2018 ASEE Annual Conference & Exposition
June 24-27, 2018
Salt Palace Convention Center
Salt Lake City, UT

Paper ID #21437

Cluster Analysis Methods and Future Time Perspective Groups of Second-Year Engineering Students in a Major-Required Course

Dr. Justine Chasmar, Goucher College

Justine Chasmar is an Assistant Professor in the Center for Data, Mathematical, and Computational Sciences and the Director of the Quantitative Reasoning Center at Goucher College. Her research focuses on tutoring, student learning, motivation, and professional identity development. Through her background in learning centers, she has applied this research to undergraduate students and peer tutors. Her education includes a B.S. and M.S. in Mathematical Sciences and Ph.D. in Engineering and Science Education from Clemson University.

Ms. Katherine M. Ehlert, Clemson University

Katherine M. Ehlert is a doctoral student in the Engineering and Science Education department in the College of Engineering, Computing, and Applied Sciences at Clemson University. She earned her BS in Mechanical Engineering from Case Western Reserve University and her MS in Mechanical Engineering focusing on Biomechanics from Cornell University. Prior to her enrollment at Clemson, Katherine worked as a Biomedical Engineering consultant in Philadelphia, PA. Her research interests include identity development through research experiences for engineering students, student pathways to engineering degree completion, and documenting the influence of co-op experiences on academic performance.

Cluster Analysis Methods and Future Time Perspective Profiles of Second-Year Engineering Students in a Major-Required Course

Introduction

This paper meets our two goals of (1) identifying homogeneous groups of second-year engineering student FTPs and (2) introducing commonly used cluster analysis techniques and providing an example of how to implement said techniques within an engineering education context. One specific aspect of motivation, Future Time Perspective (FTP) [1], has been shown to have a connection to student strategies and how they approach learning in the present [2]–[4]. One way of evaluating FTP is quantitatively through a survey instrument like the Motivation and Attitudes in Engineering survey [5]–[8]; however, it is often difficult to select appropriate analysis methods for such quantitative data, and there is a lack of literature for engineering educators comparing types of quantitative analytic methods. Thus, the second purpose of this paper is to fill this gap by discussing how to implement different types of cluster analysis (CA) techniques to create homogenous groups and how to select the best clustering method and solution based on reported results. This paper builds on the cluster analysis considerations of Ehlert, et al [9] with the following research questions for this paper: *1. What cluster analysis technique is the best fit to determine the motivational (FTP) characterizations of undergraduate engineering majors within the context of a major-required course? 2. What are the motivational (FTP) characterizations of undergraduate engineering majors within the context of a major-required course?*

Background

FTP is often defined as the "present anticipation of future goals" [10] (p. 122), and FTP can be contextualized for undergraduates as students' goals, views of the future, and the impact these goals and views have on actions in the present. FTP as a theory is important because a well-developed FTP has been quantitatively and qualitatively linked to goal-setting, self-regulation, and success in engineering programs [2], [6], [10]–[13]. In this paper, domain-general (*Connectedness, Value*), domain-specific (*Perceptions of the Future, Present on Future, Future on Present*), and context-specific constructs (*Perceived Instrumentality*) were considered. In general, *Value*, often termed valence, is the "anticipated subjective value"[14] (p. 567) of future goals for a person; thus students may place a higher value or hold one goal in higher regard than another goal. The second domain-general FTP construct, *Connectedness,* is "general feeling of connectedness to and planfulness about the future" [15] (p. 116). *Perceived Instrumentality* (PI) [15]–[17] is a context-specific variation of connectedness and is described as the importance a person places on a current task (e.g. engineering course) towards future goals. This importance, or perception of instrumentality, may be considered *endogenous*, directly related to a person's future goals, or *exogenous*, tangentially related but being seen as something to overcome towards a future goal [18].

Within the domain of engineering [19], *Perceptions of the Future* (PoF) is described using three terms: Relative distance of a students' goals into the future (*extension*); their positive to negative attitude regarding the future (*time attitude*), and "habitual time space" [15] (p. 115) (*time orientation*). The impact of current or previous tasks on goal creation is considered PoF. Similarly, a long extension supports the view of future goals impacting the present [15], which is described as the construct *Future on Present* (FoP). Overall, these domain-general, domain-specific, and context-specific FTP constructs can be utilized to qualitatively describe and quantitatively determine the future views and motivations of undergraduate students within engineering.

Cluster analysis

CA is the "art of finding groups in data" [20] (p. 1) and is the best method for this research due to its "person-centered" approach, as it allows a "one-to-many" look at dimensions [21] (p. 901). To select a CA method for a study, three questions should be considered [22]: Which similarity/dissimilarity measure (measurement of distance between data points) is appropriate? How should the data be normalized? How should domain knowledge (theory and input parameters) be utilized when clustering data? Additionally, external (fit of clustering solution compared to theory), internal (fit of the clustering solution compared to the data), and relative (fit of multiple clustering solutions) quality should be considered [23].

Figure 0: depicts an overview of CA methods available for selection and breaks CA into two categories: hierarchical and partitioning [22]. Hierarchical methods are used when little theory is available to frame the research [24], [25], allowing the data to drive the results. Partitioning methods, on the other hand, are more methodologically sound when there is strong theory to support the required *a priori* inputs [23], [26], [27]. For more detailed discussion of the different algorithms one can use in CA, see [9], [22], [23], [28]. In this paper, we use Ward's and k-means as these are very common and robust algorithms [9], [22].

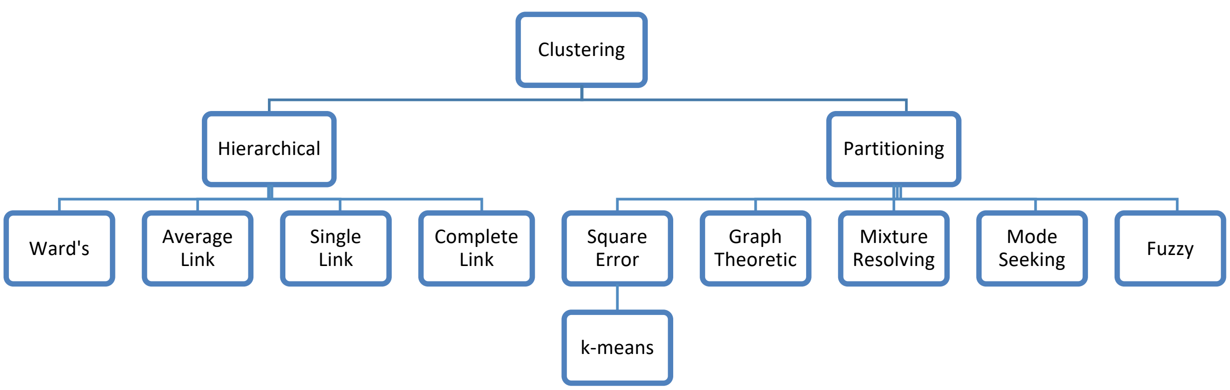

Figure 0: A taxonomy of clustering approaches [22]

Cluster Analysis of Student Motivation

Several studies of multiple populations have utilized CA to analyze and characterize student motivation and learning [21], [29], [30], and some specifically Future Time Perspective (FTP) [1], [2], of engineering undergraduate students. In particular, some studies have utilized the Motivation and Attitudes in Engineering (MAE) to cluster undergraduate engineers [31]–[33] and have discussed results where three characteristics future views of undergraduate engineers have been shown: *sugar* students with a clear future view; *waffle* students with conflicting ideal and realistic futures; and *cake* with open views of the future. Several quantitative studies cluster first and second year undergraduate engineering students based on their FTPs [6], [32], [33] typically seeing three groups:

Group 1: high F, PI, and FoP scores (sugar)

Group 2: lower F, PI, FoP scores than Group 1 and a low PI score overall (waffle)

Group 3: lower future scores, high PI scores, and overall low FoP scores (cake)

While k-means has primarily been used to identify homogeneous groups of engineering students in terms of their motivation and/or learning attributes, this paper seeks to select the most appropriate CA method and will compare both hierarchical and partitioning methods. The chapter specifically includes the solutions from the Ward's and k-means clustering algorithm to select the most fitting cluster solution. The results will be used for participant selection in future chapters.

Methods

Motivation and Attitudes in Engineering Survey

The MAE survey [7], [8] consists of 5 sections with 86 items related to goal orientation [34], FTP and Expectancy (E), task specific metacognition, problem-solving self-efficacy [35], and demographic information. This paper presents a CA of the domain- and context-specific Future Time Perspective (FTP) items utilizing the FTP and Expectancy section. The FTP items contain five theoretical factors: Perceived Instrumentality (PI), Perceptions of the Future (F), Future on Present (FoP), Value (V), and Connectedness (C). The Value and Connectedness items, adapted from Husman and Shell [1], [12], were added based on previous qualitative FTP work [7], [32], [33]. Other items were original and based on findings from prior qualitative studies [7], [32], [33], or adapted from the Motivated Strategies for Learning Questionnaire (MSLQ) [36], [37].

Items in the FTP and E section were 7-point Likert-type items with anchors "0-Strongly Disagree" and "6-Strongly Agree" [38] as anchored scales make statistical testing more valid, and allows for an easier interpretation of numeric responses [39]. Normalization was not necessary as all items were on the same scale. E items for this population are typically high and generally rank the same on a Likert scale across clusters as students in engineering have high hopes in their coursework [33]. As such, E will not be included in the CA as it does not help to differentiate students. Additionally, this research focuses on domain-specific (F, FoP, PoF) and context-specific (PI) FTP constructs.

Participants

The MAE survey was distributed in class and submitted online by students enrolled in one section of a sophomore-level materials science and engineering (MSE) course required for industrial engineering (IE), BME, and ME undergraduates at a four-year, land grant institution in the southeast (n=97). Additionally, the survey was completed in one section of a required, sophomore-level IE course (n=205) during the same semester. Both sets of students received class credit for completing the survey during class time. Prior to merging of the two groups, they were compared using robust statistical analysis (Fisher's Exact and Chi-squared tests) to ensure no differences in the two samples existed.

Exploratory Factor Analyses

An exploratory factor analysis (EFA) was conducted to assess the latent correlation structure of the survey items. This analysis validated new items that were added to the MAE (C and V) and validated the survey for a new population. Prior to the EFA, incomplete entries were listwise deleted. A total of N=223 completed entries were used for the EFA and subsequent analysis. A scree plot test [40], [41], and the FTP literature were used to determine the appropriate number of factors. Eigenvalues of the correlation matrix using a promax rotation [42] were plotted in a scree plot (Figure 1). A promax rotation of factors allows factors to be correlated, provides the simplest solution, and permits items to load into one, and only one, factor[43], [44]. The data's skew (absolute value not higher than 2) and kurtosis (value not higher than 7) were evaluated to assure assumptions of multivariate normality were met [45]. Items that had a factor loading below 0.4 during the EFA were removed [46]. In addition to an overall Chi-squared test (non-significant at $p<0.05$), the root mean square error of approximation (RMSEA) was calculated to test model fit [47]. After the EFA was completed, Cronbach's alpha [48] was used to confirm the internal consistency of the factors [49].

Cluster Analyses

A CA of the FTP constructs was conducted in order to group homogenous participants into k subgroups, or k clusters [50]. The data for the CA consisted of composite FTP survey scores (F, PI, FoP) for each participant. Two types of CA were compared: Ward's hierarchical and k-means. The most common and generally accepted distance measurement, Euclidean distance [22], [51], was used and is standard for Likert-type data. All data was analyzed using R [52] unless otherwise specified.

Results and Discussion

Aggregation of Data Sets

First, the MSE and IE data sets were cleaned by eliminating any participants who did not appear to complete the survey (list-wise deletion). Some students (N=8) were registered in both the MSE and IE courses and were removed from the IE sample so the MSE data may be used for future participant selection. Responses to each item of the eight students were compared to the remaining IE group using the statistical software JMP [53] as it runs comparisons of every item at once. Results of the comparisons indicated only one item, C40 ("It's not really important to have future goals for where one wants to be in five or ten years.") was statistically different for the two groups. This item was deleted for all future analysis. Since all other items for the survey section did not appear to be different for both groups, the responses from IE course of the students who were enrolled in both the IE and MSE course were deleted. The students' MSE responses were still included in the main data.

To merge the remaining data, the two classes were compared. JMP was utilized to run Pearson's Chi-squared test [54] to test for significant differences between items' scores for both groups. The tests were not statistically significant, and the null hypothesis was not rejected for any of our comparisons, allowing our data to be aggregated.

Exploratory Factor Analysis

An EFA was conducted on the cleaned responses (N=223) to items in the FTP and E section of the MAE survey. For this EFA, items using negative language (FoP21, PI26, V30, C36, C39, C40, C41, C43, C46) were reverse scored [55] and a scree plot was created (Figure 1) to select the number of factors.

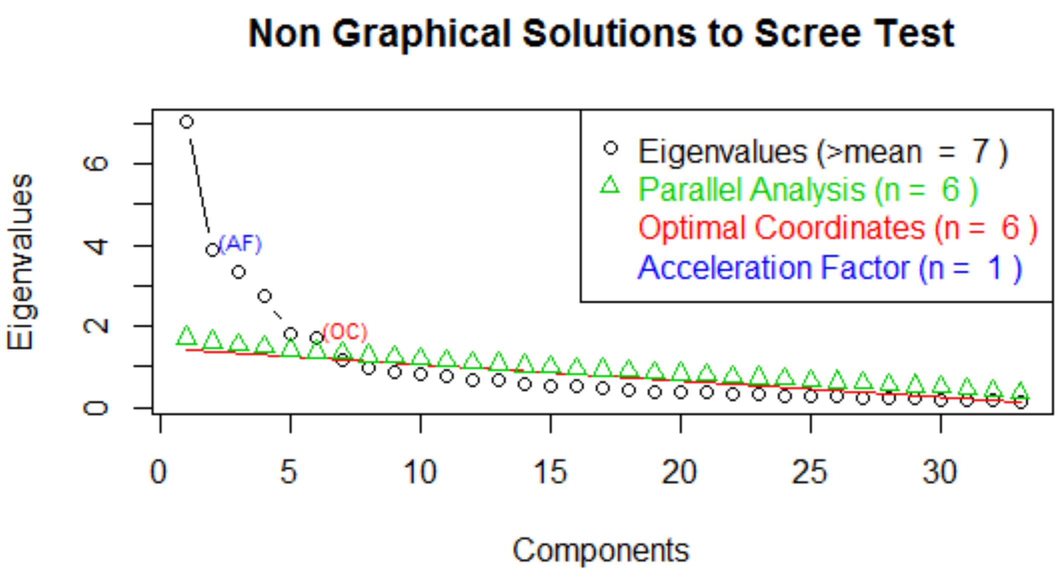

Figure 1: Scree plot for entire section of MAE including FTP and E items.

According to the scree plot, six factors is optimal, agreeing with the literature. Skewness ranged between -1.821 and -.252. The kurtosis ranged from 2.124 and 6.240. Both sets of scores indicated some non-normality but were within the level of acceptability for EFA or maximum likelihood factor analysis [44], [56]–[58]. Detailed standardized factor loadings for each item may be seen in Appendix A. Although the Chi-square statistic for this section of the survey was 499.04 (p-value = 7.54×10^{-16}, 270 degrees of freedom) was statistically significant (i.e. the six factors are not an ideal fit), the Root Mean Square Error of Approximation (RMSEA=0.0927) indicated an acceptable fit [46], [47]. Since the RMSEA was in the acceptable range, and previous studies support the six-factor model, six factors were selected. As the domain- and context-specific constructs (PI, FoP, F) have been shown to be valid for similar populations, the lack of goodness of fit was likely due to the new domain-general factors, which were not utilized in the cluster analyses.

Items that did not meet the following criteria were removed from the analysis: Item reliability (R2) ≥ 0.50, Construct Reliability ≥ 0.70, and Average Variance Extracted ≥ 0.50 [44], [56]–[58]. Additionally, Cronbach's alpha was calculated for each construct and was determined to be between 0.8 and 0.91, indicating strong internal consistency for the remaining items in each construct [59], [60]. FTP construct name, survey item number, item wording, final standardized factor loadings, uniqueness, item reliability, and construct reliability can be found in Table A1 in the Appendix and a summary of final items and factors are included in Table 1.

Table 1: The final MAE survey factors used for analysis

Factor Number	Factor Name	Number of items	High Score Definition	Factor Items	Alpha (α)
1	Connectedness (C)	7	The person plans and thinks about what they want to do in the future.	C36, C37, C38, C39, C41, C43, C45	0.86
2	Expectancy (E)	5	The person has expectations of success.	E24, E25, E27, E28, E29	0.91
3	Value (V)	5	The person believes that goals attainable in the future are important.	V31, V32, V33, V34, V35	0.82
4	Perceived Instrumentality (PI)	5	The student views what they are doing in the present as useful.	PI14, PI19, PI20, PI26, FoP21	0.82
5	Perceptions of the future (F)	4	The student has a positive and clear outlook about the future.	F15, F16, F17, F18	0.84
6	Effect of future on present (FoP)	2	The student believes the future has a high impact on what the student does in the present.	FoP22, FoP23	0.80

Cluster Analysis

<u>Hierarchical Cluster Analysis</u>

Participants were removed (N=4) who had missing responses in the domain-specific FTP factors (PI, FoP, F). Composite scores of the factors were created so that each participant had a single score for each factor, and Euclidean distances were used to determine the distance between participants. Multiple hierarchical clustering algorithms were run and dendrograms created. Ward's appeared to be a strong candidate for this data and was selected for additional analysis.

<u>Ward Clustering Algorithm</u>

A clustering dendrogram for Ward's (Figure 2) along with two additional plots, graphs plotting between sum of squares error (bss, an estimate of the distance between clusters, should be high) and within sum of squares error (wss, an estimate of the distance between points within a cluster, should be low) (Figure 3), were created to determine the appropriate number of clusters. The significant height difference between the "trees" in the clustering dendrogram (illustrated by the dashed line) in Figure 2 supports k=3. The two "elbows" of the bss and wss in Figure 3 (illustrated by the circles) show k=3 as an ideal clustering solution. Agreement between the dendrogram, the wss plot, the bss plot, and previous literature [31]–[33] show that a three cluster solution is likely. When selecting k=3, the total sum of squares is 991.00; total within sum of squares is 527.11; and between sum of squares is 463.89. The average scores and standard deviations for each factor (F, PI, and FoP) are detailed in Table 2, as well as the size of each cluster. A visual representation of the Ward's cluster analysis for k=3 can be seen in Figure 4.

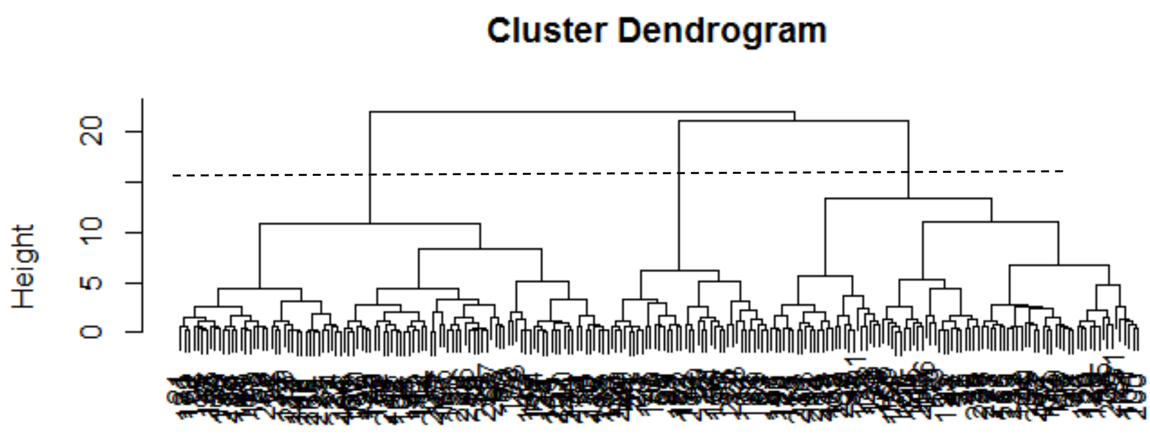

Figure 2: Ward's CA (ward.D2 in R) Dendrogram depicting three distinct clusters for the domain- and context-specific FTP items of the MAE survey for the IE and MSE responses

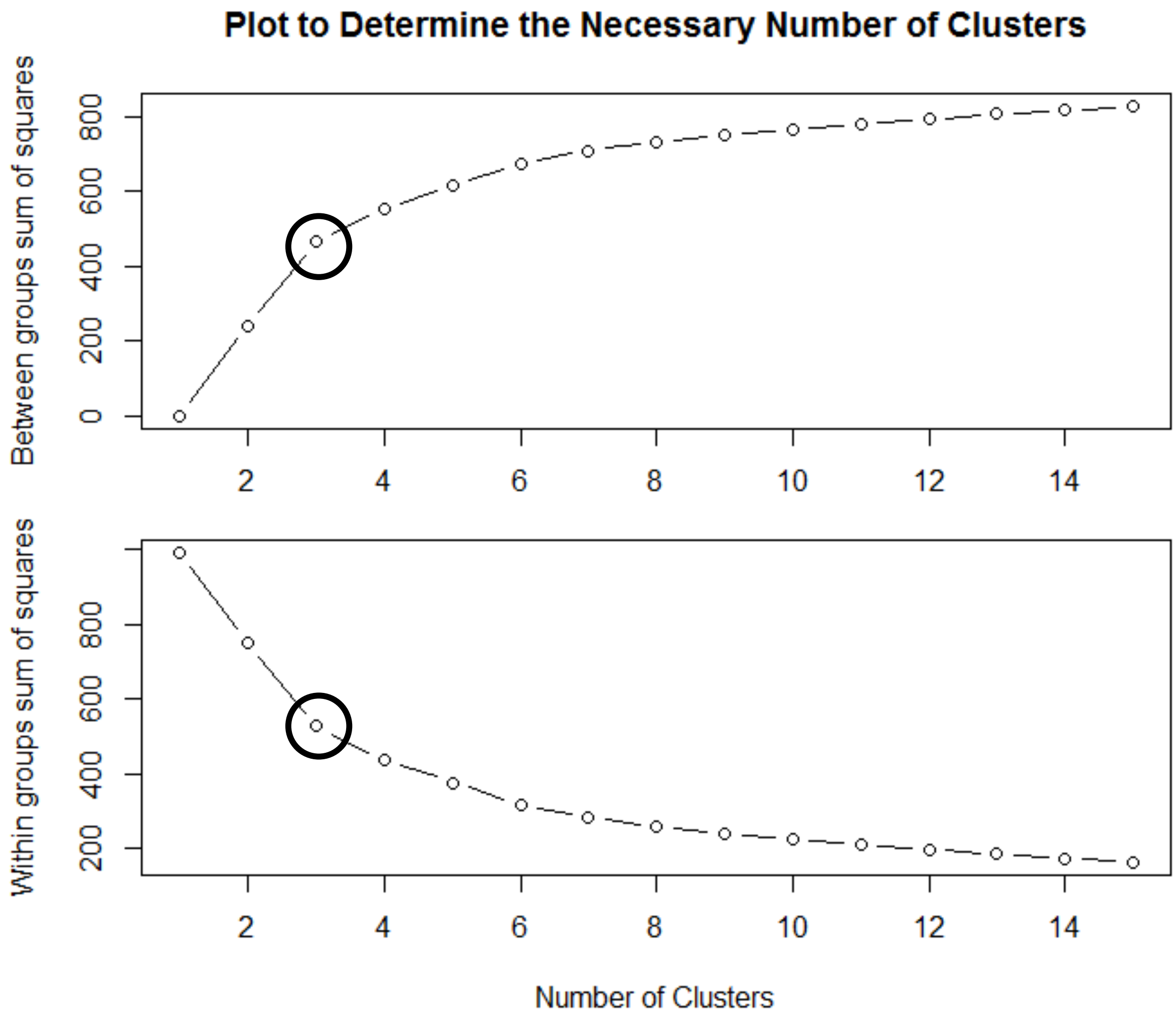

Figure 3: Ward's CA (ward.D2 in R) plots of "Between group sum of squares", and "Within group sum of squares", both depicting a three cluster solution for the domain- and context-specific FTP items of the MAE survey for the IE and MSE responses

Table 2: Clusters and average cluster variable scores for three variable Ward's CA in R with k=3

Cluster	N	Perceptions of the Future	Perceived Instrumentality	Future on Present	Cluster Type
1	86	5.01 ± 1.10	4.56 ± 0.82	4.08 ± 1.21	Waffle
2	100	6.12 ± 0.76	6.18 ± 0.62	5.07 ± 0.93	Sugar
3	37	6.14 ± 0.82	6.32 ± 0.63	2.03 ± 0.79	Cake

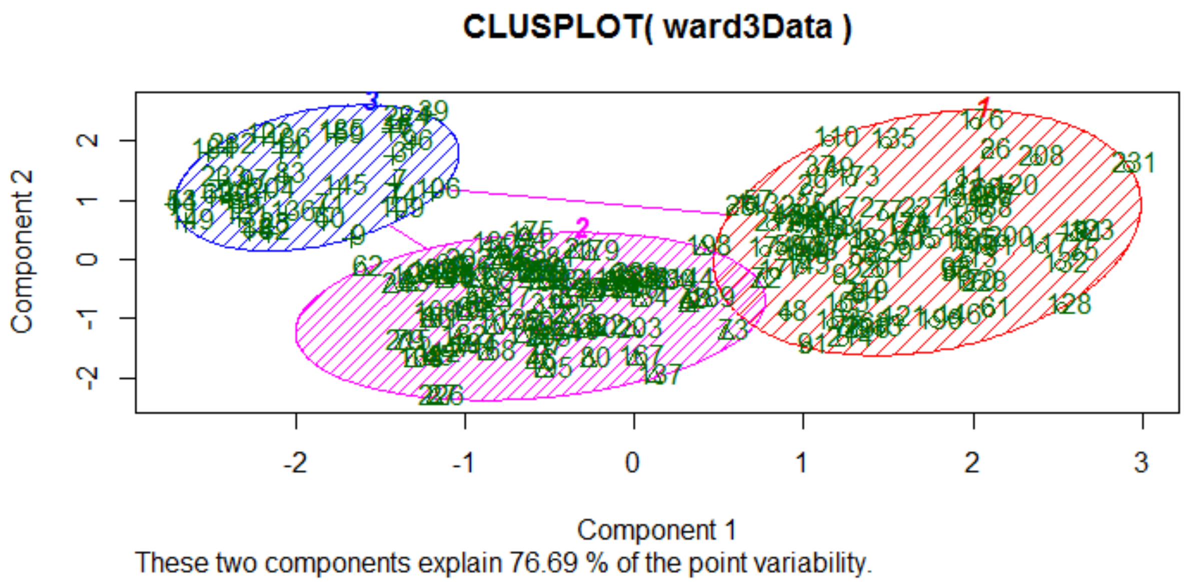

Figure 4: A Ward's three cluster solution two-dimensional visual representation using CLUSPLOT in R explaining 76.69% of the point variability

Cluster scores were first compared using a MANOVA and then ANOVA analysis on each construct to ensure there were differences between groups prior to pairwise comparisons. By running a MANOVA and ANOVA prior to pairwise comparisons, we create a "protected inference" situation, preventing the inflated Type I error that can occur if only multiple t-tests are used [40]. MANOVA and ANOVA results indicated statistical significance with the largest p-value being $p = 2.72x10^{-15}$. Pairwise t-tests were run to look for significant differences between factor scores for each cluster. Clusters 1 and 2 differ significantly in terms of F, PI, and FoP ($p = 1.4x10^{-14}$, $p < 2.0x10^{-16}$, $p = 5.0x10^{-10}$, respectively). Additionally, Clusters 1 and 3 differ in the F, PI, and FoP constructs ($p = 1.9x10^{-9}$, $p < 2.0x10^{-16}$, $p < 2.0x10^{-16}$, respectively). However, Clusters 2 and 3 only differ on the FoP construct ($p < 2.0x10^{-16}$).

<u>K-means Clustering Algorithm</u>

A scree plot of wss was created and the elbow (k=3) used to select the number of clusters (Figure 5). The tss, wss, and bss are 991.00, 482.81, and 508.19, respectively. The k=2, 3, and 4 clustering solutions were run for testing purposes and k=3 appeared to be the best fit, with the least overlap in clusters and tightest cluster solution. Table 3 shows the k-means three cluster solution, and Figure 6 displays a two-dimensional visual representation explaining 76.4% of the point variability. Three dense clusters, with few outliers and little to no overlap are shown. Cluster scores were compared using a MANOVA and then ANOVA analysis to ensure there were differences between groups prior to pairwise comparisons. MANOVA and ANOVA results indicated statistical significance with all tests reporting $p < 2.0x10^{-16}$. Pairwise t-tests showed significant differences between all three FTP factors, F, PI, and FoP, for Clusters 1 and 2 ($p < 2.0x10^{-16}$, $p = 1.2x10^{-11}$, $p < 2.0x10^{-16}$, respectively) and between Clusters 1 and 3 ($p < 2.0x10^{-16}$, $p < 2.0x10^{-16}$, $p = 5.3x10^{-10}$, respectively). However, Clusters 2 and 3 only differed significantly in student views of the impact of the future on the present (FoP, $p < 2.0x10^{-16}$).

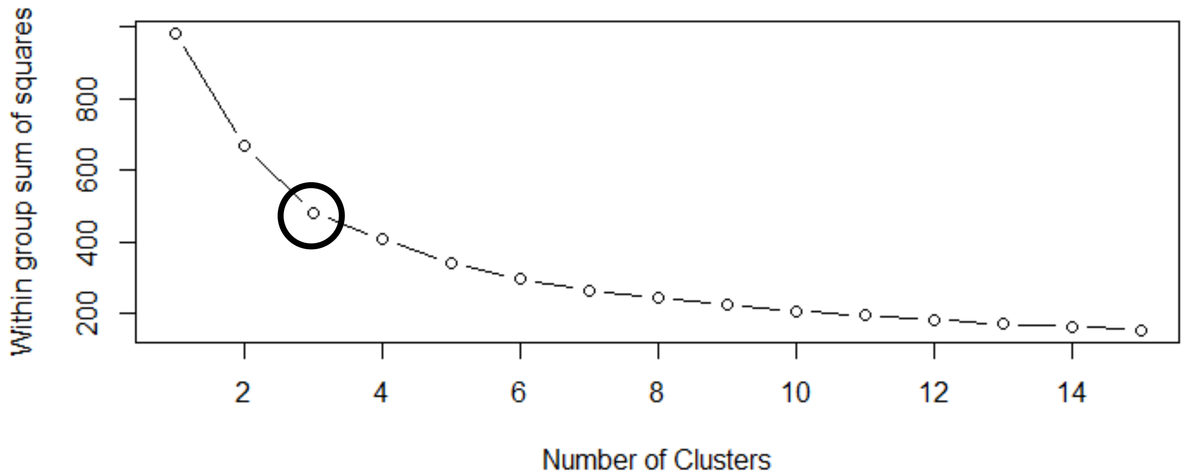

Figure 5: K-means CA wss plot, depicting a three-cluster solution for the domain- and context-specific FTP items of the MAE survey for the IE and MSE responses

Table 3: Clusters and average cluster variable scores for three variable k-means CA in R with k=3

Cluster	N	Perceptions of the Future	Perceived Instrumentality	Future on Present	Cluster Type
1	56	4.48 ± 0.98	4.60 ± 0.96	4.26 ± 0.83	Waffle
2	58	5.99 ± 0.83	5.83 ± 1.03	2.25 ± 0.83	Cake
3	109	6.17 ± 0.66	5.95 ± 0.82	5.17 ± 0.88	Sugar

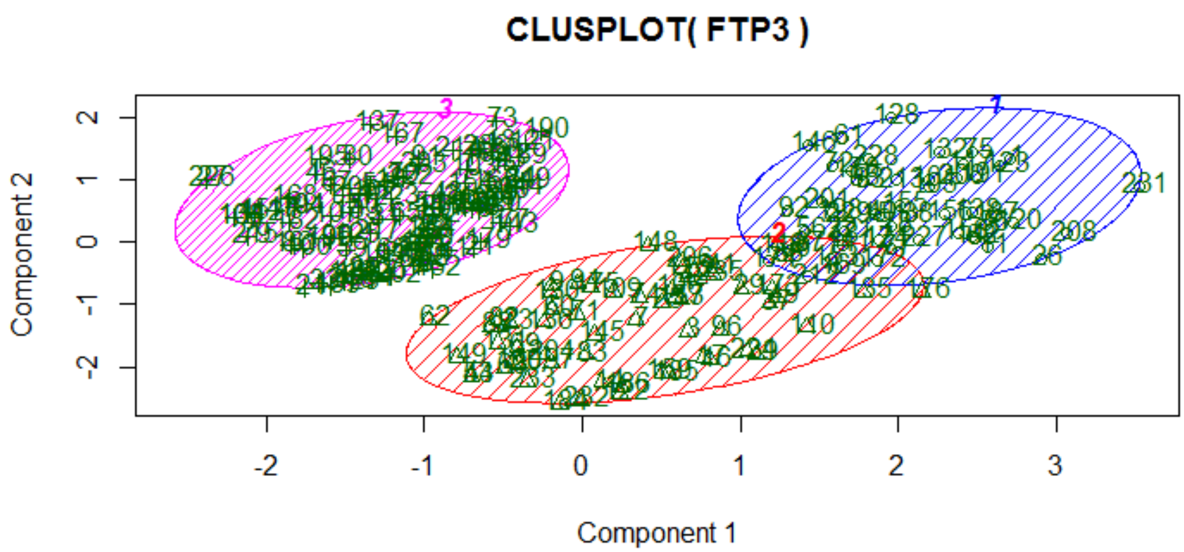

Figure 6: A k-means three cluster solution two-dimensional visual representation using CLUSPLOT in R explaining 76.4% of the point variability

Selection of Clustering Algorithm

The k-means solution appears extremely similar to the Ward's clustering solution in mean scores. Ward's Cluster 1 and k-means Cluster 1 have medium scores; Ward's Cluster 2 and k-means Cluster 3 have high average scores; and Ward's Cluster 2 and k-means Cluster 3 have high perceptions of the future and PI scores but low FoP scores. Additionally, the t-tests showed the same dissimilarities between the clustering solutions in regards to average scores across clusters. However, the size (number of participants) of these "matching" clusters appear different for the two methods. The difference in the number of participants in each cluster is likely due to the fact that the k-means algorithm will reconsider participants during the iterations, while Ward's CA algorithm does not allow for a change in cluster once a data point has been placed into a cluster.

The goodness of fit scores showed that k-means was the best fit for this set of data as the wss was smaller, showing more compact clusters, and the bss was larger, showing a more distinct clustering solution. Additionally, prior results and theory support a three-cluster solution is for MAE survey data. Due to the lower wss and higher bss value, the match between results and theory, the fitting visual representation, and the robustness of the method, the k-means clustering solution was selected.

Considering the final clusters in Table 5, the results aligned with previous research looking at homogeneous groupings of mid-year engineering majors at similar institutions [6], [32], [33]. Cluster 3 mimicked the all high scores (high F, PI, FoP) of previously documented sugar students, reflecting clearly developed future goals, high sense of instrumentality of current coursework, and a feedback loop between future goals and present actions. Cluster 1 (waffle students), featured lower average F, PI, and FoP scores than Cluster 3 (sugar students). The waffle students' scores appear lower on average due to their (often two) conflicting views of the future and thus less concrete sense of instrumentality (lower PI) and impact of the future on their present actions (lower FoP). Finally, Cluster 2 encompassed cake students, who have extremely open, but positive, views for the future, thus decreasing their perception of the impact of the future on their present actions (low FoP score). However, these students often realize, which is clear in the clustering solution, that their current engineering coursework will be important toward their future (high PI). Similarly, Cluster 3 (sugar) held the highest number of students, which also occurs in previous studies.

Conclusions, Future Work, and Limitations

This paper meets our two goals of (1) identifying homogeneous groups of second-year engineering student FTPs and (2) introducing commonly used cluster analysis techniques and providing an example of how to implement said techniques within an engineering education context. The best cluster analysis method is heavily dependent on the data set, so ensuring documentation of the rationale during analysis is necessary. In our analysis, each step allowed the data to drive our choices with support from theory. While Ward's and k-means provided similar solutions with strong results, k-means provided a more robust measure, possibly because it allowed for movement of data between iterations. The goodness of fit scores showed that k-means was the best fit for this set of data. Visually, both methods appeared to have achieved distinct clusters with strong clustering solutions. While the results from each CA were not identical, similar solutions did occur with enough evidence to select three clusters and to group similar students together. Similar average scores for each cluster were achieved.

The theory of three FTPs (domain- and context-specific) of undergraduate engineers appears to fit this population, and the clustering solution was similar to previous research. The clustering solution from this work will be utilized as a piece of participant selection for future qualitative studies. The MAE survey currently considers endogenous, and not exogenous, PI, but waffle students appear to have a high exogenous PI. For this survey, waffle students have all medium scores, all high for sugar, and high for F & PI but low FoP for cake. However, a future version of the MAE should include exogenous PI items to provide a better measure of the PI of engineering students. Moreover, PI is a domain-specific way to look at connectedness; the MAE survey would provide a stronger look at domain-specific FTP by including domain-specific value/valence items to complete this picture. Furthermore, while this work analyzed three domain- and context-specific FTP constructs (F, FoP, and PI), future work should look at a more domain-general approach to FTP and the distinctions between undergraduate engineers. Additional work could include looking at different majors, transfer versus non-transfer students, genders, and other classifications.

In future work, the MAE survey should be further studied for this population. The C and V constructs should be evaluated and refined utilizing student focus groups and other means. Additionally, item FoP21, originally intended to be included in the FoP construct, loaded into PI in this analysis. The negative language "I do not connect my future career with what I am learning in this course" lends itself to the PI construct and may have confused students. The item should be altered or reconsidered for future analysis.

This work is particularly valuable as FTP has been shown, quantitatively and qualitatively, to have an impact on goal-setting and metacognitive strategies in the present [2], [6], [10]–[13], as stated in the Background section. By clustering students into homogeneous groups, practitioners can better understand students' goals, perceptions of their future, and the perceived utility of class content. By understanding these aspects of FTP, practitioners can better motivate students within their class by customizing course instruction and materials reflective of their students' future goals. With this additional motivation, students are more likely to use self-regulatory study strategies and behaviors, which has been shown to be a positive predictor of classroom success [61]–[64].

References

[1] J. Husman and D. F. Shell, "Beliefs and perceptions about the future: A measurement of future time perspective," *Learn. Individ. Differ.*, vol. 18, no. 2, pp. 166–175, 2008.

[2] S. E. Tabachnick, R. B. Miller, and G. E. Relyea, "The relationships among students' future-oriented goals and subgoals, perceived task instrumentality, and task-oriented self-regulation strategies in an academic environment.," *J. Educ. Psychol.*, vol. 100, no. 3, pp. 629–642, 2008.

[3] C. Gutiérrez-Braojos, "Future time orientation and learning conceptions: effects on metacognitive strategies, self-efficacy beliefs, study effort and academic achievement," *Educ. Psychol.*, vol. 3410, no. March 2014, pp. 1–21, 2013.

[4] J. de Bilde, M. Vansteenkiste, and W. Lens, "Understanding the association between future time perspective and self-regulated learning through the lens of self-determination theory," *Learn. Instr.*, vol. 21, no. 3, pp. 332–344, 2011.

[5] L. Benson, B. Morkos, and J. Husman, "Motivation of first year engineering students and its relation to persistence in engineering," 2013.

[6] J. Chasmar and L. Benson, "Future time perspective and self-regulated learning: Multiple case studies in industrial engineering," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2016, vol. 2016–June.

[7] A. Kirn and L. Benson, "Quantitative assessment of student motivation to characterize difference between engineering majors," in *Frontiers in Education Conference*, 2013, pp. 69–74.

[8] C. Faber, S. Grigg, A. Kirn, J. Chasmar, and L. Benson, "Engineering Student Motivation and Perceived Metacognition in Living-Learning Communities," in *121st ASEE Annual Conference and Exposition*, 2014.

[9] K. M. Ehlert, C. J. Faber, M. S. Kennedy, and L. Benson, "Utilizing cluster analysis of close-ended survey responses to select participants for qualitative data collection," in *ASEE Annual Conference and Exposition, Conference Proceedings*, 2017, vol. 2017–June.

[10] J. Simons, M. Vansteenkiste, W. Lens, and M. Lacante, "Placing motivation and future time perspective theory in a temporal perspective," *Educ. Psychol. Rev.*, vol. 16, no. 2, pp. 121–139, 2004.

[11] S. E. Tabachnick, "The impact of future goals on students' proximal subgoals and on their perceptions of task instrumentality," University of Oklahoma, 2005.

[12] J. C. Hilpert, J. Husman, G. S. Stump, W. Kim, W.-T. Chung, and M. A. Duggan, "Examining students' future time perspective: Pathways to knowledge building," *Jpn. Psychol. Res.*, vol. 54, no. 3, pp. 229–240, 2012.

[13] J. Chasmar, “Connections between future time perspectives and self-regulated learning for mid-year engineering students : a multiple case study,” Clemson University, 2017.

[14] M. L. de Volder and W. Lens, “Academic achievement and future time perspective as a cognitive-motivational concept.,” *J. Pers. Soc. Psychol.*, vol. 42, no. 3, pp. 566–571, 1982.

[15] J. Husman and W. Lens, “The role of the future in student motivation,” *Educ. Psychol.*, vol. 34, no. 2, pp. 113–125, 1999.

[16] M. Bong, “Between- and within-domain relations of academic motivation among middle and high school students: Self-efficacy, task value, and achievement goals.,” *J. Educ. Psychol.*, vol. 93, no. 1, pp. 23–34, 2001.

[17] J. O. Raynor, “Future orientation and achievement motivation: Toward a theory of personality functioning and change,” in *Cognition in Human Motivation and Learning*, G. D’Ydewalle and W. Lens, Eds. Leuven, Belgium & Hillsdale, NJ: Leuven University Press & Lawrence Erlbaum Associates, Inc., 1981, pp. 199–231.

[18] J. Husman, W. Pitt Derryberry, H. Michael Crowson, and R. Lomax, “Instrumentality, task value, and intrinsic motivation: Making sense of their independent interdependence,” *Contemp. Educ. Psychol.*, vol. 29, no. 1, pp. 63–76, Jan. 2004.

[19] J. Husman, J. C. Hilpert, and S. K. Brem, “Future Time Perspective Connectedness to a Career: The Contextual Effects of Classroom Knowledge Building,” *Psychol. Belg.*, vol. 56, no. 3, pp. 210–225, 2016.

[20] L. Kaufman and P. J. Rousseeuw, *Finding Groups in Data: An Introduction to Cluster Analysis*, vol. 33, no. 1. 2005.

[21] D. F. Shell and L. K. Soh, “Profiles of Motivated Self-Regulation in College Computer Science Courses: Differences in Major versus Required Non-Major Courses,” *J. Sci. Educ. Technol.*, vol. 22, no. 6, pp. 899–913, 2013.

[22] A. K. Jain, M. N. Murty, and P. J. Flynn, “Data Clustering : A Review,” *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.

[23] A. K. Jain, “Data clustering: 50 years beyond K-means,” *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.

[24] J. H. Ward, “Hierarchical grouping to optimize an objective function,” *Journal of the American Statistical Association*, vol. 58, no. 301. pp. 236–244, 1963.

[25] P. D. Antonenko, S. Toy, and D. S. Niederhauser, “Using cluster analysis for data mining in educational technology research,” *Educ. Technol. Res. Dev.*, vol. 60, no. 3, pp. 383–398, 2012.

[26] R. Tibshirani, G. Walther, and T. Hastie, “Estimating the number of clusters in a data set via the gap statistic,” *J. R. Stat. Soc. Ser. B (Statistical Methodol.*, vol. 63, pp. 411–423,

2001.

[27] B. S. Everitt and T. Hothorn, *A Handbook of Statistical Analyses using R*, vol. 57, no. 2. 2003.

[28] A. Jackson and N. Mentzer, “Cluster Analysis in Engineering Education,” in *ASEE Annual Conference and Exposition*, 2017.

[29] K. G. Nelson, D. F. Shell, J. Husman, E. J. Fishman, and L.-K. Soh, “Motivational and Self-Regulated Learning Profiles of Students Taking a Foundational Engineering Course,” *J. Eng. Educ.*, vol. 104, no. 1, pp. 74–100, 2015.

[30] D. F. Shell and J. Husman, “Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon,” *J. Educ. Psychol.*, vol. 100, no. 2, pp. 443–459, 2008.

[31] J. Chasmar and L. Benson, “Future Time Perspective and Self-Regulated Learning: Multiple Case Studies in Industrial Engineering,” in *American Society for Engineering Education*, 2016.

[32] A. Kirn and L. Benson, “Engineering Students ’ Perceptions of the Future: Exploratory Instrument Development,” in *122nd ASEE Annual Conference & Exposition*, 2015.

[33] A. N. Kirn, “The Influences of Engineering Student Motivation on Short-Term Tasks and Long-Term Goals,” 2014.

[34] G. Schraw, C. Horn, T. Thorndike-Christ, and R. Bruning, “Academic goal orientations and student classroom achievement,” *Contemporary Educational Psychology*, vol. 20, no. 3. pp. 359–368, 1995.

[35] A. Mason and C. Singh, “Surveying graduate students’ attitudes and approaches to problem solving,” *Phys. Rev. Spec. Top. - Phys. Educ. Res.*, vol. 6, no. 2, p. 20124, 2010.

[36] P. R. Pintrich and T. Garcia, “The Motivated Strategies for Learning Questionnaire,” in *Annual Meeting of the American Educational Research Association*, 1995.

[37] P. R. Pintrich, D. A. F. Smith, T. Garcia, and W. J. McKeachie, “A Manual for the Use of the Motivated Strategies for Learning Questionnaire (MSLQ),” Ann Arbor, MI, 1991.

[38] J. a. Krosnick and S. Presser, “Question and Questionnaire Design,” in *Handbook of Survey Research*, 2nd ed., Emerald Group Publishing Limited, 2010, p. 886.

[39] T. R. Knapp, “Treating ordinal scales as interval scales: an attempt to resolve the controversy,” *Nurs. Res.*, vol. 39, no. 2, pp. 121–123, 1990.

[40] A. C. Rencher and W. F. Christensen, *Methods of Multivariate Analysis*, 3rd ed. Hoboken, NJ: John Wiley & Sons, 2012.

[41] R. B. Cattell, “The scree test for the number of factors,” *Multivariate Behav. Res.*, vol. 1, no. 2, pp. 245–276, 1966.

[42] G. Raiche, “nFactors: an R package for parallel analysis and non graphical solutions to the Cattell scree test.” 2010.

[43] J. W. Osborne, “What is Rotating in Exploratory Factor Analysis ?,” *Pract. Assessment, Res. Eval.*, vol. 20, no. 2, pp. 1–7, 2015.

[44] A. B. Costello and J. W. Osborne, “Best Practices in Exploratory Factor Analysis: Four Recommendations for Getting the Most From Your Analysis,” *Pract. Assess. Res. Eval.*, vol. 10, no. 7, pp. 1–9, 2005.

[45] P. J. Curran, S. G. West, and J. F. Finch, “The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis.,” *Psychol. Methods*, vol. 1, no. 1, pp. 16–29, 1996.

[46] B. Thompson, “Exploratory and confirmatory factor analysis: Understanding concepts and applications.” American Psychological Association, Washington, DC, 2004.

[47] R. C. MacCallum, M. W. Browne, and H. M. Sugawara, “Power analysis and determination of sample size for covariance structure modeling,” *Psychol. Methods*, vol. 1, no. 2, pp. 130–149, 1996.

[48] L. J. Cronbach, “Coefficient alpha and the internal structure of tests,” *Psychometrika*, vol. 16, no. 3, pp. 297–334, 1951.

[49] M. Tavakol and R. Dennick, “Making sense of Cronbach’s alpha,” *Int. J. Med. Educ.*, vol. 2, pp. 53–55, 2011.

[50] D. Witten, G. James, R. Tibshirani, and T. Hastie, *An Introduction to Statistical Learning: with Applications in R*. New York: Springer, 2013.

[51] B. S. Everitt, S. Landau, and M. Leese, *Cluster analysis*, 4th ed. London: Arnold, 2009.

[52] R Core Team, “R: A language and environment for statistical computing,” *R Foundation for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2016.

[53] JMP, “Version 12.” SAS Institute Inc., Cary, NC, pp. 1989–2007.

[54] K. Larntz, “Small-Sample Comparisons of Exact Levels for Chi-Squared Goodness-of-Fit Statistics,” *J. Am. Stat. Assoc.*, vol. 73, no. 362, pp. 253–263, 1978.

[55] R. F. DeVellis, *Scale Development: Theory and Applications*, 3rd ed. SAGE Publications, Inc, 2012.

[56] A. Godwin, “The Development of a Measure of Engineering Identity,” *123rd Am. Soc. Eng. Educ. Annu. Conf. Expo.*, p. 15, 2016.

[57] W. Dillon and M. Goldstein, *Multivariate Analysis: Methods and Applications*. New York: Wiley, 1984.

[58] J. Hair, R. Anderson, R. Tatham, and W. Black, *Multivariate Data Analysis*, 5th ed. Upper Saddle River, NJ: Prentice Hall, 1998.

[59] J. A. Gliem and R. R. Gliem, "Calculating, interpreting, and reporting Cronbach's alpha reliability coefficient for Likert-type scales," *Midwest Res. to Pract. Conf. Adult, Contin. Community Educ.*, no. 1992, pp. 82–88, 2003.

[60] D. George and P. Mallery, *SPSS for Windows step by step: A simple guide and reference. 11.0 update*, 4th ed. Boston, MA: Allyn & Bacon, 2003.

[61] P. R. Pintrich and E. V. de Groot, "Motivational and Self-Regulated Learning Components of Classroom Academic Performance," *J. Educ. Psychol.*, vol. 82, no. 1, pp. 33–40, 1990.

[62] B. J. Zimmerman and M. Martinez-Pons, "Development of a Structured Interview for Assessing Student Use of Self-Regulated Learning Strategies," *Am. Educ. Res. J.*, vol. 23, no. 4, pp. 614–628, 1986.

[63] M. Credé and L. A. Phillips, "A meta-analytic review of the Motivated Strategies for Learning Questionnaire," *Learn. Individ. Differ.*, vol. 21, no. 4, pp. 337–346, 2011.

[64] D. H. Schunk and B. J. Zimmerman, Eds., *Self-regulated learning: From teaching to self-reflective practice*. Guilford Press, 1998.

Appendix A

Table A1: Exploratory Factor Analysis Results and Cronbach's Alpha for FTP and E Items on the MAE Survey (N=223)

Construct	Item #	Item	Final Standardized Factor Loadings	Uniqueness	Item Reliability (R^2)	Construct Reliability
Connectedness	C36	I don't think much about the future.	0.79	0.44	0.84	0.86
	C37	I have been thinking a lot about what I am going to do in the future.	0.67	0.45	0.85	
	C38	What will happen in the future is an important consideration in deciding what action to take now.	0.63	0.47	0.85	
	C39	I don't like to plan for the future.	0.88	0.41	0.84	
	C41	One shouldn't think too much about the future.	0.75	0.46	0.84	
	C42	It is important to have goals for where one wants to be in five or ten years.	0.61	0.62	0.85	
	C43	Planning for the future is a waste of time.	0.66	0.52	0.84	
	C45	One should be taking steps today to help realize future goals.	0.48	0.59	0.85	
	C46	What might happen in the long run should not be a big consideration in making decisions now.	0.44	0.65	0.87	
	C40	Item removed from analysis due to negative Chi-squared test of comparisons	NA	NA	NA	
Expectancy	E24	I expect to do well in this engineering course.	0.78	0.38	0.84	0.91
	E25	I am certain I can master the skills being taught in this engineering course.	0.75	0.41	0.82	
	E27	I believe I will receive an excellent grade in this engineering course	0.92	0.18	0.91	

Construct	Item #	Item	Final Standardized Factor Loadings	Uniqueness	Item Reliability (R^2)	Construct Reliability
	E28	I am confident I can do an excellent job on the assignments in this engineering course.	0.91	0.19	0.91	
	E29	Considering the difficulty of this engineering course, the teacher, and my skills, I think I will do well in this engineering course.	0.76	0.39	0.82	
Perceptions of the Future	F15	I am confident about my choice of major.	0.51	0.65	0.69	0.84
	F16	Engineering is the most rewarding future career I can imagine for myself.	0.86	0.26	0.88	
	F17	My interest in an engineering major outweighs any disadvantages I can think of.	0.79	0.30	0.87	
	F18	I want to be an engineer.	0.8	0.37	0.83	
Future on Present	FoP22	My future career determines what is important in this course.	0.78	0.41	0.9	0.80
	FoP23	My future career influences what I learn in this course.	0.89	0.26	0.92	
Perceived Instrumentality	PI14	I will use the information I learn in my engineering course in other classes I will take in the future	0.81	0.43	0.71	0.82
	PI19	I will use the information I learn in this engineering course in the future.	0.96	0.18	0.82	
	PI20	What I learn in my engineering course will be important for my future occupational success.	0.79	0.31	0.78	
	PI21 (FoP21)	I do not connect my future career with what I am learning in this course.	0.45	0.58	0.78	
	PI26	I will not use what I learn in this engineering course.	0.45	0.57	0.77	

Construct	Item #	Item	Final Standardized Factor Loadings	Uniqueness	Item Reliability (R^2)	Construct Reliability
Value/Valence	V30	Immediate pleasure is more important than what might happen in the future.	NA	NA	NA	
	V31	It is better to be considered a success at the end of one's life than to be considered a success today.	0.61	.6	0.75	0.82
	V32	The most important thing in life is how one feels in the long run.	0.65	.49	0.79	
	V33	It is more important to save for the future than to buy what one wants today.	0.59	.62	0.67	
	V34	Long range goals are more important than short range goals.	0.82	.41	0.79	
	V35	What happens in the long run is more important than how one feels right now.	0.83	.34	0.84	