



P-ISSN 2355-2794
E-ISSN 2461-0275

Evaluating Paper-based TOEFL Preparation Program Using the Context, Input, Process, and Product (CIPP) Model

Abdul Manan^{*1}
Muhammad Arif Fadhillah²
Kamarullah¹
Habiburrahim³

¹Islamic History and Culture Department, Adab and Humanities Faculty, Universitas Islam Negeri Ar-Raniry, Banda Aceh 23111, INDONESIA

²Indonesian Language Education, Teacher Training and Education Faculty, Universitas Negeri Samudra, Langsa 24354, INDONESIA

³English Language Education, Teacher Training and Education Faculty, Universitas Islam Negeri Ar-Raniry, Banda Aceh 23111, INDONESIA

Abstract

An evaluation for the development of a paper-based test of English as a Foreign Language (TOEFL) for English Education Department students at one of the colleges in Langsa, Aceh, Indonesia, is necessary, considering their unsatisfactory scores even though the preparation program has been conducted. This research employed context, input, process, and product (CIPP) model developed by Stufflebeam upon the program. Methodologically, interviews, questionnaires, and observation were executed to five TOEFL lecturers and 34 students, while the data obtained were analyzed with the qualitative analysis method of Miles and Huberman. Based on the evaluation of context, the program had an appropriate background, goals, and objectives. Regarding the input, it still had problems related to the availability of resources and facilities. There were no standardized learning materials available, including the broken language laboratory. It affected the process since the lecturers could not maximize the facilities to create an effective learning environment. Eventually, it affected the product since most students could not reach an expected score. Most of the participants agreed that the program needed to be improved. Several improvements can be made to overcome these problems by providing required

* Corresponding author, email: abdul.manan@ar-raniry.ac.id, mananaceh@yahoo.com

Citation in APA style: Manan, A., Fadhillah, M. A., Kamarullah, & Habiburrahim. (2020). Evaluating paper-based TOEFL preparation program using the Context, Input, Process, and Product (CIPP) model. *Studies in English Language and Education*, 7(2), 457-471.

Received April 19, 2020; Revised July 18, 2020; Accepted August 7, 2020

<https://doi.org/10.24815/siele.v7i2.16467>

©Syiah Kuala University. All rights reserved.

learning facilities, designing standardized learning materials, and clustering classes in accordance with students' abilities.

Keywords: CIPP, language testing, paper-based TOEFL.

1. INTRODUCTION

English proficiency test has become indispensable for English department students to acquire chances of advanced education, apply for jobs, and get promotions. Nowadays, English proficiency test has come into several types, such as TOEFL, International English Language Testing System (IELTS), Pearson Test of English Academic (PTE-A), Cambridge English Language Assessment, and so the like. Each has its own types, functions, levels, score ranges, and payment. As the firstborn provider of language proficiency test among the aforementioned ones, TOEFL, which was established in 1964, contributed a reflection toward test-takers' proficiencies in English – specifically for English non-native speakers, institutions, and the provider itself.

One test was Paper-based TOEFL or known as PBT, in the previous time this kind of test was the main choice of many test-takers as it was widely conducted in many areas. Yet, Paper-based TOEFL (PBT) was discontinued in 2017 due to invalidity to measure the four skills in language proficiency because there was no section to test the writing proficiency; it still becomes a choice to measure English proficiency in the level of the university, specifically to make placement, entrance or exit decision (ETS, 2019). Later, the paper-based test was revised with the optional-additional Test of Writing English (TWE) (ETS, 2019). It is even implemented in the level of high schools, such as in Indonesia (Akmal et al., 2020), Japan (Kaneko, 2014), and China (Liu, 2014). At the university level, PBT is prevailed as a substitute placement tool for the English language, for instance, in the University of Illinois, the United States of America (Kokhan, 2012), although a new version of TOEFL, the Internet-based Test (iBT), is considered more comprehensive than Institutional Testing Program (ITP) in measuring English non-native speakers' four skills (Manganello, 2011). Nevertheless, some countries still utilize PBT in the form of TOEFL ITP test as the entrance examination for university students, such as in the University of Tokyo, Japan (Takeuchi, 2020), the University of the East, Philippines (Salimi & Farsi, 2016), and Mahidol University, Thailand (Amurao, 2011). In fact, several universities in Taiwan (Pan, 2015) oblige their students to hand their TOEFL ITP certificate as an exit requirement from English-language programs, although some develop their own English language tests, such as Malaysia (Singh et al., 2014) and Mexico (Walsh, 1995). This perceives a strong background that TOEFL, whether it is iBT, PBT, or even ITP, should be well prepared to cover the objectives, resources, infrastructures, preparations, results, and other related programs.

The PBT preparation program is one of the main subjects for English Education Department students at one of the colleges, Institut Agama Islam Negeri Langsa (abbreviated as IAIN Langsa), in Aceh, Indonesia. It is given in accordance with the requirement of the TOEFL ITP test as a proficiency test. Thus, considering the importance of the TOEFL score, it can be assumed that the institution has put great efforts to make sure that students master English skills.

However, the mastery of each skill within PBT is still disappointing. Based on the TOEFL resume data examined from the Language Development Unit (LDU) of IAIN Langsa, students' average score was within the range of 350-400. The TOEFL tests that had been held for 25 clusters indicated that the students still face difficulties since their scores showed that they only correctly answered 10 to 15 questions for each section. In fact, to pass the minimum score of 450 for non-English department students and 500 for English department students (see [Farkhan et al., 2019](#)), as required at most educational institutions in Indonesia, a student has to answer 25 questions correctly in each section. As a comparison, considered as the most superior university in Aceh Province, Indonesia, Universitas Syiah Kuala requires final semester students to prove their English proficiency through a certificate of TOEFL equivalent with minimum passing grade score of 450 for non-English department students ([Kasim, 2016](#)), even its current condition demands them to hand the minimum score of 478. This indicates that IAIN Langsa still practices the unstandardized requirement for TOEFL scores from their students. Even so, an investigation on to what extent IAIN Langsa has prepared its TOEFL program for its students should be conducted as the students still encounter various obstacles in achieving a 420 score.

In detail, the nature of the difficulties faced by students varies. In the listening part, the difficulties faced by students are very likely related to listening characteristics as proposed by [Brown \(2001\)](#). He deduced eight characteristics of listening – clustering, redundancy, reduced forms, performance variables, colloquial language, and rate of delivery, stress, rhythm, intonation, and interaction ([Brown, 2001, p. 252](#)). Moreover, other areas of listening, such as the quality of recorded materials, cultural differences, accent, unfamiliar vocabularies, length and speed, physical condition, and lack of concentration, may add the difficulties. Furthermore, memories, both long term, and short term, also have an important role in the process of listening ([Flowerdew & Miller, 2005](#)). The success upon listening also depends on the strategies used by learners during listening. Indeed, in the TOEFL listening section, such can be overcome by anticipating the topic discussed, determining the topic or main idea conveyed and predicting the questions offered ([Silviyanti et al., 2020](#)). Concerning the grammar section, the constraints possibly faced by students relate to inherent complexity for rules, salience grammar form in the input, the communicative force of grammar form, input processing strategies, learners' developmental stages, language transfer, and individual differences in language aptitude ([Shiu, 2011](#)). In details, [Ananda \(2016\)](#) discovers that inversions, subject-verb agreements, adverb clause connectors, passives, reduced adjectives clauses, parallel structures, and use of verbs, were the toughest grammatical questions. In the reading comprehension part, the problems commonly faced by students encompassed difficulty in understanding text, lack of focus, laziness, availability of aids, and vocabulary problem ([Garcia et al., 2014](#)). Additionally, unstated and stated details, as well as main ideas, contribute to the students' difficulties ([Samad et al., 2017](#)). Besides, the influence of basic skills, less practice, unmotivated intention to learn, and individual differences may add the problems in learning TOEFL ([Mahmud, 2014](#)).

Therefore, it can be assumed that many factors can cause problems in the process of teaching and learning skills for TOEFL. Unfortunately, based on the interview with the secretary of the English Education Department of IAIN Langsa, there were no formal evaluations ever held by any party to discover the problems during the teaching and learning process of the TOEFL preparation program. Similarly, the LDU head of

IAIN Langsa also clarified such fact; there was no evaluation conducted to deal with difficulties of teaching and learning TOEFL preparation program even though the students' TOEFL score has not yet reached intended outcomes.

To deal with a low TOEFL score, an evaluation is undoubtedly needed since it provides data to improve the students' TOEFL score. According to [Scheerend et al. \(2003\)](#), an evaluation covers systematic information gathering and offers judgments based on the information. [Stufflebeam \(2000\)](#) defines evaluation as the process of delineating, obtaining, and providing useful information for judging alternative decisions. Concerning the case at IAIN Langsa, which indicates the students' problems in learning the material for TOEFL, an evaluation is urgently needed. The gained information as the result of the evaluation process can serve as the basis of policies that can be considered by the institution to develop improvement strategies to help students gain an expected TOEFL score. Then, in evaluating the TOEFL preparation course, the CIPP model is considered as appropriate to be implemented to evaluate the program and to reveal contributing factors resulting in low TOEFL ITP scores among students in IAIN Langsa.

2. LITERATURE REVIEW

First introduced in 1966, the CIPP model assists evaluators to cover the needs, problems, opportunities, and decision-making of a program. It is suggested as a framework to direct the idea, structure, application, and valuation of the service-learning program methodically and offer responses and decisions of the program's value for constant advance ([Hakan & Seval, 2011](#)). The four aspects – context, input, process, and product – of the CIPP model ease evaluators to generate the decisions that should be made. Ideally, those aspects must be considered in conducting a language course to ensure that the course would achieve the goals and objectives of the study.

Firstly, in the aspect of context, the CIPP model evaluates the entire concrete readiness of the system being evaluated. Its goals are analyzed, to analyze whether the system meets the determined needs or not ([Stufflebeam, 2003](#)). Therefore, this aspect encloses on the foundation of any program, including language programs. The term context here refers to the needs in accordance with the objectives and goals within a program, which should be indicated by the collected data during the learning activities ([Ulum, 2015](#)). If it is related to the background of this study, the TOEFL preparation program at IAIN Langsa 'theoretically' should meet the needs of a minimum score of language proficiencies required.

The next aspect, input, back in delineating a program in the point where to indicate compulsory changes ([Stufflebeam, 2003](#)). This aspect maps the possible and potential ways to structure and fix a new, ideal system. Some patronized outlines, as listed by [Ornstein and Humkins \(2004\)](#), may pop out, if the aspect of input takes place, such as availabilities of the program objectives stated, the relationship between the main objective with sub-objectives of the program, the suitability of instructions showcased, supporting alternative strategies of a program, and so the like. According to those ideas, input as the second aspect of the CIPP model, in this study, denotes the approaches, designs, and materials used to serve the needs and objectives of the TOEFL preparation program.

Furthermore, the aspect of process pervades probabilities to criticize the range of a program being applied suitably and competently (Stufflebeam, 2003). Hence, critically, this aspect may suggest outlines, such as the appropriateness of program activities' schedules, planned program activities being conducted, the use of existing properties, and the effective roles of users in the program activities (Stufflebeam & Shinkfield, 1985). To do so, three strategies can be applied as suggested by Ornstein and Humkins (2004), namely identifying the practical scheme of a program, displaying information for conclusions, and retaining a record of trials appearing. In short, the aspect of the process relates to the implementation of plans for guided activities. In this case, as for study, particular existing activities of the TOEFL preparation program including its learning process has to be proper in achieving the goals and objectives.

The aspect of the product in the CIPP model encompasses program results, both intended and unintended ones (Stufflebeam, 2003). To be specific, the crucial purpose of the CIPP product is to assess, manifest, and criticize the completion of a program (Stufflebeam & Shinkfeld, 1985). The judged "product" principally must offer direction for a better modification upon the program to help the necessities of the program's participants and unquestionably, to cut the unnecessary expenses (Tunc, 2010, p. 27). Thus, the product relates to the outcome of learning, both planned and unplanned outcomes, without forgetting the positive and negative results as well.

In a nutshell, rational frameworks for designing each aspect of CIPP model can be patterned by following *Stufflebeam's (2003)* phases, via scoping the program evaluation, gathering the related information of the program, arranging the information gathering, analyzing the information arranged, reporting the information analyzed, and finally, administrating the evaluation. By following the framework, owing to the fact mentioned earlier, the researchers utilized the four aspects of the CIPP evaluation model to scrutinize the TOEFL preparation program offered by IAIN Langsa. The results of this research can be used to improve the program and to reduce any inhibiting factors that may affect students' achievement concerning TOEFL.

As stated before, the CIPP model had been widely implemented as a model to evaluate English programs. A study by *Chen (2009)* evaluated 20 English courses offered at the Applied English Department in Southern Taiwan. This study revealed that the improvement was required on the aspect of context to adjust to the program's needs. A similar study was also conducted by *Karimnia and Kay (2015)* to evaluate the entire Teaching English as a Foreign Language (TEFL) Program at Azad University. The result indicated that the material required improvements in the aspect of readability. *Tunc (2010)* also utilized the CIPP model to evaluate language teaching as a preparatory program for students at Ankara University. The result indicated that improvement was required upon physical conditions, content, materials, and assessment. Similarly, evaluating the distance language-learning program, *Shih and Yuan (2019)* agreed to recommend the CIPP model as a fruitful principle of program evaluation since it offers beneficial references to a private university in northern Taiwan. A comparable study was deployed by *Manan and Azizah (2016)*, where they found the proper relevance of the existing language structure of English curriculum at Universitas Islam Negeri Ar-Raniry, yet the students' TOEFL scores were unimpressive presumably due to de-emphasizing of TOEFL test items in the English language structure syllabus.

3. METHODS

This study was conducted as a qualitative study by applying a descriptive method as explained by [Cohen et al. \(2007\)](#). In addition, to gain a comprehensive understanding of the data collected, the quantitative process was applied in arranging charts for presenting data from questionnaires; it is known as descriptive statistics ([Cohen et al., 2007](#)). This research applied the CIPP model as an evaluative tool encompassing aspects of context, input, process, and product. The summary of evaluation within each aspect reflects the entire state of the program.

The participants of this research were lecturers who taught the preparation program, consisting of five lecturers, and 34 students of the Paper-based TOEFL preparation program within four units. Their ideas and perspective could reflect the entire state of the program. In collecting the data, interviews, questionnaires, and observations were used. The standardized open-ended interview containing set sequences and basic questions ([Cohen et al., 2007](#)) was conducted to five lecturers of the program, including the head of units and the English Department. Then, the structured questionnaire to observe and compare patterns ([Cohen et al., 2007](#)) was used to collect data that relates to the respondents' perceptions of the program. The questionnaire encompasses the aspect of context, input, process, and product. The structured observation was finally applied to gather data from a naturally occurring situation in the area of context, input, and process.

The data analysis is conducted by applying qualitative analysis. It encompassed three processes – data reduction, data display, and conclusion drawing ([Miles & Huberman, 1992](#)). The results of the questionnaire are presented in the form of tables to show the differences of each constraint. It serves as the basis of descriptive analysis, then, triangulation toward the data from the questionnaire was done by comparing and contrasting them by the data gathered from the observation and interviews. This kind of description encompasses particular, general, and interpretative commentary as the interpretation of the existing phenomena.

4. RESULTS

The result highlights the initial description of the PBT preparation program at IAIN Langsa, including its learning materials, process, and result.

4.1 Initial Description of the Program

LDU conducts various services in relation to foreign language programs and services at the university. Concerning the program, LDU arranges the proficiency test, as well as the preparation program for all students of the campus. However, the PBT preparation program is also a compulsory subject in the English Department as a supplementary program for those who want to get an extra lesson about it.

As one of the units of developments at IAIN Langsa, LDU receives annual financial support from the university. In conducting language programs, this unit cooperates with the English Department; lecturers at English Department teach in the PBT preparation program.

4.2 Learning Materials

Based on the interview, lecturers provided their own materials; they arranged it based on the syllabus given. The materials were collected from various resources, which had been adapted to concur with the goals of the program. To examine if the quality of learning materials was appropriate and up to the standard, the questionnaire was given to respondents. Figure 1 shows the result.

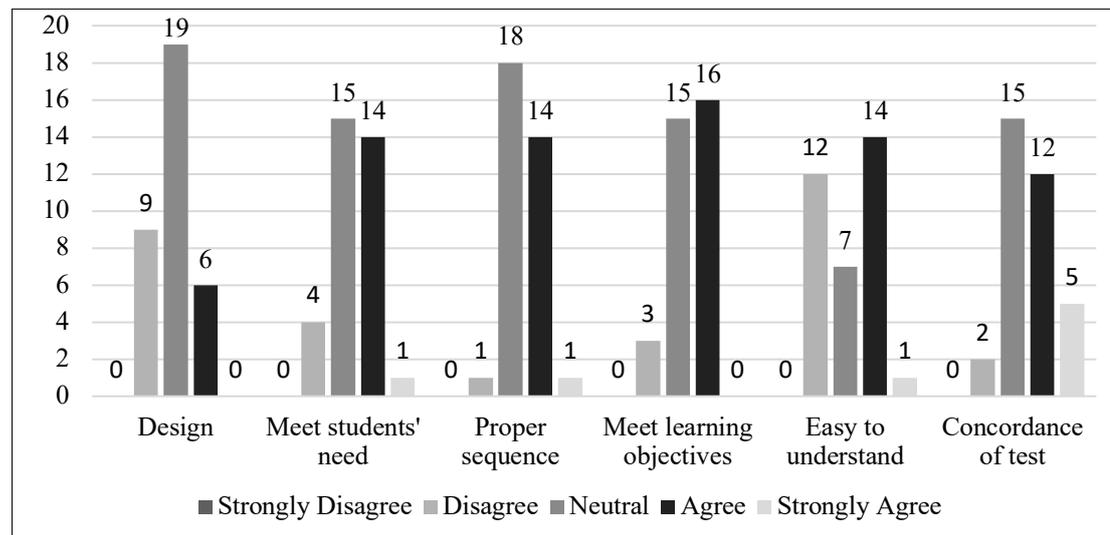


Figure 1. Questionnaire result on the quality of learning material.

Based on Figure 1, half of the students (56%) shared no arguments to the learning materials' design. Forty-one percent of the students agreed that the materials met their needs, while 44% of the rest did not provide answers. Half of them (53%) also had no ideas regarding the proper sequence of the materials. Fortunately, 47% of the students believed that the materials given by the lecturers met the learning objectives, although 35% of them found those hard to understand. Additionally, 44% of the students had nothing to share related to the concordance of the test.

4.3 Learning Process

The learning process relates to the available facilities and resources that can be accessed by students and lecturers during the program. The observation was done to check the availability of resources and materials for the program. From the observation, it was found that the facilities, particularly the language laboratory, could not be used due to technical issues. Then, the lecturers only used portable speakers in teaching listening, which produced unclear sound. In addition, the questionnaire was also distributed to assess students' perspectives on available resources and facilities for learning. Figure 2 shows the result of the availability of facilities and resources.

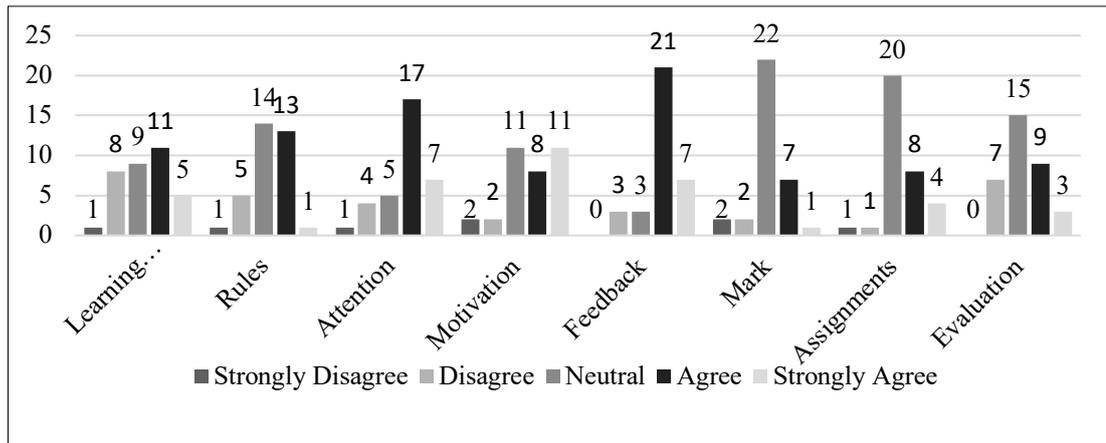


Figure 2. Questionnaire result on the process of learning.

In Figure 2, the students' responses towards the learning process were varied. They dominantly agreed with the learning strategies (32%), attention (50%) and feedback (62%). Then, a neutral option dominated the results, particularly in responding the rules (41%), mark (65%), assignments (59%), and evaluation (44%). Coincidentally, the same percentage, 32%, was appeared in the Neutral and Strongly Agree options as the students thought that the learning process had a possibility to raise their motivation.

4.4 Learning Result

Learning result correlates to the outcome, or product, from the learning process. The data for this particular aspect can be examined from several sources, such as the result of the TOEFL test, as shown in Figure 3 students' opinion, and the concordance to the program objectives, which could be gathered from the lecturers' perspectives.

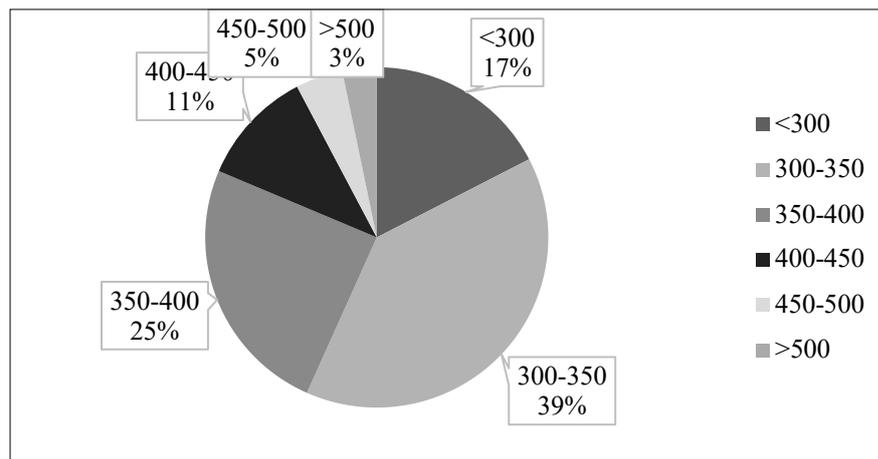


Figure 3. IAIN Langsa students' TOEFL scores in 2019.

From Figure 3, out of 247 students, 97 of them (39%) scored their TOEFL test around 300-500. This became the dominant score range obtained by the students. Next, sixty-one students (25%) scored around 350-400. The third dominant one was 43 students (17%) who scored under 300. The pool was then dominated by 11% of the

students, who scored 400-450 and 5% of the students who obtained 450 to 500 total score. Only 3% (8 students) achieved score above 500.

5. DISCUSSION

By displaying the results, the evaluation of all aspects of the CIPP model – context, input, process, and product – upon the program can be discussed. The analyses draw further insights for a better improvement to the language program.

5.1 Evaluation of the Aspect of Context

The students taking the PBT preparation program are expected to meet the minimum requirement of 420 TOEFL scores in order to graduate. The goal of the program was specified in the syllabus of the program, that is “To prepare students upon the appropriate skill and knowledge on the listening, structure, that can improve their mastery on English, particularly to get sufficient score on the test”. This statement indicates that the TOEFL preparation program conducted by LDU was designed to help students get a maximum score on the test, as well as to improve their English skills in general. This goal has met the criteria of good goals for the program since it is not only aimed at improving the score, but the program also focuses to improve students’ English skills in general.

Students’ perspectives regarding the context of learning can be traced from the result of questionnaires about learning outcomes as in Figure 1. Sixty-five percent of students stated that the program could not meet the aim. Furthermore, all of them expected that the program needed improvement. Not many students got an expected score. Despite this fact, students stated that the lecturers had used various methods in the program delivery to reach an expected score.

5.2 Evaluation of the Aspect of Input

The aspect of input relates to the available resources such as books, modules, and any other relevant sources of learning. Based on the observation, it was found that LDU experienced a lack of resources for learning. Even worse, the only language laboratory available was broken since 2018, as exposed during the interview with the head of the unit. However, LDU had tried to facilitate students by preparing a portable speaker, which could be used by the lecturers while teaching listening for the program. Based on the observation of the facilities, the condition of the classroom was not suitable for learning to listen. The fact that the classroom used as a laboratory for teaching listening was not soundproof resulted in the loss of students’ concentration and difficulties in listening to the materials given. This was because the noise from the outside classroom would interfere during the teaching-learning process. As far as the quality of the recorded materials was concerned, the materials were sufficient. The unit had transferred the recorded materials to the digital form in the form of a flash drive to prevent damage. In fact, language learners, specifically in the level of the university, should advance their language mastery in dedicated language services like self-access centers in a language laboratory (Farooqui, 2007), or a digitalized one (Wagener,

2006). By having such, competitiveness may be generated by the learners for either practicing listening or speaking skills, debating issues, getting icebreakers, or conducting language tests, for sure.

While learning materials have been digitalized, it was found out during the interview with one of the lecturers that there was no specific module used in teaching listening. As far as Wang (2019) is concerned in her study, four teachers teaching the program fully rely on the two TOEFL textbooks, namely ‘Longman Introductory Course for the TOEFL Test’ and ‘The Complete Guide to the TOEFL Test PBT Edition’. Wang amplified the use of specific modules, which was not available in the preparation course at IAIN Langsa. Nevertheless, even though LDU had provided syllabus as the guidance for teaching and learning process, which was principally integrated with the two textbooks mentioned, there were no specific modules available. Interestingly, students did not consider module unavailability as serious matters. This can be observed from the result of a questionnaire showing that 41% of them stated that the learning materials had met their needs. The absence of designed module might have enabled lecturers to adapt learning materials to suit the condition and ability of their students. A similar interesting case was also noted by Souriyavongsa et al. (2013) that the designed and “rigid” learning material could lower the students’ achievements in language learning. The teaching strategies of the lecturers that were done within the program might contribute to the satisfactory responses of the students. In relation to supporting learning materials, the head of LDU stated that they experienced a lack of supporting materials for TOEFL. Available textbooks on the program in the library were so limited that the students could not borrow it; they could ask Xerox it if needed.

To conclude the evaluation of the aspect of input, it was clear that materials and facilities to support the teaching-learning process of the program were still insufficient. This extends the outlines coined by Souriyavongsa et al. (2013) that the insufficient and inappropriate, as well as a “rigid” material could lower the achievement toward language learning. Improper facilities, such as the broken laboratory for teaching listening, the unavailability of designed and TOEFL modules, are indeed imperative input for a language program like TOEFL. Consequently, all these undoubtedly affected students’ achievements. Then, the language course had to prepare and provide the proper material as a form of feasibility in the aspect of input.

5.3 Evaluation of the Aspect of Process

The aspect of the process correlates to the implementation of the goals and objectives of the program, as well as the integration of facilities and resources toward the program. From the questionnaire result, it was revealed that the lecturers had given their best, despite the lack of facilities and resources. The result of the questionnaire indicated the effort of the lecturers in conducting the program; they taught using a variety of approaches, methods, and strategies. The aforementioned three strategies proposed by Ornstein and Humkins (2004) through the instruments were associated to evaluate the program’s process. Based on the identifying, displaying, and retaining the findings, two problems were raised – students’ lack of motivation and absence of resources and facilities (Mahmud, 2014).

During the interview, one of the lecturers explained that she had to motivate students repeatedly to learn the program. The result of the questionnaire confirmed a similar pattern. Students admitted that the tutor gave them much attention and

motivation, along with feedback. These were aimed at improving students' motivation in learning TOEFL. This also corresponds to the result in Figure 2. Such factors were important in learning a foreign language as stated by McGinnis (2007) in her study. It was realized that the students took this program for the sake of graduation; they could not graduate if their score did not reach the minimum passing criteria (Farkhan et al., 2019). Moreover, as explained by the tutor, they only attended the meeting to get the score of attendance, with the hope that it would help them pass the test. Furthermore, the unavailability of resources and facilities also affected the teaching-learning process. One of the students stated in the questionnaire that they need more access to audio since listening materials were considered difficult to get. It was in line with the study of Lodhi et al. (2019) which stated that the availability of sufficient facilities and materials affecting the learning achievement of students.

Summarizing the problems within the aspects of the process, there are two major problems. The first was the students' lack of motivation. The second problem was related to the aspect of input as learning materials were not sufficient.

5.4 Evaluation of the Aspect of Product

The evaluation of the aspect of the product is related to the result of the learning process. From the result of the TOEFL prediction test during 2019, it was revealed that most test-takers only got the score ranging from 300 to 400, which was not enough to pass the minimum required score of 420 for the undergraduate degree. As displayed in Figure 3, the 3% of them retrieving score above 500 indicated that the program was ineffective and needed to be improved. The data from the questionnaire also revealed that all participants expected improvements in the delivery of the program. Most of them were not satisfied with the program.

Students' low scores could be related to the two previous aspects: input and process. Input, related to the availability of resources and facilities as in the study of Wang (2019), indeed affected the teaching-learning process. Lack of learning facilities and materials had undoubtedly crippled program delivery.

With regard to the aspect of the process, the occurring problem was also caused by insufficient learning input. Students' motivation to learn was somehow affected by the unavailability of facilities, which was irrelevant to the findings of Mahmud (2014) noting that the appropriate facilities provided a positive impact on learning outcomes. Despite optimal efforts from lecturers, it could not be denied that lack of facilities has added to the problem.

5.5 Feedback for Further Improvement

Based on the CIPP model, it can be concluded that there were several problems taking place in the delivery of the PBT preparation program at IAIN Langsa. With the information gathered by using the CIPP model, it is then possible to propose strategies to improve the result of students' TOEFL scores.

The first problem laid upon the aspect of input. There were no sufficient facilities and resources for learning. The second problem was on the aspect of the process; since the facilities and resources were insufficient, the learning process was not effective. At last, problems regarding both aspects affected the learning result; students' achievement was low and they expected improvement within the programs.

The lack of facilities and resources in many ways is related to the financial issue of the university. As a public university, funding to procure new facilities to improve the teaching-learning process requires leadership commitment. The process of planning is also time-consuming. To make the best out of the situation, the students can be encouraged to use their own devices. Nowadays many students possess their own gadget, laptop, or smartphone. To work around a lack of teaching-learning facilities, it is possible to ask students to bring their own devices during learning. While this is not an ideal permanent solution, it would temporarily help students during the class. The lecturers can distribute and share digital learning materials to the students. The use of students' own devices also benefits them. They can save learning materials, bring them home, and learn the materials when they need them.

The second strategy is by designing a standardized module, which is crucial to address students' low achievement. LDU can invite lectures to discuss learning materials for the program. Then, the result of the discussion can be used as a guideline in designing a new module for the program. The module's content may be adapted from existing books and materials used by lecturers; yet, it must suit the learners' ability and condition. Such material will ensure the commonality of materials, as well as skills that will be mastered by students.

The last strategy proposed in this research is restructuring the class. The program of TOEFL preparation that had been conducted by LDU was done in a mixed class. As a result, the students with good mastery had to learn with the students of low-level English mastery; it caused several problems including constraints in selecting learning materials. Thus, the restructuring of the classroom is necessary for further improvement. Students from a different level of English mastery should be grouped separately. Reducing class size may also suggest a solution as suggested by [Kasim \(2016\)](#). The program can at least be divided into three groups – beginner, intermediate, and advance. A placement test can be conducted at the beginning of the program to examine the ability of students. Those who get the score under 300 can be placed into the beginner level. The students who get a score of 300 to 400 can be placed into the intermediate level. Meanwhile, those who get a score over 400 can be placed into an advanced level class. As a result, the tutor will be easier to determine and to design learning materials as the students are at a similar level. It is expected that students will feel more comfortable to study since they are on the same level as their classmates. Also, it will be easier for LDU to evaluate the students' achievement and development.

6. CONCLUSION

The evaluation of the PBT preparation program through the CIPP model reflected some inefficiencies. Although it had an appropriate background, goals, and objectives, the program still had problems related to the lack of resources and facilities. There were no standardized learning materials available and the broken language laboratory. Subsequently, the lecturers could not deliver the program effectively, followed by the fact that most students could not reach an expected score. Both groups agreed that the program needed to be improved.

Based on the analysis, several improvements can be suggested. First, the program should provide students with appropriate facilities for learning, including repairing the laboratory and alternating students' gadgets. Second, there should be

standardized learning materials, and third, the students should be separated based on their mastery level. The implemented CIPP model has significantly portrayed the essential concepts within the program's preparation, application, and evaluation. Yet, the improvement toward the program must be traced to ensure the development of the weakness of the program. Hence, the evaluation must be done as a routine activity by the end of the program or in the periodical term. In other words, a similar study can be done to re-evaluate the improvement of the program after two years.

REFERENCES

- Akmal, S., Risdaneva, R., Habiburrahim, H., & Sari, M. (2020). The English teachers' challenges in TOEFL preparation for senior high school students. *Journal on English as a Foreign Language*, 10(1), 25-45.
- Amurao, A. L. (2011). The types and nature of questions vis-à-vis students' test-taking skills as significant indicators of second language examinees' performance on the TOEFL-ITP Reading Comprehension sub-test. *The International Journal of Research and Review*, 6(2), 39-56.
- Ananda, R. (2016). Problems with Section Two ITP TOEFL Test. *Studies in English Language and Education*, 3(1), 37-51.
- Brown, H. D. (2001). *Teaching by principles*. Longman.
- Chen, C.F. (2009). *A case study in the evaluation of English training courses using a version of the CIPP Model as an evaluative tool* [Unpublished doctoral thesis], Durham University.
- Cohen, L., Manion, L., & Morrison, K. (2007). *Research methods in education* (6th edition). Routledge.
- ETS. (2019). *About the TOEFL ITP assessment series*. www.ets.org/toeflitp
- Farkhan, M., Nurlia, V., & Haucsa, G. M. (2019). University students' perception on TOEFL as a graduation requirement: A case in UIN Syarif Hidayatullah Jakarta. *Elite Journal*, 1(1), 51-64.
- Farooqui, S. (2007). Developing speaking skills of adult learners in private universities in Bangladesh: Problems and solutions. *Australian Journal of Adult Learning*, 47(1), 94-110.
- Flowerdew, J., & Miller, L. (2005). *Second language listening: Theory and practice*. Cambridge University Press.
- Garcia, Y. P., Ramayan, A. J., Sepe, A. P., & Silor, A. C. (2014, March 4-5). *Difficulties in reading comprehension and metacognitive strategies for technology livelihood education students* [Paper Presentation]. Global Summit on Education, Kuala Lumpur, Malaysia.
- Hakan, K., & Seval, F. (2011). CIPP evaluation model scale: Development, reliability and validity. *Procedia-Social and Behavioral Sciences*, 15, 592-599.
- Kaneko, M. (2014). Is the vocabulary level of the reading section of the TOEFL internet-based test beyond the lexical level of Japanese senior high school students?. *Vocabulary Learning and Instruction*, 3(1), 44-50.
- Karimnia, A., & Kay, E. (2015). An evaluation of the undergraduate TEFL program in Iran: A multi case study. *International Journal of Instruction*, VIII(2), 83-98.
- Kasim, U. (2016, November 12-13). *The implementation of TOEFL score as a requirement for script examination at Syiah Kuala University* [Paper

- Presentation]. English Education International Conference, Banda Aceh, Indonesia.
- Kokhan, K. (2012). Investigating the possibility of using TOEFL scores for university ESL decision-making: Placement trends and effect of time lag. *Language Testing*, 29(2), 291-308.
- Liu, O. L. (2014). Investigating the relationship between test preparation and TOEFL iBT performance. *ETS Research Report Series*, 2014(2), 1-13.
- Lodhi, M. A., Sahar, A. H., Qayyum, N., Iqbal, S., & Shareef, H. (2019). Relationship of school environment and English language learning at government schools. *Public Administration Research*, 8(1), 1-13.
- Mahmud, M. (2014). The EFL students' problems in answering the test of English as a foreign language (TOEFL): A study in Indonesian context. *Theory and Practice in Language Studies*, 4(12), 2581-2587.
- Manan, A., & Azizah, A. (2016, November 12-13). *Formative evaluation of the English structure of English department curriculum at teacher training faculty, the State Islamic University (UIN) Ar-Raniry* [Paper Presentation]. English Education International Conference, Banda Aceh, Indonesia.
- Manganello, M. (2011). *Correlations in the new TOEFL era: an investigation of the statistical relationships between iBT scores, placement test performance, and academic success of international students at Iowa State University* [Unpublished master's thesis]. Iowa State University.
- McGinnis, A. M. (2007). *The role of attention in second language development* [Unpublished master's thesis]. University of Montana.
- Miles, M. B. & Huberman H. M. (1992). *Analisis data kualitatif: Buku sumber tentang metode-metode baru* [Qualitative data analysis: A source book on new methods]. UI Press.
- Ornstein, A. C., and Humkins F. P. (2004). *Curriculum: Foundations, principles and issues*. Prentice-Hall.
- Pan, Y. C. (2015). Test impact: English certification exit requirements in Taiwan. *TEFLIN Journal*, 20(2), 119-139.
- Salimi, E. A., & Farsi, M. (2016). Program evaluation of the English language proficiency program for foreign students a case study: University of the East, Manila Campus. *English Language Teaching*, 9(1), 12-21.
- Samad, I. A., Jannah, M., & Fitriani, S. S. (2017). EFL students' strategies dealing with common difficulties in TOEFL reading comprehension section. *International Journal of Language Education*, 1(1), 29-36.
- Scheerend, J., Glas, C., & Thomas, S. M. (2003). *Educational evaluation, assessment, and monitoring: A systematic approach*. Swets and Zeitlinger Publishers.
- Shih, Y. C., & Yuan, Y. P. (2019). Evaluating an English Elite Program in Taiwan using the CIPP Model. *Journal of Asia TEFL*, 16(1), 200-219.
- Shiu, L. J. (2011). *EFL learners' perception of grammatical difficulty in relation to second language proficiency test* [Unpublished doctoral dissertation] University of Toronto.
- Silviyanti, T. M., Rahmadhani, R., & Samad, I. A. (2020). EFL students' strategies in answering the Listening Section of the Longman TOEFL. *Studies in English Language and Education*, 7(1), 237-246.

- Singh, H., Thamilvili, G., Zuraimi, N. A. B., & Veliapandian, Y. (2014). How to improve on the Malaysia University English Test (MUET). *Asian Journal of Education and e-Learning*, 2(06), 485-489.
- Souriyavongsa, T., Rany, S., Abidin, M. J. Z., & Leong, L. M. (2013). Factors causes students low English language learning: A case study in the National University of Laos. *International Journal of English Language Education*, 1(1), 179-192.
- Stufflebeam, D. L. (2000). The CIPP Model for evaluation. In D. L. Stufflebeam, G. F. Madaus, & T. Kellaghan (Eds.), *Evaluation models: Evaluation in education and human services*, (pp. 279-317). Springer.
- Stufflebeam, D. L. (2003). The CIPP Model for evaluation. In T. Kellaghan, & D. L. Stufflebeam (Eds.), *International handbook of educational evaluation* (pp. 31-62). Springer.
- Stufflebeam, D. L., & Shinkfield, A. J. (1985). *Systematic evaluation*. Kluwe-Nijhoff.
- Takeuchi, W. (2020). *Guide to the 2020 Entrance Examination Department of Civil Engineering Graduate School of Engineering*. The University of Tokyo.
- Tunc, F. (2010). Evaluation of an English language teaching program at a public university using CIPP model [Unpublished master's thesis]. Middle East Technical University.
- Ulum, Ö. G. (2015). A needs analysis study for preparatory class ELT students. *European Journal of English Language Teaching*, 1(1), 14-29.
- Wagener, D. (2006). Promoting independent learning skills using video on digital language laboratories. *Computer Assisted Language Learning*, 19(4-5), 279-286.
- Walsh, M. B. G. (1995). *A comparison of Instituto Tecnológico y de Estudios Superiores de Monterrey's EFL exit examinations and the TOEFL* [Unpublished master's thesis]. Instituto Tecnológico y de Estudios Superiores de Monterrey
- Wang, Y. T. (2019). The impact of TOEFL on instructors' course content and teaching methods. *TESL-EJ*, 23(3), 1-18.