

Executive Summary:
Amira BOY/MOY Data Quality Findings

Brooke McConathy
April 17, 2025

Table of Contents

Introduction.....	3
Issue 1: Missing Component Tests.....	3
Transformations	3
Exploratory Data Analysis (EDA): Missing Component Tests	3
Concern: Missing Component Tests	4
School and Class Trends: Missing Component Tests	4
TL;DR/Suggested Next Steps	5
Issue 2: Data Quality and Impact on ARM.....	6
Transformations	6
Tests: Data Quality and Impact on ARM.....	6
ARM Growth By Class Data Completeness	7
TL;DR/Next Steps.....	8

Introduction

This summary includes an exploratory analysis of two datasets reporting beginning of year (BOY) and middle of year (MOY) Amira Benchmark Assessment scores and ARM scores for students in grades Kindergarten through Fifth Grade. This summary will examine the data for quality and performance trends. All analysis was performed in Python. Cursor assisted in refactoring and debugging the code.

Two main issues observed:

Issue 1: Many students are missing Benchmark Assessment component test data.

Issue 2: Missing component test data affects ARM scores and growth trajectories.

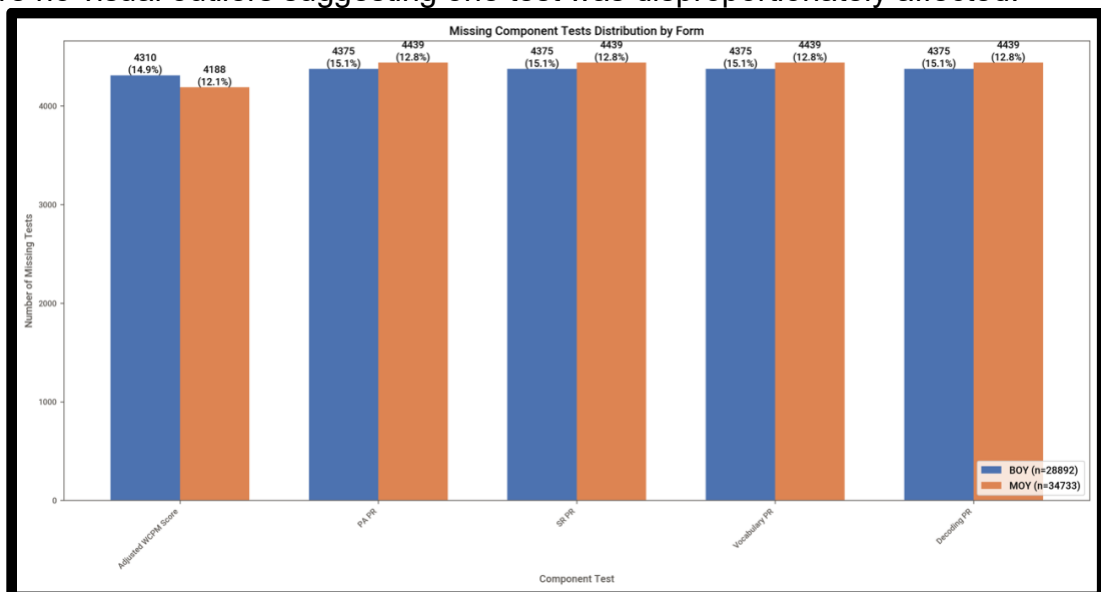
Issue 1: Missing Component Tests

Transformations

The BOY and MOY datasets were concatenated, and a “Form” attribute (“BOY” or “MOY”) was added to enable exploration of score trends over time. This produced a dataset with 13 columns and 63,625 records for analysis.

Exploratory Data Analysis (EDA): Missing Component Tests

The dataset is nearly complete, except for PR (Percentile Rank) values in the Benchmark component tests. Of the five tests, four (Phonological Awareness, Sight Reading, Vocabulary, and Decoding) are missing similar counts of PR values. There were no visual outliers suggesting one test was disproportionately affected.



Concern: Missing Component Tests

The missing PR scores are a concern. Although ***there were no missing ARM composite scores, it is unclear how ARM scores could be accurately calculated without complete PR input.*** PR scores provide valuable diagnostic information to teachers and must be included in student reporting.

School and Class Trends: Missing Component Tests

I investigated the distribution of missing scores by school and class to determine whether the issue stems from technical errors, system passback issues, or implementation challenges. Initial investigation shows:

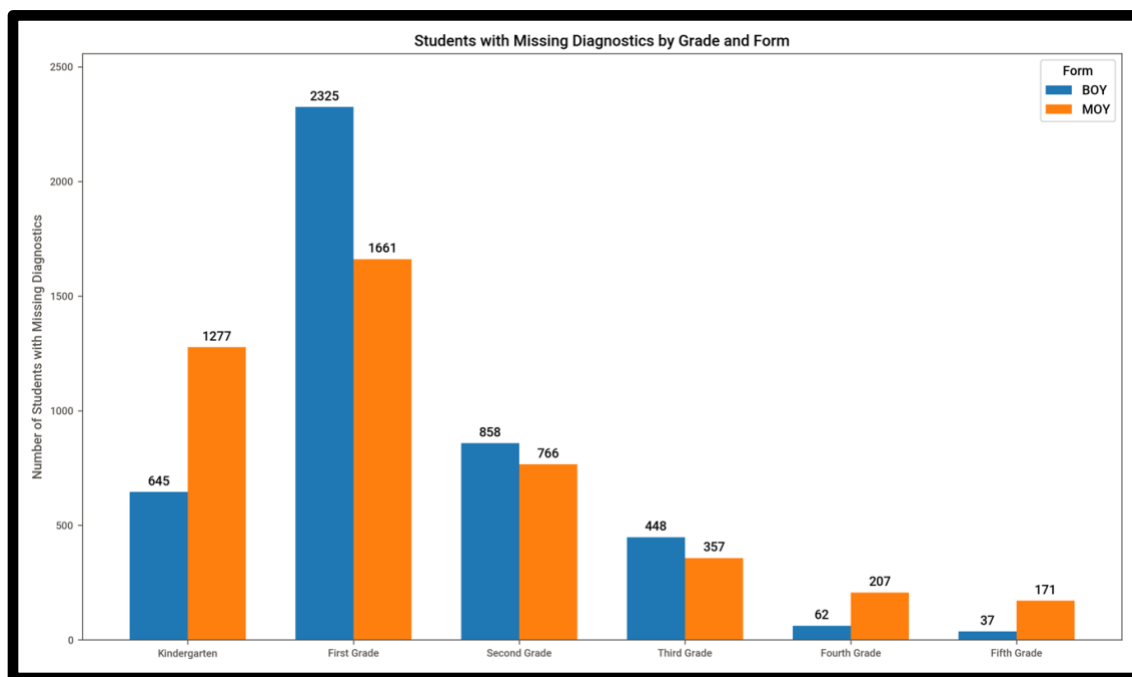
- No schools were missing all student component test scores across all classes.
- 81 total classes had all students missing component scores.
- 309 total classes had 50% or more students missing component scores.

```
--- Summary Counts ---  
=====
```

Total School Records (>50% Students Missing Any Test):	0
Total School Records (100% Students Missing Any Test):	0
Total Class Records (>50% Students Missing Any Test):	309
Total Class Records (100% Students Missing Any Test):	81

```
=====
```

Most ***missing tests were concentrated in earlier grades, specifically Kindergarten and First Grade***, suggesting that additional test administration support for early grade classrooms may be needed.



An application was developed to identify and drill down into classrooms above the 50% missing threshold. This report curates a list of classrooms that may need technical or implementation support to improve data completeness. See [full report here](#).

Filters

Select Form(s)

BOY x MOY x

Select School (for Table)

All

Select Grade (for Table)

All

Export Table Data

Download Filtered Table Data

Share ☆ ↻

Classes with High Missing Diagnostic Test Rates (>50%)

Detailed Class View School Summary

Filtered Class Data (>50% Students Missing Any Test)

Displaying 309 classes meeting the criteria based on current filters.

School ID	Grade	Class ID	Form	Total Students	Students Missing Any Test	Overall Missing Student %	Adjusted % Miss	PA % Miss	SR % Miss	Vocabulary % Miss	Decoding % Miss
033b71112d415f78	Kindergarten	f15304c3c72d33e9	BOY	4	3	75.0%	75.0%	75.0%	75.0%	75.0%	75.0%
033b71112d415f78	Kindergarten	c4dfb68e880c233a	MOY	14	9	64.3%	64.3%	64.3%	64.3%	64.3%	64.3%
044d4774e711a6f5	First Grade	d52e6af3870bbe1c	BOY	17	11	64.7%	64.7%	64.7%	64.7%	64.7%	64.7%
0552fc363beed545	First Grade	f6c4cba72d3fa5a	BOY	20	13	65.0%	65.0%	65.0%	65.0%	65.0%	65.0%
0552fc363beed545	First Grade	23d628cd68dc509d	BOY	19	11	57.9%	57.9%	57.9%	57.9%	57.9%	57.9%
0552fc363beed545	Kindergarten	9f0d172fe243407	MOY	19	15	78.9%	78.9%	78.9%	78.9%	78.9%	78.9%
085f33b3a8e7ec3f	Kindergarten	f1cc8d7ede0ecb64	MOY	21	19	90.5%	90.5%	90.5%	90.5%	90.5%	90.5%
085f33b3a8e7ec3f	Kindergarten	30d3845d90521609	MOY	16	12	75.0%	75.0%	75.0%	75.0%	75.0%	75.0%
085f33b3a8e7ec3f	Kindergarten	c5934d99d655e690	MOY	14	9	64.3%	64.3%	64.3%	64.3%	64.3%	64.3%
085f33b3a8e7ec3f	Kindergarten	e446058fa46188b8	MOY	21	13	61.9%	61.9%	61.9%	61.9%	61.9%	61.9%

TL;DR/Suggested Next Steps

- Dataset missing many PR values for Benchmark component tests.
- Concern for both ARM calculation and student reporting accuracy.
- Most students missing all 5 PR values, no indication one test more problematic than other.
- Data gaps are concentrated at the class level, not the school level.
- Lower grades (K, 1st) are more frequently missing PR data.
- Recommend identifying and supporting classes above the 50% missing threshold.

Issue 2: Data Quality and Impact on ARM

Transformations

To investigate whether missing diagnostic data impacts ARM scores, BOY and MOY data were combined across grades. Students were grouped based on the diagnostic data completeness of their classes:

- Complete (0% missing)
- Partial (1–50% missing)
- High Missing (>50% missing)

Average ARM scores were calculated at the class level for BOY, MOY, and ARM growth.

Tests: Data Quality and Impact on ARM

Independent t-tests were conducted to compare ARM performance across completeness groups. Analyses were performed across all grades combined to assess overall trends. Statistical significance was evaluated at $p < 0.05$.

BOY Score						
	Group 1	Group 2	Statistic	P-Value	Significant (p<0.05)	Message
	Complete (0%)	High Missing (>50%)	37.238	0.0000		*
	Complete (0%)	Partial (1–50%)	24.369	0.0000		*
	High Missing (>50%)	Partial (1–50%)	-22.502	0.0000		*

Change (MOY-BOY)						
	Group 1	Group 2	Statistic	P-Value	Significant (p<0.05)	Message
	Complete (0%)	High Missing (>50%)	-1.277	0.2037		
	Complete (0%)	Partial (1–50%)	-3.284	0.0011		*
	High Missing (>50%)	Partial (1–50%)	-0.262	0.7939		

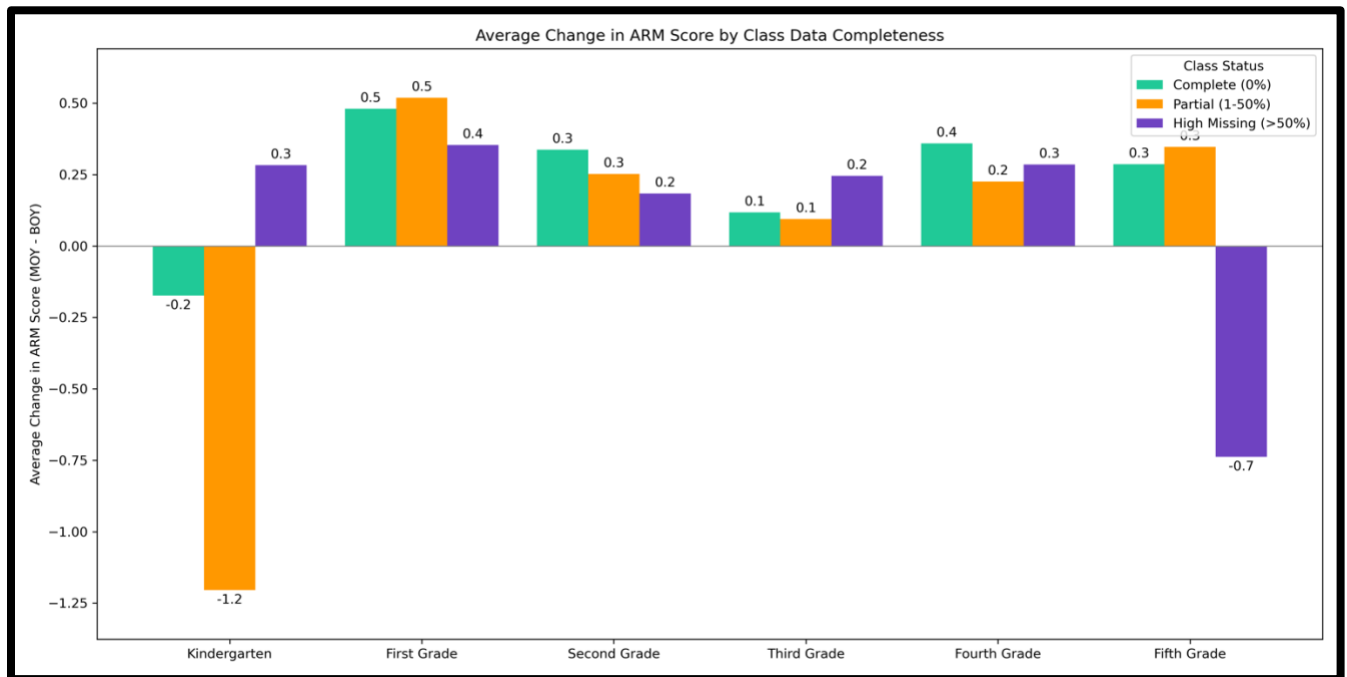
MOY Score						
	Group 1	Group 2	Statistic	P-Value	Significant (p<0.05)	Message
	Complete (0%)	High Missing (>50%)	39.010	0.0000		*
	Complete (0%)	Partial (1–50%)	25.550	0.0000		*
	High Missing (>50%)	Partial (1–50%)	-23.898	0.0000		*

Findings

- **Missing component test data is strongly associated with lower ARM scores.** Students in classes with incomplete diagnostics performed significantly worse on ARM BOY and MOY assessments.
- **Classes with complete data outperformed both partial and high-missing classes.** Differences were statistically significant ($p < 0.0001$) at both BOY and MOY.

- **Partial missing data negatively impacted ARM scores** but still performed better than high-missing classes.
- **BOY to MOY ARM growth was less sensitive to missing diagnostics.** Growth rates between groups were comparable, suggesting missing data impacts starting levels more than growth potential.

ARM Growth By Class Data Completeness



Next, I investigated BOY-to-MOY ARM score changes by grade and class data completeness. Findings include:

- Overall, ARM growth from BOY to MOY was positive across most grades and completeness groups.
- **Classes with complete data generally showed positive growth**, except in Kindergarten.
- **Classes with partial data** had mixed results; in Kindergarten, ARM scores **declined sharply**.
- **Classes with high levels of missing data** showed inconsistent performance, with negative ARM growth in Kindergarten and Fifth Grade.

TL;DR/Next Steps

- Classes with complete test data demonstrated stronger ARM growth across most grades compared to classes with partial or missing data.
- Kindergarten and Fifth Grade showed risk of ARM score declines for classes with partial and high-missing data.
- Negative ARM growth was most pronounced in classes with incomplete diagnostics.
- The primary risk linked to missing test data is lower achievement growth rates, not just lower starting points.
- Suggest data audits to identify and address missing test data.
- Investigate factors contributing to Kindergarten and Fifth Grade discrepancies.