

# The Discipline of Handwriting Quality and Essay Scoring

Ryan Dinh<sup>1</sup> and Ana Dinh<sup>1#</sup>

<sup>1</sup>Garden Grove High School, Garden Grove, CA, USA

#Advisor

## ABSTRACT

In the United States, many high school students take various Advanced Placement (AP) courses offered by the non-profit organization, CollegeBoard. Specifically, AP English Language and AP English and Literature requires students to write three essays on the day of exam. These essays allow students to demonstrate their penmanship and quality of handwriting. However, it is unknown whether the quality of penmanship plays a role in essay scoring. Very few research has been done on this discipline, specifically handwriting and intelligence. To fill this research gap, this paper will explore and attempt to demonstrate an understanding of the potential correlation between the illegible ratio of handwriting to the overall essay rubric score. 68 samples were taken from CollegeBoard and analyzed through a content analysis with a mathematical equation, the ILR equation. The data was further analyzed with a Line of Best Fit, Pearson Correlation Coefficient, and P-value. The results showed that there were no clear significant differences between neat and messy handwriting and how well it scored on the AP exam. However, the overall length of the essay and its word counts had a strong positive correlation with the essay score. Due to various limitations of this research study, further research must be conducted and tailored to this study to find a clearer understanding of this phenomenon. Nevertheless, this research can serve as the basis and foundation for later studies in this field.

## Introduction

Handwriting is simple, delicate, yet complicated. Also known as penmanship, it requires a high degree of coordination and practice before it is established as an acquisition (Gesell, 1906). Penmanship is unique and can come in many forms and variations. From a very young age, people tend to develop their own styles of handwriting, whether print or signature. There are many factors that contribute to the uniqueness of handwriting such as the roundness or sharpness of letters, the varied spacing, the shape, the pressure of the writing instrument to the paper, the average size, and the thickness of the letters (Srihari, Huang, Srinivisan, 2008). Handwriting is used in daily life, whether making shopping lists, creating works of art, copying down notes from a business meeting, or more specifically writing for a school essay.

Across the United States, many students take many Advanced Placement (AP) courses that are offered by the non-profit organization CollegeBoard. These courses are rigorous and help students develop skills including the ability to break down pieces of rhetoric, arguments, insights, and narratives. All AP courses offer exams every year during May. Students can take the exam and if they pass, they can obtain college credit for the same courses that would be otherwise taken at the college itself. In many of the AP courses such as AP English and Composition and AP English and Literature, these exams require students to write three essays about the content and skills they were taught throughout the school year. As stated earlier, the ability to write is required to answer these essay questions. Since every student has their own distinctive penmanship, this study aims to find a correlation between the uniqueness of handwriting, more specifically the ratio of illegible handwriting, to the score they received on their essay. Overall, the project goals include determining whether there is a correlation between messier handwriting and higher essay scores in AP English classes. Since the relationship between handwriting and intelligence is very complex and vague,

by researching a much smaller topic such as penmanship, I hope to discover an abstract yet concrete relationship with fuller meaning about general intelligence.

## Literature Review

### The Variables of Penmanship

Arnold L. Gessel, American psychologist and professor at Yale University, conducted a study. In his study, he discovered that children's handwriting varied with intelligence, gender, and mental willingness. But rather, it indicates a kind of intelligence rather than a grade of intelligence. Gessel describes this kind of intelligence as the "relative accuracy of the boys' and girls' writing movements... and relative willingness and capacity to maintain such a visual consciousness (Gessel, 1906)." In his study of a thousand students, he separates them into four distinctive groups. In Group III and Group IV, 64.6% of the best penmanship was written by girls, and 70% of the worst penmanship was written by boys. The majority of the teachers report the girls were twice as much pain staked in writing compared to boys and reported boys were four times more careless than the number of girls. Gessel also noticed that painstaking writing qualities did not correlate to exceptional writing. Sixty-five students were classified careless in writing but exceptional otherwise, and forty-two were classified as painstaking in writing yet unexceptional. However, those numbers should be taken in light. The majority (90%) followed the pattern where "painstaking or careless qualities in writing bespoke similar qualities in general schoolwork." Overall, painstaking or careless qualities in writing show similar qualities on school intelligence, sex differences, and accuracy in handwriting. Results from Gessel's studies introduce the question whether handwriting quality has a strong influence on one's intelligence.

### Accountability of Handwriting Speed

Prevely, a neuropsychologist at Columbia University, describes the characteristics of a well-versed writer. A great writer must generate ideas quickly and write these ideas as quickly as possible in order to use what he calls the "metacognitive process." This process includes the ability to plan, translate, and revise one's idea into a complex written composition. However, the speed of handwriting also plays a role in the length and quality of their written compositions. Transcription speed is based on the availability of information and levels of monitoring recall to compete for the resources of a limited capacity working memory (WM). The WM can temporarily store information in the cognitive system and acts as the key function for reasoning and guidance of decision-making behavior. In a 2002 study, researchers found that "adults with naturally slow handwriting may be hampered in their ability to execute higher level processes because of the strain placed on WM... (Prevely, 2006)." The significance of this study shows that slower handwriting may correlate to a limited working memory. Thus, it is implied that writers with slower transcription speed have a weaker metacognitive process and are unable to write as effectively. In a separate 2005 study, results indicated that handwriting speed "correlated positively and significantly with tutors' marks, overall number of words written, and the overall rubric score." Thus, faster handwriting may correlate to a higher essay grade. With this research finding, I can assume that faster handwriting is messier, therefore more illegible compared to slower handwriting. In essence, messier handwriting is a sign of higher scores and WM process which symbolizes a kind of intelligence, as stated in the previous section.

### Contradictory Studies on the Benefit of Doubt

Powers, Foweles, Farnum, and Ramsey, members of the Educational Testing Services who aim to advance learning and education, conduct a study between handwritten essays and word-processed essays. Essentially, word processed essays are neater and longer than their pen and paper counterparts. Hypothetically, the word-processed version of a

student's essays should score higher, but that is incorrect. The members found that “contrary to initial expectations, student papers that were converted to word-processed versions received lower scores than did the original handwritten versions (Powers, 1994).” Much more researchers have also found that “most essay readers actually preferred scoring the handwritten versions over the word-processed essays, even though the former were more difficult to read.” This is because most graders gave the students the “benefit of the doubt (BTD)” when they read essays with more illegible handwriting. Thus, these graders “often mentally transformed students' texts or filled in perceived gaps when reading handwritten essays.” To confirm these findings from previous researchers, Powers and his group members also conducted their own study. They also found that graders are essentially more lenient in scoring messier handwriting samples compared to their word-processed counterparts. “Thus, it was hypothesized that there were probably fewer instances of unreadable or difficult-to-read words among the transcribed handwritten essays than among the original handwritten essays of examinees, and thus there was less opportunity to give writers the benefit of any doubt.” The variable BTD must be considered in essays and may vary from one essay to another. It is concluded that essay grading can be more subjective than objective.

On the other hand, the studies of Hughes, Keeling, and Tuck have questioned the correlation between handwriting quality and essay scores. Although numerous research traced back to the 1920s investigated the connection between essay grades and handwriting quality, their findings of their study are inconsistent with those of the studies that were previously reported. Hughes' present study of thirty-eight high school students revealed that there was no recorded evidence of quality of penmanship and achievement expectations. As suggested, more research is required to establish the reliable connection of handwriting quality and scorer achievement expectations.

## Varying Perspectives

As investigated, many studies including the studies of Preverly and Powers discuss that untidy handwriting may have a higher chance for a higher essay grade. Messy penmanship can be said to be two birds in one stone: it shows that the student may have a higher WM and can assist them by giving the scorers the benefit of the doubt. On the contrary, many other studies such as Gessel and Hughes et. al say otherwise. Overall, these studies show mixed results and do not create one big picture for us to comprehend. Therefore, as my job as a researcher, I aim to find a correlation between handwriting quality and essay scoring to prove and disprove their past findings.

## Final Hypothesis

I hypothesize that messier handwriting is rather a kind of intelligence and cannot be used as a standardized scale to measure one's intelligence. Also, messier handwriting may influence the scorer's expectations and give them the benefit of the doubt, therefore leading to higher essay scores.

## Methodology

The overall project goal of this study is intended to determine whether there is an existing correlation between penmanship quality and essay scores. The research that will take place will utilize a cross-sectional study. The cross-sectional study falls under the category of observational study designs and is used to measure the outcome and exposure of the participants (Setia Maninder, 2016). It is like case-control studies and cohort studies except for the fact that all the participants were selected and studied at the same time. This method was chosen because by picking participants who will be graded on their timed essays under the same testing conditions, the cross-sectional study is the only way to study their exposure and outcome at the same time to estimate the prevalence of the correlation between penmanship quality and essay scores. Since their testing environment and essay scores are standardized, the

cross-sectional study will not require any manipulation of the dependent variables, allowing me to look at a variety of features of the participants' handwriting to find the prevailing characteristics in the population sample.

## Design Overview

This study has four parts which can be replicated by anyone and anytime: sampling, content analysis, analysis of the data, and interpretation of the data.

### *Sampling*

AP English and Language and AP English and Literature categories. By using essay samples from CollegeBoard, it allows my data to be standardized, removing the risk of bias or potential skewed data. The exam questions are standardized, meaning that none of the students have any previous knowledge regarding the exam questions, and had to produce their papers with the skills they have mastered throughout the school year. The people who grade the essays are trained professionally by the CollegeBoard organization, which removes any bias from essay grading and confirms every student gets the score corresponding to the mastery of skill demonstrated. For this research, 68 essays were sampled from both AP courses between the years 2014-2019. This specific time period is essential to this method because all the exams after 2019 changed their scoring rubric from a 9-point prong to a 6-point prong. Looking only at the exams after 2019, there are 18 essay samples from 2021 and 2022 which results in a confidence level less than 95%; therefore, the essay samples must be from the 2014-2019 exams in order to ensure a greater representation.

### *Content Analysis*

The second part of this study uses correlational content analysis in order to view the relationship between handwriting quality and essay scoring. Many of the past studies, such as Powers, et. al, Preverly, and Hughes, et. al, used experimental methods, such as having a control group and changing the dependent variables. Significantly, these studies resulted in mixed findings. By using a different research method like the content analysis, it enables me as a researcher to hopefully discover new findings concerning past studies in the literature review. With the content analysis, the qualitative data of handwriting can be converted into quantitative data in order to examine the relationship numerically. A mathematical equation, the ILR (Illegibility Ratio) score equation, will be used in order to conduct the content analysis. This equation will be explained in the Results section. The ILR score as well as other variables such as Word Count, Letter Size, and Letter spacing will be compared to the ES (essay score) because these variables are factors in the quality of handwriting. Thus, a potential correlation can be discovered between penmanship and essay writing

### *Analysis of Data*

The third part of this study includes analyzing the numbers provided from the ILR equation. On the Desmos Graphing Calculator program, the ILR, Word Count, Letter Size, and Letter Spacing will be compared individually to the ES graphically on the x-y plane. This allows for the usage of the Line of Best fit, accompanied with the Pearson product-moment correlation coefficient. After analyzing the data through those two methods, a p-value will be calculated. The p-value will reveal whether the relationship and results are statistically of major theoretical, clinical, or practical importance.

### Interpretation of the Data

The final part of this study uses the strength of the correlation, determined by the Line of Best Fit, Pearson product-moment correlation coefficient, and the p-value in order to answer the research question of this study. A strong relationship between penmanship quality and essay scoring will be accompanied by a strong Line of Best Fit, high Pearson coefficient, and a low p-value, and vice versa.

### Results

To analyze the results to determine whether there is an existing relationship between penmanship quality and essay scores, the ILR was compared to the essay score (ES). The ILR was calculated with the mathematical formula which can be seen in Figure 1. All the qualitative data, which is in Figure 2, was converted into quantitative data, which was used to record all the factors that affect the ILR, including illegible words, letter size, and letter spacing. The whole process is described in Figure 3. Originally, the method sought to gather the data by means of a content analysis with 108 sample papers. However, data collection stopped at the count of 68 sample papers. This is since as more and more data was collected over time, there was an obvious conclusion that there was no clear correlation.

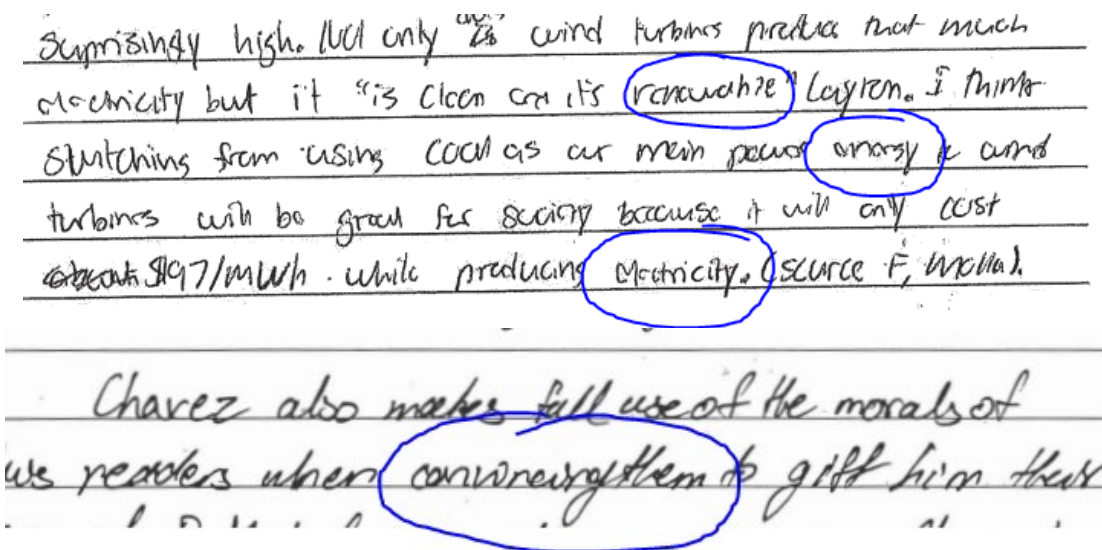
$$ILR = \frac{a}{a - b} (c \cdot d)$$

**Figure 1: ILR Equation**

ES = Essay Score  
a = Word Count  
b = # of Illegible Words

Definition of Illegible: Impossible or very difficult to read.  
Range: (1-a)

The images below are some illustrative examples of illegible handwriting, variable "a". The illegible words are circled in blue. Each illegible word is counted as one point towards the ILR equation.



C = Letter Size

Range: (1-5)

Score of 1 has big handwriting style

Score of 5 has small handwriting style

The images below are some examples of how the Letter Size was scored.

Score	
1	his readers to learn from the past, "who gets killed in the case of violent revolution? The poor, the workers." (78-79). In most cases the people
3	When the treatment of the black man within the United States became so abhorrent, so disturbing to the everyday security and pursuit of happiness for these individuals,
5	you may not know if she is or isn't there but she knows you better than you think.

D = Letter Spacing

Range: (1-3)

Score of 1 has spacious handwriting style

Score of 3 has compact handwriting style

The images below are some examples of how the Letter Spacing was scored.

Score	
1	throughout its history and even goes back to when it wasn't even the
2	specified the main audience of the poem as fathers. By incorporating both these literary techniques, Harper
3	A family unit can have a profound impact on people's lives. It can be the glue that holds them together, or the chaos that tears them apart. Family

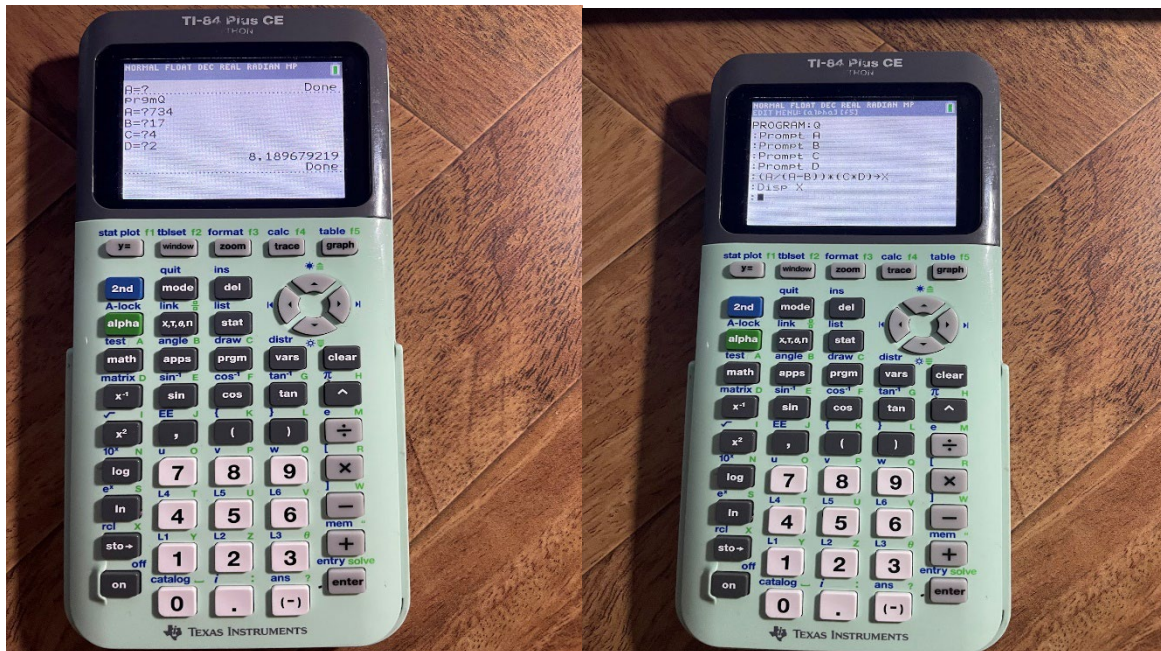
Figure 2: Data

AP LANG#1	ILR	ES	Word Count (A)	Illegible Word (B)	Letter Size (C)	Spacing (D)
2019 #1A	8.19	8	734	17	4	2
2019 #1B	2.019	6	656	6	2	1
2019 #1C	10.397	4	536	72	3	3
2018 #1A	8.181	8	902	20	4	2
2018 #1B	2.014	6	436	3	2	1
2018 #1C	1.00248	4	403	1	1	1
2017 #1A	2.00286	8	699	1	2	1
2017 #1B	2.0034	6	588	1	1	2
2017 #1C	1.00279	3	359	1	1	1
2016 #1A	8.13285	8	857	14	4	2
2016 #1B	1.00122	6	819	1	1	1
2016 #1C	2.0043	4	465	1	2	1
2015 #1A	8.0111	7	720	1	4	2
2015 #1B	1.001956	4	512	1	1	1
2015 #1C	2.076	2	273	10	2	1
2014 #1A	1.00154	8	648	1	1	1
2014 #1B	4.0081	6	495	1	2	2
2014 #1C	1.00154	4	648	1	1	1
2019 #2A	6.1314	9	980	21	3	2
2019 #2B	8.01	5	918	1	4	2
2019 #2C	1.00571	1	176	1	1	1
2018 #2A	12.456	8	546	20	4	2
2018 #2B	1.002739	5	366	1	1	1
2018 #2C	9.3957	2	276	41	1	2
2017 #2A	1.00178	8	562	1	2	1
2017 #2B	8.6886	5	429	34	4	2
2017 #2C	3.0349	2	434	5	1	3
2016 #2A	6.4135	8	760	49	3	2
2016 #2B	1.00143	6	699	1	1	1
2016 #2C	2.009708	2	207	1	2	1
2015 #2A	6.5644599	8	628	54	2	3
2015 #2B	1.0054	6	553	3	1	1
2015 #2C	1.54437	3	522	15	1	1.5

<b>AP LANG#1</b>	<b>ILR</b>	<b>ES</b>	<b>Word Count (A)</b>	<b>Illegible Word (B)</b>	<b>Letter Size (C)</b>	<b>Spacing (D)</b>
2014 #2A	1.001620746	8	618	1	1	1
2014 #2B	2.005633	6	712	2	2	1
2014 #2C	1.002849003	3	352	1	1	1
2019 #1A	2.003527337	8	568	1	2	1
2019 #1B	6.18181	5	646	19	3	2
2019 #1C	15.33149	2	185	4	5	3
2018 #1A	6.16216	8	638	1	3	2
2018 #1B	2.04926	5	416	10	1	2
2018 #1C	1.004784	3	210	1	1	1
2017 #1A	6.195888	8	854	27	3	2
2017 #1B	6.101204819	5	422	7	3	2
2017 #1C	1.00336	3	298	1	1	1
2016 #1A	12.31578	9	624	16	4	3
2016 #1B	3.001918	6	472	3	3	1
2016 #1C	10.22727	2	180	4	5	2
2015 #1A	1.001975	9	528	1	1	1
2015 #1B	2.003327	6	602	1	2	1
2015 #1C	4.141176	3	264	9	2	2
2014 #1A	3.04373	8	696	10	3	1
2014 #1B	4.0062	6	637	1	4	1
2014 #1C	10.5071	4	663	32	5	2
2019 #2A	9.05844	9	620	4	3	3
2019 #2B	4.00873364	6	459	1	2	2
2019 #2C	2.00431	4	464	1	1	2
2018 #2A	12.5786	7	413	19	4	3
2018 #2B	1.00252	5	397	1	1	1
2018 #2C	5.1503	2	137	4	5	1
2017 #2A	2.02898	7	700	10	1	2
2017 #2B	2.00309119	5	648	1	2	1
2017 #2C	1.009219	3	368	3	1	1
2016 #2A	1.001490313	9	672	1	1	1
2016 #2B	5.025706941	6	782	4	5	1
2016 #2C	3.01639443	2	184	1	3	1
2015 #2A	2.07854	7	739	27	1	2



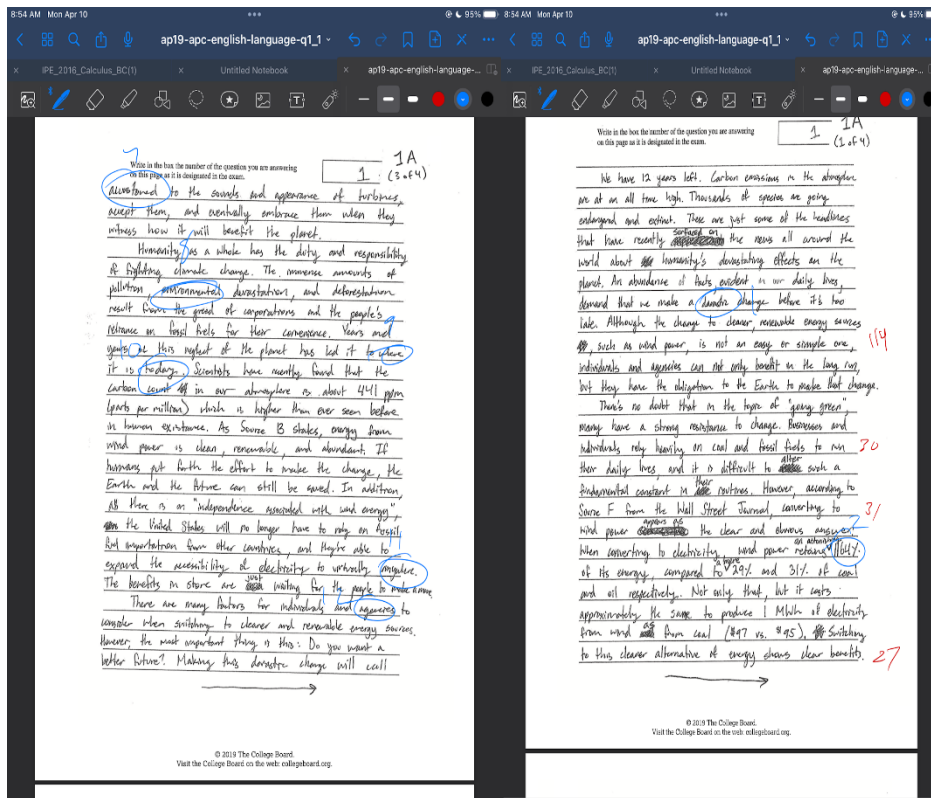
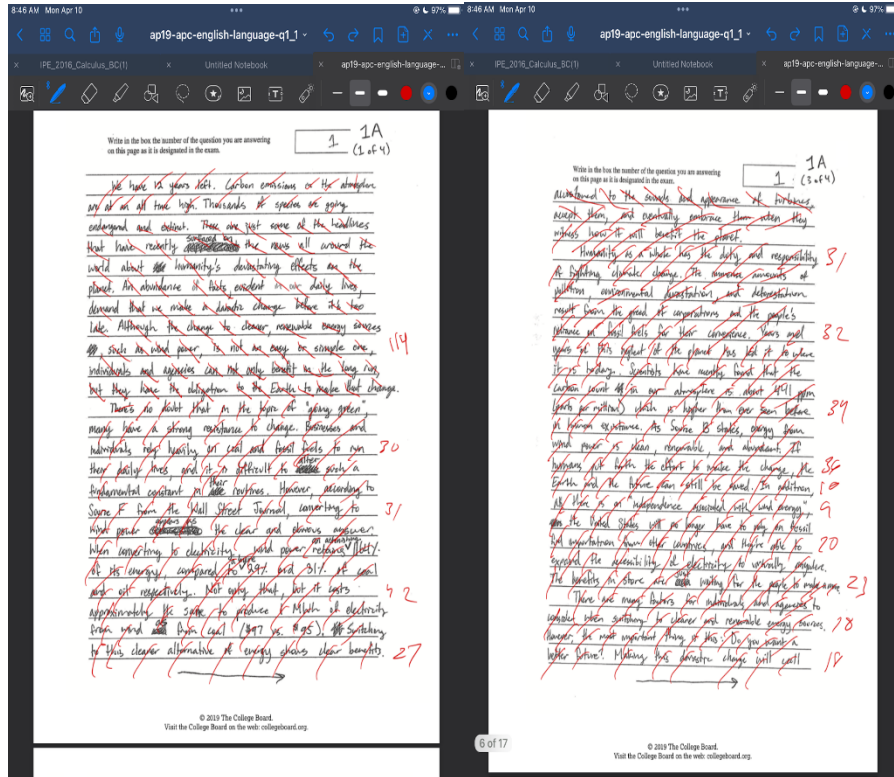
AP LANG#1	ILR	ES	Word Count (A)	Illegible Word (B)	Letter Size (C)	Spacing (D)
2015 #2B	2.0028059	5	714	1	2	1
2015 #2C	6.1237	3	297	6	3	2
2014 #2A	1.0916	8	728	14	1	1
2014 #2B	6.0084	6	720	1	2	3
2014 #2C	4.01459854	4	550	2	2	2



**Figure 3: Data Calculation**

The data was calculated in the programmed equation with a Ti-84 Plus CE Python.

The next 4 images show how the content analysis was conducted.

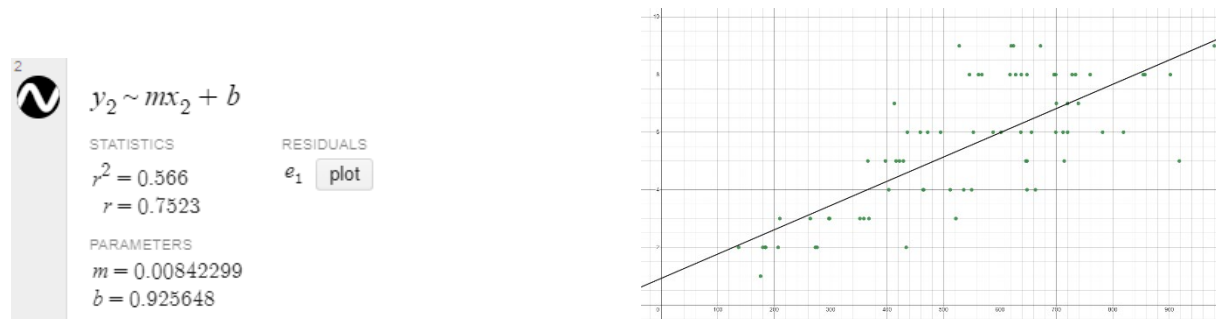


Next, the ILR and ES numbers were graphed on the graphing website program on <https://www.desmos.com/>. The data was presented on the graph and then was analyzed through the Pearson correlation, which will be discussed in the next section. With the built-in Desmos function called the “Line of Best Fit,” a linear function was generated in **Figure 4**.

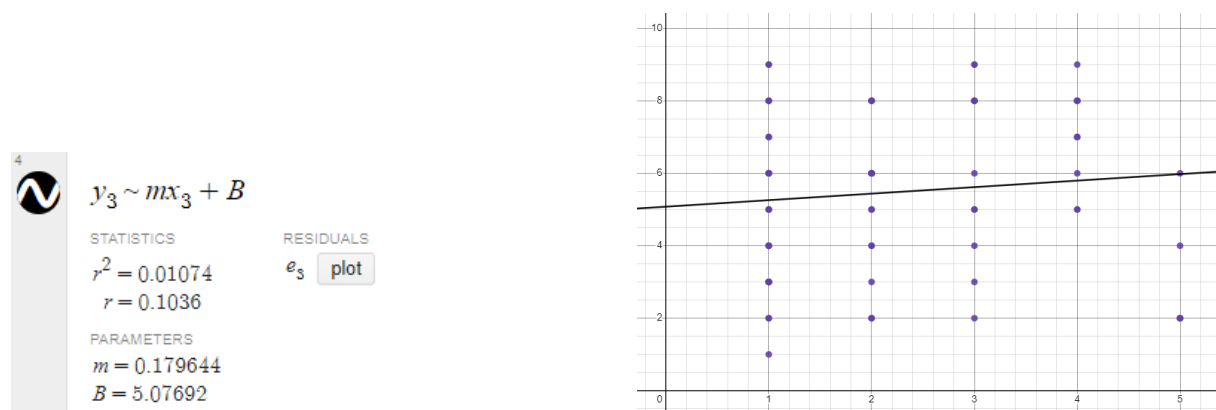


**Figure 4: ILR v. ES**

Additionally, other factors that may affect the ILR such as word count, letter size, and letter spacing, will be compared independently to the ES to see if there are any further potential external correlations or outside findings. The Pearson correlation was used to determine the strength of the correlation for Word Count v. ES (Figure 5), Letter Size v. ES (Figure 6), and Letter Spacing v. ES (Figure 7).



**Figure 5: Word Count v. ES**



**Figure 6: Letter Size v. ES**

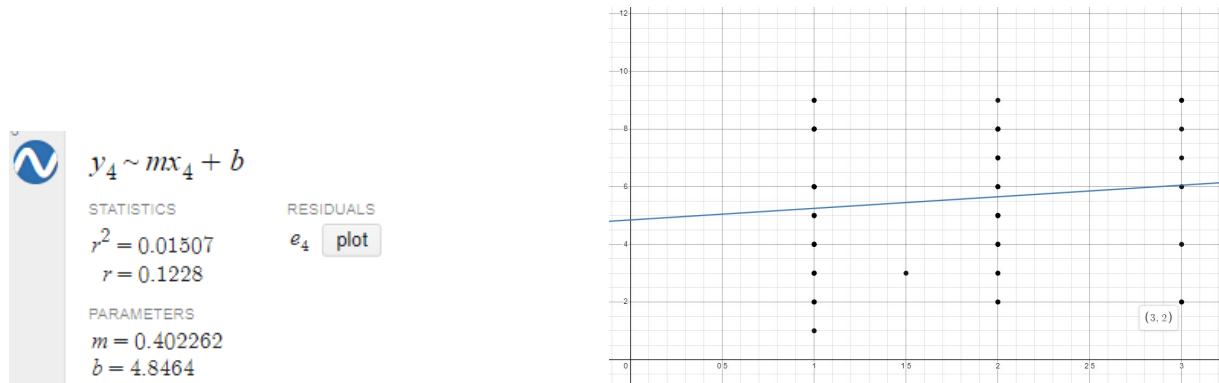


Figure 7: Letter Spacing v. ES

### What is the Pearson product-moment correlation coefficient?

The following data that was sampled from the 68 essays were mathematically translated with a statistical correlation known as the Pearson product-moment correlation coefficient. Also known as the Pearson correlation for short, it measures the strength of the linear relationship between two run random variables, represented by the symbol  $r$ . The Pearson correlation coefficient, or  $r$ , measures how far away all of these data points are from the line of best fit that a Pearson product-moment correlation seeks to draw across the data of two variables (Laerd Research, 2018). Typically, a Pearson coefficient with a strength of the range 0.1 to 0.3 shows a weak association between the two variables. Additionally, a range of 0.3 to 0.5 and the final range of 0.5 to 1.0 will result in a medium association and a strong association, respectively. In this research, all  $r$  values will be a positive value since there are no real negative values. Accompanied by the  $r$  value, the  $r^2$  value, also known as the coefficient of determination, will measure the percentage of variability within the y-values associated with the line of best fit (Penn State, 2023). Thus, an  $r^2$  value of 1.0, which is 100% in terms of percentage, strongly indicates that the line of best fit is a useful model. While on the other hand, an  $r^2$  value of 0, or 0%, indicates that the line of best fit is a useless model.

### Findings

Essentially, the Pearson correlation attempted to reveal an association between the two variables, ILR and ES (Figure 4) with the data that had been provided. The Pearson correlation created the “line of best fit” with the linear function  $y = mx + b$  with the parameters as  $m = 0.0639187$  and  $b = 5.2164$ . It also provided the value of  $r$  and  $r^2$  as being  $r = 0.1028$  and  $r^2 = 0.01056$ , measuring the average distance between the scattered data points and the actual line of best fit. With the data that was collected with the essay samples from AP Language and Composition students and AP Literature and Composition students from the years 2014 through 2019, the  $r$  number of 0.1028 reveals that there is no significant correlation between penmanship quality and the essay score that students receive. The  $r^2$  value of 0.01056 also shows that there are no good models for the data that was provided.

In Figure 5, the data for Word Count v. ES was plotted on a separate Desmos plane. On the graph, the data points were relatively close together, creating a noticeable pattern. Next, The Desmos program generated a separate Pearson correlation for the data. With an  $r$  value of 0.7523 and  $r^2$  value of 0.566, the Pearson correlation suggests that there is a strong relationship between the length of an essay and the score it received. As repeated, the data from Figure 6 for Letter Size v. ES ran through the Pearson correlation and received the  $r$  value of 0.1036 and  $r^2$  value of 0.01074. In Figure 7, the Pearson coefficient for Letter Spacing v. ES was 0.1228 and the correlation coefficient value was 0.01507. The Pearson coefficients and correlation coefficient values for Letter Size v. ES and Letter Spacing v. ES both reveal that potential factors that may affect legibility, when examined independently, do not also show a strong relationship with its corresponding essay score.

## Discussion

This research has been conducted via Pearson product-moment correlation. Overall, the sum of the two types of coefficients,  $r$  and  $r^2$ , indicates that there is a weak relationship between the ILR and ES. Placed into context, the quality of the penmanship does not play a role in whether the student receives a proficient score. The data has proven that most students who do receive high scores on these essays have mixed styles of handwriting, and that one specific type of handwriting does not correlate with a higher intellectual ability. The findings of this study agree with the studies conducted by Hughes, Kneeling, and Tuck, which suggest that there is no evidence of penmanship quality and achievement expectations. In this study, the ILR and ES had a  $r$  coefficient of 0.1028, which, as mentioned repeatedly, showed almost no correlation between the two variables. The results for ILR v. ES generally agree with the studies by Hughes and his other team members, which prove that there may be no correlation whatsoever. However, the results for ILR v. ES have shown to disagree with the studies of Preverly and Powers, which both researchers have suggested that messier handwriting may have a higher chance to receive a higher score. This creates conflict, in whether whose research is reliable and not reliable. To improve on these findings, further extensive research must be done.

Although the results for ILR v. ES in a general sense would disagree with the research by Powers, the results for Word Count v. ES depict a completely different story. As stated in Powers' study, students with faster handwriting have better working memories (WM) that can execute higher level processing tasks, compared to students with slower handwriting. Their study also indicated that handwriting speed had a strong correlation with the overall word count and rubric score. This can be confirmed by the results in this research. For the Pearson coefficient for Word Count v. ES, the  $r$  value was 0.7523, which strongly suggests that there is a significant association between the length of the essay and the score it received. Thus, in most cases, students who write longer papers often receive higher scores. To think it from a logical stance, students who write longer papers will have faster handwriting since they have a strong working memory; thus, they can process information quicker and write more effectively. One point that needs to be mentioned is that many people assume that students who write messier usually write faster and score higher. Although in some cases this may be true, many essay samples that have a higher ES score and word count also have a low ILR score. The significance of this finding is that some students can have a fast transcription speed while maintaining their quality of penmanship. Overall, the data and results from this study has proven that a higher word count correlates with a higher essay score, which in turn, can be interpreted as a higher intellectual ability.

Power also suggests that papers who are more illegible tend to be given the advantage of the benefit of the doubt (BTD). However, this part of Power's study cannot be proved nor disproved. In this research, there have been many essay samples that do have a higher ILR but still maintain a low ES. For example, in AP Lit 2016 #1A, this essay received an ILR of 12.31578 and an ES of 9, but in AP Lit 2014 #1C, this essay received an ILR of 10.5071 and an ES 4. Although these ILRs are relatively high, it doesn't necessarily mean that the grader will give them the benefit of the doubt and be less harsh on these graders. This is probably since Advanced Placement graders are trained and taught what parts of the essay to give points for, and what parts of the essay to not give points for. Therefore, all Advanced Placement essays that are graded fairly, without any bias, will receive their score based on their content and nuanced responses. This is conflicting since in this research; the essays are graded by standardized tutors. In many different schools and administrations, they probably do not train the people who grade the essays to an extensive level, so they may have different opinions and expectations for their writers. To sum it up, none of the findings in this study can confirm the theory of the benefit of doubt.

Finally, the p-values in Figure 8 suggest that the null hypothesis for ILR v. ES, Letter Size v. ES, and Letter Size v. Spacing failed to be rejected, as due to the reason that these 3 p-values are greater than 0.05. Therefore, there may not be enough significant evidence to support the findings for this research. However, the p-value of Word Count v. ES was less than 0.05 meaning that the observation is supported and can be used to support Preverly's study.

Figure	P-value
ILR v. ES (Figure 3)	0.18641878
Word Count v. ES (Figure 4)	0.000000000113732
Letter Size v. ES (Figure 5)	0.18695048
Letter Spacing v. ES (Figure 6)	0.18695048

**Figure 8: P-Values for Figures 3, 4, 5, & 6**

## Hypothesis

As stated in the introduction of this research paper, an early hypothesis was formed. The hypothesis predicted that a messier handwriting would have a positive correlation with the essay score and can be used as a tool for intelligence. However, the findings of this study have rejected the original hypothesis. The results of this study have concluded that there is no clear correlation between the quality of penmanship and the essay score it received. However, the data has proved that students who wrote more and had a higher working memory score tend to receive higher marks on their papers.

## Conclusion

In conclusion, the findings and results of this paper tried to address the research gap. As what has been discussed, the evidence proved that there is no clear correlation between the legibility of handwriting and how well a paper scores on the AP English and Language and AP English and Literature exams. However, the data suggest that there is a strong relationship between the word count of students' essays and how they perform as a whole. Thus, the recent findings from these AP English exams lean towards the argument that handwriting cannot be used as a tool to measure one's intellectual ability.

## Limitations

Nevertheless, there are some limitations to this study. First, the data that was collected and converted with the ILR mathematical equation is very generalized. This equation tries to balance four variables, which is normally difficult to measure a clear correlation. Normally, there are only two variables in correlation and regression. But adding two confounding variables only results in a high p-value, thus making this research skewed to a certain extent. The equation was also created through trial and error with logical considerations, without any professional justification like a math proof would, making it overgeneralized. Second, only 68 samples were gathered which is a smaller sample size. This means that the data that was collected is not guaranteed to be representative. Similarly, College Board only uploads only 3 sample papers for each question every year, which questions the availability of the samples that were not uploaded. Third, the results may be biased. What is considered illegible, legible, big, small, or etc., in this research may be different to others. Therefore, these findings are not standardized and only appear to be a loose generalization.

## Future Research

For future researchers who do want to further explore this field, there are gaps that still need to be addressed. For instance, what would change if the essays were non-standardized unlike CollegeBoard essays? Further study needs to be conducted to answer this question and expand the results of this study. Additionally, a better mathematical equation must be created to obtain a better understanding about the characteristics of handwriting and essay scoring. An equation that can manipulate four or more variables without any huge errors or fallacies is ideally what is wanted for this type of study. These suggestions should be left to the professionals who have more knowledge and experience.

## Implications

Starting this year, CollegeBoard has offered students the opportunity to take their AP exams digitally in schools instead of the usual paper tests. On the paper tests though, every exam has a written response portion and requires students to handwrite their responses. These questions are known as FRQs SAQs, DBQs, and LEQs. However, the controversy behind the switch from paper to computer can be correlated to Powers, et. al studies where word-processed versions scored worse than their paper counterparts. This begs the question, will the students who take their exams digitally be offered the same opportunity to display their mastery like students who do take old-fashioned paper exams? Only if this research proved that there was indeed a strong relationship between the quality of handwriting and essay scoring, it would greatly benefit CollegeBoard to debate the pros and cons for switching to digital exams.

## Acknowledgements

I would like to thank my advisor, my AP Capstone Teacher, Mrs. Dinh who gave me invaluable advice and supported me throughout the whole process. Her motivation and dedication for her students like me has helped contribute to the successful completion of this research paper.

## References

- Cohen, I., Huang, Y., Chen, J., Benesty, J., Benesty, J., Chen, J., ... & Cohen, I. (2009). Pearson correlation coefficient. *Noise reduction in speech processing*, 1-4.
- Gesell, A. L. (1906). Accuracy in handwriting, as related to school intelligence and sex. *The American Journal of Psychology*, 394-405.
- Hughes, D. C., Keeling, B., & Tuck, B. F. (1983). Effects of achievement expectations and handwriting quality on scoring essays. *Journal of Educational Measurement*, 65-70.
- Laerd Statistics. (2018). Pearson product-moment correlation coefficient. <https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php>
- Peverly, S. T. (2006). The importance of handwriting speed in adult writing. *Developmental Neuropsychology*, 29(1), 197-216.
- Powers, D. E., Fowles, M. E., Farnum, M., & Ramsey, P. (1994). They Think Less of My Handwritten Essay If Others Word Process Theirs? Effects on Essay Scores of Intermingling Handwritten and Word-Processed Essays. *Journal of Educational Measurement*, 31(3), 220-233.
- Srihari, S., Huang, C., & Srinivasan, H. (2008). On the discriminability of the handwriting of twins. *Journal of Forensic Sciences*, 53(2), 430-446.