

Examination of Quantitative Studies Published in National Journals in the Field of English Language Education in Terms of Methodology

Atilla ÖZDEMİR*

İdil SAYIN**

Ayşenur ULUYOL***

Abstract. English language education is considered extremely important in many countries, and several countries make large investments regarding the issue. Therefore, it is essential to develop evidence-based policies and practices to effectively and efficiently use these resources. For this reason, it is crucial to ensure the reliability of this evidence, which is the product of scientific research. In order to achieve this, it is necessary to identify the frequently made mistakes in the studies and offer solutions to prevent these mistakes. Therefore, this study aims to examine the quantitative studies in the field of English language education from a methodological point of view, report the findings, and offer solutions. In order to achieve this aim, this study adopts phenomenology design, one of the qualitative research methods. The data obtained were examined through the document analysis method using an article review form developed by the researchers. Data were collected from 30 articles that used inferential statistics and published in the field of English language education between 2013-2021 in 15 different journals in the Ulakbim journal base. Results show that in some of the studies, the research method was not specified, mistakes were made in statistical terms, the problem of missing data was not mentioned or the missing data methods were not used, and the assumptions of parametric analysis were not tested or reported. Additionally, results also indicate that researchers performed these analyzes with the help of package programs without having sufficient theoretical knowledge for statistical analysis. In this sense, it is recommended that researchers receive basic training on statistical analysis and manage the research process with an expert in measurement and evaluation for data analysis and reporting.

Keywords: Quantitative research, missing data, statistical assumptions, robust statistics.

* Orcid ID: <https://orcid.org/0000-0003-4775-4435>, Assist. Prof. Dr., Süleyman Demirel University, Department of Mathematics and Science Education, Turkey, atillaozdemir@sdu.edu.tr

** Orcid ID: <https://orcid.org/0000-0001-5546-2673>, Res. Ass., Süleyman Demirel University, Department of Foreign Language Education, Turkey, idilsayin@sdu.edu.tr

*** Orcid ID: <https://orcid.org/0000-0002-4358-6836>, English Teacher, Ministry of National Education, Turkey, aysenuruluyol3@gmail.com

1. INTRODUCTION

In addition to being the language with the highest number of speakers globally, English language plays a vital role in all levels of education due to its advantages in education, employment, international communication, and access to information (Reddy, 2016). In order to get optimum efficiency from the language education provided, researchers conduct various studies and make recommendations in accordance with the results of their research. In education, where evidence-based practices are encouraged, such research must be planned with methods and techniques appropriate to the adopted paradigms. There are three main paradigms in educational research; quantitative, qualitative and mixed research (Johnson & Christensen, 2012). Among these paradigms, quantitative research is often preferred due to its advantages, such as saving time and resources, being considered scientific due to statistical analysis, and giving generalizable results compared to qualitative research (Bryman, 2001; Gorard, 2001; Connolly, 2007). Although they are frequently preferred, choosing the statistical analysis suitable for data and research questions and determining a study design suitable for the study are considered to be complicated by researchers (İlhan, 2016). In quantitative research, descriptive and inferential statistics, two areas of applied statistics, are generally used (Mann, 2016; Sheskin, 2000). Descriptive statistics consist of editing, displaying and explaining techniques that allow us to see the general characteristics of the collected data (Cramer & Howitt, 2004; Mann, 2016; Sheskin, 2000). Tables, graphs, measures of central tendency and variability are frequently used in descriptive statistics (Mann, 2016; Sheskin, 2000). On the other hand, inferential statistics generalize the population with sample data, to make decisions or predictions about the population (Cramer & Howitt, 2004; Mann, 2016; Sheskin, 2000). Procedures in inferential statistics can be divided into parametric and non-parametric tests. Parametric tests are significance tests that assume that population values have a normal distribution (Cramer & Howitt, 2004). Non-parametric tests consist of many inferential techniques that do not involve evaluating the characteristics of the population from the characteristics of the sample, and they do not include the use of standard error estimates, which is the characteristic of parametric statistics (Cramer & Howitt, 2004). Non-parametric tests have several advantages over parametric tests. These advantages can be listed as follows; not needing to have a specific distribution to have validity, accordingly having a wider application area (Blischke et al., 2011), being easy to be understood, not needing sample theories, easy to be applied, the main population from which the sample is taken does not need assumptions about the frequency function and being able to be applied to categorical data measured with nominal scale, data that do not show normal distribution and data that indicates rank. Disadvantages can be listed as follows; only can be applied to nominal or ordinal data, and even with nominal or ordinal data, parametric tests should be preferred over non-parametric tests if there is a parametric test that is stronger than the non-parametric test, there is no non-parametric test that examines the relationships in ANOVA, non-parametric tests can be applied to test hypothesis but not to estimate parameters (Sahu, Pal & Das, 2015, p. 145-146). All parametric tests have two basic

assumptions: normal distribution and variance homogeneity. Many researchers are unaware that parametric tests can yield undesirable results like incorrect calculations when the assumptions of parametric tests are not fully met. These incorrect calculations encountered when assumptions are not met lead to erroneous investigations (Erceg-Hurn, Wilcox & Keselman, 2013; Larson-Hall, 2016). When the assumptions are not met, using parametric tests may result in false conclusions; even if accurate conclusions are reached, they may miss some data relationships. Necessary corrections can be made or alternative analytical strategies can be applied when the assumptions are violated (Fidell & Tabachnick, 2003). Non-parametric tests are not without assumptions; they only have fewer assumptions than parametric tests (Marascuilo & McSweeney, 1977). On the other hand, when the assumptions of parametric tests are not met, various researchers recommend using robust statistics that allow parametric operations to be performed without assumptions of parametric tests rather than using non-parametric tests (Loewen & Plonsky, 2015; Larson-Hall, 2016). Table 1 presents commonly used comparison tests and their assumptions.

Table 1

Commonly Used Comparison Tests

Comparison	Dependent (output) Variable	Independent (descriptive) Variable	Parametric Test			Non-parametric Test
			Name of the test	Assumptions	Ho to test assumptions?	
Averages of 2 independent groups	Scale	Nominal (Binary)	Independent Samples T-test	Normally distributed data Data collected from independent groups/samples Equal group variances (Variance Homogeneity) No outliers	Normal Distribution: Q-Q plots and Histograms of both groups can be examined. Shapiro Wilk and Kolmogorov Smirnov Tests can be applied. Variance Homogeneity: Levene's Test can be applied.	Mann Whitney U test/ Wilcoxon Rank-Sum Test
Averages of more than three independent groups	Scale	Nominal	One-Way ANOVA	Normally distributed residuals Equal variances (Variance Homogeneity)	Normal Distribution: Q-Q Plot and Histograms can be examined. Shapiro Wilk and Kolmogorov Smirnov Tests can be applied. Variance Homogeneity: Levene's Test or Bartlett's test can be applied.	Kruskal-Wallis Test

Average difference of dependent samples	Scale	Time/Condition	Paired Samples T-test	Normally distributed data Paired groups/samples Equal group variances (Variance Homogeneity) No outliers	Normal Distribution: Q-Q plots and Histograms of both groups can be examined. Shapiro Wilk and Kolmogorov Smirnov Tests can be applied. Variance Homogeneity: Levene's Test can be applied.	Wilcoxon Signed-Rank Test
More than three measurements in the same sample	Scale	Time/Condition	Repeated Measures ANOVA	Normally distributed residuals Sphericity; the differences between each pair of repeated measurements must be equal	Normal Distribution: Histograms of residuals can be examined. Sphericity: Mauchly's sphericity test can be applied.	Friedman Test

(Source: Marshall & Boggis, 2016; Larson-Hall, 2016; Hatch & Lazaraton, 1991; Peacock & Peacock, 2011)

In Table 1, parametric tests that can be used according to the number of groups compared and their assumptions are presented. Additionally, the ways to test assumptions and the non-parametric equivalent of the parametric test that can be applied in case of assumption violation are suggested. Similarly, Table 2 displays commonly used association tests and their assumptions.

Table 2

Commonly Used Association Tests

Association	Dependent (output) Variable	Independent (descriptive) Variable	Parametric Test			Non-parametric Test
			Name of the test	Assumptions	Ho to test assumptions?	
Relationship between 2 continuous variables	Scale	Scale	Pearson's Correlation Coefficient	Linearly related variables Independently collected data Data collected from either interval or ratio scale should be continuous. Normally distributed data Homoscedasticity (equal statistical variance)	Normal Distribution: Q-Q Plot and Histograms can be examined. Shapiro Wilk and Kolmogorov Smirnov Tests can be applied.	Spearman's Rank Correlation Coefficient

Estimating the value of a variable through the value of a predictor variable or seeking significant relationships	Scale	Any variable	Simple Linear Regression		
	Nominal (Binary)	Any variable	Logistic regression		
Examining the relationship between 2 categorical variables	Categorical	Categorical	80% of the expected values in each cell should be greater than 5 No cell with a frequency below 1 The expected value is greater than 5 in all cells in 2x2 matrices.	When the assumptions are not met, Fisher's Exact Chi-Square Test can be applied or the assumptions can be met by combining cells.	Chi-Square Test

(Source: Marshall & Boggis, 2016; Peacock & Peacock, 2011)

Table 2 shows commonly used parametric association tests and their assumptions and non-parametric equivalents that should be used when the stated assumptions are not met.

Before starting the analysis, it is important to check whether the collected data are correct and suitable for analysis (Blischke et al., 2011; Wilkinson, 1999). Before analyzing data, the process of preparing data for analysis is very important. IBM employee Wilf Hey describes the preparation phase for analysis with the statement "Garbage In, Garbage Out", known as the GIGO rule. According to this statement, the quality of the data to be analyzed directly affects the analysis results. Therefore, during the preliminary data analysis, it is important to identify missing data, conduct necessary analyses accordingly, and test the assumptions of the planned analyses. Plonsky (2017) stated that missing or unreported data that affect research transparency limit readers' ability to interpret findings. The best method to prevent missing data in the data set is to plan the study in detail (Wisniewski et al., 2006). Various methods can be applied by researchers in case of missing data in the data set. Some steps researchers should follow before using these methods are as follows (Salgado et al., 2016);

- Determining the pattern and reasons of missing data,
- Calculating the missing data rate,
- Determining the most appropriate method of missing data imputation or deletion.

Missing data types can be analyzed in three main categories (Curley et al., 2017);

Missing Completely at Random (MCR): Missing data is independent of the properties of observed or non-observed values in the data set.

Missing at Random (MR): Missing data can be explained by the observed values.

Not Missing at Random (NMAR): Missing data depends on non-observed values and cannot be explained by observed values.

Missing data handling methods can be divided into three groups: deletion, imputation, and machine learning techniques (García-Laencina et al., 2009).

In the deletion method, the complete-case and available-case operations are applied. Although the deletion method is preferred frequently, it is not recommended because it causes data loss. Despite its disadvantages, the deletion method is a powerful method used in MCR and MR situations (Demir & Parlak, 2012). In the complete-case operation, cases with missing data in any variable used in the analysis are removed. In the available-case operation, only cases that have variables of interest are used in the analysis (García-Laencina et al., 2009).

In the imputation method, a value is assigned to the missing data. There are various techniques used to calculate this value (García-Laencina et al., 2009). In the simple imputation method, the imputation process is performed by assigning an estimated value to the missing data. In the mean substitution method, the value calculated by taking the mean of all cases is assigned to the missing data. In addition to these, missing data imputation can be made using the conditional mean imputation method, which is applied using regression analysis (Demir & Parlak, 2012). Among the machine learning techniques, Multi-Layer Perceptron Assignment, K-Nearest Neighbours Assignment or Self-Organizing Maps algorithms can be applied (García-Laencina et al., 2009). Besides the mentioned missing data deletion, imputation, and machine learning techniques, approaches such as maximum likelihood (ML) and multiple imputation (MI) methods are widely used (Allison, 2009).

The purposes of performing the preliminary analysis on the data are as follows (Blischke et al., 2011, p.159);

- Verifying the source of the data
- Verify that the data contains the specified variables
- Verifying the units of measure
- 'Clearing' data by deleting or correcting the obviously incorrect data
- Identifying outliers and unusual results
- Checking for missing data
- Identifying other unusual data features

Data collection tools commonly used in quantitative studies are questionnaires, scales and tests. Although the terms *questionnaire* and *scale* are often used interchangeably, they are different data collection tools. The term "*scale*" should be used in cases where the variables are measured and the total score has a meaning. Questionnaire is a data collection tool used to collect participants' information about a topic, where the total score is insignificant (Erkuş, 2007; 2010). Since the total score in questionnaires is not meaningful, they cannot be examined by looking at properties such as validity and reliability. On the other hand, scales should have reliability and validity properties

(Erkuş, 2010; Ercan & Kan, 2004). In multidimensional scales, the Cronbach's alpha value for each dimension and the stratified alpha value for the whole scale can be used as the reliability coefficient. In the analysis of data collected by scales, using inferential statistical methods such as paired/independent samples t-test and ANOVA is suitable. Descriptive statistics and statistical tests such as chi-square can be used to analyze data collected with questionnaires. The development processes of these two data collection tools also differ. Additionally, if existing questionnaires and scales are planned to be reused, they must be adapted to the requirements of the new study and new participants (Iwaniec, 2020). A similar conceptual and terminological confusion exists around Likert-type question and Likert scales. Likert-type question and Likert scale are two different terms. Misuse of these two terms in research may result in using inappropriate statistical tests. In Likert-type questions, total score is not calculated. In a Likert scale, an inference is made by calculating total score (Turan, Şimşek & Aslan, 2015). Another important point to be considered in research is the characteristics of the group which the data are collected from. Therefore, researchers need to know the concepts such as universe, sample, and study group. In studies, the number of individuals covered by the researched subject and whose results are generalized to is called *the universe* (Nachimas & Nachimas, 1996). According to Karasar (2012), there are two types of universes. The general universe is an abstract concept that is difficult or even impossible to reach. Since it is very difficult, even impossible to reach the general universe, the concept of the study universe (target universe) has been developed (Karasar, 2012). Studies to be carried out on the universe are quite expensive and time-consuming for researchers and very difficult to control. For this reason, researchers generally prefer to work on small subgroups of the universe. These subgroups are called samples. The selection process of samples that can represent the universe is defined as sampling. Two important features should be considered in sampling. The first of these is the size of the sample; the second is how this sample should be selected (Peacock & Peacock, 2011). There are many techniques in sampling. These techniques are divided as random and non-random sampling. In random sampling, all individuals in the universe have an equal chance of being sampled. In non-random sampling, individuals are selected according to certain characteristics. In this sampling technique, a rationale should be given to determine which individuals will be in the sample and who will not (Taherdoost, 2016). Sampling methods are an important feature that affects the design of the study. For example, quasi-experimental studies are a subtype of non-experimental studies. The main feature distinguishing non-experimental research from true experimental research is the absence of random sampling in non-experimental research (Rogers & Révész, 2020; Plonsky, 2017). Quasi-experimental studies try to imitate true experimental studies without random sampling (Cook & Wong, 2008; Kirk, 2009). Besides, in cases where there is no random sampling, Balcı (2013) suggests that the subgroup selected from the population should be named as "study group" instead of sample.

Although *statistically significant* and *meaningful* expressions, which are frequently used in scientific studies, have different meanings, they are used interchangeably with the

words '*manidar*' and '*anlamlı*' in Turkish. The term *significance* is used daily to mean significant or meaningful, but it has a technical meaning in testing hypotheses in statistics (Stout & Ruble, 1995). Statistically significance, which can be interpreted with p-value, is concerned with the significance of the difference between the results of a study and whether the observed difference occurred by chance, it does not fully explain the results of the study and does not state that the results are scientifically important or meaningful (Stout & Ruble, 1995; Norman & Streiner, 2003; Wasserstein et al., 2019). P-value below .05 as a result of statistical analysis indicates that the results are statistically significant. Statistically meaningful, also known as practical significance or clinical significance, expresses the importance of the phenomenon researched in the study and it is specified by interpreting the effect size (Sullivan & Feinn, 2012; Kraemer, 2014). Cohen's d is frequently used in reporting the effect size (Schuele & Justice, 2006). For this reason, effect size must be presented to interpret the practical significance of the difference in studies with statistically significant differences between the results. However, since the terms '*manidar*' and '*anlamlı*' in Turkish are interchangeably used and lost their original meanings, this may cause studies not to report their results clearly enough. The processes that should be considered in quantitative studies have been summarized and expressed.

Purpose

In this context, the main purpose of this study is to examine the quantitative studies conducted in the field of English Language Education in terms of methodology and to report the findings. Therefore, in the examined studies, characteristics of the data collection tools, preliminary data analysis process and the statistical techniques used in the data analysis phase were examined and the appropriateness of these techniques was explained in detail.

Significance of The Study

Being the language with the most speakers globally, English has made English language education a norm in every country thanks to the individual and social privileges it provides to its speakers (Mahu, 2012; Ethnologue, 2019). Therefore, the quality of English language education and efforts to improve this education have gained importance. The European Commission (2019) advocates evidence-based policymaking and practices adopted in language education in this sense. In order to ensure the reliability of the evidence, which comes from scientific research, the quality of the research must be ensured by using the right methods and techniques. Therefore, it is thought that the study will contribute to the literature by offering solutions to researchers by revealing the theoretical foundations of inferential statistics-based research and problems in using theoretical knowledge about data collection tools and data analysis.

2. METHOD

This study adopts the phenomenology design and data were examined using the document analysis method. Figure 1 shows the research process.

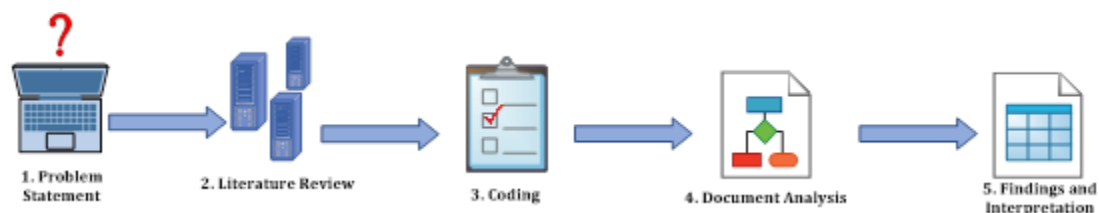


Figure 1. The Research Process

The process followed in the research is presented in Figure 1. Accordingly, the steps followed are as follows.

1. Problem Statement: The purpose of the study and the subproblems that need to be answered in order to achieve this purpose are explained.

2. Literature Review: In the literature review phase of the study, the studies having the following criteria were selected. The criteria determined according to the purpose of the study are as follows:

- a. Conducted in the field of English Language Education
- b. Applied inferential statistics
- c. Published in journals indexed in the ULAKBIM index
- d. Published between 2013 and June 2020
- e. Open-access

Accordingly, 30 articles in 15 different journals in the ULAKBIM index that met these criteria were analyzed with document analysis.

3. Coding: The data were analyzed with the study review form developed by the researchers (Appendix. Data Analysis Form). This form consists of five questions developed in accordance with the purpose of the study and related explanations. The questions consist of the research method, the group from which the data are collected, the missing data statement, the data collection tool and the analyzes performed. In addition, the form includes an explanation column in which the researchers write their detailed opinions on the relevant questions. The draft form was presented to two researchers, who had document review studies, for expert opinion. According to expert opinions, the form was finalized by making necessary developments. Five randomly selected articles were evaluated by three researchers for the credibility and transferability of the review form. Apart from the researchers, two experts were asked to evaluate the articles according to the form. It was found that both researchers and

experts expressed similar results among themselves and in their mutual evaluation. The following numerical measure of the agreement suggested by Miles-Huberman (1994) was taken into account for the findings obtained from expert opinions:

$$\text{Reliability} = \frac{\text{consensus}}{(\text{consensus} + \text{disagreement})}$$

If the percentage of agreement obtained from this calculation is over 70%, it indicates that it is reliable (Miles-Huberman, 1994; Tavşancıl & Aslan, 2001). It was found that the percentage of agreement between the expert opinions was 100% in the first round. Thus, it was concluded that the data analysis form used was clear and understandable.

4. Document Analysis: In this study, in order to find answers to the problems in the study review form developed in accordance with the purpose of the study, the studies selected according to the criteria mentioned in item 2 were analyzed by document analysis.

5. Findings and Interpretation: The data obtained from the descriptive analysis were tabulated and interpreted according to the purpose of the study and the problem situation.

Validity and Reliability

This study adopted qualitative research method, and in line with the nature of these studies, credibility (internal validity), transferability (external validity), dependability (reliability) and confirmability (objectivity) were analyzed in terms of reliability and validity.

In order to meet the specified criteria, all the details of the process at every stage of the study were taken into consideration. For this purpose, first of all, detailed reporting was conducted. The theoretical framework of the variables to be examined within the scope of the study, the criteria addressed in the selection process of the studies to be examined, and the process design of the study was tried to be expressed in a detailed and simple language. The study sample was selected by the convenience sampling method, one of the non-random sampling methods, in accordance with the determined criteria. The main purpose of selecting studies to be examined according to certain criteria is to include the suitable ones for the study. In the examination of these studies, the review form developed by the researchers by taking expert opinion was used. The three researchers who conducted the study simultaneously and independently used the same form and came together at the end of the analysis and the differences in the results were compromised. The findings are presented descriptively in the tables. It is aimed to present the findings without any subjective opinion reflected on them. The findings are analyzed within the scope of the aim of the study, theoretical framework and related literature, and presented in the comments, conclusion and discussion sections.

3. FINDINGS

In the study, 30 articles regarding English language education between 2013-2020 were examined. The distribution of the studies according to journals and the years they were published are given in Table 3.

Table 3

Distribution of The Studies According to Journals and The Years They Were Published

Journals	Publication Year								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	1	-	-	-	-	1
Çukurova Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	-	2	-	-	-	2
Cumhuriyet International Journal of Education	-	-	-	-	-	-	1	1	2
Eğitimde Nitel Araştırmalar Dergisi -ENAD	-	-	-	-	-	1	-	-	1
Eurasian Journal of Applied Linguistics	-	-	-	-	-	-	-	1	1
Eurasian Journal of Educational Research	-	1	-	-	-	-	-	-	1
Fırat Üniversitesi Sosyal Bilimler Dergisi	-	-	-	-	-	-	1	-	1
Hacettepe Üniversitesi Eğitim Fakültesi Dergisi	-	-	1	-	-	-	-	-	1
International Online Journal of Education and Teaching (IOJET)	-	-	-	-	-	-	-	3	3
Journal of Language and Linguistic Studies	1	-	-	2	1	5	2	-	11
Marmara Üni. Atatürk Eğitim Fak. Eğitim Bil. Dergisi	-	-	1	-	-	-	-	-	1
Novitas-ROYAL	-	2	-	-	-	-	-	-	2
Pamukkale Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	-	-	-	-	1	1
Turkish Online Journal of Qualitative Inquiry	-	1	-	-	-	-	-	-	1
Turkish Studies Educational Sciences	-	-	-	-	-	1	-	-	1
Total	1	4	2	3	3	7	4	6	30

Table 3 shows that in recent years the usage of inferential statistics (correlation, regression, t-test, ANOVA, Chi-square, etc.) has increased in number compared to the first years in studies regarding English language education. A total of 30 publications in 15 different journals were examined in the study. Table 4 displays the methods adopted in the studies.

Table 4

The Methods Adapted in The Examined Studies

Method	Years								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
Quantitative	1	-	2	1	1	1	2	1	9
Mixed	-	4	-	2	2	3	1	-	12
Unspecified	-	-	-	-	-	3	1	5	9
Total	1	4	2	3	3	7	4	6	30

As can be seen in Table 4, mixed method was used in twelve studies.

However, in nine studies, only analyses were mentioned in the method section and no detailed information was provided on the research method. When these studies were examined, they either used both quantitative and qualitative analyses or only quantitative analyses. Table 5 shows the information about the group the data were collected from in the studies.

Table 5

Terms Used for the Group the Data Collected From

Terms	Years								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
Sample	1	2	2	1	2	4	-	3	15
Study Group (Participant)	-	-	-	-	-	1	4	-	5
Sample/Participant	-	-	-	1	-	-	-	1	2
Unspecified	-	2	-	1	1	2	-	2	8
Total	1	4	2	3	3	7	4	6	30

Group that data collected from is important for research's both internal and external validity. Assigning individuals to the groups with random sampling methods affects the internal validity, and the generalizability of the results obtained from the research affects the external validity. External validity decreases, especially when the group that data collected from is not selected from the universe by probability sampling. In Table 5, it is seen that, while half of the studies used the term *sample*, eight studies did not include any statements regarding sampling. In two of the studies, *sample-participant* terms were used, while in five studies only *participant* term was used. However, terms such as *sample* or *study group* were not used to refer to the group that data were

collected from in eight studies. Only in some of these studies, various demographic information about the people in the groups that data were collected from was presented. This raises doubts about the external validity of the examined studies.

Reporting missing data was another point that was examined. Twenty-nine of the thirty studies did not address the problem of missing data. During examining, it was observed that three studies had inconsistencies between the numbers in the abstract, the number of participants and the numbers reported in the analysis tables. The findings related to this are given in Table 6.

Table 6

Examination of Missing Data

Missing Data	Years								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
There is missing data	1	-	-	-	-	1	3	2	6
There is no missing data	-	4	2	3	3	6	1	4	24
Total	1	4	2	3	3	7	4	6	30

In Table 6, it is seen that six studies have missing data, however, only one of these studies reported this issue and it was not stated which of the missing data handling methods was applied in that study. Another important aspect examined in the studies was data collection tools. Data collection tools are important in determining the analyses that can be applied. Data collection tools used in the examined studies are presented in Table 7.

Table 7

Data Collection Tools

Data Collection Tools	Years								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
Questionnaire	1	2	-	2	-	1	-	2	8
Semi-structured interview form	-	2	-	1	1	2	1	-	7
Weekly written reflections	-	1	-	-	-	-	-	-	1
Intercultural tasks	-	1	-	-	-	-	-	-	1
Open-ended question	-	1	-	-	-	-	-	-	1

Stimulated recall	-	1	-	-	-	-	-	-	1
Essay	-	1	1	-	-	-	-	-	2
Scale	-	-	1	2	3	4	6	4	20
Observation checklist	-	-	-	1	-	-	-	-	1
Analytical scoring rubric	-	-	-	-	1	-	-	-	1
Writing Tasks	-	-	-	-	-	1	-	-	1
Peer reflective discussion form	-	-	-	-	-	1	-	-	1
Checklist of speaking	-	-	-	-	-	-	-	1	1
Focus group interview	-	-	-	-	-	1	-	-	1
Multiple choice test	-	-	-	-	-	2	-	1	3

The studies examined consist of quantitative and mixed studies. This can also be observed from the variety of data collection tools given in Table 7. It is noteworthy that quantitative data were collected mostly through scales and questionnaires. Qualitative data were obtained from semi-structured interview form, focus group interview, stimulated recall, open-ended question, weekly written reflections and peer reflective discussion form. In four studies, which are described as mixed studies, researchers used the data obtained from interview forms to support their quantitative data. Therefore, it may be more appropriate to express such studies as qualitatively supported quantitative studies rather than mixed studies. Another common thing observed in examined studies is the use of the terms scale and questionnaire interchangeably. In the introduction part, these terms and the differences were emphasized. The statistical analyses that can be applied to the data collected from these two data collection tools, which have different characteristics, are also different. In addition, while scales should have stricter statistical-based rules about validity and reliability, it was observed that these properties were not included in many studies. In calculating the reliability of scales with more than one dimension, the reliability coefficient should be calculated with techniques such as stratified alpha, but no such analyses were found in the examined studies. It was observed that there is an absence of analyses such as exploratory and confirmatory factor analysis, which should be done in cases where scales are culturally adapted or developed to fit into study's context. In six studies, the data obtained from the questionnaires were analyzed correctly by giving the percentages and frequencies, but in some studies, the data collected by questionnaires from different groups were compared over the total score, and incorrect results were obtained. It has also been observed that studies using multiple-choice tests do not include the psychometric properties of the tests. The data obtained from data collection tools were analyzed using parametric and non-parametric tests. Tests used in the analysis of the data in the examined studies are given in Table 8.

Table 8

Data Analysis Methods

Data Analysis Methods	Years								Total
	2013	2014	2015	2016	2017	2018	2019	2020	
Parametric Test	1	4	1	3	2	5	2	6	24
Non-Parametric Test	-	-	-	-	-	-	1	-	1
Parametric and Non-Parametric	-	-	1	-	1	2	1	-	5
Total	1	4	2	3	3	7	4	6	30

As seen in Table 8, twenty-four studies analyzed data by parametric tests, one study analyzed by non-parametric tests, and five studies analyzed by both parametric and non-parametric tests. Problems such as not checking analyzed data whether the assumptions of parametric tests are met or not, using parametric tests despite having insufficient number of samples, not presenting any detail although stating that assumptions are checked, not reporting the effect sizes of the significant differences obtained from the comparison tests, improperness between the data collection tool and the analysis that can be done with the data collected from this tool (for example, applying parametric or non-parametric tests instead of chi-square on data collected from questionnaires).

In some studies where scale data were used, it was observed that parametric tests were preferred when the normality assumptions regarding the sub-dimensions were met, and non-parametric tests were preferred when they were not. In most of the studies examined, it is seen that the name of the program in which the analysis was carried out was expressed. It can be stated that it is more important to highlight the issues such as the accuracy of statistical analysis, checking the assumptions of statistical tests, suitability between the data collection tool and the analysis because there are many programs that can be used free for data analysis.

4. RESULTS, DISCUSSIONS AND SUGGESTIONS

In this study, a total of 30 studies using inferential statistics published between the years of 2013 and 2020 in 15 different journals indexed in the Ulakbim journal index in the field of English language education were examined. The studies were examined in terms of the methodology, sample, missing data problem, data collection tool and data analysis method. Later, the problems related to the examined aspects were described. The biggest challenge encountered during the study examination process was some of the studies' insufficiency and deficiencies in reporting. This situation made it difficult to find information about the aspects determined for examination and made it necessary to

read and analyze the studies in detail and to classify data into different categories. Bakker and Wicherts (2011) stated a similar issue in their study.

In the study, firstly, the method of the studies was examined. However, it was observed that nine studies did not make any explanation on this issue. All examined studies were either quantitative or mixed research since studies containing inferential statistics constitute the scope of this study. Erdem's study (2011) shows that most of the studies in the field of education in Turkey (82%) use quantitative research methods. While there may be many reasons for this, one of the most important reasons is statistical package programs, which are becoming widespread due to developing information technologies. Thanks to these programs, researchers can make individual analyses without knowing the conceptual and theoretical foundations of the statistical techniques they use. However, this situation can sometimes cause the tools to become the purpose (Arı, Armutlu, Tosunoğlu & Toy, 2009). Other reasons for the widespread use of quantitative studies may be due to the idea that objectivity cannot be discussed in positivist-empirical studies (Reis, 2009) and the belief that science can only progress as long as mathematics is used (Lafargue, 2014, p. 105). In this context, it is obvious that the preference of mixed methods in educational sciences will positively affect the validity and reliability as it will increase the diversity in multiple methods and data collection tools (Sözbilir & Kutu, 2008).

Secondly, the terms preferred for the sample group were examined. In most of the studies examined, it is seen that sufficient information was not provided regarding the sampling, which is of great importance in terms of internal validity and external validity of the study. It was also observed that the terms *sample* and *study group* were frequently used interchangeably. This makes the generalizability of the study questionable. It was also observed that the sample size of a study was below 30, which is the expected number for 5% confidence level (Altunışık et al., 2012).

Missing data was another aspect that was examined within the scope of this study. It was seen that only one of the examined studies reported missing data and this study did not state which of the missing data handling methods was applied. In the studies examined, analyses outputs and findings were examined to find out whether there was missing data or not. Performing the inferential analyses without any correction regarding the missing data may cause results to be biased or erroneous (Kang, 2013; Frisell, 2016). Demir and Parlak (2012) found similar results in their studies and stated significant deficiencies regarding conducting studies and reporting the results in the field of education in Turkey.

The most important finding regarding data collection tools is that scales and questionnaires are the most frequently used data collection tools. Similar results have been shown in different studies (Altunışık et al., 2012). When the data collection tools were examined, it was found that the most common problem is the use of scale and questionnaire terms interchangeably. These two data collection tools have different psychometric properties. In addition, the analyses that can be applied to data collected

using these data collection tools also have different features. In the studies, it was seen that the terms questionnaire and scale are used interchangeably in the method and findings sections. However, in most of the studies, the psychometric properties of the data collection tools were not mentioned. This may cause reliability problems regarding the information obtained from the collected data.

Finally, the techniques used to analyze the data in the studies were examined. It was observed that parametric analyses were used in the majority of the examined studies. Also, the most used parametric analysis was found to be the t-test. It also corresponds with the results of Hsu (2005)'s study. Although parametric tests were used in most of the studies, studies either never mentioned anything or only provided verbal statements regarding meeting the assumptions that must be met to use these tests. This has been stated in the study conducted by Demir, Saatçiođlu, and İmrol (2016) and in many different studies in the literature (Bakker & Wicherts, 2011; Özsoy & Özsoy, 2013). Probably, one of the main reasons for this situation, as stated in the studies, is that the analyses are made using a package program; therefore, the statistical knowledge of the researchers on this subject is insufficient. Additionally, mentioning the names of the package programs used for statistical analysis does not prove that the analyses are carried out without any problems. Lastly, the effect size value was not reported in any of the significant differences obtained in the comparative studies.

As a result of this study, various problems encountered in quantitative studies were addressed and solutions were presented. Considering that the data collected in this study are limited to publications in the field of English language education and the Ulakbim index, the number of studies can be increased in future studies. Studies conducted abroad in a similar field can be compared with studies conducted in Turkey. It would be beneficial for researchers to gain a theoretical basis before analyzing with package programs, to get consultant support on how to report the results obtained, and beneficial for journals to have a consultant who is an expert in the field of measurement and evaluation.

References

- Allison, P. D. (2009). *Missing data*. R. E. Millsap, A. Maydeu-Olivares (Eds.), *Quantitative methods in psychology içinde* (pp. 72-89). London: SAGE Publication.
- Altunışık, R., Coşkun, R., Bayraktarođlu, S., & Yıldırım, E. (2012). *Sosyal bilimlerde araştırma yöntemleri [Research methods in social sciences]* (7th edition). Sakarya, Sakarya Yayıncılık.
- Arı, G., Armutlu, C., Tosunođlu, N., & Toy, B. (2009). Nicel arařtırmalarda metodoloji sorunları: yüksek lisans tezleri üzerine bir araştırma [Methodology problems in quantitative research: a research on master's theses]. *Ankara Üniversitesi SBF Dergisi*, 64(4), 15-37.
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior research methods*, 43(3), 666-678.

- Balcı, A. (2013). *Sosyal bilimlerde araştırma: Yöntem, teknik ve ilkeler* [Research in the social sciences: Methods, techniques and principles]. Ankara, Pegem A.
- Blischke, W. R., Karim, M. R., & Murthy, D. P. (2011). *Preliminary data analysis. In: Warranty Data Collection and Analysis*. Springer Series in Reliability Engineering. Springer, London.
- Bryman, A. (2001). *Social research methods* (1th edition). New York: Oxford University Press.
- Connolly, P. (2007). *Qualitative data analysis in education: A critical introduction using spss*. London, Routledge.
- Cook, T., & Wong, V. (2008). *Better quasi-experimental practice*. P. Alasuutari, L. Bickman, J. Brannen (Yay. haz.), *The Sage handbook of social research methods içinde* (ss. 134-164). London, Sage.
- Cramer, D., & Howitt, D. L. (2004). *The Sage dictionary of statistics: a practical resource for students in the social sciences*. Sage.
- Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for Stratified-Parallel Tests. *Educational and Psychological Measurement*, 25(2), 291-312.
- Curley, C., Krause, R. M., Feiock, R., & Hawkins, C. V. (2017). Dealing with missing data: A comparative exploration of approaches using the integrated city sustainability database. *Urban Affairs Review*, 55(2), 591-615.
- Demir, E., & Parlak, B. (2012). Türkiye’de eğitim araştırmalarında kayıp veri sorunu. [The problem of missing data in educational research in Turkey] *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(1), 230-241.
- Demir, E., Saatçioğlu, Ö., & İmrol, F. (2016). Uluslararası dergilerde yayımlanan eğitim araştırmalarının normallik varsayımları açısından incelenmesi [Examination of educational research published in international journals in terms of normality assumptions]. *Current Research in Education*, 2(3), 130-148.
- Ercan, İ., & Kan, İ. (2004). Ölçeklerde güvenilirlik ve geçerlik [Reliability and validity in scales]. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(3), 211-216.
- Erceg-Hurn, D. M., Wilcox, R. R., & Keselman, H. J. (2013). *Robust statistical estimation*. In *The Oxford Handbook of Quantitative Methods in Psychology* (pp. 1-34).
- Erdem, D. (2011). Türkiye’de 2005-2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. [Examining the articles in educational sciences journals published in Turkey between 2005 and 2006 in terms of some characteristics: A descriptive analysis] *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(1), 140-147.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar [Problems encountered in scale development and adaptation studies]. *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2010). Psikometrik terimlerin Türkçe karşılıklarının anlamları ile yapılan işlemlerin uyumsuzluğu [Inconsistency of the meanings of the Turkish equivalents of psychometric terms and the operations performed]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 72-77.

- Ethnologue. (2019, September 24). *What is the most spoken language?* Retrieved from <https://www.ethnologue.com/guides/most-spoken-languages>.
- European Commission. (2019). *Education and training. Education and Training - European Commission*. Retrieved from https://ec.europa.eu/education/policies/multilingualism/evidence-based-policy_en
- Fidell, L. S., & Tabachnick, B. G. (2003). Preparatory data analysis. *Handbook of psychology: Research methods in psychology, 2*, 115-141.
- Frisell, T. (2016). Why missing data is a problem, and what you shouldn't do to solve it. *Annals of the Rheumatic Diseases, 75*(2), 44-45.
- García-Laencina, P. J., Morales-Sánchez, J., Verdú-Monedero, R., Larrey-Ruiz, J., Sancho-Gómez, J., & Figueiras-Vidal, A. R. (2009). *Classification with incomplete data*. T. Ganchev, M. Sokolova, R. Rada, P. J. García-Laencina, V. Ravi (Eds.), *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques içinde* (pp. 147-175). IGI Global.
- Gorard, S. (2001). *Quantitative methods in educational research: The role of numbers made easy*. London, The Tower Building.
- Hatch, E. M., & Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York, NY, Newbury House Publishers.
- Hsu, T. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research and Method in Education, 28*(2), 109-133.
- İlhan, M. (2016). Araştırmacıların nicel veri analizinde karşılaştıkları güçlüklerin ikili karşılaştırmalar yoluyla incelenmesi [Examining the difficulties faced by researchers in quantitative data analysis through pairwise comparisons]. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 7*(1), 73-84.
- Iwaniec, J. (2020). *Questionnaires, implications for effective implementation*. J. McKinley, H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics içinde*. New York, Routledge.
- Johnson, R. B., & Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches (4th edition)*. SAGE Publications.
- Karasar, N. (2012). *Bilimsel araştırma yöntemi*. [scientific research method] Ankara, Nobel Yayıncılık.
- Kirk, R. (2009). *Experimental design*. R.E. Millsap, A. Maydeu-Olivares (Eds.), *The Sage handbook of quantitative methods in psychology içinde* (pp. 23-45). London, Sage.
- Kraemer, H. (2014). Effect size. *The Encyclopedia of Clinical Psychology, 1-3*.
- Lafargue, P. (2014). *Burjuvazi Öldürür ve Toplu Denemeler*. [Bourgeoisie Kills and Collective Trials] Zeplin Yayınları.
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R (2nd edition)*. Routledge.
- Loewen, S., & Plonsky, L. (2015). *An A-Z of applied linguistics research methods*. Macmillan International Higher Education.

- Mahu, D. P. (2012). Why learning is English so beneficial nowadays? *International Journal of Communication Research*, 2(4), 374-376.
- Mann, P. S. (2016). *Introductory statistics* (9th edition). John Wiley and Sons, Inc.
- Marascuilo, L. A., & McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA Brooks/Cole.
- Marshall, E., & Boggis, E. (2016). The statistics tutor's quick guide to commonly used statistical tests. *Statstutor Community Project*, 1-57.
- Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis*. Thousand.
- Nachmias, C., & Nachmias, D. (1996). *Research methods in the social sciences (5th edition)*. New York, St. Martin's Press.
- Norman, G. R., & Streiner, D. L. (2003). *PDQ statistics*. PMPH-USA.
- Özsoy, S., & Özsoy, G. (2013). Eğitim araştırmalarında etki büyüklüğü raporlanması [Effect size reporting in educational research]. *İlköğretim Online*, 12(2), 334-346.
- Peacock, J. L., & Peacock, P. J. (2011). *Oxford handbook of medical statistics (1th edition)*. Oxford University Press
- Plonsky, L. (2017). *Quantitative research methods*. S. Loewen, M. Sato (Eds.), The Routledge Handbook of Instructed Second Language Acquisition. New York: Routledge.
- Reddy, M. S. (2016). Importance of English language in today's world. *International Journal of Academic Research*, 3(4), 179-184.
- Reis, H. (2009). *Quantitative methods in relationship research*. H. T. Reis, S. Sprecher (Eds.), Encyclopedia of human relationships içinde (pp. 1319-1321). SAGE Publications, Inc.
- Rogers, J., & Révész, A. (2020). *Experimental and quasi-experimental designs*. J. McKinley, H. Rose (Eds.), The Routledge Handbook of Research Methods in Applied Linguistics içinde. New York, Routledge.
- Sahu, P. K., Pal, S. R., & Das, A. K. (2015). *Estimation and inferential statistics*. Springer.
- Salgado, C., Azevedo, C., Proença, H. G., & Vieira, S. (2016). Missing data. *Secondary Analysis of Electronic Health Records* (pp. 143-162). Springer.
- Schuele, C. M., & Justice, L. M. (2006). The importance of effect sizes in the interpretation of research. *The ASHA Leader*, 11(10), 14-27.
- Sheskin, D. J. (2000). *Handbook of parametric and Nonparametric statistical procedures (2nd edition)*. Chapman and Hall/CRC.
- Sözbilir, M., & Kutu, H. (2008). Development and current status of science education research in Turkey. *Essays in Education*, 24(1), 1-22.
- Stout, D. E., & Ruble, T. L. (1995). Assessing the practical significance of empirical results in accounting education research: The use of effect size information. *Journal of Accounting Education*, 13(3), 281-298.
- Sullivan, G., & Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282.

- Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. *International Journal of Academic Research in Management*, 5(2), 18- 27.
- Tavşancıl, E., & Aslan E. (2001). *İçerik analizi ve uygulama örnekler [Content Analysis and Application Examples]*. Ankara, Epsilon Yayınları.
- Turan, İ., Şimşek, Ü., & Aslan, H. (2015). Eğitim araştırmalarında likert ölçeği ve likert-tipi soruların kullanımı ve analizi. [Use and analysis of likert scale and likert-type questions in educational research] *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 30(1), 186-203.
- Wasserstein, R. L., Schirm, A. L., & Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(1), 1-19.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, 54(8), 594.
- Wisniewski, S. R., Leon, A. C., Otto, M. W., & Trivedi, M. H. (2006). Prevention of missing data in clinical research studies. *Biological Psychiatry*, 59(11), 997-1000.

Supp-1

Data Analysis Form

Volume	Publication Year	Journal Name	Article Name	Method	Comment	Study Group Expression (Universe-Samples/Study Group)	Study Group Statement Relevance (External Validity Issue)	Study Group and Number of Participants	Missing Data	Data Collection Tool Used in the Study	Comment	Data analysis	Data Analysis Critique

Scientific, ethical and citation rules were followed during the writing process of the study titled "Examination of Quantitative Studies Published in National Journals in the Field of English Language Education in Terms of Methodology". It has been committed by the authors of this study that no falsification has been made on the collected data and the "Sakarya University Journal of Education Journal and its editor" has no responsibility for all ethical violations. All responsibility belongs to the authors, and that the study has not been sent to any other academic publication medium for evaluation.

Statement of Contribution of Researchers to the Article:

1st author contribution rate: %50

2nd author contribution rate:%40

3rd author contribution rate:%10

Conflict of Interest Statement

There is no conflict of interest

Statement of Financial Support or Acknowledgment:

No financial support was received from any institution for this study. No Acknowledgment.

İngiliz Dili Eğitimi Alanında Ulusal Dergilerde Yayımlanan Nicel Çalışmaların Yöntem Açısından İncelenmesi

Atilla ÖZDEMİR* İdil SAYIN** Ayşenur ULUYOL***

Öz. İngilizce eğitimi birçok ülkede önemli görülmekte ve bu konuda ülkeler büyük kaynaklar kullanmaktadır. Kaynakların etkili ve verimli kullanılmasında ise kanıta dayalı politikalar ve uygulamalar geliştirilmesi önemlidir. Bu nedenle bilimsel araştırmaların ürünü olarak kabul edilebilecek kanıtların güvenilirliklerini sağlayabilmek önemlidir. Bunu sağlayabilmek için alanyazında bulunan çalışmalarda sıklıkla yapılan hataların belirlenmesi ve çözüm önerilerinin sunulması önemlidir. Bu çalışma İngiliz dili eğitimi alanında yapılan nicel çalışmaları yöntemsel açıdan inceleyerek bulguları raporlamayı ve çözüm önerileri sunmayı amaçlamaktadır. Bu amaca ulaşmak için nitel araştırma yöntemlerinden olgu bilim deseni kullanılmıştır. Elde edilen veriler doküman incelemesi ile araştırmacılar tarafından geliştirilen bir makale inceleme formu kullanılarak incelenmiştir. Çalışmanın verileri, Ulakbim dergi tabanında yer alan 15 farklı dergide, 2013-2021 yılları arasında İngiliz dili eğitimi alanında basılan ve çıkarımsal istatistiğin kullanıldığı 30 makaleden toplanmıştır. İncelemeler sonucunda çalışmaların belirli bir kısmında araştırma yönteminin belirtilmediği, istatistiksel terimlerin kullanımında hataların yapıldığı, kayıp veri sorunundan bahsedilmediği veya kayıp veri atama yöntemlerinin kullanılmadığı ve parametrik analizlerin kullanılması için ön şart olan varsayımlardan bahsedilmediği görülmüştür. Sonuç olarak araştırmacıların istatistiksel analizler için yeterli teorik bilgiye sahip olmadan paket programlar yardımıyla bu analizleri gerçekleştirdikleri görülmüştür. Bu anlamda, araştırmacıların istatistiksel analiz konusunda temel düzeyde eğitimler almaları, araştırma süreçlerinde verilerin analizi ve raporlanması konusunda ölçme ve değerlendirme alanında uzman bir danışmandan destek alarak süreci yönetmeleri önerilmektedir.

Anahtar Kelimeler: Nicel araştırma, kayıp veri, istatistiksel varsayımlar, sağlam istatistikler.

* Orcid ID: <https://orcid.org/0000-0003-4775-4435>, Dr. Öğr. Üyesi, Süleyman Demirel Üniversitesi, Matematik ve Fen Bilimleri Eğitimi Bölümü, Türkiye, atillaozdemir@sdu.edu.tr

** Orcid ID: <https://orcid.org/0000-0001-5546-2673>, Arş. Gör., Süleyman Demirel Üniversitesi, Yabancı Diller Eğitimi Bölümü, Türkiye, idilsayin@sdu.edu.tr

*** Orcid ID: <https://orcid.org/0000-0002-4358-6836>, İngilizce Öğretmeni, Milli Eğitim Bakanlığı, Türkiye, aysenuruluylol3@gmail.com

1. GİRİŞ

İngilizce, dünyada en çok konuşmacıya sahip olan dil olmasının yanı sıra eğitimde, iş bulmada, bilgiye ulaşmada ve uluslararası iletişimde sağladığı avantajları nedeniyle, eğitimin her kademesinde önemli bir yere sahiptir (Reddy, 2016). Verilen dil eğitiminden ideal verimi alabilmek adına alan araştırmacıları çeşitli araştırmalar yapmakta ve araştırmalarının sonuçlarına uygun şekilde önerilerde bulunmaktadırlar. Kanıta dayalı uygulamaların özellikle teşvik edildiği eğitim alanında bu tarz araştırmaların güncel yaklaşımlara uygun yöntem ve teknikler ile kurgulanması önemlidir. Eğitim araştırmalarında nicel, nitel ve karma araştırmalar olmak üzere üç araştırma yaklaşımı öne çıkmaktadır (Johnson ve Christensen, 2012). Bu araştırma yaklaşımlarından içinden nicel araştırmalar, eğitim gibi birçok alanda zaman ve kaynak tasarrufu, istatistik analizlerin kullanılması nedeniyle bilimsel bulunması ve nitel araştırmalara kıyasla genellenebilir sonuçlar vermesi gibi çeşitli avantajlarından dolayı sıklıkla tercih edilmektedir (Bryman, 2001; Connolly, 2007; Gorard, 2001). Sıklıkla tercih edilmelerine karşın, veriye uygun ve araştırma sorusuna cevap verecek istatistik analizinin seçimi ve çalışmanın amacına uygun çalışma deseni tercihinin zor olması gibi özellikleri nedeniyle araştırmacıları süreçte zorlamaktadır (İlhan, 2016). Nicel araştırmalarda genellikle uygulamalı istatistiğin iki alanı olan betimleyici ve çıkarımsal istatistikler kullanılır (Mann, 2016; Sheskin, 2000). Betimleyici istatistikler toplanan verilerin genel özelliklerini görmemizi sağlayan düzenleme, görüntüleme ve açıklama tekniklerinden oluşur (Cramer ve Howitt, 2004; Mann, 2016; Sheskin, 2000). Betimleyici istatistiklerde sıklıkla tablo, grafik, merkezi eğilim ve değişkenlik ölçümleri kullanılır (Mann, 2016; Sheskin, 2000). Çıkarımsal istatistikler ise örneklem verileriyle evrene genelleme, evren hakkında karar vermek veya tahmin yapmak için kullanılır (Cramer ve Howitt, 2004; Mann, 2016; Sheskin, 2000). Çıkarımsal istatistiklerdeki prosedürler parametrik ve parametrik olmayan (nonparametrik) testler olarak, ikiye ayrılabilir. Parametrik testler popülasyon değerlerinin genelde normal dağılıma sahip olduğunu varsayan anlamlılık testleridir (Cramer ve Howitt, 2004). Parametrik olmayan testler örneklemin özelliğinden popülasyonun özelliklerini değerlendirmeyi içermeyen çok sayıda çıkarımsal teknikten oluşur ve parametrik istatistiklerin karakteristiği olan standart hata tahminlerinin kullanılmasını içermezler (Cramer ve Howitt, 2004). Parametrik olmayan testlerin, parametrik testlere göre çeşitli avantajları bulunmaktadır. Bu avantajlar şu şekilde sıralanabilir; geçerliliğe sahip olmak için belirli bir dağılıma sahip olmalarına gerek olmaması, buna bağlı olarak daha geniş bir kullanım alanına sahip olması (Blischke, Karim ve Murthy, 2011), kolaylıkla anlaşılabilir olması, örnek teorilerine ihtiyaç duymaması, uygulamalarının kolay olması, örneklemin alındığı ana popülasyonun frekans fonksiyonuyla ilgili varsayımlara ihtiyaç duymaması, nominal ölçek ile ölçülen sınıflandırma olan verilerde, normal dağılım göstermeyen verilerde ve sıralama belirten verilerde uygulanabilmesi olarak sıralanabilir. Dezavantajları ise parametrik olmayan testlerin ölçümün nominal veya sıralı olduğu durumlarda uygulanabilir olması, böyle bir durumda bile eğer parametrik olmayan testten daha güçlü bir parametrik test mevcutsa parametrik test uygulanması gerekliliği, ANOVA'da etkileşimleri özel varsayımlar

yapılmadığı sürece test eden parametrik olmayan bir test olmaması, parametrik olmayan testlerin hipotezleri test etmesi ve parametreleri tahminlememeleri olarak sıralanabilir (Sahu, Pal ve Das, 2015, s. 145-146). Bütün parametrik testlerin normal dağılım ve varyans homojenliği olmak üzere temelde iki varsayımı vardır. Birçok araştırmacı parametrik testlerin varsayımları tam anlamıyla karşılanmadığında yanlış hesaplamalar gibi istenmeyen sonuçlar verebileceğinden habersizdir. Varsayımların karşılanmadığı durumlarda karşılaşılan hatalı hesaplamalar ise hatalı araştırmaların ortaya çıkmasına neden olmaktadır (Erceg-Hurn, Wilcox ve Keselman, 2013; Larson-Hall, 2016). Varsayımların ihlali durumunda parametrik testlerin kullanımı sonucunda yanlış çıkarımlar meydana gelebilir ya da doğru çıkarımlar meydana gelse bile veri içinde bulunan ilişkiler kaçırabilir. Varsayım ihlalinde gerekli düzeltmeler yapılabilir veya alternatif analitik stratejiler uygulanabilir (Fidell ve Tabachnick, 2003). Parametrik olmayan testler ise varsayımları olmayan testler değil, sadece parametrik testlere oranla daha az varsayıma sahiptir (Marascuilo ve McSweeney, 1977). Öte yandan, çeşitli araştırmacılar parametrik testlerin varsayımlarının karşılanmadığı durumlarda parametrik olmayan testlerden ziyade varsayımların karşılanmadığı durumlarda parametrik operasyonların gerçekleştirilmesine olanak sağlayan sağlam (robust) istatistiklerin kullanılmasını önermektedir (Larson-Hall, 2016; Loewen ve Plonsky, 2015). Yaygın kullanılan karşılaştırma testleri ve varsayımları Tablo 1’de verilmiştir.

Table 1

Yaygın Kullanılan Karşılaştırma Testleri

Karşılaştırma	Bağımlı (çıkıtı) Değişken	Bağımsız (açıklayıcı) Değişken	Parametrik Test			Parametrik olmayan Test
			Testin İsmi	Varsayımlar	Varsayımlar nasıl kontrol edilebilir?	
2 bağımsız grubun ortalamaları	Ölçek	Nominal (İkili)	Bağımsız Örneklem t-Testi	Veriler normal dağılım göstermeli Veriler bağımsız gruplardan toplanmış olmalı (Bağımsız örneklem) Grup varyanslarının eşit olmalı Aykırı değer olmamalı	Normal Dağılım: İki grubun da histogramları incelenebilir. Shapiro Wilk ve Kolmogorov Smirnov Testleri uygulanabilir. Varyans Homojenliği: Levene’s Testi uygulanabilir.	Mann Whitney U Testi/Wilcoxon Sıra Toplamı Sınaması
3’den fazla bağımsız grubun ortalamaları	Ölçek	Nominal	Tek Yönlü ANOVA	Residuals/Kalıntılar normal dağılım göstermeli Varyans eşitliği olmalı	Normal Dağılım: Q-Q Plot ve Histogramlar incelenebilir. Shapiro Wilk ve Kolmogorov Smirnov Testleri uygulanabilir. Varyans Homojenliği: Levene’s Testi veya Bartlett’s Testi uygulanabilir.	Kruskall-Wallis Testi

Bağımlı örneklerin ortalama farkı	Ölçek	Zaman/Durum değişkeni	Bağımlı Örnekler t-Testi	Veriler normal dağılım göstermeli (Bağımsız örnekler) Grup varyanslarının eşit olmalı Aykırı değer olmamalı	Normal Dağılım: İki grubun da histogramları incelenebilir. Shapiro Wilk ve Kolmogorov Smirnov Testleri uygulanabilir. Varyans Homojenliği: Levene's Testi uygulanabilir.	Wilcoxon İşareti Sıralar Testi
Aynı örneklerde 3'den fazla ölçüm	Ölçek	Zaman/Durum değişkeni	Tekrarlayan Ölçümlerde ANOVA	Residuals/Kalıntılar normal dağılım göstermeli Sferiklik; her bir tekrarlanan ölçüm çifti arasındaki farkların varyanslar eşit olmalı	Normal Dağılım: Kalıntıların histogramları incelenebilir. Küresellik: Mauchly'nin sferiklik testi uygulanabilir.	Friedman Testi

(Kaynak: Hatch ve Lazaraton, 1991; Larson-Hall, 2016; Marshall ve Boggis, 2016; Peacock ve Peacock, 2011)

Tablo 1' de karşılaştırılan grup sayısına göre sıklıkla kullanılan parametrik yöntemler ile bu yöntemlerin kullanılabilmesi için karşılanması gereken varsayımlar verilmiştir. Bununla birlikte ilgili varsayımların nasıl kontrol edileceği ve varsayım ihlalinde ise kullanılması planlanan parametrik testin parametrik olmayan karşılığı verilmiştir. Benzer şekilde yaygın olarak kullanılan ilişkilendirme testleri ve varsayımları Tablo 2'de verilmiştir.

Table 2

Yaygın Kullanılan İlişkilendirme Testleri

Karşılaştırma	Bağımlı (çıkıtı) Değişken	Bağımsız (açıklayıcı) Değişken	Parametrik Test			Parametrik olmayan Test
			Testin İsmi	Varsayımlar	Varsayımlar nasıl kontrol edilebilir?	
2 sürekli değişken arasındaki ilişki	Ölçek	Ölçek	Pearson Korelasyon Katsayısı	Veriler doğrusal olarak ilişkili olmalı	Normal Dağılım: Q-Q Plot ve Histogramlar incelenebilir. Shapiro Wilk ve Kolmogorov Smirnov Testleri uygulanabilir.	Spearman Korelasyon Katsayısı
				Veriler bağımsız olarak toplanmış olmalı		
				Puan veya sıralı olarak ölçülen veriler sürekli olmalı		
				Veriler normal dağılım göstermeli		
				Homoscedastic (yani eşit istatistiksel varyansa sahip) olmalı		

Bir yordayıcı değişkenin değeri yardımıyla bir değişkenin değerini tahmin etmek veya önemli ilişkiler aramak	Ölçek	Herhangi bir değişken	Basit Doğrusal Regresyon		
2 kategorik değişken arasındaki ilişkiyi incelemek	Nominal (İkili)	Herhangi bir değişken	Lojistik Regresyon	Hücrede beklenen değer sayısının %80'i 5'ten büyük olmalı	Varsayımlar karşılanmadığında Fisher'ın Kesin Ki Kare Testi kullanılabilir veya hücreler birleştirilerek varsayımlar karşılanabilir.
	Kategorik	Kategorik	1'in altında frekansa sahip hücre bulunmamalı	2x2 matrislerinde tüm hücrelerde beklenen sayı 5'ten büyük olmalı	Ki Kare Testi

(Kaynak: Marshall ve Boggis, 2016; Peacock ve Peacock, 2011)

Tablo 2 incelendiğinde yaygın olarak kullanılan parametrik ilişkilendirme testleri ve varsayımları ile belirtilen varsayımlar karşılanmadığında kullanılması gereken parametrik olmayan karşılıkları verilmiştir.

Toplanan verilerin analize başlamadan önce analiz için doğru ve uygun olduklarını kontrol etmek önemlidir (Blischke ve diğerleri, 2011; Wilkinson, 1999). Veriler analiz edilmeden önce analize hazırlanma süreci oldukça önemlidir. IBM çalışanı olan programcı Wilf Hey, analize hazırlama aşamasını GIGO kuralı olarak da bilinen “Çöp İçeri, Çöp Dışarı/Garbage In, Garbage Out” ifadeleriyle belirtilmiştir. Bu ifadeye göre analize girecek olan verinin kalitesi alınacak sonuçları etkilemektedir. Verilerin analize hazırlama aşamasında kayıp verilerin teşhis edilip gerekli işlemlerin yapılması ve planlanan analizlerin varsayımlarının test edilmesi bu nedenle önemlidir. Plonsky (2017) araştırmada şeffaflığı etkileyen kayıp veya bildirilmemiş verilerin okuyucuların bulguları yorumlama becerilerini sınırladığını belirtmiştir. Veri setinde kayıp veri olmasını önlemek için uygulanabilecek en iyi yöntem çalışmanın detaylı bir şekilde kurgulanmış olmasıdır (Wisniewski, Leon, Otto ve Trivedi, 2006). Veri setinde kayıp veri olması durumunda araştırmacıların uygulayabileceği çeşitli yöntemler vardır. Bunun için araştırmacıların takip etmesi gereken adımlar (Salgado, Azevedo, Proença ve Vieira, 2016);

- Kayıp verinin desenini ve nedenlerini belirlemek,
- Kayıp veri oranını hesaplamak,
- En uygun kayıp veri atama veya silme yöntemini belirlemek.

Kayıp veri türleri üç ana kategoride incelenebilir, bunlar (Curley, Krause, Feiock ve Hawkins, 2017);

Tamamen Rastgele Eksik (TRE): Kayıp olan veri, veri setinde bulunan gözlenen veya gözlemlenmeyen değerlerin özelliklerinden bağımsızdır.

Rastgele Eksik (RE): Kayıp olan veri, gözlemlenen değerler ile açıklanabilir, gözlemlenen veriler ele alınarak kayıp rastgele dağıtılabilir.

Rastgele Eksik Olmayan (REO): Kayıp olan veriler, gözlemlenmeyen değerlere bağlıdır ve gözlemlenen değerler ile açıklanamaz.

Kayıp veri atama yöntemleri ise silme, atama ve makine öğrenmesi teknikleri olmak üzere üç gruba ayrılabilir (García-Laencina ve diğerleri, 2009).

Silme yönteminde complete-case ve available-case işlemleri uygulanır. Silme yöntemleri sıklıkla kullanılan yöntemler olsalar da veri kaybına yol açarlar ancak önerilmez. Veri kaybı gibi dezavantajlarına rağmen TRE ve RE durumlarında güçlü bir yöntem olduğu ifade edilir (Demir ve Parlak, 2012). Complete-case işlemi analizde kullanılan herhangi bir değişkende eksik verilere sahip durumlar analizden çıkarılır. Available-case işlemi ise sadece ilgilenilen değişkenlere sahip olan durumlar analizde kullanılır (García-Laencina ve diğerleri, 2009).

Atama yönteminde ise kayıp veriye değer ataması yapılır. Bunun için çeşitli teknikler vardır (García-Laencina ve diğerleri, 2009). Basit atama tahmini bir değer kayıp veriye atanması ile olur. Ortalama atamada ise verilerin ortalaması alınarak kayıp veriye değer atanır. Bunların yanı sıra regresyon analizinin kullanılması ile uygulanan koşullu ortalama atama yöntemi ile de atama yapılabilir (Demir ve Parlak, 2012). Makine öğrenmesi teknikleri ise Çok Katmanlı Algılayıcı Atama, K-En Yakın Komşu Atama veya Öz Düzenleyici Haritalar Ataması algoritmaları uygulanabilir (García-Laencina ve diğerleri, 2009). Belirtilen kayıp veri silme ve atama yöntemlerinin yanı sıra en çok olabilirlik (maximum likelihood-ML) ve çoklu veri atama (multiple imputation-MI) gibi yaklaşımlar da oldukça yaygın kullanılmaktadır (Allison, 2009).

Verilerin ön incelenmesinin amaçları aşağıdaki gibidir (Blischke ve diğerleri, 2011, s.159);

- Verilerin kaynağını doğrulamak
- Verilerin belirtilen değişkenleri içerdiğini doğrulamak
- Ölçü birimlerini doğrulamak
- Verileri silerek veya mümkünse açıkça yanlış olan sonuçları düzelterek "temizlemek"
- Aykırı değerleri veya başka şekilde olağandışı sonuçları belirlemek
- Eksik verileri kontrol etmek
- Diğer olağandışı veri özelliklerini tanımlamak

Bununla birlikte nicel çalışmalarda sıklıkla kullanılan veri toplama araçları anketler, ölçekler ve testlerdir. Ancak anket ve ölçek terimleri sıklıkla birbirleri yerine kullanılıyor olsalar da bunlar farklı veri toplama araçlarıdır. Değişkenlerin ölçüldüğü ve toplam puanın bir anlam taşıdığı durumlarda veri toplama aracı için *ölçek* ifadesi kullanılmalıdır. Anket terimi ise toplam puanın anlam taşımadığı, bir konu hakkında katılımcıların bilgilerini toplamak amacı ile kullanılan veri toplama araçlarıdır (Erkuş, 2007; 2010). Anketlerde toplam puan anlam taşımadığı için geçerlik ve güvenilirlik gibi ölçütlere

bakılarak değerlendirilmezler, ölçeklerde ise güvenilirlik ve geçerlik özelliklerine sahip olması gerekir (Erkuş, 2010; Ercan ve Kan, 2004). Çok boyutlu ölçeklerde güvenilirlik katsayısı olarak her bir boyuta ilişkin cronbach alfa değeri, ölçeğin tamamına ilişkin ise tabakalı alfa değeri kullanılabilir. Ölçeklerden toplanan verilerin analizlerinde bağımlı/bağımsız örneklem t-testi, ANOVA gibi çıkarımsal istatistik yöntemlerinin kullanılması uygundur. Anketler ile toplanan verilerin analizlerinde ise betimleyici istatistikler ve chi-square gibi istatistik testler kullanılabilir. Bu iki veri toplama araçlarının geliştirme süreçleri de farklılık gösterir. Ayrıca, alanda mevcut olan anketlerin ve ölçeklerin yeniden kullanılması durumunda yeni yapılacak çalışmanın ve yeni katılımcıların gerekliliklerine göre uyarlanmalıdır (Iwaniec, 2020). Benzer bir karışıklık ise likert tipi soru ve likert ölçekleri arasında yaşanmaktadır. Likert tipi soru ve likert ölçeği birbirlerinden farklı iki terimdir. Bu iki terimin araştırmalarda hatalı kullanılması, uygun olmayan istatistik testlerin kullanılmasına neden olabilir. Likert tipi sorularda puanlanan maddelerin toplam puanının hesaplanması gibi bir amaç güdülmez. Likert ölçeğinde ise bu toplam alınarak bir çıkarım yapılır (Turan, Şimşek ve Aslan, 2015).

Araştırmalarda dikkat edilmesi gereken bir diğer önemli husus ise verilerin toplandığı grubun özellikleridir. Bunun için araştırmacıların evren, örneklem ve çalışma grubu kavramlarıyla ilgili bilgi sahibi olmaları önemlidir. Çalışmalarda, araştırılan konunun kapsadığı ve sonuçların genellenmek istediği bireylerin toplamına evren denir (Nachimas ve Nachimas, 1996). Karasar'a (2012) göre evren ikiye ayrılmaktadır. Genel evren ulaşılması güç hatta imkansız olan soyut bir kavramdır. Genel evrene ulaşmak oldukça zor olduğundan çalışma evreni (hedef evren) kavramı geliştirilmiştir (Karasar, 2012). Evren üzerinde gerçekleştirilecek olan çalışmalar para ve zaman açısından çalışmacılar için oldukça pahalı ve kontrol edilmesi oldukça zordur. Bu nedenle araştırmacılar genellikle evrenin küçük alt grupları üzerinde çalışmayı tercih etmektedirler. Bu alt gruplar örneklem olarak adlandırılmaktadır. Bu şekilde evreni temsil edebilecek örneklem seçimi süreci örnekleme olarak tanımlanmaktadır. Örneklemede 2 önemli özelliğe dikkat edilmelidir. Bunlardan ilki örneklemin büyüklüğü, ikincisi ise bu örneklemin nasıl seçildiğidir (Peacock ve Peacock, 2011). Örneklemede birçok teknik vardır. Bu teknikler rastgele ve rastgele olmayan örnekleme olarak ikiye ayrılır. Rastgele örneklemede evrendeki tüm kişilerin örnekleme seçilmesi için eşit şansları vardır. Rastgele olmayan örneklemede ise kişiler belirli özelliklere göre seçilirler. Bu örnekleme tekniğinde hangi bireylerin örnekleme olacağı ve hangi bireylerin olmayacağını belirlemek için bir gerekçe gösterilmelidir (Taherdoost, 2016). Örneklem seçim yöntemleri çalışmanın desenini de etkileyen önemli bir özelliktir. Örneğin, yarı-deneysel çalışmalar deneysel olmayan çalışmaların alt türüdür. Deneysel olmayan çalışmalar ile tam deneysel olan çalışmaları ayıran ana özellik deneysel olmayan çalışmalarda rastgele örnekleme olmamasıdır (Rogers ve Révész, 2020; Plonsky, 2017). Yarı deneysel çalışmalar rastgele örnekleme olmadan tam deneysel çalışmaları taklit etmeye çalışır (Cook ve Wong, 2008; Kirk, 2009). Ayrıca, rastgele örnekleme olmadığı durumlarda Balcı (2013) çalışmalarda evren içerisinden seçilen alt grubun örneklem yerine *çalışma grubu* olarak adlandırılmasını önermiştir.

Çalışmalarda sıklıkla kullanılan ifadelerden olan *statistically significant* ve *statistically meaningful* ifadeleri farklı anlamları taşıyan ifadeler olsalar da Türkçe'de *manidar* ve *anlamlı* kelimeleri ile birbirlerinin yerine kullanılmaktadır. *Significance* terimi günlük kullanımda önemli veya anlamlı anlamında kullanılırken istatistik alanında hipotez test etmede teknik bir anlama sahiptir (Stout ve Ruble, 1995). *P* değeri ile yorumlanabilen bir durum olan *statistical significance*, bir çalışmanın sonuçları arasındaki farkın belirginliği ile ilgilidir, çalışma sonuçlarını tam anlamıyla açıklamaz gözlemlenen farkın tesadüfen ortaya çıkıp çıkmadığı ile ilgilenirken bilimsel anlamda önemli veya anlamlı olduğunu belirtmez ve çalışma sonuçlarını tam anlamıyla açıklamaz (Norman ve Streiner, 2003; Stout ve Ruble, 1995; Wasserstein ve diğerleri, 2019). İstatistiksel analiz sonucunda *p* değerinin .05 altında çıkması sonuçların *istatistiksel olarak anlamlı (statistically significant)* olduğunu gösterir. Alanyazında pratik anlamlılık veya klinik anlamlılık olarak da bilinen *statistically meaningful* ifadesi ise çalışmada incelenen olgunun önemini ifade eder ve bu etki değeri yorumlanarak belirtilir (Kraemer, 2014; Sullivan ve Feinn, 2012). Etki değerinin raporlanmasında ise sıklıkla Cohen's *d* kullanılır (Schuele ve Justice, 2006). Bu nedenle, sonuçlar arasında istatistiksel olarak anlamlı fark bulunmuş olan çalışmalarda bu farkın pratik anlamlılığının ifade edilebilmesi için mutlaka etki büyüklüğü değerlerinin de verilmesi gerekmektedir. Ancak, Türkçe'de *manidar* ve *anlamlı* ifadeleri birbirlerinin yerlerine kullanılarak özgün anlamlarını kaybettikleri için çalışmaların raporlamalarında buna bağlı yaşanan sorunlar sonuçların yeterince açık bir şekilde ifade edilememesine neden olabilir. Nicel çalışmalarda temelde dikkat edilmesi gereken süreçler bu aşamaya kadar özetlenerek ifade edilmiştir.

Amaç

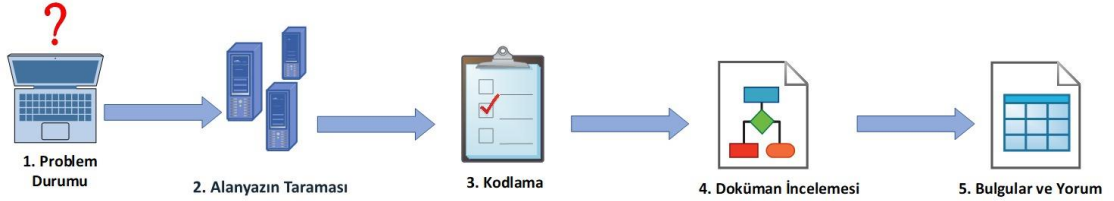
Bu bağlamda bu çalışmanın temel amacı İngilizce eğitimi alanında yapılan nicel çalışmaların metodolojik açıdan incelenerek bulguların ortaya konulmasıdır. Bu amaca ulaşabilmek için incelenen çalışmalarda veri toplama aracının niteliği, verilerin analize hazırlanması ve veri analizi aşamasında kullanılan istatistiksel tekniklerin uygunluğu detaylandırılarak açıklanmıştır. Ayrıca elde edilen bulgular ışığında araştırmacılar için çeşitli önerilerin sunulması amaçlanmaktadır.

Önem

Dünyada en çok konuşmacıya sahip dil olan İngilizce, konuşanlarına sağladığı bireysel ve toplumsal ayrıcalıklar ile İngilizce eğitimini dünyanın her ülkesinde bir norm haline getirmiştir (Ethnologue, 2019; Mahu, 2012). Dolayısıyla İngilizce eğitiminin niteliği ve bunu geliştirme çabaları önem kazanmıştır. Bu bağlamda, Avrupa Komisyonu (2019) dil eğitiminde benimsenecek politika ve uygulamaların kanıta dayalı geliştirilmesini savunmaktadır. Bilimsel araştırmaların temelini oluşturduğu bu kanıtların güvenilirliğini sağlamak için ise araştırmaların kalitesi doğru yöntem ve tekniklerin kullanılması ile sağlanmalıdır. Bu nedenle çalışmanın yordamsal istatistiklere dayalı araştırmaların kuramsal temelleri, veri toplama araçları ve veri analizleri ile ilgili teorik bilgilerin kullanımında yaşanan sorunları ortaya koyarak araştırmacılara çözüm önerileri sunması açısından alanyazına katkı sağlayacağı düşünülmektedir.

2. YÖNTEM

Bu çalışma kapsamında ele alınan araştırmalar nitel araştırma yöntemlerinden olgu bilim desenine göre ele alınmış ve veriler doküman incelemesi ile analiz edilmiştir. Araştırmada izlenen süreç Şekil 1’de gösterilmiştir.



Şekil 1. Araştırmada İzlenen Süreç

Şekil 1’de ifade edilen aşamalar aşağıdaki gibidir:

1. *Problem Durumu*: Çalışmanın amacı ve bu amacı gerçekleştirmek için cevaplanması gerekli alt problemlerde problem durumu açıklanmıştır.

2. *Alanyazın Taraması*: Çalışmanın alanyazın taramasında aşağıdaki kriterlere uygun olarak araştırmalar yapılmıştır. Buna göre belirlenen kriterler çalışmanın amacına uygun olarak şu şekilde ifade edilmiştir:

- İngiliz Dili Eğitimi alanında yapılmış olması
- Çıkarımsal istatistiklerin kullanılmış olması
- ULAKBİM dizininde indekslenen dergilerde yayımlanmış olması
- 2013 ile 2020 Haziran arasında yayımlanmış olması
- Elektronik ortamda açık erişim (open-access) sağlıyor olması

Buna göre ULAKBİM veri tabanında bu kriterlere uygun olarak belirlenen 15 farklı dergide 30 makale araştırmada doküman analiziyle incelenmiştir.

3. *Kodlama*: Araştırma kapsamında eldeki veriler araştırmacılar tarafından geliştirilen makale inceleme formu ile analiz edilmiştir (Ek. Veri Analiz Formu). Bu form araştırmanın amacına uygun olarak belirlenmiş beş soru ve bunlara ilişkin açıklamalardan oluşmaktadır. Sorular araştırmanın yöntemi, verilerin toplandığı grup, kayıp veri ifadesi, veri toplama aracı ve yapılan analizler başlıklarından oluşmaktadır. Burada araştırmacıların ilgili soru başlığına ilişkin detaylı görüşlerini de yazdıkları birer açıklama sütunu bulunmaktadır. Hazırlanan taslak form daha önce doküman inceleme çalışması olan iki farklı öğretim üyesine uzman görüşü için sunulmuştur. Uzmanlardan gelen dönütlere uygun düzenlemeler yapılarak form nihai haline kavuşturulmuştur. İnceleme formunun inandırıcılık ve aktarılabirliğine yönelik kanıt elde edebilmek için rastgele seçilen beş makale üç araştırmacı tarafından değerlendirilmiştir. Araştırmacılar dışında iki uzmandan da forma göre makaleleri değerlendirmeleri istenmiştir. Hem

araştırmacılar hem de uzmanların kendi aralarında ve karşılıklı değerlendirmelerinde sonuçları benzer şekilde ifade ettikleri bulunmuştur. Uzman görüşlerinden elde edilen bulgular için Miles-Huberman (1994) tarafından önerilen aşağıdaki uyuşum yüzdesi dikkate alınmıştır:

$$\text{Güvenirlik} = \frac{\text{Uzlaşma Sayısı}}{(\text{Uzlaşma Sayısı} + \text{Uzlaşmama Sayısı})}$$

Bu hesaplama sonucu elde edilen uyuşum yüzdesinin %70'in üzerinde çıkması güvenilir kabul edilmektedir (Miles-Huberman,1994; Tavşancıl ve Aslan, 2001). Uzman görüşleri arasındaki uyuşum yüzdesinin ilk turda %100 olduğu görülmüştür. Böylece kullanılan veri analiz formunun açık ve anlaşılır olduğu sonucuna varılmıştır.

4. Doküman İncelemesi: Bu çalışmada araştırmanın amacına uygun olarak hazırlanan problemlere cevap bulabilmek için eldeki yazılı materyaller doküman incelemesi ile analiz edilmiştir.

5. Bulgular ve Yorum: Betimsel analiz sonucunda elde edilen veriler araştırmanın amacına ve problem durumuna uygun olarak tablolaştırılarak yorumlanmıştır.

Geçerlik ve Güvenirlik Çalışmaları

Bu çalışmada nitel araştırma yöntemi benimsendiği için bu araştırmaların doğasına uygun olarak *inandırıcılık (iç geçerlik), aktarılabilirlik (dış geçerlik), tutarlılık (güvenirlik) ve teyit edilebilirlik (objektiflik) özellikleri güvenilirlik ve geçerlik içeriğinde değerlendirilmiştir.*

Belirtilen özelliklerin sağlanabilmesi için araştırmacılar tarafından çalışmanın her aşamasında sürece ilişkin tüm detaylara dikkat edilmeye çalışılmıştır. Bunun için ilk olarak detaylı bir raporlama sürecine gidilmiştir. Araştırma kapsamında incelenecek olan değişkenlerle ilgili kuramsal çerçeve, incelenecek olan makalelerin seçim sürecinde ele alınan ölçütler ve araştırmanın süreç tasarımının detaylı ve sade bir dille ifade edilmesine özen gösterilmiştir. Bununla birlikte çalışmanın örnekleme, seçkisiz olmayan örnekleme yöntemlerinden uygunluk (elverişlilik) örneklemesine göre belirlenen kriterlere uygun olarak seçilmiştir. Burada temel amaç araştırmanın amacı doğrultusunda elverişlilik açısından uygun olan araştırmaların çalışmaya dahil edilmesidir. Bununla birlikte araştırmacılar tarafından uzman görüşü alınarak geliştirilen inceleme formu tüm çalışmaların incelenmesinde kullanılmıştır. Çalışmayı yürüten üç araştırmacı da eş zamanlı ve bağımsız olarak aynı formu kullanmış ve analiz sonucunda bir araya gelerek elde edilen sonuçlardaki farklılıklar uzlaşmayla giderilmiştir. Analizler sonucunda elde edilen bulgular betimsel olarak tablolarda verilmiştir. Burada amaç bulgulara herhangi bir yorum yapılmaksızın incelenen makalelerden elde edilenlerin yansıtılmasıdır. Elde edilen bulgular araştırmacılar tarafından çalışmanın amacı, kuramsal çerçeve, ilgili alanyazın ışığında sorgulanarak yorum, sonuç ve tartışma bölümlerinde ifade edilmiştir.

3. BULGULAR

Araştırmada İngiliz dili eğitimi alanında 2013-2020 yılları arasında yayımlanmış toplam 30 makale incelenmiştir. Makalelerin dergilere ve yayımlandıkları yıllara ait bilgiler Tablo 3'de verilmiştir.

Tablo 3

İncelenen Makalelerin Dergilere ve Yıllara Göre Dağılımı

Dergiler	Yayın Yılı								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	1	-	-	-	-	1
Çukurova Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	-	2	-	-	-	2
Cumhuriyet International Journal of Education	-	-	-	-	-	-	1	1	2
Eğitimde Nitel Araştırmalar Dergisi -ENAD	-	-	-	-	-	1	-	-	1
Eurasian Journal of Applied Linguistics	-	-	-	-	-	-	-	1	1
Eurasian Journal of Educational Research	-	1	-	-	-	-	-	-	1
Fırat Üniversitesi Sosyal Bilimler Dergisi	-	-	-	-	-	-	1	-	1
Hacettepe Üniversitesi Eğitim Fakültesi Dergisi	-	-	1	-	-	-	-	-	1
International Online Journal of Education and Teaching (IOJET)	-	-	-	-	-	-	-	3	3
Journal of Language and Linguistic Studies	1	-	-	2	1	5	2	-	11
Marmara Üni. Atatürk Eğitim Fak. Eğitim Bil. Dergisi	-	-	1	-	-	-	-	-	1
Novitas-ROYAL	-	2	-	-	-	-	-	-	2
Pamukkale Üniversitesi Eğitim Fakültesi Dergisi	-	-	-	-	-	-	-	1	1
Turkish Online Journal of Qualitative Inquiry	-	1	-	-	-	-	-	-	1
Turkish Studies Educational Sciences	-	-	-	-	-	1	-	-	1
Toplam	1	4	2	3	3	7	4	6	30

Tablo 3 incelendiğinde son yıllarda İngiliz dili eğitimi alanında yapılan çalışmalarda çıkarımsal (yordamsal) istatistiklerin (korelasyon, regresyon, t-testi, ANOVA, Ki-kare, vb.) ilk yıllara nazaran sayıca çokluğu görülmektedir. Toplam 15 farklı dergide 30 yayın araştırmada incelenmiştir. İncelenen çalışmalarda kullanılan yöntem bilgisi Tablo 4'de verilmiştir.

Tablo 4

Araştırmada Kullanılan Yöntem

Yöntem	Yıllar								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Nicel	1	-	2	1	1	1	2	1	9
Karma	-	4	-	2	2	3	1	-	12
Belirtilmemiş	-	-	-	-	-	3	1	5	9
Toplam	1	4	2	3	3	7	4	6	30

Tablo 4 incelendiğinde çalışmalardan on iki tanesinde karma yöntem uygulandığı görülmüştür. Bununla birlikte dokuz çalışmanın yöntem kısmında sadece yapılan analizlerden bahsedilmiş ve araştırma yöntemi ile ilgili detaylı bilgi verilmemiştir. Bu çalışmaların bulgularına bakıldığında hem nicel hem de nitel ya da sadece nicel analizlerin yapıldığı görülmüştür. Araştırmada verilerin toplandığı grupla ilgili bilgiler ise Tablo 5'de ifade edilmiştir.

Tablo 5

Araştırmada Verilerin Toplanıldığı Grup İfadesi

Veri Toplanan Grup İfadesi	Yıllar								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Örneklem	1	2	2	1	2	4	-	3	15
Çalışma Grubu (Katılımcı-Participant)	-	-	-	-	-	1	4	-	5
Örneklem/Katılımcı (Participant)	-	-	-	1	-	-	-	1	2
Belirtilmemiş	-	2	-	1	1	2	-	2	8
Toplam	1	4	2	3	3	7	4	6	30

Araştırmalarda çalışılan grup iç ve dış geçerlik için önemlidir. Katılımcıların gruplara atanmasında yansız atama yapılıp yapılmadığının iç geçerliği, araştırmadan elde edilen sonuçların genellenebilirliği ise dış geçerliği ifade etmektedir. Özellikle evrenden olasılıklı örneklem seçilmediğinde dış geçerlik düşmektedir. Tablo 5 incelendiğinde araştırmaların yarısında örneklem ifadesi yer alırken sekiz araştırmada örneklem seçimine yönelik herhangi bir ifade yer almamıştır. Araştırmalardan iki tanesinde ise örneklem-katılımcı ifadesi yer alırken beş araştırmada sadece katılımcılar ifadesi yer almıştır. Bununla birlikte sekiz araştırmada herhangi bir şekilde veri toplanan grup için örneklem ya da çalışma grubu ifadesi kullanılmamıştır. Bu çalışmaların bir bölümünde çeşitli demografik bilgiler sunulmuştur. Bu da incelenen araştırmaların dış geçerlikleri ile ilgili şüphe uyandırmaktadır.

Bununla birlikte veri toplanılmasında incelenen bir başka durum kayıp verinin raporlanmasıdır. İncelenen toplam otuz çalışmanın yirmi dokuzunda kayıp veri sorunundan bahsedilmemiştir. Makalelerin incelenmesi sırasında özet, katılımcı sayıları ve analiz tablolarında üç makalede sayılar arasında tutarsızlıklar olduğu görülmüştür. Bununla ilgili elde edilen bulgular ise Tablo 6’da verilmiştir.

Tablo 6

Kayıp Veri İncelemesi

Kayıp Veri	Yıllar								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Kayıp veri var	1	-	-	-	-	1	3	2	6
Kayıp veri yok	-	4	2	3	3	6	1	4	24
Toplam	1	4	2	3	3	7	4	6	30

Tablo 6 incelendiğinde yapılan çalışmalardan altı tanesinde kayıp verilerin olduğu sadece bir çalışmada bu durumun raporlandığı ancak bu çalışmada da kayıp veri atama yöntemlerinden hangisinin uygulanarak analize devam edildiği ifade edilmemiştir. Araştırmada incelenen bir diğer önemli husus veri toplama araçlarıdır. Veri toplama araçları uygulanabilecek analizlerin belirlenmesi açısından önemli görülmüş ve bulgular Tablo 7’de sunulmuştur.

Tablo 7

Veri Toplama Araçları

Veri Toplama Araçları	Yıllar								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Anket	1	2	-	2	-	1	-	2	8
Yarı yapılandırılmış görüş formu	-	2	-	1	1	2	1	-	7
Haftalık görüş yazıları	-	1	-	-	-	-	-	-	1
Kültürlerarası görevler	-	1	-	-	-	-	-	-	1
Açık uçlu sorular	-	1	-	-	-	-	-	-	1
Çağrışım tekniğine dayalı görüşme	-	1	-	-	-	-	-	-	1
Deneme yazıları	-	1	1	-	-	-	-	-	2
Ölçek	-	-	1	2	3	4	6	4	20
Gözlem kontrol listesi	-	-	-	1	-	-	-	-	1
Analitik puanlama rubriği	-	-	-	-	1	-	-	-	1
Yazma görevleri	-	-	-	-	-	1	-	-	1

Akran yansıtıcı tartışma formları	-	-	-	-	-	1	-	-	1
Konuşma becerisi kontrol listesi	-	-	-	-	-	-	-	1	1
Odak grup görüşmeleri	-	-	-	-	-	1	-	-	1
Çoktan Seçmeli Test	-	-	-	-	-	2	-	1	3

İncelenen çalışmalar nicel ve karma çalışmalardan oluşmaktadır. Bu durum Tablo 7 incelendiğinde veri toplama araçlarında da görülmektedir. Burada ilk dikkat çeken husus özellikle çalışmalarda nicel verilerin ölçek ve anketlerle toplandığıdır. Nitel veriler ise yarı yapılandırılmış görüş formlarından, odak grup görüşmelerinden, çağrışım tekniğine dayalı görüşmelerden, açık uçlu sorulardan, haftalık görüş yazılarından, akran yansıtıcı tartışma formlarından elde edilmiştir. Karma çalışmalar olarak ifade edilen dört çalışmada araştırmacıların nicel verilerini desteklemek için görüş formlarından aldıkları verileri kullandıkları görülmüştür. Dolayısıyla bu türden çalışmaların karma çalışmalardan ziyade nitel ile desteklenmiş nicel çalışmalar olarak ifade edilmesi daha uygun olabilir. Bununla birlikte incelenen çalışmalarda görülen en yaygın durumun anket ve ölçek kavramlarının birbirinin yerine kullanılmasıdır. Giriş bölümünde de bu kavramlar ve aralarındaki farklar vurgulanmıştır. Birbirinden farklı özelliklere sahip bu iki veri toplama aracından toplanan verilere uygulanabilecek istatistiksel analizler de farklıdır. Ayrıca ölçeklerin geçerlik ve güvenilirlikle ilgili istatistiğe dayalı daha katı kuralları olması gerekirken birçok çalışmada bu bulgulara yer verilmediği görülmüştür. Birden fazla boyuttan oluşan ölçeklerin güvenilirlik hesaplamalarında tabakalı alfa gibi tekniklerle güvenilirlik katsayısının hesaplanması önemli iken bununla ilgili herhangi bir bulguya rastlanmamıştır. Ölçeklerin geliştirilmesi ya da kültürel uyarlamasının yapılarak kullanıldığı durumlarda yapılması gerekli olan açımlayıcı ve doğrulayıcı faktör analizi gibi çalışmalarda da eksiklikler olduğu görülmüştür. Altı çalışmada ise anketler doğru bir şekilde değerlendirilerek yüzde ve frekans verilmiş ancak farklı gruplardan toplanan veriler toplam puan üzerinden karşılaştırılarak yanlış sonuçlar elde edilmiştir. Yine çoktan seçmeli testlerin kullanıldığı çalışmalarda testlerin de psikometrik özelliklerine yer verilmediği söylenebilir. Veri toplama araçlarından elde edilen veriler parametrik ve parametrik olmayan testler kullanılarak analiz edilmiştir. Bununla ilgili elde edilen bulgular Tablo 8 de verilmiştir.

Tablo 8

Veri Analiz Yöntemleri

Veri Analiz Yöntemi	Yıllar								Toplam
	2013	2014	2015	2016	2017	2018	2019	2020	
Parametrik Test	1	4	1	3	2	5	2	6	24
Parametrik olmayan Test	-	-	-	-	-	-	1	-	1
Parametrik ve Parametrik olmayan	-	-	1	-	1	2	1	-	5
Toplam	1	4	2	3	3	7	4	6	30

Tablo 8 deki sonuçlar incelendiğinde 24 makalede toplanan verilerin sadece parametrik, 1 tanesinde non parametrik, 5 tanesinde ise parametrik ve parametrik olmayan testler ile verilerin analiz edildiği görülmüştür. Parametrik testlerle analiz edilen verilerde varsayımların kontrol edilmemesi, örneklem sayısının yetersiz olmasına rağmen parametrik testlerin kullanılması, varsayımların kontrol edildiği ifade edilmesine rağmen buna dair herhangi bir bulgunun verilmemesi, kullanılan karşılaştırma testlerinden elde edilen anlamlı farkların etki büyüklüklerinin raporlanmaması, veri toplama aracı ile bu araçtan toplanan verilerle yapılabilecek analiz arasında uyumsuzluk (örneğin anketten toplanan verilere kay kare yerine parametrik veya parametrik olmayan analizler yapılması) gibi sorunlar tespit edilmiştir.

Bununla birlikte ölçek verilerinin kullanıldığı bazı çalışmalarda alt boyutlara ilişkin normallik varsayımlarının sağlandığı durumlarda parametrik, sağlanmadığı durumlarda parametrik olmayan testlerin tercih edildiği görülmüştür. İncelenen çalışmaların büyük bir bölümünde ise analizlerin yapıldığı programın isminin yaygın olarak ifade edildiği görülmektedir. İstatistiksel analizlerin doğruluğu, varsayımların kontrolü, veri toplama aracı ve yapılan analizin uygunluğu gibi konuların çalışmalarda ön plana çıkarılmasının daha önemli olduğu, veri analizinde ücretli ve ücretsiz kullanılacak birçok programında bulunduğu ifade edilebilir.

4. SONUÇ, TARTIŞMA VE ÖNERİLER

Bu çalışmada ULAKBİM dergi tabanında yer alan 15 farklı derginin, 2013-2021 yılları arasında İngilizce eğitimi alanında yaptığı toplam 30 çıkarımsal istatistiğin kullanıldığı makale incelenmiştir. Makaleler kullanılan yöntem, örneklem, kayıp veri sorunu, veri toplama aracı ve veri analiz yöntemi başlıklarında incelenmiştir. Çalışmanın devamında incelenen başlıklara ilişkin sorunlar betimlenmiştir. Bu süreçte araştırmacıların yaşadığı en önemli problem incelenen bazı makalelerin raporlaştırmasındaki yetersizlik ve eksikliklerdir. Bu süreçteki belirlenen başlıklarla ilgili bilgilerin bulunmasını güçleştirmiş ve makalelerin detaylı okunması, analiz edilmesi ve içerikte bu bilgilerin tasnif edilmesini gerekli kılmıştır. Bakker ve Wicherts (2011) yılında yaptıkları çalışmada benzer durumu ifade etmişlerdir.

Araştırmada ilk olarak incelenen makalelerin türüne bakılmıştır. Ancak dokuz çalışmada bu konuda herhangi bir açıklama yapılmamıştır. Yordayıcı istatistikleri içeren çalışmalar araştırmanın temel amacını oluşturduğu için tüm çalışmalar nicel ya da karma araştırma türündedir. Bununla ilgili Erdem (2011) tarafından yapılan bir çalışmada Türkiye’de eğitim alanındaki çalışmaların çoğunun (%82) nicel araştırma yöntemleri kullanılarak yapıldığını göstermektedir. Bu durumdaki artışın birçok sebebi olabileceği gibi en önemlilerinden bir tanesi gelişen bilgi teknolojilerine bağlı olarak kullanımı yaygınlaşan istatistik paket programlarıdır. Bu programlar sayesinde araştırmacılar kullandıkları istatistiksel tekniklerin kavramsal ve teorik temellerini bilmeden bireysel olarak analizlerini yapabilmektedirler. Fakat bu durum bazen araçların amaçsallaşmasına neden olacak durumlar ortaya çıkarabilir (Arı, Armutlu, Tosunoğlu ve Toy; 2009). Nicel

çalışmaların yaygınlaşmasının bir diğer nedeni de pozitivist-ampirik çalışmalarda nesnelliğin tartışılmayacağı (Reis, 2009) ve bir bilimin ancak matematiği kullandığı ölçüde ilerleme kaydedebileceği düşüncesidir (Lafargue, 2014, s.105). Eğitim bilimlerinde ise karma yöntemlerin tercih edilmesinin çoklu yöntem ve veri toplama araçlarındaki çeşitliliği artıracığı için geçerliliği ve güvenilirliği olumlu yönde etkileyeceği ortadadır (Sözbilir ve Kutu, 2008).

Araştırmada ikinci olarak incelenen örneklem grubuna dair ifadelerdir. Araştırmanın iç geçerliliği ve dış geçerliliği açısından büyük öneme sahip örneklem seçiminde ise çoğu araştırmada yeteri kadar bilgi verilmediği görülmektedir. Araştırmalarda örneklem, çalışma grubu ifadelerinin birbiri yerine sıklıkla kullanıldığı görülmüştür. Bu durumda araştırmanın genellenebilirliğini şüpheli hale getirmektedir. Bir çalışmada seçilen örneklem büyüklüğünün ise %5 belirlilik düzeyi için beklenen sayı olan 30'un altında olduğu görülmüştür (Altunışık, Coşkun, Bayraktaroğlu ve Yıldırım, 2012).

Çalışmada incelenen bir diğer önemli husus ise kayıp veri sorunudur. İncelenen makalelerden sadece bir tanesinde kayıp veri olduğundan bahsedilmiş ancak söz konusu makalede kayıp veri ile başa çıkma yöntemlerinden hangisinin kullanıldığı ifade edilmemiştir. İncelenen makalelerde veri setinde kayıp verinin olup olmadığı analiz çıktıları ve bulgular bölümleri incelenerek karar verilmiştir. Yordayıcı istatistiklerin yapıldığı çalışmalarda kayıp veri sorununa ilişkin herhangi bir düzeltme yapılmaksızın analizlerin gerçekleştirilmesi sonuçların yanlış ya da hatalı olmasına sebep olabilir (Kang, 2013; Frisell, 2016). Benzer sonuçlar Demir ve Parlak (2012) tarafından yapılan çalışmada da ortaya konulmuş ve Türkiye’de eğitim alanında yapılan çalışmalarda ve sonuçların raporlanmasında önemli eksikliklerin olduğu ifade edilmiştir.

Araştırma sonuçlarından veri toplama araçlarına ilişkin elde edilen en önemli bulgu en çok ölçek ve anketlerin kullanıldığıdır. Benzer sonuçlar farklı çalışmalarda da ortaya konulmuştur (Altunışık ve diğerleri, 2012). Veri toplama araçları incelendiğinde en çok karşılaşılan sorun anket ve ölçek kavramlarının birbirinin yerine kullanılmasıdır. Her iki ölçme aracı birbirinden farklı psikometrik özelliklere sahiptir. Bununla birlikte bu veri toplama araçları kullanılarak yapılabilecek analizlerde birbirinden farklı özellikler göstermektedir. Çalışmalarda bu konuda yöntem ve bulgular bölümlerinde ifadelerin birbiri yerine kullanıldıkları görülmüştür. Çalışmaların birçoğunda kullanılan veri toplama araçlarının psikometrik özelliklerinden bahsedilmemiştir. Bu durum toplanan verilerden elde edilen bilgilerle ilgili güvenilirlik sorunu oluşturabilir.

Araştırmada son olarak verilerin hangi tekniklerle analiz ettikleri incelenmiştir. Çalışmaların büyük bir bölümünde parametrik analizlerin kullanıldığı görülmüştür. En çok kullanılan parametrik analizler ise t testleri olmuştur. Bu durum Hsu (2005) çalışmasındaki sonuçlarla benzerlik göstermektedir. Çalışmaların yirmi dört tanesinde parametrik testler kullanılmasına rağmen bu testlerin kullanılabilmesi için kontrol edilmesi gereken varsayımların sağlandığına dair ya sadece sözel olarak ifade kullanılmış ya da hiç bahsedilmemiştir. Bu durum Demir, Saatçioğlu ve İmrol (2016) tarafından yapılan çalışmada ve alanyazındaki pek çok farklı çalışmada da ortaya konulmuştur

(Bakker ve Wicherts, 2011; Özsoy ve Özsoy, 2013). Söz konusu durumun temel sebeplerinden bir tanesinin çalışmalarda da ifade edildiği gibi analizlerin sadece paket program kullanılarak yapılmış olması dolayısıyla araştırmacıların bu konudaki istatistik bilgilerinin yetersizliği olduğu düşünülmektedir. Özellikle istatistiksel analizler için kullanılan paket programların isimlerinin yazılması analizlerin sorunsuz olarak gerçekleştirildiğinin kanıtı değildir. Bununla birlikte karşılaştırma çalışmalarında elde edilen anlamlı farkların hiçbirinde etki büyüklüğü değeri raporlanmamıştır.

Sonuç olarak bu çalışmada nicel çalışmalarda karşılaşılan birçok sorun ele alınarak çözüm önerileri sunulmaya çalışılmıştır. Araştırmada toplanan veriler İngilizce eğitimi alanındaki yayınlar ve Ulakbim veri tabanı ile sınırlandırıldığından gelecek çalışmalarda incelenen makale sayısı artırılabilir. Benzer alanda yurt dışında yapılan çalışmalarla yurt içinde yapılan çalışmalar karşılaştırılabilir. Özellikle araştırmacıların istatistiksel analizler noktasında teorik olarak bir temel kazanmaları sonrasında paket programlar kullanılarak analiz yapmaları ve elde edilen sonuçların nasıl raporlanacağı noktasında danışman desteği almaları, bununla birlikte dergilerde ölçme ve değerlendirme alanında uzman bir danışmanın bulunmasının faydalı olacağı düşünülmektedir.

Kaynaklar

- Allison, P. D. (2009). *Missing data*. R. E. Millsap, A. Maydeu-Olivares (Yay. haz.), Quantitative methods in psychology içinde (ss. 72-89). London: SAGE Publication.
- Altunışık, R., Coşkun, R., Bayraktaroğlu, S. ve Yıldırım, E. (2012). *Sosyal bilimlerde araştırma yöntemleri* (7. baskı). Sakarya, Sakarya Yayıncılık.
- Arı, G., Armutlu, C., Tosunoğlu, N. ve Toy, B. (2009). Nicel araştırmalarda metodoloji sorunları: yüksek lisans tezleri üzerine bir araştırma. *Ankara Üniversitesi SBF Dergisi*, 64(4), 15-37.
- Bakker, M. ve Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior research methods*, 43(3), 666-678.
- Balcı, A. (2013). *Sosyal bilimlerde araştırma: Yöntem, teknik ve ilkeler*. Ankara, Pegem A.
- Blischke, W. R., Karim, M. R. ve Murthy, D. P. (2011) *Preliminary data analysis. In: Warranty Data Collection and Analysis*. Springer Series in Reliability Engineering. Springer, London.
- Bryman, A. (2001). *Social research methods* (1. baskı). New York: Oxford University Press.
- Connolly, P. (2007). *Qualitative data analysis in education: A critical introduction using spss*. London, Routledge.
- Cook, T. ve Wong, V. (2008). *Better quasi-experimental practice*. P. Alasuutari, L. Bickman, J. Brannen (Yay. haz.), *The Sage handbook of social research methods* içinde (ss. 134-164). London, Sage.

- Cramer, D. ve Howitt, D. L. (2004). *The Sage dictionary of statistics: a practical resource for students in the social sciences*. Sage.
- Cronbach, L. J., Schönemann, P. ve McKie, D. (1965). Alpha coefficients for Stratified-Parallel Tests. *Educational and Psychological Measurement*, 25(2), 291-312.
- Curley, C., Krause, R. M., Feiock, R. ve Hawkins, C. V. (2017). Dealing with missing data: A comparative exploration of approaches using the integrated city sustainability database. *Urban Affairs Review*, 55(2), 591-615.
- Demir, E. ve Parlak, B. (2012). Türkiye’de eğitim araştırmalarında kayıp veri sorunu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 3(1), 230-241.
- Demir, E., Saatçioğlu, Ö. ve İmrol, F. (2016). Uluslararası dergilerde yayımlanan eğitim araştırmalarının normallik varsayımları açısından incelenmesi. *Current Research in Education*, 2(3), 130-148.
- Ercan, İ. ve Kan, İ. (2004). Ölçeklerde güvenirlik ve geçerlik. *Uludağ Üniversitesi Tıp Fakültesi Dergisi*, 30(3), 211-216.
- Erceg-Hurn, D. M., Wilcox, R. R. ve Keselman, H. J. (2013). *Robust statistical estimation*. In *The Oxford Handbook of Quantitative Methods in Psychology* (ss. 1-34).
- Erdem, D. (2011). Türkiye’de 2005-2006 yılları arasında yayımlanan eğitim bilimleri dergilerindeki makalelerin bazı özellikler açısından incelenmesi: Betimsel bir analiz. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 2(1), 140-147.
- Erkuş, A. (2007). Ölçek geliştirme ve uyarlama çalışmalarında karşılaşılan sorunlar. *Türk Psikoloji Bülteni*, 13(40), 17-25.
- Erkuş, A. (2010). Psikometrik terimlerin Türkçe karşılıklarının anlamları ile yapılan işlemlerin uyumsuzluğu. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(2), 72-77.
- Ethnologue. (2019, Eylül 24). *What is the most spoken language?* <https://www.ethnologue.com/guides/most-spoken-languages> adresinden erişilmiştir.
- European Commission. (2019). *Education and training. Education and Training - European Commission*. https://ec.europa.eu/education/policies/multilingualism/evidence-based-policy_en adresinden erişilmiştir.
- Fidell, L. S. ve Tabachnick, B. G. (2003). Preparatory data analysis. *Handbook of psychology: Research methods in psychology*, 2, 115-141.
- Frisell, T. (2016). Why missing data is a problem, and what you shouldn’t do to solve it. *Annals of the Rheumatic Diseases*, 75(2), 44-45.
- García-Laencina, P. J., Morales-Sánchez, J., Verdú-Monedero, R., Larrey-Ruiz, J., Sancho-Gómez, J. ve Figueiras-Vidal, A. R. (2009). *Classification with incomplete data*. T. Ganchev, M. Sokolova, R. Rada, P. J. García-Laencina, V. Ravi (Yay. haz.), Handbook of

- Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques: Algorithms, Methods, and Techniques içinde (s. 147-175). IGI Global.
- Gorard, S. (2001). *Quantitative methods in educational research: The role of numbers made easy*. London, The Tower Building.
- Hatch, E. M. ve Lazaraton, A. (1991). *The research manual: Design and statistics for applied linguistics*. New York, NY, Newbury House Publishers.
- Hsu, T. (2005). Research methods and data analysis procedures used by educational researchers. *International Journal of Research and Method in Education*, 28(2), 109-133.
- İlhan, M. (2016). Araştırmacıların nicel veri analizinde karşılaştıkları güçlüklerin ikili karşılaştırmalar yoluyla incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 7(1), 73-84.
- Iwaniec, J. (2020). *Questionnaires, implications for effective implementation*. J. McKinley, H. Rose (Yay. haz.), The Routledge Handbook of Research Methods in Applied Linguistics içinde. New York, Routledge.
- Johnson, R. B. ve Christensen, L. (2012). *Educational research: Quantitative, qualitative, and mixed approaches (4. baskı)*. SAGE Publications.
- Karasar, N. (2012). *Bilimsel araştırma yöntemi*. Ankara, Nobel Yayıncılık.
- Kirk, R. (2009). *Experimental design*. R.E. Millsap, A. Maydeu-Olivares (Yay. haz.), The Sage handbook of quantitative methods in psychology içinde (s. 23-45). London, Sage.
- Kraemer, H. (2014). Effect size. *The Encyclopedia of Clinical Psychology*, 1-3.
- Lafargue, P. (2014). *Burjuvazi Öldürür ve Toplu Denemeler*. Zeplin Yayınları.
- Larson-Hall, J. (2016). *A guide to doing statistics in second language research using SPSS and R (2nd edition)*. Routledge.
- Loewen, S. ve Plonsky, L. (2015). *An A-Z of applied linguistics research methods*. Macmillan International Higher Education.
- Mahu, D. P. (2012). Why learning is English so beneficial nowadays? *International Journal of Communication Research*, 2(4), 374-376.
- Mann, P. S. (2016). *Introductory statistics (9. baskı)*. John Wiley and Sons, Inc.
- Marascuilo, L. A. ve McSweeney, M. (1977). *Nonparametric and distribution-free methods for the social sciences*. Monterey, CA Brooks/Cole.
- Marshall, E. ve Boggis, E. (2016). The statistics tutor's quick guide to commonly used statistical tests. *Statstutor Community Project*, 1-57.
- Miles, M. B. ve Huberman, A. M. (1994). *Qualitative Data Analysis*. Thousand.
- Nachmias, C., Nachmias, D. (1996). *Research methods in the social sciences (5. baskı)*. New York, St. Martin's Press.
- Norman, G. R. ve Streiner, D. L. (2003). *PDQ statistics*. PMPH-USA.

- Özsoy, S. ve Özsoy, G. (2013). Eğitim araştırmalarında etki büyüklüğü raporlanması. *İlköğretim Online*, 12(2), 334-346.
- Peacock, J. L. ve Peacock, P. J. (2011). *Oxford handbook of medical statistics (1. baskı)*. Oxford University Press
- Plonsky, L. (2017). *Quantitative research methods*. S. Loewen, M. Sato (Yay. haz.), The Routledge Handbook of Instructed Second Language Acquisition. New York: Routledge.
- Reddy, M. S. (2016). Importance of English language in today's world. *International Journal of Academic Research*, 3(4), 179-184.
- Reis, H. (2009). *Quantitative methods in relationship research*. H. T. Reis, S. Sprecher (Yay. haz.), Encyclopedia of human relationships içinde (s. 1319-1321). SAGE Publications, Inc.
- Rogers, J. ve Révész, A. (2020). *Experimental and quasi-experimental designs*. J. McKinley, H. Rose (Yay. haz.), The Routledge Handbook of Research Methods in Applied Linguistics içinde. New York, Routledge.
- Sahu, P. K., Pal, S. R. ve Das, A. K. (2015). *Estimation and inferential statistics*. Springer.
- Salgado, C., Azevedo, C., Proença, H. G. ve Vieira, S. (2016). Missing data. *Secondary Analysis of Electronic Health Records* (ss. 143-162). Springer.
- Schuele, C. M. ve Justice, L. M. (2006). The importance of effect sizes in the interpretation of research. *The ASHA Leader*, 11(10), 14-27.
- Sheskin, D. J. (2000). *Handbook of parametric and Nonparametric statistical procedures (2nd edition)*. Chapman and Hall/CRC.
- Sözbilir, M. ve Kutu, H. (2008). Development and current status of science education research in Turkey. *Essays in Education*, 24(1), 1-22.
- Stout, D. E. ve Ruble, T. L. (1995). Assessing the practical significance of empirical results in accounting education research: The use of effect size information. *Journal of Accounting Education*, 13(3), 281-298.
- Sullivan, G. ve Feinn, R. (2012). Using effect size—or why the p value is not enough. *Journal of Graduate Medical Education*, 4(3), 279-282.
- Taherdoost, H. (2016). Sampling methods in research methodology; how to choose a sampling technique for research. *International Journal of Academic Research in Management*, 5(2), 18- 27.
- Tavşancıl, E. ve Aslan E. (2001). *İçerik analizi ve uygulama örnekler [Content Analysis and Application Examples]*. Ankara, Epsilon Yayınları.
- Turan, İ., Şimşek, Ü. ve Aslan, H. (2015). Eğitim araştırmalarında likert ölçeği ve likert-tipi soruların kullanımı ve analizi. *Sakarya Üniversitesi Eğitim Fakültesi Dergisi*, 30(1), 186-203.

- Wasserstein, R. L., Schirm, A. L. ve Lazar, N. A. (2019). Moving to a world beyond “ $p < 0.05$ ”. *The American Statistician*, 73(1), 1-19.
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American psychologist*, 54(8), 594.
- Wisniewski, S. R., Leon, A. C., Otto, M. W. ve Trivedi, M. H. (2006). Prevention of missing data in clinical research studies. *Biological Psychiatry*, 59(11), 997-1000.

EK-1

Veri Analizi Formu

Sayı	Yayın Yılı	Dergi Adı	Makale Adı	Yöntem	Eleştiri	Çalışılan Grup İfadesi (Evren- Örnekler/Çalışma Grubu)	Çalışılan Grup İfadesi Uygunluğu (Dış Geçerlik Sorunu)	Çalışma Grubu ve Katılımcı Sayısı	Kayıp Veri	Çalışmada Kullanılan Veri Toplama Aracı	Eleştiri	Veri Analizi	Veri Analizini Eleştirisini

“İngiliz Dili Eğitimi Alanında Ulusal Dergilerde Yayımlanan Nicel Çalışmaların Yöntem Açısından İncelenmesi” başlıklı çalışmanın yazım sürecinde bilimsel, etik ve alıntı kurallarına uyulmuş; toplanan veriler üzerinde herhangi bir tahrifat yapılmamış, karşılaşılabilecek tüm etik ihlallerde “Sakarya University Journal of Education Dergisi ve Editörünün” hiçbir sorumluluğunun olmadığı, tüm sorumluluğun yazarlara ait olduğu ve çalışmanın herhangi başka bir akademik yayın ortamına değerlendirme için gönderilmemiş olduğu bu çalışmanın yazarları tarafından taahhüt edilmiştir.

Araştırmacıların Makaleye Katkı Oranı Beyanı:

1. yazar katkı oranı: %50

2. yazar katkı oranı: %40

3. yazar katkı oranı: %10

Çıkar Çatışması Beyanı:

Yok.

Finansal Destek veya Teşekkür Beyanı

Bu çalışma için herhangi bir kurumdan finansal destek alınmamıştır.