STAT 930 Final Report

Carson Trego

Due May 9th, 2025

Abstract

The mental health of students attempting a graduate degree is an ongoing area of interest [1] [3]. One study found that for a variety of different disciplines, students were more likely to seek psychiatric medication as their time in school increased [3]. The aforementioned study looked at broad areas of interest like technology and medicine, but research that looks at dental students' mental health specifically is a bit more sparse [3]. Making a distinction between dental students and other areas of study is important, as it has been shown that the area of interest one pursues can impact the effect graduate studies have on their mental health [3]. Our Domain Experts are interested in obtaining more information on the effect of pursuing a dental degree on mental health, and have gathered two years of survey data from dental students at all four grade levels. These surveys, which took place in the 2023-2024 and 2024-2025 academic years, included three mental health metrics: the PHQ-9, Burnout Mini Z 1.0 Q3, and the GAD-7. Results showed that there are significant differences within the 2024-2025 academic year for both the PHQ-9 and GAD-7, and significant differences across years were found for all three metrics.

1 Domain Expert Background

The domain experts (DEs) involved in this project were Anna Campbell and Shayla Meyer, both of whom were dental students at the UNMC College of Dentistry. Both DEs were assigned a research project overseen by Dr. Steven Wengal and Dr. Sarah Fischer. As part of this project, Anna Campbell and Shayla Meyer were expected to collect data, submit a written report, and present their findings as part of a poster presentation event.

For their project, the DEs took inspiration from a project that had been completed the previous year, which assessed the mental health of UNMC dental students using three metrics in order to compare the relative mental state of dental students in each grade.

Both DEs had expressed an interest in learning the statistics and computer programming that was required for the project. Some of the later meetings were centered around learning more about the methods used and how to report the results.

While meeting with the DEs, they had never expressed interest in a specific outcome or statistically significant effect. The DEs were not attached to any outcome and were purely interested in the method being statistically valid. Overall, the DEs showed a great scientific attitude, and working with them was an overall very positive experience.

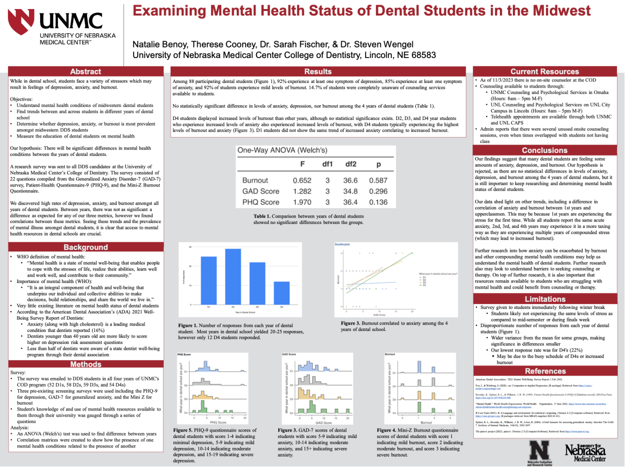
2 Timeline

- January 28th: Initial Meeting
  - In this meeting, we discussed the overall problem and goals for the project. The DE had stated that they needed a basic outline of the results within a week. This project had staged submissions, meaning that students would submit their abstract (with results), analysis, and then final paper, with the abstract being due that week.
- January 31st: Abstract Results Overview
  - After putting together a quick analysis, we met to discuss the results and what they meant. All that was needed for this part was a simple table with the estimates and p-values.
- February 13th: Analysis
  - A full analysis paper was prepared for the DE, along with visualizations. During this meeting, we went over the analysis method used, how it works, and why it was chosen.
- February 20th: Poster and Presentation Preparation

- During this meeting, we discussed the details that were added to the poster for the poster presentation part of their project. In addition to planning visuals and the structure of the poster, we had gone over a few reporting guidelines so they could confidently and accurately report the results. One example of these guidelines was to avoid conflating insignificance with equivalence, and using wording like "significance was not found" and "the relationship is unclear" rather than "these two groups are the same" (an error which was seen in other posters). Although the DEs did not believe statistical questions would be the focus, we prepared responses to questions about the analysis, such as "we chose this method as we were unsure if some classical assumptions were not met, and this method is more robust to violations of those assumptions".

- March 27th: Discuss Final Analysis Plan
  - Both DEs were very happy with how the presentation turned out, and we began discussing how the final paper would be prepared. During this meeting, I was linked to the shared document so I could add some analysis details and correct any potential statistical interpretation errors. This was the final meeting, and all further communication was done over email.

3 Introduction



UNMC UNIVERSITY OF NEBRASKA MEDICAL CENTER™

Examining Mental Health Status of Dental Students in the Midwest

Natalie Benoy, Therese Cooney, Dr. Sarah Fischer, & Dr. Steven Wengel
University of Nebraska Medical Center College of Dentistry, Lincoln, NE 68583

Abstract

While in dental school, students face a variety of stressors which may result in feelings of depression, anxiety, and burnout.

Objectives:
- Understand mental health conditions of midwestern dental students
- Find trends between and across students in different years of dental school
- Determine whether depression, anxiety, or burnout is most prevalent amongst midwestern DDS students
- Measure the education of dental students on mental health

Our hypothesis: There will be significant differences in mental health conditions between the years of dental students.

A research survey was sent to all DDS candidates at the University of Nebraska Medical Center's College of Dentistry. The survey consisted of 22 questions compiled from the Generalized Anxiety Disorder-7 (GAD-7) survey, Patient-Health Questionnaire-9 (PHQ-9), and the Mini-Z Burnout Questionnaire.

We discovered high rates of depression, anxiety, and burnout amongst all years of dental students. Between years, there was not as significant a difference as expected for any of our three metrics, however we found correlations between these metrics. Seeing these trends and the prevalence of mental illness amongst dental students, it is clear that access to mental health resources in dental schools are crucial.

Background

- WHO definition of mental health:
  - "Mental health is a state of mental well-being that enables people to cope with the stresses of life, realize their abilities, learn well and work well, and contribute to their community."
- Importance of mental health (WHO):
  - "It is an integral component of health and well-being that underpins our individual and collective abilities to make decisions, build relationships, and share the world we live in."
- Very little existing literature on mental health status of dental students
- According to the American Dental Association's (ADA) 2021 Well-Being Survey Report of Dentists:
  - Anxiety (along with high cholesterol) is a leading medical condition that dentists reported (16%)
  - Dentists younger than 40 years old are more likely to score higher on depression risk assessment questions
  - Less than half of dentists were aware of a state dentist well-being program through their dental association

Methods

Survey:
- The survey was emailed to DDS students in all four years of UNMC's COD program (52 D1s, 58 D2s, 59 D3s, and 54 D4s)
- Three pre-existing screening surveys were used including the PHQ-9 for depression, GAD-7 for generalized anxiety, and the Mini Z for burnout
- Student's knowledge of and use of mental health resources available to them through their university was gauged through a series of questions

Analysis:
- An ANOVA (Welch's) test was used to find difference between years
- Correlation matrices were created to show how the presence of one mental health conditions related to the presence of another

Results

Among 88 participating dental students (Figure 1), 92% experience at least one symptom of depression, 85% experience at least one symptom of anxiety, and 92% of students experience mild levels of burnout. 14.7% of students were completely unaware of counseling services available to students.

No statistically significant difference in levels of anxiety, depression, nor burnout among the 4 years of dental students (Table 1).

D4 students displayed increased levels of burnout than other years, although no statistical significance exists. D2, D3, and D4 year students who experience increased levels of anxiety also experienced increased levels of burnout, with D4 students typically experiencing the highest levels of burnout and anxiety (Figure 3). D1 students did not show the same trend of increased anxiety correlating to increased burnout.

One-Way ANOVA (Welch's)

	F	df1	df2	p
Burnout	0.652	3	36.6	0.587
GAD Score	1.282	3	34.8	0.296
PHQ Score	1.970	3	36.4	0.136

Table 1. Comparison between years of dental students showed no significant differences between the groups.

Figure 1. Number of responses from each year of dental student. Most years in dental school yielded 20-25 responses, however only 12 D4 students responded.

Figure 3. Burnout correlated to anxiety among the 4 years of dental school.

Figure 5. PHQ-9 questionnaire scores of dental students with score 1-4 indicating minimal depression, 5-9 indicating mild depression, 10-14 indicating moderate depression, and 15-19 indicating severe depression.

Figure 3. GAD-7 scores of dental students with score 5-9 indicating mild anxiety, 10-14 indicating moderate anxiety, and 15+ indicating severe anxiety.

Figure 4. Mini-Z Burnout questionnaire scores of dental students with score 1 indicating mild burnout, score 2 indicating moderate burnout, and score 3 indicating severe burnout.

Current Resources

- As of 11/3/2023 there is no on-site counselor at the COD
- Counseling available to students through:
  - UNMC Counseling and Psychological Services in Omaha (Hours: 8am – 5pm M-F)
  - UNL Counseling and Psychological Services on UNL City Campus in Lincoln (Hours: 8am – 5pm M-F)
  - Telehealth appointments are available through both UNMC and UNL CAPS
- Admin reports that there were several unused onsite counseling sessions, even when times overlapped with students not having class

Conclusions

Our findings suggest that many dental students are feeling some amounts of anxiety, depression, and burnout. Our hypothesis is rejected, as there are no statistical differences in levels of anxiety, depression, and burnout among the 4 years of dental students, but it is still important to keep researching and determining mental health status of dental students.

Our data shed light on other trends, including a difference in correlation of anxiety and burnout between 1st years and upperclassmen. This may be because 1st years are experiencing the stress for the first time. While all students report the same acute anxiety, 2nd, 3rd, and 4th years may experience it in a more taxing way as they are experiencing multiple years of compounded stress (which may lead to increased burnout).

Further research into how anxiety can be exacerbated by burnout and other compounding mental health conditions may help us understand the mental health of dental students. Further research also may look to understand barriers to seeking counseling or therapy. On top of further research, it is also important that resources remain available to students who are struggling with mental health and could benefit from counseling or therapy.

Limitations

- Survey given to students immediately following winter break
  - Students likely not experiencing the same levels of stress as compared to mid-semester or during finals week
- Disproportionate number of responses from each year of dental students (Figure 1).
  - Wider variance from the mean for some groups, making significance in differences smaller
  - Our lowest response rate was for D4's (22%)
    - May be due to the busy schedule of D4s or increased burnout

References

American Dental Association. "2021 Dentist Well-Being Survey Report. 1 Feb. 2022.

Fox, J., & Weisberg, S. (2020). car: Companion to Applied Regression. [R package]. Retrieved from https://cran.r-project.org/package=car.

Kroenke, K., Spitzer, R. L., & Williams, J. B. W. (1999). *Patient Health Questionnaire-9 (PHQ-9)* [Database record]. APA PsycTests. https://doi.org/10.1037/t06165-000

"Mental Health." *World Health Organization*, World Health Organization, 17 June 2022, https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response.

R Core Team (2021). R: A Language and environment for statistical computing. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

Spitzer, R. L., Kroenke, K., Williams, J. B. W., Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. Archives of Internal Medicine, 166(10), 1092-1097.

The jamovi project (2022). jamovi. (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org.

Nebraska Evaluation and Research Center
UNIVERSITY OF Nebraska Medical Center

Figure 1: Poster presentation for the project that was conducted last year.

The original project was designed to investigate if a dental school students grade (grade as in year in dental school, not grade as in GPA) had an association with their average mental health, as measured by the following three metrics:

- PHQ - 9 (PHQ)
  - Patient Health Questionnaire, 9 questions. This questionnaire, which was designed by the American Psychological Association, includes 9 questions designed for surveyors to quickly and easily assess someone's risk for depression [10]. Each question is on a zero to three integer scale, where zero indicates "not at all" and three indicates "nearly every day". Although this was designed originally to have 9 questions, I had noticed that the 9th question was missing in both surveys. The question was "Thoughts that you would be better off dead or of hurting yourself in some way" so I assumed the removal was out of ethical concerns. I had let the DEs know about this alteration.
- Burnout Mini Z 1.0 Q3 (BURN)
  - This is a single question taken from the Burnout Mini Z Questionnaire, which was created by the Institute for Professional Worklife as a means to quickly assess someone's risk for burnout [13]. This is a single question with a zero to four-point integer scale in which zero indicates complete denial of burnout and four indicates that the person reports burnout to the extent of needing help.
- GAD - 7 (GAD)
  - Generalized Anxiety Disorder, 7 questions. The GAD-7 is a questionnaire with 7 questions designed to assess someone's risk of generalized anxiety disorder, each answered using the same zero to three integer scale as the PHQ-9 [15]. All questions in the GAD-7 were retained in the survey.

The survey was conducted by placing fliers around the campus, which students could use to find an online survey. The previous project did not find any significant difference between the grade levels for any of the metrics stated. Our DEs have performed the same methods to collect data for this year, with the goal being to see if there is a significant difference between the grade levels this year, as well as see if there has been any change within grades from last year to this year.

The survey did not keep any identifying information on the participants, so it is unclear how many students had taken the survey both times. To ensure that our test across years involves completely different groups, we have decided to only test the across-year difference for grades D1 and D2.

4 Methods

4.1 Data Format and EDA

	D1	D2	D3	D4	Total
2023-2024	25	29	22	12	88
2024-2025	35	23	32	26	116

Table 1: Data for 2023-2024 and 2024-2025. Overall, 204 responses were recorded in the first and second samples, combined.

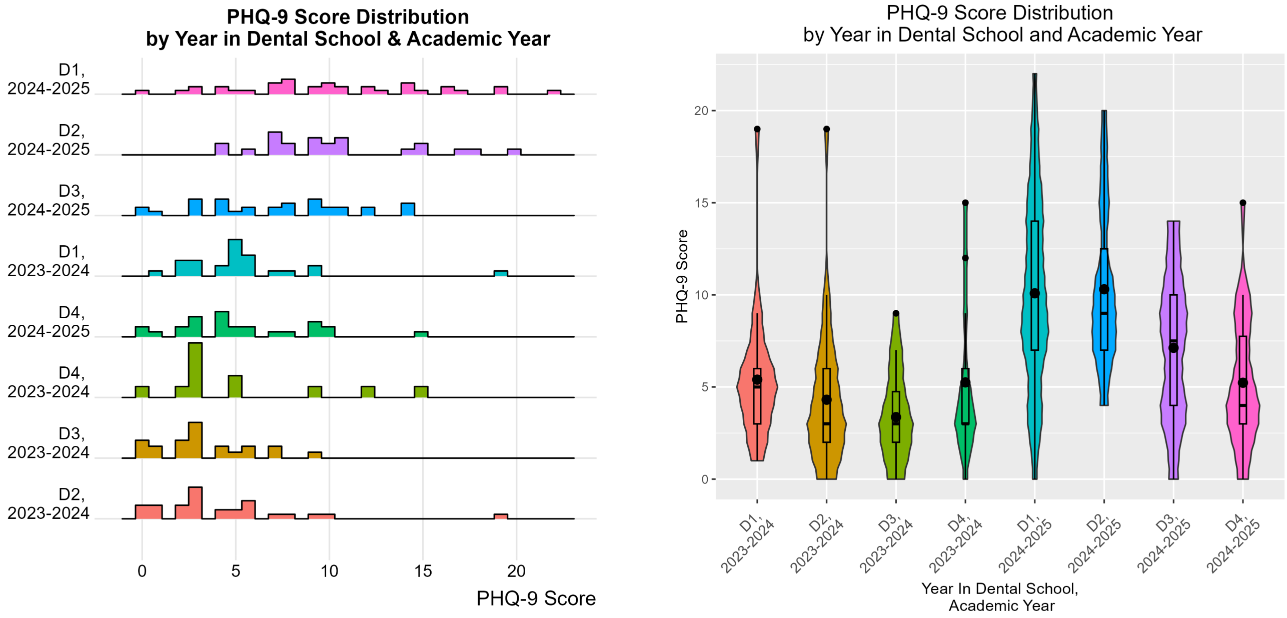


Figure 2: Comparison of PHQ score distributions using ridge (left) and discrete violin (right) visualizations.

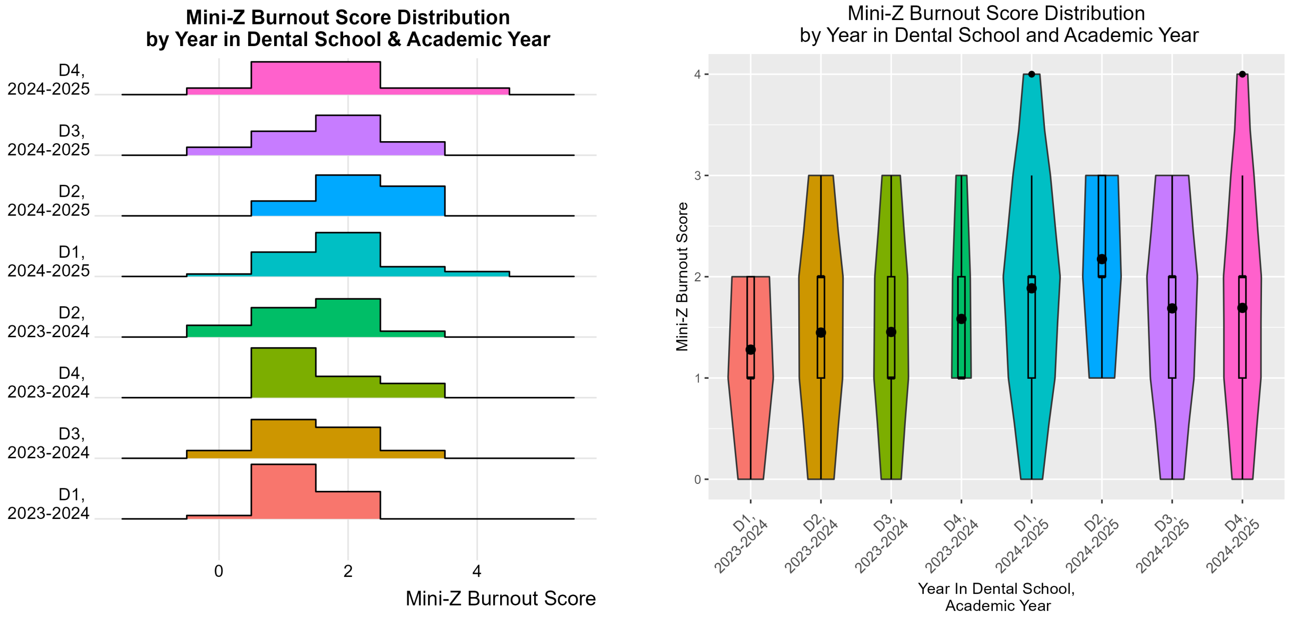


Figure 3: Comparison of BURN score distributions using ridge (left) and discrete violin (right) visualizations.

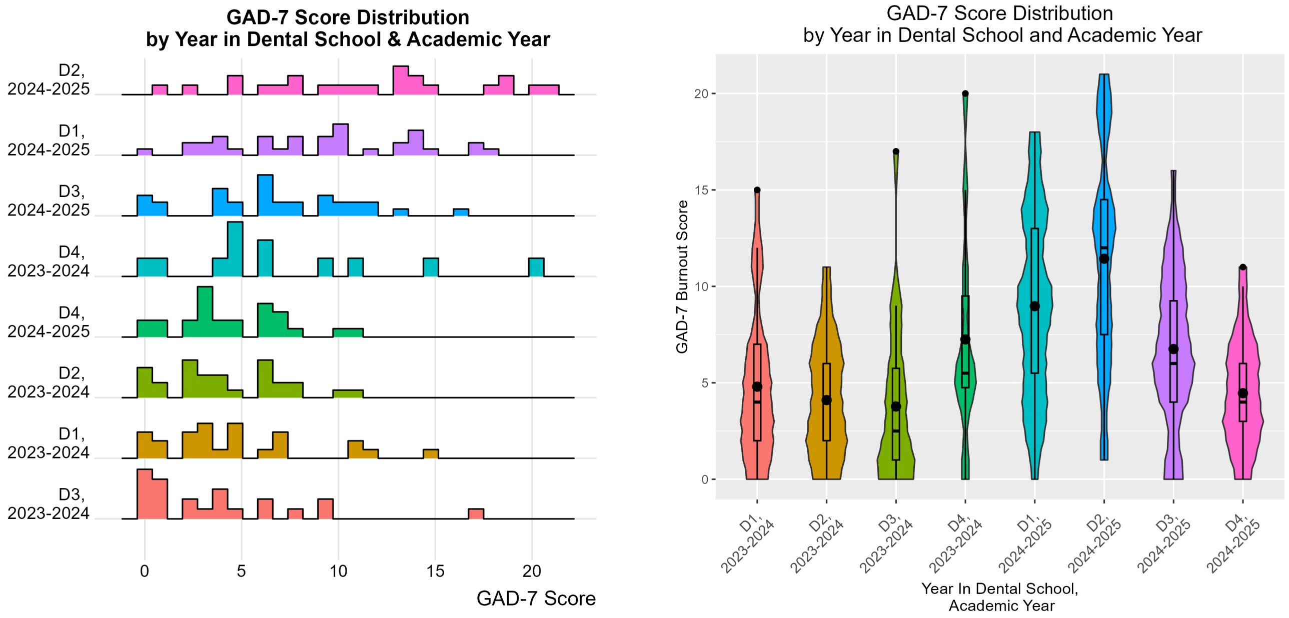


Figure 4: Comparison of GAD score distributions using ridge (left) and discrete violin (right) visualizations.

Table 2: Descriptive statistics for PHQ score by dental year and academic year. High skew values are labeled in yellow, and very high skew levels are labeled in red. Rows with a dark grey shade were removed, and not included at any part of this analysis.

Dental Year	PHQ Mean	PHQ Median	PHQ Skew	PHQ Kurt	PHQ Var
D1.2024	5.4	5	2.31	9.99	12.42
D1.2025	10.09	10	0.24	2.49	27.61
D2.2024	4.31	3	1.97	8.19	15.01
D2.2025	10.30	9	0.66	2.61	18.86
D3.2024	3.36	3	0.55	2.61	6.59
D3.2025	7.13	7.5	0.02	2.08	16.31
D4.2024	5.25	3	1.13	3.10	19.84
D4.2025	5.23	4	0.77	3.36	12.66

Table 3: Descriptive statistics for BURN score by dental year and academic year. Rows with a dark grey shade were removed, and not included at any part of this analysis.

Dental Year	BURN Mean	BURN Median	BURN Skew	BURN Kurt	BURN Var
D1.2024	1.28	1	0.14	2.49	0.29
D1.2025	1.89	2	0.50	3.48	0.75
D2.2024	1.45	2	-0.22	2.43	0.68
D2.2025	2.17	2	-0.25	2.04	0.51
D3.2024	1.45	1	0.15	2.61	0.64
D3.2025	1.69	2	-0.28	2.53	0.74
D4.2024	1.58	1	0.86	2.25	0.63
D4.2025	1.69	2	0.64	3.24	1.02

Table 4: Descriptive statistics for GAD score by dental year and academic year. High skew values are labeled in yellow. Rows with a dark grey shade were removed, and not included at any part of this analysis.

Dental Year	GAD Mean	GAD Median	GAD Skew	GAD Kurt	GAD Var
D1.2024	4.8	4	0.93	3.18	16
D1.2025	8.97	9	0.09	2.11	22.56
D2.2024	4.10	4	0.45	2.37	9.91
D2.2025	11.43	12	-0.03	2.10	32.62
D3.2024	3.77	2.5	1.55	5.48	17.52
D3.2025	6.75	6	0.09	2.58	16.06
D4.2024	7.25	5	0.96	3.21	32.75
D4.2025	4.46	4	0.43	2.64	8.18

4.2 Standard Assumption Checking

To see if the assumptions of classical ANOVA are met, we first check the assumptions for the part of the analysis that looks at the scores within the 2024-2025 academic year.

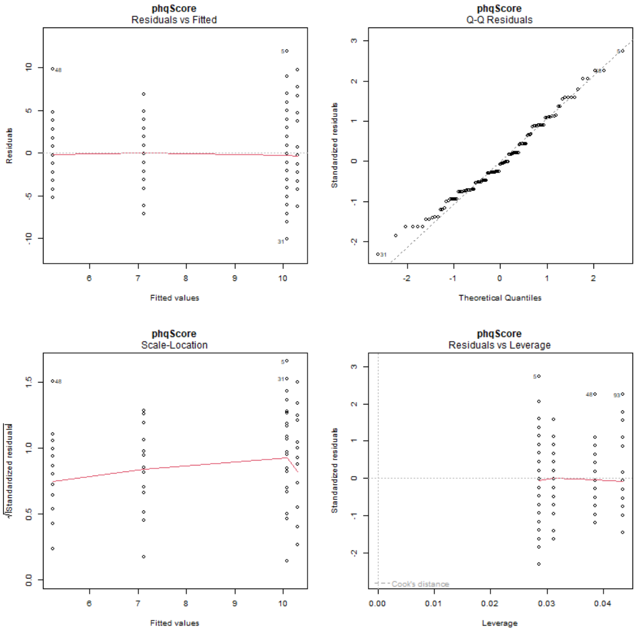


Figure 5: The top left panel is intended to check for linearity and homoscedasticity, the top right panel is intended to check for normality, the bottom left panel is intended to check for homoscedasticity, and the bottom right panel is intended to check for strongly influential points in the dataset.

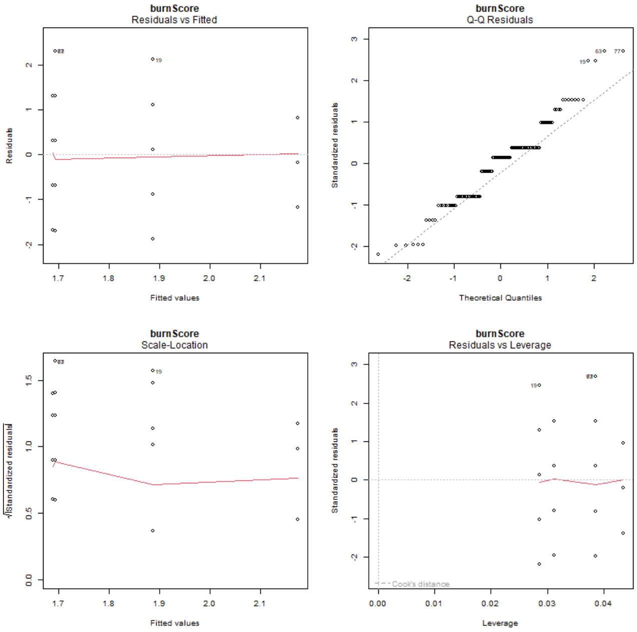


Figure 6: The top left panel is intended to check for linearity and homoscedasticity, the top right panel is intended to check for normality, the bottom left panel is intended to check for homoscedasticity, and the bottom right panel is intended to check for strongly influential points in the dataset.

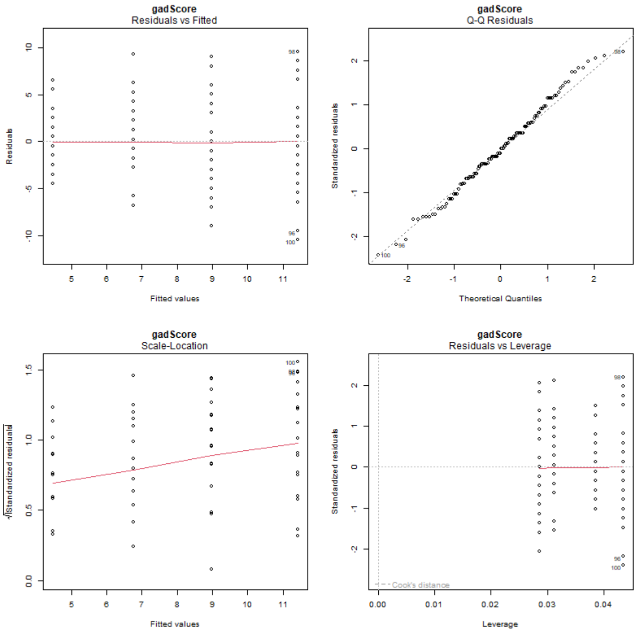


Figure 7: The top left panel is intended to check for linearity and homoscedasticity, the top right panel is intended to check for normality, the bottom left panel is intended to check for homoscedasticity, and the bottom right panel is intended to check for strongly influential points in the dataset.

Comparison	Shapiro-Wilk (SW)	Levene's Test (L)	Breusch-Pagan (BP)
PHQ Score	0.195	0.200	0.110
BURN Score	0.018*	0.547	0.471
GAD Score	0.631	0.012*	0.005*

Table 5: Significant values ($p < 0.05$) are bolded and marked with an asterisk. The Shapiro-Wilk Test tests for non-normality, Levene's Test tests for homogeneity of group variances, and the Breusch-Pagan Test tests for heteroskedasticity of residuals. Note that a significant result provides evidence that an assumption has been violated, but non-significance does not provide evidence for the assumption. Also note that significant assumption violations do not necessarily imply that the violation will strongly affect the model or conclusions derived from it.

Next, we will check if the assumptions are met for the contrasts across the 2023-2024 and 2024-2025 academic years.

Comparison	Shapiro-Wilk (SW)	Levene's Test (L)	Breusch-Pagan (BP)
PHQ (D1)	0.090	**0.009***	0.102
PHQ (D3)	0.505	**0.006***	**0.012***
BURN (D1)	**0.003***	0.178	0.066
BURN (D3)	**0.004***	0.951	0.658
GAD (D1)	0.285	0.248	0.280
GAD (D3)	**0.048***	0.918	0.878

Table 6: Normality and variance assumption checks for D1 and D3, comparing 2024 vs 2025 within each trait. Significant values ($p < 0.05$) are bolded and marked with an asterisk. The Shapiro-Wilk Test tests for non-normality, Levene's Test tests for non-homogeneity of group variances, and the Breusch-Pagan Test tests for heteroskedasticity of residuals. Note that a significant result provides evidence that an assumption has been violated, but non-significance does not provide evidence for the assumption. Also note that significant assumption violations do not necessarily imply that the violation will strongly affect the model or conclusions derived from it.

4.3 Subjective Final Conclusion

The extent to which a distortion in the expected diagnostic plot is considered problematic for a model assumption is, to some extent, subjective [4]. While the tests for normality and homoscedasticity violations, such as the Shapiro-Wilk, Levene's, and Breusch-Pagan tests, provide hypothesis tests with numeric results that have direct interpretations, some have deprecated the use of these hypothesis tests for assumption checking [14]. The first issue is that, fundamentally, hypothesis tests for violations of normality and homoscedasticity only provide information when the assumption is violated, and do not provide any evidence that the assumption is met or is reasonable for the model [14]. Furthermore, these tests only inspect for a violation of an assumption, not an impactful violation of said assumption [14]. This means that small and inconsequential violations can result in rejection [14]. As the sample size increases and the distribution of the data is better "understood", small nuances in the data can be observed, which result in non-normal and non-homoscedastic distribution [14]. This is at odds with the fact that as sample sizes increase, the sampling distribution for the means more closely **resembles** a normal distribution, meaning the normality assumption is a better fit with the model [6].

Ultimately, this is all to say that the diagnostic plots are evidently subjective, and while the tests for violations of normality and homoscedasticity provide numeric results with direct interpretations, they are far from an objective and clean answer to the question "are our assumptions reasonable for our analysis". With the plots appearing to violate our assumptions of normality and homoschedasticity, the sample only having a modest, but not particularly large, sample size for each group, and multiple of the tests for these assumptions yielding significant results, **I have decided to NOT assume normality, nor equality of variance**.

The assumptions of linearity and additivity are cleanly met, as this problem is dealing with a single categorical explanatory variable.

4.4 Assumption Implications and Client Goals

The DEs are interested in comparing the averages of each group specifically. This means that some non-parametric tests, such as Kruskal-Wallis, Dunn's, or Mann-Whitney U, would not fit with our client's goals [9]. Tests like Kruskal-Wallis, Dunn's, or Mann-Whitney U are often intended to be used to compare central tendency, but do not compare the group means and are not valid when large differences in variance across groups are found [9]

Given that the assumptions of normality and homoscedasticity are not assumed to be met, methods such as a classical ANOVA or pairwise equal variance t-test should not be used. To address the possible non-normality issue, permutation testing was used. Permutation testing does not make an assumption of a null distribution, so it can be appropriately used when non-normality is a concern [5][7][12][8]. Permutation testing, in general, assumes *exchangeability of the null*, which assumes equal variance as a consequnce [5][7][12][8]. This issue can be addressed by using studentized statistics, which are asymptotically exact even under different variances and sample sizes [5][7][12][8]. Caution is recommended when the distribution is highly skewed, and a heurstic recomendation of 30 or more smaples per group is recommended when unequal variance, skew, and unequal samples sizes are simultaneously present [8]. The only groups that showed concerning skew were D1 in the 2023-2024 academic year and D3 in the 2023-2024 academic year. As shown by Janseen, the type 1 error rate of studentized permutation tests for small sample sizes (8), highly skewed distributions, and large variance ratios can increase by approximately 1 percent, while less skewed distributions or sets of distribtuions with lower variance ratios tend to have a negligable change in type 1 error rate [7]. The variance ratio of D1 2023-2024 to D1 2023-2024 is approximately 1:2, while the variance ratio of D3 2023-2024 to D3 2024-2025 is approximately 1:1. The D1 groups have a sample size of 25 and 35 respectively, while the D2 groups have a sample size of 22 and 32 respectively. Similar to the classic tests of validity, the validity conditions of this method are somewhat subjective and heuristic. Given these circumstances, I have decided to assume that the risk of an increased type 1 error for this method is low, and this analysis will use studentized permutation testing.

4.5 Loading and Adding Metrics

Although the clients are using a pre-existing metric, we had decided to still check the loadings for each variable that was created by adding several variables together. This was done for the PHQ-9 and GAD-7, but not for the Burnout Mini Z score, which was only one question. None of the loadings or Cronbach's alpha values were below our minimum threshold, but some were close such as the PHQ-9's question 8, which had a loading of 0.419.

Table 7: Cronbach's Alpha for Each Scale (Threshold: $\alpha > 0.70$)

Scale	Cronbach's Alpha
PHQ-9	0.85
BURN	Not Applicable
GAD-7	0.90

Table 8: PHQ-9 Factor Loadings on First Component (Threshold: Loading > 0.40)

Item	Loading
1	0.739
2	0.768
3	0.515
4	0.723
5	0.624
6	0.764
7	0.664
8	0.419

Table 9: GAD-7 Factor Loadings on First Component (Threshold: Loading > 0.40)

Item	Loading
1	0.826
2	0.849
3	0.909
4	0.823
5	0.572
6	0.605
7	0.678

5 Results

Trait	Comparison	Global p	Pairwise p	Adjusted p
PHQ Score	(Global)	0.000058	–	0.000174*
	D1.2025 vs D2.2025	–	0.852650	1.000000
	D1.2025 vs D3.2025	–	0.011751	0.070506
	D1.2025 vs D4.2025	–	0.000061	0.000366*
	D2.2025 vs D3.2025	–	0.008235	0.049410*
	D2.2025 vs D4.2025	–	0.000056	0.000336*
	D3.2025 vs D4.2025	–	0.063501	0.381006
BURN Score	(Global)	0.112988	–	0.338964
GAD Score	(Global)	0.000001	–	0.000003*
	D1.2025 vs D2.2025	–	0.092642	0.555852
	D1.2025 vs D3.2025	–	0.042333	0.253998
	D1.2025 vs D4.2025	–	0.000029	0.000174*
	D2.2025 vs D3.2025	–	0.001518	0.009108*
	D2.2025 vs D4.2025	–	0.000001	0.000006*
	D3.2025 vs D4.2025	–	0.014441	0.086646

Table 10: Global Welch ANOVA and all within-year pairwise studentized permutation test results for 2025. Only traits with significant global effects are shown. Global p-values are Bonferroni-adjusted over 3 traits; pairwise p-values are adjusted over 6 comparisons per trait. Adjustments are done for each family of questions and at each level of inference, meaning that the 3 global p-values for the within the 2024-2025 academic year are adjusted by multiplying the p-values by 3, and then, if the result is significant, all 6 comparisons are adjusted by multiplying the p-values by 6.

Comparison	Trait	Raw p	Adjusted p
D1.2024 vs D1.2025	PHQ Score	0.00013	0.00077*
D3.2024 vs D3.2025	PHQ Score	0.00015	0.00091*
D1.2024 vs D1.2025	BURN Score	0.00139	0.00833*
D3.2024 vs D3.2025	BURN Score	0.29467	1.00000
D1.2024 vs D1.2025	GAD Score	0.00064	0.00386*
D3.2024 vs D3.2025	GAD Score	0.01231	0.07384

Table 11: Between-year pairwise studentized permutation test results for D1 and D3 (2024 vs 2025). Adjusted p-values reflect Bonferroni correction for 6 comparisons. Adjustments are done for each family of questions and at each level of inference, meaning the contrasts made to investigate the difference across the 2023-2024 and 2024-2025 academic years were all adjusted as one family by multiplying each p-value by 6.

Within the 2024-2025 academic year, it was shown that there is a significant difference in the PHQ-9 average between D1 and D4, D2 and D3, and D3 and D4. No significant difference within the 2024-2025 academic year was found for the Burnout Mini Z 1.0 Q3. Within the 2024-2025 academic year, a significant difference in the average GAD-7 score was found between D1 and D4, D2 and D3, and D2 and D4. Across years, significant differences were found for both grades tested (D1 and D3) for the PHQ-9 score, only the D1 grades for the Burnout Mini Z 1.0 Q3, and only the D1 grades for the GAD-7 score.

6 Discussion

Across the 2023-2024 to 2024-2025 academic years, there was a significant increase in PHQ-9 scores for both grades and GAD-7 scores for D1 students. Higher scores indicate worse mental health symptoms, so these groups appear to have worse anxiety and depression symptoms. For all significant differences within the 2024-2025 academic year, the fewer years spent in dental school, the higher the score on both the PHQ-9 and GAD-7 inventories.

The previous study, from which we obtained the data for the first year, which examined the differences in the three mental health scores across the four grades of dental school, did not find any significant difference between the grades. From the first set of tests, it appeared that the mental health of students has changed within the last year. In the second set of tests, multiple significant differences between grade levels were observed. It is possible that an increase in anxiety and depression overall makes the differences in the grade levels more apparent.

While it is unclear what could be causing the increased psychological distress among students, multiple studies have shown that election years are associated with increased reported anxiety and decreased mental well-being [2][11].

It is unclear why some higher grade levels appear to be less shifted than the lower grade levels. It is possible that going through dental school encourages students to develop better coping strategies, but this difference could also be caused by aging in general. It is also possible that being closer to completion of dental school gives students more hope for the future. The effect we observed may be specific to dental school students and may not apply to other disciplines. One longitudinal study tracked the change in psychiatric medicine use across multiple disciplines and found an increase in psychiatric medicine use following the beginning of a PhD program in natural sciences, technology, social sciences, and humanities, but not medicine [3]. Our study showed a negative association between grade level and mental health severity, but this trend may not hold for other disciplines.

7 Limitations and Considerations

While the data collected included tests that were designed to score a student's mental well-being, there was no data on what the students believed was causing their stress and anxiety. This was not a longitudinal study, so it is unclear how individuals' mental health is changing over time.

8 Conclusion

There appears to be a decrease in mental well-being among UNMC dental students, which is not affecting all grade levels to the same degree. It is currently unclear if this change represents an effect that is here to stay or if the effect is part of a regular cycle. Further research could investigate the state of students in future years to better understand if this decrease in mental well-being is part of a cycle or a consistent trend.

9 Final Notes

Several Programs were used for this analysis, you can find the complete code and datasets in the following GitHib Repository: https://github.com/seatreeg/930Final.

References

[1] American Dental Association. 2021 dentist well-being survey report, February 2022. Available from: https://www.ada.org/resources/research/health-policy-institute/dentist-well-being.

[2] American Psychological Association. Stress in america™ 2024: A nation in political turmoil, 2024. Available from: https://www.apa.org/news/press/releases/stress/2024.

[3] Sanna Bergvall, Clara Fernström, Eva Ranehill, and Anna Sandberg. The impact of phd studies on mental health – a longitudinal population study. `https://ssrn.com/abstract=4920527`, 2024. Available at SSRN.

[4] Rok Blagus, Jakob Peterlin, and Janez Stare. Goodness-of-fit testing in linear regression models, 2019.

[5] EunYi Chung and Joseph P. Romano. Exact and asymptotically robust permutation tests. *The Annals of Statistics*, 41(2):484–507, 2013.

[6] Marie Delacre, Christophe Leys, Youri L. Mora, and Daniël Lakens. Taking parametric assumptions seriously: Arguments for the use of welch's f-test instead of the classical f-test in one-way anova. *International Review of Social Psychology*, 32(1):13, 2019.

[7] Arnold Janssen. Studentized permutation tests for non-i.i.d. hypotheses and the generalized behrens–fisher problem. *Statistics & Probability Letters*, 36(1):9–21, 1997.

[8] Julian D. Karch. Choosing between the two-sample t test and its alternatives: a practical guideline. PsyArXiv, 2021.

[9] Cynthia M Kroeger, Benjamin A Hannon, Thomas M Halliday, Sydney N Hinkle, Dianna M Thomas, Jo L Freudenheim, Annette M Ferris, Julia Klein, Marissa Burgermaster, Katherine L Tucker, et al. Evidence of misuse of nonparametric tests in the presence of heteroscedasticity within obesity research [version 1; peer review: 2 approved]. *F1000Research*, 10:391, 2021.

[10] K. Kroenke, R. L. Spitzer, and J. B. W. Williams. Patient health questionnaire-9 (phq-9) [database record]. `https://doi.apa.org/doiLanding?doi=10.1037%2Ft00788-000`, 1999. APA PsycTests.

[11] Sankar Mukhopadhyay. Elections have (health) consequences: Depression, anxiety, and the 2020 presidential election. *Economics & Human Biology*, 47:101191, 2022.

[12] Kengo Noguchi, Frank Konietschke, Fernando Marmolejo-Ramos, and Daniel Oberfeld. Permutation tests are robust and powerful at 0.5% and 5% significance levels. *Behavior Research Methods*, 53:2712–2724, 2021.

[13] David Shaholli, Chiara Bellenzier, Corrado Colaprico, Francesca Vezza, Giovanna Carluccio, Luca Moretti, Maria Vittoria Manai, Vanessa India Barletta, Alice Mannocci, and Giuseppe La Torre. Mini-z validation for burnout and stress evaluation: an observational study. *Rivista di Psichiatria*, 59(2):60–68, Mar-Apr 2024. PMID: 38651774.

[14] Itamar Shatz. Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, 56:826–845, 2024.

[15] R. L. Spitzer, K. Kroenke, J. B. W. Williams, and B. Löwe. A brief measure for assessing generalized anxiety disorder: The gad-7. *Archives of Internal Medicine*, 166(10):1092–1097, 2006.