Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2024
A Regional Conference of the Society for Industrial and Systems Engineering

Optimization for Diverse Team Assignment with a Case Study at West Point

Bryan Frost[1], Phillip Schmedeman[1], and Jonathan Paynter[2]

[1]Department of Systems Engineering,
United States Military Academy,
West Point, New York 10996

[2]Department of Mathematical Sciences,
United States Military Academy,
West Point, New York 10996

Author Note: The views expressed herein are those of the authors and do not reflect the position of the United States Military Academy, the Department of the Army, or the Department of Defense.

Abstract: Diversity is often a priority when assigning individuals to teams. The United States Military Academy (USMA) aims to create diverse groups of students for rooming and training. However, the current method focuses on evenly distributing personal attributes across teams rather than optimizing diversity within each. This paper presents a goal programming approach to optimize diversity when conducting team assignments while considering interactions between multi-attribute profiles. Using data from past and synthetic USMA classes, we demonstrate improved outcomes compared to current assignment methods. When measuring the performance of between-group differences, the proposed method yielded a 50% improvement over a heuristic that mirrors the current method. Compared to random assignment, the proposed method yielded improvements of 69%. The proposed formulation enables users to explore the trade-off between solution quality and computation time, by varying the degree of attribute interactions considered. The resulting algorithm is outlined to enable application across other organizations.

Keywords: Optimization, Assignment, Teaming, Diversity

1. Introduction

The United States Military Academy at West Point (USMA) enrolls approximately 1,200 students (i.e., cadets) per academic year. The students are distributed evenly among 36 teams, referred to as “companies”, with each company consisting of about 34 students per year group. USMA, like many professional institutions, recognizes that teams composed of a diverse population committed to a common goal generate long-term benefits for the individual (e.g., character development) and the organization (e.g., innovation) (Yeager & Nafukho, 2012). At West Point, each company of students has a leadership structure that provides students with character development opportunities outside of the classroom (Franklin, 2007). The challenge for USMA lies in employing the best methods to assign students to companies so as to best achieve diversity.

1.1. Literature Review

1.1.1. Beginnings of Team Formation

Assignment problems have been studied in academic literature since the 1950s, with approaches including random assignment, heuristics, free-market assignment, and optimization. Beheshtian-Ardekani & Mahmood, 1986 discussed instruments for assigning students to group projects that take into consideration the importance of the diverse backgrounds of students. While our goal is to assign students to groups that serve an organizational purpose, Beheshtian-Ardekani & Mahmood, 1986 assigned project work to students, based on diversity of experience. Their method, which would be later known as the “people-sequential heuristic,” ranks students by experience in descending order, and then assigns them one at a time to each project group.

The team assignment process was further aided by computer tools and applications upon the Weitz & Jelassi, 1992 development of a multi-criteria allocation decision support system (MCADSS) (Weitz & Jelassi, 1992). MCADSS has been applied to student assignment problems in addition to problems related to military manpower planning and corporate programs. The demonstration of MCADSS consisted of over 220 students at the European Institute of Business Administration (INSEAD) being assigned into study groups of six or seven. With these groups, the intent was to maximize the differences within study groups to simulate an international business environment. To meet their goal, each working group would consist of entirely unique nationalities and occupational backgrounds. MCADSS was able to achieve heuristic solutions for the INSEAD assignment problem, but arriving at a solution was not guaranteed on every trial, depending on the distribution of student attributes. Therefore, any given solution was likely not optimal. Nevertheless, MCADSS provided easy-to-use, comprehensible solutions that paved the way for future conversations in team assignments by diversity.

1.1.2. Refining the Team Assignment Process

Baker & Powell, 2002 demonstrated that maximizing diversity can be represented mathematically either as maximizing within-group distance or as minimizing between-group distance. Baker and Powell concluded that both methods were essentially equivalent. However, the size of optimization problems became computationally demanding very quickly, which prompted the exploration of heuristic solution procedures. Heuristic procedures work to circumvent the limits of computing power by way of construction and improvement procedures. Construction procedures, select the first student in a group arbitrarily, then find the most similar student to the first, and place them in the second group. Improvement procedures start with a completed case of assignments, but then make marginal improvements on diversity by swapping students between groups. While new ground was uncovered on future team assignment problems, Baker and Powell also demonstrated the need to refine team assignment problems to make the optimization less demanding.

Srinivas, Alizadeh, & Bastian, 2017 used Goal Programming (GP) optimization models in a two-stage process to determine how to best assign students at the Holy Family Academy to different working teams. The objective of their GP models is to minimize the differences between groups, as seen in Baker & Powell, 2002, to balance their student workforce across the organization. The Holy Family Academy considers attributes such as gender, race, and grade level when performing their optimization. By formulating an objective function that minimizes the sum of the weighted goal deviation variables, the potential academic performance of the various work-study teams is optimized.

In their study, Schulz, 2021 proposed a simplified approach to solve the maximally diverse team assignment problem using block constraints. These constraints allow certain students to be assigned to the same groups, thereby reducing the solution space of the problem and minimizing computational strain. The researchers used a linear objective function to balance attributes across different groups and developed the Balanced Maximally Diverse Grouping Problem with Block Constraints (BMDGPBC) as the method for solving. BMDGPC can solve problems with 100 students characterized by five attributes each in less than a second.

1.2. Motivation and Overview

In the current assignment process at USMA, five categories of personal attributes are considered to quantify team diversity: gender, race, recruitment for NCAA Division 1 athletics, graduation from the USMA Preparatory School (USMAPS), and College Entrance Examination Rank (CEER), which is a general measure of success gathered from high school performance. A simple heuristic is used to assign the students by considering one characteristic at a time and distributing students across the teams until no students with the characteristic remain. This method has been in practice with minimal changes since 2014.

While the execution of the current algorithm is fast—assigning the class of 1,200 in approximately two minutes—the probability of yielding an optimal solution is low. After performing the assignment in SQL, the administrator spends several days manually adjusting the solution to identify improvements in terms of leveling attributes averages across the teams.

The solution resulting from the sequential assignment heuristic depends on the order each characteristic is distributed, with decreasing influence, making it nearly impossible to achieve optimality. USMA does not provide guidance on which attributes are most important to diversify, therefore the results of the assignment are highly dependent on the order in which the distribution is performed . A method that distributes all attributes simultaneously is preferred.

Additionally, sequential distribution ignores the fact that many of the attributes are not mutually exclusive. This paper makes a novel contribution by demonstrating the benefit of incorporating the interaction of attributes into the problem formulation. By incorporating interactions into the formulation, we are able to explore trade-offs between solution quality and computation time.

2. Proof of Concept

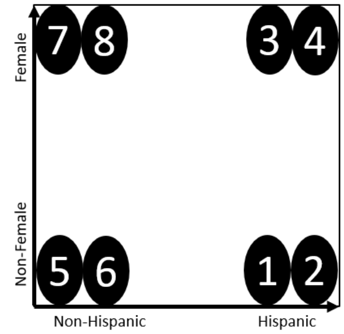
Here we present a simple example to demonstrate the benefit to diversity of incorporating combinations of attributes during team assignments. While it is important to distribute singular attributes across teams, distributing combinations of attributes (referred to here as interactions) further accomplishes our goal. Consider 8 students with 2 attributes each, assigned to two teams, 1 and 2. The attributes of the 8 students of the toy problem are depicted in Figure 1a, with 1 representing possession of the attribute, and 0 otherwise.

We can visualize the attributes of these eight students within a two-dimensional space as depicted in Figure 1b. There are various ways to assign the 8 students to 2 teams, and we illustrate our approach by depicting two solutions in Figure 2, where a student assigned to team 1 is in red, and team 2 is in blue. Assignment solutions are measured according to between-group differences and in-group diversity. Between-group differences compare the number of students with a certain attribute between the two teams and sum those differences across all attributes. For example, assigning two Hispanic students to each of the two

ISBN: 97819384962-4-0

Student	Female	Hispanic
1	1	0
2	1	0
3	1	1
4	1	1
5	0	0
6	0	0
7	0	1
8	0	1

(a) Example dataset of eight students described by two attributes.

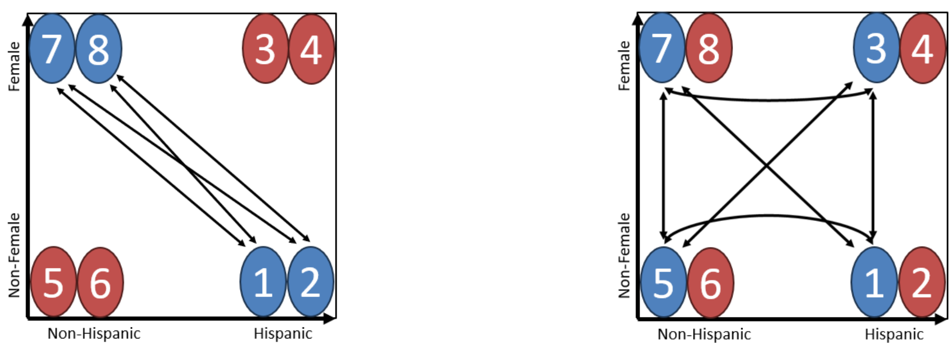


(b) Graphical representation of the eight students, with the horizontal axis representing Hispanic in binary encoding and the vertical axis representing female.

Figure 1: Consider a simple problem with eight students, each characterized by two binary attributes. Our aim is to create two teams from this population while maximizing diversity.

teams means no between-group difference for that attribute. In-group diversity measures the sum of the Euclidean distance between all assignment pairings of students in the same team, as depicted by the black lines in Figure 2.

We seek a method that minimizes between-group differences, ensuring a balance of diverse attributes in each team, while also maximizing in-group diversity, depicted here with Euclidean distance. To maintain computational feasibility, instead of directly optimizing for Euclidean distance, we aim to balance the assignments between teams not only for single attributes (e.g., Hispanic) but also for interactions of attributes (e.g., female-Hispanics). Considering interactions in this scenario would lead us to balance Hispanic females between the two teams, resulting in scenario B's assignment.



(a) Scenario A: Interactions neglected.

(b) Scenario B: Interactions considered.

Figure 2: These figures represent assignment solutions for two scenarios. Students in blue are assigned to Team 1 and students in red are assigned to Team 2. The arrows represent the calculation of distance between students assigned to one of the teams.

Assignment solutions are measured according to between-group differences and in-group diversity. Figure 2a has a between-group diversity of 0, meaning that the attributes of students are effectively balanced across Company 1 (Red) and Company 2 (Blue). However, the assignment is not ideal, as the Euclidean Distance for both teams is 5.656. Figure 2b is also able to minimize between-group differences at 0. Additionally, Scenario B introduces penalties for interactions, discouraging two students with the same attributes from being assigned to the same team. This increases the Euclidean Distance measurement to 6.828.

In order to ensure that we achieve Scenario B, as well as greater diversity, we must incorporate interaction terms into our goal program. We can add a penalty to our goal program if two people with duplicate attributes (for example, two Hispanic Males or two Non-Hispanic Females) end up in the same team. Including these penalty terms as part of our objective value allows us to achieve the ideal solution on the right in 2.

Once the interaction effects are included in the goal program, the solution generated by minimizing between group

ISBN: 97819384962-4-0

differences approximates the solution of maximizing within-team differences. In both scenarios, diversity is still optimized. We can measure the diversity of each of the teams by taking an Euclidean distance—that is, the sum of the length of the line segments between each student in the team. In 2, we demonstrate how teams assigned without interaction effects have a smaller Euclidean distance than teams with interaction effects.

3. Methodology

To implement a new method of team assignment that incorporates interactions, we created a goal programming algorithm named Frost's Optimization Goal Program for the Assignment of Teams (FOGPAT for short). Goal programming provides a means to optimize our assignment of students to teams without the same computational strain of a Quadratic Assignment Problem (Baker & Powell, 2002). Within FOGPAT, our objective function will be to minimize the sum of our positive and negative goal deviation variables, based on the decision variable of which team each student has been assigned to.

When writing FOGPAT as a linear program, we utilized Julia for mathematical optimization (JuMP), and Gurobi as the external solver. While Julia is a newer and less widely recognized language in the world of coding, it is advantageous due to its easy-to-read syntax and superior performance with heavy computations. Additionally, the Gurobi solver is an optimization solver that is reputable for producing fast and feasible solutions.

3.1. Inputs and Pre-Processing

Prior to conducting the optimization, the FOGPAT accepts three inputs. One for the number of teams, a constant k to denote the amount of interactions being calculated and an array of the attributes of each student. If student s exhibits attribute a, then $\alpha_{sa} = 1$. Otherwise, $\alpha_{sa} = 0$. Next, several steps of pre-processing occur:

- The number of rows in the attributes array are enumerated as the set of s students.
- The number of columns in the attributes array are enumerated as the set of g Goals.
- Left and right bounds are created to evenly divide the students among the number of teams.
- Goal sets are created by identifying the sets of all African-American, Hispanic, female, USMAPS-graduated, football, NCAA D1 athlete, and Asian students within the class of 2027.
- The size of each Goal Set is divided by the number of teams to determine a target parameter to balance the students of each goal set.

3.2. Formulating Interactions

Lastly, the Interactions are pre-processed according to the constant k. The interactions constant will denote to what order interactions will be processed within the code. For example, if $k = 1$, then FOGPAT will process only the goal sets processed above. If $k = 2$, then FOGPAT will process all the original goal sets, along with all the pairwise interactions (e.g. students that are Hispanic and Female). While this amendment to FOGPAT will increase the diversity of student teams and minimize the distances between teams, it will come at the cost of generating more constraint variables and continuous variables, which will be computationally straining and take significant time to compute, given that it remains feasible. Once the interaction effects are included within FOGPAT, the solution generated by minimizing between-group differences approximates the solution of maximizing within-team differences.

3.3. Optimization

For our optimization formulation, the goal program is described by the following variables:

1. **Plus Deviations:** measuring the absolute value of *positive* differences between all goals and all teams.
2. **Minus Deviations:** measuring the absolute value of *negative* differences between all goals and all teams. These are distinguished from positive deviations in order to compile the absolute values of all differences (1).
3. **Interaction Deviations:** measuring the absolute value of *interactive* differences between all goals and all teams, both positive and negative.

4. **Decision Variables:** If student s is assigned to team c, then $x_{sc} = 0$. Otherwise, $x_{sc} = 1$. Therefore, we are left with binary decision variables (2).

$$d^+_{gc}, d^-_{gc} \geq 0 \qquad \forall \quad g \in G \qquad \forall \quad c \in C \tag{1}$$

$$x_{sc} = \{0, 1\} \qquad \forall \quad s \in S \qquad \forall \quad c \in C \tag{2}$$

These variables are present in the objective function formulation (3), which is to minimize the sum of all deviation variables for all goals and all teams. Additionally, our objective function will be subject to several constraints. First, each student will be assigned to 1 team, and 1 team only (4). Second, each team will contain between the specified minimum and maximum number of students (5). (This is typically 34 or 35 students per team for the USMA case study.) Last, the difference between the number of students in a specified goal set and team as well as the total of the goal deviation variables and the parameter for the relevant goal should equal 0 for all goals, within all teams (6). This constraint is additionally applied to interactions.

$$min \quad Z = \sum_{c \in C} \sum_{g \in G} (d^-_{gc} - d^+_{gc}) \tag{3}$$

$$\sum_{c \in C} x_{sc} = 1 \qquad \forall \quad s \in S \tag{4}$$

$$N_{min} \leq \sum_{s \in S} x_{sc} \leq N_{max} \qquad \forall \quad c \in C \tag{5}$$

$$d^-_{gc} - d^+_{gc} + \sum_{s \in S_g} x_{sc} - P_g = 0 \qquad \forall \quad c \in C \qquad \forall \quad g \in G \tag{6}$$

3.4. Performance Measurements

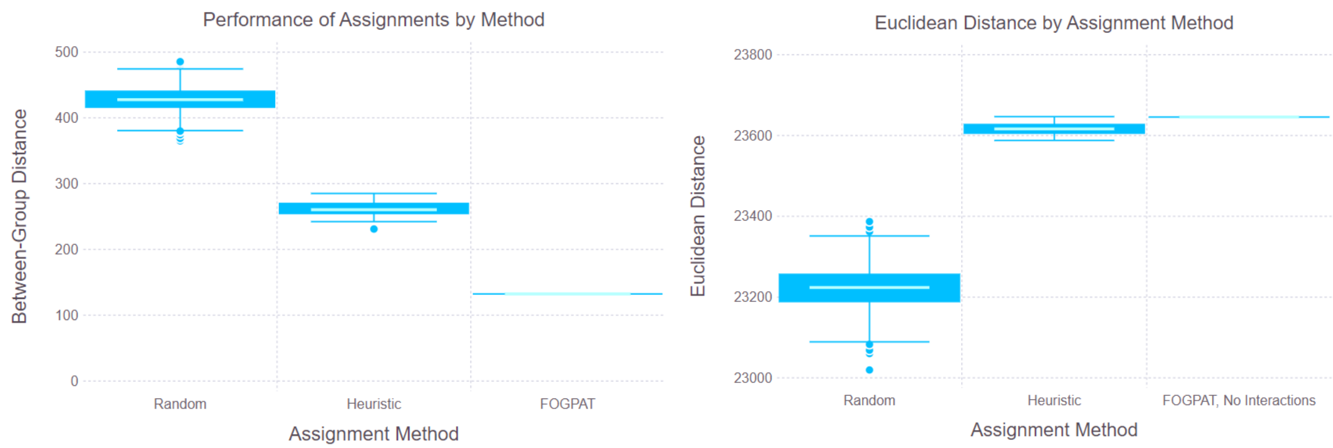
Performance of FOGPAT is measured first in terms of the diversity between student teams, which will be the same as the objective function value in Equation (3). This will generate the same parameters as in pre-processing, and find the difference in the amount of each attribute displayed in each team, which has been minimized. This measurement is important for diversity as it demonstrates how each team assignment is a microcosm of the whole population.

Our second performance measurement analyzes the diversity within each individual team, known as a Euclidean Distance. This measurement finds all students that are assigned to team c, and then calculates the distances between each student in the dimensional space defined by the amount of attributes in the program. This distance is summed all students across all teams to generate a cumulative Euclidean Distance for that assignment solution. The greater the value of the Euclidean Distance, the more diversity that is exhibited within that team as each student presents unique attributes.

4. Results & Discussion

To measure the performance of FOGPAT, we created a synthetic dataset that approximates the personal attributes of the USMA class of 2027. The objective is to assign these 1250 USMA cadets to 36 companies while maximizing diversity. The data was created using a Bernoulli random variable to approximate the percentage of USMA cadets exhibiting that binary attribute in the data. The first five records are represented in Table 1. In approximately twenty seconds, FOGPAT successfully assigned the simulated class so that all companies exhibited higher diversity, and demonstrated in our objective of minimizing between-group distances as shown in Figure 3a as well as the additional performance metric of maximizing within-group distances graphed in Figure 3b. The results of all metrics are depicted in Table 2.

The between-group measurement represents the performance of our objective function, which is our ability to evenly distribute all of the attributes of students incorporated into the assignment. By assigning 36 companies with FOGPAT, we further minimize our objective function value to 132. Compared to an average of 264 on the constructed heuristic that mirrors the current assignment method at USMA, FOGPAT achieves a 50% improvement in our objective. Additionally, we achieved a 69% improvement, on average, in our distribution of attributes when compared to random assignment, where the objective function value measured 428, on average. When examining performance in terms of Euclidean distance, expressed in Table



(a) Between-Group Distances (Goal: Minimize)

(b) Within-Group Distances (Goal: Maximize)

Figure 3: These figures graph the performance measurements of the three different methods of assigning the simulated class of USMA cadets. First, 1000 random iterations evenly assigned the class to 36 companies. Next, 50 iterations of the current method were conducted, but the order of students in the database were shuffled. Last, the FOGPAT, the proposed algorithm, was conducted.

Table 1: To test FOGPAT against random assignment and heuristic assignment, a simulated class of 1250 USMA cadets was created. Table 1 represents the first 5 records of data.

FEMALE	BLACK	ASIAN	NATIVE AMERICAN	POLYNESIAN	HISPANIC	FOOTBALL	NCAA D1	USMAPS
0	0	0	0	0	0	0	0	1
1	0	0	0	0	0	0	1	0
0	0	0	1	0	0	0	0	0
0	1	0	0	0	0	1	0	0
0	0	0	0	0	0	0	0	0

2 as within-group distance, FOGPAT makes marginal improvements, demonstrating a 1.8% increase over random assignment by increasing from 23,225 to 23,646. Compared to the heuristic, FOGPAT demonstrates a 0.1% increase over the constructed heuristic, which measured 23,613.

An advantage in the performance of FOGPAT is the ability to remove variability when conducting team assignments. Due to the nature of the heuristic operating sequentially, the results of the assignment are dependent on the order in which the students appear in the data matrix. The result is a standard deviation of 3.6% in the heuristic method from the average. FOGPAT's simultaneous optimization of all 36 groups effectively mitigates this uncertainty. Given the same attributes, teams, and interactions constant, FOGPAT will conduct the same optimal team assignment, every time.

When interaction terms were added, FOGPAT encountered increased computational strain due to the increased number of decision variables to compute. To run when $k = 2$, the optimality gap was lowered to 0.7. However, on toy problems with smaller amounts of data, implementation of interactions maintained between-group distances, and additionally incrementally increased within-group distances. Alternatively, improvement procedures can be combined with the current heuristic procedure in place by changing the order in which students appear in the database before conducting an additional iteration.

5. Conclusion

The incremental improvements made as we increase the columns of data suggest that assigning students to companies using 10 attributes per person without interactions will be less optimal than integrating interaction effects with only 5 or 6 attributes. This comparison of assignments is done in order to account for the lack of computing power necessary to perform the optimization. Once all interaction effects are accounted for with the full attributes of each student, then we will likely demonstrate measured improvement over the current method of sequential assignment being utilized. As of now, the implementation

ISBN: 97819384962-4-0

Table 2: Descriptive statistics for the three methods of assignment for the simulated class of USMA cadets. Measurements are depicted are unitless, but they offer comparisons between group assignments as they demonstrate the maximization of diversity relative to the population's potential for diversity.

		Random	Heuristic	FOGPAT (k=1)
Between-Group	Average	428	264	132
	Standard Deviation	18.3	9.7	0
Within-Group	Average	23,225	23,613	23,646
	Standard Deviation	49.6	17.6	0

process is possible, starting with the incoming USMA class of 2028, which will be comprised of similar demographics to the simulated data. Future work could examine methods to reduce the number of variables in the program, exploring the amounts of variables generated in the program, and finding the most efficient optimal solution based on how many interaction constraints are generated. Lastly, incorporating qualitative attributes, such as CEER score, and special assignments into the optimization, is a likely progression in the team assignment process. These can be accomplished by standardizing the columns or binning the attributes.

6. References

Baker, K. R., & Powell, S. G. (2002). Methods for assigning students to groups: A study of alternative objective functions. , *53*(4), 397–404. Retrieved 2022-09-28, from `https://doi.org/10.1057/palgrave.jors.2601307` doi: 10.1057/palgrave.jors.2601307

Beheshtian-Ardekani, M., & Mahmood, M. A. (1986). Education development and validation of a tool for assigning students to groups for class projects. , *17*(1), 92–113. Retrieved 2024-01-16, from `https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-5915.1986.tb00215.x` (_eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1540-5915.1986.tb00215.x) doi: 10.1111/j.1540-5915.1986.tb00215.x

Belford, P. C., & Ratliff, H. D. (1972). A network-flow model for racially balancing schools. , *20*(3), 619–628. Retrieved 2024-01-16, from `https://pubsonline.informs.org/doi/abs/10.1287/opre.20.3.619` (Publisher: INFORMS) doi: 10.1287/opre.20.3.619

Franklin, J. P. (2007). *Building leaders the west point way: Ten principles from the nation's most powerful leadership lab*. HarperCollins Leadership. (Google-Books-ID: 1Z6GCDBMdtAC)

Schulz, A. (2021). The balanced maximally diverse grouping problem with block constraints. , *294*(1), 42–53. Retrieved 2024-01-23, from `https://www.sciencedirect.com/science/article/pii/S037722172100059X` doi: 10.1016/j.ejor.2021.01.029

Srinivas, S., Alizadeh, M., & Bastian, N. D. (2017). Optimizing student team and job assignments for the holy family academy. , *47*(2), 163–174. Retrieved 2023-08-31, from `https://pubsonline.informs.org/doi/10.1287/inte.2016.0868` (Publisher: INFORMS) doi: 10.1287/inte.2016.0868

Weitz, R. R., & Jelassi, M. T. (1992). Assigning students to groups: A multi-criteria decision support system approach. , *23*(3), 746. Retrieved 2022-09-28, from `https://www.proquest.com/docview/198054771/abstract/6A5CF76022A74CF7PQ/1` (Num Pages: 12 Place: Atlanta, United States Publisher: American Institute for Decision Sciences)

Yeager, K. L., & Nafukho, F. M. (2012). Developing diverse teams to improve performance in the organizational setting. , *36*(4), 388–408. Retrieved 2022-09-28, from `https://doi.org/10.1108/03090591211220320` (Publisher: Emerald Group Publishing Limited) doi: 10.1108/03090591211220320