were marginally significant at the .05 level). From these 52 features, 33 features were associated with grade, 32 features with Grit, 5 with Leadership, 22 with Teamwork, 4 with Resilience, and 7 with Curiosity. The statistically significant correlations of greatest magnitude (up to ten) are presented in Figure 1. It is notable that with the exception of Leadership and Resilience, all SE skills, as well as grade, had positive correlations with LMS activity measures. Grade had the largest correlations with LMS activities, followed by Grit, which is to be expected given that this course was not designed with SE skills in mind. There were different activities associated with each of the SE skills (with the exception of number of sessions, which was significantly related to several of the SE skills), thus providing additional validity evidence for ACT Tessera; the skills are distinct and are expressed in different student activities.
Figure 1. Correlation of LMS Activities with Grade and SE Skills

**Grade**
- Number of Attempts on Discussion Board: 0.358
- Time spent on Discussion Board: 0.316
- Variability of time spent on Discussion Board: 0.313
- Number of Sessions: 0.307
- Time spent on Lecture Class Activities: 0.243
- Time spent on Announcements: 0.221
- Variability of time spent on Announcements: 0.206
- Time spent on Gradebook: 0.191
- Variability of time spent on Lecture Class Activities: 0.185
- Variability of time spent on Information: 0.177

**Grit**
- Variability of time spent on Assignment: 0.335
- Time spent on Assignment: 0.233
- Number of Attempts on Assignment: 0.212
- Number of Sessions: 0.18
- Time spent on Lecture Class Activities: 0.175
- Number of Attempts on Announcements: 0.173
- Variability of time spent on Announcements: 0.146
- Time spent on Orientation: 0.145
- Time spent on Announcements: 0.137

**Leadership**
- Number of Attempts on General Content: 0.136
- Variability of time spent on Exam Review Materials: 0.12
- Time spent on Exam Review Materials: 0.11
- Variability of time spent on Practice Exam: 0.109
- Time spent on Answer Key: 0.097
- Time spent on Answer Key: 0.211

**Teamwork**
- Time spent on Email Messages: 0.156
- Time spent on Assignment: 0.141
- Number of Attempts on Announcements: 0.137
- Number of Sessions: 0.135
- Time spent on Orientation: 0.135
- Variability of time spent on General Content: 0.133
- Variability of time spent on Announcement: 0.125
- Number of Attempts on Assignments: 0.124
- Time spent on Announcements: 0.115
- Variability of time spent on Orientation: 0.114

**Resilience**
- Grade: 0.102
- Time spent on Exam Review Materials: 0.131
- Number of Attempts on Survey: 0.123
- Variability of time spent on Exam Review Materials: 0.101
- Time spent on Discussion Class: 0.133

**Curiosity**
- Grade: 0.139
- Time spent on Email Messages: 0.136
- Variability of time spent on Learning Objectives: 0.132
- Number of Attempts on Gradebook: 0.113
- Number of Attempts on Announcements: 0.111
- Number of Attempts on Learning Objectives: 0.105
- Time spent on Learning Objectives: 0.102
- Variability of time spent on Reading Quiz: 0.097
Key Finding 2: High-level learning tactics could be extracted using sequential data mining techniques

While the relationship of individual items to outcomes is interesting, students frequently engage in a series of activities during a session in the LMS. As mentioned above, we used a model-based sequence clustering method to group activity sessions. To get the optimal number of clusters, we considered six possible values (from 3 to 8) and built a model for each choice. The best model was selected using the BIC, a common method used for model selection. Among the six models, the model with the smallest BIC value was chosen as the optimal one. As a result, we were able to identify five “learning tactics” from common patterns of activities. These patterns also had interesting relationships to SE skills and student course grade.

These learning tactics were created by identifying the most observed activities and entries in TPMs (see above). For example, in the first learning tactic, 40% of activities were either “assignment-” or “lecture or online class activity-” related. Therefore, we refer to this tactic as “Study and Assignment.” In Figure 2, the most probable pair-wise transition probabilities for this tactic are illustrated as a point of comparison. The left column indicates the initial activities. The most probable activities are on the top, and the size of the block indicates the magnitude of the probability. The width of an arrow is proportional to the probability value in the TPM. From the graph, we find that, with the Study and Assignment tactic, when students started with an “assignment” activity, they were highly likely to follow with class lecture or online class related activities. Also, lecture class activities led to “assignment” activities quite often. A second example is provided in Figure 3, which depicts the “Reading Quiz” tactic. Notice that taking a reading quiz is the most frequent activity, and the most important sequential pattern was from reading quiz to assignment.

Figure 2. Study and Assignment Learning Tactic Transition Matrix
We labeled each student session (defined as all activities that took place between a single instance of logging in and logging out) with a corresponding learning tactic. Then we generated the aforementioned activity sequence features at the student-level by computing for the number of times (frequency) a student exhibits each identified learning tactic (number of sessions). As the frequency of a learning tactic can vary widely, we also computed the relative frequency, dividing the frequency by the total number of sessions for that student. Based on the above information, we characterized the learning tactics as follows:

- **Study and Assignment**: Activities in this group were mainly related to viewing online lectures and taking assignments.
- **Reading Quiz**: In this group, taking a reading quiz was the most frequent activity. Also, in both directions, the transitions between “reading quiz” activity and “assignment” activity were the most important sequential patterns.
- **Exam-Related**: The majority of activities in this group were exam-related activities, such as reviewing exam materials and taking practice exams.
- **Gradebook and Discussion**: Most of the activities in this group were checking grade-book information and participating in the discussion board.
- **Announcement**: Sessions in this group contained only one activity, announcements that provide current information related to the course.
In Table 5, descriptive statistics are provided that characterize the number of student sessions in each of the learning tactics for the entire semester. Activities related to the weekly reading quizzes, which are counted toward student course grade, are the most frequent activity by a substantial amount (M = 90.05), while activities related to learning (“Study and Assignment,” M = 10.09) or preparing for an exam (“Exam-related,” M = 10.33) had a lower number of sessions.

Table 5. Learning Tactic Activity per Student (Semester Total Count; N = 489)

<table>
<thead>
<tr>
<th>Learning Tactics</th>
<th>Mean</th>
<th>Median</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study and Assignment</td>
<td>10.09</td>
<td>10</td>
<td>4.26</td>
<td>2</td>
<td>33</td>
</tr>
<tr>
<td>Reading Quiz</td>
<td>90.05</td>
<td>89</td>
<td>27.25</td>
<td>2</td>
<td>183</td>
</tr>
<tr>
<td>Exam Related</td>
<td>10.33</td>
<td>9</td>
<td>6.15</td>
<td>0</td>
<td>46</td>
</tr>
<tr>
<td>Gradebook and Discussion</td>
<td>26.89</td>
<td>21</td>
<td>21.77</td>
<td>2</td>
<td>225</td>
</tr>
<tr>
<td>Announcement</td>
<td>30.67</td>
<td>23</td>
<td>26.66</td>
<td>3</td>
<td>153</td>
</tr>
</tbody>
</table>

Key Finding 3: Predictions using LMS behavioral data were more accurate than those made using student family and educational background, providing an accurate and powerful basis for actionable interventions

To address the question of whether predictions using LMS behavioral data are more accurate than those made using demographic data, we built multiple predictive models using different predictor sets using both the linear regression (LR) method and the gradient boosting machine (GBM) method. The GBM models outperformed the LR models for all predictor sets at every week of the course. Hence, we only present performances of GBM models. Figure 4 shows the RMSE of models predicting final course grade and the five SE skill scores with different predictor sets. Caution should be exercised in comparing the RMSE between grades and SE skills as each outcome had different ranges. Nonetheless, in all cases, we found that the accuracy of models made using LMS behavioral data were slightly higher than models using the background demographic and academic information. Given that the LMS data reflect student activity during an entire semester, it is a better source for taking action and providing students with updated information that will allow them to self-regulate and improve their performance in a course.

In Figure 5, the predictive models for course grade and SE skills are displayed. In predicting grade, the background variable set contains cumulative college GPA, which is usually a powerful predictor of the course grade. We used the
model with the background variables as a baseline to compare other models. This baseline model already has a strong performance with an RMSE value of 0.692. For each of the predictor sets (other than BV) individually used in the model, the LMS Activity Measures (ACTV) set is the most powerful set (RMSE = 0.703), followed by Activity Sequence Features (SEQ) (RMSE = 0.813) and SE skill scores (TES; RMSE = 0.899). Model performance progressively improved by adding more types of predictor sets. When we used all the SE skills and learning behavior features, the model had the smallest RMSE of 0.677, which is less than a 2% improvement over the benchmark model. This improvement is modest and would be practically insignificant; in operational use, it is likely that only one of the LMS data sources would be used.

**Figure 4. Comparative Predictive Model Accuracy for Grade and SE Skills**
In pursuit of a predictive model, the accuracy over time is critical, as it determines how early an intervention could be implemented to improve student performance, with sufficient time for a student to recover his or her learning or address gaps in knowledge or relative grade performance. In Figure 5, the weekly performance of the consolidated LMS behavioral data set is depicted. For weekly predictive models of course grade, we used the predictor sets of TES, ACTV, and SEQ features. As more information was collected about students’ online behaviors, model performance steadily improved over time. It appears that at week 6, model performance achieves an RMSE of 0.79 (within .75 of a grade point), which could be sufficient to make a reasonably accurate prediction to identify potentially at-risk students. This could be used to inform early intervention if a student had a low predicted final grade. Predictive models of SE skills were created using a combination of ACTV and SEQ features (i.e., we analyzed the theoretical causal model in reverse; although SE skills are likely to lead to the types of behaviors observed in the LMS, here we analyzed this relationship in reverse). In contrast to predictions of grade, we observed stable model performance over time for SE skills. Collecting more behavioral data did not significantly improve model accuracy, and SE skills could be predicted by the second week of the term. The contrast between the grade and SE skill model resolution is a striking finding that merits further consideration and additional study.
Key Finding 4: LMS behavioral data partially mediated the relationship between Grit and course grades

We conducted a series of linear regression models with bootstrapping (Preacher & Hayes, 2008) to test whether student online behaviors mediate the relationship between SE skills and course grades. In testing mediation effects, only the relationship between Grit and course grade was found to be partially mediated by LMS behavioral data. Table 6 shows the results of the mediation analysis for Grit. The total effect of Grit on grades was 0.348, and the direct effect of Grit on grades was 0.205, meaning that there was a total indirect effect of 0.142, which corresponds to a 41.2% mediation effect. The specific indirect effects provide an estimate of the relative mediation effect of each of the mediators in the model. The largest mediation effect was the total number of sessions (Num_session) a student engaged in during the semester (0.052), followed by the variability of

Figure 5. Comparative Predictive Model Accuracy for Grade and SE Skills
time spent engaging in practice exam activities (Practice_exam_TimeSpentSD; 0.030), a proxy for the extent to which students accessed features consistently throughout the semester, and number of times a student engaged with the discussion board (Discussion_board_Attempt; 0.023).

Table 6. Results of Mediation Analysis: LMS Behaviors Mediating the Association between Grit and Course Grades

<table>
<thead>
<tr>
<th>Description</th>
<th>β</th>
<th>SE</th>
<th>t-value</th>
<th>p-value</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total Effect</strong></td>
<td>0.348</td>
<td>0.052</td>
<td>6.729</td>
<td>0.000</td>
<td>0.246</td>
<td>0.450</td>
</tr>
<tr>
<td><strong>Direct Effect</strong></td>
<td>0.205</td>
<td>0.048</td>
<td>4.277</td>
<td>0.000</td>
<td>0.111</td>
<td>0.299</td>
</tr>
<tr>
<td><strong>Specific Indirect Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discussion_board_Attempt</td>
<td>0.023</td>
<td>0.006</td>
<td>3.631</td>
<td>0.000</td>
<td>0.012</td>
<td>0.036</td>
</tr>
<tr>
<td>Practice_exam_TimeSpentSD</td>
<td>0.030</td>
<td>0.006</td>
<td>5.063</td>
<td>0.000</td>
<td>0.018</td>
<td>0.041</td>
</tr>
<tr>
<td>Number of Sessions (Num_session)</td>
<td>0.052</td>
<td>0.010</td>
<td>5.195</td>
<td>0.000</td>
<td>0.032</td>
<td>0.070</td>
</tr>
<tr>
<td>Exam Related Learning Tactic</td>
<td>0.018</td>
<td>0.004</td>
<td>4.331</td>
<td>0.000</td>
<td>0.010</td>
<td>0.026</td>
</tr>
<tr>
<td>Survey_Attempt</td>
<td>0.008</td>
<td>0.004</td>
<td>2.004</td>
<td>0.046</td>
<td>0.000</td>
<td>0.016</td>
</tr>
<tr>
<td>Announcements_TimeSpentSD</td>
<td>0.017</td>
<td>0.009</td>
<td>1.973</td>
<td>0.049</td>
<td>0.001</td>
<td>0.033</td>
</tr>
<tr>
<td>Practice_assignment_TimeSpent</td>
<td>-0.005</td>
<td>0.004</td>
<td>-1.394</td>
<td>0.164</td>
<td>-0.012</td>
<td>0.002</td>
</tr>
<tr>
<td><strong>Total Indirect Effect</strong></td>
<td>0.142</td>
<td>0.046</td>
<td>3.086</td>
<td>0.002</td>
<td>0.053</td>
<td>0.232</td>
</tr>
</tbody>
</table>

Key Finding 5: No evidence of demographic bias was found in predictive models

Ensuring appropriate representation of all students, including those from underserved backgrounds, is a critical obligation for researchers. To ensure the results were free from bias and appropriately characterized all subgroups, the predictive models were fit again with the subgroup criterion (e.g., gender) removed from the model. Then the distribution of prediction error between the members of the subgroup (e.g., male / female) was plotted using a density plot, and a two-sided t-test was run to evaluate if there was a significant difference between the groups. We conducted subgroup analysis based on students’ gender, race / ethnicity, first generation college student status, and transferred student status. UMBC identified these characteristics as important in their efforts to promote academic success for all students. None of the mean difference tests were statistically significant. Additional visual inspection of RMSE distributions,
as illustrated in Figure 6, which depicts gender differences, did not reveal any differences across the error distribution. These findings indicate that the predictive models apply for all UMBC students independent of their gender, race / ethnicity, first generation status, and transfer student status.

**Figure 6. Distribution of Model Bias (Error) by Gender**

![Distribution of Model Bias (Error) by Gender](image)

**Implications**

These findings demonstrate that SE skills not only have systematic relationships with LMS activity (analyzed as individual activities or aggregated into sequential learning tactics) but can also be predicted by them. In the correlation analyses, differences between the correlation coefficients and relative importance of LMS activities suggest strong construct validity; that is, different SE skills are observed in different student actions using the LMS. For example, Grit had the strongest correlation with completing assignments, the number of online sessions, and lecture activities – all of which are behaviors that one would expect from students who are attentive to detail and persistent at completing their work. In contrast, Curiosity had the strongest correlations with learning objectives and gradebook – both of which are related to activities that provide new ideas or updated course information to the student. A partial mediation effect was also found for the
relationship between Grit and course grades, which is what would be expected if the LMS activities are indeed a behavioral manifestation of the underlying SE skill.

In the predictive models, we observed high accuracy between the LMS activity features and SE skills. After the second week, the predictive model accuracy was nearly the same, which indicates that student activities for the remainder of the semester are consistently associated with the SE skills. Grit had a lower accuracy than other SE skills in the predictive models; however, the difference in accuracy was small (0.03) and may be attributed to the smaller range in Grit scores. In practical terms, this difference was negligible. Taken in sum, these findings indicate that SE skills are robustly and consistently accounted for in this course’s LMS data.

In comparing predictive models of course grade using different predictor sets, the model created using SE skills only had lower predictive accuracy than the model using LMS activity, consistent with the findings from the correlation analysis. For the LMS activity, the individual activity features had higher accuracy than the features of activity sequences. This is likely due to the larger number of ACTV features. By adding all the predictors together into a combined model (ACTV+SEQ+TES), we found a small improvement in accuracy over the model with individual activity.

The combined model achieved a similar level of accuracy as the baseline model with student demographic background data, even when background data included current college GPA. The LMS data have multiple benefits compared to background data in providing insights in how to improve student performance within a class. Namely, it is sensitive to current activities and can provide actionable feedback on changes that a student could make over the course of a semester. By contrast, background information only indicates students at risk based on their historical circumstances or prior academic performance, without any equitable interpretation behind that prediction or practical actionable insights for that student during a course (e.g., how to address one’s first generation status). Further, student background data can be difficult to obtain without significant additional effort and can contain sensitive personally identifiable information. These results show that LMS data, modeled properly, can outperform background data and provide actionable information that could be presented to students or faculty or used to build automated systems.

A benefit of predicting both SE skills and course grades is that they could be used to create personalized interventions that could suggest messages and perhaps even different activities depending on the SE skills of the student. For example, a student with low Grit whose grade is predicted to be low could be encouraged to attend to deadlines, calendar items, and other detailed items,
whereas a student with high Grit could be encouraged to focus on forum interactions and other communication activities.

In sum, this research demonstrates that SE skills can be operationalized using LMS data and learning analytics techniques. The differences in correlations and model accuracy metrics between these skills and student grades demonstrate that these are unique constructs that can provide students, instructors, and other stakeholders with useful information that they can use to improve student success.

**Limitations and Future Work**

This project was conducted with a single course with intensive use of the LMS, and as such, the findings should be considered as exploratory until replicated with additional courses and institutional contexts. Part of that future analysis should include identifying the LMS adoption threshold required for these relationships to be found; that is, it may be that courses with low usage of the LMS will not show significant relationships with outcomes. Or further, low LMS usage would prevent instructors from creating actionable feedback that could be delivered via the LMS. In addition, the manual mapping of LMS log data to course-specific features limits the potential to scale out this work without significant additional effort. Automated methods to identify features that also have fidelity with the underlying course design would enable broader testing and deployment of models.

Additional future work should explore the nature and timing of behavioral nudges or interventions that could be implemented to improve student success and the resulting impact on course grades. While the results of this study are compelling, the relationships found are correlational rather than causal; additional studies could provide evidence for the causal effects of increasing or decreasing specific LMS activities on course outcomes.
References


About ACT

ACT is an independent, nonprofit organization that provides assessment, research, information, and program management services in the broad areas of education and workforce development. Each year, we serve millions of people in high schools, colleges, professional associations, businesses, and government agencies, nationally and internationally. Though designed to meet a wide array of needs, all ACT programs and services have one guiding purpose—helping people achieve education and workplace success.

About ACT Research

ACT Research leads the field with authority and high-quality scientific evidence in support of education and workforce practices, solutions, and services. Our mission-driven team comprises a variety of backgrounds and disciplines, and offers a wide spectrum of knowledge and skills, enabling us to deliver quality, high-impact products and services aligned to ACT’s strategy and mission. Together, our research teams provide policymakers, educators, parents, and learners with research-based insights to inform their decision-making, and deliver educators and workforce development professionals with tools and services needed for education and career navigation.