

**Equal Access, Unequal Outcomes: A Clustering Analysis of Advanced Coursework and
Postsecondary Outcomes Amid Structural Barriers in North Carolina High Schools**

Yue Summer Wu

Teachers College, Columbia University

ORLA 6541: Applied Data Science in Organizations and Leadership

Dr. Alex J. Bowers

November 25th, 2024

Abstract

The current research is grounded in the premise that educational attainment is essential for individual success and societal well-being (NASEM, 2019), yet systemic structural barriers, such as economic disadvantage, limit opportunities for marginalized students (Handwerk et al, 2008; NASEM, 2019; Patrick et al., 2020). This study aims to address two questions: To what extent does access to advanced coursework explain disparities in educational attainment? To what extent do structural barriers play a role in the relationship between advanced coursework and postsecondary outcomes? Using data from the Civil Rights Data Collection and the North Carolina Department of Public Instruction, hierarchical cluster analysis was conducted on 239 high schools in North Carolina to identify patterns in advanced coursework availability, enrollment, and postsecondary outcomes. A bubble plot further examined the role of economic disadvantage in shaping these relationships.

The findings revealed four distinct clusters of schools, showing that course availability alone did not correlate with postsecondary attainment. Schools with similar resources exhibited varied outcomes based on their student demographics, with economically disadvantaged schools often showing lower course enrollment and attainment. However, some schools predominately serving disadvantaged populations appeared to overcome these barriers. These results suggest that addressing educational inequities requires more than resource equality; schools must provide tailored interventions, such as expanded access to advanced coursework and additional supports. Future research should include qualitative case studies of high-performing schools and incorporate student-level data to better understand and address systemic disparities.

Keywords: clustering analysis, data visualization, advanced coursework, postsecondary readiness, educational equity

Equal Access, Unequal Outcomes: A Clustering Analysis of Advanced Coursework and Postsecondary Outcomes Amid Structural Barriers in North Carolina High Schools

Disparities in educational attainment have long been a persistent feature of the U.S. education system, reflecting deep-rooted socioeconomic and racial inequities (National Academies of Sciences, Engineering, and Medicine [NASEM], 2019). Historically, students from white, high-income, and English-proficient families have consistently achieved higher levels of success compared to their peers from marginalized backgrounds (Gamoran, 2001; NASEM, 2019). As Gamoran (2001) predicted, in the 21st century, educational inequity remains a defining challenge, particularly as students from socioeconomically, racially, and culturally disadvantaged backgrounds represent an increasing proportion of the student population.

To address educational inequities, scholars have focused on monitoring opportunity gaps and addressing resource disparities, as well-represented groups tend to access higher-quality resources (Darling-Hammond, 2010; DeSantis, 2012). Among these efforts, access to advanced coursework has drawn significant attention, with researchers employing various methods to explore the accessibility of such coursework and its relationship to educational attainment (e.g., Karp et al., 2007; Long et al., 2009, 2012; Morgan et al., 2018). However, as NASEM (2019) highlighted, educational equity extends beyond resource distribution to addressing systemic disadvantages that hinder marginalized students. Therefore, this study aims to examine the role structural barriers play in access to advanced coursework and its relationship with educational attainment in North Carolina high schools, offering insights to promote equitable education.

Conceptual Framework

The National Academies of Sciences, Engineering, and Medicine (NASEM, 2019) proposed a comprehensive framework of 16 indicators that provide a structured lens to evaluate

educational opportunities and outcomes across schools, districts, and states. As NASEM (2019) cautions, solely monitoring opportunity gaps is insufficient without addressing deeper structural inequities, so their framework underscores the importance of examining both opportunity indicators, such as Access to and Enrollment in Rigorous Coursework, and outcome indicators, such as Postsecondary Readiness.

According to NASEM (2019), educational attainment is a critical goal not only for individual success but for broader societal well-being. Higher educational attainment benefits individuals through improved employment prospects, financial stability, and health outcomes, while society gains from a stronger workforce and reduced public expenditures on healthcare, welfare, and criminal justice (NASEM, 2019). To measure educational attainment, NASEM (2019) advocated for two key indicators: On-Time Graduation and Postsecondary Readiness.

In addition, NASEM (2019) indicates that inequities can manifest in three primary ways: excessive disparities in educational outcomes or resource access, misalignment between resources and student needs, and inadequate efforts to mitigate structural disadvantages such as economic hardship or segregation. In other words, this framework shifts the focus from merely assessing whether students meet specific benchmarks to interrogating the systemic conditions that shape those outcomes.

Types of Advanced Coursework

Advanced Placement (AP) courses are among the most widely recognized forms of rigorous coursework (Morgan et al., 2018). They are designed to mirror college-level work, allowing students to earn college credits, bypass introductory college classes, and develop skills essential for college success (Morgan et al., 2018; NASEM, 2019). To receive college credits, students must take standardized AP exams, with credit eligibility determined by their

performance (Morgan et al., 2018). Research highlights the strong predictive relationship between AP exam participation and postsecondary success (Handwerk et al, 2008; Mattern et al., 2013; Morgan et al., 2018; NASEM, 2019). Mattern et al. (2013) found that 58% of students who took AP exams graduated from college within four years, compared to 38% of those who did not take AP exams. Furthermore, 73% of students who scored high on AP exams graduated within four years, compared to just 37% of those with lower scores. Given the role of AP participation and performance in fostering academic achievement and improving college outcomes, it is crucial for schools to offer AP courses and encourage students to take them (Handwerk et al, 2008; NASEM, 2019).

Another pathway to college credit in high school is through dual enrollment programs, where students take college-level courses that appear on their high school transcripts (Karp et al., 2007; Morgan et al., 2018). These programs offer students early exposure to college coursework, which can enhance their high school GPA and reduce college costs (Karp et al., 2007). In addition, students earning college credits while completing high school tend to be more engaged in the classroom and graduate at higher rates than their peers (Karp et al., 2007). Dual enrollment also works as an effective strategy for promoting postsecondary access and persistence (Karp et al., 2007).

Certain high school STEM courses, such as Algebra II, Calculus, Biology, Chemistry, and Physics, are also rigorous coursework (NASEM, 2019). Success in these courses is associated not only with improved performance in math and science but also with achieving a cumulative GPA of 3.0 or higher, persisting in STEM majors, and earning STEM-related bachelor's degrees (Godwin, 2023). These courses also help students develop necessary skills for high- and middle-skill careers in fields such as healthcare, financial operations, and computer science (Godwin,

2023). Collectively, AP courses, dual enrollment, and advanced STEM coursework have been identified by NASEM (2019) as key measures of Access to and Enrollment in Rigorous Coursework.

Role of Advanced Coursework in Educational Attainment

Many studies have used national samples of high school students to show the positive association between increased academic intensity and standardized test scores (Attewell & Domina, 2008; Lee, 2012), high school graduation (Morgan et al., 2018; Schneider et al., 1998), postsecondary entry and performance (Attewell & Domina, 2008; Karp et al., 2007; Long et al., 2009), reduced remediation needs (Hein et al., 2013), and labor market earnings (Joensen & Nielsen, 2009). Students who engage in rigorous coursework also demonstrate better cognitive strategies necessary for college success and transferable skills such as writing, problem-solving, and study habits (Morgan et al., 2018; Woods et al., 2018). Particularly, Ogut and colleagues (2021), through clustering students based on the courses they had taken, suggested that students who have enrolled in advanced courses are more likely to attend four-year colleges.

Long and colleagues (2012) have suggested several possible explanations for the positive role of advanced coursework in educational attainment. First, such courses offer a more academically demanding curriculum and often attract students who are highly motivated, creating a peer group that fosters academic engagement. Second, schools may have assigned their most skilled teachers to advanced courses, which enhances student learning by increasing the quality of instruction. Third, advanced courses may act as signals of student capability to college admission committees, as these courses are typically limited to students who demonstrate high academic potential.

Disparities in Access to Advanced Coursework

Despite the benefits of advanced coursework, researchers have long recognized that historically disadvantaged groups, such as students of color and those from low-income backgrounds, often miss out on such critical opportunities which can pave the way for success in college and careers (Darling-Hammond, 2010; Handwerk et al., 2008; Morgan et al., 2018; Patrick et al., 2020; Solorzano & Ornelas, 2004; Theokas & Saaris, 2013). Patrick and colleagues (2020) reported that while disadvantaged students perform well in advanced courses when given the opportunity, they remain underrepresented in such courses for two salient reasons: First, schools with predominantly disadvantaged students enroll fewer students in advanced courses compared to schools with smaller disadvantaged populations, and second, schools may disproportionately restrict access to advanced courses for disadvantaged students. Similarly, Handwerk and colleagues (2008), by clustering schools based on factors such as locale, control, size, and percentage of disadvantaged students, found that across all clusters, low-income students were underrepresented among AP exam takers. Even in highly diverse schools, disadvantaged students were less likely to take AP exams than their White and Asian American peers (Handwerk et al., 2008; Solorzano & Ornelas, 2004; Theokas & Saaris, 2013).

Current Study

Previous research has extensively examined the relationship between advanced coursework enrollment and postsecondary outcomes and the disparities in access to advanced coursework (e.g., Karp et al., 2007; Long et al., 2009; Patrick et al., 2020), and some studies have clustered schools and students based on school categorical or coursework enrollment (Handwerk et al. 2008; Ogut et al., 2022). However, few studies have combined these two dimensions to explore how they intersect, leaving gaps in understanding the systemic factors influencing both access and outcomes. As NASEM (2019) points out, monitoring equity requires

identifying and addressing structural barriers, so the current study aims to understand the role of structural barriers in the relationship between access to advanced coursework and educational attainment. This approach allows to explore whether schools with similar resources show different attainment patterns and if structural barriers like percentage of economically disadvantaged students could explain the differences. Specifically, this study aims to address the following research questions:

1. To what extent do access to advanced coursework explain disparities in educational attainment?
2. To what extent do structural barriers play a role in the relationship between access to advanced coursework and educational attainment?

Data and Methods

Databases

This study utilized data from two publicly accessible databases: the Civil Rights Data Collection (CRDC) and the North Carolina Department of Public Instruction (NC DPI).

The CRDC is managed by the U.S. Department of Education's Office of Civil Rights. It collects extensive information from K-12 public schools and districts across the United States, with the primary purpose of monitoring and enforcing civil rights laws. Conducted biennially, the CRDC survey covers a broad range of topics, including student enrollment, educational programs and services, school climate and discipline, teaching and staff information, and early childhood education (U.S. Department of Education, n.d.). The most recent data for public use was collected for the 2020–2021 academic year, which is available as a downloadable zip file and contains three district-level datasets and 31 school-level datasets in CSV format.

The NC DPI hosts data for North Carolina School Report Cards (SCR), which includes

datasets on a variety of topics, such as school performance, attendance, test results, expenditures, and post-secondary outcomes (North Carolina Department of Public Instruction, n.d.). Updated annually to support research purposes, these datasets can be downloaded as a zip file, with each excel file representing a specific data category and providing a longitudinal record. As of the latest update, the data includes information through the 2022–2023 academic year.

Dataset Construction

To construct the dataset for the current analysis, data was extracted and engineered from CRDC and NC DPI, resulting in a total of 31 features (see Table 1 for descriptive statistics) for 239 schools. Among these features are 21 variables representing advanced coursework offered in NC public high schools, seven capturing school-level postsecondary outcomes, two describing school type as categorical variables, and one summarizing the percentage of economically disadvantaged students. Specifically, the advanced coursework variables were derived from 12 CRDC datasets and organized into three main categories: course availability, course enrollment, and college-level course enrollment. An additional CRDC dataset provided information on school characteristics, allowing the author to obtain indicators for whether a school is a magnet school or a charter school. Furthermore, variables related to postsecondary outcomes were derived from four SRC datasets, and percentage of economically disadvantaged students was retrieved from one other dataset. The datafiles used to construct the current dataset were listed in Table 2.

Given that data on advanced coursework is from the 2020–2021 academic year and most high schoolers take advanced and college-level courses in grades 11 and 12, the metrics for postsecondary outcomes were calculated as the average across the three possible graduation years—2021, 2022, and 2023. Missing values were ignored in the calculation by using the

`na.rm` function in R, a common approach for calculating descriptive statistics like mean (GeeksforGeeks, 2022).

Unfortunately, the CRDC and NC DPI datasets do not share a common school identifier, as the CRDC uses National Center for Education Statistics (NCES) IDs while the NC DPI relies on state-specific internal IDs. To integrate variables from these two systems, school names were used as the primary matching criterion, with county names serving as a secondary identifier in cases where school names were not unique. Only schools that could be successfully matched using this approach were included in the final dataset, which led to a sample size of 239.

This merging approach introduced an undesirable data quality issue regarding missing values. Since the two systems exhibit distinct patterns of missing values, the merge resulted in an unsystematic pattern where some schools were missing many advanced course variables, while others were missing most postsecondary outcome variables. Consequently, a decision was made to exclude all observations missing any postsecondary outcome variables, as these metrics are essential for answering the research questions. Other missing values were retained as NA, as most were due to inapplicability, which provides meaningful information about the availability of advanced coursework. For example, a missing value for the percentage of students enrolled in computer science at a school may indicate that the course is not offered.

Hierarchical Cluster Analysis

The primary analysis method employed in this research was hierarchical cluster analysis (HCA), a data mining technique used to group the observations based on their similarity across multiple variables, particularly in high-dimensional datasets (Romesburg, 1984; Jain et al., 1999; Niktin et al., 2022). Hierarchical clustering, as an unsupervised method, focuses on uncovering hidden patterns within unlabeled datasets rather than accurately predicting unseen samples from

the same probabilistic distribution (Bowers, 2007; Jain et al., 1999; Xu & Wunsch, 2005). The process begins by treating each data object as its own individual cluster, and clusters are then progressively merged based on their similarity until a hierarchical structure is formed (Jain et al., 1999; Xu & Wunsch, 2005). As such, hierarchical clustering preserves the original data points and data structure instead of summarizing it by the mean or reporting broad trends (Bowers, 2007, 2010; Eisen et al., 1998). Compared to alternative unsupervised clustering techniques such as k-means, HCA is advantageous because it does not require specifying the number of clusters in advance and thus avoids the risk of arbitrary decisions that could compromise the accuracy of the results (Eisen et al., 1998; Jain et al., 1999; Niktin et al., 2022). Therefore, the current study adopted HCA given the paucity of theoretical or empirical evidence to support the assumptions about the optimal number of clusters.

Before conducting HCA on the current dataset, all continuous variables were transformed to z-scores to standardize the data, ensuring that each variable contributed equally to the clustering process regardless of its original scale (Romesburg, 1984, as cited in Bowers, 2010). HCA requires two main metrics: a distance metric which measures the similarity or dissimilarity between observations, and an agglomeration method which determines how clusters are merged at each step of the analysis (Bowers, 2010). Following the recommendations from previous literature in the field of education, this study used uncentered correlation as the distance metric and average linkage as the agglomeration metric (Bowers, 2007, 2010).

Specifically, uncentered correlation assumes a mean of zero and differs from Pearson correlation by adjusting for patterns separated by a constant, thus is robust for comparing data trends that may be offset by a constant value (Bowers, 2010). Average linkage was chosen for two primary reasons: First, it is robust to missing values which are commonly encountered in

education datasets, as the algorithm ensures that observations with similar missing patterns cluster together because of similar computed distances (Bowers, 2007, 2010). Second, average linkage, by calculating the average distances between all pairs of observations, ensures that clustering reflects the relationship between clusters based on the overall characteristics of each single school rather than being skewed by a few outliers or extreme comparisons (Bowers, 2010; Niktin et al., 2022).

The generalized form of equation for uncentered correlation is as follows (Bowers, 2010, p.18):

$$r(x_i, y_i) = \frac{1}{n} \sum_{i=1}^n \left(\frac{x_i}{\sigma_x^{(0)}} \right) \left(\frac{y_i}{\sigma_y^{(0)}} \right)$$

in which

$$\sigma_x^{(0)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i)^2}$$

and

$$\sigma_y^{(0)} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i)^2}$$

The average linkage metric can be calculated using the equation below (Bowers, 2010, p.18):

$$D(A, B) = \frac{1}{n_A n_B} \sum_{i=1}^{n_A} \sum_{j=1}^{n_B} r(x_i, y_j)$$

HCA Heatmaps (Clustergrams)

Hierarchical clustering results can be visualized using a dendrogram, where horizontal lines represent cluster merging, shorter lines indicate higher similarity, and the dendrogram

grows hierarchically until all data points form a single cluster (Bowers, 2010). Combined with heatmaps, which represent each observation's values across variables as color blocks, a clustergram was derived, enabling intuitive pattern recognition through two contrasting colors: one representing positive values and the other negative, with greater intensity in either color indicating larger absolute values (Bowers, 2010; Eisen et al., 1998; Lee et al., 2016). In the current analysis, the clustergram visualized each school's standardized values for advanced coursework and postsecondary outcomes, with redder color denoting higher values, bluer indicating the lower, and dendrograms grouping similar schools by rows (see Figure 1). In addition, to investigate the potential structural barriers in postsecondary outcomes, annotations about whether a school is a charter school or magnet school, and percentage of economically disadvantaged students were added.

Bubble Plot

A follow-up bubble plot was created alongside the clustergram to explore patterns within clusters with better educational attainment and to analyze how structural barriers influence these patterns. Specifically, the plot focused on the relationship between AP course enrollment and college enrollment, two variables widely studied in the literature. Drawing inspiration from the visualization presented by Schwabish (2021, p.40), quadrants were incorporated into the bubble plot. These quadrants were defined by the mean values of the two variables, corresponding to a standardized value of 0 in the heatmap. To contextualize these patterns, the size of each bubble represented the percentage of economically disadvantaged students in each school, while the color indicated the clusters with better educational attainment.

R Packages

Data preparation, HCA, and visualizations were performed using R (Version 4.3.2). The

R packages used were listed in Table 3. Particularly, the HCA heatmap was created using the `ComplexHeatmap` package (Gu et al., 2016; Gu, 2022).

Results

Clustergram

As illustrated by Figure 1, four schools clusters were identified. Cluster 2 and 3 revealed interesting patterns that are insightful for understanding the relationship between access to advanced coursework and postsecondary outcomes under structural context. In Cluster 2, two groups of schools emerged with higher availability of advanced STEM courses but showed different patterns in STEM course enrollment, college-level course participation, and educational attainment outcomes. A closer examination indicated that schools with more economically disadvantaged students, represented by the top segment in Cluster 2, had lower participation in advanced coursework and college enrollment. Additionally, some schools in this segment are magnet schools, which are designed to offer specialized programs, potentially explaining the higher availability of advanced STEM courses. However, despite the availability, students in these schools were not actively enrolled in these courses, highlighting a potential disconnection between course availability and actual participation. This trend signals that the presence of advanced coursework alone might not bridge the equity gap if students from disadvantaged backgrounds do not participate in these opportunities.

Despite having limited availability of STEM courses, lower enrollment, and inactive participation in college-level courses, schools in Cluster 3 demonstrated strong educational attainment outcomes. This pattern suggested that positive postsecondary outcomes are achievable even when access to rigorous coursework is limited. However, the annotations in the clustergram indicated that most of these schools are charter schools with a low percentage of economically

disadvantaged students. Notably, several schools in this cluster had a higher percentage of economically disadvantaged students. Further investigation into these schools could provide valuable insights for school leaders into practices that support high educational attainment for students from disadvantaged backgrounds.

Schools in Cluster 1 displayed patterns opposite to the top segment of Cluster 2, as they had fewer STEM courses available, yet students actively participate in those courses as well as in college-level courses. Students in these schools also demonstrated better educational attainment, suggesting that active engagement in advanced coursework is more critical than simply the number of available courses. In contrast, schools in cluster 4 generally showed lower values across most advanced course features, except for enrollment in Algebra 2, Biology, and Geometry. However, educational attainment metrics in these schools were relatively low.

Bubble Plot

The bubble plot, Figure 2, enabled a closer exploration of the relationship between enrollment in AP courses and college enrollment. The gray dots, which represented the schools in clusters with worse postsecondary outcomes, i.e., Cluster 2 and Cluster 4, had bigger sizes, indicating higher percentage of economically disadvantaged students. Even with more students enrolled in AP course, as shown in the bottom right quadrant, the college enrollment rates in these schools remained lower than schools in Cluster 1 and Cluster 3. This demonstrated that economic disadvantages created some obstacles for college enrollment, even when opportunities for advanced coursework like AP classes were made available and accessible. The data highlighted the persistent barriers faced by economically disadvantaged students in translating academic opportunities into postsecondary success.

However, some schools in Cluster 1 and Cluster 3, such as Henderson Collegiate, Triad

Math and Science Academy, and City of Medicine Academy, had higher economically disadvantaged students but still exhibited better postsecondary attainment, regardless of their students' exposure to advanced coursework like AP courses. This observation suggested that differences in outcomes may be associated with additional factors, such as institutional practices or support systems, that contributed to higher levels of success for economically disadvantaged students.

Discussion

Through HCA, this study identified four distinct clusters of schools, each exhibiting unique configurations of advanced course offerings, enrollment in advanced coursework, and educational attainment. The clustergram (Figure 1) showed that Cluster 1 had high values in advanced coursework variables and high values in educational attainment features (High-High), and cluster 2, 3, and 4 can be respectively characterized as High-Low, Low-High, and Low-Low. These patterns provided invaluable insights into the relationship between rigorous coursework and student success after high school and how structural barriers, like economic disadvantage, played a role in this relationship.

Findings and Implications

Across all clusters, course availability, i.e., the number of STEM courses, did not appear to correlate with students' postsecondary attainment. For instance, while schools in Cluster 2 offered an abundance of courses, they exhibited two distinct patterns: The top segment had lower advanced course enrollment and postsecondary attainment, whereas the bottom segment showed higher values. This indicated that offering courses alone does not ensure participation, consistent with previous research suggesting that students may not engage in advanced coursework even when it is available, thereby failing to achieve the scores necessary for postsecondary success

(Handwerk et al., 2008). Notably, schools in the top segment of Cluster 2 had a higher percentage of economically disadvantaged students, which replicated the previous findings that schools with predominantly disadvantaged populations tend to enroll fewer students in advanced courses (Patrick et al., 2020).

What ultimately matters more for students' postsecondary outcomes was their participation in those rigorous courses. Schools in Clusters 1 and 2 demonstrated higher enrollment rates in advanced coursework, highlighting that addressing disparities in access remained a critical area for improvement (Darling-Hammond, 2010; DeSantis, 2012). However, accessibility alone did not fully explain postsecondary outcomes. For example, while both Cluster 3 and Cluster 4 had lower enrollment rates in advanced coursework, schools in Cluster 3 achieved higher educational attainment. This disparity suggested that simply making advanced courses available and accessible is necessary but insufficient; students' postsecondary success could still vary even with equal access (NASEM, 2019). A key distinction lied in the economic demographics of these clusters—Cluster 4 had a higher proportion of economically disadvantaged students, supporting prior findings that low-income students consistently lag behind their more advantaged peers (Handwerk et al., 2008; Patrick et al., 2020). This underscored the need for schools to provide tailored resources and support systems to create truly equitable educational opportunities, as advocated by NASEM (2019).

The bubble plot provided an opportunity to closely examine schools with high educational attainment, i.e., those in Clusters 1 and 3. It was evident that schools in these two clusters had fewer economically disadvantaged students compared to the other clusters, as indicated by the smaller bubble sizes. This observation highlighted the persistent structural barriers to achieving outcome equity, even under conditions of resource equality (NASEM, 2019). However, in

Cluster 3, where schools had limited resources yet still demonstrated high performance, a few schools met the eligibility criteria for the Community Eligibility Provision as designated by NC DPI—indicating that over 40% of their students are economically disadvantaged. Examples included Henderson Collegiate, Triad Math and Science Academy, and Northside High in Onslow. These schools stood out for their strong educational attainment despite predominantly serving economically disadvantaged students. This suggests that these schools might have implemented unique strategies to support their minority students, allowing them to achieve notable outcomes even without high enrollment in advanced coursework. The specific initiatives employed by these schools remained unknown from the current data. Future research should explore these questions by incorporating additional data on these schools or conducting case studies. Interviews with school leaders could provide qualitative perspectives on the practices and interventions that enabled schools with predominantly economically disadvantaged students to achieve high educational outcomes. Such studies could offer valuable insights into effective strategies for promoting equity and success in similar contexts.

Limitations and Future Directions

Despite some insights into the role of structural barriers in the relationship between access to advanced coursework and educational attainment, this analysis has some limitations. The absence of postsecondary outcome data in CRDC necessitated the use of NC DPI data, leading to two main challenges. First, the lack of a common identifier between the two systems required reliance on school names for merging, which introduced a high risk of error. Second, the decision to retain only schools with complete postsecondary outcome data unfortunately led to a substantial reduction in sample size.

Although this study incorporated many relevant features related to advanced coursework,

educational attainment, and basic school characteristics, it excluded critical factors related to structural barriers, such as detailed demographic compositions for each school. Moreover, focusing solely on advanced coursework to investigate disparities in educational attainment was inadequate, as inequities in access could manifest in various forms, including high-quality instruction, effective teaching, and non-academic support systems (Darling-Hammond, 2010; NASEM, 2019). These inequities could emerge as early as the pre-kindergarten learning environment, further compounding disparities over time (Darling-Hammond, 2010; Lee, 2012; NASEM, 2019). Therefore, future research should expand the dataset to include a broader range of variables, capturing additional school characteristics and incorporating measures for other NASEM indicators to provide a more comprehensive understanding of educational inequities.

Additionally, since this study is purely descriptive and based on a relatively small dataset from one state, its scope and explanatory power are inherently limited. As Weinstein (2008) noted, HCA provides only a first-order insight into data, meaning that complex nonlinear relationships among a small number of samples are unlikely to emerge. Furthermore, as Xu and Wunsch (2005) explained, HCA cannot offer an accurate characterization of unobserved samples, limiting the generalizability of the findings. Moreover, since this analysis was conducted at the school level, it was difficult to determine which students were enrolled in the advanced courses. Even in schools with high enrollment rates in advanced coursework, it is possible that economically disadvantaged students are still disproportionately restricted from enrollment, as suggested by Patrick and colleagues (2020). Therefore, future studies should, if possible, conduct this analysis at the student level to better capture disparities in access and participation.

Table 1*Descriptive Statistics of Variables in the Current Dataset*

Variable	Description	%NA	Min	Max	Mean	SD
num_classes_adv_math	# of advanced math classes offered.	0.42	0.00	87.00	7.65	9.39
num_classes_algebra2	# of Algebra 2 classes offered.	0.42	0.00	68.00	15.73	11.09
num_classes_calculus	# of calculus classes offered.	0.42	0.00	24.00	2.72	3.51
num_classes_biology	# of biology classes offered.	0.00	0.00	98.00	20.10	15.37
num_classes_chemistry	# of chemistry classes offered.	0.00	0.00	63.00	9.73	9.65
num_classes_geometry	# of geometry classes offered.	0.42	0.00	73.00	16.82	12.12
num_classes_cs	# of computer science classes offered.	0.00	0.00	47.00	4.82	6.95
num_classes_physics	# of physics classes offered.	0.00	0.00	31.00	2.28	3.69
pct_enr_adv_math	% of students enrolled in advanced math classes.	3.35	0.46	30.97	8.47	5.68
pct_enr_algebra2	% of students enrolled in Algebra 2 classes.	1.26	3.01	33.06	20.36	5.73
pct_enr_calculus	% of students enrolled in calculus classes.	20.50	0.10	17.28	2.31	2.18
pct_enr_biology	% of students enrolled in biology classes.	1.67	3.34	38.36	23.60	7.05
pct_enr_chemistry	% of students enrolled in chemistry classes.	1.26	1.53	33.51	11.67	5.87
pct_enr_geometry	% of students enrolled in geometry classes.	0.84	4.65	100.00	22.80	8.08
pct_enr_cs	% of students enrolled in CS classes.	26.36	0.05	42.58	4.77	5.37
pct_enr_physics	% of students enrolled in physics classes.	35.15	0.05	18.75	2.65	3.52
pct_enr_ap	% of students enrolled in at least one AP course.	2.09	0.18	56.87	15.26	10.85
pct_enr_ap_math	% of students enrolled in AP mathematics courses.	15.06	0.13	21.87	3.67	3.39
pct_enr_ap_science	% of students enrolled in AP science courses.	23.01	0.09	35.80	2.84	3.83
pct_enr_ap_cs	% of students enrolled in AP CS courses.	49.79	0.05	19.75	2.02	3.02
pct_enr_dual	% of students in dual enrollment programs.	19.67	0.07	99.51	10.69	11.56
sat_participation	% of students participating in the SAT.	0.00	0.05	0.96	0.23	0.16
sat_performance	Average SAT score among participating students.	0.00	838.50	1328.67	1095.63	76.56
ap_participation	% of students participating in AP exams.	0.00	0.01	0.57	0.16	0.11
ap_performance	% of AP Exam scores of 3 or above.	0.00	0.05	0.86	0.46	0.18
graduation_rate_4yr	Four-year high school graduation rate.	0.00	54.27	95.00	87.27	6.26
graduation_rate_5yr	Five-year high school graduation rate.	0.00	56.67	95.00	88.57	5.63
pct_collge_enrollment	% of graduates enrolling in college.	0.00	0.25	0.94	0.63	13.34
Magnet School	Indicates whether the school is a magnet school.	0.00	NA ^a	NA	NA	NA
Charter School	Indicates whether the school is a charter school.	0.00	NA	NA	NA	NA
% Eco Disadv Students	% economically disadvantaged students.	0.00	5.00	67.10	35.96	14.04

Note. This table lists all the variables used in the current dataset for hierarchical cluster analysis, and the order of variables is the same as the column order in the heatmap (see Figure 1). # represents count, and % denotes percentage.

^a NA stands for Not Applicable. It is used for the two school type indicators whose values are Yes and No.

Table 2*Public Datasets Utilized in Dataset Construction*

Datafile Name	Number of Variables Used ^a	Number of Features Derived ^b
CRDC		
Enrollment.csv ^c	4	0
Advanced Mathematics.csv	3	2
Algebra II.csv	3	2
Calculus.csv	3	2
Geometry.csv	3	2
Biology.csv	3	2
Chemistry.csv	3	2
Physics.csv	3	2
Computer Science.csv	3	2
Advanced Placement.csv	8	4
Dual Enrollment.csv	2	1
School Characteristics.csv	2	2
NC DPI		
rcd_ap.xlsx	2	2
rcd_sat.xlsx	2	2
rcd_acc_cgr.xlsx	2	2
rcd_college.xlsx	1	1
rcd_acc_eds.xlsx	1	1

Note. This table demonstrates how public datasets from CRDC and NC DPI help construct the current dataset for analysis. From each dataset, only a selection of variables was used to engineer the new features.

^a The values in this column represents the number of variables used for feature engineering purposes, excluding the variables used to identify schools.

^b The values in this column represents the number of new features derived from the public dataset, excluding the variables used to identify schools.

^c This dataset was only used to calculate how many high schoolers are in each school. Since many data points from the public datasets are number of students, this dataset helps to calculate the percentage of students.

Table 3*R Packages Used in Current Analysis*

Package	Version
base	4.3.2
Biobase	2.62.0
BiocGenerics	0.48.1
circlize	0.4.16
cluster	2.1.6
ComplexHeatmap	2.21.1
ggrepel	0.9.6
ggtext	0.1.2
grateful	0.2.10
hopach	2.62.0
janitor	2.2.0
tidyverse	2.0.0
zip	2.3.1

Figure 1
Hierarchical Cluster Analysis Heatmap of 239 NC High Schools

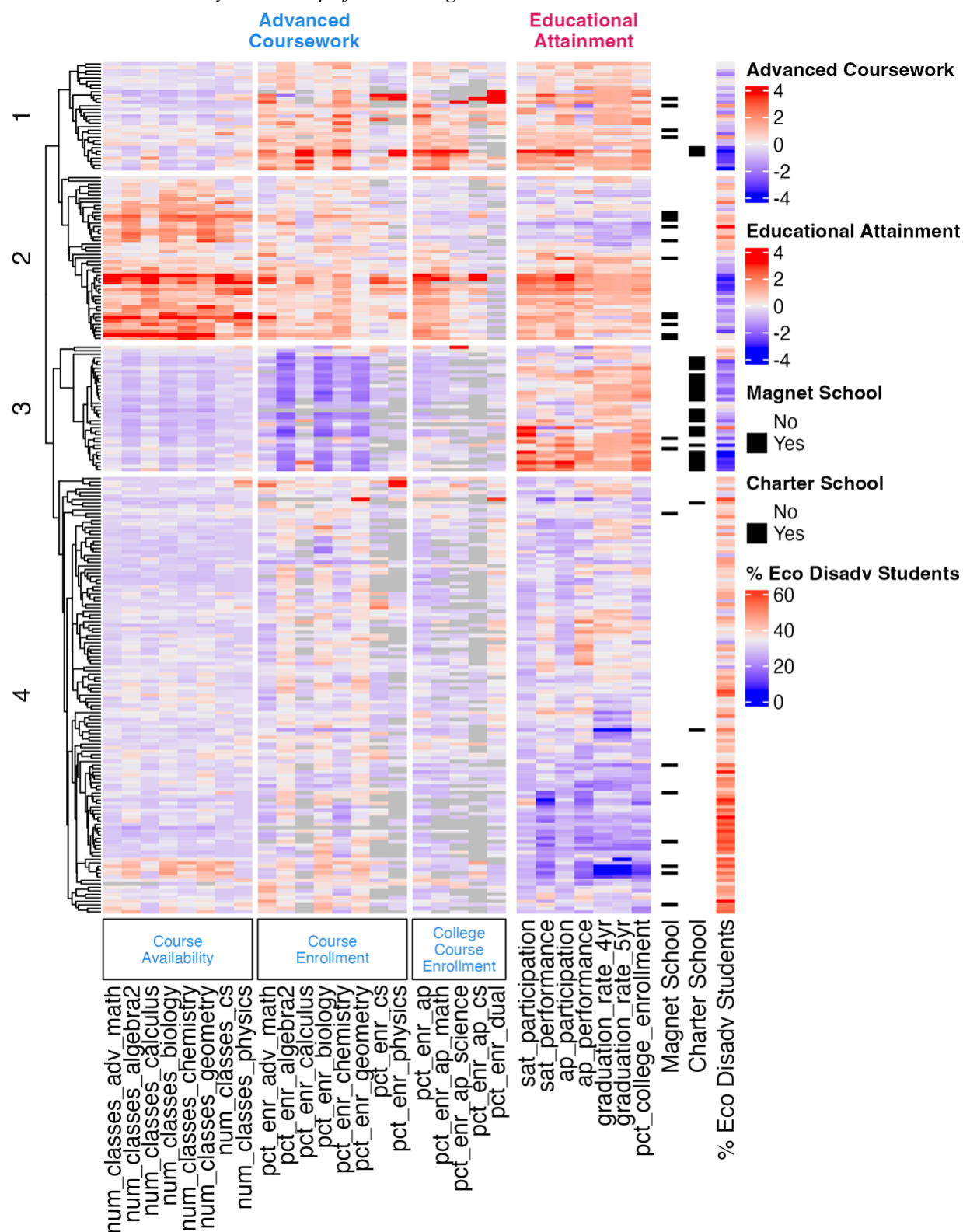
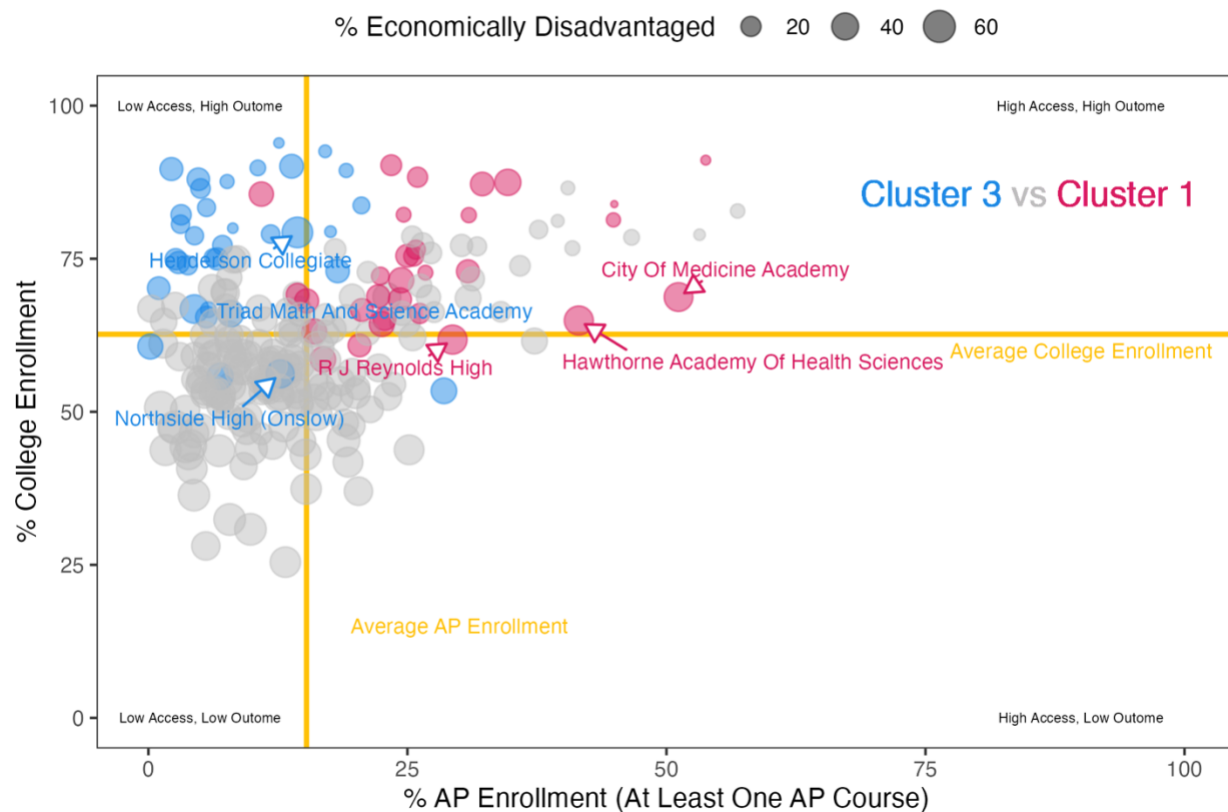


Figure 2

Bubble Plot: % AP Enrollment vs % College Enrollment

Equal Access, Unequal Outcome

Economically disadvantaged schools have lower college enrollment rates, even with similar AP course enrollment levels



NC DPI: schools with at least 40% economically disadvantaged students are eligible for Community Eligibility Provision

References

- Attewell, P., & Domina, T. (2008). Raising the bar: Curricular intensity and academic performance. *Educational Evaluation and Policy Analysis*, 30(1), 51–71.
<https://doi.org/10.3102/0162373707313409>
- Bowers, A. J. (2007). *Grades and data driven decision making: Issues of variance and student patterns* [Doctoral dissertation, Michigan State University]. MSU Libraries Digital Repository. <https://doi.org/doi:10.25335/bj21-3g32>
- Bowers, A. J. (2010). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierarchical cluster analysis. *Practical Assessment, Research, and Evaluation*, 15(7).
<https://doi.org/10.7275/r4zq-9c31>
- Darling-Hammond, L. (2010). The anatomy of inequality: How the opportunity gap is constructed. In *The flat world and education: How America's commitment to equity Will determine our future* (pp. 27–65). Teachers College Press. Derived from
<https://www.uvm.edu/~rgriffin/Darling-Hammond-Anatomy.pdf>
- DeSantis, J. (2012). The flat world and education: A review [Review of the book *The flat world and education: How America's commitment to equity Will determine our future*, by L. Darling-Hammond]. *International Journal of Progressive Education*, 8(1), 35–38.
<https://ijpe.inased.org/makale/2539>
- Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America*, 95(25), 14863–14868.
<https://doi.org/10.1073/pnas.95.25.14863>

- Gamoran, Adam. (2001). American schooling and educational inequality: A forecast for the 21st century. *Sociology of Education*, 74, 135–153. <https://doi.org/10.2307/2673258>
- GeeksforGeeks. (2022, February 1). *How to use na.rm in R?*
<https://www.geeksforgeeks.org/how-to-use-na-rm-in-r/>
- Godwin, J. (2023, November 8). Assessing the state of STEM education this National STEM Day. *ACT*. <https://leadershipblog.act.org/2023/11/national-stem-day.html>
- Gu, Z. (2022). Complex heatmap visualization. *iMeta*, 1(3), e43. <https://doi.org/10.1002/imt2.43>
- Gu, Z., Eils, R., & Schlesner, M. (2016). Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics*, 32(18), 2847–2849.
<https://doi.org/10.1093/bioinformatics/btw313>
- Handwerk, P., Tognatta, N., Coley, R. J., & Gitomer, D. H. (2008). *Access to success: Patterns of advanced placement participation in U.S. high schools*. Educational Testing Service.
<https://files.eric.ed.gov/fulltext/ED505556.pdf>
- Hein, V., Smerdon, B., & Sambolt, M. (2013). *Predictors of postsecondary success*. American Institutes for Research. <https://files.eric.ed.gov/fulltext/ED555671.pdf>
- Jain, A. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: A review. *ACM Computing Surveys (CSUR)*, 31(3), 264–323. <https://doi.org/10.1145/331499.331504>
- Joensen J. S., & Nielsen, H. S. (2009). Is there a causal effect of high school math on labor market outcomes? *Journal of Human Resources*, 44(1), 171–198.
<https://doi.org/10.3368/jhr.44.1.171>
- Karp, M. M., Calcagno, J. C., Hughes, K. L., Jeong, D. W., & Bailey, T. R. (2007). *The postsecondary achievement of participants in dual enrollment: An analysis of student outcomes in two states*. National Research Center for Career and Technical Education.

<https://files.eric.ed.gov/fulltext/ED498661.pdf>

Lee, J. (2012). College for all: Gaps between desirable and actual P–12 math achievement trajectories for college readiness. *Educational Researcher*, 41(2), 43–55.

<https://doi.org/10.3102/0013189X11432746>

Lee, J. E., Recker, M., Bowers, A. J., & Yuan, M. (2016). Hierarchical cluster analysis heatmaps and pattern analysis: An approach for visualizing learning management system interaction data. In T. Barnes, M. Chi, & M. Feng (Eds.), *Proceedings of the 9th International Conference on Educational Data Mining*, 603–604. Derived from

https://www.educationaldatamining.org/EDM2016/proceedings/paper_34.pdf

Long, M. C., Conger, D., & Iatarola, P. (2012). Effects of high school course-taking on secondary and postsecondary success. *American Educational Research Journal*, 49(2), 285–322. <https://doi.org/10.3102/0002831211431952>

Long, M. C., Iatarola, P., Conger, D. (2009). Explaining gaps in readiness for college-level math: The role of high school courses. *Education Finance and Policy*, 4(1), 1–33.

<https://doi.org/10.1162/edfp.2009.4.1.1>

Mattern, K. D., Marini, J. P., & Shaw, E. J. (2013). *Are AP[®] students more likely to graduate from college on Time*. College Board. <http://files.eric.ed.gov/fulltext/ED556464.pdf>

Morgan T. L., Zakhem, D., & Cooper W. (2018). From high school access to postsecondary success: An exploratory study of the impact of high-rigor coursework. *Education Sciences*, 8(4), 191. <https://doi.org/10.3390/educsci8040191>

National Academies of Sciences, Engineering, and Medicine. (2019). *Monitoring educational equity*. <https://doi.org/10.17226/25389>.

Nitkin, D., Ready, D. D., & Bowers, A. J. (2022). Using technology to personalize middle school

- math instruction: Evidence from a blended learning program in five public schools. *Frontiers in Education*, 7, Article 646471. <https://doi.org/10.3389/feduc.2022.646471>
- North Carolina Department of Public Instruction (n.d.). *Data & Reports*. Retrieved October 28, 2024, from <https://www.dpi.nc.gov/data-reports>
- Ogut, B., Circi, R., & Yee, D. (2021). *Why does high school coursework matter? The case for increasing exposure to advanced courses*. American Institutes for Research. <https://www.air.org/sites/default/files/2022-06/22-18169-Advanced-Coursetaking-Brief-061422.pdf>
- Patrick, K., Socol, A., & Morgan, I. (2020). *Inequities in advanced coursework: What's driving them and what leaders can do*. The Education Trust. <https://files.eric.ed.gov/fulltext/ED603195.pdf>
- Romesburg., H. C. (1984). *Cluster analysis for researchers*. Lifetime Learning Publications.
- Schneider, B., Swanson, C. B., & Riegle-Crumb, C. (1997). Opportunities for learning: Course sequences and positional advantages. *Social Psychology of Education: An International Journal*, 2(1), 25–53. <https://doi.org/10.1023/A:1009601517753>
- Schwabish, J. (2021). *Better data visualizations: A guide for scholars, researchers, and wonks*. Columbia University Press.
- Solorzano, D. G., & Ornelas, A. (2004). A critical race analysis of Latina/o and African American Advanced Placement enrollment in public high schools. *The High School Journal*, 84(3), 15–26. <https://www.jstor.org/stable/40364293>
- Theokas, C., & Saaris, R. (2013). Finding America's missing AP and IB students. The Education Trust. <https://files.eric.ed.gov/fulltext/ED578802.pdf>
- U.S. Department of Education. (n.d.). *Recent data files*. Civil Rights Data Collection. Retrieved

October 28, 2024, from <https://civilrightsdata.ed.gov/data>

- Woods, C. S., Park, T., Hu, S., & Bertrand Jones, T. (2018). How high school coursework predicts introductory college-level course success. *Community College Review*, 46(2), 176–196. <https://doi.org/10.1177/0091552118759419>
- Xu, R., & Wunsch, D. (2005). Survey of clustering algorithms. *IEEE Transactions on Neural Networks*, 16(3), 645–678. <http://doi.org/10.1109/TNN.2005.845141>