Probabilistic Motivation Profiles and Student Behaviors in Log Data

Gahyun Sung, Harvard University, gsung@g.harvard.edu

Abstract: Motivation is a multi-faceted construct that has complex relationships with behavior. To better understand student motivations in a large introductory statistics course, we cluster different aspects of student motivation and investigate their link to observed student engagement in an online textbook. A soft clustering method reveals three distinct motivation profiles in students: reluctant, motivated, and confident. Membership in the confident group is associated with GPA and financial difficulties, but not with engagement metrics that reflect student choice, such as time spent. Contrary to the simple hypothesis that better motivation will lead to higher engagement, students with "reluctant" and "motivated" profiles seem to spend similar amounts of efforts for course preparation but spend less of it progressing with learning, and more time struggling.

Introduction

Numerous theories link learner motivation to engagement, a precursor to successful learning (e.g. Elliot & Harackiewicz 1994; Ryan & Deci 2000). While studies have utilized self-reports to empirically establish the link between motivation and adaptive student behavior (e.g., Hendy, Schorschinsky & Wade, 2014), self-reports can be biased and may be prone to recall error, particularly when probing low-level behaviors. Thus, an alternative method of inquiry for computer-assisted learning environments has been to take advantage of the digital log traces that occur with student activities (e.g. clickstream) to track their behavior (e.g., Xu & Yang, 2016; Schoor & Bannert, 2011).

The current study adopts this approach to study a context where motivation is expected to vary, and is crucial for success: a large introductory statistics course where college students engage with an online textbook in a self-directed manner. Motivation is operationalized using the expectancy-value theory of motivation (Eccles et al., 1983), which has been previously successfully applied in the context of learning analytics for the goal of finding student profiles, in studies such as Schumacher & Ifenthaler (2018) as well as Templaar and colleagues (2018).

Two methodological considerations are made in response to calls for change in associated fields: the use of a person-centered approach in motivation profiles, and the use of engagement metrics that go beyond click counts. The use of a person-centered approach in estimating student motivation have been increasingly endorsed by researchers as a way of allowing how patterns of variables related to motivation have an intricate relationship with one another at the level of the individual (Linnenbrink-Garcia & Barger, 2014). The use of more sophisticated log data metrics that measure engagement and student learning behavior in ways other than summative count has been raised time and again, such as in Knight & Shum (2017), Lodge and Lewis (2012), and Fincham and colleagues (2019).

With such considerations in mind, we hope to answer the following research questions:

1. What types of distinct student motivation profiles exist in an introductory statistics course?
2. Are the probabilities of belonging to a certain motivational profile impacted by student factors, namely gender, race, GPA, and economic hardship?
3. Do the ways in which students engage with course material differ by motivational profile?

Setting and data

Log data was collected from an online textbook used in a large introductory statistics class in a U.S. research university with a highly competitive undergraduate program. The online textbook was a central component in the course. Students learned the material by working through assigned chapters in the textbook every week before coming to lectures. The log data includes all student interactions with textbook elements, such as reading material, graphics, videos, R coding exercises, and formative assessment questions. Our data comes from 166 students who filled out the pre -course survey, and did not drop out of the course.

The survey data used to create motivational profiles comes from the pre-course survey. The survey contained various demographic questions, as well as a combination of validated measures for the estimation of expectancy, value, and cost from two instruments (Kosovich, et al., 2015; Gaspard et al., 2017). To create a more holistic, person-centered motivational profile, we utilize additional survey questions relevant to discriminating different student motivations for taking a statistics course: the level of self-reported prior experience in statistics, and the level of self-reported intent to persist in statistics learning.

Methodology

Motivational profiles were created by combining confirmatory factor analysis (CFA) with gaussian mixture modeling (GMM). First, confirmatory factor analysis was applied to the previously described survey items to create composite scores. Five factors were created: expectancy, value, cost, prior experience, and intent (to persist in statistics). Model fit indices largely indicated that the CFA had acceptable fit (χ^2=539.481, p < 0.001; CFI = 0.830, RMSEA = 0.084, SRMR = 0.077). Then, the resulting factors were normalized and used to create GMM-based student motivation clusters. GMM is a "soft" clustering method, i.e., a probabilistic model that assigns a student a probability of belonging to a certain cluster. Compared to "hard" clustering methods such as k-means clustering, it provides a better theoretical fit in cases where students do not belong to clearly discrete groups. A three-cluster model was chosen based on elbow graphs of fit statistics (AIC, BIC).

Parallelly, the following week-level metrics were created from the log data to understand student behaviors expected to co-vary with motivation: time spent, ratio of questions that the student ultimately did not get correct despite unlimited tries, ratio of questions never attempted, ratio of attempts correct to total attempts, and ratio of time spent on the due date (i.e. cramming) to total time. These metrics attempt to measure not only the amount of effort or engagement the student put forth in a particular week, but also the way in which this time was spent: was a student's time on the textbook spent fruitfully progressing through the material, or spent making wrong attempts, or rushing to finish work? That is, might student motivation profiles be related to not only the amount of effort, but also the ways in which this effort is expended?

While three distinct clusters are described, subsequent analysis uses a binary variable for membership in the most advantageous cluster, termed the "confident" group (described further in the next section). This is to account for strong class imbalance – the smallest cluster, while mathematically and theoretically distinct, was approximately 1/10 of the size of the "confident" group. Two types of analysis are performed on this binary profile. First, we test whether different student factors, namely race, gender, GPA, and economic hardship, are associated with "confident" group membership. Then, behavioral metrics are connected to this binary variable using linear mixed models (LMM), a generalization of linear regression that allows modeling of random effects of correlated clusters (Byrk & Raudenbush, 1987). Since we have multiple datapoints from each student, and multiple datapoints from each week, LMMs allow us to explicitly model these effects for a less biased estimate of main effect sizes. Lastly, given the importance of prior academic history in determining motivation and student behavior, we introduce prior GPA (a binary variable, 3.5- and 3.5+) as a control variable in LMM analysis. Identical models without the control variable showed the same trends.

Results

RQ1. Student profiles in an introductory statistics course

Our first research question asked whether there are distinct motivation profiles in this introductory statistics course. The results of cluster analysis showed the emergence of three distinct clusters. Based on the inspection of variable distributions, we respectively term them the "reluctant", "motivated" and "confident" groups. As seen in figure 1, the reluctant group is characterized by a low level of intent to persist in statistics, low expectancy for success, and low value for the course, despite having some prior related experience. The motivated group stands out most for their lack of prior experience, yet they expect to do better than the reluctant group and value the class more, as well as having a higher intent. Lastly, the confident group has the highest intent, level of prior experience, expectancy, and value. The reluctant group includes 12 students, while the motivated and confident groups each consist of 59 and 95 students.

Figure 1

Factor distribution for reluctant, motivated, confident clusters

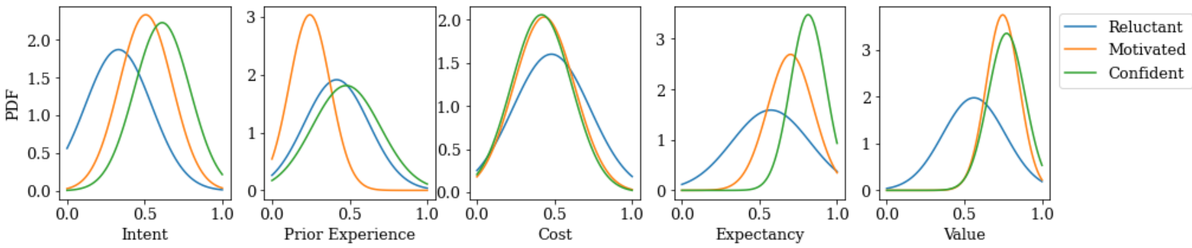

Table 1
Factor means for reluctant, motivated, confident clusters

	Intent	Prior experience	Cost	Expectancy	Value
Reluctant	0.33	0.42	0.48	0.57	0.56
Motivated	0.51	0.24	0.43	0.70	0.75
Confident	0.61	0.48	0.42	0.81	0.77

RQ2. Student factors influencing probabilities of profile membership

Next, turning to our second research question, results of contingency analysis showed that group membership is not statistically significantly associated with either student gender or race; however, they were significantly associated with prior GPA (less or more than 3.5; $\chi^2(1) = 8.039, p = 0.004$) and economic hardship (level of economic hardship reported: high, medium, low, none; $\chi^2(3) = 9.527, p = 0.023$). In the contingency table below (table 2), actual counts are accompanied by expected values in parenthesis. The table indicates that lower GPA students, and students with high- and medium-levels of financial difficulty tend to belong less than chance to the confident group.

Table 2
Contingency table between student factors and motivation clusters.

	GPA		Economic Hardship			
	3.5-	3.5+	High	Medium	Low	None
Confident group	**32 (44)**	132 (125)	**13 (18)**	**46 (51)**	56 (44)	60 (61)
Other groups	52 (40)	104 (116)	22 (17)	52 (47)	31 (42)	59 (58)

* Number in cells indicate observed count (expected count)

RQ3. Connection between student motivation profiles and learning behavior

The LMM regression results showed somewhat surprising results. Many metrics did *not* have a statistically significant association with "confident" group membership. The time spent on the textbook, the ratio of questions never attempted, and ratio of time spent cramming all had statistically nonsignificant associations. However, the ratio of correct attempts to total attempts, and the ratio of questions given up differed between the confident group and other groups. Table 3 summarizes the LMM regression results for these two metrics for which there were significant results. Belonging to the "confident" group was on average associated with a 5-percentage point increase in the ratio of correct attempts, and a 7-percentage point decrease on the ratio of questions never answered correctly, or given up, even after controlling for student GPA.

Table 3
Results of LMM regression analyses

	M1: Ratio of correct attempts	M2: Ratio of questions given up
Belonging in "Confident" group	.05 *** (3.544)	-.07 *** (-4.212)
GPA less than 3.5	-.07 *** (-4.60)	.11 *** (5.728)

*** p < 0.001; Satterthwaite approximations used for calculation of t-statistics (in parentheses)

Discussions and conclusion

Our results reveal three distinct profiles of student motivation in taking this introductory statistics course: the reluctant group sees lower value in the course and the subject despite some prior experience, the motivated group is new to the field but has high expectations for the course and for their own performance, and the confident group is surest of their success and the relevance of this course, and has highest prior experience. Bringing contextual information into the model allows us to create realistic profiles that more holistically capture the types of student motivations for taking an introductory STEM course.

These motivational profiles are shown to be related to self-reported student GPA and economic hardship. While these results are hardly surprising, it does bear note that gender and race, influential factors known to be related to motivation in STEM, lose significance in this relatively homogenous population while economic hardship retains its influence, speaking to its salience and perseverance as a barrier to academic success.

Lastly, regression results show that having a "confident" motivation profile at the start of the course may be less related to how student *decides* to engage in course preparation. That is, the time they spend, the questions attempted, and the way they divide the time throughout the week are all choices students autonomously make. Conversely, the ratio of correct attempts, and the ratio of questions given up, both indicators of the quality of experiences a student had during their engagement, differ by motivational profile. While further analysis is needed to fully understand this different experience, it seems that students with the "reluctant" and "motivated" profiles spend similar efforts for course preparation but spend less of it progressing with learning, and more time struggling with the material. This disproves the simple hypothesis that "better" motivation leads to higher engagement; rather, it seems that these holistic motivational profiles are correlated with student resources beyond their willingness to put in effort, which lead to different experiences in the course, despite their best intent.

Future work will focus on creating more sophisticated engagement metrics, making the next natural connection between engagement and performance, and then testing a larger model that includes motivation, engagement, and performance. Ultimately, we hope to understand how students that enter a STEM classroom with different types and levels of motivation engage in learning based on their process data, for the end goal of tailoring feedback for students based on both their motivation and observed engagement.

References

Bryk, A. S., & Raudenbush, S. W. (1987). Application of hierarchical linear models to assessing change. *Psychological bulletin*, *101*(1), 147.

Deci, E. L., & Ryan, R. M. (2000). The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, *11*(4), 227-268.

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meece, J. L., & Midgley, C. (1983). Expectancies, values, and academic behaviors. In J. T. Spence (Ed.), *Expectancies, values, and academic behaviors* (pp. 75–146). San Francisco, CA: Freeman.

Elliot, A. J., & Harackiewicz, J. M. (1994). Goal setting, achievement orientation, and intrinsic motivation: A mediational analysis. *Journal of personality and social psychology*, *66*(5), 968.

Fincham, E., Whitelock-Wainwright, A., Kovanović, V., Joksimović, S., van Staalduinen, J. P., & Gašević, D. (2019, March). Counting clicks is not enough: Validating a theorized model of engagement in learning analytics. In *Proceedings of the 9th international conference on learning analytics & knowledge* (pp. 501-510).

Gaspard, H., Häfner, I., Parrisius, C., Trautwein, U., & Nagengast, B. (2017). Assessing task values in five subjects during secondary school: Measurement structure and mean level differences across grade level, gender, and academic subject. *Contemporary Educational Psychology, 48*, 67-84)

Hendy, H. M., Schorschinsky, N., & Wade, B. (2014). Measurement of math beliefs and their associations with math behaviors in college students. *Psychological assessment, 26(*4), 1225.

Knight, S., & Shum, S. B. (2017). Theory and learning analytics. *Handbook of learning analytics*, 17-22

Kosovich, J. J., Hulleman, C. S., Barron, K. E., & Getty, S. (2015). A practical measure of student motivation: Establishing validity evidence for the expectancy-value-cost scale in middle school. *The Journal of Early Adolescence, 35*(5-6), 790-816.

Lodge, J., & Lewis, M. (2012). Pigeon pecks and mouse clicks: Putting the learning back into learning analytics. *Future challenges, sustainable futures. Proceedings ascilite Wellington*, 560-564.

Schoor, C., & Bannert, M. (2011). Motivation in a computer-supported collaborative learning scenario and its impact on learning activities and knowledge acquisition. *Learning and Instruction*, *21*(4), 560-573.

Schumacher, C., & Ifenthaler, D. (2018). The importance of students' motivational dispositions for designing learning analytics. *Journal of Computing in Higher Education*, *30*(3), 599-619.

Tempelaar, D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408-420.

Xu, B., & Yang, D. (2016). Motivation classification and grade prediction for MOOCs learners. *Computational intelligence and neuroscience*, *2016*.