



Experiencing problems in data collection by undergraduate EFL students as novice researchers

Faisal Mustafa¹, Merina Devira^{2*}, Heri Apriadi¹, Bonjovi Hassan Hajan³

¹ English Education Department, Faculty of Teacher Training and Education, Universitas Syiah Kuala, Banda Aceh, Indonesia

² English Education Department, Faculty of Teacher Training and Education, Universitas Samudra, Langsa, Indonesia

³ Senior High School Division, Faculty of Research, José Rizal University, Metro Manila, Philippines

*Email: merinadevira.fkip@unsam.ac.id (corresponding author)

Article history: Received 30 May 2022; Revised 23 July 2022; Accepted 31 July 2022;
Available online 7 August 2022; Published regularly September 2022

Abstract

Data collection is a research process determining the validity and reliability of a study, and novice researchers have been found to experience many problems in this process. However, little study addressed the problem in data collection among these researchers. Therefore, this study was conducted to uncover problems experienced by English as a Foreign Language (EFL) undergraduate students without any previous experience in data collection. Seventeen undergraduate EFL students were assigned to collect data using a two-item questionnaire, which involved communicating with research participants before sending the questionnaire to the participants. After the data collection, they were requested to write a brief report explaining their communication with the participants and the problems they experienced. The reports written by the students were treated as the data for this research and were subjected to qualitative analysis to understand the problems they faced in the data collection. The results show that students experienced some problems in collecting the data, i.e., lack of responses, late responses, a negative response, and uncompleted questionnaires. The results of the study suggest that potential problems in data collection should be introduced to undergraduate EFL students before they conduct their undergraduate research.

Keywords: EFL undergraduate students; data collection; novice researchers; research process

To cite this article: Mustafa, F., Devira, M., Apriadi, H., & Hajan, B. H. (2022). Experiencing problems in data collection by undergraduate EFL students as novice researchers. *Journal on English as a Foreign Language*, 12(2), 383-404. <https://doi.org/10.23971/jefl.v12i2.4111>

To link to this article: <https://doi.org/10.23971/jefl.v12i2.4111>



Copyright © 2022 THE AUTHOR(S). This article is distributed under a *Creative Commons Attribution-ShareAlike 4.0 International* license.

Introduction

Dissertation or sometimes referred to as thesis, which involves data collection, is a requirement for a university graduation in many English as a Foreign Language (EFL) undergraduate programs in many countries including Indonesia (Indrian & Ardi, 2019), Malaysia (Okafor et al., 2018), the United Kingdom (Hannigan & Burnard, 2001), and Australia (Roberts & Seaman, 2018). Dissertation process facilitates students in learning and helps them learn how to conduct research (Laosum et al., 2016). The significance of dissertation for EFL undergraduate students is also emphasized by frequent publications of undergraduate dissertation in many reputable international journals. The urgency to publish among lecturers teaching in English language teaching to pre-service EFL teachers in those countries has motivated them to write papers using the data collected by the students they supervise for their undergraduate dissertation as reflected in Rawat and Meena (2014). In most cases, the content of the dissertation is also adapted for publication. Although the finding is interesting and seems to contribute to the knowledge, the validity of the results might be questionable because the data are collected by inexperienced, novice researchers (Cutcliffe & McKenna, 2004; Johnson & Masten, 1998). To avoid such biased results, undergraduate dissertation supervisors need to have detailed information on the process of data collection by their students in order to provide better supervision. However, the process of data collection by novice researchers remains under-researched.

Many research studies in this area focus on doctoral students such as Barry and Ployé (2013), Carter-Veale et al. (2016), Khosa et al. (2020), and Shahsavar and Kourepaz (2020), to name a few. One study which focused on undergraduate students found that the problems in data collection process include research location, respondents' participations, cultural issue, and unorganized data (I'Anson & Smith, 2004). Lack of participations have been

identified by many researchers among experienced researchers. Therefore, the purpose of the present study is to find out the problems related to a lack of respondents' participations experienced by undergraduate EFL students as novice researchers. The scope of this study was limited to EFL students, and thus the results of this study are significant for lecturers teaching research method in English language teaching (ELT) for undergraduate students and for undergraduate dissertation supervisors in the same discipline. EFL undergraduate students are usually encouraged to publish their research findings in an international journal, which makes it even more necessary for the data to be accurate and the sample size to be adequate. The lecturers can address the potential problems, identified by this study, during the research method courses and provide detailed measures to avoid or solve the problems if they arise. Undergraduate supervisors and students can predict potential problems presented by the present study and discuss which solutions are suitable for their research contexts. Without identifying potential problems prior to the field research, it might be too late when they emerge.

Data collection is a systematic method of obtaining information related to research variables to answer questions outlined in the research (Kabir, 2016). It is the most crucial part of a study because it is the data collection process which determines the validity and reliability of research results (Olsen, 2012). Although approaches in data collection might differ by discipline (Carr et al., 2019), they share the same focus i.e. on maintaining a specific and accurate selection of the data regardless of data type. Due to the significance of this research process, data collection should be carefully planned by researchers who are experts in research method (Mayr & Weller, 2016; Olsen, 2012; Pawar, 2007), although this planning requires knowledge that novice researchers might not have (I'Anson & Smith, 2004). In addition, expertise in the corresponding discipline is a requirement for data collection planning (Carr et al., 2019). Experts have categorized methods of data collection (Hammersley & Gomm, 2009; Marshall et al., 2009; Pawar, 2007). However, the categorization is not unique (Plowright, 2011), but it depends on, among others, types of data, research design, and research objective. In addition, the approach in data collection keeps evolving, and internet data collection has now gained its popularity (McIntosh, 2008).

The nature of a research study determines the procedure of data collection; however, all research with the same methodology shares general guideline for the data collection. The critical and systematic procedures in the process of collecting data may vary depending on the research contexts and the possible constraints surrounding a study (Cohen et al., 2007). Prior to data collection, a

researcher needs to prepare research instrument, obtain the ethics approval, select research sample, and obtain the study participation approval (Kristjansson et al., 2013). They added that ethics approval is usually proposed to an ethics committee in institutions or government, while study participation approval is requested to potential participants or their institutions. Thus, the identification of types of data, the research instrument design and pilot testing, and the sampling technique are also important to be outlined in order to help the researchers gather accurate data (Onwuegbuzie et al., 2010). During the data collection, the instruments, tests or questionnaires are distributed to the participants after explaining the procedure participation, the participations' rights, and possible misinterpretation of research instrument (Schober et al., 2018). For interview, the first task of researchers is to create a comfortable atmosphere starting from introducing themselves. They also need to be familiar with the questions and the question orders (Ary et al., 2010). For observation, the researcher must prepare an observation sheet to complete and paper to take a comprehensive note that enables researchers to recreate the observation (Best & Kahn, 2006).

The challenges in the process of collecting data sometimes can be unprecedented by the researchers (Couper, 2017), especially when the two or more types of big data are required to answer a research question (Small, 2011). Previous studies have revealed the challenges that emerge during the process of collecting data which can be attributed to both primary factors: the selected research designs and societies (Glasgow & Emmons, 2007; I'Anson & Smith, 2004). For example, in a multi-discipline studies conducted by experienced researchers, the challenges of using the survey are substantial be it in the form of paper, telephone, or online-based responses (Benson et al., 2017; Islam et al., 2010). The problems in their data collection processes can be categorized into several cases: (1) a steady decline of the response rates, (2) difficulty in analyzing big data, (3) challenges in establishing representativeness of sample to general population, and (4) failure to find evidence. King (2009) also experienced an issue related to sample, resulting in a number of biases due to societal factors in the demographic groups of the study participants, such as failure to identify their socio-economic and education status. Furthermore, the ethics approval or gaining initial access to collect the data are also highlighted as a challenge in the multi-discipline research studies, particularly for a sensitive research topic (King, 2009). In addition to access problem and sensitive research topic, undergraduate students as novice researchers identified other challenges during data collection processes, including: (1) the location of research conduct, (2) respondents' participations, (3) cultural issue, and (4) no

important theme resulted during the data collection (I'Anson & Smith, 2004). Furthermore, recording qualitative data is a part of data collection process which presents specific challenge for researchers. The earliest form of interview record is field note, but notes cannot be replayed, creating a chance of losing detail information (Ashmore & Reed, 2000). Tape or digital recorded version can be replayed, but it needs to be transcribed. However, this version needs to be transcribed and it takes an excessive amount of time for this process (Tessier, 2012). Therefore, technology has been develop to replace human transcribers with automatic transcription software (Tessier, 2012).

Experience has been shown to help in the process of data collection. With experience in conducting research, researchers will develop knowledge and skills required to successfully collect data, predict potential problems, and solve them should they arise (Thiry et al., 2012). In addition, research has shown that experienced researchers are most likely to obtain the evidence through their investigation (Cohen et al., 2007). Even though research skills of an experienced researcher take time and efforts to naturally develop, novice researchers like undergraduate students can improve their research skill through good mentorship (Thiry et al., 2012) and practical guidance (Moser & Korstjens, 2018). Thus, Lown (1993) suggests the benefits of undergraduate students involved in their professor research projects. With a multi-year research experience, undergraduate students can be identified as experienced researchers who not only understand about the data collection techniques but comprehend more advanced scientific skills to plan experimental designs and methods (Thiry et al., 2012). In addition, an experience building on the research will enable the researchers to develop the theory within the data, to think critically for a problem solving on the process of research, and eventually to develop additional data when they are needed as the strategies in the data collection process (Wheeldon & Faubert, 2009).

Based on the review of previous studies, there are some skills required to collect the data for research in order that the data can accurately represent the real phenomena (Thiry et al., 2012). Previous studies have also in general identified problems experienced by novice researchers in data collection in research process and the causes of the problems societies (Glasgow & Emmons, 2007; I'Anson & Smith, 2004). One of such problems is related to lack of respondents' participation. However, the problems of lack of participations have not been addressed in detail in previous studies. Therefore, the current study is aimed at finding out the problems experienced by novice researchers who did not have any previous research experience in collecting data.

Method

Research design

This research utilized a qualitative approach where the data were collected through a questionnaire and analyzed it in detail, and the presentation of results was in the form of list and explanation. Qualitative method was selected because this type of study can be used to obtain an in-depth information of a phenomenon (Lune & Berg, 2017), which represents the objective of the present study. Since this study was motivated by an assumption that students experience a problem of lack of respondents' participations when collecting research data, the design of this qualitative research was a case study, as defined by Lune and Berg (2017). The source data for this research were reports prepared by undergraduate students on their experience in collecting data for an assignment in the research methodology course at the department of English education of a well-established state university in Aceh, Indonesia.

Participants

The participants of this study were undergraduate students majoring English language teaching, and thus they were pre-service teachers, studying at English Education Department, Faculty of Teacher Training and Education at one of the universities in Aceh Province, Indonesia. The number of students involved in this research was 17, consisting of nine male and eight female students between 22 and 26 years of age. The participants involved in this research are detailed in Table 1.

The students were in the sixth and eight semesters, enrolled in Research Methods in English Language Teaching, which is a compulsory course in any English pre-service teacher program at an undergraduate level. All students gave their consent to use their report as the source of data for this research.

Instruments and data collection

In this research, students as the research participants were treated as the research assistants. The instrument used for this research was an instruction to collect data and write a report consisting of problems experienced during the process. The participants were assigned to collect data regarding types of cheating in an English online exam using a questionnaire which they delivered using a google form, consisting of two questions. One is a closed-ended question (see Appendix) asking respondents to choose one or more types of cheating provided on the list, and the other requires them to write other forms of cheating that they knew their classmates did, or they themselves did on the

exam. This questionnaire is not the instrument of the present study, but it was the instrument used by the participants to collect the data so that they could write their report. This type of questionnaire was selected to provide students, who were the participants or research assistants in this research, with adequate challenges in data collection. Each participant was assigned to send the questionnaire to four respondents. Before sending the questionnaire, the participants communicated with their potential respondents to explain the purpose of the questionnaire and to ask them to complete the questionnaire. The participants were also requested to ask the respondents to confirm whether they had completed the questionnaire, and this also served as a reminder for the respondents who had ignored the questionnaire. The participants were given three days to collect the data, and they prepared a report describing what they did, the problem they experienced, and how they approached the problems.

Table 1. The demographics information of the research participants

No	Initials	Code	Gender	Age	Semester	GPA
1	AA	S1	M	23	4 th	3.01
2	AM	S2	F	22	4 th	3.34
3	AN	S3	M	22	4 th	3.08
4	BCI	S4	M	22	4 th	3.38
5	EM	S5	F	22	4 th	3.37
6	FR	S6	F	23	4 th	3.26
7	IF	S7	M	26	6 th	NA
8	LMN	S8	F	22	4 th	3.53
9	NA	S9	F	23	4 th	3.51
10	RH	S10	M	23	4 th	3.32
11	SSF	S11	F	23	4 th	3.35
12	TFA	S12	M	22	4 th	3.46
13	VNH	S13	M	23	4 th	3.44
14	W	S14	M	23	4 th	3.38
15	WP	S15	M	23	4 th	3.19
16	WW	S16	F	22	4 th	3.41
17	WIS	S17	F	23	4 th	3.28

Trustworthiness

The students, as the research participants, were enrolled in Research Methods course taught by one of the researchers, which could potentially suffer from biases. However, the biases were expected and those the instruction was given in a way which eliminated the effect of researchers. When assigning students to collect the data from respondents, the participants were treated as research

assistants, and thus the number of responses that the students got was not considered in grading the students' work. They were graded based on how complete their report was. Therefore, the data obtained from this research process are considered similar to the data collected in a context where the researcher were not the course instructors. Therefore, the study captured the real phenomena intended in the research purpose, which is a way to establish trustworthiness in a qualitative research study (Kaminski & Pitney, 2004).

Data analysis

After the reports were submitted to the researchers, they were analyzed qualitatively, the process of which consists of three steps, i.e., data reduction, data display, conclusion, and verification (Lune & Berg, 2017). In the first step, the problems that the students faced were extracted from the students' reports using simple coding technique consisting of five steps as suggested by Creswell (2012), i.e. reading text, segmenting text, labeling text, reducing replicated code, and categorizing the codes. An ID was also assigned to each student report in the form of S1 to S17. In the second step, the researchers categorized problems and displayed the results in a table. In this step, codes, i.e., the problems that students experienced based on their reports, were categorized into themes before tabulating them into tables or figure which facilitates the drawing of conclusion. Finally, the researchers drew conclusions, i.e., creating the list of problems faced by students in data collection.

Findings

All students, who were treated as the sample of this research, were assigned to collect data using an online questionnaire, and they completed their task. The number of respondents from whom the data should be collected were initially 68 respondents. However, some respondents could not be contacted because their phone numbers were no longer registered, thus some students requested additional respondents. As a result, the number of respondents successfully contacted by the students was 67, and 48 of them completed the questionnaire. As expected, almost all students experienced problems during the data collection. This section describes all the problems found during this data collection, grouped into six problems, as listed in Table 2.

Table 2. Problems in data collection

No	Problems	Description	No. of participants
1	-	Respondents responded to the communication and completed the questionnaire	46 (68.7%)
2	Problem 1	Respondents responded to the communication but refused to complete the questionnaire	1 (1.5%)
3	Problem 2	Respondents responded to the communication, and agreed to complete the questionnaire, but did not complete it.	1 (1.5%)
4	Problem 3	Respondents responded to the communication but did not complete the questionnaire	6 (9.0%)
5	Problem 4	Respondents did not respond to the communication but completed the questionnaire	2 (3.0%)
6	Problem 5	Respondents did not respond to the communication nor complete the questionnaire	11 (16.4%)
7	Problem 6	Respondents responded to the communication very late.	8 (11.9%)

Table 2 shows that the return rate of the questionnaire was satisfactorily high, and only 20 out of 68 respondents who refused or did not complete the questionnaire although some of them responded to the communication initiated by the participants, where problem 2 and problem 3 can be combined. Table 2 shows that there are five primary problems that the participants experienced in collecting data. The detail for each problem above is described in the following sub-sections.

Ignoring communication and questionnaire

The most notable problem during the data collection was that the students as the data collector were ignored, meaning that their message was not replied. Some students contacted the potential respondents on the phone before sending them a text message through WhatsApp, which is a default messaging application used in Indonesia. Most students sent them a WhatsApp message without calling them because, according to the students' report, it is impolite to call them without previous communication. Table 1 shows that 20.6% of the respondents did not pick up the phone, nor return the call, or reply to the messages. The students reported that many of the respondents who ignored the initial communication read the first WhatsApp message sent by the students but did not respond to it even after the students sent further messages. The extract from one of the students' reports illustrated this problem.

S3: I sent a WhatsApp message to Respondent 4, greeted him, introduced self, and explained why I texted him. I asked him if he was willing to complete the survey and sent him the survey link. He did read my text but did not respond. Then, I texted him again twice, but he did not read or respond my texts.

In the extract, Student 3 (initial here) initiated several communication efforts with the respondent, but the respondent did not response to any messages sent by the students. Therefore, the participant did not know whether this respondent completed the questionnaire.

Responding to initiated communication but ignoring questionnaire

The second most common problems experienced by the students as the novice researchers in distributing their questionnaire was that the potential respondents did respond to the communication they initiated, but they did not complete the questionnaire although the students had reminded them multiple times. Students reported that many of them replied the first WhatsApp message that the students sent. Some of the respondents asked many questions before ignoring further messages. One of the respondents even agreed to complete the questionnaire, but he did not do so. The following are the extracts from two of the students' reports.

S12: I texted Respondent 2 to introduce myself and explain my intention. She responded to my message and asked what the survey was about. I explained more about the survey, but she did not respond my message, so I immediately sent her the survey link and gave her the instruction, but she did not read my message. Later, I asked whether there was any problem with the survey, but she still did not respond. Finally, I reminded her to inform me after she finished completing the survey, but she still did not respond.

S17: I got a lot of problems with Respondent 4. The first problem was that it was very difficult to contact him. The second problem was that he delayed responding my message. The third problem was that when he did pick up his phone, he asked a lot of questions. I politely answered his questions in detail. Finally, he understood and agreed to complete the survey (but he did not complete it).

Students 12 and 17 reported, as shown in the excerpts above, that their respondents responded their message and asked about the survey. However, after the students sent the questionnaire, the respondents did not reply to the message, neither did they complete the questionnaire.

Ignoring initiated communication but completing questionnaire

A limited number of respondents ignored the communication initiated by the students but ended up completing the survey without telling or confirming to the students. However, students listed it as one of the problems in data collection because, in this anonymous online survey, they could not conclude whether the respondents would complete the survey. Students reported that the respondent did read the message because the indicator on WhatsApp shows double green ticks, but the respondent did not respond. The result of the survey shows that the respondent had completed the questionnaire. Student 3 wrote the following in his report.

S3: I sent a WhatsApp message to Respondent 3 to introduce myself and explain why I texted him. He did read my text, but he did not respond. Then, I sent him the survey link and I asked him if he was willing to complete it. Finally, I asked the respondent if he had already completed the survey, but I did not receive any response.

Based on the excerpt, Student 3 wrote that although no responses were given for any texts sent, the respondent completed the questionnaire. Yet, the respondent did not confirm that he had completed the questionnaire when the student asked them.

Responding but refusing to complete the questionnaire

Out of 19 respondents who did not complete the questionnaire, only one of them frankly stated that he refused to complete the questionnaire. The students reported that the potential respondent did not reply to his message until he sent the second message. The extract of the student's report is as follows.

S12: The respondent only read the message but did not respond. So, I sent a second message along with the survey link using nice and polite language asking whether or not the respondent was willing to complete the survey. The respondent replied the message saying that he was too busy to complete the survey. Thus, I thanked and apologized for disturbing him.

The case presented in the excerpt above shows that, although very rare, the respondent explicitly stated that he could not complete the questionnaire. However, the respondent did not immediately refuse, but he did it after the student sent the questionnaire.

Participants responded the communication very late

Although the delayed response from the potential respondents does not affect their participation, seven students informed that the potential respondents replied their WhatsApp text quite late using such keywords as *a (little) bit late, quite slow responding, waited for hours, and a long time*. One of them even considered a three-hour delayed response as a long wait. The following extracts show how the students were concerned about not receiving an immediate response from their potential respondents.

S2: She was quite slow in responding my message. I had been waiting for her response for about 3 hours. I sent her the message at 04:12 pm, and she replied it at 07:26 pm.

S5: Respondents delayed responding my message for hours. I texted the respondent again, and he finally responded, and he completed the survey afterwards.

Student 2 in the excerpt considered that delayed response of 3 hours as unacceptable. Student 5 did not provide the exact length of time but he or she expected that the respondent replied the message faster.

Discussion

The objective of this research was to explore the problems experienced by EFL undergraduate students as the novice researchers in collecting research data using an online questionnaire. The results show that the EFL students experienced five major problems which include, among others, ignorance in communication, lack of participation and delayed response.

For the first problem, students informed that the participants ignored them. They neither responded to the communication initiated by the students nor participated in completing the questionnaire. Students usually sent the potential respondents' messages multiple times until they gave up. This problem is expected in an online survey, where Vehovar and Manfreda (2008) reported that the response rate is generally low when a questionnaire is delivered online. On another occasion, the students wrote that many potential respondents replied their message, especially the first message, but ignored further messages and did not complete the questionnaire. Students reported that they reminded them to complete the questionnaire, but the effect was low. This finding opposes the results of a study conducted by Toepoel (2017), which found that a reminder is effective to improve the response rate in an online

survey. Because of such common problem experienced by researchers, experiments have been conducted to seek effective ways of delivering an online survey to improve participations (Göriz, 2006; Singer et al., 2000). The results varied, but incentive is promising in improving participation. However, for student research, which is generally self-funded, providing incentive is unfeasible. In connection, Gaiser and Schreiner (2009) suggested that not all respondents need to be rewarded. The researchers can randomly selected candidates to receive the reward, and the number of selected respondents can be adjusted to the funding that the students can afford.

Another problem that the students reported is that their respondents ignored the message sent to initiate the communication, but these respondents completed the questionnaire. Students considered this lack of response a problem because they could not conclude whether the respondents needed to be re-reminded. This lack of response might be due to the fact that the respondents did not read the initial message, but they read the second message where the students included the survey information. The respondents did not consider replying messages as more important than completing the questionnaire. Therefore, as the novice researchers, students need to be informed that lack of response to messages should be expected, and thus researchers usually allow participants a week or more to complete an online questionnaire (Gu, 2018; Tran et al., 2021). Therefore, a plan should be devised to check whether the respondents have completed the questionnaire regardless of their response to messages. With the use of technology, it is possible to create unique link to each respondent and check whether they have clicked the link and complete the survey, as implemented by Mei and Brown (2018). Therefore, the respondents' reply to text messages can be treated as unnecessary.

This study also found that the students reported one respondent refusing to complete the questionnaire because he claimed that he was too busy to participate in the research. Considering the small number of respondents who explicitly refused to participate, it can be concluded that, although respondents do not have any intention to participate in the survey, explicit refusal of a survey completion is uncommon. Therefore, students should inform the potential respondents that refusing to complete the survey or withdrawing at any time after starting to answer the survey is allowable. This voluntary nature of participation in a survey for research is in line with the ethical consideration which should be included into an informed consent (American Psychological Association, 2002), as also applies in many disciplines such as health (Weissinger & Ulrich, 2019). In terms of the quality of the data, respondents who complete the questionnaire out of coercion might pay less attention to

questions or their answers, hence endangering data validity. Thus, Göritz (2006) warns about the threat of invalid data when incentive is used to increase a survey response.

Finally, students informed that the potential respondents did not respond to their message immediately. Some respondents delayed response for about 3 hours, and the others one day. Based on the comparison between delayed text response and survey response, all respondents about whom the students complained for responding to their texts late did complete the questionnaire. Therefore, students as novice researchers need to be informed that it is common for the potential respondents to not respond to the initial communication immediately. In fact, the data on survey response rate presented by Toepoel (2017) show that some responses were still given on the second and third day, and if a reminder were sent on the fourth day, the response would have surged on the fifth day. In the current study, the students reported that they sent the reminder on the same day of the initial communication. This immediate reminder might play minimal effect because the students can still expect many responses on the second day. The reminder should be sent after the effect of initial communication dies out. Friedlander et al. (2020) allowed the potential respondents 24 hours before the first reminder was sent, and the second reminder was sent 72 hours afterwards. The similar timeline can be introduced to students prior to data collection.

The results of this study have provided a significant implication for data collection in the field of education, in addition to ELT. Urgency to publish research studies in an international journal among EFL undergraduate students emphasizes that the data collection needs to be comprehensively planned. One of the research weaknesses which are often neglected among undergraduate students is small sample size. While internet survey was found to result in low response rate (Toepoel, 2017; Vehovar & Manfreda, 2008), the response rate in the current study was rather high, i.e., 72%. Online survey was usually delivered through an email sent to multiple potential participants, but in the current study, the students who delivered the questionnaire personally texted the respondents and asked for their willingness to participate in completing the questionnaire before sending the questionnaire link. This initiated communication allows researchers to adjust communication to be more personal and friendlier rather than academic and formal. Potential respondents can reply to the communication briefly and receive response from researchers immediately. This type of communication builds relationship between researchers and respondents, ensuring a better response rate. The response rate is also predicted to be higher when this type of communication is accompanied

by a reward to some randomly selected respondents than when the message is sent in the form of mass communication (Gaiser & Schreiner, 2009). The same effectiveness of text message as a media to be sent online survey was also reported by Friedlander et al. (2020). Therefore, the results of this study can be an additional section in research method on ELT books.

Conclusion

The objective of this study was to find out the problems experienced by undergraduate EFL students in data collection for their dissertation. This study has revealed the problems experienced by EFL undergraduate students as the novice researchers in collecting data using a questionnaire delivered online after an initial communication with the potential participants. The results of the study showed that there were some problems experienced during the data collection. The most frequently reported problem was that the students found their respondents ignoring their invitation and their questionnaire was not completed. They got responses from some of the respondents when they invited them, but their further messages such as reminders were ignored. In addition, the students considered it a problem when their invitation was ignored by some respondents although the questionnaire was completed. Another problem was that one respondent explicitly refused to complete the questionnaire. Finally, students also reported that many respondents did not reply to their initial message immediately. Some of them replied to it after three hours, and some after 24 hours.

Based on the results, students should be introduced to the process of data collection prior to collecting data for their undergraduate dissertation. They should also be familiarized with the problems found in this research because these are potential problems that the students are likely to experience during the data collection. Finally, with knowledge of potential problems, they can anticipate the problems or devise solutions long before the problems surface. Thus, they will be able to effectively collect valid and reliable data within reasonable length of research time.

This research has revealed potential problems faced by novice researcher in data collection using an online survey. However, the generalizability of the results of this study is limited due to several factors. First, the survey was to find out the types of cheating that the participants' classmates used in a previously taken online English proficiency test. This type of information requested from the participants was considered sensitive. The results might be

different if the information requested by the questionnaire was more public. In addition, the questionnaire used in this research was very brief, consisting of one closed-ended question and one open-ended question, and it was developed by the researchers. Therefore, the results cannot be used to generalize surveys consisting of several pages. In the case of the current study, it was less likely that the participants withdrew their participation after starting the survey. Meanwhile, withdrawing participation is more possible in a long survey. Therefore, further research needs to consider using a standard research instrument such as student academic self-efficacy scale delivered to larger sample size to be able to analyze the data after grouping it to several categories.

Acknowledgements

The authors would like to express our gratitude to the research participants and students who had participated during the data process.

Disclosure statement

No potential conflict of interest was reported by the authors.

ORCID

Faisal Mustafa <https://orcid.org/0000-0001-8619-5117>

Merina Devira <https://orcid.org/0000-0003-1533-142X>

Heri Apriadi <https://orcid.org/0000-0002-7538-2328>

Bonjovi Hassan Hajan <https://orcid.org/0000-0003-2911-5824>

References

- American Psychological Association. (2002). Ethical principles of psychologists and code of conduct. *The American Psychologist*, 57(12), 1060–1073. <https://doi.org/10.1037/0003-066X.57.12.1060>
- Ary, D., Jacobs, L. C., Sorenson, C., & Razavieh, A. (2010). *Introduction to research in education* (8th ed.). Wadsworth, Cengage Learning. <https://doi.org/10.1038/modpathol.2014.11>

- Ashmore, M., & Reed, D. (2000). Innocence and nostalgia in conversation analysis: the dynamic relations of tape and transcript. *Forum: Qualitative Social Research, 1*(3), 1–19. <http://nbn-resolving.de/urn:nbn:de:0114-fqs000335>
- Barry, V., & Ployé, A. (2013). Obstacles and challenges in writing a doctoral dissertation training: tracks for teachers. *Recherche et Formation, 73*(2), 57–72. <https://doi.org/10.4000/rechercheformation.2092>
- Benson, A. P., Bowen, T. S., Ferguson, C., Murgatroyd, S. R., & Rossiter, H. B. (2017). Data collection, handling, and fitting strategies to optimize accuracy and precision of oxygen uptake kinetics estimation from breath-by-breath measurements. *Journal of Applied Physiology, 123*(1), 227–242. <https://doi.org/10.1152/jappphysiol.00988.2016>
- Best, J. W., & Kahn, J. V. (2006). *Research in education* (10th ed.). Pearson Education Inc.
- Carr, E. M., Zhang, G. D., Ming, J. (Hung) Y., & Siddiqui, Z. S. (2019). Qualitative research: an overview of emerging approaches for data collection. *Australasian Psychiatry, 27*(3), 307–309. <https://doi.org/10.1177/1039856219828164>
- Carter-Veale, W. Y., Tull, R. G., Rutledge, J. C., & Joseph, L. N. (2016). The dissertation house model: doctoral student experiences coping and writing in a shared knowledge community. *CBE Life Sciences Education, 15*(3), 1–12. <https://doi.org/10.1187/cbe.16-01-0081>
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th ed.). Routledge Taylor & Francis Group. <https://doi.org/10.4324/9780203029053>
- Couper, M. P. (2017). New developments in survey data collection. *Annual Review of Sociology, 43*, 121–145. <https://doi.org/10.1146/annurev-soc-060116-053613>
- Creswell, J. W. (2012). *Educational research: planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson Education, Inc.
- Cutcliffe, J. R., & McKenna, H. P. (2004). Expert qualitative researchers and the use of audit trails. *Journal of Advanced Nursing, 45*(2), 126–133. <https://doi.org/10.1046/j.1365-2648.2003.02874.x>
- Friedlander, E. K. B., Soon, R., Salcedo, J., Tschann, M., Fontanilla, T., & Kaneshiro, B. (2020). Text message link to online survey: a new highly effective method of longitudinal data collection. *Contraception, 101*(4), 244–248. <https://doi.org/10.1016/j.contraception.2019.11.008>
- Gaiser, T. J., & Schreiner, A. E. (2009). Survey research on the internet. In T. J. Gaiser & A. E. Schreiner (Eds.), *A guide to conducting online research* (pp. 68–

- 81). SAGE Publications Ltd. <https://doi.org/10.4135/9780857029003.d7>
- Glasgow, R. E., & Emmons, K. M. (2007). How can we increase translation of research into practice? types of evidence needed. *Annual Review of Public Health*, 28, 413–433. <https://doi.org/10.1146/annurev.publhealth.28.021406.144145>
- Göritz, A. S. (2006). Incentives in web studies: methodological issues and a review. *International Journal of Internet Science*, 1(1), 58–70. http://www.ijis.net/ijis1_1/ijis1_1_goeritz.pdf
- Gu, P. Y. (2018). Validation of an online questionnaire of vocabulary learning strategies for ESL learners. *Studies in Second Language Learning and Teaching*, 8(2), 325–350. <https://doi.org/10.14746/ssllt.2018.8.2.7>
- Hammersley, M., & Gomm, R. (2009). Introduction. In R. Gomm, M. Hammersley, & P. Foster (Eds.), *Case study method* (pp. 1–16). SAGE Publications Ltd. <https://doi.org/10.4135/9780857024367>
- Hannigan, B., & Burnard, P. (2001). Preparing and writing an undergraduate dissertation. *Nurse Education in Practice*, 1(4), 175–180. <https://doi.org/10.1054/NEPR.2001.0028>
- l’Anson, R. A., & Smith, K. A. (2004). Undergraduate research projects and dissertations: issues of topic selection, access and data collection amongst tourism management students. *The Journal of Hospitality Leisure Sport and Tourism*, 3(1), 19–32. <https://doi.org/10.3794/johlste.31.54>
- Indrian, R. D., & Ardi, P. (2019). Rhetorical structures of English-major undergraduate thesis introduction chapters. *Indonesian Journal of EFL and Linguistics*, 4(2), 195–214.
- Islam, N. S., Khan, S., Kwon, S., Jang, D., Ro, M., & Trinh-Shevrin, C. (2010). Methodological issues in the collection, analysis, and reporting of granular data in Asian American populations: historical challenges and potential solutions. *Journal of Health Care for the Poor and Underserved*, 21(4), 1354–1381. <https://doi.org/10.1353/hpu.2010.0939>
- Johnson, B., & Masten, D. (1998). Understand what others don’t. *Design Management Journal (Former Series)*, 9(4), 17–22. <https://doi.org/10.1111/j.1948-7169.1998.tb00224.x>
- Kabir, S. M. S. (2016). *Basic guideline for research: an introductory approach for all disciplines*. Book Zone Publication.
- Kaminski, T. W., & Pitney, W. A. (2004). Strategies for establishing trustworthiness in qualitative research. *Athletic Therapy Today*, 9(1), 26–28. <https://doi.org/10.1123/att.9.1.26>
- Khosa, A., Burch, S., Ozdil, E., & Wilkin, C. (2020). Current issues in PhD supervision of accounting and finance students: evidence from Australia

- and New Zealand. *The British Accounting Review*, 52(5), 100874. <https://doi.org/10.1016/j.bar.2019.100874>
- King, E. (2009). From data problems to data points: challenges and opportunities of research in postgenocide Rwanda. *African Studies Review*, 52(3), 127–148. <https://doi.org/10.1353/arw.0.0295>
- Kristjansson, A. L., Sigfusson, J., Sigfusdottir, I. D., & Allegrante, J. P. (2013). Data collection procedures for school-based surveys among adolescents: the youth in Europe study. *Journal of School Health*, 83(9), 662–667. <https://doi.org/10.1111/josh.12079>
- Laosum, T., Kanjanawasee, S., & Pitayanon, T. (2016). Development of a dissertation quality value-added model for humanities and social sciences programs for private higher education institutions in Thailand. *Kasetsart Journal of Social Sciences*, 37(3), 138–143. <https://doi.org/10.1016/j.kjss.2016.08.010>
- Lown, J. M. (1993). Involving undergraduate students in faculty research. *Advancing the Consumer Interest*, 5(2), 29–31. <http://www.jstor.org/stable/23862837>
- Lune, H., & Berg, B. L. (2017). *Qualitative research methods for the social sciences* (9th ed.). Pearson Education Limited.
- Marshall, B., Heinzen, T., & Roberts, K. (2009). Methods of data collection. In F. J. Fowler (Ed.), *Survey research methods* (4th ed.) (pp. 68–85). SAGE Publications, Inc. <https://doi.org/10.4135/9781452230184.n5>
- Mayr, P., & Weller, K. (2016). Think before you collect: setting up a data collection approach for social media studies. In P. Mayr & K. Weller (Eds.), *The SAGE handbook of social media research methods* (pp. 107–124). SAGE Publications Ltd. <https://doi.org/10.4135/9781473983847.n8>
- McIntosh, L. (2008). Internet data collection. In S. Boslaugh (Ed.), *Encyclopedia of epidemiology* (pp. 555–557). SAGE Publications, Inc. <https://doi.org/10.4135/9781412953948.n237>
- Mei, B., & Brown, G. T. L. (2018). Conducting online surveys in China. *Social Science Computer Review*, 36(6), 721–734. <https://doi.org/10.1177/0894439317729340>
- Moser, A., & Korstjens, I. (2018). Series: practical guidance to qualitative research. Part 3: Sampling, data collection and analysis. *European Journal of General Practice*, 24(1), 9–18. <https://doi.org/10.1080/13814788.2017.1375091>
- Okafor, L. E., Khalid, U., & Then, T. (2018). Common unofficial language, development and international tourism. *Tourism Management*, 67, 127–138. <https://doi.org/10.1016/J.TOURMAN.2018.01.008>
- Olsen, W. (2012). *Data collection: Key debates and methods in social research*. SAGE

- Publications, Inc.
- Onwuegbuzie, A. J., Leech, N. L., & Collins, K. M. T. (2010). Innovative data collection strategies in qualitative research. *Qualitative Report, 15*(3), 696–726.
- Pawar, M. (2007). Data collection. In N. J. Salkind (Ed.), *Encyclopedia of measurement and statistics* (pp. 220–225). Sage Publications, Inc. <https://doi.org/10.4135/9781412952644.n123>
- Plowright, D. (2011). Data collection: an overview. In D. Plowright (Ed.), *Using mixed methods: frameworks for an integrated methodology* (pp. 49–62). SAGE Publications, Inc. <https://doi.org/10.4135/9781526485090.n5>
- Rawat, S., & Meena, S. (2014). Publish or perish: where are we heading? *Journal of Research in Medical Sciences, 19*(2), 87–89.
- Roberts, L. D., & Seaman, K. (2018). Good undergraduate dissertation supervision: perspectives of supervisors and dissertation coordinators. *International Journal for Academic Development, 23*(1), 28–40. <https://doi.org/10.1080/1360144X.2017.1412971>
- Schober, M. F., Suessbrick, A. L., & Conrad, F. G. (2018). When do misunderstandings matter? evidence from survey interviews about smoking. *Topics in Cognitive Science, 10*(2), 452–484. <https://doi.org/10.1111/tops.12330>
- Shahsavari, Z., & Kourepaz, H. (2020). Postgraduate students' difficulties in writing their theses literature review. *Cogent Education, 7*(1). <https://doi.org/10.1080/2331186X.2020.1784620>
- Singer, E., Van Hoewyk, J., & Maher, M. P. (2000). Experiments with incentives in telephone surveys. *Public Opinion Quarterly, 64*(2), 171–188. <https://doi.org/10.1086/317761>
- Small, M. L. (2011). How to conduct a mixed methods study: recent trends in a rapidly growing literature. *Annual Review of Sociology, 37*, 57–86. <https://doi.org/10.1146/annurev.soc.012809.102657>
- Tessier, S. (2012). From field notes, to transcripts, to tape recordings: evolution or combination? *International Journal of Qualitative Methods, 11*(4), 446–460. <https://doi.org/10.1177/160940691201100410>
- Thiry, H., Weston, T. J., Laursen, S. L., & Hunter, A. B. (2012). The benefits of multi-year research experiences: differences in novice and experienced students' reported gains from undergraduate research. *CBE Life Sciences Education, 11*(3), 260–272. <https://doi.org/10.1187/cbe.11-11-0098>
- Toepoel, V. (2017). Online survey design. In N. G. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 184–202). SAGE Publications Ltd. <https://doi.org/10.4135/9781473957992.n11>

- Tran, T. K., Dinh, H., Nguyen, H., Le, D.-N., Nguyen, D.-K., Tran, A. C., Nguyen-Hoang, V., Nguyen Thi Thu, H., Hung, D., Tieu, S., Khuu, C., & Nguyen, T. A. (2021). The impact of the COVID-19 pandemic on college students: an online survey. *Sustainability*, 13(19), 10762. <https://doi.org/10.3390/su131910762>
- Vehovar, V., & Manfreda, K. L. (2008). Overview: online surveys. In N. Fielding, R. M. Lee, & G. Blank (Eds.), *The SAGE handbook of online research methods* (pp. 176–194). SAGE Publications, Ltd. <https://doi.org/10.4135/9780857020055.n10>
- Weissinger, G. M., & Ulrich, C. M. (2019). Informed consent and ethical reporting of research in clinical trials involving participants with psychotic disorders. *Contemporary Clinical Trials*, 84. <https://doi.org/10.1016/j.cct.2019.06.009>
- Wheeldon, J., & Faubert, J. (2009). Framing experience: concept maps, mind maps, and data collection in qualitative research. *International Journal of Qualitative Methods*, 8(3), 68–83. <https://doi.org/10.1177/160940690900800307>

Appendix. Questionnaire used by the participants

Survey for Research Course

This survey is a part of Research Method course to observe cheating behavior among students when they participated in TOEFL Preparation course as a graduation requirement at Universitas Syiah Kuala.

The use of data

The data obtained from this survey will be used as a material for class discussion and as data for research. Therefore, the respondents are requested to complete the form honestly.

Confidentiality

The respondents do not need to register their names in this questionnaire, so their identity is not recorded. Students who are collecting the data do not have access to the responses given by the respondents. When the data need to be reported, it will be presented in aggregate, so individual respondents are not identified.

Follow up

Students prepared a brief report regarding the process of survey administration and they will receive grade based the quality of their report.

Questions

1. Do you wish to participate in completing this questionnaire?

- A. Yes
- B. No

2. Please select the name of student who sent you this questionnaire.

- A. Student 1
- B. Student 2
- C.

Many students in TOEFL Preparation course obtained very high score in the online post-test compared to their scores in the pre-test. Based on research, this is probably because:

- They used a dictionary.
- They used Google Translate or other translation tools.
- They discussed with classmates or other students.
- They have prepared answer key from other students who had taken the test.
- They received help from other students.
- Their English proficiency has significantly improved.

3. From the possible causes above, which ones did your classmates (or yourself) in the TOEFL Preparation course? (You may select more than one option)

- A. Using a dictionary
- B. Using Google Translate or other translation tools
- C. Discussing with classmates or other students
- D. Using prepared answer key
- E. Receiving help from other students
- F. I really do not know.

4. Please provide other types of cheating that your classmates (or yourself) did when taking the online posttest in TOEFL Preparation course.
