

Longitudinal Analysis of Academic Growth Patterns at Morningside Academy: A Five-Year Assessment Study

Saratessa Palos, Will Arangelov, Marcus Cumberbatch

December 11, 2024

1 Abstract

This study analyzed longitudinal assessment data from Morningside Academy, a private lab school specializing in behavioral education approaches for students with academic challenges. We examined student performance across five academic years (2008-2009, 2015-2016, 2016-2017, and 2023-2024) using secondary data analysis of scores from multiple standardized assessments including Woodcock-Johnson (WJ), Iowa Test of Basic Skills (ITBS), and Iowa Assessments (IA). The dataset comprises student assessment scores from four academic years (2023-2024: n=91, 2016-2017: n=88, 2015-2016: n=91, and 2008-2009: n=82). Growth calculations were performed on reading, writing, and mathematics measures to analyze progress between fall and spring testing periods. Data processing involved standardizing student identifiers, calculating growth scores, and merging datasets for comprehensive analysis. Results provide insights into the effectiveness of the MMGI educational approach through examination of academic growth patterns across different domains and years.

2 Introduction

The Morningside Model of Generative Instruction (MMGI) is a research based, empirically validated educational model shown to be extremely effective in developing competent learners. Morningside Academy, established in Seattle, Washington, utilizes the MMGI approach to help elementary and middle school children who are one or more years behind expected grade level to catch up and excel academically.

The school guarantees that students will gain at least two grade levels in their area of greatest deficit within one academic year. Unlike traditional special education models that emphasize compensatory strategies, Morningside directly addresses student deficits through performance-based placement and systematic instruction.

This study aims to validate these educational approaches through analysis of standardized assessment data collected over four academic years. The use of multiple assessment systems (WJ, ITBS, and IA) provides complementary measures of student achievement across reading, writing, and mathematics domains. Understanding patterns of academic growth is crucial for evaluating educational effectiveness and informing instructional practices.

Table 1: Table 1: Average Growth by Academic Domain

Year	Academic Growth Averages		
	Mean Reading	Mean Writing	Mean Math
2008-2009	1.52	1.91	1.90
2015-2016	2.43	2.57	2.52
2016-2017	2.39	2.46	NaN
2023-2024	1.42	NaN	2.38

Note: Note: NaN indicates missing data for that academic year.

To better assess the efficacy of the MMGI, we analyzed 1) patterns of academic growth across different assessment types (math, reading, and writing) between fall and spring testing periods, and 2) how patterns for these domains vary across years analyzed.

3 Methods

This study analyzed student assessment data from five academic years at Morningside Academy (2008-2009, 2015-2016, 2016-2017, and 2023-2024). Data cleaning and analysis were conducted using R statistical software with tidyverse, readxl, corrplot, viridis, and gridExtra packages. For each academic year, fall and spring assessment scores were processed separately. All datasets required standardization of ceiling scores (converting ">12.9" to "12.9" and "<3.0" to "3.0"), and numeric conversion of assessment scores. Growth scores were calculated by subtracting fall scores from spring scores across multiple academic measures including Reading Comprehension (RC), Letter-Word Identification (LWID), Word Attack (WA), Reading Fluency (RF), Writing Fluency (WF), Writing Samples (WS), Calculation (CALC), and Math Fluency (MF). Student names were standardized across datasets using a custom function to handle different name formats (LastName, FirstName vs. FirstName LastName). The datasets were then merged using bind_rows after selecting standardized names, year identifiers, and growth measures.

4 Results

Our analytic strategy first involved obtaining aggregate scores of different growth measure types across participants. Reading comprehension, letter-word identification, and reading fluency measures were coalesced into composite scores. Next, we repeated this process for writing, by combining writing fluency, writing samples, and word attack growth scores. Finally, we combined the two assessment measures from mathematics, CALC as well as math fluency, into averaged scores. We then obtained grand means for each of the three academic domains (Reading, Writing, and Math). Refer to table 1 for a breakdown of these means.

In addition to means analyses, we conducted statistical tests to observe mean comparisons of academic domains, on aggregate. These tests allowed us to account for non-unique participant identification numbers within the current data, and further enabled us to observe potential differences in growth patterns across various academic years.

4.0.1 Research Question 1

4.0.1.1 Analytic Strategy A repeated measures analysis of variance (ANOVA) was conducted via the `aov` function in R to quantify patterns of academic growth across different assessment measure types. Individual measure types were aggregated by domain across participants to create three, composite academic domains of math, reading, and writing.. The repeated measures approach was utilized because growth scores were measured several times for each student across several assessment types within a given year (as well as during multiple years). Thus, we treated `assessment_type` as a within-subject factor. Since the data included non-unique student IDs, the model accounted for repeated measures per student. Post-hoc pairwise comparisons with Bonferroni correction were conducted to determine specific differences between the three assessment types, to account for uneven sample sizes within year.

4.0.1.2 Results A repeated measures ANOVA was conducted to examine the effect of Assessment Type (reading, writing, and math) on academic growth scores. Results showed a significant main effect of `assessment_type`, $F(2, 259) = 5.81, p < 0.003$, indicating that growth scores varied across assessment types. Additionally, there was a significant main effect of `year`, $F(3, 259) = 9.27, p < 0.001$, suggesting that academic growth differed across academic years. However, the interaction between `assessment_type` and `year` was not significant, $F(1, 402) = 0.62, p < 0.43$, indicating that growth patterns across assessments were consistent over time.

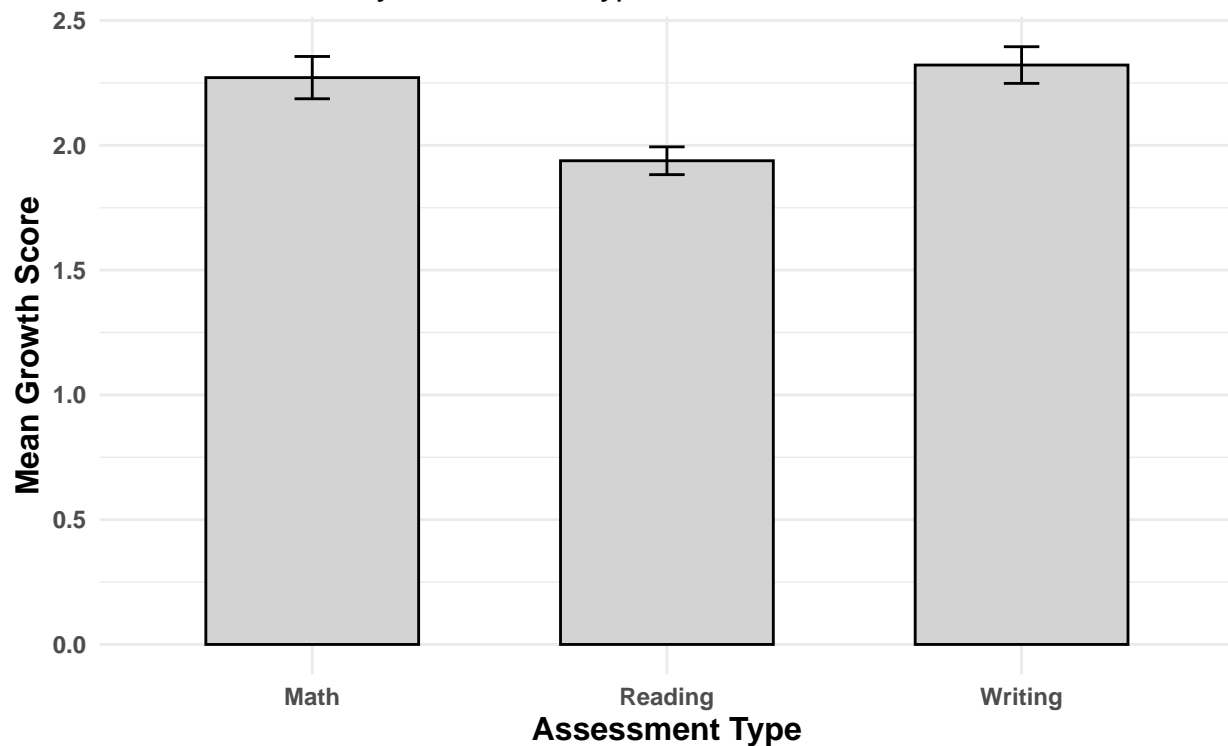
Post-hoc pairwise comparison tests using a Bonferroni correction revealed that growth in Reading was significantly lower than in Math ($M_{\text{difference}} = -0.92, p = 0.015$) and Writing ($M_{\text{difference}} = -1.10, p = 0.003$). No significant differences emerged between Math and Writing ($p = 1.00$), however. These comparisons suggest that students experienced the least academic growth in reading, while growth in math and writing was statistically similar (refer to the ANOVA model output below for full transparency). Results are shown in Fig. 1.

```
##
## Error: id
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Assessment_Type  2   24.5  12.275    5.806 0.00341 **
## year            3   58.8  19.603    9.273 7.59e-06 ***
## Residuals      259  547.5    2.114
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: id:Assessment_Type
##               Df Sum Sq Mean Sq F value    Pr(>F)
## Assessment_Type  2   22.3  11.159    7.416 0.000687 ***
## year            1    0.9   0.927    0.616 0.432913
## Residuals      402  604.9    1.505
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Error: id:year
##               Df Sum Sq Mean Sq F value    Pr(>F)
```

```
## year      1    0.92  0.9153    0.81  0.372
## Residuals 53   59.87  1.1296
##
## Error: id:Assessment_Type:year
##           Df Sum Sq Mean Sq F value Pr(>F)
## Residuals 33   35.64    1.08
##
##
## Pairwise comparisons using t tests with pooled SD
##
## data: analysis_long$Growth_Score and analysis_long$Assessment_Type
##
##      Math   Reading
## Reading 0.0147 -
## Writing 1.0000 0.0031
##
## P value adjustment method: bonferroni
```

Fig. 1

Academic Growth by Assessment Type



4.0.2 Research Question 2

4.0.2.1 Analytic Strategy To examine how academic growth varied across different academic years, a linear mixed-effects model (LMM) was conducted, using the `lmer` and `lm` functions within

R, in addition to `car`, `lme4`, and `lmerTest` for relevant statistics of significance in the LMM outputs. Academic year was utilized as a fixed effect, with student ID included as a random effect to account for repeated measurements, due to certain students appearing in the data set multiple times across years. This method allowed us to model within-student variability while estimating the average effect of academic year on academic growth in reading, writing, and math. The LMM was fit using restricted maximum likelihood (REML) for unbiased estimates of fixed and random effects.

4.0.3 Results

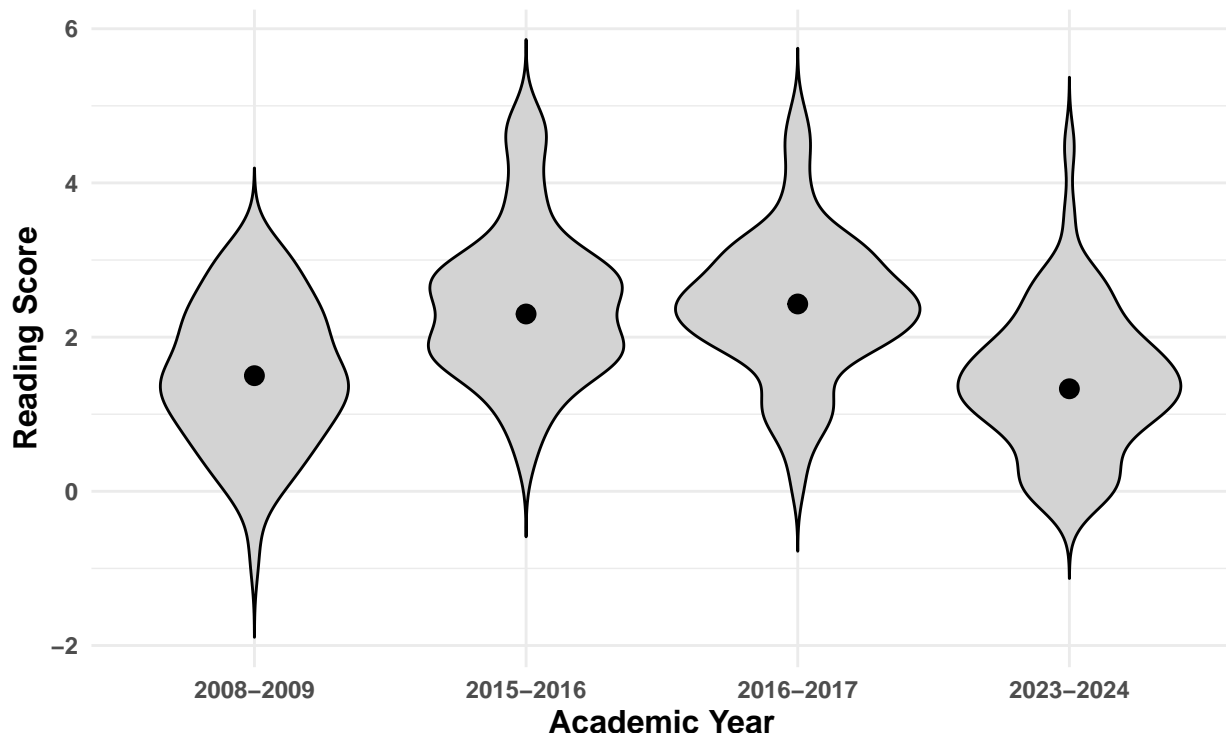
4.0.3.1 Reading A linear mixed-effects model (LMM) was conducted to assess the effect of academic year on reading growth scores, with year 2008-2009 serving as a reference, or baseline year for comparisons. Academic year was treated as a fixed effect, and student ID was entered as a random intercept to account for repeated measurements. Results suggest that academic year significantly predicted reading scores, $F(3, 276) = 14.24, p < .001$. Students in the 2015-2016 academic year scored significantly higher than the reference year (2008-2009), $B = 0.91, SE = 0.15, t(276) = 6.04, p < .001$. Similarly, students in 2016-2017 also showed significant growth relative to the baseline, $B = 0.87, SE = 0.16, t(276) = 5.53, p < .001$. However, no significant difference was found between the 2023-2024 year and the reference year, $B = -0.10, SE = 0.16, t(276) = -0.62, p = .533$. Refer to the model outputs for reading, writing, and math below for full transparency. Results for reading are shown in Fig. 2.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: reading ~ year + (1 | id)
## Data: analysis_avgs
##
## REML criterion at convergence: 759.4
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.5706 -0.6456 -0.0186  0.5856  3.2849
##
## Random effects:
## Groups   Name                Variance Std.Dev.
## id      (Intercept)  7.725e-16  2.779e-08
## Residual                    8.624e-01  9.287e-01
## Number of obs: 280, groups: id, 246
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)    1.51724    0.10653  276.00000    14.243 < 2e-16 ***
## year2015-2016    0.91236    0.15115  276.00000     6.036 5.08e-09 ***
## year2016-2017    0.87120    0.15756  276.00000     5.529 7.45e-08 ***
## year2023-2024   -0.09785    0.15690  276.00000    -0.624  0.533
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
## Correlation of Fixed Effects:
##          (Intr) y2015- y2016-
## yr2015-2016 -0.705
## yr2016-2017 -0.676  0.477
## yr2023-2024 -0.679  0.479  0.459
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')
```

Fig. 2

Reading Score Distribution by Academic Year



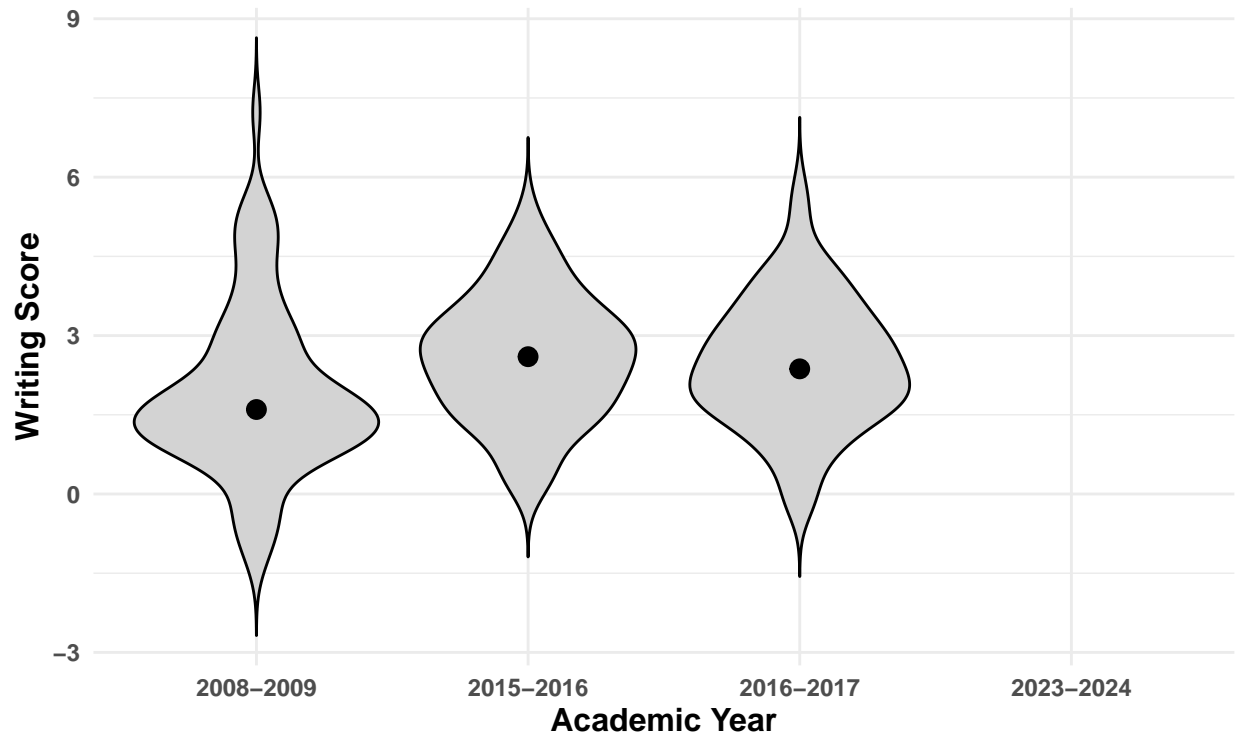
4.0.3.2 Writing A linear mixed-effects model was applied to examine the effect of academic year on writing growth scores. Similar to reading, academic year was treated as a fixed effect, and student ID was entered as a random intercept to account for repeated measurements. Academic year significantly predicted writing scores, $F(2, 242) = 12.55, p < .001$. Writing scores improved significantly in 2015-2016, $B = 0.67, SE = 0.21, t(242) = 3.11, p = .002$, and 2016-2017, $B = 0.55, SE = 0.21, t(242) = 2.62, p = .009$ compared to 2008-2009. Due to missing data, comparisons between 2008-2009, and 2023-2024 could not be analyzed. Results for writing are shown in Fig. 3.

```
## Linear mixed model fit by REML. t-tests use Satterthwaite's method [
## lmerModLmerTest]
## Formula: writing ~ year + (1 | id)
## Data: analysis_avgs
##
```

```

## REML criterion at convergence: 842.9
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.3664 -0.6248 -0.1401  0.6041  3.9599
##
## Random effects:
##   Groups   Name      Variance Std.Dev.
##   id       (Intercept) 0.000    0.000
##   Residual                1.805    1.344
## Number of obs: 245, groups: id, 192
##
## Fixed effects:
##              Estimate Std. Error    df t value Pr(>|t|)
## (Intercept)    1.9095     0.1521 242.0000  12.552 < 2e-16 ***
## year2015-2016  0.6650     0.2138 242.0000   3.110  0.00209 **
## year2016-2017  0.5488     0.2095 242.0000   2.619  0.00937 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Correlation of Fixed Effects:
##              (Intr) y2015-
## yr2015-2016 -0.712
## yr2016-2017 -0.726  0.517
## optimizer (nloptwrap) convergence code: 0 (OK)
## boundary (singular) fit: see help('isSingular')

```

Fig. 3*Writing Score Distribution by Academic Year*

4.0.3.3 Math A linear regression model was conducted to examine the effect of academic year on math growth scores, due to lack of unique participant IDs relative to the number of observations. Academic year significantly predicted math scores, $F(2, 229) = 3.42, p = .034$. Math scores were significantly higher in the 2015-2016 academic year compared to the reference year (2008-2009), $B = 0.62, SE = 0.25, t(229) = 2.52, p = .012$. In contrast, students' math scores in 2023-2024 did not significantly differ from the reference year, $B = 0.48, SE = 0.26, t(229) = 1.85, p = .065$. Due to missing data, comparisons between 2008-2009, and 2016-2017 could not be analyzed. Results for math are shown in Fig. 4. The low variance values for each model between reading, writing, and math suggests that most of the variability in domain scores can be attributed to the effects of year, rather than between-student effects. Results for math are shown in Fig. 4.

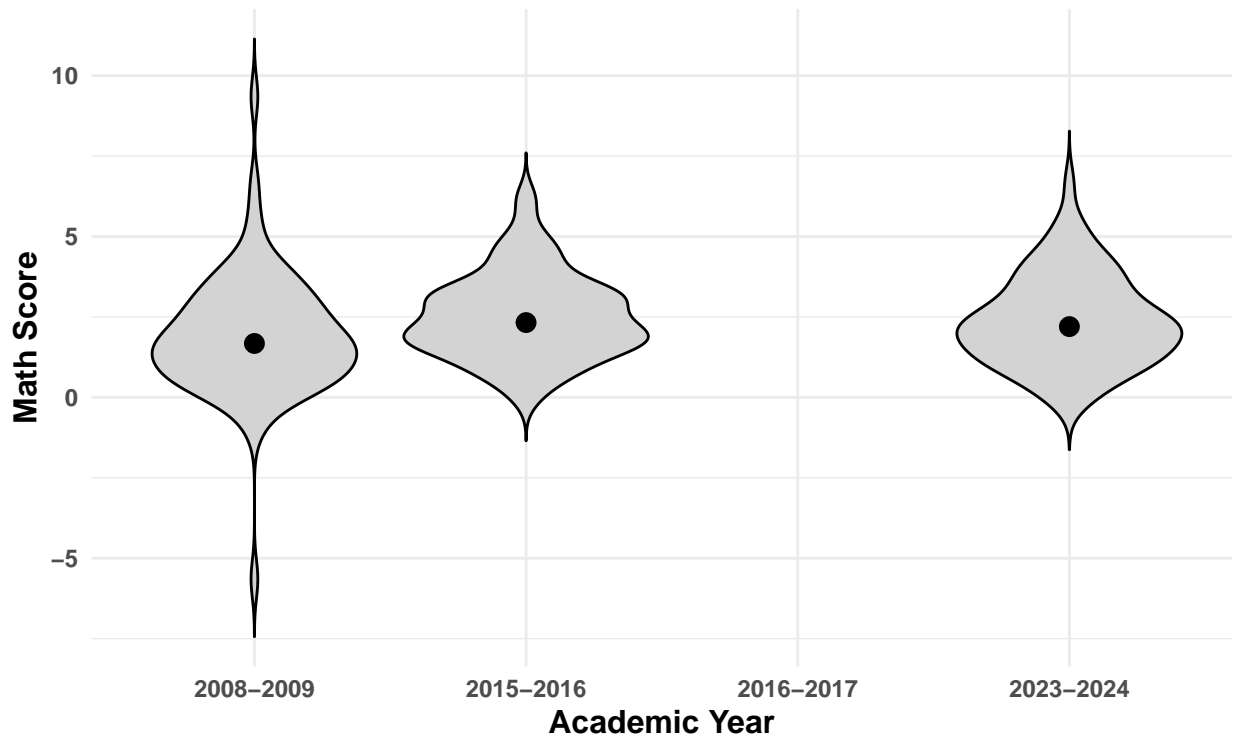
```
##
## Call:
## lm(formula = math ~ year, data = analysis_avgs)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.553  -0.980  -0.180   0.872   7.447
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.9032     0.1769  10.761  <2e-16 ***
```



```
## year2015-2016    0.6188      0.2456    2.519    0.0124 *
## year2023-2024    0.4768      0.2572    1.854    0.0650 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.562 on 229 degrees of freedom
## (115 observations deleted due to missingness)
## Multiple R-squared:  0.02898,    Adjusted R-squared:  0.0205
## F-statistic: 3.417 on 2 and 229 DF,  p-value: 0.03448
```

Fig. 4

Math Score Distribution by Academic Year



5 Discussion

The main goal of this research paper is to provide evidence for the effectiveness of the Morningside Model of Generative Instruction for children behind in their educational capabilities. We used four waves of longitudinal data from the Morningside Academy in Seattle, Washington, specifically 2008-2009; 2015-2016; 2016-2017, and 2023-2024. Focusing on multiple waves allowed us to establish a more precise view of the educational process in this institution and account for fluctuations across the years and domains. Our main focus of the analysis was on the macro level of testing used, the pre- and post-test, as this is the participants' most important measurement of growth (Street & Johnson, 2014). The descriptive statistics showed that this model is effective in educational settings and provided evidence that using scientific learning and a task-oriented

environment supports the progress of students, similar findings have been documented in previous research (Johnson & Street, 2013). Both reading, writing, and math domain scores are higher compared to the initial ones, those result echoes previous research conducted in the scholarship (Johnson & Street, 2004). Being effective in this setting and providing the promised educational advances is crucial for the funding of the academy, considering the vast dependency on tuition. This model can be sufficient, however, more governmental support might be beneficial in the long term, reducing the threat of possible closure due to a year or two of lower results, which can be based on external factors and not on the quality of education.

Our first research question concerns the growth in the academic domain for the chosen period, we found support that growth scores varied across assessment types. The data showed significantly lower reading growth in comparison to math and writing, which were on similar levels. Special education is beneficial for students with disabilities and clinical disorders and improves their reading, math, and writing (Mayes & Calhoun, 2006). Focusing on math and writing growth, we can conclude that the MMGI model yields a positive effect on students in these institutions, improving their abilities. These findings contradict emerging evidence on the effectiveness of special education that questions the practice (Kvande et al., 2019), or found a negative correlation between special education and reading and math abilities in the case of preschoolers (Sullivan & Field, 2013), and math and language skills (Lekhal, 2018). When considering reading, improvement has been documented (Ehrhardt et al., 2013), contradicting our data. The reason for the limited reading growth in our sample might be the ceiling effect, due to the measurements. Additionally, math and writing are usually domains more likely to require interventions. There are vast differences in how special education is positioned and conducted in different countries and contexts, therefore more research will be needed to expand on these contradictory findings.

Our second research question considers the waves of the data, revealing significant variations across the sample years. There was a substantial improvement in growth in the years 2015 to 2017 in all three domains of interest, while 2023-2024 is more similar to the baseline period of the study (2008-2009). We can explain these results in part with the global events influencing the educational system, such as the economic crisis of 2008 and the COVID-19 pandemic. Research has found that education is suffering in times of crisis (Barakat et al., 2010), resources are limited and children are suffering disproportionately (Shafiq, 2010), especially those in marginalized communities (Lovell & Isaacs, 2008). Similarly, the pandemic had an overall negative effect on the educational system due to the constant changes, uncertainty, digital learning (Pokhrel & Chhetri, 2021), and especially for the children with special needs (Mete Yesil et al., 2022). It is unclear whether the mentioned global crises are the sole reasons for the base and current results to be significantly lower than the ones from the 2015-2017 period, more research will be required to answer this question.

5.1 Implications

This study provides additional knowledge to the literature in the field of special education, especially scientifically-based learning tools and practical tasks. The findings can be used to inform practices and promote these educational methods for additional implementation across institutions and countries. The study provides data about the student's growth and capabilities that can be used in educational conversation, policy discussions, and development and has implications for the funding of such academies. Our findings can be beneficial for governmental and state agencies that are centering their efforts to establish and work on educational practices.

5.2 Limitations

One of the limitations of our study is failing to account for individual differences can be beneficial for providing a more detailed picture of the effectiveness of MMGI with the inclusion of comparative data across demographics. Additionally, there were a variety of testing methods used for the macro-level establishment of the student's growth in the Morningside Academy in Seattle. We decided to choose only one of the used testing methods as a primary, future research can look to create a more comprehensive contribution including all testing methods. Examining the reasons for the students to join this type of school is an additional research question that requires attention, as there are contradicting findings surrounding the effectiveness of these institutions, it is important to investigate the characteristics of the participants, their backgrounds, and lived experiences. Follow-up with participants after their time in the academy and evaluating their progress and integration into the traditional school system is another valuable research inquiry.

6 References

- Barakat, B., Holler, J., Prettnner, K., & Schuster, J. (2010). The impact of the economic crisis on labour and education in Europe (Working Paper 6/2010). Vienna Institute of Demography Working Papers. <https://www.econstor.eu/handle/10419/96975>
- Ehrhardt, J., Huntington, N., Molino, J., & Barbaresi, W. (2013). Special Education and Later Academic Achievement. *Journal of Developmental & Behavioral Pediatrics*, 34(2), 111. <https://doi.org/10.1097/DBP.0b013e31827df53f>
- Johnson, K. R., & Street, E. M. (2013). Response to intervention and precision teaching: Creating synergy in the classroom. Guilford Press.
- Johnson, P. D. K., & Street, P. D. E. M. (2004). The Morningside Model of Generative Instruction: What It Means to Leave No Child Behind. Cambridge Center for Behavioral Studies.
- Kvande, M. N., Bjørklund, O., Lydersen, S., Belsky, J., & Wichstrøm, L. (2019). Effects of special education on academic achievement and task motivation: A propensity-score and fixed-effects approach. *European Journal of Special Needs Education*, 34(4), 409–423. <https://doi.org/10.1080/08856257.2018.1533095>
- Lekhal, R. (2018). Does special education predict students' math and language skills? *European Journal of Special Needs Education*, 33(4), 525–540. <https://doi.org/10.1080/08856257.2017.1373494>
- Mayes, S. D., & Calhoun, S. L. (2006). Frequency of reading, math, and writing disabilities in children with clinical disorders. *Learning and Individual Differences*, 16(2), 145–157. <https://doi.org/10.1016/j.lindif.2005.07.004>
- Mete Yesil, A., Sencan, B., Omercioglu, E., & Ozmert, E. N. (2022). The Impact of the COVID-19 Pandemic on Children With Special Needs: A Descriptive Study. *Clinical Pediatrics*, 61(2), 141–149. <https://doi.org/10.1177/00099228211050223>
- Pokhrel, S., & Chhetri, R. (2021). A Literature Review on Impact of COVID-19 Pandemic on Teaching and Learning. *Higher Education for the Future*, 8(1), 133–141. <https://doi.org/10.1177/2347631120983481>
- Shafiq, M. N. (2010). The Effect of an Economic Crisis on Educational Outcomes: An Economic Framework and Review of the Evidence (SSRN Scholarly Paper 1709045). Social Science Research Network. <https://papers.ssrn.com/abstract=1709045>
- Street, E. M., & Johnson, K. (2014). The Sciences of Learning, Instruction, and Assessment as Underpinnings of the Morningside Model of Generative Instruction. *Acta de Investigación Psicológica*, 4(3), 1773–1793. [https://doi.org/10.1016/S2007-4719\(14\)70979-2](https://doi.org/10.1016/S2007-4719(14)70979-2)
- Sullivan, A. L., & Field, S. (2013). Do preschool special education services make a difference in kindergarten reading and mathematics skills?: A propensity score weighting analysis. *Journal of School Psychology*, 51(2), 243–260. <https://doi.org/10.1016/j.jsp.2012.12.004>