2021 ASEE ANNUAL CONFERENCE
Virtual Meeting | July 26–29, 2021 | Pacific Daylight Time
Paper ID #33867

Proof of Concept: An Algorithm for Consideration of Students' Personalities in Team Formation

Mr. Jeong Hin Chin, University of Michigan

Bachelor of Science in Statistics and Data Science, Statistics Department, College of Literature, Science and the Arts.

Herbert Li, University of Michigan

A student studying at the University of Michigan. Majoring in Computer Science and Data Science. Passionate in data analytics and inspired to utilize data to find more useful discoveries.

Dr. Robin Fowler, University of Michigan

Robin Fowler is a lecturer in the Program in Technical Communication at the University of Michigan. She enjoys serving as a "communication coach" to students throughout the curriculum, and she's especially excited to work with first year and senior students, as well as engineering project teams, as they navigate the more open-ended communication decisions involved in describing the products of open-ended design scenarios. She is also one of the faculty innovators behind Tandem, a tool to support teams with equity in teamwork.

Proof of Concept: An Algorithm for Consideration of Students' Personalities in Team Formation

Abstract

Team-based pedagogy is common across engineering, especially in introductory courses and senior design projects. Although team-based learning provides students with many benefits, not all students benefit equally from it. Due to the fact that students of a team may come from vastly different backgrounds (academically and non-academically), faculty members find it difficult to predict barriers to team success. Better prediction of student and team outcomes would allow for better pedagogical decision-making in the formation and support of team-based learning. This methodological paper describes a dataset including student personality and team preference inputs, self-and-peer-assessments of teamwork, and perceptions of teamwork outcomes. We use student- and team-level inputs to consider students' likelihood of dominating group discussions. Using such information, we propose a method for clustering students for analysis with the aim of finding ways to group students into teams more effectively to help achieve their particular goals or alleviate their concerns.

Introduction

Team-based work can provide students with a wide array of benefits, such as developing leadership skills, learning how to be a listener, or adapting to working with people who are drastically different from one another. Team-based work serves as a microcosm of the real world, in which people of vastly different backgrounds aggregate together to solve common problems. Inevitably, people composing a team can have different experiences, aspirations, or personalities. At times, such differences lead to inequitable practices within teams such as unfair distributions of the workload or type of work, often in problematic ways. For example, active and dominant students might advocate for themselves to take on the more challenging and interesting parts of a project, leaving the more mundane and menial work to their passive teammates. Mix-gendered teams in engineering education have been found to have unbalanced work distribution: women may do more work related to planning or communication, while men may do more technical work [1]. Such disproportionate allocation of work and assumption of roles lead to students obtaining unequal gains from team projects, for which the more active student might have had the more enjoyable and useful experience from the project, at the expense of the more passive student [1], [2]. Ideally, team projects should offer students the opportunity to get involved in many different aspects of the experience of achieving a common goal, as opposed to just concentrating in one area.

Learning analytics is the collection and analysis of data about learners and their contexts, for the purposes of optimizing learning experiences [3]. First defined in 2011 [3], learning analytics has come to the fore based partially on the digital traces learners leave via learning management systems. “Big data” provides an opportunity to investigate small relationships not detectable in small samples, such as an individual teacher observing patterns of experiences of their students would see, while doing so in authentic spaces such as real classrooms. Online team assessment tools, such as the Comprehensive Assessment of Team Member Effectiveness (CATME) [4], provide a similar opportunity. Multiple team assessment tools exist; in this paper, we use data available via Tandem [5], a team-support tool being developed at the University of Michigan.

Tandem is designed to provide formative feedback to individuals and teams, and it includes algorithms intending to identify equity issues on teams. First deployed in a single class in 2019, it is still under development and its dataset is still fairly small (it’s now being used by ~1500 students/semester). However, we recognize the potential for detecting patterns of inequity as its sample grows, and we propose a methodology here for doing so.

This research aims to consider different types of students based on their personalities and preferences as reported on an initial survey. Moreover, through ratings that students receive from their teammates and teams over the semester, students of each type and groupings of types can be connected to certain team experiences, which can help instructors form teams to promote fairness within the team or to alleviate specific concerns. The purpose of this research is to propose a method of analyzing students to promote equity within a team, specifically regarding how much students speak up in group conversations. Currently, in many instances, instructors can assign teams randomly or with certain criteria in mind, but often instructors do not have quantitative, predictive tools or information from students regarding their preferences or goals [6]. For example, faculty can choose to group students who identify as women together on teams or distribute them equally across teams; these two practices are each informed by a particular philosophy. Such a practice can be facilitated via team formation software, such as CATME or Tandem, or can be done by hand. Given the degrees of freedom involved in forming teams, however, most faculty are only able to consider a few such “decision rules” in team formation. Further, there is often not (yet) a feedback loop in tools, to help instructors understand whether the teams as formed performed in line with instructor philosophy (e.g., an instructor who makes sure that students who identify as women are not isolated on teams may be following [7] and worry that women who are isolated on teams will be less heard on their teams and feel less belonging. “Big Data” practices would allow teams to be formed and to generate outcome predictions: a team with students who [set of variables] would be predicted to have [more, less, same] equity of idea enactment, for example. Then, actual outcomes can be used to inform the tool, such that predictions of outcomes continually improve. Eventually, instead of asking the tool to “pair people who identify as women” on teams, the tool can instead “decrease likelihood of students reporting low idea enactment and low belonging,” and can consider much more nuance in addition to reported gender identification (e.g. incoming sense of belongingness, extraversion, willingness to speak up in groups, self efficacy). This project is a methodology paper; our goal is to utilize “Big data” to statistically identify students’ qualities, providing instructors an alternative to guessing what a student is like, and to demonstrate how such information could be used to iteratively inform team formation.

Data Set

Tandem is one of many team support tools. It is designed for student teams both within engineering and outside of engineering; analyses will need to be conducted using nested models to determine whether there are differences in groups by discipline (as well as by other factors, such as grade status, and also to capture uncertainty based on characteristics).

Tandem includes student information provided via three routes; the resulting data set can be combined with other institutional information (demographic information, GPA, concurrent enrollment, etc).

- Surveys
    - BoT: a **“beginning of term” survey** (before students have met their teams) that asks about individual characteristics relevant in teamwork literature, such as personality characteristics, as well as about teamwork preferences and previous teamwork experiences
    - An **“early check” survey** (administered after one or two team meetings) that asks students to rate themselves and their teammates on a variety of behaviors. This survey is designed to understand students’ initial perceptions of each other as well as to detect imbalances in power and communication concerns, such as students who do not feel their ideas will be valued by teammates
    - MoT: a **“mid-project” assessment** (which can be administered multiple times during a semester as desired) that asks students to rate themselves and their teammates on a variety of behaviors. Behaviors are chunked into “contributions to the teamwork product” and “pro-teamwork behaviors” such as listening to teammates and contributing ideas. This survey also includes open-ended feedback on self and teammates, including what they should be most proud of and what they should work on.
    - EoT: an **“end of project” assessment** that asks students to rate themselves and their teammates on a variety of behaviors, again chunked as above. This survey also includes open-ended feedback on self and teammates, including what they should be most proud of and what they should work on. Finally, this survey also repeats items from the beginning survey, including teamwork preferences, to understand how students’ teamwork skills and attitudes might have changed.
- Team Checks
    - Designed to be mobile-friendly and fast, team checks are provided weekly to students by default. Students are asked to **rate the team overall on five items**

(overall “working well,” “sharing of work,” “sharing of ideas,” “team confidence,” and “logistics/challenges.”) If a student rates any of those lower than the midpoint, they see a second page with some boxes that can be checked (if particular common problems apply) as well as an optional text-entry space to alert instructors regarding issues the team is facing.

Tandem is envisioned as a tool to assist our institution and classrooms in our goals of inclusive teaching and promoting diversity, equity, and inclusion goals across the College of Engineering. Tandem’s messaging often includes gentle equity nudges such as making space for quieter voices, and it sometimes includes messages more explicitly related to diversity, equity, and inclusion (DEI), such as messaging regarding how more diverse groups make better decisions.

Method

This study makes use of data from Tandem, an online team-support tool that provides self-and peer-assessment, and tailored instruction to teams and team members. Specifically, Tandem uses a “Beginning of Term” survey (BoT), “Middle of Term” survey (MoT), and “End of Term” survey (EoT) to collect information from students along with a number of metrics, which it can use to build team profiles and to identify potential equity concerns on teams. The beginning of term survey includes students’ own ratings of themselves on relevant experiences and self-efficacy for project-related tasks in the class and preferences for approaching teamwork (including procrastination, academic orientation, and extraversion). When responding to the surveys, students often use a numerical scale to express different levels of agreement or agreement to an assertion.

Below are some of the BoT survey items. Students move a slider over seven points, from identifying completely with one end of the scale to identifying completely with the other side. The slider initially sits on the center “neutral” position, but it must be moved for the student to advance in the survey (it can be moved off the neutral and returned to neutral, though). Items in the BoT are inspired by validated scales in the literature for constructs relevant to teamwork, but to keep the surveys short, they are single-item and sometimes even double-barreled, based on user testing conducted by the Center for Academic Innovation.

Where would you place yourself on the following scales? [7 stops on the scale]			
[Extraversion]	In groups, I tend to listen more than speak.	⟵⟶	I often speak up groups.
[WorkEarly]	I usually do work close to a deadline.	⟵⟶	I get working on a project as soon as it is assigned.
[BelongConcern]	I expect to fit right into $Course.	⟵⟶	I expect to feel pretty out of place in $Course.
[Control]	I think it's good to share work, even if my team might finish tasks differently than me.	⟵⟶	I'd rather pick up extra work so I know it's done right.
[SpeakUp]	I'd rather hold back ideas or preferences if my group stays happy.	⟵⟶	It's easy for me to speak up about my ideas or preferences even if it disrupts my group.

Figure 1. Students can use numbers between 1-7 to quantify their agreement/disagreement with each statement after reflecting on themselves. $Course is replaced by tailored text.

Cluster analysis can be used to divide the students into clusters based on students' patterns of agreement and disagreement with the statements above. Each cluster represents students of a certain type or behavior. Each cluster will be marked as having the highest or lowest average ratings of certain metrics. For example, if students in Cluster 1 on average used higher numbers when responding to the statements regarding "WorkEarly" compared to other clusters, Cluster 1 will be marked as high in the metric of "WorkEarly". On the other hand, if students in Cluster 1 on average used lower numbers when responding to the statements regarding "SpeakUp" compared to other clusters, Cluster 1 will be marked as low in the metric of "SpeakUp." To further justify the personality or characteristics of students in a cluster, the ratings that students in each cluster received will be further analyzed. In the MoT and EoT surveys, students are asked to rate other members of their team on statements shown in the following figure:

Where would you place $TeamMember on each of these scales? [slider with 9 stops]			
[ET_PeerIdeas]	I didn't hear many ideas from $TeamMember.	⟵⟶	$TeamMember offered up most of our team's creative solutions.
[ET_PeerTeacher]	$TeamMember did not explain what they were doing for the project or actively share their skills and knowledge.	⟵⟶	$TeamMember actively teaches others and shares their skills and knowledge.
[ET_PeerListener]	$TeamMember was more likely to speak up with their own ideas than to listen and encourage others.	⟵⟶	$TeamMember was a great listener who helped encourage our other members.
[ET_PeerEnacted]	Our project didn't include many ideas from $TeamMember.	⟵⟶	Many of $TeamMember's ideas were used in our project.
[ET_PeerEffort]	$TeamMember didn't put in as much effort as they should have.	⟵⟶	$TeamMember did more than their fair share of work for our project.
[ET_PeerQuality]	$TeamMember's work often needed to be redone or wasn't good enough.	⟵⟶	$TeamMember's work for our project was exceptional.
[ET_PeerBelon gin g]	$TeamMember sometimes clashes with members of our group.	⟵⟶	$TeamMember fits in well with the whole group.
[ET_PeerReliabil ity]	$TeamMember was often late to meetings, was distracted while we were collaborating, or was generally unreliable.	⟵⟶	$TeamMember always showed up, responded to messages, and was generally reliable.
[ET_PeerValuable]	$TeamMember was still gaining the skills needed for our project.	⟵⟶	The skills $TeamMember brings to the team are incredibly valuable

Figure 2. Students can use numbers between 1-9 to quantify their agreement/disagreement with each statement regarding another member on the student's team. $TeamMember is replaced by tailored text.

Noticeable patterns in ratings that differentiate the clusters will be recorded and checked for significance. If the correlation between a cluster and certain ratings is significant, we will consider that cluster as having connections to such characteristics. For example, if through statistical analysis, it is noted that students of Cluster 1 are more likely to receive ratings leaning towards the statement "$TeamMember was more likely to speak up with their own ideas than to listen and encourage others", the statement is potentially useful to describe students of Cluster 1. Because different clusters might have different sizes, bootstrap analyses will be applied to each group to more accurately reflect the characteristics of students. Moreover, to confirm significance, a statement will only be connected to a cluster if there are patterns in ratings for the statement in both MoT and EoT surveys.

Cluster analysis

Cluster Analysis is an approach that groups a set of objects such that those objects in the same group are more similar to each other than to those in another group. Although there are a variety of cluster models that can be used, this paper will focus on the centroid models which can be performed using the k-means algorithm. K-means is a prototype-based, simple partitional clustering algorithm that attempts to find K non-overlapping clusters [8]. The mathematics behind K-Means:

$$\min_{\{m_k\},1\le k\le K} \sum_{k=1}^{K} \sum_{\boldsymbol{x}\in C_k} \pi_{\boldsymbol{x}} \mathrm{d}ist(\boldsymbol{x}, \boldsymbol{m}_k), \tag{1}$$

π_x is the weight of **x** ,

n_k is the number of data objects assigned to cluster C_k ,

$m_k = \sum_{x \in C_k} \frac{\pi_x x}{n_k}$ is the centroid of cluster C_k ,

K is the number of clusters set by the user, function “dist” computes the distance between object **x** and centroid m_k ,$1 \le k \le K$. Usually is the Euclidean Distance.

Bootstrap analysis

Bootstrap analysis is a general approach to quantify uncertainty without requiring any distributional assumptions and is used because the cluster analysis will likely result in clusters of different sizes. It works by treating the sample as a population and drawing samples from it with replacement. The bootstrap procedure:

Goal: Estimate mean/variance/standard error of ***T***.

1) Compute $\tilde{T}$ from original data
2) Draw bootstrap samples of size n with replacement
3) Compute T_r from bootstrap sample
4) Repeat 2-3 from r = 1,...,R where R is large (at least 1000, preferably more than 10000)

Distribution of T_r is empirical distribution of T

Simulation

To simulate how the cluster analysis would work, we can generate random integer values in the range [1,7] to represent students' responses in BoT. We will create simulated data of size 500 to imitate the data collected from BoT. This simulation section provides an explanation on how the codes are implemented, which is useful for future research reproduction.

````
```{r}
rep <- function(){sample(1:7, 500,replace = TRUE)}
Data_Simul <- data.frame(Control =  rep(),
                    SpeakUp = rep(),
                    WorkEarly = rep(),
                    BelongConcern = rep(),
                    Extraversion = rep())
```
````

Since we are unable to predict how the students will answer the survey, we will use the sample() function provided by R to mimic the randomness in students' responses. sample() will help us pick the values in the range from 1 to 7 (inclusively) with replacement. Then, in order to look at the clustering of data, a dendrogram is drawn and the cluster analysis plot is generated. The Euclidean distance is used in the cluster analysis.

````
```{r}
dcan <- dist(Data_Simul, method = "euclidean")
hcan <- hclust(dcan)
par(mar = c(0,0,0,0))
plot(hcan, axes = FALSE, ann = FALSE, main = NA, labels = FALSE, hang = 0.01)
```
````

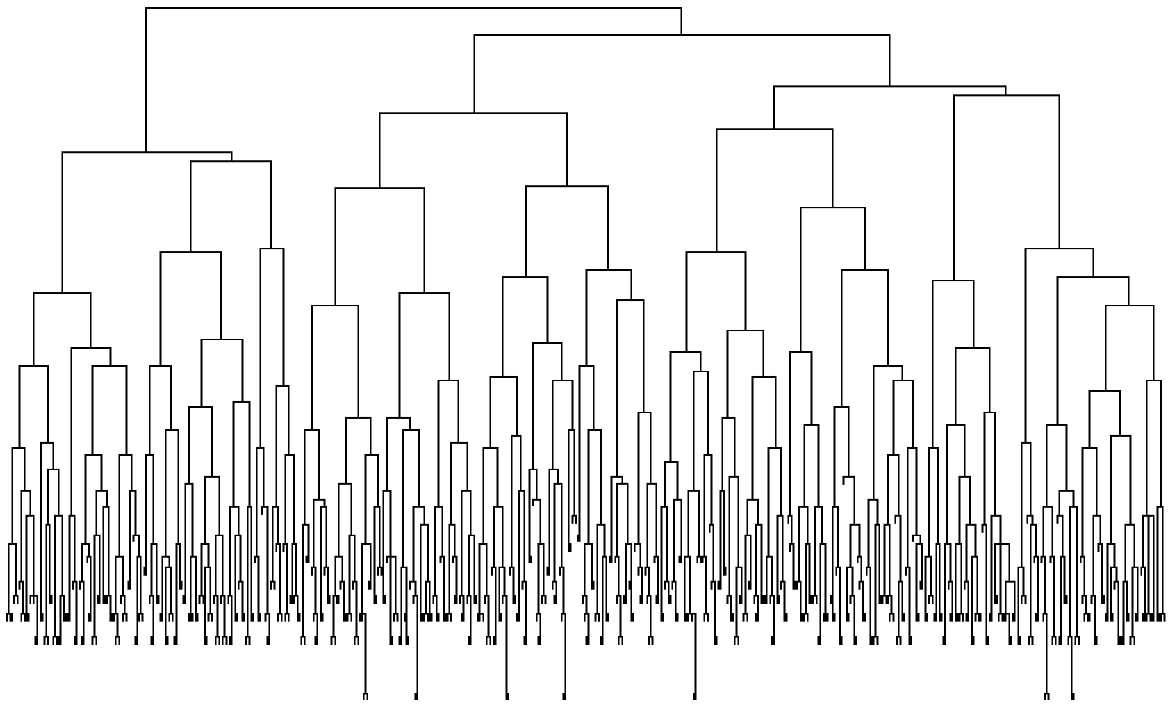

Figure 3. Dendrogram of the clustering performed on the simulated data suggesting 3 different clusters.

We can use K-Means clustering to perform clustering analysis on the data.

```
```{r}
Data_Simul_Kmeans <- Data_Simul
k3 <- kmeans(Data_Simul_Kmeans, centers = 3, nstart = 25)
fviz_cluster(k3, data = Data_Simul_Kmeans,geom = "point")
Data_Simul_Kmeans_Result <- Data_Simul_Kmeans %>%
                mutate(Cluster = k3$cluster) %>%
                group_by(Cluster) %>%
                summarise_all("mean")
Data_Simul_Kmeans <- Data_Simul_Kmeans %>% mutate(Cluster = k3$cluster)
```
```

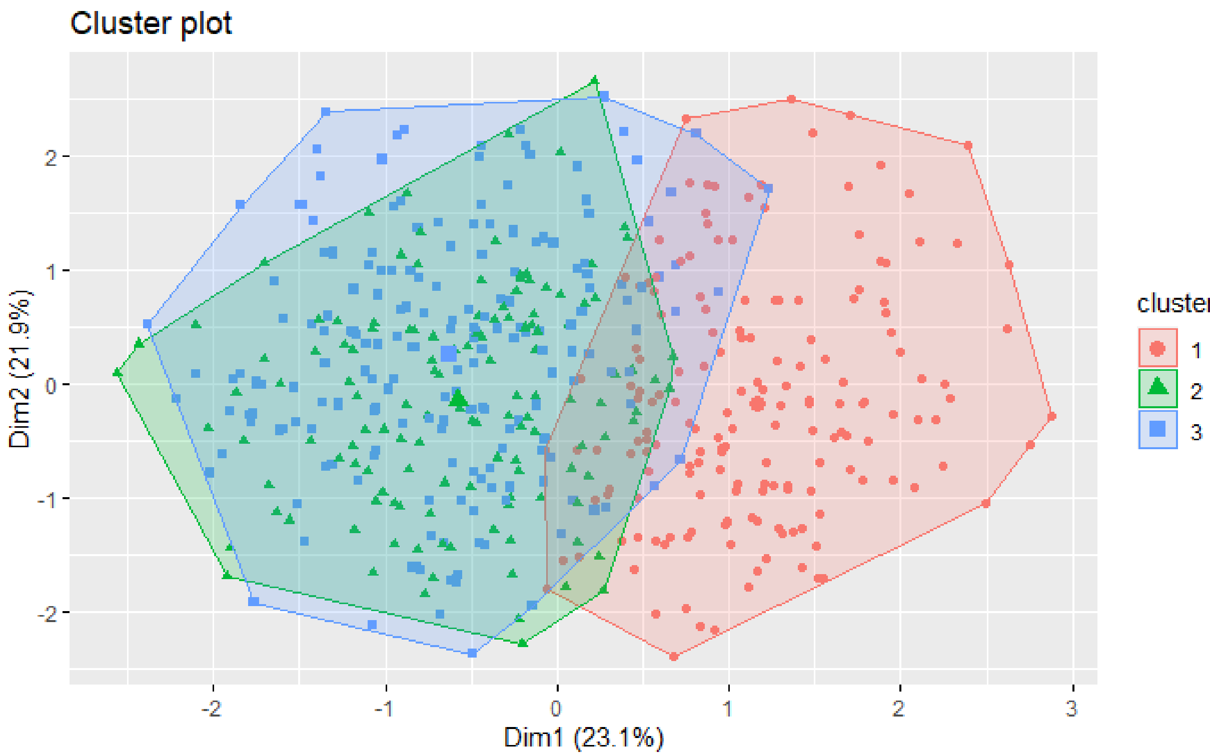

Figure 4. Simulated students' data separated into the three different clusters.

Cluster	Control	SpeakUp	WorkEarly	BelongConcern	Extraversion
1	2.3509	3.8363	4.8421	3.4094	5.2515
2	5.5303	4.5682	2.6212	4.3106	5.3788
3	4.6701	3.9188	4.1472	3.8528	2.0152

Figure 5. Average ratings that the students gave to themselves on the BoT survey.

Looking at Figure 5, we notice that Cluster 2's Control score is higher than Control scores of Clusters 1 and 3. We can say with certainty that there is a difference in Control score between Cluster 1 and Cluster 2. Nonetheless, the Control scores in Cluster 2 and 3 are almost the same. Therefore, when there is ambiguity in the value in the table, we can use bootstrap analysis to help us determine if there is a difference in the control variable between Cluster 2 and Cluster 3.

````
```{r}
# We just want Cluster 2 & 3
control <- Data_Simul_Kmeans %>% filter(Cluster != 1)

# Setting up the functions and constants
````
````

~~~
B <- 10000


# Function to calculate the difference in mean in the “Control” variable
Control_mean_diff <- function(x, index) {
      xstar <- x[index, ]
      mean(xstar$Control[xstar$Cluster == 2 ], na.rm = TRUE) -
      mean(xstar$Control[xstar$Cluster == 3 ], na.rm = TRUE)
}

boot_control_2_3 <- boot(control, statistic = Control_mean_diff, R = B)
boot.ci(boot_control_2_3, type = "norm")  # Normal Confidence Interval
```
~~~

```
BOOTSTRAP CONFIDENCE INTERVAL CALCULATIONS
Based on 10000 bootstrap replicates

CALL :
boot.ci(boot.out = boot_control_2_3, type = "norm")

Intervals :
Level      Normal
95%   ( 0.5313,  1.1850 )
Calculations and Intervals on Original Scale
```

Figure 6. Result of Bootstrap Analysis on the Difference in the mean of Control between Cluster 2 and 3.

As we can see, the value 0 is not present in the 95% confidence interval. This suggests that we are 95% confident that there is indeed a difference of at least 0.5313 in the mean of the “Control” variable between Cluster 2 and 3. Similar bootstrap analysis can be performed on other variables that the researchers are interested in.

Using the values from the data and bootstrap analysis, we can observe that:
Cluster 1: High in metrics of “WorkEarly” and “Extraversion”, low in metrics of “Control”
Cluster 2: High in metrics of “Control” and “Extraversion”, low in metrics of “WorkEarly”
Cluster 3: Low in metrics of “Extraversion”

Therefore, students in each cluster will have the characteristics associated with each cluster. Knowing the characteristics of each cluster is crucial; using incoming students’ characteristics, Tandem can then make predictions regarding how that student will interact with their peers; teams can be designed to maximize or minimize the likelihood of various patterns.

As mentioned in the previous section, besides BoT survey, Tandem also sends out MoT and EoT surveys to students during the middle and end of semester. Both surveys include identical questions for peer ratings on the student. After separating the students into different clusters, we will calculate the mean scores of each variable in the MoT and EoT by cluster groups. This will allow us to triangulate, too; since BoT data is provided by individuals via self-report, it is critical that we consider whether peer-reported values suggest that the self-report includes important signal.

Example on Real Data

Below is an example demonstrating the usage of the codes mentioned in the Simulation section. Based on the BoT survey, data from 501 students in engineering courses are divided into three clusters after applying cluster analysis. After identifying the three clusters, each cluster represents different characteristics of students across the seven metrics on average, as shown below.

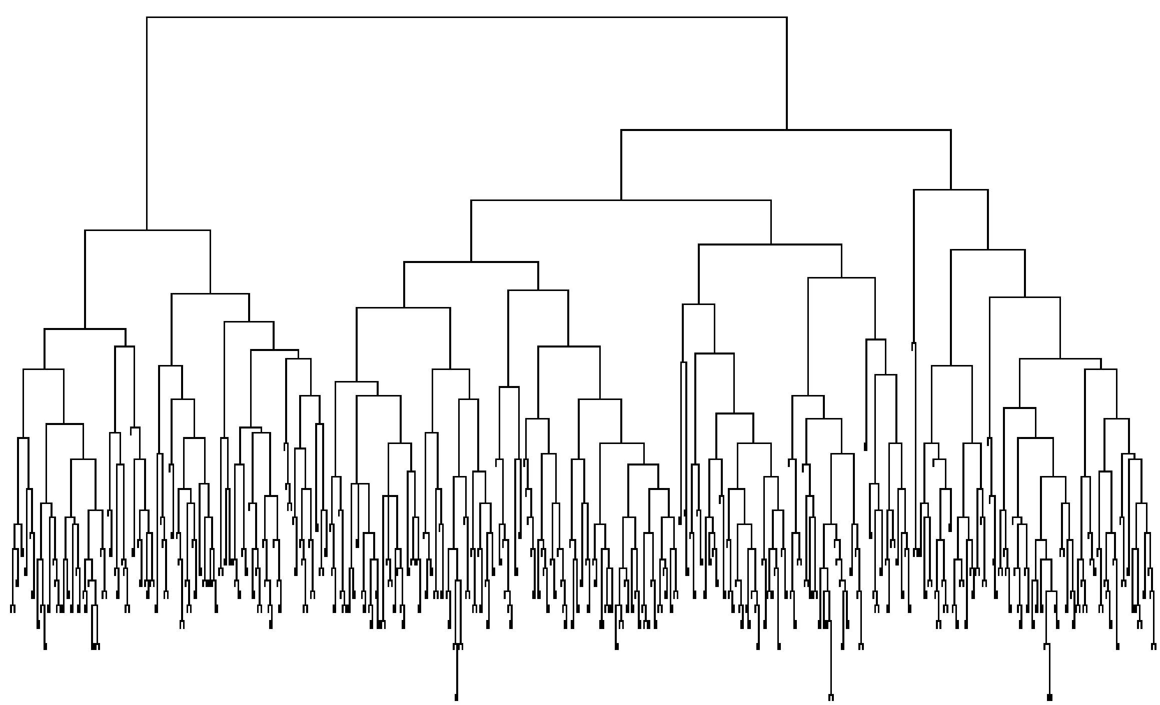

Figure 7. Dendrogram of the clustering performed on the BoT survey data (n = 501) suggesting that there might be 3 different clusters.

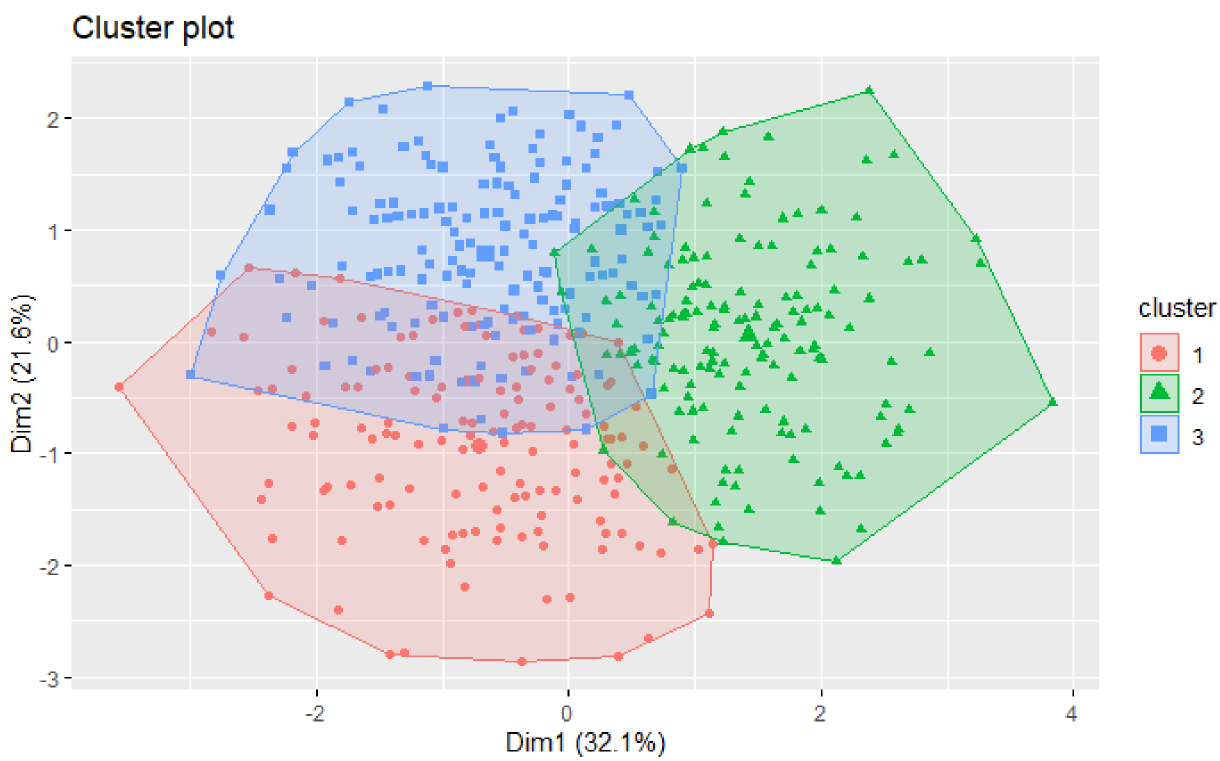

Figure 8. The red circles represent students of cluster 1. Green triangles represent students of Cluster 2. Blue squares represent students of Cluster 3. Overlaps does not indicate that a student is in multiple clusters: each student is only in one of the clusters.

Cluster	Control	SpeakUp	WorkEarly	BelongConcern	Extraversion
1	3.7625	4.8500	5.4500	2.8688	5.1813
2	3.7055	3.3620	3.7362	3.9141	2.9018
3	2.9888	4.9551	2.7865	2.7247	5.1067

Figure 9. The table represents the average ratings that the students gave to themselves on the BoT survey.

Cluster	Control	SpeakUp	WorkEarly	BelongConcern	Extraversion
1	1.4900	1.1774	0.7994	1.0528	1.0271
2	1.3284	1.1157	1.2163	1.2293	0.9764
3	1.2888	1.0567	0.8021	1.2109	1.0866

Figure 10. The table represents the standard deviation for each variable in each cluster on the BoT survey.

Using the values from the data and bootstrap analysis, we can observe that:
Cluster 1: High in “Control”, “WorkEarly”, and “Extraversion”
Cluster 2: High in “BelongConcern” and “Control”, low in metric of “Extraversion”
Cluster 3: High in “SpeakUp”, low in “Control”, “WorkEarly”, and “BelongConcern”

As cluster analysis requires, context experts (in this case, engineering faculty and/or engineering students) should look at the clusters and characterize them according to their characteristics.

As described earlier, this analysis is interested in predicting patterns of domination on teams. It has been our experience that some team members wield more decision-making power than perhaps they should, and other students enter team situations wary. This is one reason that instructors in Engineering are sometimes suggested to avoid isolating women [7]. This cluster analysis might allow us to predict these problems before they occur, considering more nuance than solely students’ reported gender identity: for example, students in Cluster 1 are probably more likely to be “powerful” in teams, because of their feeling of belonging, their likelihood of starting a project early and wanting to make sure it’s done in the ways they want, and their extraversion/willingness to speak up. In contrast, we might worry most that students in Cluster 2 will feel “run over” by those students. Students in Cluster 2 have more concerns they do not belong in engineering and are also less willing to speak up. The cluster analysis could be combined with information on students’ reported gender to make this analysis even richer.

We would like to examine concurrent validity of these groups by seeing how the responses of students pattern on MoT and EoT variables; ideally we would do something like hierarchical linear modeling (HLM) rather than simple ANOVAs, because students are nested in teams. We expect that students in Cluster 1 might be more likely to be rated as “more likely to speak up with their own ideas than to listen to others,” and there might be a unique effect, where students from Cluster 2 might be less satisfied with their teams depending on the number of Cluster 1 students assigned.

Conclusion

When assigning groups, instructors tend to group students randomly or at times group students based on some criteria, such as GPA, exam grades, or even gender [6]. However, such methods of grouping students are fairly arbitrary and ignore the intrinsic qualities and nuance of each individual student; they also undervalue the growing differences among students [9]. In this research proposal, we offer an alternative approach to grouping students randomly or based on objective criteria that omit many details of the students. To make things simple, students should be divided into clusters depending on their responses to an initial survey that reveals details about their personalities. Hence, knowing which cluster a student belongs to provides instructors a clearer overview of the personality of an individual student, as opposed to just guessing on the surface level. Going from the averages of the metrics of each cluster, instructors can have a general sense of the behavior or personality of the student as expressed by the students. Given other questions in the BoT survey that asks students to disclose their goals and concerns for the courses, instructors or the tool can take into account the cluster that a student belongs to and assign students together to try to fulfill their goals or alleviate their concerns. As a result of using clustering analysis, instructors are more likely to achieve the goal of creating more diverse and equitable groups, since instructors will have a clearer understanding of each student after knowing which cluster the student belongs to. For example, having a group of students coming from different clusters will likely lead to the diversification of personalities/goals within the team, since students coming from different clusters are marked by their differences in characters and preferences. It would be important to note the values of the variables of each cluster and plan accordingly. For example, if students from cluster 1 have a significantly higher score on BelongConcern and lower score on SpeakUp, instructors ideally might want to group students of cluster 1 with students from another cluster that has a lower Control score. As a result, instructors are better able to take into account equity when creating groups, since students who might not have a strong sense of belongingness or willingness to share ideas will be less likely to be dominated by other students.

Moreover, through ratings from students' team members, we can better understand students in each group to iteratively improve predictions. As a result, the goal would be to have a database of different clusters of students and their associated characteristics as described by their team members. Subsequently, when new clusters appear as similar to clusters in the database, the instructors would have a more comprehensive understanding of the personalities of students in such clusters.

In the BoT survey on Tandem and in many other team formation tools, students respond to a variety of questions that reveal information about themselves. However, because of the sheer volume of questions, using student responses directly to try to group students together could impose substantial amounts of work for instructors. Tandem, CATME, and other team formation tools let instructors specify how to use such input items to "optimize" teams, such as to group people with similar schedules or to avoid isolating women. The tool can handle the computational complexity of trying to optimize across multiple variables.

However, to the best of our knowledge, no tool currently considers interactions among variables. That is, perhaps "not isolating women" is more important for women who report low belonging upon entering a class but not for women who enter with a high sense of belongingness, high self efficacy, and strong extraversion. This analysis is a first step towards considering the complicated ways that demographic, identity, personality, and team preference variables might interact in a way that can be used to improve team formation algorithms. Hence, we propose to use cluster analysis to understand students and better take into account the students' personalities. With an analysis such as described above, team formation tools can better understand the nuance of each student's personality and plan accordingly.

References

[1] R. R. Fowler and M. P. Su, “Gendered Risks of Team-Based Learning: A Model of Inequitable Task Allocation in Project-Based Learning,” *IEEE Trans. Educ.*, vol. 61, no. 4, pp. 312–318, Nov. 2018, doi: 10.1109/TE.2018.2816010.

[2] L. Meadows and D. Sekaquaptewa, “The Influence of Gender Stereotypes on Role Adoption in Student Teams,” in *2013 ASEE Annual Conference & Exposition Proceedings*, Atlanta, Georgia, Jun. 2013, p. 23.1217.1-23.1217.16, doi: 10.18260/1-2--22602.

[3] “What is Learning Analytics?,” *Society for Learning Analytics Research (SoLAR)*. https://www.solaresearch.org/about/what-is-learning-analytics/ (accessed Mar. 07, 2021).

[4] M. L. Loughry, M. W. Ohland, and D. J. Woehr, “Assessing Teamwork Skills for Assurance of Learning Using CATME Team Tools,” *Journal of Marketing Education*, vol. 36, no. 1, pp. 5–19, Apr. 2014, doi: 10.1177/0273475313499023.

[5] “Tandem.” https://tandem.ai.umich.edu/welcome (accessed Mar. 07, 2021).

[6] B. Benson, A. Danowitz, J. Callenes, and P. Hummel, “Perceived Benefits and Drawbacks of Group Assignment Methods,” in *2019 ASEE Annual Conference & Exposition Proceedings*, Tampa, Florida, Jun. 2019, p. 33166, doi: 10.18260/1-2--33166.

[7] B. Oakley, R. M. Felder, R. Brent, and I. Eljajj. (2004). Turning student groups into effective teams. *Journal of student centered learning*, *2*(1), 9-34.

[8] J. Wu, “Cluster Analysis and K-means Clustering: An Introduction,” in *Advances in K-means Clustering: A Data Mining Thinking*, J. Wu, Ed. Berlin, Heidelberg: Springer, 2012, pp. 1–16.

[9] S. “Stan” Lan, “Participation, Class Types, and Student Performance in Blended-learning Format,” in *2011 ASEE Annual Conference & Exposition Proceedings*, Jun. 2011, p. 21, [Online]. Available: https://www.asee.org/public/conferences/1/papers/1670/view.