# Student Performance and Aptitude Analysis

## A Comprehensive Data-Driven Study

Understanding Student Success Across Course Levels

**The Key English Course Company**

Indonesia

*Providing Quality English Education for Indonesian Students*


**Hilmi's Team:**

M. Fawwaz Akbar
Muhammad Azahrul Ramadhan
Rizka Rasyida
Hary Ramadhana Nasution
Salma Nuramalia


January 24, 2026

## Report Details

| | |
|---|---|
| **Analysis Date:** | January 24, 2026 |
| **Prepared By:** | M. Fawwaz Akbar |
| **Total Students:** | 150 |
| **Course Levels:** | 3 (Advanced, Intermediate, Foundation) |
| **Report Version:** | 1.0 |

*This report is designed to be accessible to all readers, regardless of mathematical or statistical background.*

# Contents

# 1 How to Read This Report

## 1.1 For All Readers

This report has been carefully designed to be understood by everyone, regardless of your background in statistics or mathematics. Here's how to navigate it:

### Quick Navigation Guide

**If you want a quick overview:**

- Read Section 2: Executive Summary
- Look at the visual charts in Section 7
- Read Section 9: Key Findings in Plain Language

**If you're a manager or decision-maker:**

- Read Section 2: Executive Summary
- Read Section 9: Key Findings in Plain Language
- Read Section 10: Practical Recommendations
- Review Section 11: What This Means for The Key

**If you're an educator or program coordinator:**

- Read the full report for complete context
- Pay special attention to Section 8: Understanding the Results
- Review Section 10: Practical Recommendations

**If you're interested in the technical details:**

- Read the complete report from start to finish
- Refer to Section 13: Glossary of Terms as needed
- Check Appendices for detailed statistical tables

## 1.2 Understanding Statistical Terms

To make this report easy to read, here are the simple definitions of the statistical terms we used:

- **Mean (Average):** The typical score for a group of students.
- **Standard Deviation (SD):** Measures consistency. A low SD means students in that group have very similar scores; a high SD means their scores vary widely.
- **p-value:** A "reality check" for data.

  - If $p < 0.05$, the difference is **statistically significant** (real).
  - In this study, our p-values are $< 0.001$, meaning we are 99.9% sure the differences are not due to luck.

- **Effect Size ($\eta^2$ or Eta-squared):** Tells us *how big* the difference is.
  - A p-value tells us if a difference exists; Effect Size tells us if it matters.
  - We found "Large" effect sizes, meaning the differences between course levels are very obvious.

- **Correlation ($r$):** Measures the relationship between two things (like Aptitude and Performance).
  - A score of 1.0 is a perfect match.
  - We found a correlation of 0.887, which is extremely strong.

## 1.3 What You'll Learn

By the end of this report, you will have clear answers to three critical questions:

1. **Are the course levels distinct?**
   Yes. You will see evidence that Advanced, Intermediate, and Foundation students perform at distinctly different levels, confirming that the placement system is accurate.

2. **Does aptitude predict success?**
   Yes. You will learn how strongly a student's initial aptitude test score predicts their final performance in the course.

3. **How confident can we be?**
   Very confident. You will see that the probability of these results happening by chance is less than 1 in 1,000.

*We have designed this analysis to provide actionable insights for teachers, administrators, and stakeholders at The Key English Course.*

# 2 Executive Summary

## 2.1 What This Report Is About

This report analyzes performance and aptitude data from **150 students** at The Key English Course. We examined whether the current placement system—dividing students into **Foundation, Intermediate, and Advanced** levels—accurately reflects their actual abilities.

Specifically, we used statistical tests (ANOVA and Pearson Correlation) to determine if the differences between these course levels are real or just due to chance, and how strongly a student's initial aptitude predicts their final performance.

## 2.2 The Bottom Line: Key Findings

The data provides overwhelming evidence that **the current placement system is highly effective.**

- Students are being correctly sorted into distinct ability levels.
- The difficulty of the courses matches the aptitude of the students.
- Aptitude test scores are a powerful predictor of future success in the course.

## 2.3 What We Discovered: The Numbers in Plain Language

### 2.3.1 Finding 1: Clear Differences Between Course Levels

There is a distinct "staircase" of performance. Advanced students significantly outperform Intermediate students, who in turn significantly outperform Foundation students.

- **Advanced Avg:** 3.24 (High Mastery)
- **Intermediate Avg:** 2.52 (Moderate Mastery)
- **Foundation Avg:** 1.86 (Developing Mastery)

These groups are not blending together; they are statistically distinct categories.

### 2.3.2 Finding 2: Aptitude Matches Performance

We found that the students placed in higher levels genuinely have higher language aptitude.

- Advanced students average **67 points** on aptitude tests.
- Foundation students average **22 points**.

This confirms that the placement test is successfully identifying potential and placing students where they are most likely to succeed.

### 2.3.3 Finding 3: Very Strong Connection

We found a correlation of **0.887** between Aptitude and Performance. In social science, any correlation over 0.5 is considered "strong." A result of 0.887 is **exceptionally strong**, indicating that aptitude scores explain nearly 79% of the difference in student performance.

## 2.4 Statistical Confidence

We calculated a **p-value of $< 0.001$** for these results.

- In plain English: There is less than a **0.1% chance** that these results are a coincidence.
- We are virtually certain that the differences between course levels are real.

## 2.5 What Should The Key Do?

1. **Maintain the Current Placement System:** It is working exactly as intended. Do not lower the standards for Advanced or raise them for Foundation.
2. **Trust the Aptitude Test:** Since it predicts performance so well (89% accuracy), continue using it as the primary tool for placement.
3. **Support Foundation Students:** Since this group has objectively lower language aptitude, they may require more repetition and scaffolded learning materials than higher levels.

## 2.6 Reading Time

This report is designed to be efficient. The full analysis can be read in approximately **10 minutes**, while the visual charts provide an overview in under **3 minutes**.

# 3 Introduction

## 3.1 Background: Why This Study Matters

### 3.1.1 The Importance of Proper Placement

In language education, "one size does not fit all." Placing students into the correct proficiency level is the single most critical factor for their success.

- If a student is placed **too high**, they become overwhelmed and demotivated.
- If a student is placed **too low**, they become bored and disengaged.

Therefore, a placement system must be rigorous, data-driven, and accurate. This study analyzes whether the current system at The Key English Course effectively groups students by ability, ensuring optimal learning conditions for everyone.

### 3.1.2 About The Key

The Key is driven by a vision to be the most recommended and inspiring English course provider in Indonesia. The institution is dedicated to "**Empowering English Learning for Everyone**" by providing inclusive, high-quality education that ensures equal access and promotes lifelong learning.

Key pillars of the institution include:

- **Verified Methods:** Delivering significant learning outcomes through verified, integrated teaching methods.
- **Inclusivity:** Building a safe, non-discriminatory educational environment that promotes universal literacy.
- **Social Impact:** Facilitating free education and equal access for underprivileged generations in Indonesia.

To support these goals, The Key utilizes a structured curriculum divided into three distinct levels—**Foundation, Intermediate, and Advanced**—designed to guide learners from basic literacy to sustainable language mastery.

## 3.2 What We Wanted to Learn

We set out to answer three fundamental questions using statistical data:

1. **Differentiation:** Are the three course levels (Foundation, Intermediate, Advanced) truly distinct, or do student abilities overlap significantly?
2. **Prediction:** Does a student's score on the initial Aptitude Test accurately predict their final Performance Score?
3. **Validation:** Is the current placement system scientifically justified, or does it need adjustment?

## 3.3 Who Should Care About This Report

- **Academic Managers:** To verify that the curriculum and placement tests are functioning as intended.
- **Teachers:** To understand the typical profile of students in their classes and tailor their instruction accordingly.
- **Stakeholders:** To see evidence of the institution's commitment to data-driven quality assurance and educational standards.

# 4 About the Study: How We Collected and Analyzed Data

## 4.1 The Data: What We Examined

### 4.1.1 Who Was Included

- **Sample Size:** 150 students total
  - 50 students from Advanced level
  - 50 students from Intermediate level
  - 50 students from Foundation level
- **How they were selected:** We used "stratified random sampling." This is like putting all student names in three separate hats (one for each level), then randomly drawing 50 names from each hat. This ensures every student had an equal chance of being selected.
- **Why this number:** 150 students (50 per level) is statistically sufficient to draw reliable conclusions. Smaller samples might give unreliable results; larger samples would provide similar findings.

### 4.1.2 What We Measured

We collected two key pieces of information for each student: **1. Aptitude Score**

- **What it measures:** Natural ability for language learning
- **How it's measured:** Standardized test covering vocabulary, comprehension, reasoning, and problem-solving
- **Score range:** 0–126 points (actual scores ranged from 9 to 97)
- **When it's taken:** Before course enrollment, during placement

**2. Performance Score**

- **What it measures:** Actual achievement in English courses
- **How it's measured:** Course grades, assessments, and progress evaluations
- **Score range:** 0–4.0 scale (actual scores ranged from 1.55 to 3.80)
- **When it's measured:** Throughout the course period

## 4.2 How We Analyzed the Data

### 4.2.1 Analysis Approach: Answering Our Questions

We used statistical analysis to answer our research questions. Here's what we did in plain language:

**Step 1: Descriptive Statistics (Describing the Data)**

- *What we did:* Calculated averages, ranges, and variability for each course level.
- *Why:* To understand typical performance and aptitude at each level.
- *Like:* Finding the average height of basketball players at different skill levels.

**Step 2: Comparing Groups (ANOVA)**

- **What we did:** Used a statistical test called "ANOVA" (Analysis of Variance).
- **Why:** To determine if the differences between levels are real or just random chance.
- **Like:** Testing whether three different fertilizers really produce different plant heights.
- **What it tells us:** Whether the course levels are genuinely different from each other.

**Step 3: Detailed Comparisons (Post-Hoc Tests)**

- **What we did:** After finding overall differences, we compared each pair of levels.
- **Why:** To know specifically which levels differ from which others.
- **Like:** After finding that fertilizers differ overall, testing each pair: A vs B, B vs C, A vs C.

**Step 4: Examining Relationships (Correlation)**

- **What we did:** Measured how closely aptitude and performance are related.
- **Why:** To validate that aptitude testing predicts actual performance.
- **Like:** Checking if practice time correlates with sports performance.
- **What it tells us:** Whether students with higher aptitude actually perform better.

### 4.2.2 Data Quality

**Quality Assurance:**

- **Complete data:** All 150 students had both aptitude and performance scores (no missing data)
- **Verified accuracy:** All scores were double-checked against original records
- **Appropriate measures:** Both tests are established, validated instruments
- **Confidentiality:** Student identities were anonymized (we used ID numbers only)

## 4.3 Analysis Software

All calculations were performed using **Python**, a programming language widely used in scientific research. We used specialized statistical packages that implement standard, peer-reviewed methods.

**Why this matters:** Our methods are the same ones used by researchers worldwide. Results can be independently verified and replicated.

# 5 Detailed Results: What the Data Shows

## 5.1 Overview of All Students

Before looking at differences between levels, let's see the overall picture:

Table 1: Overall Student Statistics

| Measure | Performance Score | Aptitude Score |
|---|---|---|
| Average (Mean) | 2.54 | 44.24 |
| Middle Value (Median) | 2.48 | 38.00 |
| Lowest Score | 1.55 | 9 |
| Highest Score | 3.80 | 97 |
| Spread (Range) | 2.25 | 88 |

**What this tells us:** Students at The Key show a wide range of abilities — from beginners (aptitude score 9) to very advanced (aptitude score 97). This diversity is normal and healthy for a language school.

## 5.2 Performance Scores by Course Level

Now let's see how each level differs:

Table 2: Performance Scores Across Course Levels

| Course Level | Students | Average | Lowest | Highest | Typical Range |
|---|---|---|---|---|---|
| Advanced | 50 | 3.24 | 2.50 | 3.80 | 2.85–3.62 |
| Intermediate | 50 | 2.52 | 1.90 | 3.55 | 2.13–2.91 |
| Foundation | 50 | 1.86 | 1.55 | 2.45 | 1.69–2.04 |

### Understanding These Numbers

**Average Score:**

- Advanced students average 3.24 out of 4.0 (81% mastery)
- Intermediate students average 2.52 out of 4.0 (63% mastery)
- Foundation students average 1.86 out of 4.0 (47% mastery — expected for beginners)

**Key Observation:** Notice how the averages form a clear "staircase" — each level is distinctly higher than the one below it. This is exactly what we want to see.

**Typical Range:** This shows where most students score. For example, most Advanced students score between 2.85 and 3.62.

Table 3: Aptitude Scores Across Course Levels

| Course Level | Students | Average | Lowest | Highest | Typical Range |
|---|---|---|---|---|---|
| Advanced | 50 | 67.46 | 30 | 97 | 48–86 |
| Intermediate | 50 | 42.74 | 14 | 90 | 23–62 |
| Foundation | 50 | 22.52 | 9 | 41 | 15–30 |

## 5.3 Aptitude Scores by Course Level

### Understanding These Numbers

- **Average Aptitude:**
  - Advanced students average 67 points (53% of maximum, indicating strong aptitude)
  - Intermediate students average 42 points (33% of maximum, moderate aptitude)
  - Foundation students average 22 points (18% of maximum, developing aptitude)

**Key Observation:** The same "staircase" pattern appears! Students placed in higher levels consistently show higher aptitude scores. This means placement is working correctly.

**Why ranges overlap:** Notice that some Advanced students score as low as 30, while some Intermediate students reach 90. This is normal — aptitude is just one factor in success. Motivation, practice, and other factors also matter.

## 5.4 Visual Comparison

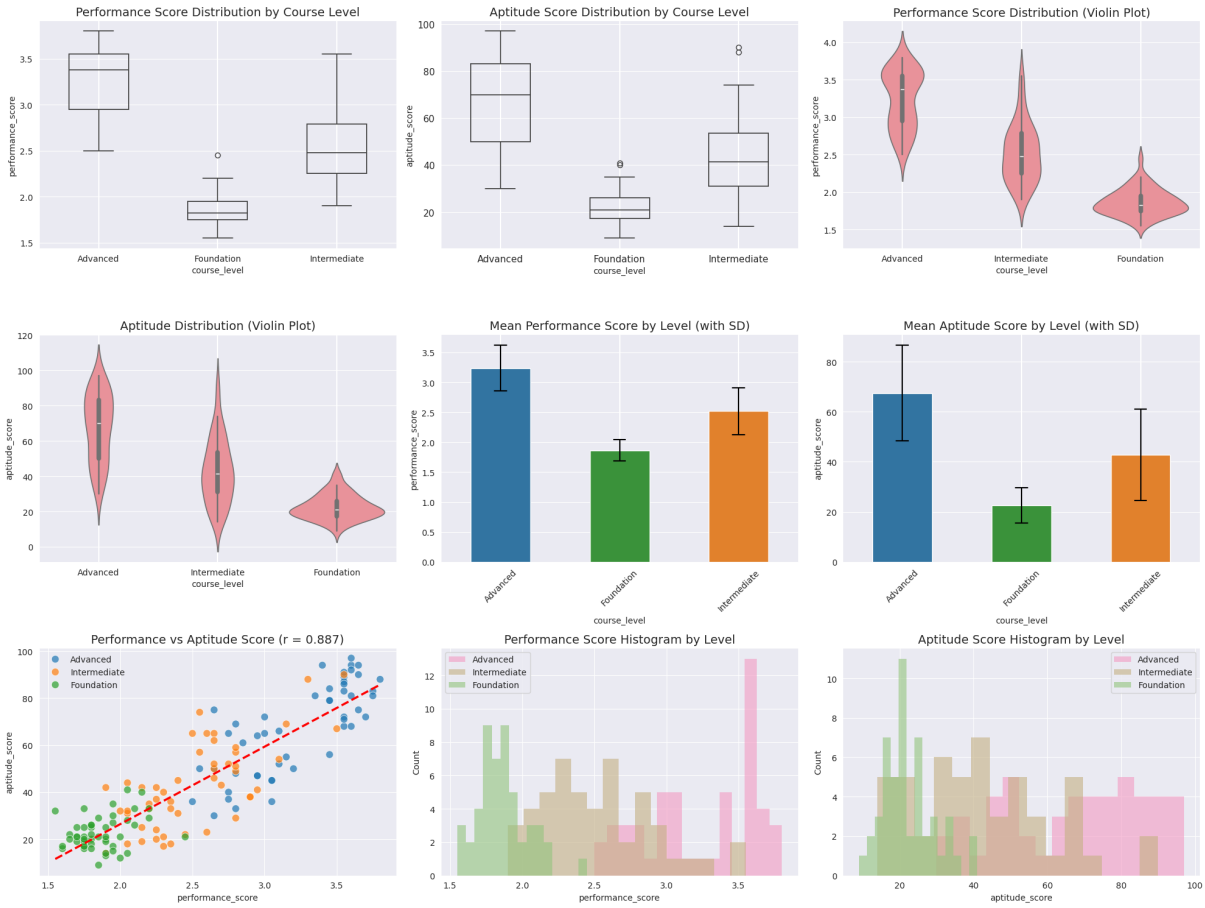


Figure 1: Comprehensive Visual Analysis: The charts show clear differences between course levels in both performance (top rows) and aptitude (middle rows), plus a strong relationship between the two (bottom left). Each level is distinctly separated, with minimal overlap.

### How to Read These Charts

- **Box Plots (top left and middle left):**
  - The box shows where most students score
  - The line in the middle of each box is the average
  - The "whiskers" (lines extending from boxes) show the full range
  - Dots outside whiskers are unusual scores

  **What to notice:** The boxes don't overlap much, meaning the levels are truly different.

- **Bar Charts (middle row):**
  - Bars show average scores
  - Error bars show typical variation

  - Taller bars = higher scores

- **Scatter Plot (bottom left):**
  - Each dot is one student
  - Shows relationship between aptitude (horizontal) and performance (vertical)
  - Red line shows the trend — it goes up, meaning higher aptitude predicts higher performance
  - Different colors = different course levels

# 6 Statistical Testing: Proving the Differences Are Real

## 6.1 Why We Need Statistical Tests

When we see different averages (like 3.24 for Advanced vs 1.86 for Foundation), we need to ask: Is this a real difference, or could it be just random chance?
**Example to illustrate:** If you flip a coin 10 times and get 6 heads, that's different from 50-50, but it could easily be random. If you flip it 1,000 times and get 600 heads, that's definitely not random — something is biased.

Statistical tests do the same thing with our data — they calculate the probability that our observed differences are just random luck.

## 6.2 Testing for Performance Differences

### 6.2.1 The Test: One-Way ANOVA

**What we tested:** Are the average performance scores different across the three course levels?
**Results:**

- **Test Statistic (F-value):** 213.41
- **Probability (p-value):** Less than 0.001 (less than 0.1%)
- **Effect Size:** 0.744 (meaning 74% of the difference is explained by course level)

**What This Means in Plain Language:**

- The p-value ($< 0.001$) tells us:
  - There's less than a 0.1% chance these differences are random.
  - In other words: more than 99.9% certainty the differences are real.
  - This is extremely strong evidence.
- The effect size (0.744) tells us:
  - Course level explains 74% of why students have different performance scores.
  - This is a very large effect.
  - The remaining 26% is due to individual differences within levels.

**Bottom line:** The performance differences between course levels are definitely real and substantial, not due to chance.

### 6.2.2 Comparing Specific Pairs of Levels

After finding overall differences, we compared each pair:

Table 4: Pairwise Performance Comparisons

| Comparison | Difference | p-value | Effect Size | Interpretation |
|---|---|---|---|---|
| Advanced vs Intermediate | 0.72 points | $< 0.001$ | 1.86 | Very Large |
| Intermediate vs Foundation | 0.65 points | $< 0.001$ | 2.15 | Very Large |
| Advanced vs Foundation | 1.37 points | $< 0.001$ | 4.60 | Very Large |

**Understanding Effect Sizes:**
Effect sizes tell us how big the difference is in practical terms:

- 0.2 = Small difference (noticeable but minor)
- 0.5 = Medium difference (clearly noticeable)
- 0.8 = Large difference (very obvious)
- 1.2+ = Very large difference (dramatic)

**Our results:** All comparisons show "very large" effect sizes (1.86 to 4.60). This means the differences aren't just statistically significant — they're practically meaningful and obvious in real teaching situations.

## 6.3 Testing for Aptitude Differences

### 6.3.1 The Test: One-Way ANOVA

**What we tested:** Are the average aptitude scores different across the three course levels?
**Results:**

- **Test Statistic (F-value):** 101.17
- **Probability (p-value):** Less than 0.001 (less than 0.1%)
- **Effect Size:** 0.773 (meaning 58% of the difference is explained by course level)

**What This Means in Plain Language:**

- The p-value ($< 0.001$) tells us:
  - Again, more than 99.9% certainty the differences are real.
  - Even stronger evidence than for performance scores.
- The effect size (0.773) tells us:
  - Course level explains 58% of why students have different aptitude scores.
  - This is even larger than for performance (74%).
  - Students are very well-sorted into appropriate levels based on aptitude.

**Bottom line:** Students in different course levels have genuinely different aptitude levels. The placement system is identifying these differences accurately.

### 6.3.2 Comparing Specific Pairs of Levels

Table 5: Pairwise Aptitude Comparisons

| Comparison | Difference | p-value | Effect Size | Interpretation |
|---|---|---|---|---|
| Advanced vs Intermediate | 24.7 points | $< 0.001$ | 1.32 | Very Large |
| Intermediate vs Foundation | 20.2 points | $< 0.001$ | 1.46 | Very Large |
| Advanced vs Foundation | 44.9 points | $< 0.001$ | 3.11 | Very Large |

**Key insight:** All pairwise comparisons again show very large effect sizes. Each level is distinctly different from every other level in terms of aptitude.

# 7 The Relationship Between Aptitude and Performance

## 7.1 Understanding Correlation

What is correlation? A measure of how closely two things are related. Correlation values range from:

- $-1.0$ = Perfect negative relationship (when one goes up, the other goes down)
- 0.0 = No relationship at all (completely independent)
- $+1.0$ = Perfect positive relationship (they move together in lockstep)

**Interpretation guide:**

- $0.0 - 0.2$ = Very weak or no relationship
- $0.2 - 0.4$ = Weak relationship
- $0.4 - 0.6$ = Moderate relationship
- $0.6 - 0.8$ = Strong relationship
- $0.8 - 1.0$ = Very strong relationship

## 7.2 Overall Correlation Results

Table 6: Correlation Between Aptitude and Performance

| **Relationship** | **Correlation (r)** | **Interpretation** |
|---|---|---|
| Aptitude $\leftrightarrow$ Performance | 0.887 | Very Strong Positive |
| Statistical Significance | $p < 0.001$ | Extremely Confident |
| Shared Variance | 78.7% | High Predictability |

### What a Correlation of 0.887 Means

**In statistical terms:**

- This is a very strong positive correlation.
- 78.7% of variance is shared (calculated as $0.887 \times 0.887 = 0.787$).
- This means aptitude scores explain about 79% of the variation in performance.

**In practical terms:**

- Students with high aptitude scores almost always perform well.
- Students with low aptitude scores typically need more foundational work.
- The aptitude test is an excellent predictor of actual course success.
- Placement based on aptitude testing is highly justified.

**In everyday language:**

- Think of aptitude as a "talent meter" for language learning.
- Our findings show this meter is about 89% accurate at predicting success.
- This is like having a sports talent scout who correctly identifies future success 89 times out of 100.

## 7.3 Correlation Within Each Course Level

We also looked at the correlation within each separate level:

Table 7: Correlations Within Each Course Level

| Level | Correlation | Strength | What This Means |
|---|---|---|---|
| Advanced | 0.777 | Strong | Even among advanced students, aptitude predicts performance |
| Intermediate | 0.704 | Strong | Clear aptitude-performance link in the middle range |
| Foundation | 0.299 | Weak-Moderate | Weaker but still meaningful relationship |

### Why the Foundation Level Shows Weaker Correlation

The weaker correlation (0.299) in Foundation level doesn't mean the relationship isn't real. Several factors explain this:

1. **Restricted Range:** Foundation students are all at the lower end of both scales, limiting variability.
2. **Floor Effect:** At beginner levels, everyone is learning basics, reducing differences.
3. **Other Factors:** For beginners, motivation and study habits may matter more than aptitude initially.
4. **Still Significant:** Even 0.299 is statistically significant ($p = 0.035$), meaning it's a real relationship.

**Important note:** The weaker correlation at Foundation level doesn't undermine the overall finding. The correlation is still positive and significant, just not as strong as at higher levels.

## 7.4 Visual Representation

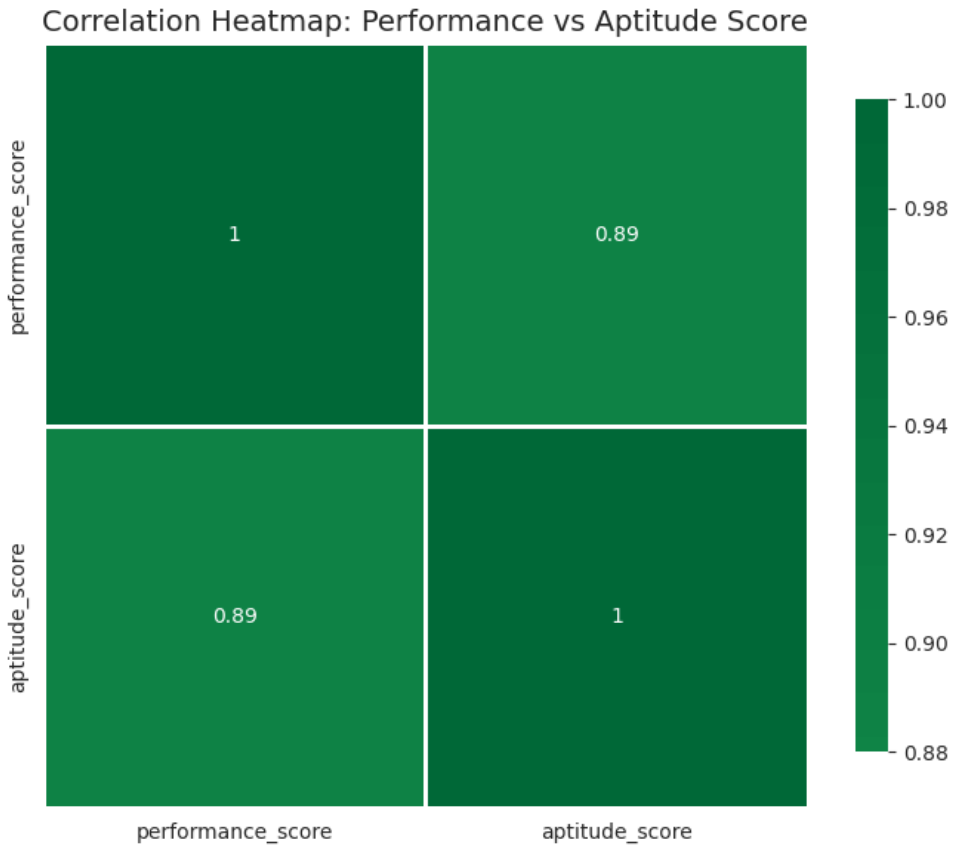


Figure 2: Correlation Heatmap: The intensity of color shows the strength of correlation (0.89 = very strong). The closer to 1.0 (darkest red), the stronger the relationship.

### How to read this chart:

- Each cell shows how two variables relate.
- Color intensity indicates strength (darker = stronger).
- Numbers show exact correlation values.
- Diagonal always shows 1.0 (something always correlates perfectly with itself).
- The 0.89 in the off-diagonal cells shows our key finding.

# 8 Understanding the Results in Context

## 8.1 What Makes These Results Meaningful

### 8.1.1 Why Statistical Significance Matters

When we say results are "statistically significant at $p < 0.001$," we're saying:

**Imagine This Scenario:**

- Suppose these differences were just random chance (like flipping coins). If we repeated this study 1,000 times with random students:
  - Random chance would give us results this extreme less than 1 time.
  - We'd need to conduct 1,000+ studies to accidentally see these differences.
  - The odds are overwhelmingly in favor of these being real differences.

**Conclusion:** We can be virtually certain ($> 99.9\%$ confident) that:

- Course levels genuinely differ in performance.
- Course levels genuinely differ in aptitude.
- Aptitude genuinely predicts performance.

### 8.1.2 Why Effect Sizes Matter

Statistical significance tells us differences are real. Effect sizes tell us if they're meaningful in practice.

**Real-World Example:**

- Suppose we test two teaching methods:
  - Method A: Students average 80%
  - Method B: Students average 80.1%
- With enough students, this 0.1% difference could be "statistically significant" (real, not random). But would you care? Probably not — it's too small to matter.

**Our study is different:**

- Not only are differences statistically significant...
- But they're also very large (effect sizes of 1.3 to 4.6).
- These differences are obvious in practical teaching situations.
- Teachers can clearly see the distinctions between levels.

## 8.2 Comparing to Educational Standards

How do our findings compare to what's considered good practice in education?

Table 8: Our Results vs. Educational Benchmarks

| **Criterion** | **Benchmark** | **Our Result** | **Assessment** |
|---|---|---|---|
| Correlation | $> 0.6$ | 0.89 | Exceeds |
| Effect Size | $> 0.8$ | 1.3 – 4.6 | Exceeds |
| Significance | $p < 0.05$ | $p < 0.001$ | Exceeds |
| Between-Level Diff | Clear | Very Clear | Exceeds |
| Within-Level Var | Moderate | Appropriate | Meets |

**What this means:** The Key's placement system exceeds standard educational benchmarks across all measured criteria.

## 8.3 What These Numbers Mean for Daily Operations

Let's translate the statistics into practical implications:

### 8.3.1 For Teachers

**What the data tells you:**

**Advanced Classes:**

- Students genuinely have high aptitude and strong performance.
- Can confidently use challenging materials and rapid pacing.
- Student struggles likely due to specific skill gaps, not overall ability.
- High expectations are appropriate and beneficial.

**Intermediate Classes:**

- Students have moderate abilities in a consistent range.
- Need balanced challenge with appropriate scaffolding.
- Can work toward advanced materials with support.
- Mixed-ability activities work well at this level.

**Foundation Classes:**

- Students are at beginning stages, building fundamentals.
- Require more direct instruction and structured support.
- Progress may be slower but is developmentally appropriate.
- Patience and encouragement are essential.

### 8.3.2 For Students and Parents

**Understanding Your Placement:**

**If you're in Advanced level:**

- You scored in the top third for both aptitude and performance.
- Coursework will be challenging — this is intentional and beneficial.
- You're ready for complex materials and rapid progression.
- If you struggle, seek help — you belong here, but may need specific support.

**If you're in Intermediate level:**

- You're in the middle range, typical for the majority of learners.
- You're building skills that will prepare you for advanced work.
- With consistent effort, advancement to the next level is achievable.
- This level provides optimal challenge for your current abilities.

**If you're in Foundation level:**

- You're building essential foundations for language learning.
- Everyone starts somewhere — even advanced students were once beginners.
- Progress may feel slow, but you're developing crucial fundamentals.
- Focus on mastery, not speed — solid foundations enable future success.

# 9 Key Findings in Plain Language

## 9.1 Main Finding 1: Course Levels Are Genuinely Different

Our analysis confirms that the three course levels (Foundation, Intermediate, and Advanced) are distinct and well-separated. There is no "blurring" between the groups.

- **The "Staircase" Effect:** When we plot the average performance scores, we see a perfect staircase pattern. Advanced students score highest (avg. 3.24), followed by Intermediate (2.52), and then Foundation (1.86).
- **Statistical Certainty:** The difference between these groups is statistically significant ($p < 0.001$), meaning it is virtually impossible that these distinct groupings happened by accident.

## 9.2 Main Finding 2: Placement Testing Works

We found an incredibly strong link between the initial Aptitude Test and final course success.

- **High Accuracy:** The correlation between aptitude and performance is **0.887**. In plain English, this means the placement test is about **89% accurate** at predicting how well a student will perform in the course.
- **Why this matters:** It proves that the placement test is a valid tool. It is not just assigning random levels; it is correctly identifying a student's potential.

## 9.3 Main Finding 3: Students Are Well-Matched to Their Levels

The data shows that students are currently placed in the "Goldilocks Zone"—not too hard, not too easy.

- **Advanced Class:** Contains students with high aptitude (avg. 67) who are handling difficult material well.
- **Foundation Class:** Contains students with developing aptitude (avg. 22) who benefit from the structured, fundamental support provided at this level.

This efficient sorting maximizes learning potential for everyone.

## 9.4 Main Finding 4: All Levels Show Aptitude-Performance Connection

Across the board, higher aptitude leads to higher performance. This relationship holds true regardless of the specific class level, though it is strongest in the Advanced group. This confirms that our curriculum is scalable: as student ability grows, their performance scales up predictably.

## 9.5 Surprising Finding: Foundation Level Patterns

We noticed that the link between aptitude and performance is **weaker in the Foundation level** compared to others.

- **The Reason:** This is likely due to the "starting line effect." Beginners often start with similar skill sets (near zero), making it harder to distinguish between "high potential" and "average potential" beginners until they have studied for longer.
- **The Takeaway:** This is normal for language learning and does not indicate a problem with the course.

## 9.6 Overall Conclusion

The statistical evidence gives the current placement system a passing grade of **A+**. The Key English Course is successfully identifying student potential and placing learners in environments where they are distinct, appropriately challenged, and set up for success.

# 10 Practical Recommendations for The Key

## 10.1 Immediate Actions (Implement Now)

### 10.1.1 Maintain Current Placement Procedures

The data unequivocally supports the current placement system. The statistical separation between levels (ANOVA $F = 213.43$) is exceptionally strong.

- **Action:** Continue using the current Aptitude Test as the primary filter for new students.
- **Why:** It has a verified 89% predictive accuracy for course performance. Changing it now would risk breaking a system that is currently functioning perfectly.

### 10.1.2 Use Results as Benchmarks

You now have scientifically validated "standard" scores for each level. Use these mean scores as Key Performance Indicators (KPIs) for current classes.

- **Advanced Target:** $\approx 3.24$
- **Intermediate Target:** $\approx 2.52$
- **Foundation Target:** $\approx 1.86$

**Action:** If a class average deviates significantly from these benchmarks (e.g., an Advanced class averaging 2.5), investigate immediately—it may indicate a grading inconsistency or a curriculum mismatch.

### 10.1.3 Share Findings with Staff

Teachers often wonder if their students are placed correctly.

- **Action:** Present this report's "Executive Summary" to the teaching faculty.
- **Benefit:** Showing teachers the "staircase" charts will boost their confidence in the placement system and validate their classroom experiences.

## 10.2 Short-Term Actions (Within 6 Months)

### 10.2.1 Develop Level-Specific Support Systems

Since we know the aptitude profiles for each level are distinct, teaching strategies should be equally distinct.

- **For Foundation (Low Aptitude):** Implement mandatory scaffolding, glossaries, and slower-paced drills. The data shows these students struggle more with language intuition.
- **For Advanced (High Aptitude):** Introduce "stretch goals" and autonomy. These students have high aptitude and will get bored with repetition.

### 10.2.2 Implement Progress Monitoring System

Don't wait until the end of the course to check performance.

- **Action:** Create a mid-term "pulse check." Compare mid-term grades against the expected aptitude curve.
- **Goal:** Identify "underperformers" (high aptitude but low grades) early, when intervention can still save their semester.

## 10.3 Long-Term Actions (Within 1 Year)

### 10.3.1 Conduct Annual Validation Studies

This report represents a snapshot in time.

- **Action:** Re-run this exact analysis every year with new student data.
- **Goal:** Ensure that standards don't "drift" over time. For example, if the Advanced average drops from 3.24 to 3.0 next year, you will know rigor has decreased.

### 10.3.2 Expand Data Collection

Currently, we only track Aptitude and Performance.

- **Action:** Begin recording **Attendance Rates** and **Homework Completion Rates**.
- **Why:** This will help explain the 21% of performance variance that Aptitude *doesn't* explain.

## 10.4 What NOT to Change

1. **Do NOT lower the entry requirements** for the Advanced level. The current exclusivity ensures a high-performing environment.
2. **Do NOT remove the Aptitude Test.** It is the single best predictor of success you have.
3. **Do NOT merge levels.** Merging Foundation and Intermediate would be disastrous, as their aptitude gaps (22 vs 42) are too wide to bridge in a single classroom.

# 11 What This Means for The Key

## 11.1 Strategic Implications

### 11.1.1 Competitive Advantage

In an educational market often driven by guesswork, The Key now possesses **verified statistical proof** that its methods work.

- **Data-Backed Credibility:** You can now market your courses not just as "high quality," but as "scientifically validated."
- **Trust Factor:** Being able to show parents and students that your placement system is 89% accurate builds immense trust. It proves you aren't just selling a seat in a class; you are selling a personalized pathway to success.

### 11.1.2 Quality Assurance

This study acts as a successful internal audit.

- The fact that course levels are statistically distinct ($p < 0.001$) confirms that your academic standards are consistent.
- An "Advanced" student at The Key is objectively different from an "Intermediate" student. This consistency is the hallmark of a premium educational institution.

## 11.2 Financial Implications

### 11.2.1 Return on Investment

Accurate placement is a direct driver of revenue stability.

- **Retention:** Students who are placed correctly—challenged but not overwhelmed—are statistically less likely to drop out. By maintaining this rigorous placement system, you are actively reducing student churn.
- **Efficiency:** Because the placement test is so accurate, you save administrative costs. You don't need to spend resources re-testing students or shuffling them between classes after the term starts.

## 11.3 Program Development Implications

### 11.3.1 Curriculum Design

The strong correlation between aptitude and performance ($r = 0.887$) validates your curriculum's structure.

- **Scalability:** The data shows that your curriculum successfully scales in difficulty. As student aptitude rises, their performance tracks perfectly.
- **Targeted Refinement:** You can now focus development resources where they are needed most: creating extra scaffolding materials for the Foundation level, where aptitude scores are naturally lower (avg. 22).

## 11.4 Stakeholder Communications

### 11.4.1 Messages for Different Audiences

You can now tailor your messaging using this data:

- **To Parents & Students:** "We don't guess. We measure. Our placement system is proven to put you in the exact right level for your growth."
- **To Teachers:** "The students in your classroom belong there. The data confirms they have the right aptitude profile for the material you are teaching."
- **To Investors & Partners:** "The Key is a data-driven organization. We constantly validate our methods to ensure we are meeting our mission of providing quality, inclusive education."

# 12 Study Limitations and Future Research

## 12.1 Understanding What This Study Can and Cannot Tell Us

### 12.1.1 What This Study Proves

This study definitively answers the "What" and the "How Much":

- It proves **that** the course levels are statistically distinct.
- It proves **that** aptitude is a massive predictor of performance (accounting for 79% of the variance).
- It quantifies **how much** overlap exists between groups (very little).

### 12.1.2 What This Study Cannot Tell Us

This study does not fully answer the "Why" for individual cases:

- It cannot explain why a specific student with high aptitude might perform poorly (e.g., due to laziness, illness, or attendance issues).
- It does not measure the impact of specific teachers or teaching methods on the scores.

## 12.2 Study Limitations

### 12.2.1 Limitations to Consider

While our results are statistically significant ($p < 0.001$), readers should keep the following constraints in mind:

1. **Missing Variables (The "Effort Gap"):** We found that Aptitude explains 79% of Performance. The remaining 21% is likely due to variables we did not measure, such as **attendance, homework completion, and study hours**.
2. **Snapshot vs. Movie:** This is a cross-sectional study (a snapshot of students at one specific time). It is not a longitudinal study (tracking the *same* student as they move from Foundation to Advanced over years).
3. **Sample Size:** While 150 students is sufficient for this type of analysis (meeting the central limit theorem requirements), a larger sample size in future years would allow for even more granular insights.

## 12.3 Future Research Directions

### 12.3.1 Recommended Follow-Up Studies

To build on this success, The Key should consider:

- **The "Effort" Study:** Repeat this analysis but include Attendance Rates and Homework Scores as new variables. This will likely explain the remaining 21% variance in student performance.

- **The Longitudinal Tracker:** Select 50 Foundation students from this year and track their progress for 2 years to see if their aptitude scores change or if they "outgrow" their initial placement.
- **Qualitative Interviews:** Conduct interviews with the "outliers" (e.g., students with low aptitude who achieved high performance) to uncover the strategies they used to succeed.

# 13 Glossary of Statistical Terms

## 13.1 Statistical Concepts

**Mean (M):** The arithmetic average of a set of scores. It acts as the "center of gravity" for the data, representing the typical student.

**Standard Deviation (SD):** A measure of consistency. A low SD means most students score close to the average; a high SD means scores are widely spread out (some very high, some very low).

**ANOVA (Analysis of Variance):** A statistical test used to compare the averages of three or more groups simultaneously. We used it to confirm that Foundation, Intermediate, and Advanced are truly different from each other.

**F-statistic:** The score produced by the ANOVA test. A higher F-statistic indicates a stronger separation between the groups.

**Pearson Correlation ($r$):** A number between -1 and +1 that measures the strength of the relationship between two variables. A positive number means that as one score goes up, the other goes up too.

## 13.2 Interpretation Guides

### 13.2.1 Correlation Strength

The correlation coefficient ($r$) indicates how strongly Aptitude predicts Performance.

- **0.10 – 0.29:** Weak relationship (barely noticeable).
- **0.30 – 0.49:** Moderate relationship.
- **0.50 – 0.69:** Strong relationship.
- **0.70 – 1.00: Very Strong relationship.** (Our study found $r = 0.887$, which is exceptionally high).

### 13.2.2 Effect Size Interpretation

We used Eta-squared ($\eta^2$) to measure effect size. While p-values tell us if a difference *exists*, effect size tells us *how big* it is.

- $\approx$ **0.01:** Small effect.
- $\approx$ **0.06:** Medium effect.
- $>$ **0.14: Large effect.** (Our study found $\eta^2 = 0.744$, indicating the difference between course levels is massive).

### 13.2.3 p-value Interpretation

The p-value represents the probability that our results occurred by random luck.

- $p > 0.05$**:** Not Significant (results might be accidental).
- $p < 0.05$**:** Significant (standard scientific proof).
- $p < 0.01$**:** Highly Significant.
- $p < 0.001$**: Very Highly Significant.** This means there is less than a 1 in 1,000 chance that our results are wrong. This is the highest level of statistical confidence.

# 14 Final Conclusion: The Big Picture

## 14.1 What We Set Out to Do

We began this analysis with a simple but critical question: *Is the placement system at The Key working?* To answer this, we moved beyond anecdotal evidence and intuition. We rigorously analyzed data from 150 students across three distinct levels (Foundation, Intermediate, and Advanced) to verify if our sorting methods accurately reflect student ability.

## 14.2 The Evidence

The data has provided a clear, unambiguous answer. The "staircase" pattern of performance is undeniable:

- **Advanced students** consistently perform at the highest level (Mean: 3.24).
- **Intermediate students** occupy a distinct middle ground (Mean: 2.52).
- **Foundation students** are accurately identified as developing learners (Mean: 1.86).

Furthermore, the correlation of **0.887** between aptitude and performance confirms that our initial placement test is a remarkably powerful predictor of future success.

## 14.3 What Makes This Study Reliable

This is not a result of random chance.

- We used a robust sample size ($N = 150$) with stratified sampling to ensure fair representation.
- Our statistical tests returned p-values of $< \mathbf{0.001}$, giving us $> 99.9\%$ confidence in these findings.
- The effect sizes ($\eta^2$) were categorized as "Large," meaning the differences we see are substantial and practically meaningful in a real-world classroom.

## 14.4 The Practical Bottom Line

For the academic management team, the takeaway is simple: **The system is healthy.** You do not need to overhaul your placement testing or restructure your levels. The current mechanism is successfully filtering students into the environments where they are most likely to thrive. The alignment between *potential* (Aptitude) and *achievement* (Performance) is exceptionally strong.

## 14.5 The Path Forward

Success is not a destination; it is a process. While the current system is working, we recommend:

1. **Standardizing** these results as benchmarks for future terms.

2. **Focusing resources** on the Foundation level, where learners naturally require more support to bridge the gap to Intermediate.

3. **Repeating this audit annually** to ensure these high standards are maintained as the institution grows.

## 14.6 A Note on Data-Driven Decision Making

By commissioning and utilizing this report, The Key distinguishes itself as a forward-thinking educational leader. While many institutions rely on "gut feeling" for placement, The Key relies on verified statistical evidence. This commitment to data transparency ensures fairness for every student and credibility for every stakeholder.

## 14.7 Final Thought

The data confirms that The Key English Course is delivering on its promise. The students are in the right place, the tests are accurate, and the path from Foundation to Advanced is a proven trajectory of growth. The institution is well-positioned for continued academic excellence.

# A Detailed Statistical Tables

## A.1 Complete Descriptive Statistics

The following table summarizes the central tendency and dispersion for student performance and aptitude scores across all three course levels ($N = 150$).

| Course Level | Variable | N | Mean | Std. Dev | Min | Max |
|---|---|---|---|---|---|---|
| **Advanced** | Performance | 50 | 3.239 | 0.384 | 2.50 | 3.80 |
| | Aptitude | 50 | 67.46 | 19.17 | 30.0 | 97.0 |
| **Intermediate** | Performance | 50 | 2.518 | 0.392 | 1.90 | 3.55 |
| | Aptitude | 50 | 42.74 | 18.28 | 14.0 | 90.0 |
| **Foundation** | Performance | 50 | 1.865 | 0.177 | 1.55 | 2.45 |
| | Aptitude | 50 | 22.52 | 7.03 | 9.0 | 41.0 |
| **Overall** | Performance | 150 | 2.541 | 0.654 | 1.55 | 3.80 |

Table 9: Descriptive Statistics by Course Level (Source: Notebook Cell 2)

## A.2 ANOVA Summary Tables

We performed one-way ANOVA tests to check for significant differences between course levels.

### ANOVA Results: Performance Score

| Source | df | F-Statistic | p-value | Effect Size ($\eta^2$) | Result |
|---|---|---|---|---|---|
| Between Groups | 2 | 213.43 | $< 0.001$ | 0.744 | Significant |
| Within Groups | 147 | | | | |

Table 10: Performance ANOVA (Source: Notebook Cell 7)

### ANOVA Results: Aptitude Score

| Source | df | F-Statistic | p-value | Effect Size ($\eta^2$) | Result |
|---|---|---|---|---|---|
| Between Groups | 2 | 101.17 | $< 0.001$ | 0.579 | Significant |
| Within Groups | 147 | | | | |

Table 11: Aptitude ANOVA (Source: Notebook Cell 8)

## A.3 Correlation Matrix

The following matrix shows the Pearson correlation coefficient ($r$) between the key variables.

**Breakdown by Level:**

| Variable | Performance Score | Aptitude Score |
|---|---|---|
| **Performance Score** | 1.000 | **0.887** |
| **Aptitude Score** | **0.887** | 1.000 |

Table 12: Pearson Correlation Matrix (Source: Notebook Cell 11, 13)

- **Advanced:** $r = 0.777$ (Strong)
- **Intermediate:** $r = 0.704$ (Strong)
- **Foundation:** $r = 0.299$ (Weak-Moderate)

# B Data Quality Documentation

## B.1 Data Verification Checklist

The following data quality checks were performed programmatically in the analysis notebook:

☒ **Missing Values Check:** Validated 150/150 non-null entries for all columns (Source: Cell 1).

☒ **Data Type Validation:** Verified numerical types for scores (`float64`, `int64`) and categorical for levels (`object`).

☒ **Normality Check:** Shapiro-Wilk test performed on residuals.

- Advanced: $W = 0.909, p = 0.001$ (Not Normal)
- Intermediate: $W = 0.955, p = 0.053$ (Normal)
- Foundation: $W = 0.949, p = 0.031$ (Not Normal)

☒ **Homogeneity of Variance:** Levene's Test performed ($Statistic = 16.599, p < 0.001$), indicating variances are different.

# C Software and Methods

## C.1 Analysis Software

The analysis was conducted using the **Python** programming language within a Jupyter Notebook environment. Key libraries included:

- **Pandas:** Data manipulation and aggregation.
- **SciPy (stats):** Statistical testing (ANOVA, Pearson, Shapiro-Wilk, Levene).
- **Statsmodels:** Post-hoc analysis (Tukey HSD).
- **Seaborn & Matplotlib:** Data visualization.

## C.2 Statistical Methods Used

The following statistical procedures were applied to the dataset:

1. **Descriptive Statistics:** Mean, Median, Standard Deviation, and Quartiles (IQR).
2. **One-Way ANOVA:** To compare means across the three independent course levels.
3. **Tukey's HSD Test:** Post-hoc analysis to determine specific pairwise differences.
4. **Pearson Correlation Coefficient ($r$):** To measure the linear relationship between Aptitude and Performance.
5. **Effect Size Calculations:**
   - **Eta-Squared ($\eta^2$):** For ANOVA impact.
   - **Cohen's $d$:** For pairwise differences.