OXFORD JOURNAL
OF STUDENT
SCHOLARSHIP

Graphomotor Friction and Linguistic Complexity: A Computer Vision Analysis

Sophie Baryalai
sophiebaryalai@gmail.com

ABSTRACT

Transcription Bottleneck theory suggests that mechanical writing demands consume working memory, limiting the extent to which cognitive ability is reflected in written output. This study investigates whether graphomotor friction, measured using a computer vision–based framework, is associated with changes in linguistic complexity and output efficiency. It was hypothesized that increased motor friction would reduce linguistic complexity and transcription performance.

Using a within-subjects design ($n = 30$), three transcription conditions were compared: dominant-hand baseline, non-dominant handwriting, and speech-to-text production. Linguistic complexity was measured using Flesch-Kincaid (FK), Automated Readability Index (ARI), and Type-Token Ratio (TTR).

Results showed a significant decline in output efficiency, with words per minute decreasing from 147.67 to 8.88 ($t(29) = 24.11$, $p < .001$), alongside reductions in linguistic complexity, as FK scores fell from 7.55 to 3.99 ($p < .001$) and ARI from 6.29 to 3.77 ($p = .008$). These differences are consistent with a transcription bottleneck effect, though condition-level prompt variation is acknowledged as a limitation. A moderate correlation between friction and transcription errors was also observed ($R = 0.37$, $p = .042$). High-friction conditions were associated with convergence in performance across participants, suggesting motor constraints may limit the observable expression of underlying cognitive ability.

INTRODUCTION

In classroom writing assessments, students often exhibit slowed transcription, extended pauses, and irregular pacing despite being able to explain their ideas clearly when speaking (Gregg, 2018; Hayes & Olinghouse, 2015). These patterns reveal a disconnect between cognitive potential and manual execution, suggesting written output may be constrained by transcription mechanics rather than conceptual understanding. Graphomotor friction, defined as variability in motor execution during handwriting, captures this strain as a measurable constraint on written expression.

A study in the *Journal of Educational Psychology* found writing fluency, measured by pause length and words produced per unit of time, was associated with writing quality and mediated the relationship between early transcription skills and later performance (Gregg, 2018). These findings indicate reduced transcription efficiency can limit the extent to which cognitive ability is reflected in written work. The pattern aligns with the Transcription Bottleneck theory, proposing writing performance can be constrained by the mechanical demands of transcription. When the physical act of letter formation requires ongoing effort, it competes with the cognitive resources needed for idea development. A bottleneck is created when transcription processes are not automatic which restricts higher-order composition.

Cognitive Load theory provides a parallel explanation: working memory is finite, and effort devoted to mechanical transcription reduces the capacity available for complex reasoning. As lower-level motor demands increase, fewer resources remain for sentence construction and conceptual depth. Such a resource-dependency model aligns with findings that handwriting effort disrupts fluency and prevents cognitive ability from being fully reflected in the final product (Kim, 2024).

This relationship is supported by prior research indicating that "persistence with writing tasks is interrelated with an individual's affective response and cognitive processing abilities" (Gregg, 2018). Gregg's findings therefore suggest that observable disengagement during high-friction writing reflects processing strain rather than diminished motivation, a distinction with direct implications for how educators interpret student behavior during timed assessments.

LITERATURE REVIEW

Across multiple studies, transcription automaticity is consistently identified as a prerequisite for higher-order writing processes. Writing operates through a layered cognitive system where lower-level skills, letter formation and spelling, must reach a state of invisibility before conceptualization can fully develop. "Having to think about how to form a letter may tax children's processing capacity, resulting in the loss of one or more writing ideas" (Skar et al., 2022). When sub-lexical processes are not automated, "most WM resources are devoted to graphomotor and orthographic aspects," leaving a deficit of resources for composition (Da Vita et al., 2021). Consequently, the physical act of writing becomes a cognitive burden rather than a neutral tool. Children then become less motivated to write since it is harder for them to transcribe (Hayes & Olinghouse, 2015). Such foundational constraints on cognitive resources directly inform the role of working memory limitations in shaping linguistic output.

Building on the premise of competing cognitive demands, working memory theory provides a framework for understanding how transcription demands constrain linguistic sophistication. Working memory is defined as a "limited capacity system allowing the temporary storage and manipulation of information necessary for such complex tasks" (Baddeley, 2000). Because a "general limit" exists on an individual's ability to perform mental work, cognitive effort does not accumulate but instead reallocates (Bruya & Tang, 2018). Sophisticated language relies on transcription requiring minimal attention, as "execution can, when well-practiced, proceed virtually automatically," while conceptual formulation remains effortful (Kellogg, 1996). When motor execution remains manual rather than automatic, it reduces the cognitive resources available for lexical selection.

The result of such resource competition is a phenomenon which can be conceptualized as a "Lexical Paradox," where a student's internal intellectual sophistication is masked by the physical demands of the page. In twice-exceptional students, "their gifts mask their disabilities and their disabilities mask their gifts," causing them to appear average despite high intelligence (Krochak & Ryan, 2007). The dual-masking effect is further exacerbated by rater subjectivity in traditional assessment, where "[r]aters awarded higher scores to responses presented in handwritten form as compared to the exact same responses presented as computer-printed text" (Russell & Tao, 2004). This bias compounds the disadvantage faced by high-friction writers, who may be penalized twice, once for reduced complexity and again for reduced legibility, necessitating a shift toward objective, kinematic measurement.

Computational analysis addresses the limitations of traditional assessments, which are often "restricted to the analysis of the final, static handwriting product and [do] not consider... movement dynamics" (Gargot et al., 2020). By age 14, handwriting is typically automated; therefore, measurable friction in older cohorts indicates a persistent neurological bottleneck rather than developmental immaturity (Danna et al., 2023). Modern computational models can now classify these motor features with high precision, "reaching 96% accuracy" in predicting transcription difficulties (Ikermane & El Mouatasim, 2023). The technology allows for the isolation of motoric strain, which can then be bypassed through alternative modalities.

Speech-to-text (STT) may serve as a means of reducing transcription-related cognitive load by removing the mechanical burden of transcription. Kraft (2023) notes STT allows students to write "without having to focus on spelling." Cognitive capacity is effectively freed for higher-level planning and increased vocabulary diversity (Almgren Bäck et al., 2024).

Transcription research has established handwriting's role in supporting composition, yet the measurement of handwriting difficulty remains largely observational. In their meta-analytic review, existing assessment tools prioritize visible outcomes rather than measurable motor dynamics (Feng et al., 2019). This limitation is reflected in the observation that "the evaluation of handwriting legibility is more subjective and qualitative than the evaluation of handwriting fluency, potentially leading to inaccurate measurements" (Feng et al., 2019). The reliance on visual judgement positions handwriting difficulty as something observed rather than quantified. As a result, researchers can identify messy output but lack a standardized numerical metric to measure the degree of motoric friction occurring during transcription. The absence of such a metric leads to a clear methodological gap: although transcription effort is widely theorized to influence writing performance, research rarely measures it as a concrete and continuous variable. Without an objective way to quantify graphomotor strain, researchers cannot directly test whether increased motor effort corresponds with measurable changes in linguistic sophistication. As a result, the relationship between motor load and lexical complexity is often inferred rather than experimentally isolated.

To address the limitation, the study introduces a Computer Vision framework using OpenCV to translate spatial irregularity into a measurable Friction Score, allowing transcription strain to be quantified rather than inferred. By utilizing a within-subjects design featuring three transcription conditions, dominant hand (baseline motor load), non-dominant hand (elevated motor load), and speech-based response (transcription bypass), each condition represents a different degree of mask over underlying cognitive potential, directly operationalizing the Lexical Paradox framework.

By combining an objective friction metric with controlled manipulation of transcription load, this study addresses a methodological gap by quantifying graphomotor friction as a continuous variable and directly linking measured motor strain to linguistic complexity, moving beyond the behavioral inference that characterizes prior research. The framework leads to the central research question of the study: To what extent does increased graphomotor friction, as quantified by computer vision analysis, correlate with a reduction in lexical diversity and linguistic complexity among high school students?

RESEARCH DESIGN AND METHODOLOGY

Quantitative research entails the systematic collection and statistical analysis of numerical data to identify patterns and relationships. Such a framework is exemplified here through the measurement of variables such as graphomotor friction scores, words per minute, and readability indices analyzed across controlled experimental conditions. The study utilizes a within-subjects experimental design to examine the relationship between graphomotor friction and lexical complexity. A quantitative design was necessary to operationalize graphomotor friction as a continuous variable and statistically evaluate its relationship with linguistic complexity across controlled conditions.

To ensure internal validity, defined as “the extent to which [a study's] design and the data it yields allow the researcher to draw defensible conclusions about cause-and-effect” (Leedy & Ormrod, 2019), the researcher selected a within-subjects approach where each of the thirty participants serves as their own internal control. Utilization of a within-subjects design choice facilitates the isolation of the primary variable by mitigating the influence of participant-specific factors, such as baseline cognitive ability or pre-existing vocabulary depth, which would otherwise act as confounding variables in a between-groups comparison. Consequently, shifts in linguistic performance can be evaluated as functions of varying motor demands instead of individual aptitude.

The experimental framework applies the Transcription Bottleneck theory (Berninger et al., 1992). To evaluate the impact of mechanical barriers on linguistic choice, the researcher implemented a protocol designed to model high-friction writing environments. By requiring participants to use their non-dominant hand, the study intentionally disrupts transcription automaticity to simulate motor load (Kellogg, 1996). Transitioning from an automated process to a conscious motor task enables analysis of how increased physical effort may influence the complexity of the produced text, although it may not fully replicate clinical dysgraphia.

To establish a baseline for linguistic potential, the design incorporates a Speech-to-Text (STT) modality as a low transcription condition. Removing the mechanical transcription requirement allows for an assessment of the participant's lexical capability independent of motor constraints. The study thus examines the Independent Variable (IV) of transcription modality across three levels, consisting of the dominant hand, non-dominant hand, and audio output trials. The Dependent Variables (DV) are defined as lexical complexity, quantified through established readability indices such as Type-Token Ratio and ARI scores, and graphomotor friction, quantified via kinematic scores extracted through computer vision. By comparing these conditions, the study seeks to empirically investigate the Lexical Paradox to assess whether the physical demands of transcription may obscure a participant's true linguistic sophistication (Krochak & Ryan, 2007). It should be noted that distinct prompts were used across conditions to minimize practice effects, a methodological tradeoff that may have allowed prompt topic to influence linguistic complexity independently of transcription modality.

The study utilized a sample of thirty participants (n = 30) recruited through purposive sampling. Purposive sampling is defined as a process of “choosing participants... for a particular purpose” (Leedy & Ormrod, 2021). The primary selection objective was to isolate students who had surpassed basic handwriting development, ensuring observed friction reflected cognitive rather than maturational factors; participants were recruited across diverse academic tracks, including Advanced Placement, Dual Enrollment, and standard curricula. Integrating varied academic tracks ensured the sample provided a balanced representation of the 14–17 age range. The chosen age bracket was informed by the findings of Danna et al. (2023), who observe handwriting automation is typically almost complete by age 14. By focusing on a cohort of experienced writers across different educational backgrounds, the researcher could ensure letter formation no longer required primary conscious focus, aligning with the “automaticity” theories (Kellogg, 1996). Consequently, any graphomotor friction suggested within the data is unlikely to reflect developmental immaturity and is instead more indicative of a persistent cognitive constraint.

Ethical safeguards included secured informed consent and parental assent for all minor participants. Participation was voluntary, with an explicit right to withdraw without penalty. To ensure privacy, names were replaced with alphanumeric codes (P01–P30) and all data was stored on a secure, password-protected drive restricted to the researcher.

To facilitate the high-resolution analysis required for the present investigation, a custom Python-based script processed the participants' writing samples. Transitioning from manual observation to a computerized model ensures data integrity while eliminating observer bias, a phenomenon where a researcher might subconsciously over-interpret poor handwriting as a sign of diminished cognitive ability. By removing subjective human interpretation, the researcher established a standardized environment where “the data of these three indicators were obtained through a Corpus Tool... and an online software”

(Liu & Dou, 2023). Consequently, such an automated framework allowed for the objective extraction of linguistic data, transforming raw handwritten text into quantifiable indicators associated with the participant's cognitive processing during the writing task. The computational pipeline was developed through iterative prototyping with technical support for implementation.

As demonstrated in a model study, "handwriting is a complex perceptual motor task" involving multiple cognitive and motor processes (Gargot et al., 2020). Their use of computational feature extraction supports objective analysis of handwriting variability, a framework applied through OpenCV to quantify graphomotor friction.

In addition to linguistic analysis, the script incorporated computer vision techniques via the OpenCV (Open Source Computer Vision Library) to analyze visual information from digital images (Appendix A). Within the present study, the script utilized thresholding, a digital processing technique, which converts grayscale images into binary black-and-white pixels, to isolate ink strokes from the paper background. Following stroke isolation, pixel intensity analysis was used to quantify stroke darkness, which serves as a proxy for pen pressure. The methodology assumes higher pixel density correlates with greater physical force applied to the writing surface, a measurement approach supported by prior research (Horie & Shibata, 2018).

The script was programmed to calculate specific indices, which serve as proxies for cognitive management and linguistic maturity. Primary among these was the Number of Different Words (NDW) and the Type-Token Ratio (TTR). The inclusion of NDW is justified because "NDW in written narratives may be an efficient, educationally relevant, marker of language maturity" (Wood et al., 2019). These metrics allowed the researcher to track lexical diversity, which "refers to the variation in vocabulary use within a text... measured by the value of the type-token ratio (TTR)" (Liu & Dou, 2023). While TTR can be sensitive to text length, it remains a standard metric for comparing short-form writing samples of similar duration.

Furthermore, these factors were chosen specifically to test the "Transcription Bottleneck" hypothesis. The theoretical basis for the current study is "writing speed and accuracy are not compatible... [and] affect the 'trade-off' correlation" (Horie & Shibata, 2018). By analyzing these trade-offs through automated extraction, the researcher could pinpoint how mechanical friction disrupts cognitive flow.

The methodology assumes participants possess baseline handwriting automaticity, ensuring observed friction reflects cognitive constraint rather than developmental limitation. OpenCV-derived metrics are treated as relative measures of motor variability, while linguistic indices represent cognitive complexity. Controlled conditions isolate transcription modality as the primary variable. The study hypothesizes that increased graphomotor friction will significantly reduce linguistic complexity and output efficiency while increasing transcription errors, indicating an inverse relationship between motor-load and cognitive expression.

EXPERIMENTAL PROCEDURE

An experimental approach utilized a structured process for establishing a baseline before active testing began. To help control for internal validity, the procedure began with a comprehensive pre-test survey consisting of 20 distinct variables. Such a survey served to identify potential confounding variables in the sample, specifically participant fatigue or existing word avoidance habits, which could otherwise skew lexical diversity results.

Furthermore, the survey collected data regarding academic background; measured physiological and mental states; recorded cognitive and motor self-perceptions; and accounted for graphomotor history and habits. Such variables were not directly manipulated but were recorded to contextualize individual differences during analysis.

The core of the experiment involved a sequential, three-stage task designed to manipulate transcription friction. As shown in Figure 1, the handwriting samples produced across conditions reflect visible differences in motor execution, illustrating the contrast between dominant and non-dominant writing. To control for order effects, tasks were counterbalanced across participants, with condition sequences varied systematically across the sample. This design choice strengthens internal validity by ensuring observed differences in linguistic output reflect transcription modality rather than task familiarity or fatigue. Each condition is outlined below.

Condition 1 (Dominant Hand — Control): Participants transcribed a standardized sentence to establish baseline graphomotor speed and accuracy: “The scientific method requires objective observation and precise data collection to ensure the validity of experimental results.”

Condition 2 (Non-Dominant Hand — Increased Motor Load): Participants completed a timed (2-minute) writing task using their non-dominant hand in response to the prompt: “Use the most advanced, professional language possible to write the steps for preparing a standard sandwich.”

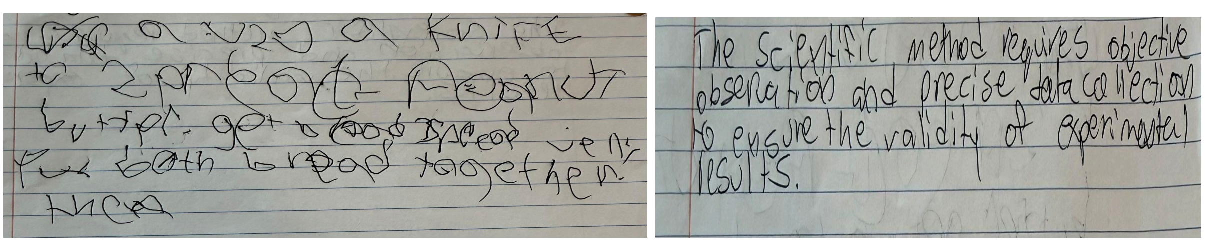


Figure 1. Comparison of dominant and non-dominant handwriting samples. The non-dominant sample shows visibly reduced legibility and consistency, reflecting the increased motor demand of the high-friction condition.

Condition 3 (Speech-to-Text — Reduced Motor Load): Participants completed a timed (1-minute) response using speech-to-text software (Google Voice Typing) in response to the prompt: “Use the most advanced, professional language possible to explain the steps of your morning routine from waking up to leaving the house."

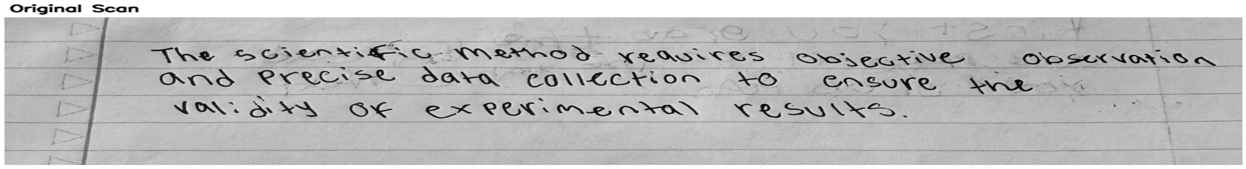
Distinct prompts were used to minimize practice effects and prevent pre-loading of vocabulary from earlier conditions, a deliberate methodological tradeoff. Because conditions utilized different prompts, variation in topic may have influenced linguistic complexity independent of transcription modality. Future studies should ensure prompts are counterbalanced across handwriting and speech-to-text conditions, so that each prompt appears with equal frequency in both modalities. Time limits also differed across conditions to normalize total linguistic output, as speaking produces a higher word rate than non-dominant writing. The design choice allowed for more comparable token counts across trials, although differences in modality may still influence production speed. Instructions were standardized and timing was controlled using a digital stopwatch.

Once the live experiment was completed, the raw data underwent a meticulous transcription and coding phase. The audio recordings from the Speech-to-Text task were reviewed to identify Verbal Disfluencies, such as vocal fillers or mid-sentence stalls, which indicate cognitive load during verbalization. Simultaneously, the handwritten samples were transcribed into a digital text format. During the phase, the researcher manually calculated Transcription Errors specifically for the non-dominant hand trials, noting any mechanical spelling or character errors occurring due to motor strain.

Four primary linguistic metrics were calculated for each trial to quantify written output. Type-Token Ratio (TTR), defined as the ratio of unique words to total words, was used to assess lexical diversity. Flesch-Kincaid (FK) Grade Level evaluated language maturity based on sentence length and syllable complexity. Words Per Minute (WPM), calculated from total word count over time, measured production speed and processing efficiency. Automated Readability Index (ARI), a character-based readability metric, was included to cross-validate structural complexity across conditions.

The process then shifted to a digital lab environment. All physical handwriting samples were scanned as high-resolution JPEGs and placed into a directory for processing. Using a custom OpenCV (Open Source Computer Vision) script, the researcher applied Otsu's Binarization to each image. As illustrated in Figure 2, Otsu's Binarization, a clustering-based thresholding algorithm, converted grayscale scans into a binary mask by calculating the optimal intensity separator to isolate ink strokes from paper backgrounds while filtering sensor noise.

Original Scan



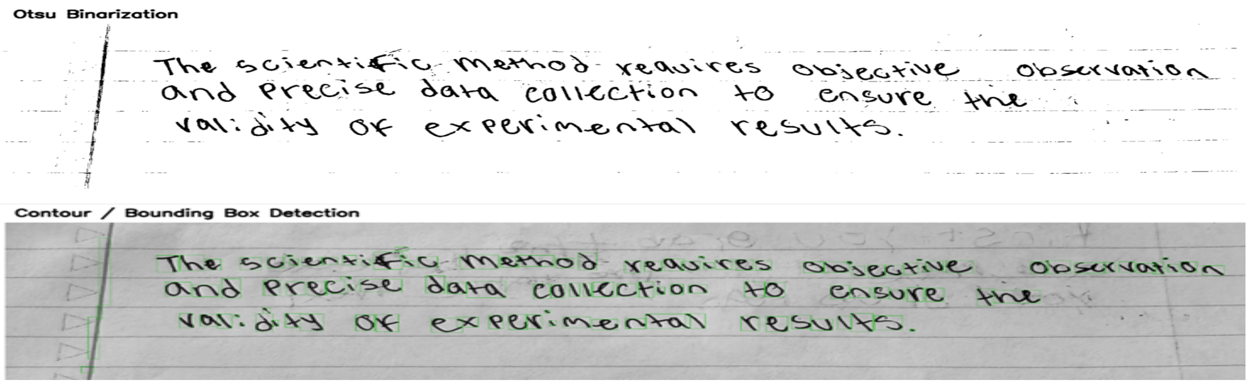


Figure 2. Computational pipeline for dominant-hand handwriting analysis. (A) Original scan; (B) binarized image using Otsu thresholding; (C) contour-based segmentation for feature extraction. Each stage progressively isolates handwriting features used to calculate the friction score, ink intensity, and legibility index.

The application performed three automated visual extractions for each sample. Friction score was calculated as the coefficient of variation (standard deviation/mean) of bounding-box heights to capture motor instability across varying character sizes. Character height inconsistency was treated as an indicator of diminished motor stability, since greater variability in letter sizing suggests that handwriting execution becomes less controlled as physical demand increases. This interpretation is grounded in the assumption that a writer with stable motor control will produce letters of relatively consistent height, while increased friction disrupts that consistency in measurable ways.

Ink intensity was derived from grayscale pixel analysis within segmented strokes, with darker values indicating higher ink concentration and serving as a proxy for pen pressure. Legibility was assessed using a composite index combining y-position stability, aspect ratios, stroke area, and pixel density, weighted equally to reflect overall structural consistency. As shown in Figure 3, bounding-box height variation across conditions illustrates differences in motor stability between dominant and non-dominant writing samples.

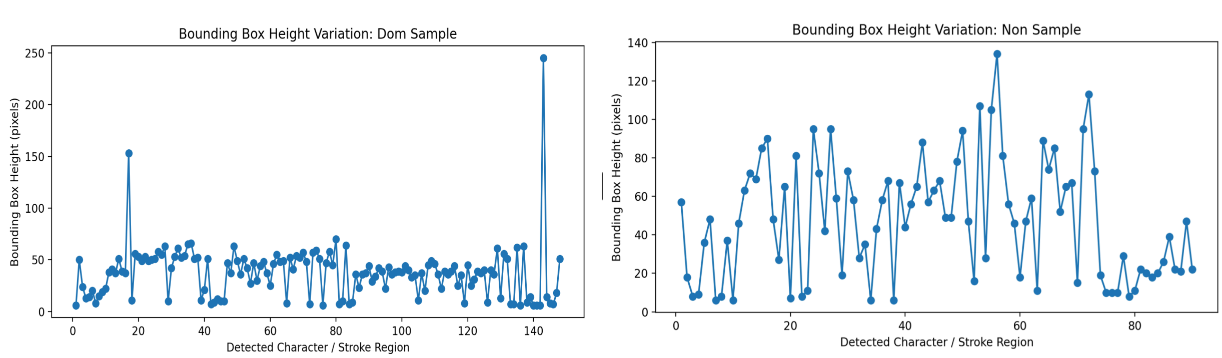


Figure 3. Bounding-box height variation across dominant and non-dominant handwriting samples, illustrating increased variability under high-friction conditions. Greater variability in detected

character heights under the non-dominant condition is consistent with its classification as a high-friction writing environment.

RESULTS

The results examine whether increased graphomotor friction produces a measurable reduction in linguistic complexity and output efficiency while increasing transcription errors across transcription modalities.

The disparity in production speed across modalities was measured by Mean Words Per Minute (WPM) to establish a baseline for efficiency. As illustrated in Figure 4, the Audio modality achieved a mean output of 147.67 WPM, while the Non-Dominant condition fell to a mean of 8.88 WPM. A t-test (Paired Two Sample for Means) indicated the observed 94% loss in efficiency is statistically significant, yielding a t-stat of 24.11 and a $P < .001$. The difference in bar heights visible in Figure 4 reflects the magnitude of this efficiency gap across modalities.

To further examine the motor demands associated with this reduction, physical resistance was quantified using a friction score across transcription conditions. The dominant condition produced a mean friction score of 92.14, while the non-dominant condition increased to 112.75. A paired t-test yielded $t = -1.99$ with $p = .056$; while close to the conventional threshold, this value does not meet the standard criterion for statistical significance ($\alpha = .05$). Although the non-dominant condition showed a higher mean friction score, this difference approached but did not reach statistical significance. This finding should be interpreted cautiously: the observed increase in mean friction score from 92.14 to 112.75 represents a directional trend consistent with the hypothesis, but cannot be treated as confirmed. The friction score should therefore be regarded as a preliminary indicator rather than a fully validated measure of motor load, and warrants replication with a larger sample.

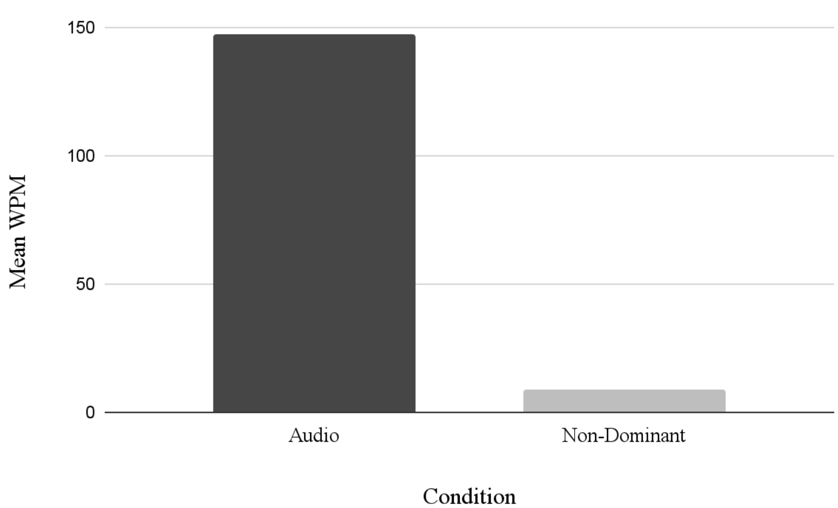


Figure 4. Mean words per minute (WPM) across transcription conditions, illustrating a substantial decrease in output efficiency under increased motor load. The audio condition produced a notably higher output rate than the non-dominant condition, consistent with the predicted reduction in production efficiency under increased motor load.

Linguistic depth was assessed using the Flesch–Kincaid (FK) and Automated Readability Index (ARI) scales. The audio modality exhibited higher linguistic complexity, with a mean FK grade level of 7.55 and an ARI of 6.29. In contrast, the non-dominant condition showed reduced complexity, with FK decreasing to 3.99 and ARI to 3.77, a regression toward early elementary writing sophistication suggesting motor constraint actively simplified syntactic construction rather than merely slowing output. As shown in Figure 5, both FK and ARI scores were consistently lower under the non-dominant condition, and paired t-tests indicated that these reductions were statistically significant for both FK ($p < .001$) and ARI ($p = .008$). It should be noted that multiple dependent variables were tested across conditions without a family-wise error correction such as Bonferroni adjustment; reported p-values should therefore be interpreted with appropriate caution.

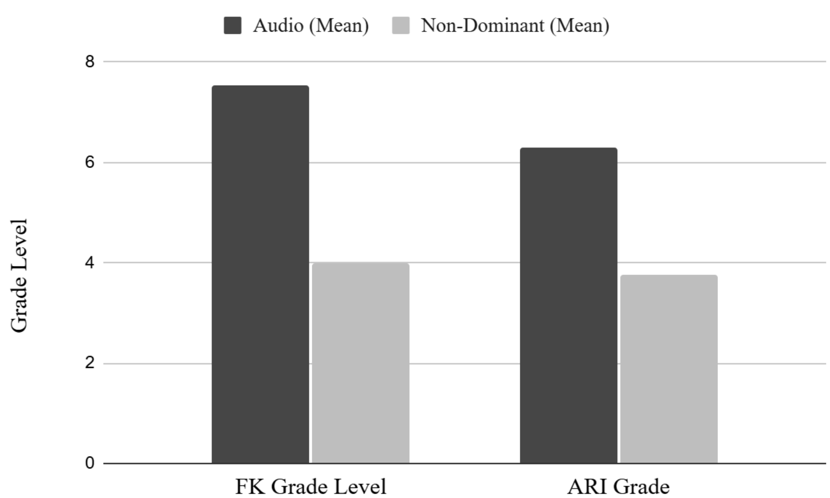


Figure 5. Mean Flesch–Kincaid (FK) and Automated Readability Index (ARI) grade levels across transcription conditions. Both measures show lower scores under the non-dominant condition, indicating a reduction in linguistic complexity associated with increased motor demand.

To determine if physical motor-load directly influences accuracy, a linear regression analysis was conducted between Friction Scores and Transcription Errors within the Non-Dominant condition. The analysis yielded a Multiple R of 0.37, indicating a moderate positive correlation. As shown in Figure 6, the upward trend line across the scatter plot reflects this relationship, with higher friction scores corresponding with greater transcription error frequency. The regression model produced a statistically significant P-value of .042, suggesting motor instability is associated with increased error frequency.

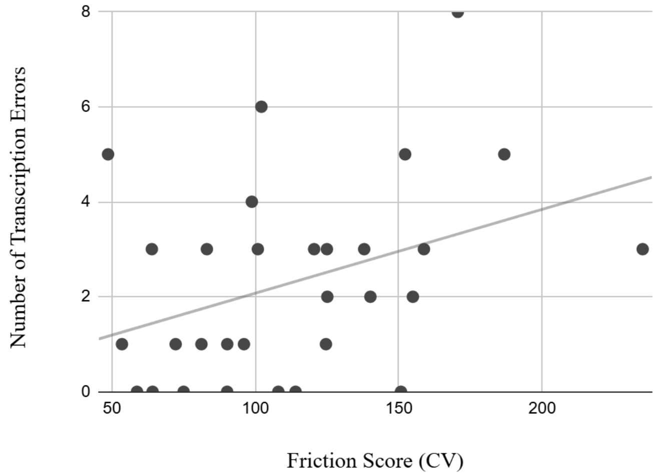


Figure 6. Relationship between friction score and transcription errors in the non-dominant condition. The positive trend line indicates that higher friction scores correspond with greater error frequency, suggesting motor instability is associated with reduced writing accuracy.

Across these measures, consistent patterns emerge linking motor constraint to changes in linguistic output. Reductions in WPM align with declines in both FK Grade Level and ARI, indicating that increased motor constraint corresponds with simplified sentence construction rather than solely slower transcription. Moderate correlations in readability measures contrast with the weaker association between baseline speaking speed and non-dominant writing performance ($r = 0.25$), suggesting speaking ability does not transfer to non-dominant writing, and motor constraint reduces the influence of individual variation.

The Flesch–Kincaid Grade Level shows a significant regression from a mean of 7.54 to 3.98 ($p < .001$), with a t-stat of 4.66 and a moderate Pearson correlation ($r = 0.41$). Similarly, ARI decreases from 6.28 to 3.77 ($p = .007$), with a moderate correlation ($r = 0.49$).

In contrast, type-token ratio (TTR) presents an inverse pattern, increasing from 47.86 to 85.28 ($p < .001$) with a weak Pearson correlation ($r = 0.21$). This pattern reflects a sample-length artifact rather than genuine lexical diversity, as limited word count reduces repetition and inflates diversity ratios.

This pattern is further reflected at the individual level. Advanced academic training offers little protection against the mechanical bottleneck. Participants P03 and P05, despite demonstrating post-graduate baseline complexities, showed a marked decline in linguistic complexity under high motor-load conditions, with output measures regressing to early-elementary levels. Subject P03, with a baseline ARI of 20.95 and 159 WPM, regressed to an ARI of 5.1 and 9 WPM during the non-dominant trial. Similarly, P05 dropped from an audio baseline of 96 WPM and 8.72 ARI to 6.5 WPM and 2.11 ARI. These cases suggest that students with higher baseline linguistic complexity may still show substantial simplification under high motor-load conditions, indicating that intellectual expression may remain constrained by motor fluency regardless of baseline ability.

An inverse relationship between mechanical speed and vocabulary diversity emerges during non-dominant trials, suggesting a statistical artifact. As output slows, the Type-Token Ratio (TTR) increases; however, this trend does not necessarily indicate intentional word choice or improved vocabulary. Instead, it reflects the “Brevity Effect,” where extreme motor friction restricts subjects to a sample size so small that word repetition becomes mathematically less likely. Participants produced limited linguistic output, reducing the likelihood of natural word repetition, resulting in an artificially inflated diversity score.

Support for this pattern appears in the Pearson correlation ($r = -0.684$; $R^2 = 0.468$, $p < .001$), indicating that a substantial proportion of the variance in vocabulary diversity is associated with mechanical speed rather than intentional lexical selection. As a result, motor limitations shape the statistical structure of the output, reinforcing the impact of the “Brevity Effect” under high-friction transcription.

To assess whether external variables contribute to these patterns, correlations were evaluated between psychological measures and both motor performance and linguistic output. Analysis indicates near-zero associations between fatigue ($r = 0.06$), burnout ($r = 0.036$), and sleep ($r = -0.048$) with the friction gap, suggesting that variation in motor difficulty is not explained by environmental or psychological conditions. Although moderate relationships exist among secondary variables—such as fatigue and burnout ($r = 0.44$)—these do not extend to objective motor performance.

Under baseline (audio) conditions, fatigue demonstrates a moderate positive relationship with vocabulary diversity (TTR; $r = 0.54$) and a smaller association with readability (FK; $r = 0.29$). However, under non-dominant conditions, these relationships weaken or reverse, with fatigue showing a weak negative correlation with TTR ($r = -0.14$) and FK ($r = -0.23$). Overall, psychological variables show limited and inconsistent associations with performance under motor constraint, while remaining largely uncorrelated with the friction gap.

A similar pattern emerges in the relationship between lexical diversity and syntactic complexity under constrained output conditions. Lexical diversity, as measured by TTR, becomes decoupled from underlying linguistic ability under these conditions. Although the Low Complexity group displays a higher average TTR, the lower value observed in the High Complexity group reflects how advanced language production depends on sufficient output length to sustain variation. When output is restricted, higher-level synthesis cannot fully materialize, leading to an artificial suppression of measurable diversity.

To examine this pattern, participants were grouped into “Top 8” and “Bottom 8” cohorts based on baseline ARI scores, representing higher and lower initial linguistic complexity. Differences in average Type-Token Ratio (TTR) between these groups imply an inverse relationship between initial complexity and observed lexical diversity. The “Top 8” (high complexity) group regressed to an average TTR of 0.76 ($SD = 0.12$), while the “Bottom 8” (low complexity) group maintained a higher average TTR of 0.88 ($SD = 0.05$). The pattern is supported by a negative Pearson correlation of $r = -0.26$ and a t-statistic of 4.13 (p

< .001) within this exploratory subsample analysis. Because these cohorts were identified post-hoc rather than pre-specified, these results are best understood as hypothesis-generating, pointing toward a convergence effect that warrants direct testing in future research with larger, pre-defined groups.

This divergence reflects the Brevity Effect, where reduced output length limits repetition and inflates ratio-based measures rather than indicating true increases in lexical variation. In contrast, shorter responses limit repetition, inflating TTR independent of underlying lexical sophistication.

In addition, despite initial differences in baseline complexity, both high- and low-complexity groups converged to nearly identical ARI levels under non-dominant conditions, indicating that increased motor constraint reduces observable differences in linguistic performance.

DISCUSSION

The findings fill the identified gap by establishing transcription-related strain as a measurable variable and demonstrating that mechanical constraints on writing can be empirically quantified. Results support the hypothesis that increased graphomotor friction is associated with reduced linguistic complexity and output efficiency while increasing transcription errors across modalities. The comparison between dominant, non-dominant, and audio conditions provides a basis for interpreting how transcription modality influences written expression. By comparing conditions across different transcription modalities, the data are consistent with linguistic complexity being reduced as motor demands increase, a pattern predicted by the Transcription Bottleneck framework. Because conditions differed in both modality and prompt content, however, these findings are best understood as correlational evidence supporting the bottleneck hypothesis rather than direct causal demonstration.

As demonstrated in the results, increased graphomotor friction is associated with reduced external expression of cognitive ability, not by reducing what a student knows, but by limiting the working memory resources available to translate knowledge into complex written language. The pattern aligns with the model of working memory, where increased motor demands correspond with the reduced capacity available for higher-order language processing (Baddeley, 2000). The findings also support the conclusion that "writing fluency … mediated the relations of … transcription skills to writing quality" (Kim, 2024). However, the present study extends prior research by demonstrating that motor-load can be quantified and directly linked to measurable declines in cognitive output, suggesting transcription constraints may compress observable differences in intellectual performance.

Building on these findings, motor friction correlates with a convergence of cognitive output, as data from the high-complexity and low-complexity cohorts is consistent with a significant narrowing of performance gaps. Despite a 5.6-point difference in baseline potential, the high-friction environment resulted in both groups regressing toward a similar complexity floor, suggesting physical constraints are associated with reduced observable performance differences between higher- and lower-aptitude students under elevated motor load conditions.

The brevity effect relates to a measurable shift in lexical variety, where higher-aptitude participants produced less varied vocabulary under motor stress, as the brevity of output may have increased reliance on frequently repeated common words. While a reduced word count can mathematically skew certain diversity ratios, the range of vocabulary selection appears to be notably reduced, aligning with theory positioning lexical selection as secondary to primary metabolic and cognitive demands of managing a high-friction transcription medium.

In parallel, physical friction functions as a stronger constraint than biological and academic variables, as the case study of participants with contrasting friction profiles indicates factors such as sleep volume and intellectual independence may be secondary to mechanical difficulty. Participant P04, despite reported rest (7+ hours) and high baseline potential, exhibited a substantial linguistic decline alongside a friction score of 235.03. While this case is consistent with the hypothesis that friction magnitude may function as a stronger constraint on output than individual preparedness, single-participant observations cannot support generalizable conclusions and are presented here as illustrative rather than evidential.

Across these patterns, cognitive resource allocation prioritizes motor stability over qualitative synthesis. Data regarding neatness priority and self-reported effort suggests increased conscious attention to the physical act of writing does not yield higher linguistic complexity. Instead, when participants prioritized legibility under high-friction conditions, there was a measurable decline in syntax and grade-level metrics. Cognitive trade-offs suggest that motor control and higher-order composition compete for the same limited cognitive resources; when motor demands increase, cognitive resources available for higher-order language production appear to be redirected toward managing physical execution.

The findings within the present research suggest potential implications for educational assessment and the evaluation of cognitive potential. Within the constraints of this sample, timed handwritten examinations may disproportionately measure motor dexterity and friction tolerance rather than a student's actual command of the subject matter. Students with high cognitive potential but high motor friction may be systematically misrepresented in traditional settings, as their measured linguistic output may substantially underrepresent their underlying cognitive ability. Because academic and professional evaluations frequently equate simplified output with lower aptitude, a disparity may exist between a student's demonstrated and actual ability when assessment relies on high-friction transcription modalities. Capturing student ability accurately likely requires transition toward low-friction modalities, such as speech-to-text or high-speed typing, ensuring physical transcription difficulty does not limit intellectual expression.

At the same time, several limitations constrain the scope and interpretation of these findings. While results provide empirical evidence of the graphomotor bottleneck, data extraction faced challenges during non-dominant hand trials due to significantly reduced writing speed and script size. Smaller word counts in the high-friction condition limited the robustness of lexical variety metrics compared to the expansive audio samples, introducing brevity bias as shorter samples correspond with the reduced reliability of lexical metrics.

A further limitation concerns prompt non-equivalence across conditions. Because distinct prompts were used to control for practice effects, the study cannot fully disentangle the effect of transcription modality from variation in prompt topic or difficulty. Future designs should counterbalance prompt assignment across conditions to address both threats to internal validity simultaneously.

Demographic homogeneity further restricts generalizability, as the sample consisted of 30 participants aged 14 to 17 within a single geographic and socioeconomic corridor. Such a focus excludes developmental data from younger children still acquiring motor automation and does not account for adults with different cognitive compensation strategies. The absence of neuroimaging tools limits direct observation of brain activity, and without EEG or fMRI data, the theory of resource competition between the motor and prefrontal cortex remains a behavioral inference based on output patterns.

The study also did not include formal clinical screening for undiagnosed motor processing disorders or dysgraphia, meaning undisclosed differences may influence baseline friction levels of certain subjects while the current methodology relied on self-reporting.

Acknowledging these constraints provides necessary context for the biological and mechanical complexities involved in the transcription process. Addressing these constraints requires further investigation into the conditions under which motor limitations can be reduced, measured, and potentially overcome.

FUTURE DIRECTIONS

Subsequent research should expand into longitudinal designs and more diverse demographics. A primary direction could involve analyzing younger primary school students and cohorts with diagnosed motor impairments to determine if the observed motor constraint is a fixed neurological constraint or a variable reducible through targeted training.

Furthermore, a longitudinal study tracking participants over months of non-dominant hand practice could demonstrate the extent to which the transcription bottleneck adapts as high-friction tasks become automated.

Future studies should also consider a three-way comparison between handwriting, typing, and speech-to-text (STT) to investigate if alternative modalities consistently mitigate the brevity effect for high-friction writers.

Finally, incorporating neuroimaging tools such as fMRI or EEG would allow researchers to move past behavioral inference to explore any direct relationship between neural activity and motor friction.

ACKNOWLEDGEMENTS

Several individuals contributed to the completion of this research. Appreciation is extended to those who assisted with managing checkpoints, reviewing early drafts, and providing critical feedback during later stages of revision.

This study would not have been possible without the participants who volunteered their time and effort. Additional thanks to those who assisted in coordinating participant recruitment, which contributed to the development of the samples.

REFERENCES

Almgren Bäck, G., Mossige, M., Bundgaard Svendsen, H., Rønneberg, V., Selenius, H., Berg Gøttsche, N., … Svensson, I. (2024). Speech-to-text intervention to support text production among students with writing difficulties: a single-case study in nordic countries. *Disability and Rehabilitation: Assistive Technology*, *19*(8), 3110–3129. https://doi.org/10.1080/17483107.2024.2351488

Baddeley A. (2000). The episodic buffer: a new component of working memory?. *Trends in cognitive sciences*, *4*(11), 417–423. https://doi.org/10.1016/s1364-6613(00)01538-2

Berninger, V. W., Yates, C., Cartwright, A., Rutberg, J., Remy, E., & Abbott, R. (1992). Role of transcription in written language expression revisited: Next steps in recovering from the structuralist error. *Reading and Writing*, *4*(3), 257–274. https://doi.org/10.1007/BF01027151

Bruya, B., & Tang, Y. Y. (2018). Is Attention Really Effort? Revisiting Daniel Kahneman's Influential 1973 Book *Attention and Effort*. *Frontiers in psychology*, *9*, 1133. https://doi.org/10.3389/fpsyg.2018.01133

Chernov, Y., & Caspers, C. (2020). Formalized Computer-Aided Handwriting Psychology: Validation and Integration into Psychological Assessment. *Behavioral sciences (Basel, Switzerland)*, *10*(1), 27. https://doi.org/10.3390/bs10010027

Danna, J., Puyjarinet, F., & Jolly, C. (2023). Tools and Methods for Diagnosing Developmental Dysgraphia in the Digital Age: A State of the Art. *Children (Basel, Switzerland)*, *10*(12), 1925. https://doi.org/10.3390/children10121925

De Vita, F., Schmidt, S., Tinti, C., & Re, A. M. (2021). The Role of Working Memory on Writing Processes. *Frontiers in psychology*, *12*, 738395. https://doi.org/10.3389/fpsyg.2021.738395

Feng, Lin, et al. "The Roles of Handwriting and Keyboarding in Writing: A Meta-Analytic Review."

Reading and Writing, vol. 32, no. 1, 2019, pp. 33–63. *SpringerLink*, https://doi.org/10.1007/s11145-017-9749-x.

Gargot, T., Asselborn, T., Pellerin, H., Zammouri, I., M Anzalone, S., Casteran, L., Johal, W., Dillenbourg, P., Cohen, D., & Jolly, C. (2020). Acquisition of handwriting in children with and without dysgraphia: A computational approach. *PloS one*, *15*(9), e0237575. https://doi.org/10.1371/journal.pone.0237575

Gregg, Noel, and J. Nelson. "Empirical Studies on the Writing Abilities of Adolescents and Adults with Learning Difficulties." *Writing Development in Struggling Learners: Understanding the Needs of Writers across the Lifecourse*, edited by Barbara Miller et al., vol. 35, Brill, 2018, pp. 73–96. *JSTOR*, www.jstor.org/stable/10.1163/j.ctv3znwkm.8.

Hayes, J. R., & Olinghouse, N. G. (2015). Can Cognitive Writing Models Inform the Design of the Common Core State Standards? *The Elementary School Journal*, *115*(4), 480–497. https://doi.org/10.1086/681909

Horie, S., & Shibata, K. (2018). Quantitative evaluation of handwriting: factors that affect pen operating skills. *Journal of physical therapy science*, *30*(8), 971–975. https://doi.org/10.1589/jpts.30.971

Ikermane, M., & El Mouatasim, A. (2023). Digital handwriting characteristics for dysgraphia detection using artificial neural network. *Bulletin of Electrical Engineering and Informatics*, *12*(3), 1693–1699. https://doi.org/10.11591/eei.v12i3.4571

Kellogg, R. T. (1996). A model of working memory in writing. In C. M. Levy and S. Ransdell (Eds.), The science of writing: Theories, methods, individual differences, and applications (pp. 57-71). Lawrence Erlbaum Associates, Inc.

Kim Y. G. (2024). Writing Fluency: Its Relations With Language, Cognitive, and Transcription Skills, and Writing Quality Using Longitudinal Data From Kindergarten to Grade 2. *Journal of educational psychology*, *116*(4), 590–607. https://doi.org/10.1037/edu0000841

Kraft, S. (2023). Revisions in written composition: Introducing speech-to-text to children with reading and writing difficulties. *Frontiers in Education*, *8*, 1133930. https://doi.org/10.3389/feduc.2023.1133930

Krochak, L., & Ryan, T. (2007). The challenge of identifying gifted/learning disabled students. *International Journal of Special Education*, *22*(3), 44–53. https://files.eric.ed.gov/fulltext/EJ814509.pdf

Leedy, P. D., & Ormrod, J. E. (2021). *Practical Research: Planning and Design* (12th ed.). Pearson.

Liu, Z., & Dou, J. (2023). Lexical density, lexical diversity, and lexical sophistication in simultaneously interpreted texts: A cognitive perspective. *Frontiers in Psychology*, *14*, 1276705. https://doi.org/10.3389/fpsyg.2023.1276705

Russell, M., & Tao, W. (2004). Effects of handwriting and computer-print on composition scores: A follow-up to Powers, Fowles, Farnum, and Ramsey. *Practical Assessment, Research, and Evaluation*, *9*(1), Article 1. https://doi.org/10.7275/vbw6-re29

Skar, G. B., Lei, P. W., Graham, S., Aasen, A. J., Johansen, M. B., & Kvistad, A. H. (2022). Handwriting fluency and the quality of primary grade students' writing. *Reading and Writing*, *35*(2), 509–538. https://doi.org/10.1007/s11145-021-10185-y

Wood, C. L., Bustamante, K. N., Schatschneider, C., & Hart, S. A. (2019). Relationship Between Children's Lexical Diversity in Written Narratives and Performance on a Standardized Reading Vocabulary Measure. *Assessment for effective intervention : official journal of the Council for Educational Diagnostic Services*, *44*(3), 173–183. https://doi.org/10.1177/1534508417749872

APPENDICES

Appendix A. Computer Vision Analysis Script (Python/OpenCV)

This appendix includes the Python script used to process handwriting data and extract quantitative features for analysis.

```
import cv2
import numpy as np
import pandas as pd
import os

# --- 1. SETUP ---
EXCEL_FILE = "Giant_Miner.xlsx"
PHOTO_FOLDER = "photos"

MASTER_CHART_DATA = {
  "1": {"Audio": [173.0, 0.428, 5.34, 2.93, 0], "Non": [6.50, 1.000, 4.91, 5.36, 6]},
  "2": {"Audio": [184.0, 0.478, 6.66, 5.53, 0], "Non": [10.50, 0.952, 1.19, 0.46, 3]},
  "3": {"Audio": [159.0, 0.472, 18.59, 20.95, 2], "Non": [9.00, 0.889, 5.85, 5.10, 8]},
  "4": {"Audio": [170.0, 0.447, 5.73, 4.14, 0], "Non": [10.00, 0.850, 3.65, 1.47, 3]},
  "5": {"Audio": [96.0, 0.646, 9.01, 8.72, 11], "Non": [6.50, 1.000, 3.28, 2.11, 3]},
  "6": {"Audio": [187.0, 0.428, 6.46, 4.98, 0], "Non": [10.50, 0.762, 6.65, 6.79, 3]},
  "7": {"Audio": [159.0, 0.434, 7.00, 6.10, 0], "Non": [12.00, 0.750, 9.01, 7.64, 5]},
```

```
    "8": {"Audio": [132.0, 0.523, 10.87, 11.16, 3], "Non": [8.50, 0.882, 4.67, 4.95, 4]},
    "9": {"Audio": [114.0, 0.465, 9.19, 7.61, 5], "Non": [9.50, 0.789, 4.86, 5.92, 5]},
    "10": {"Audio": [147.0, 0.449, 12.63, 12.81, 1], "Non": [8.50, 1.000, 2.30, 1.38, 0]},
    "11": {"Audio": [109.0, 0.477, 3.80, 1.48, 1], "Non": [10.50, 0.857, 1.99, 0.42, 3]},
    "12": {"Audio": [134.0, 0.470, 3.29, 1.02, 0], "Non": [10.50, 0.905, 1.43, 1.09, 0]},
    "13": {"Audio": [141.0, 0.468, 11.63, 11.83, 1], "Non": [6.50, 0.846, 1.28, 4.63, 0]},
    "14": {"Audio": [134.0, 0.537, 7.21, 6.61, 3], "Non": [17.50, 0.629, 11.88, 11.68, 2]},
    "15": {"Audio": [145.0, 0.483, 5.42, 3.19, 0], "Non": [14.00, 0.536, 2.51, 1.89, 2]},
    "16": {"Audio": [100.0, 0.540, 7.67, 5.76, 3], "Non": [5.50, 1.000, 5.86, 3.34, 3]},
    "17": {"Audio": [204.0, 0.397, 7.65, 5.53, 1], "Non": [10.00, 0.850, 3.65, 2.65, 3]},
    "18": {"Audio": [124.0, 0.484, 6.08, 4.86, 1], "Non": [7.50, 0.867, 11.50, 13.39, 1]},
    "19": {"Audio": [118.0, 0.398, 6.61, 5.10, 0], "Non": [5.50, 1.000, 5.86, 6.76, 1]},
    "20": {"Audio": [208.0, 0.404, 6.15, 4.02, 1], "Non": [12.50, 0.760, 6.90, 7.27, 1]},
    "21": {"Audio": [185.0, 0.524, 10.09, 9.12, 0], "Non": [8.50, 0.882, 5.62, 5.91, 3]},
    "22": {"Audio": [120.0, 0.442, 4.47, 3.12, 1], "Non": [7.00, 0.786, 2.31, 2.26, 1]},
    "23": {"Audio": [110.0, 0.591, 7.54, 7.48, 0], "Non": [7.00, 0.786, 5.04, 7.44, 1]},
    "24": {"Audio": [204.0, 0.446, 5.80, 3.76, 0], "Non": [7.50, 0.933, 5.21, 2.08, 5]},
    "25": {"Audio": [107.0, 0.477, 4.82, 1.92, 1], "Non": [8.00, 0.938, -0.97, -3.46, 0]},
    "26": {"Audio": [139.0, 0.504, 9.51, 7.35, 9], "Non": [7.00, 0.929, 4.20, 3.74, 0]},
    "27": {"Audio": [148.0, 0.581, 8.28, 6.93, 0], "Non": [9.50, 0.789, 0.54, 1.32, 2]},
    "28": {"Audio": [159.0, 0.484, 8.99, 8.38, 0], "Non": [6.00, 0.833, 0.52, 0.41, 0]},
    "29": {"Audio": [169.0, 0.444, 5.04, 2.85, 0], "Non": [7.00, 0.786, -1.06, -0.44, 1]},
    "30": {"Audio": [151.0, 0.437, 4.98, 3.45, 1], "Non": [7.50, 0.800, -0.86, -0.41, 0]}
}

df = pd.read_excel(EXCEL_FILE)
results = []
actual_files = os.listdir(PHOTO_FOLDER) if os.path.exists(PHOTO_FOLDER) else []

print(f" Running MASTER Injector (Improved Matching)...")

for index, row in df.iterrows():
    # Robust Participant ID cleaning
    raw_id = str(row['Participant #']).split('.')[0].strip().upper().replace('P', '').lstrip('0')
    cond = str(row['Condition']).strip().lower()

    row_data = row.to_dict()
    v_err, t_err = 0, 0

    # Match against dictionary
    if raw_id in MASTER_CHART_DATA:
        data = MASTER_CHART_DATA[raw_id]
```

```
        if "audio" in cond:
            row_data['WPM'], row_data['TTR'], row_data['FK Grade'], row_data['ARI Grade'], v_err =
data["Audio"]
        elif "non" in cond:
            row_data['WPM'], row_data['TTR'], row_data['FK Grade'], row_data['ARI Grade'], t_err =
data["Non"]

    # Image Analysis logic
    f_cv, i_cv, l_cv = 0.0, 0.0, 0.0
    if "audio" not in cond:
        s_cond = "Non" if "non" in cond else "Dom"
        match_f = next((f for f in actual_files if f"P{raw_id}_" in f.upper() and s_cond.upper() in f.upper()),
None)
        if match_f:
            img = cv2.imread(os.path.join(PHOTO_FOLDER, match_f), 0)
            if img is not None:
                _, th = cv2.threshold(img, 0, 255, cv2.THRESH_BINARY_INV + cv2.THRESH_OTSU)
                cnts, _ = cv2.findContours(th, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)
                hts = [cv2.boundingRect(c)[3] for c in cnts if cv2.boundingRect(c)[3] > 5]
                if hts: f_cv = round((np.std(hts)/np.mean(hts))*100, 2)
                pxs = img[th > 0]
                if len(pxs) > 0: i_cv = round((np.std(pxs)/np.mean(pxs))*100, 2)
                # 4-factor Legibility
                bxs = [cv2.boundingRect(c) for c in cnts if cv2.contourArea(c) > 30]
                if len(bxs) > 5:
                    y_p = [y+(h/2) for (x,y,w,h) in bxs]
                    rts = [w/h for (x,y,w,h) in bxs]
                    ars = [w*h for (x,y,w,h) in bxs]
                    dns = [np.count_nonzero(th[y:y+h, x:x+w])/(w*h) for (x,y,w,h) in bxs]
                    l_cv = round(((np.std(y_p)/np.mean(y_p)) + (np.std(rts)/np.mean(rts)) +
(np.std(ars)/np.mean(ars)) + (np.std(dns)/np.mean(dns)))*25, 2)

    row_data["Friction Score (CV)"] = f_cv
    row_data["Ink Intensity (CV)"] = i_cv
    row_data["Legibility Rating:"] = l_cv
    row_data["Verbal Disfluencies"] = v_err
    row_data["Transcription Errors"] = t_err
    results.append(row_data)

pd.DataFrame(results).to_excel("Step_3_Final_Data.xlsx", index=False)
print("\n🏁 Master Sheet Finished! Please check 'Step_3_Final_Data.xlsx'")
```

ABOUT THE AUTHOR

Sophie Baryalai is a rising senior at Lake Travis High School in Lakeway, Texas. Her academic interests span English, writing, and cognitive science, and she is drawn to questions about how the brain processes and produces language. She plans to study English in college, with broader interests in education, the social sciences, and the neurological foundations of human learning.