Academic Success in Online Colleges: The Role of Self-Regulated Learning Profiles

Timothy J. Cleary1, Jason M. Bryer2, & Elie ChingYen3

1 Rutgers University

2 City University of New York

3 University at Albany

Author note

DAACS was developed under grants #P116F150077 and #R305A210269 from the U.S. Department of Education. However, the contents do not necessarily represent the policy of the U.S. Department of Education, and you should not assume endorsement by the Federal Government.

Correspondence concerning this article should be addressed to Timothy J. Cleary, Postal address. E-mail: [timothy.cleary@rutgers.edu](mailto:timothy.cleary@rutgers.edu)

Abstract

One or two sentences providing a **basic introduction** to the field, comprehensible to a scientist in any discipline. Two to three sentences of **more detailed background**, comprehensible to scientists in related disciplines. One sentence clearly stating the **general problem** being addressed by this particular study. One sentence summarizing the main result (with the words “**here we show**” or their equivalent). Two or three sentences explaining what the **main result** reveals in direct comparison to what was thought to be the case previously, or how the main result adds to previous knowledge. One or two sentences to put the results into a more **general context**. Two or three sentences to provide a **broader perspective**, readily comprehensible to a scientist in any discipline.

*Keywords:* self-regulated learning, diagnostic assessments, profile analysis

*Word count:* X

Academic Success in Online Colleges: The Role of Self-Regulated Learning Profiles

# Methods

## Participants

Data for this study were retrieved from a larger randomized control involving 13,196 newly enrolled students in an online school with students from all over the United States (Authors). From that study, half, or 6,536 students, were assigned to the treatment group. Of the students assigned to treatment, 6,376 (97.60%) completed all four assessments and are therefore represent the data used for this study.

Students in this study attended an online competency based college. Students enroll in for a six-month term where they are expected to complete a minimum of 12 credits within the six months. Competencies are completed by students at their own pace. They meet regularly with academic mentors (approximately every two weeks) to check on progress. When necessary, students can meet with subject matter experts, otherwise all course work is completed only asynchronously.

## Measurements

DAACS assess students in self-regulated learning, reading, writing, and mathematics. The self-regulated learning was developed specifically designed for DAACS and includes 57 questions across metacognition (with subdomains planning, monitoring, and evaluation), motivation (including subdomains self-efficacy, managing test anxiety, mindset, and mastery orientation), and strategies (including subdomains strategies for understanding, help seeking, managing time, and managing environment). Chronbach’s alpha (Cronbach, 1951) for SRL using the three subscales were 0.77, mathematics with 166 items, for mathematics with six subdomains was 0.64, reading with five subdomains was 0.66, and writing with ten subdomains was 0.75. The six DAACS measures will serve as the independent variables. For simplicity of interpretation, all independent variables will be converted to standard scores.

We analyze how the DAACS measures of college readiness predict two outcomes: feedback views and on-time progress. The DAACS system records all student clicks which serves as a proxy for student engagement with feedback. The feedback views measure removes any pages viewed for completing assessments and therefore capture only the pages which have feedback. Feedback views is positively skewed so will be log transformed before analysis.

All students in this study are considered full-time and expected to complete a minimum of 12 credits within six months. On-time progress is a dichotomous variable which indicates whether a student has successfully completed a minimum of 12 credits within the six-month term.

## Data analysis

To provide context for cluster analysis we will first use a variable centered approach. Specifically, regression models will be estimated the two dependent variables, feedback views and on-time progress, on the six DAACS measures using linear and logistic regression, respectively.

Cluster analysis is a statistical procedure that attempts to group observations such they are similar across a number of observations. *k*-means clustering (Lloyd, 1982; MacQueen, 1967) partitions observations in *k* clusters where each observation belongs to the cluster with the closest mean as measured by the squared-Euclidean distance. The goal is to minimize the variance (of the squared-Euclidean distances) within each cluster. This study will cluster students using the six DAACS measures of student readiness.

Once students have been partitioned into *k* clusters, null hypothesis tests will be conducted to determine if there are statistically significant differences between clusters. ANOVA and chi-squared tests will be used for feedback views and on-time progress, respectively.

# Results

Table 1 summarizes the linear and logistic regression models for feedback views and on-time progress, respectively. With the exception of motivation for the on-time progress model, all of the DAACS measures are statistically significant predictors. Of particular note, metacognition and motivation have negative coefficients.

Table 1: Regression results for predicting feedback views and on-time progress from DAACS measures.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Feedback Views | | On Time Progress | |
| (Intercept) | **1.045 \*\*\*** | **(0.097)** | **-3.100 \*\*\*** | **(0.285)** |
| Motivation | **-0.179 \*\*\*** | **(0.027)** | -0.005 | (0.078) |
| Metacognition | **-0.139 \*\*\*** | **(0.020)** | **-0.151 \*** | **(0.059)** |
| Strategies | **0.208 \*\*\*** | **(0.027)** | **0.216 \*\*** | **(0.080)** |
| Mathematics | **0.586 \*\*\*** | **(0.058)** | **1.322 \*\*\*** | **(0.169)** |
| Reading | **0.571 \*\*\*** | **(0.081)** | **2.245 \*\*\*** | **(0.232)** |
| Writing | **1.202 \*\*\*** | **(0.064)** | **1.065 \*\*\*** | **(0.180)** |
| N | 6376 |  | 6376 |  |
| R2 | 0.124 |  |  |  |
| logLik | -7287.883 |  | -3799.798 |  |
| AIC | 14591.766 |  | 7613.596 |  |
| \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. | | | | |

## Determining desired number of clusters

There are numerous procedures for determining the optimal number of clusters to use. The goal is to find a balance between statistical properties and theory. Figure 1 provides three common metrics used to determine the optimal number of clusters. The *x*-axis for all of these is the number of clusters and the *y*-axis is the statistic. As described above, *k*-means clustering attempts to minimize within cluster sum of squares of the Euclidean distances. The first plot provides the average within cluster sum of squares for varying number of clusters. When using this plot to determine the number of clusters one must balance minimizing the within sum of squares and over fitting. Typically this is referred to as finding the “elbow.” That is, at which point would increasing the number of clusters reduces the withing cluster sum of squares minimally. For the silhouette (Rousseeuw, 1987) and Gap (Tibshirani, Walther, & Hastie, 2001) scores larger values are desired. Based on the methods we examined 3, 5, and 6 cluster solutions[[1]](#footnote-26) and concluded that the five cluster solution was optimal in providing a strong theoretical interpretation, reasonable cluster sizes, and justifiable given these statistics.

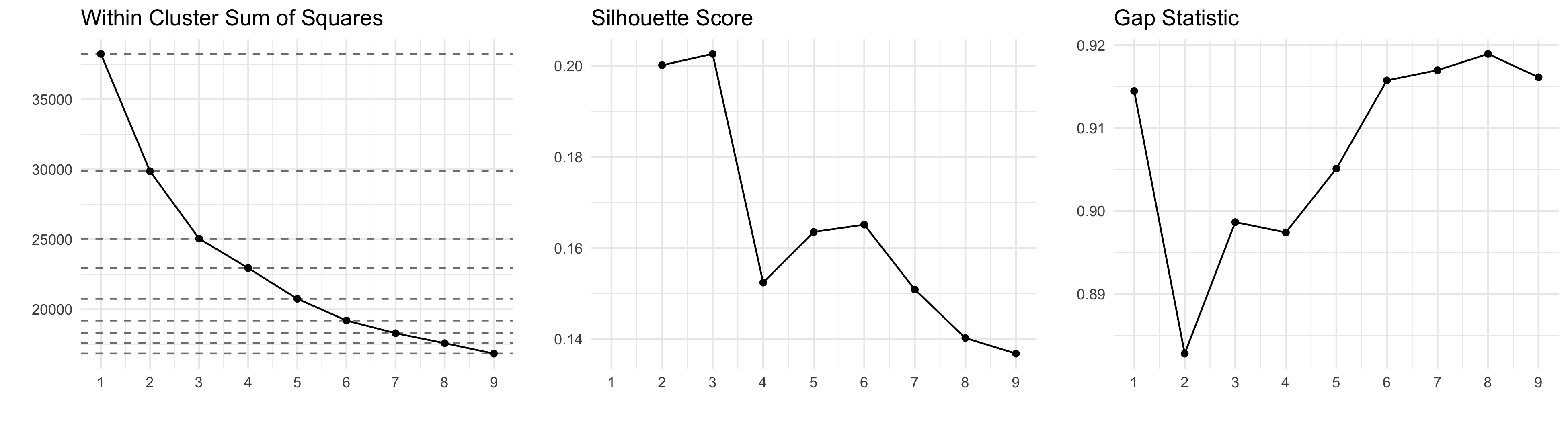


Figure 1: Scree plots to determine optimal number of clusters

## Cluster Analysis

*k*-means cluster was estimated using a five cluster solution. The total sum of squares was 20,747 with an average within cluster sum of squares of 4,149. Figure 2 provides a bar plot with the cluster sizes.

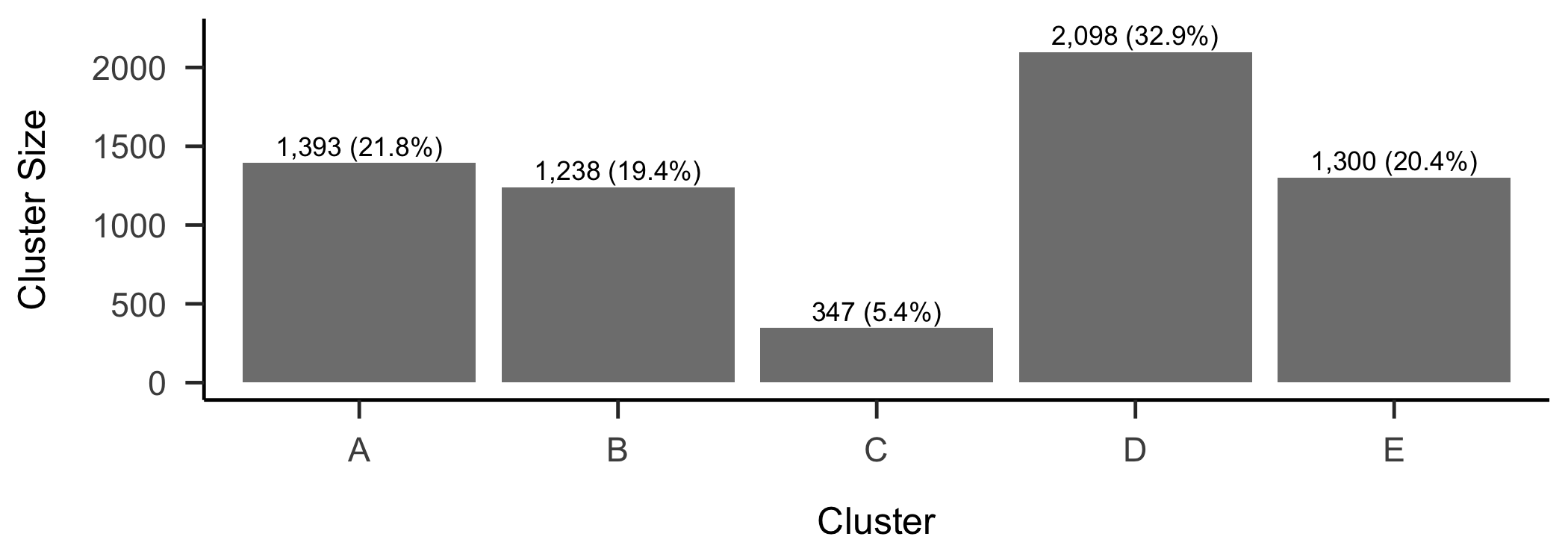


Figure 2: Cluster sizes for a five cluster solution.

3

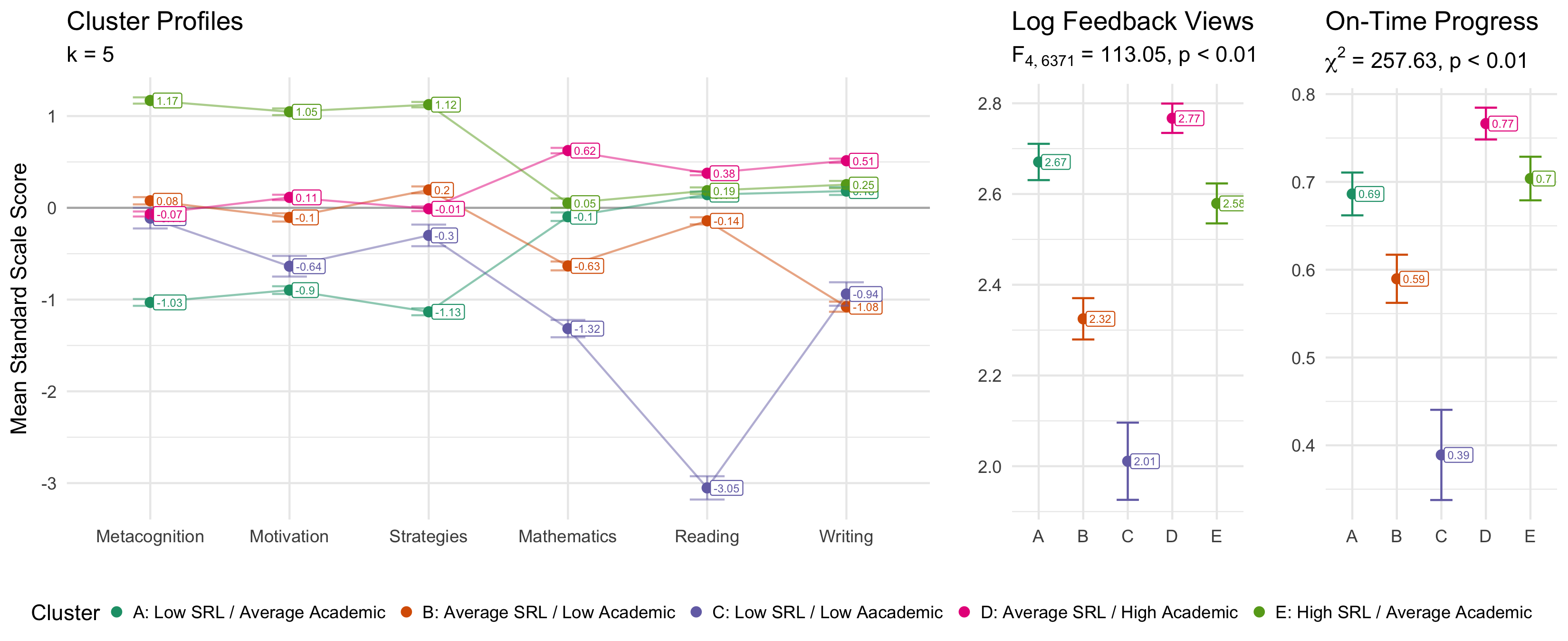


Figure 3: Cluster profiles (left) with feedback views and on-time success outcome variables (right)

2

Table 2: Regression results for predicting feedback views and on-time progress from clusters.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Feedback Views | | On Time Progress | |
| (Intercept) | **2.671 \*\*\*** | **(0.021)** | **0.783 \*\*\*** | **(0.058)** |
| clusterB | **-0.346 \*\*\*** | **(0.031)** | **-0.420 \*\*\*** | **(0.082)** |
| clusterC | **-0.660 \*\*\*** | **(0.047)** | **-1.234 \*\*\*** | **(0.124)** |
| clusterD | **0.096 \*\*\*** | **(0.027)** | **0.406 \*\*\*** | **(0.077)** |
| clusterE | **-0.091 \*\*** | **(0.030)** | 0.083 | (0.084) |
| N | 6376 |  | 6376 |  |
| R2 | 0.066 |  |  |  |
| logLik | -7489.907 |  | -3866.692 |  |
| AIC | 14991.813 |  | 7743.384 |  |
| \*\*\* p < 0.001; \*\* p < 0.01; \* p < 0.05. | | | | |

# Discussion

# References

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. <https://doi.org/10.1007/bf02310555>

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>

MacQueen, J. (1967). *Some methods for classification and analysis of multivariate observations*. Retrieved from <https://api.semanticscholar.org/CorpusID:6278891>

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. <https://doi.org/10.1016/0377-0427(87)90125-7>

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *63*(2), 411–423. <https://doi.org/10.1111/1467-9868.00293>

1. The online supplemental materials include the results from the three and six cluster solutions. [↑](#footnote-ref-26)