A Grading Analysis Tool that Empowers Educators via Data-Driven Decision Making

ABSTRACT
Universities generate vast amounts of student performance data, but often lack the tools to extract actionable insights from the data. Our Grading Analysis Tool (GAT) is an open-source, publicly accessible, Python-based software package that facilitates data-driven analysis of student course grade data. GAT provides an easy-to-use graphical user interface that supports analyses at the department, course, student, and instructor level. This paper describes GAT and its capabilities. In addition, it provides several sample analyses using actual education data to show the utility of the tool and the types of visualizations that it can generate.

Keywords
Grade Analysis Tool, Educational Analysis, University Education

1. INTRODUCTION
Most universities and colleges were established prior to the start of the Information Age and have been relatively slow to adopt modern data-driven approaches to academic decision making and the setting of academic policies. Although industries such as finance and healthcare have embraced advanced analytics, higher education has often relied on manual reports and spreadsheets to evaluate student performance, faculty effectiveness, and course outcomes.

Universities generate large amounts of student data that is often distributed to deans and department chairs without effective tools for analyzing the data. Ellucian Banner, one of the most widely used enterprise student information systems by higher education [3] primarily functions as a student records management system rather than an analytical tool. While it provides some structured reports on grades, enrollment, and student performance to general administrators, it lacks interactive visualizations or exploratory data analysis features that allow for deeper insights into grading trends. Furthermore, access to certain Banner screens with high level aggregations is highly restricted, with most data centralized within institutional research offices. Database technology and scripting languages can be a great tool for analyzing the data, but require a level of programming knowledge that makes them inaccessible to most academic administrators. The lack of customized analytical tools poses a significant challenge, as data-driven insights can influence departmental funding, curricula, and faculty evaluations. Without an effective way to analyze this data, universities may make poor decisions that will impact efficiency, resource allocation, and educational quality.

The Grading Analysis Tool (GAT) addresses these challenge by providing a tool specifically designed as an open source solution to analyze educational data. The development of GAT was inspired by the research presented in [6]. Their work highlighted the need for an accessible and user-friendly tool for analyzing educational data, as their figures were done in an ad-hoc manner using code written specifically for the purpose, which may not be reused. Built on Pandas [4], it includes an easy-to-use graphical user interface (GUI) that simplifies data exploration and analysis. Traditional spreadsheet software lacks the ability to handle large datasets efficiently, often requiring manual filtering and formula-based calculations that can introduce errors. GAT, in contrast, provides automated aggregation capabilities tied to common education elements (e.g., course sections, courses, instructors, departments) and visualization features tied to these elements. The tool offers the performance and scalability to handle the large datasets common in higher education.

GAT is a targeted tool, as it focuses mainly, but not exclusively, on student grades and grading patterns. A department chair could use GAT to compare instructor grading over all sections of a course to identify instructors that grade leniently or harshly, while a dean could use it to identify academic departments with grading patterns out of the norm. This paper describes the GAT tool and then illustrates some of its capabilities through real-world case studies.

2. TOOL DESCRIPTION
GAT's core functionality lies in its ability to aggregate data across various dimensions and calculate a variety of metrics along these dimensions. These metrics include the number of sections, number of courses, average GPA, and a variety of statistical measures for various metrics, such as the standard deviation or kurtosis [2] associated with a GPA distribution.

For example, if the chosen dimension is “Department” (or “Instructor”) then GAT will calculate the average GPA and number of sections and courses associated with each department (or instructor). This uniformity makes the tool easier to use.

2.1 Accessing the Tool

The Grading Analysis Tool is publicly available via GitHub [1](anonymous.4open.science/r/Grading-Analysis-Tool), where the installation process is laid out. The tool is written in Python and is open-source so that it can be extended and modified. The GitHub repository also includes a manual and a sample data set so that users can immediately explore the tool. The dataset used in our analysis consists of 446,067 student records, across 63 departments covering 2,636 unique courses and 2,237 unique instructors. While this dataset was used in our analysis, the version uploaded to our GitHub repository has been further anonymized to protect privacy and represents only a fraction of the real data due to GitHub file size limits.

2.2 Input

This tool is designed to accept post-processed academic data at the student–course level. Each row of the dataset should represent a single student’s record for one course, including the grade received and associated metadata such as semester, major, and department. Because raw data from universities is often not formatted this way, some manual preprocessing—either via a custom script or spreadsheet software—is usually required before using the tool. The expected input format is summarized in Table 1.

Table 1: Dataset Fields Expected by GAT

Field	Example
Student ID	S12345
Department	Theatre
Course Number	4920
Credit Hours	4
Term	Fall 2020
Class Size	12
Faculty ID	F26352
Course Title	Intro To Sociology
Student Level	Freshmen
Major	Dance
Letter Grade	A
Numeric Grade	4.0
Course Code	Theatre 4920
Semester	201810: Fall 2018

2.3 Output

The analyses that GAT supports generates a variety of output. The primary output are graphical figures, such as histograms, scatter plots, and line charts. Some options can output heatmaps and parallel plots.

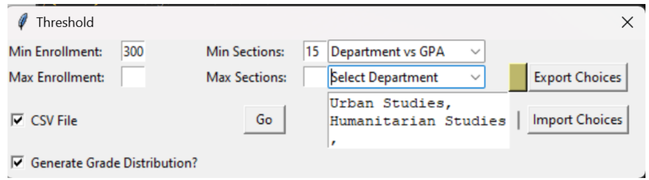

Figure 1: Pre-analysis Customization Option Box

After selecting an analysis type, the tool displays to users the threshold popup seen in Figure 1. This will always be displayed for any analysis type and will usually look the same. These range from 2 to 7 different visualization options depending on the selected analysis type (Sec. 2.4) with a total of 41 separate queries across all analysis types.

Users can assign colors to different categories by selecting the color box, which then prompts them to label the category. This categorization is not necessary, but if done, it allows for higher level views provided by the Create Groups option referenced in Table 2, and provides a cleaner and easier to interpret plot. After creating a custom query, users can save their predefined categories and reuse them across any analysis option that supports these inputs via the “Import Choices” button.

The figures are displayed after pressing ‘Go’ (Fig. 1). The tool generates a primary output window displaying the results of the selected visualization, tailored to the chosen analysis type (Sec. 2.4). If the ’grade distribution’ option is enabled, a second window showing a parallel plot of letter grade distribution, which visualizes the frequency of each letter grade in descending order (A, B+, B, B-, etc.) (displayed in Figure 7).

The primary output window provides seven more customization options, detailed in Table 2, allowing users to interactively explore their data from multiple perspectives, alongside refining the visualizations to focus on specific aspects of the data, facilitating a deeper understanding of the underlying trends. Users can also save figures directly from this window to various file formats (.png, .jpeg, .pdf, .svg) and destinations, facilitating easy sharing and storage of visualizations.

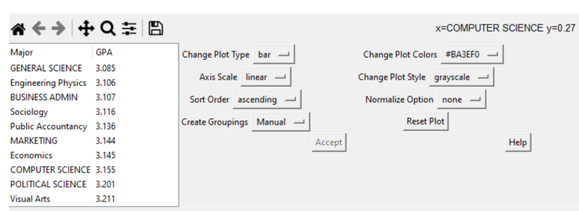

Figure 2: Plotting Options

GAT automatically appends a plot customization panel (shown in Figure 2) to the resulting visualization with its options described in Table 2. Alongside these options, we have a table representing the current state of what is being plotted. Here, you can see the data that is being viewed, and if you click on a item in the table, the selected item is highlighted on the plot you choose as a red dot.

Table 2: GAT Plot Customization Options

Option	Description
Change Plot Type	Select bar, scatter, or line plot
Axis Scale	Y-Axis Scale: Linear, Log, Asinh, Symlog
Sort Order	Ascending, Descending, Random, None
Create Groupings	Group data: Manual, by Colors
Change Plot Colors	Modify category colors
Change Plot Style	Select Matplotlib background style
Normalize Option	MinMax, Zscore, Robust, MaxAbs, Log

Users can also export the processed data as a CSV file if the 'CSV File' checkbox is checked (Fig. 1). The saved data can be especially useful as it can represent the original data after various filtering, thresholding, and aggregation operations are applied. The output format is explained in Table 3.

Table 3: Definitions of Aggregated Metrics Provided by GAT

Metric	Definition
Dimension	Primary grouping variable (e.g., Major, Department).
GPA	Average GPA within dimension.
GPAW	Credit weighted GPA for dimension.
Standard Deviation	Variability of GPA (or GPAW).
Kurtosis	Tailedness of the GPA distribution, indicating outlier influence.
Skewness	Asymmetry of the GPA distribution.
CoV (%)	Coefficient of variation, calculated as the standard deviation divided by the mean GPA.
ModeGPA	Most frequently occurring GPA value.
Grade Count	Count of letter grade (A, A-, B+, etc.).
GPA(W) Change	Average change in weighted GPA across semesters.

2.4 Capabilities & Features

GAT supports 9 different high-level analysis options, each of which supports a variety of queries. Below we briefly describe each type of analysis and provide an example of one specific query that it can satisfy.

- **Department Analysis:** Analyze aggregated course statistics by department.
  - **Example:** Compare the average GPA across different departments.
- **Instructor Analysis:** Examine relative grading trends and statistics for instructors.
  - **Example:** Evaluate the distribution of grades given by instructors.
- **Major Analysis:** Compare academic performance across different majors.
  - **Example:** Assess the average change over time between majors and enrollment.
- **Course Analysis:** Study aggregated statistics of sections for individual courses.
  - **Example:** Analyze how course enrollment size correlates with average GPA of the course.
- **Section Analysis:** Investigate performance variations at the section level.
  - **Example:** Examine the impact of class size on GPA.
- **Student Level Analysis:** Explore GPA trends based on students' academic levels (e.g., Freshman, Sophomore).
  - **Example:** See how average GPA changes with student level.
- **Course Level Analysis:** Examine performance trends across different course levels, derived from Course Number (Table. 1) (e.g., 1000 level vs. 4000 level).
  - **Example:** Compare enrollment trends between lower- and upper-level courses.
- **Student-Course Level Analysis:** Perform multi-dimensional analysis of student-course interactions.
  - **Example:** See the average GPA of each course level for each student level (e.g., Average GPA of sophomores in 2000-level courses).
- **Student Analysis:** Investigate GPA distributions and course loads among students.
  - **Example:** Examine how student GPA changes as more courses are taken.

3. SAMPLE ANALYSES

This section presents a few sample analyses to illustrate GAT's capabilities. The analyses in this section are based on real course-grade data from a large metropolitan university. Individual students or faculty cannot be identified based on the information presented in this section. The analyses presented in this section are not comprehensive and do not demonstrate the full range of analysis capabilities provided by the tool, but are sufficient to show its basic capabilities.

3.1 Department Level Analysis

In this example, GAT's "Department Analysis" command is used to identify departments with significantly higher or lower average GPAs than the institutional average. It does this by plotting the z-score normalized weighted GPA on the y-axis. This means that a positive (negative) y-value of 1.0 standard deviation is ahead (behind) from the mean departmental GPA. The Z-score formula is shown below.

$$z = \frac{x - \mu}{\sigma}$$

Where: x: raw score, μ: mean value, σ: standard deviation, z: normalized score,

The results are displayed in Figure 3. The tool also has the capability to display the departments with their raw GPA scores, which in many cases may be preferable as they are easier to interpret. However, Figure 3 is appropriate when trying to quantify the departmental grading differences.

Figure 3 shows many interesting things. First, it shows that there are large variations in departmental GPA as many departments have GPA averages that differ from the mean by more than one standard deviation. The GAT figures are interactive and in this image the user is hovering the mouse over the mathematics department entry so the tool displays additional information about the department in the yellow "tip box." The figure shows that STEM departments generally exhibit a noticeably lower GPA value than the average department, a result that is consistent with the widely

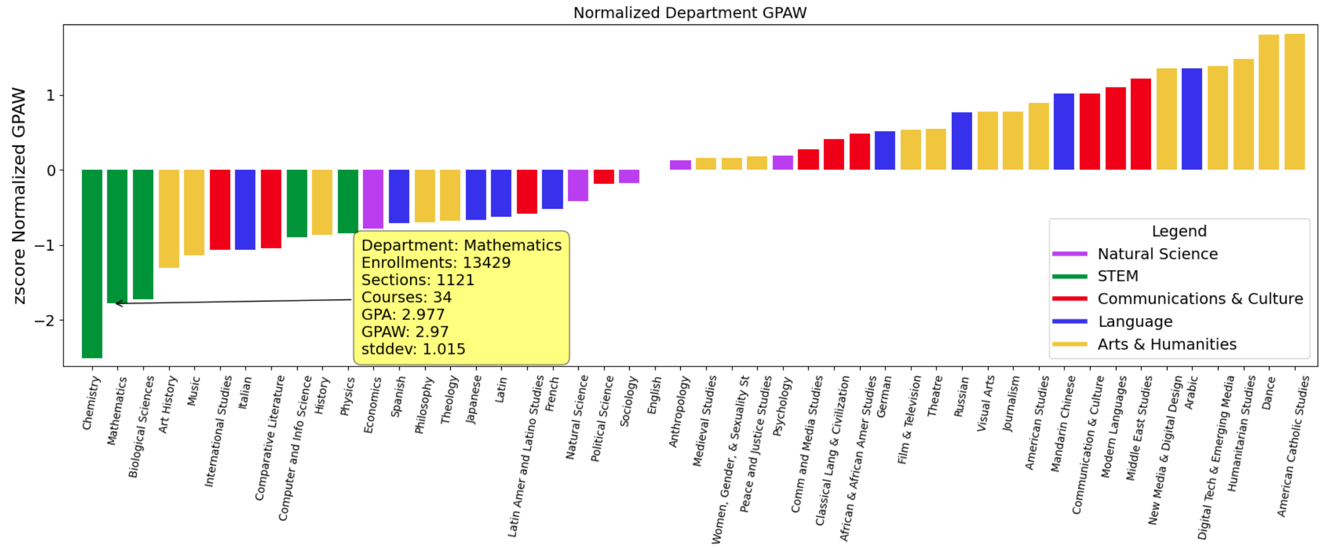

Figure 3: Distribution of Normalized Weighted GPA by Department (Mathematics Department Highlighted)

held view that STEM coursework presents a higher level of academic difficulty [5]. Figure 3 shows greater variance in normalized weighted GPA among departments in the Language, Communications and Culture, and Natural Science categories, with some above and some below the institutional average. Departments in the Arts and Humanities category contribute most of the highest normalized GPAs.

The category GPA differences are examined in more detail in Figure 4, which utilizes the “Create Groupings” plot customization option from Table 2. The departments within each category are weighted based on the total number of enrollments, so small departments have less impact. Figure 4 highlights the clear separation in assigned grades for the different academic categories. This suggests that the disciplines in different categories may have different levels of difficulty, different student expectations, and/or different philosophies in assigning grades. STEM fields systematically produce lower average GPAs than other disciplines, which may not be due only to course difficulty, but also a grading culture that relies on objective assessments rather than more interpretive methods like essays [6].

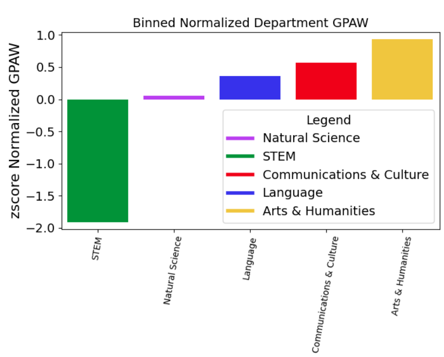

Figure 4: Normalized Weighted GPA by Academic Catetgory

While lower average GPAs in STEM fields may be expected, it is useful to analyze the grading differences from multiple perspectives to identify all the factors contributing to grading differences. The mathematics department, despite having a high enrollment, has one of the lowest GPAs among them. In the next section we examine instructor grading patterns, with a special focus on instructors in the Mathematics department, and one in particular that assigns extremely high grades.

3.2 Instructor Level Analysis

In this section we examine the grading behavior of instructors, which is summarized in Figure 5. In this figure each point corresponds to one instructor and the x-axis, GPA, represents the average grade they assigned over all course sections that they taught. Thresholds were set so that the only instructors that appear in this figure taught at minimum 20 sections. The figure shows that instructors assign vastly different grades. The primary focus should be on the x-axis value, as the grade assigned is, in this case, more significant than the standard deviation of the instructors grades. While the standard deviation does provide some information, it will by necessity decrease as the average grade increases. The standard deviation must be 0 when the GPA is 4.0, as that is the highest grade assigned.

The “tip box” displayed in Figure 5 highlights a Mathematics instructor, F71330, with a very liberal grading policy; the average grade this instructor assigned is slightly above a 3.7 (A-). This grading pattern is quite anomalous for the Mathematics department, which we previously noted assigns very low grades. The 3.702 average is also based on a substantial body of work, as the instructor taught 538 students over 3 courses and 23 course sections. While this extreme grading behavior seems hard to justify, different courses may have different averages, so we next take a more detailed look at this instructor by focusing on a single course, using the section analysis feature discussed in Section 2.4.

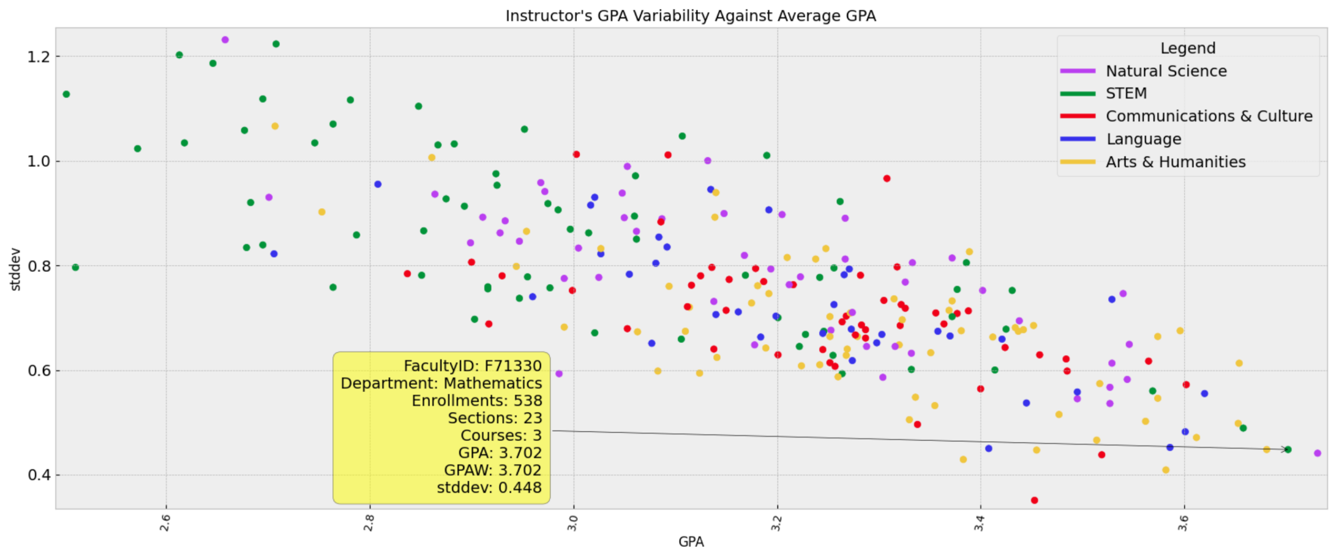

Figure 5: Distribution of Instructor Grades (Math Instructor F71330 is Highlighted)

Figure 6 shows the grades for each section of the targeted math class. This particular course was chosen as it was the course most frequently taught by instructor F71330. Only instructors who taught the course at least 3 times were included and only sections with at least 12 students were considered. The sections taught by instructor F71330 are denoted by the black points. As these points are clustered towards the right of the figure, it is clear that F71330 consistently assigns amongst the highest grades in the course. Most administrators would conclude that F71330 is too lenient a grader, although it is possible that they are just a much more effective teacher (this latter conclusion would be much more defensible with standardized exams, but these are rare in higher education). In this particular case a department chair might want to review the instructors syllabus and exams to ensure that departmental standards are being maintained.

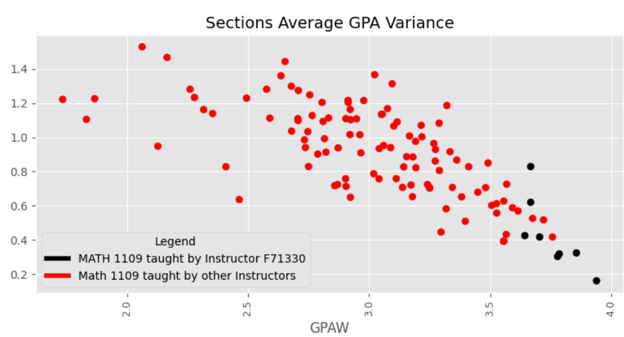

Figure 6: Distribution of Math 1109 Section Grades

3.3 Course Grade Distribution Analysis

If the “Generate Grade Distribution” option in the Pre-analysis Customization Option Box displayed in Figure 1 is selected, then a plot like Figure 7 is generated. This figure uses parallel coordinates to show each letter grade and their percent frequency for any chosen analysis using L1-normalization.

$$L = \frac{x_i}{\sum_{j=1}^{n} |x_j|}$$

Where: x_i: original value for component i, $\sum_{j=1}^{n} |x_j|$: the sum of the absolute values of all components, L: L1-normalized value

Parallel plots are ideal as they allow simultaneous visualization of multiple categorical variables (letter grades), while preserving relationships across courses. Each connected line section represents a single course (aggregated over all course sections). This provides a more granular and low-level view of grading pattern along with a visual explanation of the composition of GPA. Figure 7 shows that the grading patterns do differ between areas, with many STEM courses having more grades than other disciplines in the B- to F range. This same type of visualization can be applied to the majority of analysis levels (Sec. 2.4).

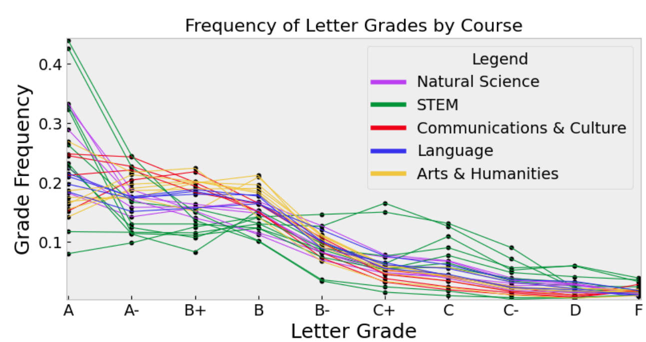

Figure 7: Course-Level Grade Distribution ($\geq$ 75 Sections)

3.4 Heatmaps

Heatmaps in GAT offer a concise way to clearly visualize how two dimensions jointly influence a particular outcome. Figure 8 shows a robust-normalized GPA across student levels (rows) and course levels (columns) (Sec. 2.4). Robust normalization was picked for this as it is resilient to outliers, which is helpful in this case where, for example, not many freshman have the ability to take a 4000 level course. This normalization, described below, ensures the data remains relevant and representative.

$$R = \frac{x - \tilde{x}}{\text{IQR}}$$

Where: x: raw score, $\tilde{x}$: median value, IQR: interquartile range (75th percentile minus 25th percentile), R: robust normalized score

Each cell's color represents how much the average GPA in that group deviates from the overall mean, with reds indicating higher-than-average performance and blues indicating lower-than-average performance. Users can configure different normalization methods (e.g., z-score, min–max, robust) to emphasize specific aspects of the data, while filtering options allows them to focus on particular semesters, or only take courses with a minimum class size into account. The heatmap ability is only avaliable for the student-course level analysis. (Sec. 2.4)

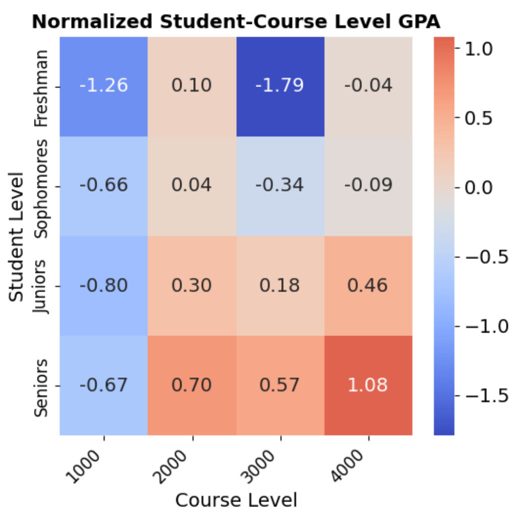

Figure 8: Heatmap between levels

4. LIMITATIONS

One limitation of the current implementation is the absence of demographic and background information about students and instructors. Table 1 does not include details such as gender, race, or age for students, nor rank or years of experience for instructors. These omissions result from university privacy policies and FERPA guidelines, which require the anonymization of student and faculty identifiers. Although this restricts some potential analyses, it does not affect the tool's core functionality. Future versions of GAT will incorporate these additional variables to enable a more comprehensive analysis of academic performance trends.

5. CONCLUSION

In this paper, we introduced the Grading Analysis Tool, an innovative, open-source solution designed to bridge the gap between raw educational data and actionable insights for academic decision-making. By integrating advanced data aggregation and interactive visualization techniques, GAT enables administrators to examine student performance across multiple dimensions—ranging from departmental overviews to instructor-specific analyses—with great ease and precision. Moreover, GAT's extensive customization options allow users to tailor analyses to their specific needs, resulting in polished, presentation-ready reports that effectively showcase key insights.

Future work should focus on incorporating additional data dimensions, such as demographic information, to refine our understanding of grading practices and their broader impact on student success. Replicating our analyses across multiple institutions with much more data would also be crucial for validating these findings and supporting more informed, equitable educational policies. Ultimately, GAT lays a foundation for ongoing research and continuous improvement in higher education, giving educators the tools to make decisions that foster student achievement and institutional excellence.

6. REFERENCES

[1] Anonymous. Grading analysis tool. https://www.anonymous.4open.science/r/Grading-Analysis-Tool, 2025. Accessed: 2025-02-17.

[2] Brad S. Chissom. Interpretation of the kurtosis statistic. *The American Statistician*, 24(4):19–22, 1970.

[3] Ellucian. Ellucian at a glance. https://www.ellucian.com/assets/en/ellucian-glance.pdf, 2023. Accessed: 2025-02-17.

[4] The pandas development team. pandas-dev/pandas: Pandas. https://www.doi.org/10.5281/zenodo.3509134, February 2020.

[5] J.H. Tomkin, West, and M. Stem courses are harder: evaluating inter-course grading disparities with a calibrated gpa model. *International Journal of STEM Education*, 9:17, 2022.

[6] Gary M. Weiss, Luisa A. L. Rosa, Hyun Jeong, and Daniel D. Leeds. An analysis of grading patterns in undergraduate university courses. In *Proceedings of the 2023 IEEE 47th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 310–315, 2023.