# The Impact of Large Language Models on Applicants' Cover Letter Scores and the Moderation Effect of Language Nativity

Ceren Kalkir

SNR: 2100029

Tilburg University

M.Sc. Work and Organisational Psychology

2023/2024

Supervisor: Dr. Djurre Holtrop

2nd reader: Dr. Bence Bago

April 7, 2024

**Abstract**

The use of the Large Language Model in the cover letter writing process by candidates increased the ethical considerations in the recruitment process. However, the outcomes of the use of the Large Language Model (LLM) by the candidates in recruiters' evaluations are limited. Additionally, LLM usage may challenge the positive effect of language nativity on the recruiter evaluations. Therefore, this research aims to answer the following: How does the use of various types of LLMs impact recruiter evaluations of Cover Letters, and is this impact influenced by language nativity? This study utilises an experimental design and two-way repeated measures ANOVA to assess the recruiters' evaluations for various cover letters representing one unique combination of Chat GPT usage types (within-subjects) by native and English as a second language groups (between-subjects). Candidates can create a new cover letter or enhance their prewritten cover letter in the cover letter writing process with the utilisation of LLMs. Recruiter evaluations were measured using the Hireability Index. Qualtrics survey tool used to collect the recruiters' evaluations on cover letters. Forty-six final responses were included in the analysis. Recruiters' cover letter evaluations increased from no Chat GPT-used cover letters to Chat GPT enhanced and Chat GPT created cover letters. However, the recruiters' evaluations did not show a significant difference between Chat GPT enhanced and Chat GPT created cover letters. The interaction effect of language proficiency and Chat GPT usage was insignificant. The main effect of language proficiency on cover letter scores was significant, with higher cover letter scores for Native speakers than for English as a second language (ESL) speakers. Post-hoc analysis revealed that recruiters' evaluations are higher for native speaker candidates than ESL group candidates when no LLM is utilised. The difference between ESL and native groups is insignificant when candidates utilise LLMs by creating or enhancing their cover letters.

## Introduction

The cover letter is often the initial business communication from a candidate. The cover letters written by a candidate affect the organisation's decision, which leads to the invitation of the right candidate for an interview (Watts, 2015, p. 10). The advancements in LLMs have affected the candidates' application process for their new jobs (Nuzula & Amri, 2023). Nuzula and Amri (2023) stated that the utilisation of LLMs by the candidates is ethically acceptable for some tasks, such as preparation for an interview and identification of candidates' skills for a specific position. However, they identified that some tasks raise ethical considerations, such as utilising LLMs in the cover letter writing process. Utilisation of LLMs by the applicants in the cover letter writing process may lead to the invitation of the wrong candidate for an interview, as the cover letter is not written by the candidate but instead by an LLM. Additionally, LLMs are also expected to challenge the previously known positive effect of the applicant's language nativity on cover letter evaluations of the recruiters (Carlsson et al., 2023). Therefore, more insights are needed on the recruiter's evaluations between human written and Chat GPT utilised cover letters to eliminate the risk of invitation of the wrong candidate for an interview.

This research is theoretically relevant as an understanding of LLMs and their outcomes in the recruitment field is limited. First, before raising the ethical considerations, it is essential to analyse whether the utilisation of LLMs in the cover letter writing process leads to better evaluations by the recruiters. Therefore, analysing the effect of Chat GPT utilisation on cover letter evaluation is the primary goal of this research. Second, candidates may use LLMs such as Chat GPT for their cover letters in two ways: they can create a new cover letter from scratch or enhance their pre-written cover letter (Strubberg et al.,2023). The literature is lacking on whether the various ways of utilising LLMs in the cover letter writing process affect recruiters' evaluations

differently. Thus, an understanding of various ways LLM's utilisation in the cover letter writing process on recruiters' evaluations will be revealed. Third, candidates' language proficiency plays a significant role in the cover letter evaluation, with higher call-back rates for candidates with higher levels of language proficiency (Carlsson et al., 2023). Language nativity and Language proficiency are used interchangeably in this study. The English as a second language (ESL) group's use of LLMs is expected to increase recruiter evaluation more than the increase of native speakers. Therefore, the candidate's language proficiency becomes less important when candidates utilise LLMs.

Therefore, based on these three considerations, this research aims to examine how the various ways of LLM utilisation influence recruiter evaluations and whether this influence varies with different levels of language proficiency. Based on the research aim, the following research question has been generated: "*How does the use of various types of LLM utilisation impact recruiter evaluations of Cover Letters, and is this impact influenced by language proficiency?"*

The practical relevance of this research is that the organisations' resources are limited to interviewing every applicant for an open position, and it is important to invite the right candidate for an interview (Watts, 2015, p. 10). Chat GPT utilisation in the cover letter writing process by the candidates may lead to the invitation of the wrong candidate as Chat GPT utilised cover letters do not reflect the candidates but are generated by the LLM. Recruiters may find themselves in a situation where candidates are coming to an interview stage or even being recruited with the utilisation of Chat GPT (Newry, 2023). The increased awareness of recruiters about the outcomes of the Chat GPT utilisation may convince them to take the required actions to avoid this problem. The findings may utilised by organisations to allow Chat GPT utilisation for a specific usage of Chat GPT if no effects on recruiters evaluations are found on that specific use of Chat GPT in the

recruiter evaluations. Findings may suggest completely restricting Chat GPT usage in the cover letter writing process if both types of Chat GPT usage are influential on recruiters' evaluations. Additionally, guiding candidates about Chat GPT usage in job descriptions may benefit the applicants. For instance, candidates may be hesitant about whether they should use Chat GPT or not when they are writing a cover letter. Some candidates may find using Chat GPT in the cover letter writing process unethical, but their thoughts about other candidates who use Chat GPT may convince them to use it. However, the ones who do not use Chat GPT may not invited for an interview even though they put more effort than the ones that utilise Chat GPT to write a cover letter, and their background is similar to the candidates invited for an interview. Therefore, more insight and regulations regarding the use of LLMs are needed in the field of human resources. The following sections introduce the key theoretical perspectives to formulate the hypothesis of this research.

## Effects of Large Language Models on Cover Letter Scores

The cover letter, often referred to as the application letter or job applicant letter, is a written submission from the candidate to get hired for a specific position. Its purpose is to showcase the candidate’s self-presentation, expressing the desire to be considered for a particular position. The cover letter comprises three or four paragraphs outlining the applicant’s experience and skills in a single page (Rudman & Glick, 2001). Anecdotal evidence indicates that recruiters do not score cover letters on a structured measure. Instead, they assess the general hireability of a candidate for a specific position.

The technological improvements in LLMs and the increased number of platforms have led people to use them in their daily tasks. LLMs are extensively trained on large volumes of text data to produce responses in natural language that mimic human conversation (Nuzula & Amri, 2023).

A widely known example of LLM is Chat GPT. Chat GPT is a model that interacts with users conversationally. The dialogue format allows Chat GPT to answer follow-up questions, admit mistakes, challenge incorrect premises, and generate texts (OpenAI, 2022). Candidates seeking a new job have started using LLMs to create cover letters in their applications. Specifically, LLM usage in the cover letter writing process raised ethical issues (Nuzula & Amri, 2023).

A recent business report revealed that 59% of job seekers who used Chat GPT to write cover letters were hired, and 78% secured an interview (Peralta, 2023). LLMs are extensively trained on large volumes of text data (Nuzula & Amri, 2023). The large volumes of text data may lead to better arguments for Chat GPT utilised cover letters then traditional way of cover letter writing. For instance, LLMs create significantly higher quality argumentative essays than humans in an online writing forum. Argumentative essays refer to essays in which students discuss a position on a controversial topic by collecting and reflecting on evidence (De Winter, 2023). Therefore, LLMs may create better arguments to convince the recruiter than humans as they are trained on large volumes of text data, and the argumentation of LLMs is better in argumentative essays. As a result, it is expected that Chat GPT utilised cover letters lead to higher cover letter scores compared to no Chat GPT used cover letters. The following argument supports this expectation but also provides an expectation about the various ways of Chat GPT utilisation.

The different ways of LLM utilisation in the cover letter writing process are expected to influence recruiters' evaluations differently. Strubberg et al. (2023) analysed the two ways of LLM, Chat GPT, usage in various stages of cover letter writing. While some participants used Chat GPT to improve their pre-written cover letters (hereafter referred to as "Chat GPT enhanced"), others used to write a cover letter from scratch (hereafter referred to as "Chat GPT created"). This research will have an additional level of Chat GPT usage as a control group where participants are

not allowed to use Chat GPT in their cover letter writing process (hereafter referred to as "No Chat GPT usage" or "Traditional way of Cover letter writing").

Chat GPT can affect recruiter scores in several ways. One of these ways is by showing desirable characteristics. It is found that Chat GPT has a default persona desired by employers, namely, conscientiousness and agreeableness (Newry, 2023). This means that text generated by Chat GPT comes across as being above average in conscientiousness and agreeableness. Conscientiousness is interpreted as "Engagement in task-related endeavours", such as working, planning, and organising (Ashton & Lee, 2007). Scoring high on conscientiousness is associated with being organised, disciplined, careful and precise (Ashton & Lee, 2008). Barrick and Mount (1991) found a positive effect of conscientiousness on positive work-related outcomes, which is performance. A study by Topor et al. (2006) focused on conscientiousness in relation to hireability. The findings revealed that HR practitioners were more inclined to hire applicants based on the trait of conscientiousness rather than cognitive ability (Topor et al., 2006). Agreeableness is interpreted as "Reciprocal altruism (tolerance)", such as patient, tolerant, and peaceful (Ashton & Lee, 2007). Scoring high on agreeableness is associated with being patient, tolerant, peaceful, mild, and agreeable (Ashton & Lee, 2008). A small correlation was found between job performance and agreeableness for manager roles where frequent interaction or cooperation with others is important (Barrick & Mount, 1991).

Recruiters are expected to be aware of the positive effects of conscientiousness and agreeableness. Additionally, many employers desire positive work-related outcomes. Therefore, recruiters' evaluations are expected to be higher when these traits are more present in the cover letters. The pre-written cover letters entered into an LLM may negatively influence these traits on generated outcomes. LLMs are expected to show more conscientiousness and agreeableness when

candidates do not provide their pre-written cover letter to an LLM. Therefore, pre-written cover letters may negatively influence LLM to show these personas on the generated cover letters. Therefore, it is expected that Chat GPT created cover letters lead to higher cover letter scores compared to Chat GPT enhanced cover letters.

Additionally, agreeableness and conscientiousness traits may not be present when no Chat GPT is used in cover letters, which supports the first argument that Chat GPT utilised (enhanced or created) cover letters lead to higher cover letter scores compared to no Chat GPT-used cover letters.

## Language Proficiency

Hulstijn (2011) defined language proficiency as a person's level of competence and skill in a specific language (in this research, English), whether it is their first language (hereafter referred to as "native") or a second language (hereafter referred to as "English as a second language/ ESL"). The following paragraph will explain how the language proficiency of the candidates may moderate the effect of Chat GPT usage on cover letter scores.

Carlsson et al. (2023) illustrated that improving the language proficiency in a cover letter enhances the call-back rate, nearly doubling it when transitioning from low to native-like proficiency, from 8% to 15%. Their findings also supported that the effect remains consistent across various occupational groups. Therefore, it is expected that Chat GPT usage may show different results between ESL and Native speaker groups.

LLMs are extensively trained on large volumes of text data, creating responses in natural language. (Nuzula & Amri, 2023). This feature of LLMs may help the ESL group more than Native speakers, as the native group has proficiency in the language. As a result, LLM utilisation (Chat

GPT enhanced or Chat GPT created) compared to no Chat GPT usage in cover letters by the ESL group may lead to a higher increase in recruitment evaluations than Native speakers.

LLMs are expected to be influenced by the pre-written cover letters in Chat GPT enhanced cover letters. A recent study found that the ESL population needed more linguistic competence to appraise and promote their skills and achievements as potential assets to the target enterprise in the cover letter writing process (Paramasivam & Rahim, 2016). Pre-written cover letters written by the ESL group may lack linguistic competence for self-promotion, while this is not expected to be a problem for native speakers. The cover letter scores for Chat GPT created groups are not expected to be influenced by self-promotion, as no pre-written cover letter putted into a LLM for this Chat GPT usage level. Therefore, the results are expected to be close to each other between ESL and Native speakers for Chat GPT created cover letters. Therefore, it is expected that the increase in cover letter scores from Chat GPT enhanced to Chat GPT created cover letters lead to a higher increase for the ESL group than native speakers.

Based on the previous arguments, the conceptual model in Figure 1 is created. The theorised effects can be seen in Figure 2.

**Figure 1**

*Conceptual Framework*

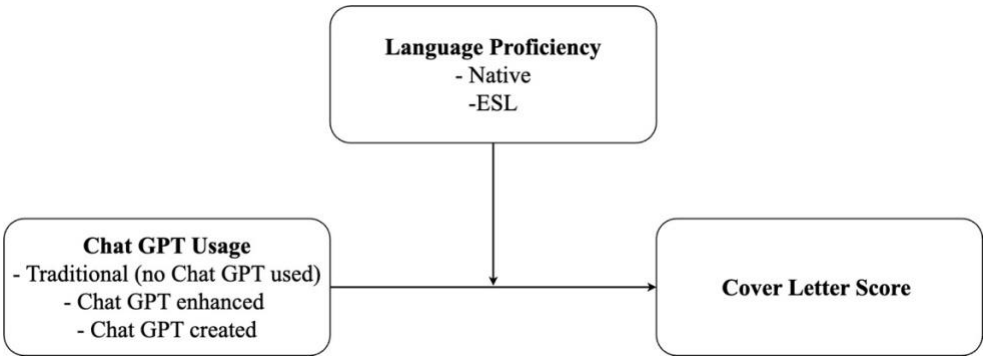
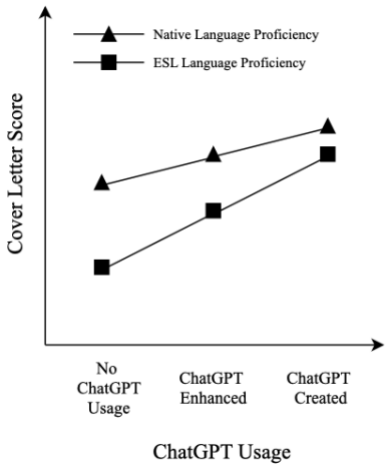


**Figure 2**

*The Theorised Effects of Chat GPT usage on Cover Letter Scores and Language Proficiency*




## Hypothesis

Based on the literature review, the following hypothesis has been formulated:

- H1a: Chat GPT created cover letters receive higher cover letter scores compared to no Chat GPT used cover letters.
- H1b: Chat GPT enhanced cover letters receive higher cover letter scores compared to no Chat GPT used cover letters.
- H1c: Chat GPT created cover letters receive higher cover letter scores compared to Chat GPT enhanced cover letters.
- H2: Language proficiency moderates the effect of Chat GPT usage on cover letter scores by strengthening the positive effect of Chat GPT utilisation for English as a second language group compared to native English speakers.

# Method

## Design

This research employed an experimental design to assess the impact of LLM usage on cover letters, with the moderation of the candidate's language proficiency. This research investigates how recruiters (i.e., the research participants) rate cover letters that come from native and ESL candidates.

Chat GPT usage was the within subjects independent variable of this research, which has three levels. These levels were no Chat GPT used, Chat GPT enhanced, or Chat GPT created. The moderator was the language proficiency of the person who wrote the cover letter. As a person cannot be in the EFL and Native group at the same time, language proficiency was the between subjects variable of the design. The recruiters evaluated six cover letters, each cover letter representing one unique combination of language nativity (native / ESL) of the candidate and Chat GPT usage (no Chat GPT used, Chat GPT enhanced, Chat GPT created).

## Creation of the cover letters

The previously collected cover letters were used in this research. In the previous research, participants were asked to self-write a cover letter for a customer support worker manager role. These cover letters were used in this research as no Chat GPT used cover letters. They were asked to make another cover letter, but this time, they were asked to utilise Chat GPT. Then, they were asked how they utilised Chat GPT. They were asked to choose from four options. Participants' cover letters included as Chat GPT enhanced in this research for two options. These two options were:

- "I used Chat GPT to make my handcrafted cover letter better."

- “I created a new cover letter with Chat GPT and manually changed some elements before submitting.”

The third option was included as Chat GPT created in this research; “I created a new cover letter with Chat GPT but practically did not change anything before submitting”. If participants stated their Chat GPT utilisation was none of the options above, their Chat GPT used cover letters were not included in this research.

Additionally in the previous research, participants were asked whether English was their primary spoken language (mother tongue) or not. Cover letters of participants whose primary spoken language is English were included in this research as native language proficiency. Cover letters of participants whose primary spoken language is not English were included in this research as ESL language proficiency.

In total, 314 cover letters were used for current research from the previous research. One hundred and fifty cover letters were used for no Chat GPT usage (ESL: 74, Native: 76). Forty-two were used for Chat GPT enhanced (ESL: 18, Native: 24), and 122 for Chat GPT created (ESL:59, Native: 63). In the current research these cover letters were used to have the participants ratings for each combination of Chat GPT usage and language proficiency.

## Procedure

The survey started with the consent form. Respondents needed to agree to participate in this study to continue with the next part of the survey. First, demographic questions such as age, gender, work experience, and level of English were asked. Next, they were asked to imagine that they were a recruiter looking for a new employee for their company. Then, information about the required position and the organisation was given. The role was the same as the previous research in which candidates wrote their cover letters. The job description can be found in Appendix 1.

Then, they were asked which skills were required for the position. This was made to check the participants' attention to the survey. Afterwards, the participants were shown a cover letter and asked to evaluate the Hireability Index questions as an operationalisation of recruiter scores. Participants in total were asked to evaluate all six combinations of cover letters (e.g., ESL and no Chat GPT used, Native and no Chat GPT used, ESL and Chat GPT enhanced). At the end of the survey, the participants were given a debriefing about the study topic.

The Qualtrics software tool was used to distribute the survey. Two randomisations were used in the Qualtrics Survey to increase the validity of this research. First, the cover letter for each group was chosen randomly from all the cover letters included in this group from the previous research (a group referred to as a combination of Chat GPT usage and language proficiency, e.g., ESL and no Chat GPT used). For instance, it was used to randomly determine which cover letter would be shown for a respondent in the no Chat GPT usage and ESL group out of 74 cover letters. Second, randomisation was used to determine the sequence of the block of the cover letters that were shown to a respondent. Therefore, every respondent evaluated the cover letters for each group in a different order.

Participants were requested to agree to participate in this study by accepting the consent form beforehand to fill out the survey. They were informed of their right to withdraw their input at any time if they requested without providing any explanation. This study was approved by the ethics review board (ERB) of Tilburg University (ERB number: TSB_RP1173). At the end of the survey, participants were informed on how and for which purpose their input would be used.

## Participants

A priori power analysis in G*Power software was used to calculate the minimum required sample size for the study. In G*Power, the required sample size was calculated using the given

alpha level, power, and effect size for the "ANOVA: repeated measure, within-between interaction" analysis. A conventional alpha level of .05 and a desired power of .80, with a small expected population effect size of 0.2 were used as input parameters. Additionally, two number of groups (language proficiency) and three number of measurements (Chat GPT usage) were used as input parameters. Based on the power analysis, a minimum sample size of 42 respondents was required for the within-between subjects two-way repeated measures ANOVA.

In total, 265 responses were recorded using the Qualtrics tools. The missing values were analysed to determine whether they occurred on the same question or whether the participants quit the survey on the same question. No systematic errors were identified in the missing values. Therefore, observations with missing values were excluded from the analysis. Additionally, one observation was removed from the sample as it was intended to test the survey. After the removal of missing values and test observation, 89 observations remained.

Exclusion criteria were applied for the remaining observations. Exclusion criteria have been set for respondents with no experience with cover letter evaluation or who do not know what the cover letters are used for. Fifty-four respondents had experience with cover letter evaluation. Thirty-five respondents knew what cover letters are, but they had no experience with cover letter evaluation. No respondents were unaware of the purpose of the cover letters. Only participants who had recruitment experience were included in the final sample. Therefore, participants might be indicated as recruiters in the following sections as the exclusion criteria applied for the persons who did not have experience with cover letter evaluation.

The survey included two attention checks, and participants who failed one were excluded. Eight respondents failed one of the two attention checks. After the exclusion criteria and attention check, 46 observations were included in the analysis.

The final sample of the survey consisted of 39.1% male and 60.9% female respondents (*N*= 46). The average age of respondents was 34.63 years (*SD* = 13.21). The sample consisted of 4.3% high school graduates, 47.8% undergraduates, 45.7 % graduates, and 2.2% hold doctoral degree (*N*= 46). The sample consisted of 6.5% students, 69.6% employed, 17.4% employed students, 2.2% out of work and 4.3% retired (*N*= 46). Figures of the sample descriptive statistics can be found in Appendix 2 - Descriptive Tables of the Sample.

## Measurements

The recruiters usually do not evaluate the cover letters with a scale in real life. For this reason, the Hireability Index was used to measure the general heritability of each cover letter (Douglas & Cole, 2016). The Hireability Index measures the likelihood of inviting candidates to interview and hiring them (Rudman & Glick, 2001). This index is constituted by a 3-item 5-point scale ("not at all likely" to "extremely likely") questionnaire for the recruiters. These items are:

1. How likely would you be to invite the applicant to interview for the customer service manager position?
2. How likely would you be to hire the applicant for the customer service manager position?
3. How likely do you think it is that the applicant was actually hired for the job he applied for?

The internal consistency within the items of the Hireability Index was tested by computing Cronbach's alpha level. Hireability Index Scale was used in the survey six times for each group of cover letters. Therefore, Cronbach's alpha test was repeated for each time used in the survey. The scale had a high level of internal consistency, as determined by alpha reliability coefficients ranging from .93 to .96. Therefore, the average hireability score was calculated for every

participants, that takes the average of the three items for a specific cover letter (e.g., ESL and no Chat GPT used). The Cronbach's alpha for each group can be found in Appendix 3.

## Plan of Analysis

A one-way repeated measures ANOVA was conducted to answer whether the use of various types of LLM utilisation impacts recruiter's cover letter evaluations. Three variables were created that take the average of ESL and Native language proficiency for each level of Chat GPT usage. These variables were named no Chat GPT used, Chat GPT enhanced, and Chat GPT created. These three variables were used in the one-way repeated measures ANOVA. Hypotheses H1a, H1b, and H1c were tested with the findings.

Additionally, this research used within-between subjects two-way repeated measures ANOVA. Whether the impact of various types of LLM utilisation on recruiter evaluations influenced by language nativity of the candidate or the writer of the cover letter. Hypotheses 2 was tested with the findings of two-way repeated measures ANOVA. Post-hoc comparisons was made based on the ANOVA results.

## Results

### Assumption Checks

A one-way repeated measures ANOVA was conducted to determine whether there were statistically significant differences in cover letter scores over various Chat GPT usage levels.

One outlier identified in no Chat GPT used level as assessed by boxplot and the test repeated with and without the outlier. The outlier does not change the effects, therefore the outlier kept in the analysis. Cover letter scores were normally distributed ($p > .05$) except for Chat GPT enhanced group, as assessed by Shapiro-Wilks's test of normality. As the ANOVA analysis is robust for non-normality, the processes followed without any actions for non-normality. Mauchly's test of sphericity indicated that the assumption of sphericity had not been violated, $\chi2(2) = 4.81$, $p = .090$.

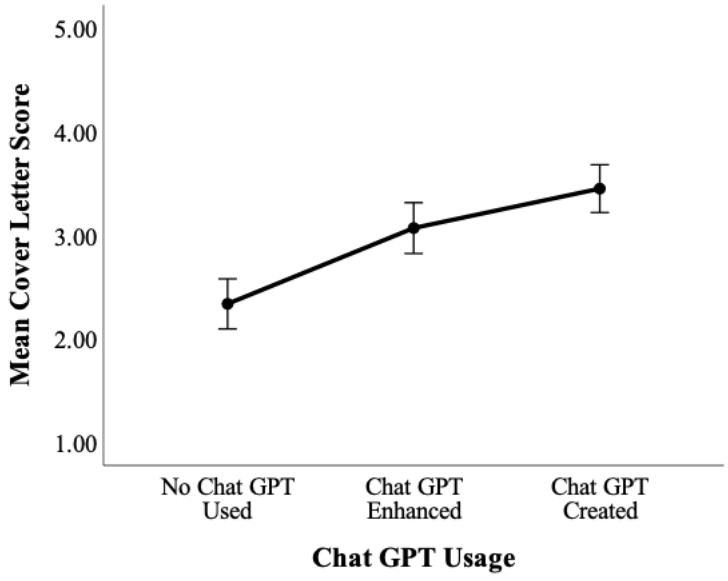
A two-way repeated measures ANOVA was run to determine the effect of Chat GPT usage levels over language proficiency on cover letter scores. Analysis of the studentized residuals showed that there was a non-normality, as assessed by the Shapiro-Wilk test of normality. The processes followed without any actions for non-normality as violations were minimal. Analysis of the studentized residuals showed that there was no outliers, as assessed by no studentized residuals greater than ± 3 standard deviations. Mauchly's test of sphericity indicated that the assumption of sphericity was met for the two-way interaction, $\chi2(2) = 4.05$, $p = .132$. Therefore, the data included in the analysis generally met for the assumptions of parametric tests.

### Main Findings

The main analysis started with visualising the effect of various Chat GPT utilisation levels on cover letter scores to test H1a, H1b and H1c. The visualisation of the main effect can be found on Figure 3.

**Figure 3**

*Visualisation of Cover letter scores for various Chat GPT usage levels*



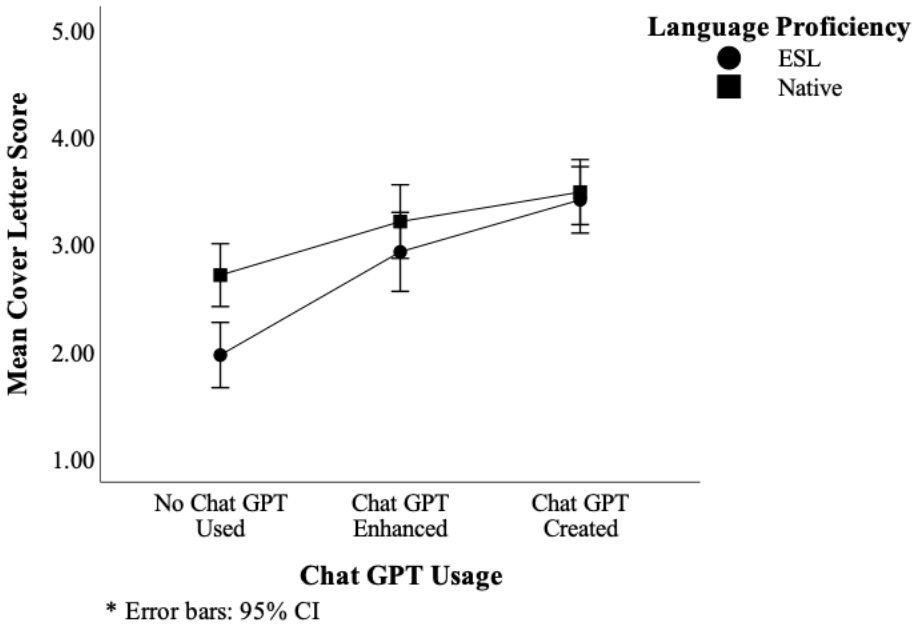

**Error bars: 95% CI*

A one-way repeated measures ANOVA results revealed that Chat GPT usage elicited statistically significant changes in cover letter scores, $F(2,90) = 24.28$, $p < .001$, partial $\eta^2 = .35$. Post-hoc comparisons made between the conditions. There was an increase in cover letter scores from no Chat GPT used ($M = 2.34$, $SD = 0.81$) to Chat GPT created cover letters ($M = 3.45$, $SD = 0.78$), a statistically significant mean increase of 1.11, 95% CI [0.77, 1.46], $p < .001$ (H1a supported). There was an increase in cover letter scores from no Chat GPT used ($M = 2.34$, $SD = 0.81$) to Chat GPT enhanced cover letters ($M = 3.07$, $SD = 0.83$), a statistically significant mean increase of 0.73, 95% CI [0.27, 1.19], $p < .001$ (H1b supported). There was an insignificant increase of 0.38 in cover letter scores from Chat GPT enhanced ($M = 3.07$, $SD = 0.83$) to Chat GPT created cover letters ($M = 3.45$, $SD = 0.78$), 95% CI [-0.02, 0.78], $p = .065$ (H1c not supported).

The moderation analysis to test H2 started with visualising the effect of Chat GPT usage on cover letter scores by different language proficiency of candidates. The visualisation of moderation effect can be found on Figure 4.

**Figure 4**

*Visualisation of Cover letter scores for various Chat GPT usage levels by Language Proficiency*




A two-way repeated measures ANOVA showed that there was no statistically significant two-way interaction between Chat GPT usage and Language Proficiency, $F(2, 90) = 2.65$, $p = .076$, partial $\eta^2 = .056$ (H2 not supported). The two-way repeated measures ANOVA results can be found in Table 1.

**Table 1**

*Two way ANOVA Results*

| Predictor | Sum of Squares | *df* | Mean Square | *F* | *p* | Partial η2 |
|---|---|---|---|---|---|---|
| Chat GPT usage | 58.89 | 2 | 29.40 | 24.28 | <.001 | .350 |
| Language proficiency | 9.30 | 1 | 9.30 | 8.56 | .005 | .160 |
| Chat GPT usage x Language proficiency | 5.47 | 2 | 2.74 | 2.65 | .076 | .056 |

*Note** Results are for sphericity assumed findings as this assumption is met previously.

The main effect of Chat GPT usage showed a statistically significant difference in cover letter scores, F(2,90) =24.28, $p < .001$, partial $\eta^2 = .35$. This effect confirms the test of the first hypothesis. The main effect of language proficiency showed a statistically significant difference in cover letter scores, $F(1,45) = 8.56$, $p = .005$, partial $\eta^2 = .16$. There was an increase in cover letter scores from ESL ($M = 2.77$, $SD = 0.09$) to native group ($M = 3.16$, $SD = 0.10$), a statistically significant mean increase of 0.37, 95% CI [0.11, 0.62], $p = .005$.

Additionally, further post hoc comparisons were made for ESL and native candidates' differences on each level of Chat GPT usage. Cover letter scores were statistically significantly different for the ESL group (M = 1.96, SD = 1.03) compared to the native group ($M = 2.71$, $SD = 0.99$) in the no Chat GPT used level, $F(1,45) = 18.20$, $p < .001$ , partial $\eta^2 = .29$. In the no Chat GPT used level, cover letter scores for the native group ($M = 2.71$, $SD = 0.99$) was 0.75 point higher than the ESL group (M = 1.96, SD = 1.03), 95% CI [0.39, 1.01], $p < .001$. However, cover letter scores were not statistically significantly different for the ESL group ($M = 2.93$, $SD = 1.24$) compared to the native group ($M = 3.21$, $SD = 1.16$) in the Chat GPT enhanced level, $F(1,45) = 1.21$, $p = .276$. Additionally, cover letter scores were not statistically significantly different for the ESL group ($M = 3.41$, $SD = 1.05$) compared to the native group ($M = 3.49$, $SD = 1.02$) in the Chat GPT created level, $F(1,45) = .130$, $p = .720$.

## Discussion

This study aims to analyse how the use of LLMs impacts recruiter evaluations of cover letters and whether this impact is influenced by language proficiency. Chat GPT is used as a LLM for the analysis. First, results suggest that both Chat GPT enhanced and Chat GPT created cover letters receive higher cover letter scores compared to no Chat GPT used cover letters by the recruiters. Additionally, Chat GPT enhanced cover letters do not significantly lead to higher cover letter scores than Chat GPT created cover letters. Second, Chat GPT usage does not show an interaction effect with the language proficiency of the candidate on cover letter scores. Third, the main effect of language proficiency revealed that cover letters written by native speakers lead to higher cover letter scores than the ESL group. Last, ESL and native speakers showed significant differences on no Chat GPT cover letters, with higher cover letter scores for native speakers than for the ESL group. However, the difference between EFL and native speakers' cover letter scores was not significantly different on Chat GPT enhanced and Chat GPT created cover letters.

This research has four theoretical implications. First, the literature suggested that the utilisation of LLMs by the candidates in the recruitment process raised ethical considerations. The use of Chat GPT in the cover letter writing process is ethically sceptical. Cover letters are expected to show the writing quality and motivation of a candidate for a specific job. Candidates fake their interests and writing quality in their application by using Chat GPT in the cover letter writing process. Additionally, Chat GPT usage in the cover letter writing process is quicker and easier than writing a cover letter without Chat GPT utilisation (Nuzula & Amri, 2023). Even though the literature raised ethical considerations, the literature was lacking on whether utilisation of Chat GPT in the writing process leads to better evaluations by the recruiters.

Therefore, this research filled this gap by creating insights about the effects of various ways of LLM utilisation by the candidates on recruiter evaluations. The results showed that Chat GPT usage in the cover letter writing process was influential in the evaluation of recruiters. Therefore, raising ethical considerations about LLM usage is important in recruitment practice, as using LLMs (creating a cover letter from scratch or enhancing a pre-written cover letter) leads to higher recruitment evaluations. For instance, a candidate might be hesitant to use a LLM in the cover letter writing process. His thoughts about the other candidate's possibility of using LLM may have led him to use it. Candidates that do not use LLMs are negatively influenced by the actions of the other candidates that use a LLM. Even though they have a similar background to the ones that used an LLM and put more effort into writing a cover letter, the recruiter's evaluations are higher for the candidates that utilised LLM in the cover letter writing process and they are not invited for an interview. Therefore, it can be concluded that the data that LLMs utilise leads to better arguments than humans to convince recruiters but with faking the candidate's interest and writing ability. Organisations need to take further action regarding the candidates' use of LLMs in the recruitment process to avoid candidates that fake their interest and writing skills in the cover letters. These actions lead to invite the candidate that has real interest for the position. Suggested actions for the organisations can be found in the practical implications section.

The second theoretical implication is about the characteristics of Chat GPT. Chat GPT shows more conscientiousness and agreeableness characteristics in generated texts than humans. These characteristics are generally related to positive work related outcomes (Barrick & Mount ,1991). The findings were complementary to the literature. From the literature, it was known that Chat GPT scored higher on these characteristics than humans. As a consequence, Chat GPT utilized cover letters and received higher evaluations from recruiters.

Third, the literature was lacking on the effects of various Chat GPT utilisation ways on recruiter evaluations. It was expected that the pre-written cover letters might negatively influence the Chat GPT generated cover letters on conscientiousness and agreeableness characteristics. However, the results showed that recruiters evaluations does not significantly changed between Chat GPT enhanced and fully Chat GPT written cover letters. Therefore, pre-written cover letters does not influence the outcomes generated by Chat GPT on the recruiter evaluations.

Fourth, the literature suggested that language proficiency had a positive effect on recruiters' evaluation with higher call-back rates for native speakers (Carlsson et al., 2023). This research tried to challenge this finding by including language proficiency as a moderator. However, no interaction effect was found between language proficiency and Chat GPT usage on cover letter scores. The main effect of language proficiency was significant, with higher cover letter scores for native speakers than ESL group, which is supportive of the findings of Carlsson et al. (2023). However, post hoc comparisons revealed that recruiters' evaluations lead to higher evaluations for native speakers than the ESL group when an LLM is not used. The difference between ESL and the native groups was not significant in Chat GPT enhanced, and Chat GPT created ways of LLM utilisation. Suggestions for further research about language proficiency can be found in the further research section.

The practical implication of this research is that organisations have limited capacity to interview every applicant. Chat GPT utilisation by the candidates brings the risk of inviting the wrong candidate for an interview as applicants fake their writing ability and interest in the job. Both Chat GPT utilisation levels create the risk of invitation of the wrong candidates. Therefore, organisations can be sceptical about the recruiter's evaluations of cover letters. They might consider not using cover letters anymore in the recruitment process, or they can regulate the cover

letter writing process by the candidates. For instance, guiding candidates whether LLM usage is restricted or not and determining LLM usage by detecting systems. Additionally, recruiters might notice agreeableness and conscientiousness characteristics on Chat GPT utilised cover letters, which lead to higher recruiter evaluations than self-written cover letters. However, the candidate fake his conscientiousness and agreeableness characteristics on his cover letter by utilisation of Chat GPT. As a result, recruiters might invite the wrong candidates as they think that the candidate has conscientiousness and agreeableness characteristics, but in reality the cover letter is created by a LLM. Therefore, their predictions about positive work-related outcomes from these characteristics may not be valid anymore as candidate may utilise Chat GPT in the cover letter writing process.

This research has three limitations. First, the outputs generated by LLMs differ for different platforms. The data used in a LLM might affect the created cover letters. Chat GPT version 3.5 was utilised in this research. Therefore, different results can be found when another platform is used. Future research can be conducted with the utilisation of another LLM platform.

The second limitation is about the position used in the research. The cover letters used in this research are created for a customer support worker manager role. In this role, coordination skills were important. Agreeableness is generally associated with being patient, tolerant, and peaceful (Ashton & Lee, 2007). The previous research found small correlation between agreeableness and job performance for manager roles (Barrick & Mount, 1991). The position used in this research was a managerial role. The agreeableness persona that the Chat GPT has might be influential on the results. Being patient, tolerant, and peaceful in managerial roles is important. Further research can be conducted using different professions in which the agreeableness trait is

not important. For instance, agreeableness is not expected to be a valid predictor of job performance for engineering roles (Barrick & Mount, 1991).

Third, this research does not include conscientiousness and agreeableness characteristics as a variable. Therefore, their effects could not analysed on the effect of LLM utilisation on further research that can be conducted, including these characteristics as a moderator on the effect of Chat GPT utilisation on recruiter evaluations.

Three suggestions were made for future research as a consequence of the limitations. Additionally, this paragraph will give advice for future research based on the findings of this study. The interaction effect of Chat GPT utilisation and language proficiency on recruiters evaluations was insignificant. Additionally, when LLMs are not used, the native group receives higher cover letter scores than the ESL group. However, there is no difference between ESL and native groups in terms of recruiters' evaluations for Chat GPT enhanced and Chat GPT created groups. Therefore, combining Chat GPT enhanced and Chat GPT created cover letter scores as Chat GPT utilised cover letters might lead to a significant interaction effect. This still indicates a trend that can be investigated further.

All in all, many studies have focused on the usage of Chat GPT for educational practices and implications for academia (Farrokhnia et al., 2023). An understanding of Chat GPT and its outcomes in different fields is needed. This research showed that the utilisation of LLMs affected the outcomes of widely used practices by organisations, namely cover letters. The findings showed that cover letters are no longer a good screening tool as the recruiter’s evaluations are affected by the LLM utilisations by the candidates. The advancements in machine learning and LLMs may have affected other organisational practices and outcomes. Therefore, further attention and

knowledge about the effect of LLMs' utilisation on organisational outcomes are required for different organisational practices.

## References


Ashton, M. C., & Lee, K. (2007). Empirical, theoretical, and practical advantages of the HEXACO model of Personality Structure. Personality and Social Psychology Review, 11(2), 150–166. https://doi.org/10.1177/1088868306294907

Ashton, M. C., & Lee, K. (2008). The HEXACO model of personality structure and the importance of the H factor. Social and Personality Psychology Compass, 2(5), 1952–1962. https://doi.org/10.1111/j.1751-9004.2008.00134.x

De Winter, J. C. (2023). Can chatgpt pass high school exams on English language comprehension? International Journal of Artificial Intelligence in Education. https://doi.org/10.1007/s40593-023-00372-z

Barrick, M. R., & Mount, M. K. (1991). The Big Five Personality Dimensions and Job Performance: A meta-analysis. *Personnel Psychology*, *44*(1), 1–26. https://doi.org/10.1111/j.1744-6570.1991.tb00688.x

Carlsson, M., Eriksson, S., & Rooth, D.-O. (2023). Language proficiency and hiring of immigrants: Evidence from a new field experimental approach. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.4361919

Douglas, S. B., & Cole, J. (2016). Romantic priming effects on the social desirability and hireability of self-promoting women. *International Journal of Psychological Studies*, *8*(4), 121. https://doi.org/10.5539/ijps.v8n4p121

Hulstijn, J. H. (2011). Language proficiency in native and nonnative speakers: An agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly*, *8*(3), 229–249. https://doi.org/10.1080/15434303.2011.565844

IBM Corp. (2012). *IBM SPSS Statistics for Windows*, Version 21.0. IBM Corp.

Newry, R. (2023). (rep.). *ChatGPT vs Personality Assessments: Does it have the right personality traits to get an interview?* Arctic Shores . Retrieved October 24, 2023, from https://www.arcticshores.com/insights/ChatGPT-vs-personality-assessments-does-it-have-the-right-personality-traits-to-get-an-interview.

Newry, R. (2023). (rep.). *ChatGPT vs Aptitude Tests: Can Generative AI Outperform Humans?* Arctic Shores . Retrieved October 24, 2023, from https://www.arcticshores.com/insights/ChatGPT-vs-aptitude-tests-can-generative-ai-outperform-humans

Nuzula, F, I., & Amri, M., M. (2023). Will CHATGPT bring a new paradigm to HR world? A critical opinion article. *Journal of Management Studies and Development*, *2*(02), 142–161. https://doi.org/10.56741/jmsd.v2i02.316

OpenAI. (2022). Introducing ChatGPT. https://openai.com/blog/

Paramasivam, S., & Rahim, M. I. B. (2016). Genre Analysis of Job Application Letters in Malaysia. *The Asian ESP Journal*, *12*(2).

Peralta, P. (2023). 59% of job seekers who used ChatGPT to write cover letters were hired. Should recruiters be alarmed? Employee Benefit News. https://www.benefitnews.com/news/what-will-the-use-of-ChatGPT-for-resumes-and-cover-letters-mean-for-recruiting-trends

Qualtrics (2020). *Qualtrics.* (Version of the 01.01.2021) [Computer Software]. Retrieved from https://www.qualtrics.com

Rudman, L. A., & Glick, P. (2001). Prescriptive gender stereotypes and backlash toward agentic women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

Strubberg, B. C., Bennett, K. C., & Nardone, C. F. (2023). How to navigate shifting tides: Mapping technical writing students' use of Artificial Intelligence. *2023 IEEE International Professional Communication Conference (ProComm)*. https://doi.org/10.1109/procomm57838.2023.00023

Topor, D. J., Colarelli, S. M., & Han, K. (2006). Influences of traits and assessment methods on Human Resource Practitioners' evaluations of job applicants. *Journal of Business and Psychology*, *21*(3), 361–376. https://doi.org/10.1007/s10869-006-9033-6

Watts, R. J. (2015). Setting the scene: Letters, standards and historical sociolinguistics. *Letter Writing and Language Change*, 1–13. https://doi.org/10.1017/cbo9781139088275.002

## Appendix 1 - Job Description

### First-Line Supervisor of Customer Support Workers

In this position you will directly supervise and coordinate the activities of clerical and administrative support workers.

Common tasks include:
- Developing and implementing procedures to deal effectively with customer requests.
- Coordinating and controlling the work of the customer support workers.
- Discussing customer responses with other managers, with a view to improving the products and services provided.

Required skills:
- Active Listening. Giving full attention to what others are saying, taking time to understand the points being made, and asking questions as appropriate.
- Coordination. Ensuring the team works together cohesively.
- Written and verbal communication. Conveying information effectively in discussions and clearly documenting instructions.
- Monitoring. Constantly evaluating performance from the team and input from customers to make improvements or take corrective action.

Required knowledge:
- Customer Service. Knowledge of principles and processes for providing customer and personal services, such as meeting quality standards for services and evaluation of customer satisfaction.
- Administrative. Knowledge of administrative and office procedures and systems such as word processing, managing files and records, designing forms, and workplace terminology.

You can download the job description here.

Note that you can look back at the job description while evaluating cover letters on the next pages.

## Appendix 2 - Descriptive Tables of the Sample

### Figure - Frequency of Experience with cover letter evaluation (Before exclusion criteria)

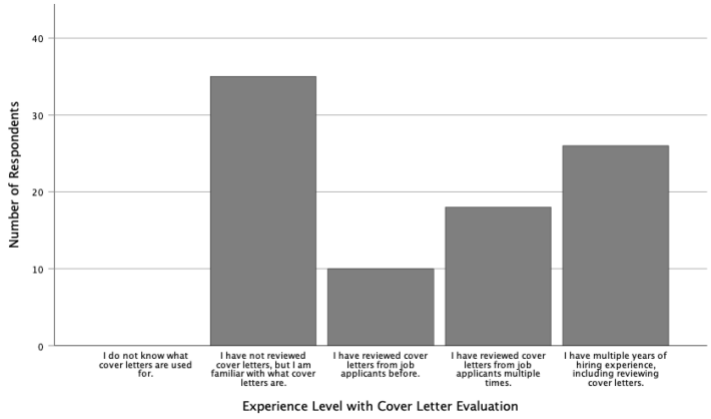


**Figure – Frequency table of Age**

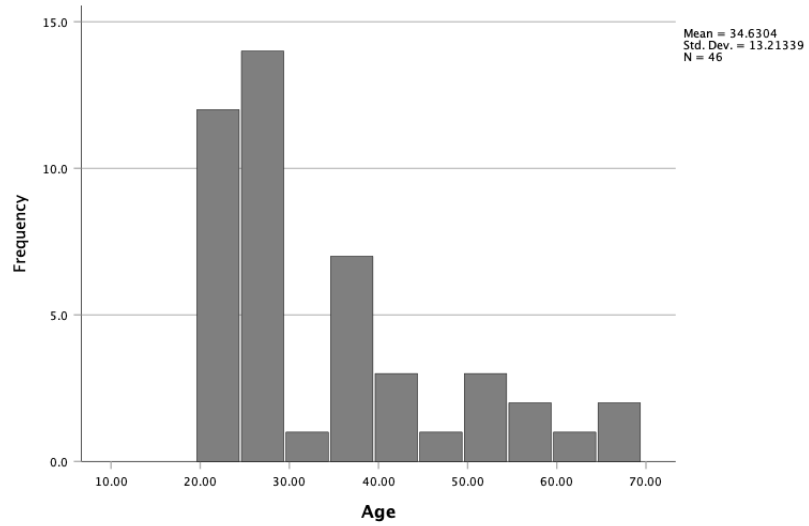



**Figure – English Proficiency of Respondents**

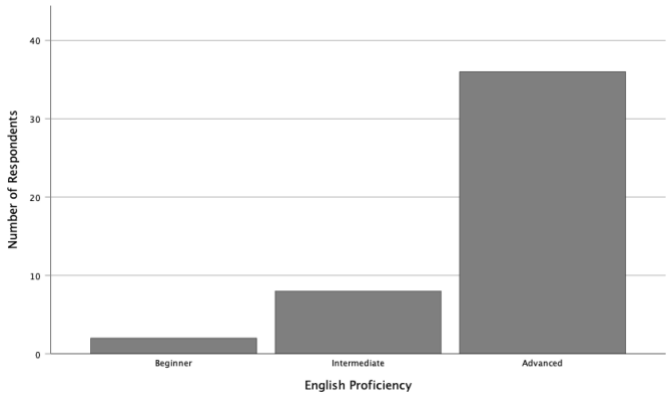



**Figure – Gender of Respondents**

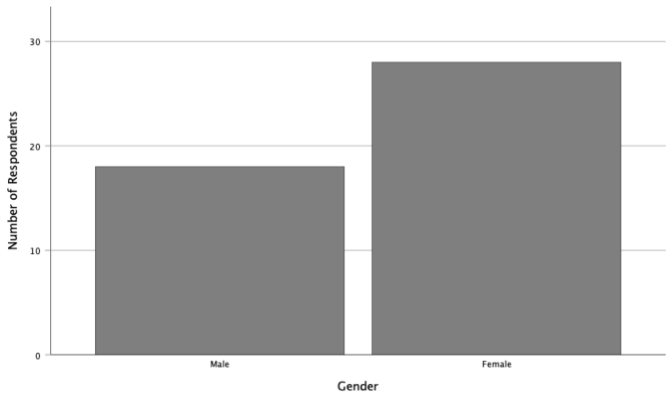
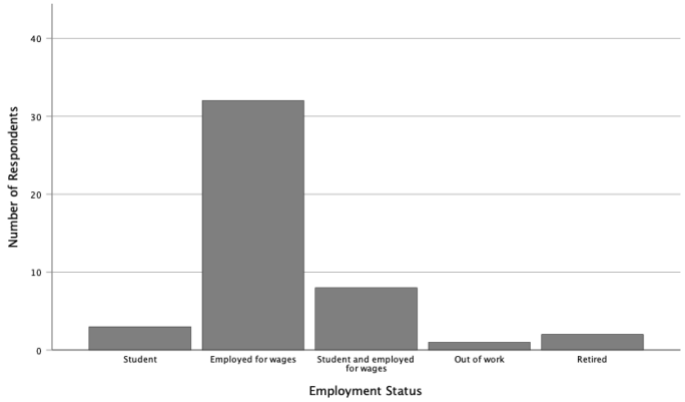


**Figure – Educational Level of Respondents**




**Appendix 3 - Hireability Index' Cronbach's alpha levels for each group of cover letters**

Hireability Index' Cronbach's alpha levels for each group of cover letters

| Group of Cover Letter | Cronbach's alpha level |
|---|---|
| ESL - No Chat GPT usage | 0.959 |
| Native - No Chat GPT usage | 0.941 |
| ESL - Chat GPT enhanced | 0.960 |
| Native - Chat GPT enhanced | 0.959 |
| ESL - Chat GPT created | 0.945 |
| Native - Chat GPT created | 0.934 |