Project

District Performance Evaluation: Math Outcomes in Grant vs. Non-Grant Schools

Author: Dheeraj Shathragna Mandhala

Tools Used: SAS Studios and Power BI

Project Objective:

The purpose of this analysis is to evaluate the impact of the Texas Education Agency (TEA)'s Strategic Compensation Grant Program on 3rd–5th grade STAAR math outcomes in SY 2022–2023. We compare performance across districts that received grant funding with those that did not.

Executive Summary:

In the 2022–2023 academic year, the Texas Education Agency (TEA) supported 200 school districts through a strategic compensation grant program to improve 3rd–5th grade math outcomes. This analysis aims to determine whether grant-participating districts outperformed non-grant districts in STAAR math results. Using SAS, we cleaned and merged STAAR performance data, demographic information, and grant participation records. The cleaned dataset was used to compare outcomes and conduct a statistical analysis. Visualizations and summary dashboards were created using Power BI to present findings clearly and interactively.

Data Cleaning and Analysis:

a) Data Import

The first step involved importing three key datasets into the SAS environment using the *PROC IMPORT* procedure. These included district-level STAAR math results for the 2022–23 school year, a list of districts that received grant funding, and a demographics file containing information such as economically disadvantaged and emergent bilingual student percentages. These files served as the foundational inputs for all subsequent data cleaning and analysis steps.

- district_staar_22-23.csv: district-level STAAR math results
- grant_district_list.xlsx: list of districts that received grant funding.
- All_districts_info_22-23.xlsx: demographic and enrollment data

b) Data Exploration

```
/* Inspect structure of the imported datasets */
proc contents data=work.staar; run;
proc contents data=work.grant_list; run;
proc contents data=work.district_info; run;
```

grant_district_list

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	District	Char	6	$6.	$6.	District

district_staar_22-23

Alphabetic List of Variables and Attributes					
#	Variable	Type	Len	Format	Informat
6	GRADE	Num	8	BEST12.	BEST32.
5	LANGUAGE	Char	7	$7.	$7.
2	district	Num	8	BEST12.	BEST32.
3	dname	Char	15	$15.	$15.
14	m_all_approgl_nm	Num	8	BEST12.	BEST32.
16	m_all_mastrgl_nm	Num	8	BEST12.	BEST32.
15	m_all_meetsgl_nm	Num	8	BEST12.	BEST32.
13	m_all_unsatgl_nm	Num	8	BEST12.	BEST32.
12	m_docs	Num	8	BEST12.	BEST32.
4	region	Num	8	BEST12.	BEST32.
9	rla_all_approgl_nm	Num	8	BEST12.	BEST32.
11	rla_all_mastrgl_nm	Num	8	BEST12.	BEST32.
10	rla_all_meetsgl_nm	Num	8	BEST12.	BEST32.
8	rla_all_unsatgl_nm	Num	8	BEST12.	BEST32.
7	rla_docs	Num	8	BEST12.	BEST32.
1	year	Num	8	BEST12.	BEST32.

All_districts_info_22-23

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
1	District	Char	6	$6.	$6.	District
2	District Name	Char	50	$50.	$50.	District Name
3	District Type	Char	11	$11.	$11.	District Type
4	ESC Region	Char	2	$2.	$2.	ESC Region
5	Enrollment	Num	8	BEST.		Enrollment
10	Overall Rating SY21_22	Char	18	$18.	$18.	Overall Rating SY21_22
9	Overall Scale Score SY21_22	Num	8	BEST.		Overall Scale Score SY21_22
7	Percent Economically Disadvantag	Num	8	BEST.		Percent Economically Disadvantaged
6	Percent Emergent Bilingual	Num	8	BEST.		Percent Emergent Bilingual
8	TEA Description	Char	29	$29.	$29.	TEA Description

c) Cleaning and Filtering

To prepare the STAAR data for meaningful analysis, the raw dataset was filtered with a new dataset to include only the most relevant records for the analysis. This included keeping only math test results for students in grades 3, 4, and 5 from the 2022–23 school year. Observations with missing or zero values in key columns such as m_docs (number of student answer documents) or m_all_meetsgl_nm (number of students meeting grade level) were excluded to ensure data quality. This filtration was performed within a DATA step to create the staar_math_clean dataset. After applying these filters, the resulting dataset included over 4000 valid records representing reliable and complete data points for analysis.

```
data work.staar_math_clean;
    set work.staar;
    where year = 23 and GRADE in (3, 4, 5) and
    not missing(m_docs) and m_docs > 0 and not missing(m_all_meetsgl_nm);
```

d) Transformation: New Performance Metric

After filtering the STAAR dataset to include valid math records, a new variable was created to represent student performance in a standardized way across districts. The variable percent_meets_math calculates the percentage of students who achieved "Meets Grade Level" or better by dividing the number of students meeting the standard (m_all_meetsgl_nm) by the total number of student math documents (m_docs), then multiplying by 100 to express it as a percentage. This metric became the core performance indicator used throughout the analysis to assess the effectiveness of grant participation. The transformation was applied within the same DATA step used for initial filtering. First 10 rows were printed to check data subset logic.

```
    /* Calculate % Meets Grade Level or better in Math */
    percent_meets_math = (m_all_meetsgl_nm / m_docs) * 100;
    /* Convert district ID to character data type for proper joining */
    length district_char $6;
    district_char = put(district, z6.);
run;
```

Converted the numeric district ID into a **6-digit character format (district_char)** using put(district, z6.) to ensure consistent merging across datasets with differently formatted identifiers.

Obs	year	district	dname	region	LANGUAGE	GRADE	m_all_meetsgl_nm	m_all_mastrgl_nm	percent_meets_math	district_char
1	23	1902	CAYUGA ISD	7	english	3	17	3	48.5714	001902
2	23	1903	ELKHART ISD	7	english	3	30	9	35.2941	001903
3	23	1904	FRANKSTON ISD	7	english	3	41	9	66.1290	001904
4	23	1906	NECHES ISD	7	english	3	8	1	42.1053	001906
5	23	1907	PALESTINE ISD	7	english	3	77	23	34.0708	001907
6	23	1908	WESTWOOD ISD	7	english	3	28	7	25.6881	001908
7	23	1909	SLOCUM ISD	7	english	3	5	1	23.8095	001909
8	23	2901	ANDREWS ISD	18	english	3	112	30	39.7163	002901
9	23	3801	PINEYWOODS COMM	7	english	3	38	10	43.6782	003801
10	23	3902	HUDSON ISD	7	english	3	101	46	52.0619	003902

e) Rationale for Selecting the “Meets Grade Level” Metric

While the STAAR dataset includes multiple performance levels in math (Unsatisfactory, Approaches, Meets, and Masters), we selected “Meets Grade Level or Better” (m_all_meetsgl_nm) as the performance benchmark for this analysis. This aligns with TEA’s typical academic accountability standards and provides a clear proficiency threshold for assessing program impact. Although the performance task prompt only stated "math" outcomes for grades

3–5, the "Meets" level represents the minimum proficiency standard generally used to evaluate instructional effectiveness, making it the most appropriate indicator for measuring improvement across districts.

f) Aggregation to District Level

To analyze the data at a district level, the individual student performance records were aggregated using SQL procedures in SAS. The goal was to compute an average math performance metric for each district by averaging the previously calculated percent_meets_math values. This was accomplished using the PROC SQL statement with a GROUP BY clause on the district identifier. The resulting dataset, math_avg_by_district, included one row per district, each containing the district's average percent of students who met or exceeded the grade-level standard. This step allowed for a clean comparison across districts and enabled a fair evaluation of the grant program's effectiveness at the district level.

Obs	district	avg_percent_meets
1	001902	51.9799
2	001903	46.1811
3	001904	66.6991
4	001906	44.9771
5	001907	35.5208
6	001908	29.9595
7	001909	39.8557
8	002901	24.5334
9	003801	32.0927
10	003902	51.9662

→ Used PROC SQL to compute:
→ mean(percent_meets_math) as avg_percent_meets group by district.
→ Resulting dataset: math_avg_by_district, capturing each district's average math performance across grades

g) Merging with Grant Participation Data

To assess whether grant participation influenced student math outcomes, the next step was to identify which districts in the cleaned dataset were part of the TEA grant program. This involved merging the math_avg_by_district dataset with the grant_district_list dataset using a common district identifier, which was standardized to a 6-digit character format. During the merge, a new binary flag variable called grant_participant was created and set to 1 for participating districts and 0 for non-participants. This indicator allowed for side-by-side comparisons in the analysis. After merging, a frequency check using PROC FREQ confirmed that 198 districts were flagged as grant participants, providing confidence that the merge was accurate and complete.

The FREQ Procedure

grant_participant	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	972	82.94	972	82.94
1	200	17.06	1172	100.00

- Matched districts to the grant list using district ID (put(district, z6.))
- Created a binary flag grant_participant = 1 or 0
- Validated match using PROC FREQ (example: 198 districts flagged as grant participants)

Obs	district_char	avg_percent_meets	grant_participant
1	001902	51.9799	0
2	001903	46.1811	0
3	001904	66.6991	0
4	001906	44.9771	0
5	001907	35.5208	0
6	001908	29.9595	0
7	001909	39.8557	0
8	002901	24.5334	0
9	003801	32.0927	0
10	003902	51.9662	0

h) Adding Demographic Data

To strengthen the context of our performance analysis, demographic characteristics were integrated into the dataset. This was achieved by renaming the district identifier column in the district_info dataset to match the format used in earlier steps and then merging it with the combined performance and grant participation data. The merge introduced two key variables for each district: the percentage of students classified as economically disadvantaged and the percentage of emergent bilingual students. These factors are critical in interpreting performance outcomes and ensuring any observed differences are analyzed within the broader socio-economic context. The result of this step was a unified dataset, math_final, that included both academic performance metrics and demographic information.

- Renamed District to district_char and merged with district_info table
- Brought in variables like:
    - % Economically Disadvantaged
    - % Emergent Bilingual
- Final dataset before cleaning: math_final

Obs	district_char	avg_percent_meets	grant_participant	District Name	District Type	ESC Region	Enrollment	Percent Emergent Bilingual	Percent Economically Disadvantag	TEA Description	Overall Scale Score SY21_22	Overall Rating SY21_22
1	001902	51.9799	0	CAYUGA ISD	INDEPENDENT	07	594	0.0134680135	0.4966329966	Rural	93	A
2	001903	46.1811	0	ELKHART ISD	INDEPENDENT	07	1194	0.0376884422	0.5896147404	Non-metropolitan Stable	92	A
3	001904	66.6991	0	FRANKSTON ISD	INDEPENDENT	07	802	0.0436408978	0.5885286783	Rural	92	A
4	001906	44.9771	0	NECHES ISD	INDEPENDENT	07	310	0.0225806452	0.5612903226	Rural	94	A
5	001907	35.5208	0	PALESTINE ISD	INDEPENDENT	07	3297	0.2302092812	0.8037609948	Independent Town	92	A
6	001908	29.9595	0	WESTWOOD ISD	INDEPENDENT	07	1414	0.0862800566	0.7446958982	Non-metropolitan Stable	92	A
7	001909	39.8557	0	SLOCUM ISD	INDEPENDENT	07	349	0.0171919771	0.5128939828	Rural	91	A
8	002901	24.5334	0	ANDREWS ISD	INDEPENDENT	18	4198	0.1841353025	0.4942829919	Other Central City Suburban	91	A
9	003801	32.0927	0	PINEYWOODS COMMUNITY ACADEMY	CHARTER	07	997	0.0651955868	0.6118355065	Charter School Districts	93	A
10	003902	51.9662	0	HUDSON ISD	INDEPENDENT	07	2768	0.099349711	0.5404624277	Non-metropolitan Stable	91	A

Alphabetic List of Variables and Attributes						
#	Variable	Type	Len	Format	Informat	Label
4	District Name	Char	50	$50.	$50.	District Name
5	District Type	Char	11	$11.	$11.	District Type
6	ESC Region	Char	2	$2.	$2.	ESC Region
7	Enrollment	Num	8	BEST.		Enrollment
12	Overall Rating SY21_22	Char	18	$18.	$18.	Overall Rating SY21_22
11	Overall Scale Score SY21_22	Num	8	BEST.		Overall Scale Score SY21_22
9	Percent Economically Disadvantag	Num	8	BEST.		Percent Economically Disadvantaged
8	Percent Emergent Bilingual	Num	8	BEST.		Percent Emergent Bilingual
10	TEA Description	Char	29	$29.	$29.	TEA Description
2	avg_percent_meets	Num	8			
1	district_char	Char	6	$6.	$6.	District
3	grant_participant	Num	8			

i) Final Cleaning and Labeling

The combined dataset underwent a final round of cleaning to ensure its quality before statistical testing and visualization. Records with missing values in either the avg_percent_meets metric or the key demographic fields (Percent Economically Disadvantaged and Percent Emergent Bilingual) were removed to maintain analytical integrity. Additionally, the avg_percent_meets values were rounded to one decimal place to standardize presentation across all reporting outputs. To enhance clarity for Power BI dashboards and analysis summaries, variable labels were added to denote the meaning of key fields. The result was a final clean dataset math_final_clean containing all necessary performance and demographic information, ready for analysis and visualization.

```
data work.math_final_clean;
    set work.math_final;
    if missing(avg_percent_meets) then delete;
    if missing('Percent Economically Disadvantag'n)
        or missing('Percent Emergent Bilingual'n) then delete;
    avg_percent_meets = round(avg_percent_meets, 0.1);
```

						Region		Emergent Bilingual	Economically Disadvantag	Description	Scale Score SY21_22	Rating SY21_22
1	001902	52.0	0	CAYUGA ISD	INDEPENDENT	07	594	0.0134680135	0.4966329966	Rural	93	A
2	001903	46.2	0	ELKHART ISD	INDEPENDENT	07	1194	0.0376884422	0.5896147404	Non-metropolitan Stable	92	A
3	001904	66.7	0	FRANKSTON ISD	INDEPENDENT	07	802	0.0436408978	0.5885286783	Rural	92	A
4	001906	45.0	0	NECHES ISD	INDEPENDENT	07	310	0.0225806452	0.5612903226	Rural	94	A
5	001907	35.5	0	PALESTINE ISD	INDEPENDENT	07	3297	0.2302092812	0.8037609948	Independent Town	92	A
6	001908	30.0	0	WESTWOOD ISD	INDEPENDENT	07	1414	0.0862800566	0.7446958982	Non-metropolitan Stable	92	A
7	001909	39.9	0	SLOCUM ISD	INDEPENDENT	07	349	0.0171919771	0.5128939828	Rural	91	A
8	002901	24.5	0	ANDREWS ISD	INDEPENDENT	18	4198	0.1841353025	0.4942829919	Other Central City Suburban	91	A
9	003801	32.1	0	PINEYWOODS COMMUNITY ACADEMY	CHARTER	07	997	0.0651955868	0.6118355065	Charter School Districts	93	A
10	003902	52.0	0	HUDSON ISD	INDEPENDENT	07	2768	0.099349711	0.5404624277	Non-metropolitan Stable	91	A

- Dropped records with missing performance or demographic values.
- Rounded avg_percent_meets to one decimal place.
- Labeled key columns for clarity in Power BI

j) Statistical Analysis: T-Test

To assess whether grant participation had a statistically significant impact on student math performance, a two-sample independent t-test was conducted using the cleaned dataset. The average percentage of students meeting grade level standards (avg_percent_meets) was compared between grant-participating districts and those that did not receive the grant. The results showed that non-grant districts had a higher mean score (40.90%) compared to grant districts (37.51%), with a mean difference of **3.39 percentage points**. The **p-values** for both the pooled and Satterthwaite methods were **< 0.01**, indicating that the observed difference is statistically significant at the 1% level. This suggests that the grant program did not yield improved math outcomes, and non-grant districts actually performed better on average in SY 2022–23. Used PROC TTEST to compare avg_percent_meets by grant participation status:

T-Test: Math % Meets Grade Level - Grant vs Non-Grant Districts

The TTEST Procedure

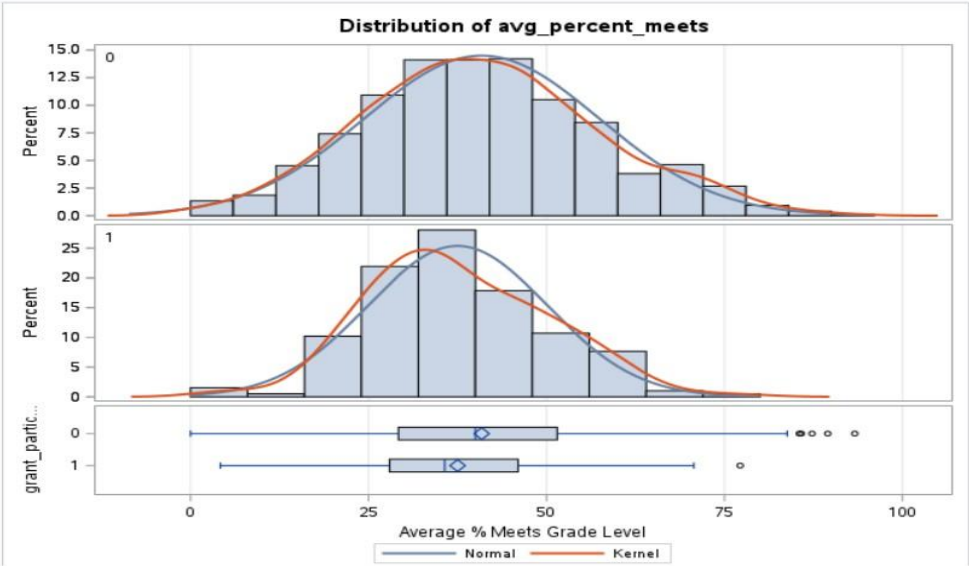
Variable: avg_percent_meets (Average % Meets Grade Level)

grant_participant	Method	N	Mean	Std Dev	Std Err	Minimum	Maximum
0		971	40.8984	16.5176	0.5301	0	93.3000
1		196	37.5087	12.5747	0.8982	4.2000	77.2000
Diff (1-2)	Pooled		3.3897	15.9258	1.2471		
Diff (1-2)	Satterthwaite		3.3897		1.0429		

grant_participant	Method	Mean	95% CL Mean		Std Dev	95% CL Std Dev	
0		40.8984	39.8581	41.9386	16.5176	15.8142	17.2870
1		37.5087	35.7373	39.2801	12.5747	11.4409	13.9599
Diff (1-2)	Pooled	3.3897	0.9429	5.8365	15.9258	15.3047	16.5999
Diff (1-2)	Satterthwaite	3.3897	1.3384	5.4410			

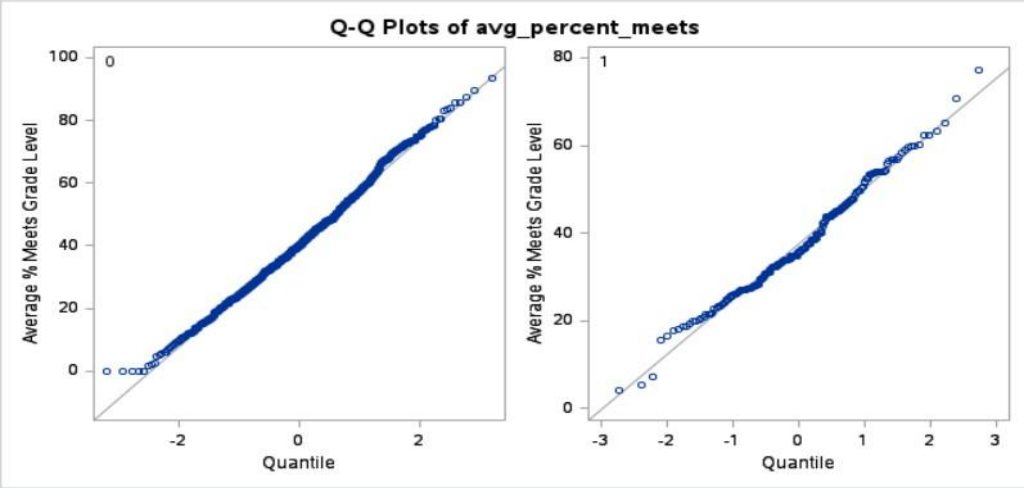
Method	Variances	DF	t Value	Pr > |t|
Pooled	Equal	1165	2.72	0.0067
Satterthwaite	Unequal	346.05	3.25	0.0013

Equality of Variances				
Method	Num DF	Den DF	F Value	Pr > F
Folded F	970	195	1.73	<.0001



The distribution of avg_percent_meets was also evaluated using Q-Q plots and histograms. The Q-Q plots for both groups followed a near-linear trend, indicating approximate normality of the data a key assumption for t-tests. Overlay histograms and kernel density plots further confirmed that while both distributions resembled a normal curve, grant districts showed slightly more variability and a lower central tendency.

These results suggest that, on average, non-grant districts outperformed those receiving grant funding in 3rd–5th grade STAAR math outcomes for SY 2022–23. This finding may warrant further investigation into the effectiveness and implementation consistency of the grant program across districts.

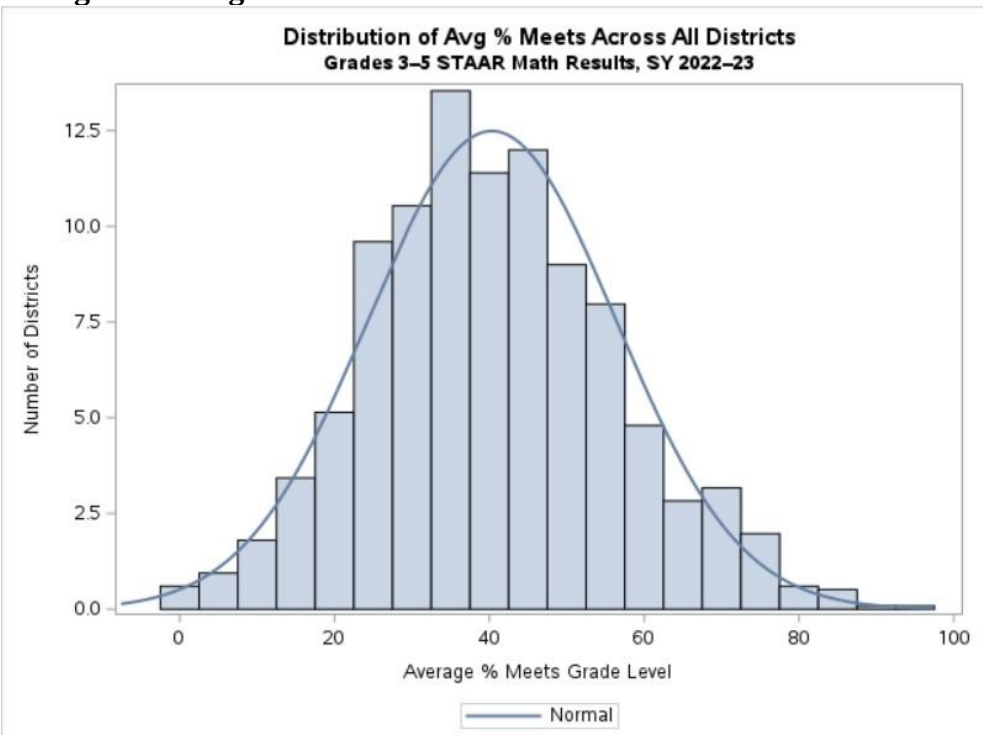


- Grant Districts avg percent meets in Math Performance:
  Grant Districts: 37.5% | Non-Grant Districts: 40.9%
- **Result:** Statistically significant difference in math performance (3.4 percentage points)

k) SAS Visualizations

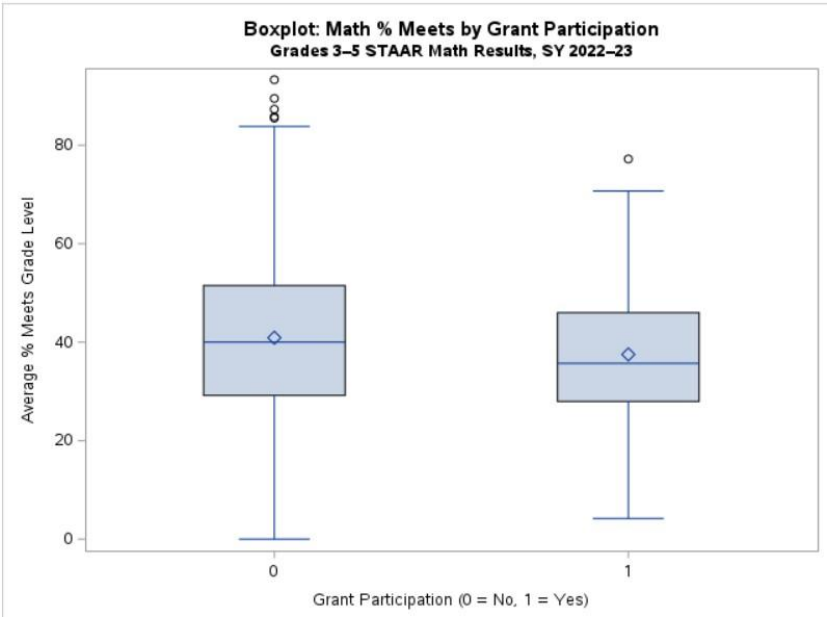
Several visualizations were created in SAS to explore district-level math performance and highlight differences between grant and non-grant districts:

- **Histogram of Avg % Meets:**



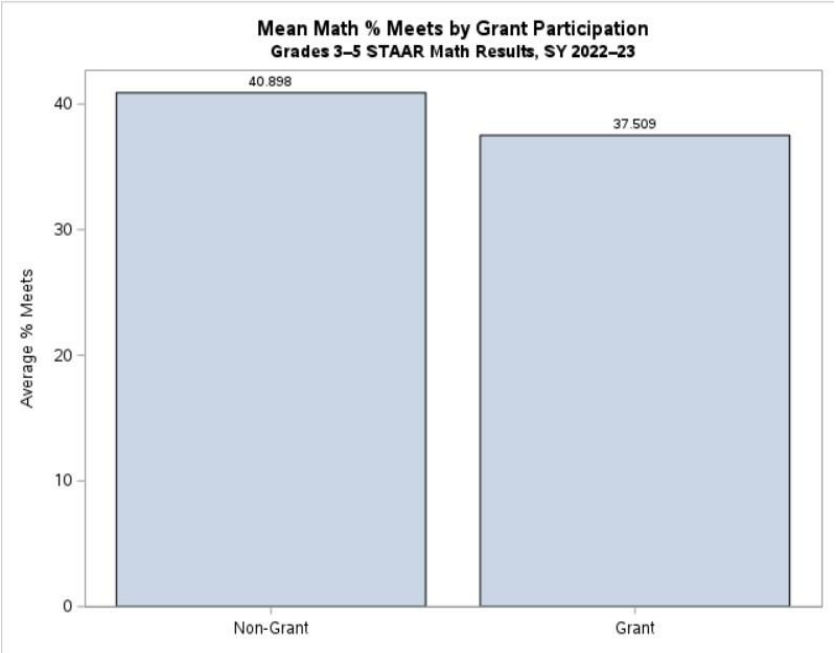
This plot displayed the distribution of avg_percent_meets across all districts. The shape resembled a bell curve, with most districts falling between 30% and 50%. A small number of districts exceeded 60%, while others performed below 20%, highlighting performance variability across the state.

- **Boxplot:**



A side-by-side boxplot compared average math performance between grant and non-grant districts. This visual clearly showed that non-grant districts had a higher median score and narrower interquartile range, suggesting more consistent performance among non-grant participants.

- **Bar Plot:**



A vertical bar chart summarizes the mean avg_percent_meets for each group. This was an effective visual reinforcement of the t-test findings, showing a 3.4 percentage difference favoring non-grant districts. It provided a straightforward comparison that could be easily understood by stakeholders.

These visuals were instrumental in validating and communicating the key insights derived from the data.

l) Statistical Summary (Descriptive)

Summary Stats: Math % Meets by Grant Participation
Grades 3–5 STAAR Math Results, SY 2022–23

The MEANS Procedure

Analysis Variable : avg_percent_meets Average %% Meets Grade Level				
Grant Participant Flag	N Obs	Mean	Std Dev	N
0	971	40.9	16.5	971
1	196	37.5	12.6	196

A summary table created using PROC MEANS confirmed the key performance differences between groups. Non-grant districts had an average STAAR math performance of 40.9%, while grant districts averaged 37.5%. Standard deviation was higher in non-grant districts (16.5% vs. 12.6%), indicating wider variation in performance. These descriptive statistics support the findings from the t-test and visual analyses.

Additional Visualizations for key insights using Power BI:

After completing the data cleaning and statistical analysis in SAS, the final dataset was imported into Power BI to create an interactive and visually engaging dashboard. The purpose of this dashboard was to allow program stakeholders to explore trends in performance outcomes across districts and regions, with specific attention to grant participation status and equity metrics.

KPI Cards:

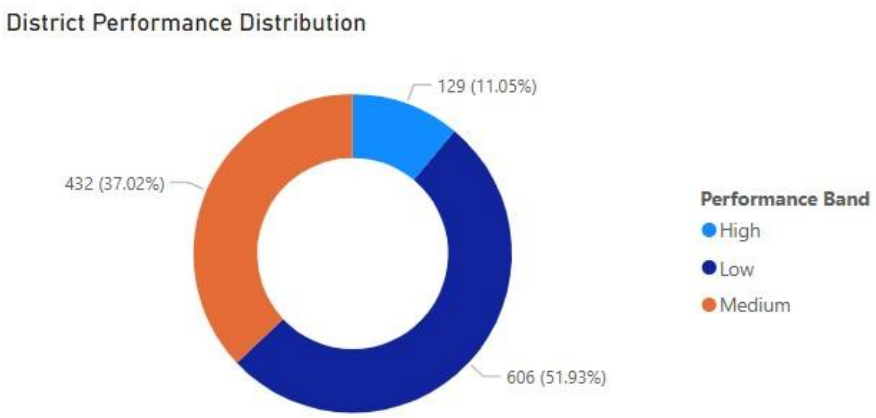
Three key performance indicators (KPIs) were created to highlight the average STAAR math performance for both grant and non-grant districts, along with their absolute difference. These metrics provide a clear snapshot of how the two groups compared:

- *Avg % Meets (Grant)*: 37.5%
- *Avg % Meets (Non-Grant)*: 40.9%
- *Performance Gap*: +3.4 percentage points (non-grant districts outperforming)

Performance Bands (Pie Chart):

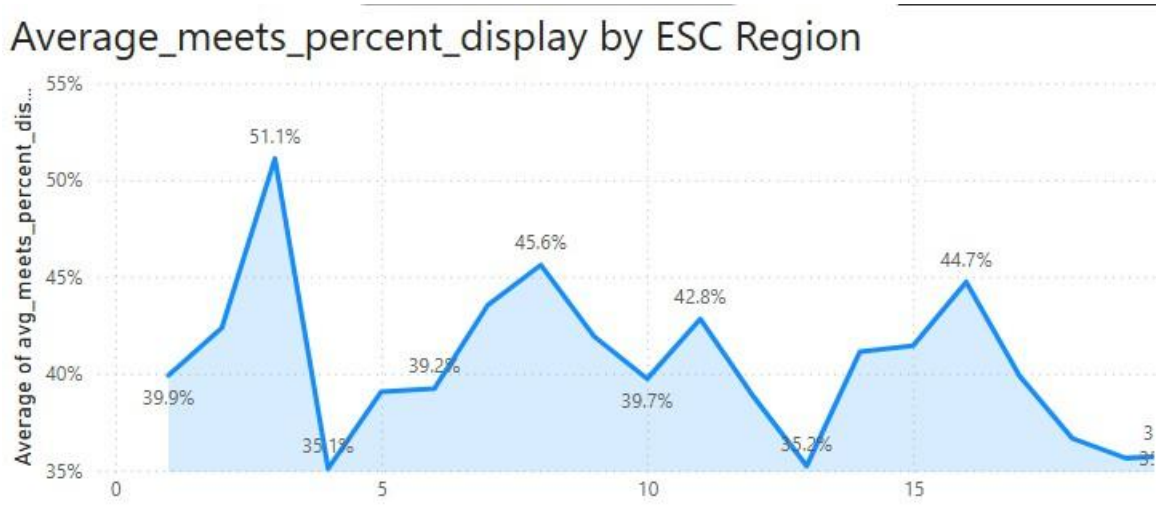
To visualize the overall distribution of district performance, a calculated column was used to group districts into performance bands (High, Medium, Low). The resulting pie chart shows that: Only a small percentage of districts achieved high performance (≥60%). The majority fall in the medium band (40–59%).

- 11% of districts achieved high performance (≥60%)
- 37% performed in the medium range (40–59%)
- 52% fell below 40%, considered low performing.



ESC Region Line Chart:

A line chart was created to display average performance by ESC region. This visual reveals significant geographic variation, with some Education Service Centers achieving averages above 45%, while others remain below 36%.

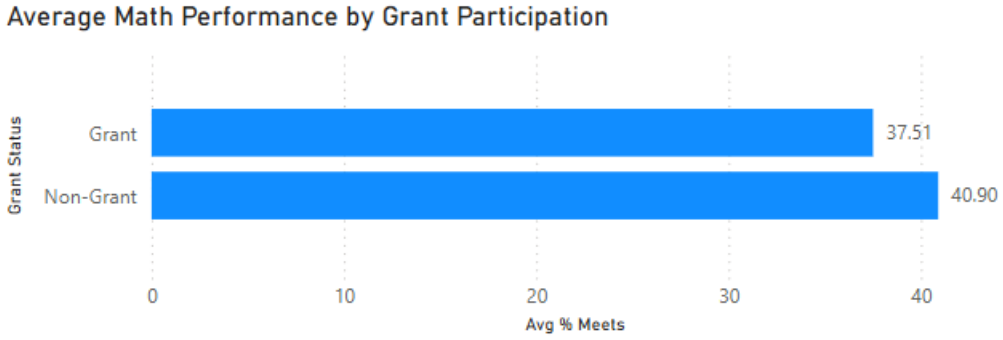


Interactive Table with Filters

A detailed table was added to allow exploration of individual district performance, demographic context (e.g., economically disadvantaged %, emergent bilingual %), and grant participation flag. This table helps stakeholders examine patterns district by district.

district_char	Grant Status	Performance Band	avg_meets_percent_display
001902	Non-Grant	Medium	52.0%
001903	Non-Grant	Medium	46.2%
001904	Non-Grant	High	66.7%
001906	Non-Grant	Medium	45.0%
001907	Non-Grant	Low	35.5%
001908	Non-Grant	Low	30.0%
001909	Non-Grant	Low	39.9%
002901	Non-Grant	Low	24.5%
003801	Non-Grant	Low	32.1%
003902	Non-Grant	Medium	52.0%
003903	Non-Grant	Low	28.0%
003904	Non-Grant	Low	27.5%
003905	Grant	Low	36.5%
003906	Non-Grant	Low	38.4%

Bar Chart: Avg % Meets by Grant Participation



This bar chart shows that non-grant districts outperformed grant-funded districts in math achievement for Grades 3–5 in SY 22–23."

Slicer Controls

Filters were incorporated using slicers for district type and ESC region. This feature enables users to focus on subsets of districts to gain more localized insights or conduct targeted comparisons.

These visuals collectively support the statistical findings from SAS and provide a powerful tool for decision-makers to explore patterns, identify disparities, and assess the effectiveness of the grant program across Texas.

Overall Conclusion

This analysis demonstrated a comprehensive, data-driven evaluation of the TEA Strategic Compensation Grant's impact on elementary math performance. By integrating multiple datasets, applying robust cleaning and transformation logic in SAS, conducting statistical tests, and building interactive Power BI dashboards, we found that non-grant districts outperformed grant recipients by an average of 3.4 percentage points. The results, though surprising, were statistically significant and reinforced through visualization and regional comparisons. These findings highlight the importance of not only funding programs, but also continuously assessing their implementation effectiveness and regional context to ensure the intended outcomes are achieved.