Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/12S114

A Statistical Approach to Designing and Conducting Studies in Translation Studies (TS)

Alalddin T. Al-Tarawneh[*]

Department of English and Translation, Faculty of Arts, Zarqa University, Zarqa, Jordan

Received: 16 Sep. 2023, Revised: 9 Oct. 2023, Accepted: 9 Nov. 2023.
Published online: 1 Dec. 2023.

Abstract: Within Translation Studies (TS), the prevalent research method is qualitative. However, investigating the translation process and outcome has seen an increase in the importance of quantitative methods. This poses a challenge as many TS scholars lack familiarity with these techniques. This article suggests a solution by outlining an effective approach to quantitative research and emphasizing key aspects of such research to address this issue. The critical component of quantitative studies is explicit research designs, which can be created through several steps to ensure thorough and accurate results. Consideration must be given to defining the sample and population under study, as well as signaling the scope of the variables involved, and stating the research hypotheses and questions. In order to ascertain the comprehensibility and significance of the resulting data, one should keep in mind which outcomes would either affirm or disprove hypotheses, as well as what types of results are significant. While quantitative research is often the focus, it is worthwhile to integrate qualitative analyses and regard the resulting product.
Keywords: Qualitative research, Quantitative research, Translation Studies (TS), Statistics, Study design, mixed methods.

1 Introduction

Early in the research design, there are two important decisions to make that will ultimately impact the outcome. The first is determining if qualitative or quantitative methods (or a mix of both) are most appropriate for addressing the research question at hand. Generally speaking, TS leans towards qualitative methods, but sometimes, topics related to evaluation of the process of translation and its product my rely heavily on quantitative methods [1]. Despite the focus on quantitative research in this article, it is believed that this article will make a compelling case for the importance of incorporating qualitative analyses of product and process.

Distinguishing between experimental and naturalistic approaches is crucial. The naturalistic approach entails the observation of the subject matter within its original setting or habitat. On the other hand, experiments involve the creation of a specific situation or task that stimulates the behavior or object of study. This dichotomy resembles the dichotomy between qualitative and quantitative methods, with naturalistic research undergoing qualitative analysis and experimental research undergoing quantitative analysis. It is important to acknowledge that quantitative research is adaptable to different methodologies and can range from naturalistic to experimental designs [2]. Seeing these two perspectives as continuum's opposite ends is helpful, rather than categorizing them as exclusive options. The definition of an experiment requires that all variables, except those being investigated, be controlled. However, in studies pertaining to human behavior, this cannot always be the case. Thus, the term "experiment" is used to describe creating an environment or task for research purposes, often with a certain degree of naturalism. Compared to psycholinguistics, TS sits closer towards the naturalistic end of the spectrum, even with experimental research [2]. This highlights the importance of modern methods of statistics that enable control within quasi-naturalistic environments in TS.

[*]Corresponding author e-mail: aaltarawneh@zu.edu.jo

2 Designing a Quantitative Research Model in TS

Every quantitative study must focus on two key aspects to ensure its success, from conception to conclusion. The initial consideration involves meticulously planning the entire study and its research design prior to any data gathering. Additionally, it is essential that the design be as transparent as possible. By doing so, it will significantly help ease the decision-making process and effective implementation of statistical analyses. As we delve into the world of quantitative study design, we have a few key steps to keep in mind. These steps, though best approached sequentially, can be seen more fluidly as a process that may require revisitation of earlier steps.

2.1 Questions and hypotheses

In order to initiate a study, one must ask oneself a "what" inquiry, usually stemming from prior research or individual observations. Consider the following question: "Do student translators under training have different translation product than professionals?" This is a valid inquiry, and has been examined in various avenues of literature. But this may not be a solid, worthy researching question. In order for it to serve as a sturdy basis for design, we require a more exact and clear-cut inquiry. It is particularly necessary to specify what exactly is meant by 'translation process' or 'translation product', which area of the product arouses our curiosity, and the methodology we plan to employ in measuring it. Furthermore, we ought to elaborate on the intended definition of 'students' and 'professional translators'.

Inquiring about the difference in gaze duration or the stylistic representation of the final product on the source text between students in a BA-program of translation and professionals with over a decade of full-time experience is a more specific and explicit approach to reformulating the initial question. The query clearly identifies the targeted groups and process/product of study, while also indicating the utilization of eye-tracking (process) or linguistic software (product) as the measuring method. Refining the initial broad inquiry to concentrate on distinct groups and a particular aspect of the process or product requires careful consideration, yet is essential in achieving specific research objectives. It is necessary for the researcher to articulate the reasons for choosing these particular groups as crucial to the investigation [3].

Reformulating the original question as a hypothesis is another option [4]. In view of prior literature or experience, we could reasonably make the assumption that there is a distinction in the amount of time students and professionals spend observing ST. This could lead us to frame it as a hypothesis rather than a question: Students enrolled in a BA program are different from professionals who have over a decade of full-time experience, when it comes to the time they spend gazing at the ST or the stylistic representation in their final product. In fact, we might even be confident enough in our hypothesis to formulate a directional one: The time spent gazing by translation students is longer as compared to the professionals with 10+ years of experience. With the implementation of directionality, we can make the inference that students studying translation will focus their attention more on the source text (ST). Alternatively, the non-directional hypothesis permits the possibility of the contrast swinging in either direction. The differentiation between hypotheses that lack direction and those with it is crucial for defining one's design and expectations. Additionally, this distinction impacts statistical analyses. If a directional hypothesis is explicit enough, one-tailed tests could be suitable to use. It is widely acknowledged that one-tailed tests can only provide results that align with a specific direction. However, it is generally deemed inappropriate to employ one-tailed tests in reputable publications. To apply a one-tailed test, a researcher must possess a comprehensive comprehension of the direction in which their hypothesis is headed [2].

2.2 Sample and Population of the Study

The people we are researching are the primary focus of our discussion. This group is commonly referred to as the population. However, it is impractical to test every individual within the theoretical population, which includes past, current, and future individuals all over the world. In lieu of this, we must select a suitable sample to research and employ inferential statistics to determine if any trends observed within that sample are generally applicable to the wider population [5].

As far as the inferential statistics is concerned, the underlying rule is that the sample is chosen randomly. Essentially, every single person within the population must have an equal chance of ending up in the study sample, and that decision must have been made randomly [6]. Unfortunately, this is not always feasible in certain fields, which means that we need to find a way to juggle both randomness and possibility. With this said, it is still crucial that statistical tests assume random sampling. In a lot of instances, we might have to rethink what the population consists of depending on the study sample we have on hand, considering that we may only be able to sample from a subsection. If that is the case, it is important to recognize that our conclusions may only apply to that subsection.

Let us say we want to study how professional translators behave. It is not possible to randomly pick from such a diverse group. Typically, scholars of translation only have a limited access to a number of translators who specialize in certain languages within a certain region. This means that any conclusions drawn from their behaviors could not be applicable to all professionals. We need to have access to studies from various samples before broad generalizations can be made. Only when similar patterns emerge across these different groups can we draw conclusions that apply to all.

It is common to wonder about the optimal sample size for research. Many factors come into play in determining this. One key factor is size: as a general rule, the larger the size of sample, the more representative it is likely to be, and the greater statistical power it has. This means that even small differences can be detected in statistical tests with greater certainty. While there are no hard and fast rules for calculating statistical power, useful guidelines can be established to provide us with a better understanding. Given the unpredictable nature of translator behavior, we suggest the use of at least ten participants in TS experiments [2]. Although psycholinguistic studies commonly involve 20 or more participants, such a number is often unachievable in TS research. For independent groups design, obtaining even more participants would be ideal [2] and [7].

The results obtained from analysis are dependent on the number of participants and items, as they are interconnected. When there are numerous items, such as words, phrases, or areas of focus in a research study, it may be beneficial to limit the number of subjects. However, if the task involves translating and post-editing production within a set time frame, a larger pool of participants may be necessary. The significance of participant numbers may not be as great when there are more items or observations to be analyzed. Drawing conclusions about the population of participants or translators is more reliable when there are a limited number of items and a sizable number of participants. Conversely, when there are more items and fewer participants, generalizations can be made confidently about the population of items, such as words or texts. It is important to keep in mind that the quantity of observations in each scenario is not necessarily equivalent.

The complexity of the design dictates the number of participants and items required to discern the effects being investigated. The more intricate the design, the more observations are needed, necessitating greater numbers of participants and items to detect these effects. Regression modelling also has a rule of thumb, as per Harrell [8], which recommends having at least 10-20 observations per parameter analyzed. When designing a study, it's crucial to consider the impact of varying factors such as the level of expertise of participants, the nature of the task they're performing, and how these variables interact with one another. To attain a sample size that is balanced, one must consider multiple factors such as the quantity of items and participants, the expected magnitude of the effects, and the design of the study's complexity [9].

2.3 Variables

Crucial for selecting a statistical approach is the central step of study design – consideration of variables. In the design process, the functions of various variables are of utmost importance. Additionally, the characteristics of each variable studied must also be considered.

2.3.1 Control, explanatory, and dependent variables.

Let us begin with the introductory inquiry and break down the types of variables into three categories: explanatory, dependent, or control variables [2]. The research question or hypothesis should reveal both the explanatory and dependent variables. Throughout the study, the dependent variable captures any measured behavior or phenomenon, whereas the explanatory variable holds any variable that could potentially describe any variations within the dependent variable. Essentially, the explanatory variables affect the dependent variable and they are responsible for elaborating on the dependent variable. Outcome or response variables can be alternative terms for dependent variables, while independent variables or predictors are often used in place of explanatory variables, specifically in regression analyses. Recognizing that the role of a given variable is flexible is crucial. While it may be identified as a dependent variable in one study, it could be a control or an explanatory variable in another. The duration of one's gaze on a given source text is the dependent variable that has been examined to determine whether there are disparities between two groups: MA students of translation and translators with wide experience as professional full-timers. The explanatory variable in this study is expertise, which is defined as a contrasting factor between the two groups.

The most crucial variables could be considered as the 'dependent' and 'explanatory' variables, nonetheless, they are insignificant if the third kind of variable called control variables is not taken into account. The control variables are essential because they prevent any apparent effect of the explanatory variable being a byproduct of another factor. In order to study the variation in gaze duration on ST in relation to the proficiency of the translator, we need to ensure that proficiency is not merged with other variables. In other words, it is important to control all factors that could interfere with the results. The objective is to pair our cluster of students with a corresponding cluster of professionals. However, this may not always be feasible, so crucial adjustment factors should still be taken into account. For instance, in our case, age and second language proficiency may prove challenging to equate between the student and professional groups. Despite this, we must make an effort to address these discrepancies as much as possible. Comparatively simpler variables can be regulated like ensuring an equitable distribution of both genders within groups. When analyzing text comparisons, a similar principle applies. We must ensure that our central explanatory variable, the aspect of the text on which our hypothesis is

based, is not muddled with other variables. Thoroughly examining literature, evaluating the study's structure, and incorporating personal insight can help establish this.
Including variables in the analysis or matching beforehand, referred to as experimental control, can achieve the required control [2]. This aligns moderately with the contrast between naturalistic and experimental designs. To handle some variables experimentally would be an adequately simple task, while to incorporate some statistically would be an uncompromising and difficult task.
Plurality is the norm for control variables, whereas dependent variables usually remain fixed, with one analysis for each. Should a study offer multiple dependent variables, the number of analyses will correspond. Explanatory variables, in contrast, can vary in number. The expertise example, for instance, manipulates only one explanatory variable. While more complex designs exist, the number of variables must be limited for practicability's sake. Explanatory variables can interact with each other, exercising effects on a certain variable depending on another. This is just one reason why managing multiple explanatory variables can quickly become an overwhelming task. As more variables are added to the mix, the complexity only increases.

2.3.2 Numerical and categorical variables.

Our design requires a consideration of variable scale and function. In this decision, two key types must be classified: numerical variables which quantify and categorical variables. Categories can be simple or complex. Take for example categorical variables that have two categories like male and female which are unordered. However, there are more complex categorical variables that have multiple categories or are ordered. As an instance, categories can be ordered such as between freshmen and seniors where seniors have more expertise than freshmen. What matters in statistical theory is whether the categories are ordered or not. As the number of category levels increase, statistical analysis becomes more complex.
The categorization of numerical variables can be divided into two distinct types: discrete and continuous. Continuous variables are composed of real numbers and can have an infinite number of decimal places, whereas discrete variables are whole numbers. For example, the frequency of an idiom's appearance in a corpus is a discrete numerical variable, as it is quantified solely in whole numbers. In contrast, variables that measure time, such as gaze duration in our given instance, are continuous variables, which can be broken down into smaller subdivisions depending on the equipment used for measurement. This means a measurement could be 584 milliseconds, but with a more basic measuring tool, it would be 0.6 seconds, and a more advanced one could determine a value of 584.4 milliseconds.
Variable classification can be problematic - often there is an innate classification, like gender being determined as male or female. However, there are times where scale is a matter of perspective, often relying on the available data. For example, knowledge is considered categorical and binary, with no hierarchy between the two categories. It could be classified as part of the other three categories as well. To illustrate, we might list three arranged groups - novice, upperclassmen, and freelance translators - which are ranked in order but are not evenly spaced out.
According to Balling [2], to conceptualize expertise in a study, we can choose between a few options. There are two distinct methods for analyzing the data on translation expertise. The first method treats the data as a discrete numerical variable, which involves the assumption that the difference between a first-year and a second-year student is equivalent to the difference between a second-year and a third-year student. On the other hand, the second method involves the quantification of the duration of time spent by the participant in the role of a translator. It also considers the level of expertise as a continuous variable. The choice of approach somewhat depends on the data we have available. If the sample consists of distinct groups, it makes sense to treat expertise as a categorical variable. However, if we can recruit participants with varying years of experience, we gain more information and statistical power by treating expertise as a continuous variable rather than forcing it into categories.
The strength of statistics is highly dependent on the type variable being analyzed, with continuous variables yielding the most potent results and unordered categories producing weaker outcomes. The root of this asymmetry stems from the fact that continuous variables provide a greater amount of information than their categorical counterparts. For example, if we measure someone's professional experience out of 300 months, we are capable of deducing a more comprehensive understanding of their skill level than if we simply separated them into two groups based on arbitrary benchmarks.

3 Data collection and preparation for statistical analysis

Before conducting statistical analyses, it is crucial to ensure the collected data truly reflects the phenomenon being studied, because we want to accurately represent professional translators and students. It is not rare for anomalies to emerge in due to a lack of behavioural sensitivity in instruments or human error while recording.

Consider excluding outliers as a first step. Out of the sample, these observations are significantly considered different in number. However, outliers may be far apart from the rest of the sample. For example, a moment of fixation could last longer than usual in a recording because the participant is actually focusing on the same area for an extended time. While this observation is genuine, it may not indicate the phenomena we want to analyze.

When it comes to outlier exclusion, it is tough to avoid using some subjective judgment. One option to remove outliers is to do it manually, but that could be quite time consuming, especially for large data sets. One method of getting rid of outliers is by doing so automatically. To establish how much data varies from a sample's average, we use the standard deviation. Should the standard deviation measure low, this indicates closely clustered observations, while the opposite suggests the presence of variability [10].

4 Statistical analyses

4.1. Inference and Description

To achieve optimal data analysis, preparation is key. The two most common statistical analysis techniques are descriptive and inferential statistics. Descriptive statistics give insights into the specific sample, whereas inferential statistics allow us to generalize to a wider population [11]. Descriptive statistics computation involves the presentation of multiple measures, such as means and medians. These measures signify the midpoint of a dataset where exactly half of the observations lay above or below it. Frequencies and standard deviations can also be used to describe the studied sample. To conduct a thorough analysis of the data relevant to our research query, it would be beneficial to utilize several descriptive measures. For instance, we could scrutinize variables such as the median and mean duration of gaze on ST per group while also calculating the deviation from the average. If 'standard deviation' seems to be significantly larger than the 'mean', this implies the existence of a substantial amount of variation within the group, which is a noteworthy characteristic of the sample. While well-crafted charts and informative answers to pertinent questions can provide a fundamental understanding of a sample, they only offer a descriptive account of it. In order for researchers to draw conclusions beyond their sample, they must utilize inferential statistics. These statistics are often used in conjunction with significance tests to analyze significant differences in group means. However, factors such as variance and sample size also come into play. Without the use of inferential tests, it is impossible to determine if the patterns observed in a sample can be generalized to a larger population. Surprisingly, appropriate inferential statistics are rarely used in TS. This is problematic, as generalizations are made solely based on descriptive statistics. While some of these conclusions may be valid, accurate assessments of their applicability require appropriate testing.

4.2 Statistical test application

4.2.1. Significance and hypothesis testing

The statistical testing process involves distinguishing between two types of effects: those that systematically apply to the entire population and those that are randomly observed within the sample. Two hypotheses are created, the null hypothesis and the alternative hypothesis. The null hypothesis tests the explanatory variable being investigated and posits that there is no correlation, variance, or impact on the dependent variable. For instance, we could hypothesize that there is no discrepancy in gaze duration on ST between students and professional translators. Although statistics can examine this baseless hypothesis, most researchers are intrigued by an alternative hypothesis, which suggests the existence of a distinction between professionals and students in the gaze duration on the source text. Nonetheless, since solid hypotheses can't be directly verified, the alternative hypothesis must be evaluated through a null hypothesis test. The hypothesis framework is immaterial, whether it is based on a hypothesis or a question.

When examining the 'null hypothesis,' the statistical tests calculate the likelihood of noticing values that are higher the ones present in the sample. This calculation is associated with a p-value. The p-value can be interpreted as the likelihood of the observed effect in the sample occurring if the population had no variation, or if the null hypothesis were valid, in more informal terms [2]. To put it plainly, the tendency that the observed effect is not applicable to the entire population is also provided. If the probability, expressed as a 'p-value,' is under a particular threshold, commonly 5%, or 0.05, it indicates a low probability and invalidates the null hypothesis. This leads to the rejection of the null-hypothesis and the acceptance of an alternative hypothesis, which is commonly known as a significant effect. However, it is crucial to remember that the statistical test solely determines the rejection or non-rejection of the null hypothesis and nothing else. With regards to the alternative and null hypotheses, straightforward acceptance or rejection is not possible. In the case of obtaining a result that is not significant, and therefore has a p-value that exceeds 0.05, the only conclusion we can make is that the null hypothesis cannot be rejected. It is not possible to affirm the truth of the null hypothesis nevertheless. Hypothetically, the effect could be significant with either a larger sample size or more meticulous measurements.

We often rely on the threshold of 5.0%, which is arbitrary, yet remains frequently used in scientific studies. Basically, this value means that we find it acceptable to make a mistake in our interpretations 5% of the time when supporting the alternate hypothesis, leading to what is known as a type I error [12]. It is crucial to note that using the same data to conduct several significance tests can result in a highly potential risk of type I errors.

4.2.2. Assumptions

Certain considerations should be made before performing statistical tests, with one crucial factor being assumptions. Independence of the sample's observations (excluding pairs for before and after-tests of the same person in repeated measures designs) is a significant assumption. In cases of non-independence, such as multiple observations or responses from the same individual, dependencies must be either eliminated or modeled. When using parametric statistical tests, their potency is rooted in an assumption which necessitates the sample come from a specific distribution type. A prime illustration of this is the t-test, which is employed with variables that are continuous and dependent, such as the time taken t make a response. This test relies on the supposition that the driven data being analyzed is derived from a normal distribution.

In case the data in the sample does not meet the assumptions of parametric tests, it is important to note that the test results may be unreliable. As they say: “garbage in, garbage out” [13]. Assessing the assumptions of statistical tests is necessary for determining whether a dataset is normally distributed or not. Though frequently used, the Kolmogorov-Smirnov test poses some issues as it may suggest that normality exists in data sets with small sample sizes; prompting flawed analyses. Conversely, when applied to larger sample sizes, these tests can prove to be too delicate [14]. Such sensitivity may lead to false rejection of the hypothesis that a dataset is normally distributed [15]. To boost the precision of a normality examination, it might prove beneficial to fashion a graphical depiction of the sample's distribution. That being said, grasping such an illustration mandates a certain level of skill. Hence, when dealing with limited sample sizes, opting for non-parametric examinations, which aren't reliant on any specific distribution, may prove more prudent, although this preference may lead to diminished statistical potential [2].

At times, it is possible to modify the relevant variable to ensure that a test's normality assumptions are satisfied. This can be achieved through techniques such as logarithmic transformation, as explained in Bordens and Abbott's work [16]. Provided that the transformation is standardized and frequently employed, it is regarded as an acceptable approach. Logarithmic transformation is utilized to balance out the substantial discrepancies found within the higher end of a range, such as the distinction between the times needed for responses of 1, 000 and 10, 000 milliseconds, and equalize them with smaller gaps found within the lower end of the same range. Due to its placement within the 100-1000 millisecond range, interpreting the variable may prove to be challenging. Nevertheless, this range has the potential to correspond with cognitive reality, a point that Smith and Levy have noted [17].

4.3. Experimental and Naturalistic research

4.3.1. Regression and factorial designs

Selecting and conducting statistical tests requires careful consideration of several key factors. One such factor is balancing design aspects of experimental and naturalistic nature. Additionally, variables must be categorized according to their role (dependent, control, or explanatory) and scale (categorical or numerical). Different types of research, such as experimental and naturalistic, may have distinct impacts on both design and statistics, leading to different approaches. Corresponding statistical analyses and factorial designs may be required for experimental research, whereas analyses and regression designs might be more appropriate for quasi-experimental or naturalistic research. Understanding the contrast between these two research methods is crucial for selecting the appropriate statistical tests.

The process of creating factorial designs involves the careful selection of items or participants to represent explanatory variables as categorical factors. Control variables are also taken into account and matched accordingly among the various levels of these factors [18]. Consider our earlier research inquiry as a basis for a factorial design, involving two closely related groups of participants and advanced control variable management. To analyze the data from factorial designs, factorial statistics such as t-tests for two groups or ANOVAs for multiple groups or explanatory variables are preferable.

Working with authentic texts and quasi-natural settings is imperative in fields like TS. However, achieving strict experimental control in such scenarios is often impractical. The solution lies in statistical control where relevant variables are measured and included in the analysis. To accomplish this, we require a regression model that may consist of a combination of numerical and categorical variables, or solely numerical variables. This approach offers another means of

control that is necessary for a more naturalistic methodology. Control in a statistical approach gives TPR researchers the flexibility to investigate more variables and obtain detailed information without needing to control the variables beforehand. Regression models have a unique ability to incorporate both categorical and numerical variables, eliminating the need for categorization and allowing for acceptance of variable nature. This characteristic makes a strong case for the use of regression designs in quantitative translation research. This is especially important in the context of TPR research, which is a relatively new field. Ultimately, the incorporation of both types of variables makes regression models a robust tool for this type of research. This decision provides a more targeted and adaptable approach to research analysis. Detecting effects is easier when using numerical explanatory variables, as they are linked with more statistical power. It is worth noting that utilizing numerical explanatory variables could bring out different shapes of an effect, whereas it may not be the case with a categorical one. Obtaining more information with numerical explanatory variables could also prove useful in developing our comprehension, as noted earlier.
For years, the factorial approach has been prominent in the field. This is mainly because the ability to perform t-tests and ANOVAs manually is effortless. However, thanks to the current computing capabilities. As a result of this progress, we believe that translation research should capitalize on this trend and embrace the added benefits of precise statistical control, more realistic environments, and the abundance of data offered by regression methods.

4.3.2. Mixed models

It is suggested to employ a specific regression model known as 'mixed models,' or 'linear mixed-effects regression' (LMER). The models that are recognized as "mixed-effects" are set apart by their combination of both random and fixed effects. In essence, the fixed effects of these models are made up of explanatory and control variables, while random effects stem from the random nature of our sample. For instance, consider the explanatory variable group, which can be categorized as either students or teachers. This variable is a random effect that has the potential to impact our analysis. We can separate fixed effects, like the participants' professional status, from random effects, like the differences between individual participants. Although we focus on specific participant characteristics, the participants themselves are not significant since they are chosen randomly. Additionally, we could consider words or sentences as items, and examine their characteristics as random effects instead of focusing on the items themselves.
By factoring in random effects of both participant and item, the model acknowledges the inevitable variation among individuals and things. This approach seems particularly sensible in the cognitive sciences, given how often people's performance can fluctuate. Additionally, including these effects serves an important technical purpose. It is one way to handle the reality that observations are rarely entirely separate from each other, as most scenarios involve numerous observations for both individuals and items. Taking into consideration the statistical model's assumption of independence between observations, it is imperative to address non-independence cases. The inclusion of random effects is an ideal solution. A conventional technique for achieving self-sufficiency is to perform statistical evaluations on the means of participants and/or items. Nonetheless, this method has several disadvantages. For instance, means present issues when there are uneven observations that contribute to each mean. Furthermore, the variations in observations that may hold significance are eliminated when means are taken. Lastly, conducting two analyses - one for participant means and the other for item means - is necessary, whereas a mixed model can work for both dependencies at once, making it sufficient [2].
Learning how to use mixed models may be more difficult than other methods, but it is simply an advanced statistical technique with no fundamental difference in difficulty level. The benefits of statistical control, exploring various variables on varying scales, and analyzing non-independent and individualized effects make the effort worthwhile. Despite the steep learning curve, the advantages make it superior to alternative approaches.

5 Quantitative process data supplementation

Supplementary methods can enhance quantitative research in numerous ways. Most significantly, incorporating qualitative approaches can augment quantitative ones to achieve optimal results. For instance, when exploring an assumed hypothesis, quantifiable differences in eye movements can signify discrepancies between groups. But, for a more comprehensive understanding of why these distinctions arise and the intricacies involved, it is imperative to supplement quantitative data with qualitative analyses. These examinations may involve interviews, questionnaires, cued retrospection or translations of the final product.
The assumption that the product of a translation process is automatically acceptable is a common misconception that we must address. While it is true that professional translators should produce adequate work, dismissing the outcome of the process ignores the potential for varying levels of quality. This oversight can compromise our ability to produce a comprehensive understanding of the translation process, which necessitates accounting for differences in the product and its quality through quantitative analysis.

Process studies have largely overlooked product and quality for two reasons. Firstly, analyzing products is usually a time-consuming endeavor. Secondly, the definition of translation quality is uncertain. If we were to give attention to both process and product, we would require greater clarity regarding what translation quality consists of and how we can gauge it. Assessing the quality of translated texts can be done through various methods. One of the approaches is to seek the opinion of panels. The panels could comprise of end-users of translated texts, teachers of language and translation, or professional translators. This approach, however, can be time-consuming and costly. On the other hand, a more feasible method of quality assessment is through machine translation. BLEU or Bilingual Evaluation Understudy, as demonstrated by Carl and Buch-Kromann [19], can evaluate human translations using probabilistic estimation, which is an automatic method of quality assessment. Assessing the quality of human translated texts can be efficiently done through the automated BLEU score, as found by Carl and Buch-Kromann. This tool is quick and low-cost, making it a potential option to consider. However, further investigation is required on the effectiveness of other methods for evaluating translation quality.

6 Conclusions

It is essential to thoroughly contemplate the research design before delving into any investigation, particularly in quantitative studies. Numerous pertinent aspects of quantitative research are brought to light in this piece; nonetheless, the central concept is unambiguous: a distinct research design lays the groundwork for quantitative studies. By delineating hypotheses and research questions, specifying the population and sample, and taking cognizance of the scale and of role variables, a strong and unambiguous design can be formulated. Especially if we consider what would confirm or reject our hypotheses, as well as which results would be meaningful and comprehensible, statistical analysis can be quite simple. Our method, although not foolproof, can aid us in achieving dependable insights and enriching our comprehension of the captivating process of translation.

References

[1] L. Meister, "n methodology: How mixed methods research can contribute to translation studies," Translation Studies, vol. 11, no. 1, pp. 66-83, 2018.

[2] L. W. Balling and K. T. Hvelplund, "Design and Statistics in Quantitative Translation (Process) Research," Translation Spaces.,**4(1)**, 170-187, 2015.

[3] M. Savin-Baden and . H. Major, Qualitative Research: The Essential Guide to Theory and Practice, London: Taylor & Francis, 2023.

[4] A. Dani, "Investigating research students' perceptions about statistics and its impact on their choice of research approach," Heliyon ., **9(1)**, 1-10, 2023.

[5] C. D. Mellinger and . A. Hanson, Quantitative Research Methods in Translation and Interpreting Studies, London: Taylor & Francis, 2016.

[6] M. Rahman, M. Tabash, A. Salamzadeh, S. Abduli and M. S. Rahaman, "Sampling techniques (probability) for quantitative social science researchers: a conceptual guidelines with examples," Seeu Review., **17(1)**, 42-51, 2022.

[7] K. T. Jensen, A. C. Sjørup and L. W. Balling, "Effects of L1 syntax on L2 translation," Copenhagen Studies in Language ., **38(1)**, 319-336, 2009.

[8] F. Harrell, Regression Modelling Strategies. With Applications to Linear Models, Logistic Regression, and Survival Analysis, New York: Springer, 2001.

[9] A. Field, J. Miles and Z. Field, Discovering statistics using R, London: Sage publications, 2012.

[10] M. A. Mansournia, M. Nazemipour, A. Naimi, G. Collins and M. Campbell, "Reflection on modern methods: demystifying robust standard errors for epidemiologists," International Journal of Epidemiology., **50(1)**, 346-351, 2021.

[11] P. Mishra, C. Pandey, U. Singh, A. Gupta, C. Sahu and A. Keshri, "Descriptive statistics and normality tests for statistical data," Annals of cardiac anaesthesia., 22(1), 67-81, 2019.

[12] X. Wang and X. Ji, "Sample size estimation in clinical research: from randomized controlled trials to observational

studies," Chest ., **158(1)**, 12-20, 2020.

[13] J. Derraik, W. Parklak, B. Albert, K. Boonyapranai and K. Rerkasem, "Fundamentals of data collection in clinical studies: simple steps to avoid “garbage in, garbage out," The International Journal of Lower Extremity Wounds., **20(3)**,183-187, 2021.

[14] N. A. L. Alhuzami, "Social Media and the Collective Mind Inductive Study," Zarqa Journal for Research and Studies in Humanities., **21(2)**, 380-394, 2021.

[15] G. Ruxton, D. Wilkinson and M. Neuhäuser, "Advice on testing the null hypothesis that a sample is drawn from a normal distribution," Animal Behaviour., **107(1)**, 249-252, 2015.

[16] K. Bordens and B. B. Abbott, Research design and methods: A process approach, New York: McGraw-Hill, 2002.

[17] N. Smith and R. Levy, "The effect of word predictability on reading time is logarithmic," Cognition ., **128(3)**, 302-319, 2013.

[18] O. D. Apuke, "Quantitative research methods: A synopsis approach," Kuwait Chapter of Arabian Journal of Business and Management Review., **33(1)**, 1-8, 2017.

[19] M. Carl and M. Buch-Kromann, "Correlating translation product and translation process data of professional and student translators," in Correlating translation product and translation process data of professional and student translators, St Raphael, France., 2010.

[20] K. T. Hvelplund, "Eye tracking and the translation process: Reflections on the analysis and interpretation of eye-tracking data," MonTI ., **1(1)**, 201–223, 2014.

[21] S. Gries, Statistics for linguistics with R: A practical introduction, Walter de Gruyter: Berlin., 2013.

[22] H. R. Baayen, D. Davidson and D. Bates, "Mixed-effects modeling with crossed random effects for subjects and items," Journal of memory and language., **59(4)**, 390-412, 2008.