The Relationship between Formative and Summative Assessments Using the Mosoteach App in College EFL Classes in China

Dianping Liu¹, Yang-Hee Kim²*
¹Associate Professor, Yanbian University, ²Associate Professor, Baekseok University

Mosoteach 앱을 이용한 중국 대학 EFL 수업 형성평가와 총괄평가와의 관계

Abstract
This study was intended to help improve assessment in college EFL classes by examining the relationship between formative assessment (FA) and summative assessment (SA). Participants for the study were 75 undergraduates, taking an eight-week optional college English course at a four-year college in China. FA data were collected in the final course of the course by a self-reporting survey using the online training platform Mosoteach. To achieve the finding, the relationship between FA scores and SA scores (final exam and performance scores) was analyzed in SPSS by means of Pearson correlation analyses. Significant positive correlations were found between FA and SA scores overall. In addition, students’ performance on chapter tests, online discussions, brainstorming, quick-responses, assignments and the number of thumbs-up clicks by teachers were significantly correlated with SA scores. The results suggested that FA administered through the Mosoteach app could improve students’ academic performance, thus providing an empirical basis for improving educational assessment. Based on these findings, implications for assessment in EFL classes were described.

Key Words: EFL, self-reporting survey, correlation, thumbs-up clicks, academic performance

요약 본 연구는 형성평가 (FA)와 총괄평가 (SA) 사이의 관계를 조사하여 대학 EFL 수업 평가 개선을 돕고자 수행되었다. 이 연구의 참여자들은 중국의 4년제 대학에서 선택적으로 8주간의 대학 영어 과정을 수행한 75명의 학생들이다. FA 데이터는 온라인 교육 플랫폼 Mosoteach를 사용하여 자기 보고식 설문 조사에 의해 코스 마지막 과정에서 수집되었다. 연구 결과를 도출하고자, Pearson 상관관계 분석을 통해 FA와 SA(기말고사 및 수행활동 점수)간의 관계를 SPSS로 분석하였다. 연관성으로 그 두 평가 점수 간 유의미한 긍정적 상관관계가 있다는 것으로 나타났다. 학생들의 단원 평가, 온라인 토론, 브레인스토밍, 신속한 답변, 그리고 과제 수행에서 교수자들의 엄지척 클릭수와도 깊은 상관관계를 보였다. Mosoteach 앱을 통해 관리되는 FA가 학생들의 수업의 학업 성과를 항상시킬 수 있다는 결론은 교육 평가 개선을 위한 경험적 기반을 제공할 수 있음을 시사하였다. 이러한 연구 결과에 따라 EFL 수업 평가에 대한 중요성이 설명되었다.

주제어: 외국어로서의 영어, 자기 보고식 설문조사, 상관관계, 엄지척 클릭, 수업의 학업 성과

©This work was supported by a General Grant on Higher Education and Teaching Reform of the NEAC, China (19080): a General Grant of Social Science Fund of Jilin Province, China (2021B225): a General Grant of Jilin Higher Education Institution (JGIX2019D39)

*Corresponding Author: Yang-Hee Kim (214julie@bu.ac.kr)

Received July 26, 2021
Accepted August 20, 2021
Published August 28, 2021
1. Introduction

In educational settings, formative assessment (FA), focusing on learning processes and students’ development began to be more highly valued than summative assessment (SA), which lays emphasis on learning results and dealing with big-scale exams in the late 1980s, particularly in western countries (Cowie & Bell, 1999)[1]. FA is defined in the College English Teaching Guidelines (2004)[2] as various methods used to track teaching, offer feedback, and promote student development toward instructional goals. Although a recent survey showed that college English teachers in China recognize the importance of FA (Jin & He, 2015)[3], these teachers still face difficulties with implementation, owing to time constraints, as well as lack of training about how to conduct FA. As a result, EFL teachers do not pay as much attention to FA in practice as they would like (Jin & Sun 2020)[4]. Thus, there is a need for research that supports the implementation of FA in college EFL classes.

Although formal policies in China, as embodied in documents such as the College English Teaching Guidelines (2020)[5], emphasize the importance of FA, EFL instruction in Chinese higher education still focuses more on summative assessment (SA) than on FA. As a result, empirical research on FA in higher education EFL classes is limited as well (Tang & Zhang, 2007; Wang et al., 2018)[6–7], even though advances in educational technology enable increasingly convenient incorporation of FA into teaching practice. Even when EFL studies consider assessment, SA and FA tend to be separated rather than integrated (Lu & Mao, 2018)[8]. In light of this gap in prior research, the present study examined the relationship between FA and SA in an undergraduate EFL class. FA data were obtained through the Mosoteach app (one of the commonly-used online education platforms in Chinese EFL instruction) and correlated with SA data. The research questions explored in this study were whether FA and SA would be related overall, and whether individual components of FA within Mosoteach would be positively related to SA. An affirmative answer to either or both questions would suggest that FA contributes to student outcomes.

2. Literature Review

2.1 A Worldwide Paradigm Shift in Assessment Culture

The benefits of feedback from FA are attributable to, inter alia, the academic guidance. Traditionally, educational assessment in China and other countries was mainly summative, yielding quantitative information about student outcomes such as grades or standardized exam performance (Herrera & Macía, 2015)[9]. SA data are useful, but they provide little information about the learning process.

In order to compensate for the shortcomings of traditional, result-oriented SAs, FAs began to be more widely used in the late 1980s. With respect to assessment there had been a paradigm shift from assessment of learning (AOL) to assessment for learning (AFL) (Wai & Hirakawa, 2001)[10]. Studies suggest that FAs have significant benefits for student learning and enhance academic as well as behavioral outcomes (Alison, 2005; Black & Wiliam, 1998; Darrell et al., 2014)[11–13]. Most researchers believe that these benefits of FAs result from opportunities they provide teachers to assess individual student progress and use this information to make beneficial changes in instruction.

An important part of FA is feedback, which helps learners become aware of their progress toward instructional goals as well as what actions are necessary to achieve those goals (Sadler, 1989)[14]. Through feedback, FA helps prevent
the destructive cycle in which students attribute poor performance to lack of ability thereby becoming discouraged and unwilling to engage in further learning (Ames, 1992; Vispoel & Austen, 1995)[15-16].

While feedback generally originates from teachers, learners can also play an important role in FA through self-evaluation and peer-evaluation. Experimental studies have shown that students’ engagement in learning and lowered anxiety can result from appropriate feedback. For example, in emphasizing the importance of incorporating active experiences into FA, Darrell et al. (2014)[13] noted that students respond to these experiences with greater involvement. Cardozo et al. (2020)[17] argued that appropriate implementation of FA results in significantly lower levels of stress and anxiety, as well as improving student performance, as compared with a traditional lecture-style pedagogy. In sum, FA can play an important, beneficial role in supporting student progress, through direct support for academic skills as well as motivational and emotional support that also contributes to academic success.

2.2 Assessment Reform in China

FA has been incorporated into Chinese college English instruction via the College English Curriculum Requirement in 2007, and it is increasingly accompanying the traditional, large-scale standardized testing (Zhang, 2019)[18]. In brief, teachers are now being asked to implement FA in their classrooms, and they are being encouraged to conduct their own research on FA (Gu & Yu, 2020)[19].

EFL studies in China have shown that FA can promote instructional quality, mobilize students’ interest in learning, increase student initiative, and cultivate students’ independent ability and writing skills (Cao, 2004)[20]. However, FA is not meant to replace SA (Alison, 2005)[11]. Rather, FA that guides day-to-day instruction is expected to support better performance on SAs such as final exams in classes and standardized national exams (Richard et al., 2008)[21].

In China, implementing FA remains challenging due to the influence of the country’s traditional, examination-driven educational system as well as instructors’ lack of training on how to conduct FA. Limited instructional time and large class sizes create more obstacles, along with lingering assumptions that SA and learning outcomes are most important. Research on foreign language FA shows that although teachers recognize the limitations and stress created to students by overemphasis on SA, SAs still gain the most attention since they are easier and less time-consuming to administer than FAs (Wang et al., 2018)[7]. Even though instructors are being asked to do FA in class, most regard it as essentially a series of frequently conducted small quizzes or assignments. As a result, instructors mostly just grade FAs as if they were SAs, instead of providing effective feedback. This prevents FA from playing its proper role in assessing student progress and then using this information to modify instruction. Finally, ensuring the reliability of FA is a challenge due to unclear standards on how it should be conducted (Guo & Yang, 2003)[22].

With the integration of internet technology into education, new trends such as mobile learning are attracting much attention. The Mosoteach online teaching platform, one of the commonly used online app assistants, is suitable for mobile network environments. With Mosoteach app, teachers can create a class and invite students to join, as well as upload class resources, post assignments and group tasks, administer questionnaires, facilitate brainstorming, engage in discussions, give tests, and so on. Students can use their mobile devices to subscribe to classes, complete assignments, receive messages, download courseware, view videos, and so on. Apps such as this can
stimulate students’ interest in learning, and enrich students’ online learning experiences (Zhang, 2020)[23].

Because Mosotech allows teachers to track each student’s learning progress, it facilitates the integration of FA into teaching practice. However, research to date on Mosotech focuses more on the app’s role in teaching methods than on assessment (Liu & Wu, 2020)[24]. Research on the relationship between FA and SA in such apps is particularly rare. Thus, this study examined the relationship between FA and SA utilizing Mosotech. The two research questions guiding this study are:

1. Does FA on the whole promote SA performance?
2. Which specific components of FA in the Mosotech app are related to SA performance?

3. Research Methods

3.1 Participants

Sampling was carried out at a comprehensive four-year university located in Yanbian Korean Autonomous Prefecture in Jilin Province, a northeastern part of China. Eighty-two undergraduates attending an optional intercultural communication college English course agreed to voluntarily complete surveys during academic year 2018–2019. Seventy-five surveys were fully completed and thus constituted the sample used for data analysis.

Seven-six percent of the sample was female, and 24% was male: 71% of the sample self-identified as Han Chinese, while 29% self-identified as Korean Chinese or some other ethnic minorities. The gender distribution of the sample is representative of university-level English classes in China, where it is normal to see more female students. The percentage of ethnic minority students (29%) was lower than the university-wide population (40–45%), perhaps owing to ethnic minority students’ lower self-confidence in their English competence.

3.2 Measures

Students impressions of the course and its FAs were obtained from a self-report survey collected on the Mosotech app at the end of the course. Survey questions were mainly drawn from validated scales used in published research and consisted of Likert-type response scales. (The actual questions were presented along with the results in the next section.)

Each of the FA components used in the course was graded by the instructor on a conventional 100-point scale. The SAs used in the course were the final exam and team project. These SAs were graded separately by the instructor, each on a conventional 100-point scale.

3.3 Data Analysis

For purpose of analysis, each participant’s response to each survey question was coded into one of two categories: mostly or somewhat agree versus uncertain, somewhat disagree or mostly disagree.

SPSS was used for all data analyses. For each student, an FA mean was derived from grades on all FA components in the course. In order to calculate descriptive statistics, all students’ mean FA scores were then averaged to create a single FA mean. The SA mean was the average of all students’ final exam and team project scores. Pearson correlations were used to determine the relationship between FA scores and SA scores. These analyses included correlations between individual FA components and SA scores, as well as between overall FA scores and SA scores.

4. Research Results & Discussion
4.1 Participants’ Views of the Course & Assessment

Concerning survey question 1 to 3, the benefit and the degree of difficulty the students faced, although just over one-fifth of the students considered the course difficult, almost all stated that they benefited from it. Bonferroni-corrected pairwise comparisons showed that “understanding the instructor’s English” and “engagement in curricular activities” were considered significantly greater causes of difficulty than “the course content” and “the assessment” (all \( p < .05 \)), a finding which hints that students’ weak English competence and unfamiliarity with student-centered instruction created challenges for them, whereas course content and FAs were not as challenging.

Regarding survey question 4 and 5, their opinions on the instructor’s teaching style and the strength of the course, despite the fact that all the students favored the instructor’s teaching style, some students still recommended strengthening the course in its intercultural communication theory, extent of resource learning, and opportunities for student involvement, indicating that there was still space for the course to be improved.

Concerning survey question 6 and 7, their opinions on the efforts they made in the course and the instructor’s assessments, most of the students claimed that they studied hard in the course, and that the instructor’s assessments were fair on the whole; both findings indicate positive views of their own efforts and the instructor’s assessments.

However, slightly fewer students indicated positively in response to survey question 8, the effectiveness of the FAs that led to their efforts. The different percentages might be a hint that students were more satisfied with the SAs than the FAs, perhaps because the FAs required more time and effort overall.

For survey question 9, interest-driven activities, “resource learning”, “sign-in”, and “online discussion” were the top three activities that aroused interest in learning, with slightly fewer students choosing other options. Given the large class size (82 students) and students’ limitations in English, it was encouraging to see that autonomous learning activities like resource learning were viewed positively, and that online discussion was a preferred way for students to express themselves and communicate with the instructor and classmates.

4.2 Descriptive Statistics and Correlational Analyses for Student Assessment Data

The descriptive statistics for each form of assessment are shown in Table 1. For the FAs, the median and the mean were both high, suggesting that most students actively participated in FA activities and performed well. However, the minimum and maximum FA scores were 25 and 92 points, respectively, indicating substantial variability across students.

The minimum and maximum values for SA were 74 points and 91 points, respectively, indicating some variability, but not as much as for FA (as also indicated by the smaller standard deviations). Although the means, medians, and maximum scores for SA and FA were quite similar, the differences in ranges and standard deviations showed that some students struggled with the FA. Of particular importance is that the lowest SA value was 74, while the lowest FA value was 25. This finding in turn suggests that for some Chinese students, it is still challenging to shift from traditional result-centered learning to process-centered learning.

Table 1. Descriptive Statistics for Participants’ Formative & Summative Assessments

<table>
<thead>
<tr>
<th>Assessment</th>
<th>N</th>
<th>Mean</th>
<th>Median</th>
<th>Min</th>
<th>Max</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative Assessment</td>
<td>75</td>
<td>82.3</td>
<td>83</td>
<td>25</td>
<td>92</td>
<td>9.29</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>75</td>
<td>80.9</td>
<td>81</td>
<td>74</td>
<td>91</td>
<td>5.14</td>
</tr>
</tbody>
</table>
4.3 Correlational Analyses

The Pearson correlation between FA and SA was .333, a significant value \( (p < .005) \), indicating that the higher the FA score, the higher the SA score. This finding addresses the first research question by demonstrating an FA–SA relationship.

Table 2 presents descriptive statistics for each FA component, as well as correlations between the FA and SA results for that component. The correlational results for chapter test, assignment, online discussion, quick-answer, brainstorming, and teacher’s thumbs-up were all significant (all \( p \)'s < .05), indicating the higher the score of each FA component, the higher the SA scores. This addresses the second research question by showing that many individual components of FA were related to SA. At the same time, the correlations between FA and SA for resource-learning and sign-in were not significant. Following is a discussion of each of the significant correlations reported in Table 2.

The significant FA–SA correlation for chapter tests \( (p < .001) \) indicated that the higher the score on the tests, the higher the score on the final exam, which suggests that FA in the form of chapter tests helped students consolidate and apply what they have learned, effectively improving their SA score on the final exam. The same pattern was observed for assignments \( (p < .05) \), which suggests that they played a similar role in consolidating student knowledge.

The significant FA–SA correlations for online-discussion \( (p < .01) \) and brainstorming \( (p < .05) \) suggest that the more students got involved in these group activities, the better they performed on the final exam, a result which suggests that students benefited from such collaborative activities.

The significant FA–SA correlation for quick-response \( (p < .005) \) suggests that student outcomes could be influenced by teachers’ recognition and respect, as well as teachers learning about individual students through interactions with them and adjusting instruction accordingly. Further research is needed to distinguish the relative combination of each.

The instructor used surveys to learn about student views of the course, and students received credit for completing the surveys. Since all data were derived from completed surveys, there was no variability, and the FA–SA correlation could not be calculated.

The FA–SA coefficients for resource-learning and sign-in were positive but not significant. For resource learning, the absence of a significant correlation may reflect the fact that students only needed to click the resources to receive credit for doing so, and thus the app did not assess the actual time and effort students put into learning. Moreover, the content of these resources was relatively broad and included material not directly related to the final exam.

As for sign-in score, because this was an optional course, the majority of the students chose it out of personal interest, resulting in a high attendance rate and therefore not much variability. Moreover, sign-in scores only showed whether students attended the class or not, and thus did not reflect student learning much.

Table 2. Descriptive Statistics for FA and Correlational Analyses for FA and SA

<table>
<thead>
<tr>
<th>Variables</th>
<th>FA Scores</th>
<th>Min</th>
<th>Median</th>
<th>Pearson Coefficient</th>
<th>( p ) value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resource learning</td>
<td>75</td>
<td>80</td>
<td>15.7</td>
<td>92.2</td>
<td>.071</td>
</tr>
<tr>
<td>Assignment</td>
<td>81</td>
<td>92.7</td>
<td>0</td>
<td>100</td>
<td>.256</td>
</tr>
<tr>
<td>Chapter test</td>
<td>70</td>
<td>69.7</td>
<td>30.3</td>
<td>96.9</td>
<td>.523</td>
</tr>
<tr>
<td>Sign-in</td>
<td>86.5</td>
<td>100</td>
<td>18.2</td>
<td>100</td>
<td>.181</td>
</tr>
<tr>
<td>Survey</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>–</td>
</tr>
<tr>
<td>Online discussion</td>
<td>77.2</td>
<td>85.9</td>
<td>0</td>
<td>88.5</td>
<td>.300</td>
</tr>
<tr>
<td>Brainstorming</td>
<td>70.5</td>
<td>85.7</td>
<td>0</td>
<td>85.7</td>
<td>.242</td>
</tr>
<tr>
<td>Quick-response</td>
<td>14.9</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>.321</td>
</tr>
<tr>
<td>Teacher’s thumbs-up</td>
<td>24.9</td>
<td>25</td>
<td>0</td>
<td>100</td>
<td>.344</td>
</tr>
</tbody>
</table>
4.4 Comparison of FA and SA Scores across Different Groups

Positive evaluation by teachers was indicated by their frequency in using the thumbs-up icon. In order to further test the influence of this FA, students were grouped according to whether they were ever affirmed by a positive thumbs-up click. An independent-sample t-test compared positive and negative evaluations of the students by their FA and SA scores. As can be seen in Table 3, the thumbs-up group outperformed the non-thumbs-up group by over 20 and over 26 points on both FA and SA scores (both \(p's < .01\)). These findings suggest that enhancement of teacher-student interaction, and student recognition by teachers, can better improve students’ scores in both FA and SA.

<table>
<thead>
<tr>
<th>Assessment</th>
<th>N</th>
<th>Thumbs-up</th>
<th>Non-thumbs-up</th>
<th>t</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formative Assessment</td>
<td>75</td>
<td>87.3</td>
<td>66.7</td>
<td>7.35*</td>
</tr>
<tr>
<td>Summative Assessment</td>
<td>75</td>
<td>89.2</td>
<td>63.2</td>
<td>6.37*</td>
</tr>
</tbody>
</table>

*\(p < .01\)

5. Conclusion & Implications

This paper intended to examine the relationship between formative assessment (FA), provided by the Mosoteach app, and summative assessment (SA), as reflected on a final exam and project. The goal of the study was to improve the assessment with a focus on improving college students’ performance for ethnic minorities in China. To accomplish this goal through the use of the Mosoteach app, two research questions were asked.

For research question 1, whether FA on the whole promotes SA, the results showed a positive correlation between FA and SA, which implied that teachers’ emphasis on students’ daily performance and progress during their learning process enhanced not only students’ behavioral outcome with respect to positive learning attitudes and engagement but also their academic outcomes in the end.

As for research question 2, the specific components of FA related to SA, scores on chapter tests, question-answering, brainstorming, online discussions, and the number of teacher’s thumbs-up clicks were positively correlated with the final exam and project scores, while relationships were not observed for resource learning and class attendance. These findings illustrated how specific components of FA contribute to students’ learning experiences. Some implications of these findings are as below.

Firstly, along with a positive overall impact of FA, individual components of FA can be distinguished, and many of these make a separate contribution to student outcomes. In order to support their learning in a whole and track individual student progress as well, teachers should clarify how to offer more effective formative feedback while using various FA components that the Mosoteach app already has, and explore how specific components of FA contribute to students’ learning experiences with teachers’ feedback concerning on students’ individual performance and learning demands.

Secondly, teachers should take advantage of the app’s data collection functionality, in order to better understand student progress and adjust teaching methods as needed. This promotes a more scientific approach to teaching, and more student-centered methods of instruction while utilizing educational technology like the Mosoteach app to develop FA and SA assessments.

Future research might consider several issues that the present study did not address, including the relative impact of different types and amounts of FA, the effectiveness of different
types of apps that provide FA, and individual differences between students in their responsiveness to formative feedback. Therefore, it can be expected that the implementation of FA and SA assessments using online apps will help teachers to improve not only their English teaching but also assessment skills.

REFERENCES


류 던 빙 (Dianping Liu)

- 1999 ~ present : College of Foreign Languages, Yanbian University, Associate Professor
- Interest : English education and intercultural communication
- E-Mail : liudianping@ybu.edu.cn

김 양 희 (Yang–Hee Kim)

- 2011년 2월 : 숭실대학교 문학 박사
- 2015년 3월 ~ 현재 : 백석대학교 어문학부 영어과 부교수
- 관심분야 : 영어교육, 언어학, 영어문학, 음악 활용 영어
- E-Mail : 214julie@bu.ac.kr