JRM JOURNAL OF RESEARCH METHODOLOGY

Journal of Research Methodology
Vol. 37 No. 3 (September–December 2024)
ISSN 0857-2933 ISSN online 2697-4835
Academic Article https://doi.org/10.14456/jrm.2024.13

# Determining the Appropriate Sample Size in EFL Pilot Studies

Patsawut Sukserm[1]

Received: September 17, 2024 – Revised: December 11, 2024 – Accepted: December 15, 2024

**Abstract**

Pilot studies are fundamental in the field of English as a foreign language (EFL) as they provide opportunities to evaluate and improve the research design and instruments. This review summarizes key researchers' suggestions for determining sample size in EFL pilot studies and offers instructions for conducting pilot studies in this context. For example, a range of 10 to 30 participants is suitable for pilot studies, while 30 participants are generally sufficient for a questionnaire. Additionally, about a dozen subjects in a group are sufficient when estimating distribution parameters. Guidelines for reliability are also formulated. This review therefore enhances EFL research by providing a practical understanding of sample size recommendations and their implications in EFL settings, thereby assisting researchers in designing effective EFL pilot studies. This will increase the credibility and relevance of EFL research and contribute to the development of more comprehensive and trustworthy studies in the future.

**Keywords:** Pilot Studies, Survey Research, EFL, Sample Size, Reliability

[1] *Corresponding Author,*
Chulalongkorn University Language Institute, Chulalongkorn University, Bangkok, 10330, Thailand
patsawut.s@chula.ac.th

JRM JOURNAL OF RESEARCH METHODOLOGY
วารสารวิธีวิทยาการวิจัย
ปีที่ ๓๗ ฉบับที่ ๓ (กันยายน–ธันวาคม ๒๕๖๗)
ISSN 0857-2933 ISSN online 2697-4835
บทความวิชาการ
https://doi.org/10.14456/jrm.2024.13

# การกำหนดขนาดตัวอย่างที่เหมาะสมในการศึกษานำร่องทางภาษาอังกฤษเป็นภาษาต่างประเทศ

พรรษวุฒิ สุขเสริม[1]

รับต้นฉบับ : 17 กันยายน 2567 – รับแก้ไข : 11 ธันวาคม 2567 – ตอบรับตีพิมพ์ : 15 ธันวาคม 2567

**บทคัดย่อ**

การศึกษานำร่องมีความสำคัญในด้านภาษาอังกฤษเป็นภาษาต่างประเทศเพราะจะช่วยให้สามารถประเมินและปรับปรุงการออกแบบการวิจัยและเครื่องมือต่าง ๆ ได้ บทความปริทัศน์นี้ได้สรุปข้อเสนอแนะสำคัญเกี่ยวกับการกำหนดขนาดตัวอย่างในการศึกษานำร่องในงานวิจัยด้านภาษาอังกฤษเป็นภาษาต่างประเทศและแนะนำขั้นตอนการทำการศึกษานำร่องในบริบทการวิจัยด้านนี้ ยกตัวอย่างเช่น ตัวอย่างวิจัยขนาด10 ถึง 30 คนเหมาะสมกับการศึกษานำร่อง ในขณะที่ตัวอย่างวิจัยจำนวน 30 คนเพียงพอสำหรับการตอบแบบสอบถาม นอกจากนี้ตัวอย่างวิจัยประมาณ 12 คน ก็เพียงพอแล้วเมื่อประมาณค่าพารามิเตอร์การกระจาย นอกจากนี้ยังนำเสนอแนวทางสำหรับค่าความเที่ยงไว้ด้วย บทความปริทัศน์นี้จึงช่วยเพิ่มความเข้าใจในเชิงปฏิบัติเกี่ยวกับข้อเสนอแนะเรื่องขนาดตัวอย่างและผลที่ตามมาในบริบทการวิจัยด้านภาษาอังกฤษเป็นภาษาต่างประเทศซึ่งจะช่วยนักวิจัยให้การออกแบบการศึกษานำร่องด้านภาษาอังกฤษเป็นภาษาต่างประเทศมีประสิทธิภาพมากขึ้น โดยแนวทางนี้จะเพิ่มความน่าเชื่อถือและความสอดคล้องของการวิจัยด้านภาษาอังกฤษเป็นภาษาต่างประเทศ และยังช่วยส่งเสริมการพัฒนาการวิจัยที่ครอบคลุมและน่าเชื่อถือมากขึ้นในอนาคต

**คำสำคัญ :** การศึกษานำร่อง, การวิจัยเชิงสำรวจ, ภาษาอังกฤษเป็นภาษาต่างประเทศ, ขนาดตัวอย่าง, ค่าความเที่ยง

[1] *ผู้รับผิดชอบบทความหลัก,*
สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย กรุงเทพมหานคร 10330
patsawut.s@chula.ac.th

## Introduction

The study of EFL is of great relevance in understanding language learning and language teaching around the world. Language as a subject seeks to achieve the goal of teaching a particular language and calls for the provision of adequate competence, development of effective language tests and other research relevant to second language learning. It is also very important for EFL studies to specify aspects which are regarded as pilot investigations aimed at testing the proposed research methodology and tools prior to the actual main investigations (In, 2017; Thabane et al., 2010).

In the same way, linguistic studies that underpin EFL approaches are concerned with the description of the language in its structure, usage and learning in various contexts around the world. The helpers are scholars who are specialized in the application of linguistics geared toward language and the EFL settings (Brown, 2014). Practical or hands on approaches to the acquisition of knowledge in this area can be found in empirical research, which entails observing behavioral phenomena that relate to defined contexts such as EFL teaching and that can later be validated by questionnaires, experimental designs or corpus analysis. When carrying out such work in EFL settings, pre-testing the research methodology is very useful because it helps to refine the instruments to make them acceptable and reliable in the specific—cultural and contextual—settings.

Thus, pilot studies in EFL research serve the purpose of anticipating challenges that may be encountered in the main study, refining data collection instruments, and, most critically, validating the methods used for the research. In a more detailed perspective, Van Teijlingen and Hundley (2001) state that the challenges of recruitment of study participants, methods of data collection and actual implementation of the study are tested in a pilot study. Therefore, pilot studies are important in deciding what changes need to be undertaken before considerable investments are made towards expansive research.

One issue that has been pointed out as very important in EFL research is the selection of appropriately sized samples in pilot studies. The sample of the pilot study does not need to be very big, but nor should it be too small. Some suggest that a sample of 10-30 subjects is often enough to bring out important aspects and perfect the study design as well as sufficient to test it (Johanson & Brooks, 2010). Ten to fifteen subjects are the numbers recommended by Bujang et al. (2024) for questionnaire-based studies in order to provide a suitable number of answers to experimentally test each question in a study.

In addition, it is necessary to understand the factors leading to variability in the responses with regard to the particular design of the study. Hertzog (2008) contends that without sufficient variability in a pilot study, a researcher will not be able to determine the sample size required for the main study and ensure that it achieves an adequate level of statistical power. This is particularly the case for EFL settings as the cultural and linguistic backgrounds of both the students and the raters may influence the use of language assessments and their interpretation. The focus of EFL research in assessing instruments has to be on the relevant methodological issues, especially the constructs being measured and the instruments' reliability and validity. Validity is concerned with the degree to which an instrument measures what it is intended to measure (Bachman & Palmer, 1996). In language assessment, for instance, it is the construct validity which seeks to ensure that all the test items that are included refer to the construct which is of interest, e.g. reading comprehension or grammar (Bachman & Palmer, 1996). Equally important is content validity, which encompasses the relevance and sufficiency of the test in relation to the ability being tested (Brown, 2014). Reliability, on the other hand, refers to the degree of consistency of the measurement instrument. Internal consistency is commonly measured using Cronbach's alpha. Values of 0.70 or more are generally considered acceptable for new measurements (Nunnally, 1978).

Pilot studies are an essential and critical phase that helps in the evaluation of the above components. They allow researchers to check the clarity of the questions, the extent to which the items are aligned with the objectives, and whether the instrument is able to produce similar results in different EFL contexts (Van Teijlingen & Hundley, 2001). For example, questionnaire-based instruments need to be examined in sufficient detail to ensure that the items are applicable and unambiguous in various cultural settings (DeVellis, 2012). When these conditions are met, researchers can improve their instruments in terms of reliable and valid measurements so that large-scale studies can be conducted (Teresi, 2022). The aim of this paper is to suggest sample size considerations in the context of conducting EFL pilot studies. This also helps to highlight a possible way forward for researchers by addressing the issues that arise in all EFL contexts, such as language and cultural differences, making it possible to improve the quality of EFL research designs in terms of reliability and usefulness, so that such research is both fruitful and practical.

## Definition and Purpose of EFL Pilot Studies

In the context of EFL research, pilot studies can be understood as short, small-scale and preliminary research conducted before any larger and more detailed studies are undertaken. They help to focus on the research methodology and enable the investigators to solve any likely difficulties which may arise before embarking on a large-scale project (Kunselman, 2024; Thabane et al., 2010). There are many other reasons why pilot studies are important in EFL research, e.g. the testing of research instruments, techniques for data collection, etc. The primary objective or purpose of pilot studies is to enhance the dependability and the authenticity of the designed research tool. This involves the testing of survey instruments, interview schedules and any other instrument focusing on the population of concern in terms of precision, usefulness and relevance (Leon et al., 2011). For instance, in EFL areas, conducting such studies enables the researchers to avoid questions which are unclear or worded such that the real ability and attitude of the respondents are reflected inappropriately (Giner-Sorolla et al., 2024).

Pilot studies also allow for the evaluation of data collection activities. This therefore helps the researchers refine the methodology—for example, by identifying recruitment strategies and essential criteria such as the time needed to undertake the tests and the conditions whereby the data should be collected (Van Teijlingen & Hundley, 2001). Noticing such logistical constraints ahead of time enables researchers to adjust their methodology so as to benefit the data collection process. Furthermore, pilot studies help in evaluating the overall reasonableness of the research plan. They shed light as to whether or not the proposed study is possible, with the resources and limitations at hand. They also help to determine a suitable time frame for recruiting participants and carrying out the study in question.

In summary, EFL pilot studies are crucial about revision of research instruments, practicing data collection and assessing the feasibility of the study. They assist in enhancing the reliability and the validity of the research findings as they help to identify problems in the very beginning so that the follow-up studies are adequately designed and resourced.

## Importance of Sample Size in EFL Pilot Studies

Especially for this EFL research, determining an adequate sample size in pilot studies is critical for obtaining reliable and valid EFL research results. The appropriate sample size could guarantee that the pilot stage can validate the research design in its entirety in testing any proposed issues (Andrade, 2020; Johanson & Brooks, 2010). If the sample size is not enough, the pilot study may not reveal even critical problems

and then the conclusions are misleading, and the research instruments developed are ineffective.

One of the most important reasons for estimating an appropriate sample size is to ensure the statistical potential of the pilot study (Andrade, 2020). Statistically, the power of a study refers to its ability to detect an effect if it exists. Suppose a pilot study has too few participants. Many relevant details would be ignored, e.g. the design of the survey questions or the associated education or training, which would jeopardize the internal validity of the entire study (Kraemer et al., 2006). In EFL research, for example, concerns about participants' language proficiency and cultural background justify consideration of an appropriate sample size that would cover the range of some of these factors. Hertzog (2008) further asserts that assessing variability within the data can also help in estimating the sample size required for the main study and guarantees that there is adequate power to detect important effects.

In addition, piloting is also very helpful in determining the logistical and practical aspects of conducting the research. With an adequate sample size, researchers are able to assess the practicality of recruitment, data collection and intervention techniques (Leon et al., 2011). This can ensure that the final version of the case study is as accurate as a money machine and ensure that appropriate changes are made before the large study begins. Julious (2005) suggests a simple guide for pilot studies to help them outline their aim; he advises a target group of around 12 participants for studies trying to find out the mean, standard deviation and other factors. This recommendation also balances the two challenges of accuracy and practicality of the pilot study.

Therefore, the selection of an appropriate sample size is crucial for the success of EFL pilot studies. Given the variability of sample sizes across studies, it is crucial to determine them correctly. Appropriate results from pilot studies can be achieved if the established criteria are adhered to but modified to reflect the diverse nature of EFL research, thus providing a basis for the success of larger studies (Azman et al., 2024).

## Understanding Scale Validity and Reliability in EFL Studies

In terms of scale validity, it is important to identify potential problems in relation to how the scale is used to measure the variables in question. This aspect of validity should be carefully considered to ensure that the results of the test are not only accurate but also interpretable. There are several types of validity that researchers should investigate.

First and foremost is the construct validity. It checks whether the scale is used to measure the intended theoretical construct to which it was originally tailored (Bachman & Palmer, 1996). This means that a valid vocabulary test in EFL, for example, should be consistent with theories of lexical knowledge. Second, the instrument should accurately capture the salient features of the skill or facet being assessed (Brown, 2014). For example, teaching a foreign language to improve a learner's speaking skills should include elements of speaking that cover fluency, accuracy and pronunciation. Third, criterion-related validity, which is the primary focus of this review, considers the correlational aspects between a standard assessment score and another criterion or standard, which may be performance-based assessments (Cohen et al., 2007). In the context of language learning, it could be the evaluation of the effectiveness of a placement or entrance test for language learners in predicting their language proficiency after studying a prescribed program over a period of time.

Pilot studies are necessary to authenticate scales in EFL environments. They enable researchers to perform item analysis and polish their instruments in such a way that construct and content validity as well as cultural and linguistic aspects are addressed (Van Teijlingen & Hundley, 2001). The turnaround of validation of scales greatly enhances the precision and reliability of the tools which are employed in EFL studies.

Reliability is the degree of consistency with which a measurement instrument performs, i.e. the ability to produce the same result when used any number of times under the same conditions (Cohen et al., 2007). In the case of EFL research instrument development, reliability means that the instruments used—tests, scales or questionnaires—provide consistent and valid results when used at different times and in different samples.

The reliability of a particular instrument can be determined in different ways. First, one type of reliability assessment, known as internal consistency, examines the extent to which the items on a scale correlate to measure the same construct. Cronbach's alpha, a measure of the internal consistency or reliability between items, ratings or measurements (Bujang et al., 2018), is the statistic most used for this purpose, and Nunnally (1978) found that values of 0.70 and above are acceptable for new instruments. The recommended reliability coefficients, which apply to more stable measures, are values above 0.80. Secondly, test-retest reliability involves measuring the same correlation on two occasions with an agreed interval between them and determining the correlation between the two values to assess the stability of the instrument over time. If the correlation is high, this means that the instrument is relatively stable over time (Brown, 2014). What's more, inter-rater reliability holds great

significance for subjective evaluations such as oral or written tests where there are several assessors of the subjects' performance. It assesses the level of consensus among the raters and is usually determined mathematically, for example, using Cohen's Kappa or Intraclass Correlation Coefficient (ICC) (McNamara, 1996) which is important. Finally, parallel forms reliability assesses the extent to which two forms of an instrument measure the same thing. According to Cohen et al. (2007), if it has been established that there is a strong correlation between them, this indicates that the instrument has a good level of parallel forms reliability.

The construction of research tools which are effective for students' research starts with the validity and reliability of the instrument (Azman et al., 2024). Therefore, it is important for the researchers to perform their assessments of validity and reliability through testing construct, content, criterion, and area validity alongside internal consistency. Furthermore, researchers should adapt and validate their instruments in the context of the language and culture of EFL learners to improve the research quality.

## Benchmarks for Cronbach's Alpha in EFL Studies

Within the realm of EFL studies it is determined that the coefficient of internal consistency evaluated by means of Cronbach's alpha is of great significance, which is similar to its use in science education (Taber, 2018). It shows the level of relationship between a collection of items and the extent to which they measure a unified construct. There are a number of well-known researchers who have proposed criteria for understanding the interpretation of Cronbach's alpha values that are useful for researchers.

Cronbach's alpha, with its focus on test item internal consistency, was first discussed by Cronbach (1951). Although he did not provide precise measures, his efforts formed the basis for later measurements. Nunnally (1978) suggested that for preliminary studies, correlation coefficient values of 0.70 should suffice but for instruments of greater calibration suggested a value above 0.80. Such values are commonplace in educational and psychological research and set standards for reliability testing. In the view of George and Mallery (2003), many researchers assess reliability with the following interpretation of Cronbach's alpha: 0.9 – Excellent; 0.8 – Good; 0.7 – Acceptable; 0.6 – Questionable; 0.5 – Poor; < 0.5 - Unacceptable. This scale is presented in Table 1. Such a scale theoretically enables the researchers to derive the reliability of their instruments with higher accuracy than usual and emphasizes the norms anticipated in their research.

**Table 1**

*Suggested Values for Cronbach's Alpha ($\alpha$)*

| Value of Alpha | Meaning | Explanation |
|---|---|---|
| $.90 \leq \alpha$ | Excellent | The data can be used for formal tests and established scales. |
| $.80 \leq \alpha < .90$ | Good | The data are reasonably acceptable for several studies and well-established scales. |
| $.70 \leq \alpha < .80$ | Acceptable | The data can be used in most cases, but scales that are already well established would be preferable. |
| $.65 \leq \alpha < .70$ | Slightly acceptable | The data should be used with latitude in exploratory cases but should not be assumed to be highly credible or accurate. |
| $.60 \leq \alpha < .65$ | Questionable | The data is more of a scale reconstruction type and the focus should be on item analysis. |
| $.50 \leq \alpha < .60$ | Poor | The data are normatively completely unacceptable. A change in scale direction is suggested. |
| $\alpha < 0.50$ | Unacceptable | The data implies a change in direction or a modification of the scale rather minimally, as the internal consistency is significantly explained. |

**Adapted from George and Mallery (2003)*

Similarly, in the context of EFL, DeVellis (2012) put forth guidelines around scale development and for the interpretation of reliability coefficients. He argued that alpha values that are lower than 0.70 are generally unacceptable, but he conceded this might not apply uniformly to all studies. For instance, low values of alpha may be considered adequate for an exploratory study on teaching foreign languages, but moderately high alpha values will be required from a test for high stakes purposes.

Consequently, the measure of internal consistency, Cronbach's alpha, is crucial in determining the reliability of EFL instruments. EFL researchers can check the quality and validity of their instruments by following certain guidelines, and as long as the reliability coefficients correspond to the goals of a specified study, such standards would not be violated.

## General Guidelines for Sample Size in EFL Pilot Studies

Numerous researchers have suggested guidelines for determining the appropriate size of the sample so as to decrease the chances of a pilot experiment being both costly and of no value. In this regard, Johanson and Brooks (2010) assert that for preliminary research, a sample of approximately 10-30 people is usually considered adequate. This range is sensible and manageable as it enables researchers to experiment only on key features of the study's design and the activities of collecting the data. This guide is particularly useful for the initial phase of feasibility testing in EFL studies, where logistical and practical considerations often make a big difference.

A more detailed consideration is provided by Bujang et al. (2024), who argue that in questionnaire-based studies it is better to take a per-item approach to sample size. In their estimation, such a subgroup analysis is cost-effective as no more than 5-10 participants per item are required. This approach helps to identify deficiencies in relation to the questions and increases the reliability of the instrument. Hertzog (2008), on the other hand, emphasizes the aim of the study in each survey as it varies from study to study along with the expected response size. On the other hand, if the pilot study is to serve as the basis for a larger study, a larger sample may generally be required to obtain reliable estimates. The size of the sample in the quantitative phase of the study depends on how much one estimates the degree of fluctuation in the results to overcome considerations of statistical power. As Julious (2005) notes, when conducting pilot studies, approximations can be used to answer questions such as the mean and standard deviation. A sample size of 12 participants per group in a pilot study is sufficient, taking cost-effectiveness into account. This number is based on the requirement to achieve sufficient accuracy in estimating the parameters while not making the hypothesis too complicated.

Regarding the aforementioned purpose, design and estimated variability of the pilot studies, the general instructions for sample size in EFL pilot studies seem vague to us. From a practical point of view, they emphasize that one should start with a range of 10-30 participants, and a more precise estimate gives 5-10 participants per item for questionnaire-based studies. By looking at the aims of the study, the researcher can filter out an acceptable and realistic sample size for their pilot study.

## Recommendations for Conducting EFL Studies

In the case of EFL studies, it is crucial to be guided by eminent experts in the field to conduct appropriate and sound research. Several well-known researchers have made suggestions that extend to the design and conduct of EFL studies, covering sample size, instrument preparation and methodology.

Johanson and Brooks (2010) consider pilot studies to be necessary in their research. For pilot studies, they recommend a sample size of 10 to 30 people, which is sufficient to broadly identify some of the most serious problems with the designs and data collection tools used in the studies. This size range is particularly useful in language learning studies where pre-planned testing of the instrument is important in preparing the instrument and the research method. In addition, Nunnally (1978) offers a set of criteria that have been extensively cited to ensure the reliability of research instruments. He mentions that the value of 0.70 of Cronbach's alpha is acceptable for analysis in the early stages of research, while the value of 0.80 is most appropriate for more mature instruments. This benchmark helps educators conduct EFL studies to determine the internal reliability of their questionnaires and tests and to ensure that the items measure the same construct.

On the other hand, Hair et al. (2018) explain the required sample size in multivariate data analysis, especially in Exploratory Factor Analysis (EFA). They state that a minimum of 5–20 participants per variable is required to support the reliability of factor identification. This recommendation should be crucial for EFL researchers designing studies that aim to explain factors that influence either language proficiency or attitudes towards language learning (Sukserm, 2024). According to Hertzog (2008), there has to be a consideration of the sample size issue when dealing with anticipated study goals and variability in the results expected. These researchers advocate for the recruitment of a greater number of subjects for pilot studies that are intended to set relevant thresholds for primary research. In spite of this, Julious (2005) suggested that 12 subjects per group should be enough. This suggestion is both practical in its implementation and exact in its requirements to ensure that pilot studies are cost-effective.

## Empirical Studies on Sample Size in Pilot Studies in EFL Settings

Empirical studies of sample sizes used in pilot studies in EFL settings give a fair degree of insight into the application of the above recommendations. Such studies emphasize the practice of respecting sample size validity recommendations but equally show the necessity of contextual tailoring. For example, Tegeh et al. (2014) initiated a pilot study of 30 subjects to prepare an EFL reading comprehension test for high school learners in Indonesia. In this case, the researchers were able to identify insufficient test items and improve their measurement instrument by adhering to Johanson and Brooks' (2010) recommendation that 10 to 30 subjects participate in a pilot study. This convergence underscores the feasibility of the range, particularly regarding the need to address design issues in smaller educational contexts.

Moreover, Tabatabaei and Loni (2015) tried to evaluate the reliability of the questionnaire by employing a sample consisting of 30 high school students who were males and females. This figure exceeds their minimum number of participants as proposed by Johanson and Brooks (2010) and also conforms to Hertzog's (2008) refrain concerning the sort of covering needed for parameter estimation. Such larger sample size could provide the researcher the opportunity to explain variability in culture, use of language and hence the range of contexts that the EFL instruments are required to be used in and hence validated and made reliable. This principle demonstrates the need for formulating operational objectives and carefully considering the sample taken. To make their pilot study more comprehensive, Vo et al. (2024) recruited 100 participants for their pilot study on EFL learners' readiness and challenges for immediate online learning. This very high number of participants helped to accurately identify the reliability of the questionnaire.

These outcomes highlight the connection between theory and practice. The perspectives of scholars are useful (Hertzog, 2008; Johanson & Brooks, 2010; Julious, 2005), though in practice application of their advice depends on the goals and context of the research. For instance, it might be necessary to substantially raise the number of participants in studies which adopt exploratory factor analysis (Hair et al., 2018) to obtain acceptable results in terms of factor identification. Yet, in small-scale or feasibility studies, the use of smaller samples may be adequate for parameter estimation. The practice and application of sample size norms as described in the studies demonstrate how such norms have been implemented within the continuum of different EFL research types. These studies also underscore the need to put together theoretical benchmarks and practical problems having to do with participant variability and research aims. Therefore, flexibility in sample sizes for EFL pilot studies has been accentuated as an important facet in EFL research.

## The Proposed Appropriate Sample Size for EFL Pilot Studies

As shown in Table 2, researchers have put forth various suggestions on estimating the sample size for pilot studies. According to Johanson and Brooks (2010), the optimal sample size should comprise around 10–30 participants, which can be useful in identifying the key shortcomings of both the methodology and the instruments. They also argue that using 5 to 10 subjects for each item of the questionnaire would be a reasonable decision. Hertzog (2008) emphasizes that one should understand the objectives of the study and the expected variance in responses, and she suggested using larger samples for parameter estimation. Julious (2005) recommends approximately 12 participants per group for parameter estimation.

**Table 2**

*Proposed Sample Size Recommendation*

| Study Type | Sample Size Recommendation |
|---|---|
| General Pilot Testing | 10–30 participants |
| Questionnaire-Based Studies | 5–10 participants per item |
| Parameter Estimation | ~12 participants per group |
| Exploratory Factor Analysis (EFA) | Minimum of 20 participants per variable |
| Variability Considerations | Larger sample sizes needed for diverse contexts and populations |

**Adapted from Bujang et al. (2024), Ghani et al., 2012; Hair et al. (2018), Hertzog (2008), Johanson and Brooks (2010), and Bujang et al. (2024)*

Based on the literature reviewed, the suggested appropriate sample size for EFL pilot studies takes into account both practical limitations and reliable data requirements. Furthermore, pilot studies need to take into account participants' language proficiency and culture, which can have an important impact on the results. This is necessary because a sufficiently large sample allows the pilot to test the applicability of the research design and other possible sources of problems. It could be assumed that there should be an optimal sample size that combines practicality and feasibility by summarizing the findings of the key researchers as follows:

1. General pilot testing: In pilot studies where the focus is on testing the design and procedures, a sample size of 10 to 30 participants is considered appropriate. This is in line with Johanson and Brooks (2010) and allows enough data to be collected to highlight critical issues without being overly burdensome.

2. Questionnaire-based studies: For research that requires the formulation and validation of questionnaires (Bujang et al., 2024; Ghani et al., 2012), it is suggested that the sample size should be between five and ten participants for each item. This helps in analyzing problems by ensuring that each question is tested, which increases the overall reliability of the instrument by identifying problem questions.

3. Estimation of parameters: For a study whose aim is to obtain some data, e.g. parameters such as mean or standard deviations to enable the design of the main study, it has been suggested that each group consists of approximately 12 subjects. This recommendation by Julious (2005) helps to solve the conflict between the estimation accuracy and the practicality of conducting the pilot study to obtain useful estimates without surveying the audience to an unrealistically large extent.

4. Exploratory Factor Analysis (EFA): Pilot studies using EFA require a minimum number of 5–20 participants for each variable, especially latent variables (Sukserm,

2024), as reported by Hair et al. (2018). This allows a valid and meaningful factor solution to be estimated and explained in a consistent manner.

The guidelines in the EFL context for estimating the sample size of pilot studies are suitable for different research purposes. That is, the choice of guidelines depends on the purpose and limitations of the study. The views of Bujang et al. (2024) are more suitable for instrument development. For limited-funded studies, the approach of Julious (2005) is useful, while for basic studies with multivariate analysis, it may be fruitful to follow Hair et al. (2018). Knowing these differences allows researchers to choose the most appropriate guidelines for their research objectives and the resources available to them.

These recommendations could also assist researchers in conducting relevant mini-studies that would be helpful for the primary exploratory research because they suggest appropriate limits for estimating the mini-study sample size, particularly for EFL research purposes. With the relevant objectives in mind, they could also determine the types of tools to be used, and the required level of estimation accuracy needed. These proposed recommendations form a sufficiently flexible but also stringent framework on how to decide the minimum sample size for pilot studies.

The decision on EFL sample size for pilot studies follows several steps with a view to achieving balance between methodological soundness and practicality. The first step is always to *"define the problem"* to be studied—exploratory, confirmatory or descriptive—as this will inform further decisions. Researchers must then "*determine the objectives*" of the study, e.g. the feasibility of testing certain research designs or the parameters required for the pilot studies.

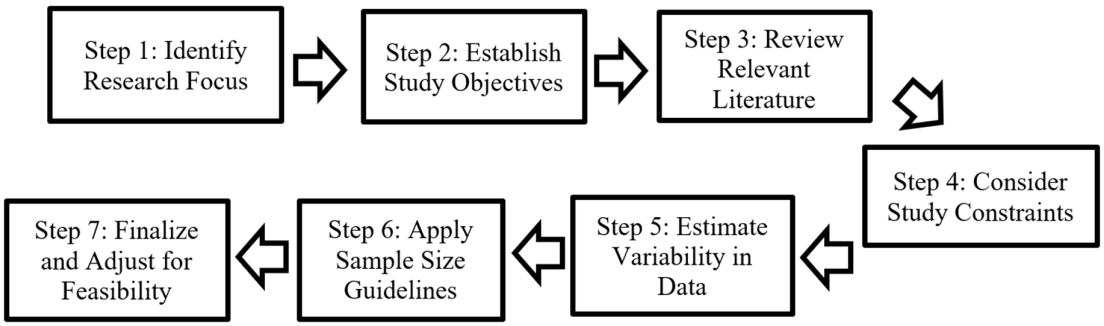
Once the focus and objectives of the study have been determined, a *"review of the relevant literature*" follows. These steps require the review of established guidelines and empirical evidence as presented in several studies for sample size determination. Then, researchers should *"consider barriers*" such as time, resources and field exposure in order to make the proposed sample size reasonably attainable.

The next step is to *"determine the degree of variation in the data"* with respect to how participants in different EFL contexts respond to the task in question. Noting this variability is essential when determining the statistical sample that will enable meaningful conclusions. Subject to these considerations, *"sample size determination guidelines"* with regards to the aims and context of the study can be applied. Finally, researchers must *"complete and amend the size of the sample"* for the sake of practicality, making theoretical expectations and the actual situation, such as, recruitment of study participants or funding, congruent. This process, illustrated in Figure 1 below, provides a simple procedure for determining the appropriate sample

size in EFL pilot studies. By adhering to these key principles, studies are strengthened and can better fit within the context of the study.

**Figure 1**

*Proposed Steps of Determining Sample Size in EFL Pilot Studies*




The proposed procedure also provides an adaptive and applicable approach which meets the needs of most EFL pilot studies. This section offers practical recommendations for researchers by consolidating the aspects of generalized testing, questionnaire surveys, estimation of parameters, and exploratory factor analysis guidelines. This provision allows them to increase their practical efficiency while remaining methodological which, in turn, makes it possible to create viable pilot studies that meet the requirements of EFL research.

## Practical Implications and Future Directions

When deciding to conduct pilot studies, an important first step for EFL researchers is to determine the appropriate sample size. More importantly, pilot studies should be designed and conducted in a way that provides reliable and valid results, following the summarized guidelines of key researchers. This assists researchers in avoiding and resolving possible difficulties at the beginning of the research, thus saving them more time and resources.

In EFL research, in which measurement is likely to be an important component, the higher suggested sample sizes would enhance measurement. For instance, in questionnaire-based studies (Ghani et al., 2012), it is best to aim towards having 5 to 10 respondents per item to achieve a desired targeting ratio ensuring clarity, accuracy and relevance for all items put on the questionnaire. This is especially the case in EFL contexts when cultural or linguistics barriers exist that can interfere with survey item meanings.

Future research in the field of foreign language teaching should further develop these recommendations in different settings and for different samples. Another direction for future research could be the effect of different sample sizes on the reliability and validity of different EFL test instruments.

Discussions about reliability and applicability in EFL pilot studies would benefit greatly by considering large-scale assessments such as PISA (Programme for International Student Assessment) and PIRLS (Progress in International Reading Literacy Study). These assessments assist in understanding how language proficiency and literacy skills differ due to context and population across institutes and countries. This is especially important for PISA as an example, which assists students in 15 countries to undertake reading, mathematics and science in the context of real life problem-solving and literacy (OECD, 2019). The fact that that achievement varies across linguistic and cultural contexts suggests the need to modify research tools for the population involved.

Adopting these large-scale assessments at the same level avoids any bias inherent in early-phase EFL studies in two ways. First, they underscore the importance of developing tools that are appropriate in context—that is, tools which are cross-culturally and cross-linguistically valid. Second, EFL measures can also be improved by using some of the more solid strategies outlined for the PISA and PIRLS projects, such as the significance attached to validity and reliability of language in the PISA project.

The significance of determining sample sizes is pertinent in improving the quality and systemic reliability of EFL pilot studies. Further studies should seek to validate the proposed guidelines and discover other pertinent spheres like new technologies and distance education for devising and extending EFL research method sample size strategies.

## Limitations and Challenges

Even though guidelines on the appropriate sample size constitute useful considerations when planning EFL pilot studies, their actual implementation is, at times, fraught with difficulties of a practical or contextual nature. Knowing these difficulties is important so that the research is not beyond the practical level.

The recruitment of participants for pilot studies is usually a difficult task, especially in some EFL contexts such as remote areas or certain school categories. Following Hair et al.'s (2018) guidelines, where they suggested a minimum of 5–20 subjects for each variable in exploratory factor analysis, may be unrealistic when the number of research subjects is limited. Likewise, studies that employ 24 participants

for questionnaire studies (Bujang et al., 2024) may also face difficulties in their recruitment process in programs with low enrolment levels or when the participants have specialized linguistic needs.

The subjects of EFL research are usually drawn from a variety of linguistic, cultural, and educational orientations. The importance of variability in parameter estimation is also backed up by Hertzog (2008). This variability, however, may cause disparity in people's responses and, as a result, causes difficulties in attaining the reliability criteria. Research instruments, for instance, may yield different results when used in multilingual classrooms as opposed to monocultural ones, meaning that the instrument or sample size would need to be changed.

It is common for pilots to be constrained by resources such as time, funds, or staff, which can result in failure to meet sample size suggestions. According to Julious (2005), in cases when limited funding is available for pilot trials, a more pragmatic rule of 12 subjects per group would provide a reasonable solution, although may not satisfy the requirements of more comprehensive studies.

It is said that there should be a compromise between the level of practicality and the level of methodological sophistication. The number of participants recommended by Johanson and Brooks (2010), namely 10–30, is acceptable for low level feasibility studies. However, this will not be sufficient in studies where empirical analysis of sampling distributions is required, or when a wider range of applicability is desired. It is essential that these guidelines be adapted to the aims and limitations existing in each study.

Challenges arise in implementing sample size criteria with respect to participant recruitment, variability across contexts, and scarcity of resources. These restrictions suggest the importance of being flexible when it comes to the application of theoretical ideals. Most importantly, researchers should recognize and mitigate these challenges so that the methodological integrity of their pilot studies can be maintained.

## Conclusion

Pilot studies have been found to be fundamental in testing and refining research designs, instruments and data collection methodologies in EFL research. This review has collated suggested criteria for sample size determination and adapted these guidelines to particular situations in EFL environments. This paper contrasts with more general summaries in its concentration on linguistic and cultural differences in EFL contexts and their impact on the application of these guidelines, and therefore assists EFL researchers in this respect. A distinctive feature of this review is that it focuses on the need to reconcile theoretical perspectives with some of the more challenging

aspects of EFL research, such as problems with participant recruitment and limited resources. By critically examining the application of these guidelines in empirical studies, this article promotes the integration of theory and practice and supports researchers in their main task of preparing a pilot study. As a starting point for improving the quality and application of EFL pilot studies, this review provides more specific factors and parameters for the determination of sample sizes. This focus helps strengthen the methodological base of EFL studies and increase the relevance of the research for EFL learners. These insights have implications for the methodological requirements of EFL studies which relate to cross-cultural, cross-national or even global assessment of research topics in applied linguistics.

## References

Andrade, C. (2020). Sample size and its importance in research. *Indian Journal of Psychological Medicine, 42*(1), 102–103. https://doi.org/10.4103/IJPSYM.IJPSYM_504_19

Azman, N. A., Hamzah, M. I., Abdul Razak, K., & Zulkifli, H. (2024). Digital competence among Islamic teachers: A pilot study on validity and reliability. *The Turkish Online Journal of Educational Technology, 23*(4), 51-62.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford University Press.

Brown, H. D. (2014). *Principles of language learning and teaching* (6th ed.). Longman.

Bujang, M. A., Omar, E. D., & Baharum, N. A. (2018). A review on sample size determination for Cronbach's alpha test: A simple guide for researchers. *The Malaysian Journal of Medical Sciences (MJMS), 25*(6), 85–99. https://doi.org/10.21315/mjms2018.25.6.9

Bujang, M. A., Omar, E. D., Foo, D. H. P., & Hon, Y. K. (2024). Sample size determination for conducting a pilot study to assess reliability of a questionnaire. *Restorative Dentistry & Endodontics, 49*(1), 1–8. https://doi.org/10.5395/rde.2024.49.e3

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2007). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Routledge.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

DeVellis, R. F. (2012). *Scale development: Theory and applications* (3rd ed.). Sage Publications.

Elder, C., & O'Loughlin, K. (2003). Investigating the relationship between intensive English language instruction and band score gain on IELTS. *IELTS Research Reports, 4*, 207–254.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference*. Allyn & Bacon.

Giner-Sorolla, R., Montoya, A. K., Reifman, A., Carpenter, T., Lewis, N. A., Aberson, C. L., Bostyn, D. H., Conrique, B. G., Ng, B. W., Schoemann, A. M., & Soderberg, C. (2024). Power to detect what? Considerations for planning and evaluating sample size. *Personality and Social Psychology Review*, *28*(3), 276–301. https://doi.org/10.1177/10888683241228328

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis* (8th ed.). Pearson.

Hertzog, M. A. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health, 31*(2), 180–191. https://doi.org/10.1002/nur.20247

In, J. (2017). Introduction of a pilot study. *Korean Journal of Anesthesiology, 70*(6), 601–605. https://doi.org/10.4097/kjae.2017.70.6.601

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394–400. https://doi.org/10.1177/0013164409355692

Julious, S. A. (2005). Sample size of 12 per group rule of thumb for a pilot study. *Pharmaceutical Statistics, 4*(4), 287–291. https://doi.org/10.1002/pst.185

Kraemer, H. C., Mintz, J., Noda, A., Tinklenberg, J., & Yesavage, J. A. (2006). Caution regarding the use of pilot studies to guide power calculations for study proposals. *Archives of General Psychiatry, 63*(5), 484–489. https://doi.org/10.1001/archpsyc.63.5.484

Kunselman, A. R. (2024). A brief overview of pilot studies and their sample size justification. *Fertility and Sterility, 121*(6), 899–901. https://doi.org/10.1016/j.fertnstert.2024.01.040

Leon, A. C., Davis, L. L., & Kraemer, H. C. (2011). The role and interpretation of pilot studies in clinical research. *Journal of Psychiatric Research, 45*(5), 626–629. https://doi.org/10.1016/j.jpsychires.2010.10.008

McNamara, T. F. (1996). *Measuring second language performance*. Longman.

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). McGraw-Hill.

OECD. (2019). PISA 2018 results (Volume I): What students know and can do. OECD Publishing. https://doi.org/10.1787/5f07c754-en

Sukserm, P. (2024). Understanding latent variables in EFL contexts. *Shanlax International Journal of Education*, *12*(4), 60–69. https://doi.org/10.34293/education.v12i4.7886

Tabatabaei, O., & Loni, M. (2015). Problems of teaching and learning English in Lorestan province high schools, Iran. *Mediterranean Journal of Social Sciences, 6*(2 S1), 47–55. https://doi.org/10.5901/mjss.2015.v6n2s1p47

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tegeh, I. M., Jampel, I. N., & Pudjawan, K. (2014). The effectiveness of the EFL reading comprehension test. *Journal of Educational Research and Evaluation, 3*(1), 12–22.

Teresi, J. A., Yu, X., Stewart, A. L., & Hays, R. D. (2022). Guidelines for designing and evaluating feasibility pilot studies. *Medical Care, 60*(1), 95–103. https://doi.org/10.1097/MLR.0000000000001664

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., et al. (2010). A tutorial on pilot studies: What, why and how. *BMC Medical Research Methodology, 10*(1), 1–10. https://doi.org/10.1186/1471-2288-10-1

Van Teijlingen, E., & Hundley, V. (2001). The importance of pilot studies. *Social Research Update, 35*, 1–4.

Vo, T. K. A., Long, N. V., & Vo, H. T. T. N. (2024). EFL learners' readiness and challenges for immediate online learning: A case study in Vietnam. *Journal of Institutional Research South East Asia, 22*(2), 141–158.