The Effect of Frequent Assessment on Language Development*

ผลของการประเมินแบบบ่อยครั้งต่อการพัฒนาทางภาษา

Thinley Wangdi**
Thanyapa Palanukulwong (ธัญภา พลานุกูลวงศ์)***

Abstract

Frequent assessment has been one of the most common teaching tools used in an educational field. However, its impact on learners' performance is unclear. This study aimed to investigate the impact of frequent assessment in the form of unit test on the midterm and final scores of the learners. Fifty Thai vocational students were recruited for the study. They were divided into two groups of 25 students each; the experimental and controlled group. A total of ten unit tests following each unit of the course were administered to the participants in the experimental group. Subsequently, feedback was given to the participants on tests' items. The controlled group neither received any unit tests nor the feedback. Then, both groups were administered the midterm and final tests. Independent sample t-test was used to compare the midterm and final mean scores of the two groups. The result of the study showed a significant increase in the midterm and final scores of the experimental group. The scores of the midterm and final tests of the two groups were significantly different at 0.05 level. The study also revealed a strong positive relationship between the experimental group's unit test scores and the midterm and final tests scores. In addition, there is a positive impact on the performance of students when the frequent assessment is conducted.

Keywords: formative assessment, summative assessment, frequent testing, academic performance.

* The article is conducted to investigate the effects of frequent assessment on language development and to fulfill the requirement for Master of Arts in Teaching English as an International Language, Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus.

** M.A student, Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus, E-mail:thinley11@gmail.com

*** Assoc. Prof.Dr. Thanyapa Palanukulwong, Faculty of Liberal Arts, Prince of Songkla University, Hatyai Campus, E-mail: thanyapa.c@psu.ac.th.

บทคัดย่อ

การประเมินแบบบ่อยครั้ง (Frequent Assessment) เป็นเครื่องมือหนึ่งที่ใช้บ่อยที่สุดในด้านการศึกษา แต่ผลกระทบต่อความสามารถของผู้เรียนยังคงมีความไม่แน่นอน งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาผลกระทบของการประเมินแบบบ่อยครั้ง (แบบทดสอบท้ายบทเรียน) ต่อผลการทดสอบกลางและการทดสอบปลายภาคของผู้เรียน ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ ประกอบด้วย นักเรียนไทยในระดับอาชีวศึกษา 50 คนแบ่งออกเป็นสองกลุ่มคือกลุ่มทดลองและกลุ่มควบคุมจำนวนกลุ่มละ 25 คน ทั้งสองกลุ่มมีความสาใรถด้านภาษาอังกฤษไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ก่อนการทดลอง กลุ่มทดลองทำแบบทดสอบของแต่ละหน่วยจำนวน 10 หน่วยของแต่ละหน่วยการเรียนรู้และครูแจ้งผลคะแนนและให้ข้อมูลย้อนกลับ (Feedback) การทดสอบของแต่ละครั้ง กลุ่มควบคุมไม่ได้รับการทดสอบของแต่ละหน่วยและผลคะแนนใดๆ จากนั้นทั้งสองกลุ่มทำแบบทดสอบระหว่างภาคและปลายภาค โดยสถิติที่ใช้ในการวิเคราะห์ข้อมูลได้แก่ ค่าสถิติ t-test ผ

ผลการศึกษาพบว่า

กลุ่มทดลองมีระดับคะแนนเฉลี่ยจากการทำแบบทดสอบระหว่างภาคและปลายภาคสูงขึ้นอย่างมีนัยสำคัญที่ระดับ .05 ซึ่งแสดงให้เห็นถึงความสัมพันธ์เชิงบวกระหว่างคะแนนการทดสอบท้ายบทของกลุ่มทดลองและคะแนนการทดสอบกลางภาคและปลายภาค โดยส่งผลในเชิงบวกต่อความสามารถของนักเรียนเมื่อทำการประเมินแบบบ่อยครั้ง นอกจากนี้ยังพบว่าการทดสอบบ่อยครั้งทำให้กลุ่มทดลองเข้าชั้นเรียนบ่อยกว่าและให้เวลากับการทบทวนบทเรียนมากกว่ากลุ่มควบคุม

คำสำคัญ: การประเมินผลระหว่างเรียน การประเมินผลสรุป การทดสอบแบบบ่อยครั้ง ผลการเรียน

Introduction

Conventionally, an assessment is defined as a process of evaluating students' work to help them pass the enrolled course (Taras, 2005). Assessment is often divided into a formative and summative assessment. Formative assessment is interchangeably used as an assessment *for* learning and summative assessment as an assessment *of* learning (DeLuca & Klinger, 2010). The use of an assessment has become the most common trends in an educational field to help students improve their learning outcome (Wiliam, Paul, & Black, 2011). In fact, an earlier researcher pointed out that 'a good teaching without a good assessment is a job only half done' (Maudsley, 1989). The statement appears to be a strong recommendation to the educators of all ages to have assessment included within the course to enhance the quality of teaching and learning.

According to Cilliers, Schuwirth, Adendorff, Herman and Vleuten (2010), assessment is an educational tool that is primarily designed to promote a meaningful learning. It is considered as one of the most influential tools in an educational field. They also highlight the impact of assessment on students learning process, believing that assessment provides extrinsic motivations and enables students to study more. A similar claim was made on students' change in efforts and dedications towards learning after the conduct of assessment (Van Etten, Freebern, & Pressley, 1997). Thus, the positive impacts of assessment on students and how it contributes to students learning process has emphasized in various quarters (Segers & Dochy, 2006).

Formative and Summative Assessment

Formative assessment is an evaluation technique used by teachers to evaluate teaching materials and students' learning progress, particularly focused on the improvement of the students learning outcome through a series of feedback (Andersson & Palm, 2017). However, the definition of formative assessment is still inconsistent. Some authors claim that formative assessment is a mere classroom assessment (Brookhart, 2001). Nonetheless, the use of formative assessment is always believed to play a vital role in enhancing the students learning ability since it enables students to evaluate their own progress of learning (McDowell, Wakelin, Montgomery, & King, 2011).

Formative assessment is beneficial in other aspects as well. It is believed that formative assessment helps students improve their cognitive intelligence which fills in as the main impetus to enhance their summative performance (Cauley & McMillan, 2010; Krasne, Wimmers, Relan, & Drake, 2006). Also, a wide range of previous studies have claimed that the use of any developmental activities (formative assessments) during the course significantly improve the students' final summative performance (e.g., Andersson & Palm , 2017; Nguyen & McDaniel, 2014; Krasne.et.al., 2006).

Most importantly, Hill, Guinea, and McCarthy (1994) emphasize the fact that students extensively feel the need of having a frequent formative assessment in the course. They reported that 89% of the students who were participating in their study acknowledged the advantage of frequent formative assessment because it enabled them to improve their learning strategy.

Over the last few decades, many studies have been done on the impact of formative assessment on students' final performance (Cauley & McMillan, 2010; Hill, Guinea, & McCarthy, 1994). For instance, William, Lee, Harrison, and Black (2004) investigated the effectiveness of the formative assessment on students' final performance. They claimed that the formative assessment during the course could improve the performance of students by 60 percent more than their actual ability. Moreover, Nguyen and McDaniel (2014) reported that frequent formative assessment helped students improve their final summative assessment. In addition, Butler and Roediger (2007) and Roediger and Louis (2014) also claimed that frequent assessment in the classroom could improve students' retention of the learned materials and their retention ability. Nevertheless, a recent study of Wiliam et al. (2011) showed that the degree of effectiveness of the formative assessment depended on the frequency of the intervention or assessment during the course.

On the other hand, *summative assessment* is defined as a process of summing up the records of students to learn his/her overall achievement of the course. It is basically carried out at certain intervals when students achievement has to be reported (Harlen & James, 1997). Summative assessment serves as a reliable evidence to the students' learning achievement since the students are assessed based on the common specified goals and criteria by the institutions (Knight, 2002). Unlike formative assessment which is intended to improve students learning, summative assessment is more of summing-up the scores for grading purposes (Brookhart, 2001).

However, it's not only assessment which helps students learn better. Feedback after the assessment is equally considered as an important tool in the framework of assessment. Feedback on assessment helps students identify learning errors and modify them accordingly (Cauley & McMillan, 2010). Feedback allows students to interpret their errors and bring change in their learning process (Blackman, 2012; Rushton, 2005). Particularly, immediate feedback after tests plays an important role in learning (Epstein & Brosvic, 2001). Immediate feedback on test materials is effective because it helps students acquire and retain tests materials in higher volume (Epstein et al., 2002). Feedback also help students understand and detect the error of the learned materials, which they can alter and learn to improve their final test performance (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991).

Frequent Testing

Frequent testing has been defined and interpreted in numerous ways by different scholars based on the conduct of the test and allotted time intervals for the tests (Gholami & Moghaddam, 2013). For instance, according to Kling, McCorkle, Miller and Reardon (2005), frequent testing is a test which is administered on a monthly basis, whereas earlier researchers have defined it as a routine examination done to assess students in a weekly (Keys, 1934) and daily basis (Dineen, Taylor & Stephens, 1989). The most common tools used in frequent assessment are short quizzes (e.g. multiple choice questions, gap-filling, and short answer questions). These tools are integrated into the courses to help students master the learned materials for the final examination (McDaniel, Anderson, Derbish & Morrisette, 2007).

Frequent testing has been the primary choice for the educators since it helps students know one's own learning progress (McDowell et al., 2011; Keys, 1934). Moreover, testing is a mandatory task in which each individual student must be interested in to qualify the enrolled course, thereby motivating students to learn more than the usual (Brown, 2005). Furthermore, the conduct of frequent testing is believed to improve students' learning consistency as well as students attendance to the class (Wilder, Flood, & Stromsnes, 2001).

In addition, McDaniel et al. (2007) reported that incorporating frequent tests in the course helps students remember the classroom materials in higher volume. Since testing allows students to have an additional exposure to the materials (Butler & Roediger, 2007). Frequent testing also encourages students to increase their frequency of study by making them revise materials periodically for the test (Karpicke & Roediger, 2007).

Leeming (2002) and Trumbo, Leiting, McDaniel, and Hodge (2016) assert that frequent testing not only helps students to improve their retention of materials but it also helps students to lower their testing anxiety by familiarizing them with both tested and non-tested content. The impact of test anxiety on students' learning is reported by Cassady and Johnson (2002). They claimed that the test anxiety has a negative impact on students' performance. Their finding shows that the students with higher level of anxiety scored less compared to those students with a moderate level of test anxiety.

On the flip side, Wooldridge, Bugg, McDaniel, and Liu (2014) expressed reservations about the idea of frequent testing and its positive impacts on higher retention of the classroom materials. They argued that the idea was not applicable unless the tests done during the course and the final test include the same items. Haberyan (2003) also recounted on having no significant difference between the group with frequent testing and a group without.

In addition, a meta-analysis study by Bangert-Drowns et al., (1991) on the effect of frequent testing in the classroom towards the students outcome showed that out of 35 studies taken for the analysis, 29 studies claimed that frequent testing relatively had a positive effect on students' outcome whereas the other 6 studies claimed the effects as negative or neutral.

Related Studies

There are many studies on the effects of formative assessment on students' performance. For instance, William et.al., (2004) investigated the effect of formative assessment practice in the classroom on learners' final achievement. The participants were 35 teachers from different schools in the UK; 21 teachers in the experimental group used formative classroom and 24 teachers in the controlled group did not. The teachers in the experimental group taught a total of 362 students and those in the controlled group taught 376 students. Both groups of teachers took classes for one year in their respective institutions. The result revealed that the students of the experimental groups performed better than the students in the controlled group. The researchers concluded that formative assessment, in general, helped learners to increase their final grades.

Similarly, Gholami and Moghaddam (2013) explored the impact of the frequent formative assessment on students' final achievement, 70 second-grade high school students participated in the study. The students were divided into two groups and were assessed differently. The experimental group received a weekly quiz test and the controlled group took the only summative assessment, the midterm examination. The result revealed that the scores of the group who took weekly quizzes were significantly higher than those who did not take quizzes.

In addition, Padilla-Walker (2006) examined the impact of daily extra credit quizzes on students' final performance. The participants were 36 undergraduate students of Midwestern State University taking advanced seminar course. The students met their teacher twice a week and were given extra credit quiz each day. The result of the study revealed that there was a significant improvement in the students' final performance. The finding also suggested that daily quiz increased students' regularity in learning classroom materials.

Roediger, Agarwal, McDaniel, and McDermott (2011) examined the effect of formative assessment on students' retention. A total of 142 sixth-standard students participated in the study. Students were divided into two groups. The experimental group received a series of classroom assessment (quizzes) and the other group received none. It was found that the students who received quizzes during the course performed relatively high in their final scores as well as in their retention test compared to the group with no quiz.

Furthermore, a positive acknowledgement from the participants was reported. Brookhart (2001) investigated the students' perception towards the formative and summative assessment practice in the classroom. Participants for this study were 990 high schools' students in the USA. Almost all participants were academically successful ones. The data were collected through interviews. The result revealed that most of the students felt the importance of assessment in learnings. Since it enabled them to revisit, master and learn the classroom materials in higher frequency. However, the finding could not reveal whether the students preferred formative or summative assessment. It suggested that assessment in general is effective in learning and for better learning outcome.

As discussed so far, although there are many previous studies on the effect of frequent testing on students' performance, the findings are still inconclusive. In addition, there was no study that looks at the interrelationship between frequent testing and students' final performance in Thailand.

Objectives

This research aimed to investigate the effects of using unit tests on students' midterm and final performance and to investigate the interrelationship between students' unit tests scores and final scores (midterm and final tests) on vocabulary and grammar.

Research Questions

This study was designed to answer the following research questions.

1. Is there any difference in performance in the midterm and final scores between the participants who receive frequent tests and those who do not?

2. Is there any relationship between the students' vocabulary and grammar's unit test scores with their midterm and the final scores?

Materials and Methods

1. Participants

A total of 50 second year certificate vocational students of two intact classes from Songkhla Vocational College in Thailand were recruited as participants for this study. The majority of the participants were females, with 42 females and 8 males of 15-17 years of age. All the participants were taking English course with a textbook entitled *English for Life*, consisting of 10 units in their second semester, 2017 academic years. The participants were assigned based on their English Grade Point Average (GPA) to the experimental and controlled group, with 25 students in each group. Both groups were comparable in terms of language proficiency as reflected by their average GPA ($\overline{X}$ = 2.54, S.D = 0.71) and ($\overline{X}$ = 2.56, S.D = 0.70) for the experimental and controlled group respectively.

2. Research Instruments

2.1. Unit Test

Ten unit tests used were based on the 10 units of the course textbook, *English for life* (Hutchinson, 2003). Each unit test consisted of 25 items, 15 vocabulary and 10 grammar items. All the items were in the form of gap-filling and matching. The tests were developed by the researcher and checked by three qualified research committee members to establish its validity. The test was administered to the test takers after the completion of each of the 10 units of the course, five unit tests before the midterm and five unit tests after the midterm exam.

2.2. Midterm and Final Test

The course's midterm and the final tests developed and administered by the institution were used to represent the participants' achievement on the course, their final performance on the course. The midterm test comprised units 1- 5 and the final test comprised of unit 6 – 10. Both the midterm and final tests had 40 items of various parts i.e. vocabulary, grammar and reading comprehension, in the form of multiple-choice and gap-

filling. Only the vocabulary and grammar parts were used in this study. The midterm test was administered in the middle of the semester after the completion of units 1 – 5 and the final tests towards the end of the semester after the completion of units 6 - 10.

3. Data Collection

The study was conducted in the second semester of 2017 academic year (October 2017- February 2018). The data were collected in the following steps.

1. Both the experimental and controlled groups were taught 5 units (units 1-5) by the class teacher, the researcher. Both groups received 2 hours' instructional time per week in different class settings. However, the type of assessment received by the two groups was different. The experimental group was given a 20-minute unit test after the completion of each unit of the course. The test participants were informed of their test scores and subsequently, feedback on the test items was given. The controlled group didn't receive any unit tests.

2. In December after the completion of unit 5, the midterm test was administered to both the experimental and controlled group. Also, the attendance of the participants in both groups was recorded for the whole semester.

3. Both groups were taught another 5 units (units 6 – 10) from the same course. The experimental group was administered a 20-minute unit test after each unit of the course. Then, towards the end of the semester, in February after the completion of unit 10, the final test was administered to both groups for the course evaluation.

4. Data Analysis

Data obtained from the unit tests and the midterm and final tests were analyzed and interpreted to answer each research question. A descriptive analysis such as mean and standard deviation was used to compare the midterm and final tests scores of both the experimental and the controlled group. Then, an independent sample t-test analysis was used to identify the difference in the midterm and the final scores of the two groups. Pearson correlation coefficient was also used to analyze the relationship between the unit test scores and the midterm and final tests scores.

Results

This section contains the results of the study, arranged based on the 2 research questions.

Research Question 1: Is there any difference in performance in the midterm and final scores between the participants who receive frequent tests and those who do not?

To answer the first research question, the participants' scores on the midterm and final tests were analyzed as shown in Table 1 below.

Table 1. Experimental and control groups' performance on midterm and final exam

Groups	Midterm Scores Total = 25 scores		Final Scores Total = 25 scores	
Experimental (n = 25)	$\overline{X}$	S. D	$\overline{X}$	S. D
	14.56	4.50	16.00	4.10
Controlled (n = 25)	11.88	3.74	13.42	3.70
Difference (D)	2.68*	0.76	2.58*	0.4

* significant at the 0.05 level

As reflected in Table 1, the difference between the average mean scores of both the midterm and final tests for the experimental group and controlled group was significantly different (D = 2.68 and D = 2.58, respectively). The participants' midterm scores analysis showed the average mean scores ($\overline{X}$ = 14.56, S.D = 4.50) and ($\overline{X}$ = 11.88, S.D = 3.74) for the experimental group and controlled group respectively. The experimental group who received frequent tests (unit tests) during the course performed significantly higher in the midterm test than the group with no unit tests. They scored 2.68 higher than the controlled group.

Similarly, the average mean scores in the final tests of the experimental group ($\overline{X}$ = 16.00, S.D = 4.10) was significantly higher than that of the controlled group ($\overline{X}$ = 13.42, S.D = 3.70). The average score of the experimental group was 2.58 higher than the controlled group. Interestingly, they performed consistently better in both the midterm and final examination.

Apart from helping participants to improve their academic performance, data collected on class attendance also showed that unit tests helped increase the class attendance of the participants in the experimental group. The average class attendance of the experimental group and the controlled group were 93.09% and 81.82% for the whole semester. In addition, the frequent conduct of tests enabled participants in the experimental group increase their study hours for the assigned course. The average time spent per week on the course reported by 25 participants in the experimental group was almost an hour more than the average time spent by the participants in the controlled group. The average time spent reported by the controlled group was low as they revised the materials only before the big tests (i.e. midterm and final tests).

Research Question 2: Is there any relationship between the students' vocabulary and grammar's unit test scores with their midterm and the final scores?

To investigate the relationship between the unit test scores and the midterm and final tests scores, the unit tests scores and the midterm and final tests scores of the 25 participants in the experimental group were taken for analysis. Table 2 A shows the relationship between the unit tests scores and the midterm test scores. Table 2 B shows the relationship between the unit tests scores and the final test scores.

Table 2 A. Intercorrelations between unit tests score and the midterm score

	Midterm
Unit Test	.781**

** significant at 0.01 level

Table 2 B. Intercorrelations between unit tests score and the final score

	Final
Unit Test	.731**

** significant at 0.01 level

As shown in table 2, there was a strong correlation between the unit tests scores and the midterm and final test scores. The intercorrelation analysis between unit tests scores and the midterm and final tests scores was **r** = .781 and **r** = .731 respectively ($p < 0.01$). Those who scored high in the unit tests would also score high in the midterm score and final score and vice versa.

Discussion/Conclusion

The findings of this study can be summarized based on the research questions.

1. The general findings of the current study suggest a positive impact of frequent assessment on participants' academic performance. The study revealed that the use of frequent tests in the course helped participants in the experimental group have better final academic performance. The participants in the experimental group outperformed the controlled group in both the midterm and final tests. The result was in line with Nguyen and McDaniel (2014); Gholami and Moghaddam (2013) and Roediger.et.al. (2011) who claimed that frequent quizzes helped students perform better than those who did not take quizzes.

In addition, the feedback on the test items the participants received after each unit tests seems to equally play an important role in participants learning. The finding was in consistent with some previous studies reporting that the feedback helped students identify the flaws in learning and overcoming them accordingly (Cauley & McMillan, 2010).

2. The result revealed a strong correlation between the unit tests scores and the midterm and final tests scores. The students who scored high on the unit tests consistently scored high on the midterm and final tests. This finding seems to suggest that the participants who performed well in frequent assessment could benefit more.

3. The frequent tests in the course helped participants increase the amount of time put in to study the target lessons. It was found that the group who received frequent testing studied almost an hour more than the controlled group per week of the semester. This could be due to the influence of the motivational force that the participants gained from frequent tests, especially the extrinsic motivation. The participants seemed to be motivated extrinsically after receiving a positive feedback and good grades in their tests. Correspondingly, it was pointed out by the Changlek & Palanukulwong (2015) that the psychological factors play an important role in language learning, especially motivation, be it intrinsic or extrinsic motivation. The result of the present study is in line with many studies such as Karpicke and Roediger (2007) and Wilder.et.al. (2001) who posited that frequent assessment increases students'

revision of the classroom materials. Furthermore, the current study also revealed that the frequent testing increased participants' class attendance. This finding was similar to that of Padilla-Walker (2006) who reported that the frequent quizzes in the course increased the students' regularity to the class.

In addition, the frequent tests helped participants evaluate their own learning progress; knowing their own learning progress seemed to help them adopt self-directed learning, particularly those participants whose prior test performance is unsatisfactory. The participants seemed to put more efforts to cope with the poor performance in their last test(s). These encouraged participants to immerse themselves in independent learning both inside and outside the classroom. This finding best fit with the claim made by Inthachot (2017) on having tremendous benefits of self-directed learning in the field of education since it provides additional space for learners to revisit the materials learned in the classroom.

The finding of this present study suggests that the integration of the frequent test in the course may be beneficial in EFL classroom to help improve the academic performance of the students. However, to confirm the findings of this research, further research is needed on the other aspects of language besides vocabulary and grammar, and research in different class settings before a conclusion can be drawn on the impact of frequent assessment.

References

Andersson, C., & Palm, T. (2017). The impact of formative assessment on student achievement: A study of the effects of changes to classroom practice after a comprehensive professional development programme. *Learning and Instruction*, *49*, 92–102.

Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, *61*(2), 213–238.

Bangert-Drowns, R. L., Kulik, J. A., & Kulik, C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, *85*(2), 89–99.

Blackman, A. (2012). The business review, cambridge * Vol. 20 * Num. 2 * The Immediate Feedback Assessment Technique (IF-AT): An innovative teaching technique for human resource management students.

Brookhart, S. M. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice*, *8*(2), 153–169.

Brown, S. (n.d.). Brown, Sally (2005) Assessment for Learning. Learning and teaching in higher education (1) assessment for learning fit-for-purpose assessment, *1*, 2004–5.

Butler, A. C., & Roediger, H. L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, *19*(4–5), 514–527.

Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, *27*(2), 270–295.

Cauley, K. M., & McMillan, J. H. (2010). Formative assessment techniques to support student motivation and achievement. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas*, *83*(1), 1–6.

Changlek, A., & Palanukulwong, T. (2015). Motivation and Grit_ Predictors of language learning achievement.

Cilliers, F. J., Schuwirth, L. W., Adendorff, H. J., Herman, N., & van der Vleuten, C. P. (2010). The mechanism of impact of summative assessment on medical students' learning. *Advances in Health Sciences Education*, *15*(5), 695–715.

DeLuca, C., & Klinger, D. A. (2010). Assessment literacy development: identifying gaps in teacher candidates' learning. *Assessment in Education: Principles, Policy & Practice*, *17*(4), 419–438.

Dineen, P., Taylor, J., & Stephens, L. (1989). The effect of testing frequency upon the achievement of students in high school mathematics course. *School Science and Mathematics*, *89*(3), 197–200.

Epstein, M. L., Epstein, B. B., & Brosvic, G. M. (2001). Immediate feedback during academic testing. *Psychological Reports*, *88*(3), 889–894.

Epstein, M. L., Lazarus, A. D., Calvano, T. B., Matthews, K. A., Hendel, R. A., Epstein, B. B., & Brosvic, G. M. (2002). Immediate feedback assessment technique promotes learning and corrects inaccurate first responses. *The Psychological Record*, *52*(2), 187–201.

Gholami, V., & Moghaddam, M. M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science*, *5*(1), 36–41.

Haberyan, K. A. (2003). Do weekly quizzes improve student performance on general biology exams? *The American Biology Teacher*, *65*(2), 110–114.

Harlen, W., & James, M. (1997). Assessment and Learning: differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, *4*(3), 365–379.

Hill, D. A., Guinea, A. I., & McCarthy, W. H. (1994). Formative assessment: a student perspective. *Medical Education*, *28*(5), 394–399.

Inthachot, M. (2017). Fostering self-directed learning ability of educational technology learners : insights from project-based learning implementation, *10*(5), 218–232.

Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*(2), 151–162.

Keys, N. (1934). The influence on learning and retention of weekly as opposed to monthly tests. *Journal of Educational Psychology*, *25*(6), 427–436.

Kling, N., McCorkle,D., Miller, C., & Reardon,J. (2005).The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business*, *81*(2), 67–72.

Knight, P. T. (2002). Summative assessment in higher education: practices in disarray. *Studies in Higher Education*, *27*(3), 275–286.

Krasne, S., Wimmers, P. F., Relan, A., & Drake, T. A. (2006). Differential effects of two types of formative assessment in predicting performance of first-year medical students. *Advances in Health Sciences Education*, *11*(2), 155–171.

Leeming, F. C. (2002). The Exam-A-Day procedure improves performance in psychology classes. *Teaching of Psychology*, *29*(3), 210–212.

Maudsley, R. F. (1989). Effective in-training evaluation. *Medical Teacher*, *11*(3–4), 285–290.

McDaniel, M. A., Anderson, J. L., Derbish, M. H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, *19*(4–5), 494–513.

McDowell, L., Wakelin, D., Montgomery, C., & King, S. (2011). Does assessment for learning make a difference? The development of a questionnaire to explore the student response. *Assessment & Evaluation in Higher Education*, *36*(7), 749–765.

Nguyen, K., & McDaniel, M. A. (2014). Using quizzing to assist student learning in the classroom: The good, the bad, and the ugly. *Teaching of Psychology*, *42*(1), 87–92.

Padilla-Walker, L. M. (2006). The impact of daily extra credit quizzes on exam performance. *Teaching of Psychology*, *33*(4), 236–239.

Roediger, H. L., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*(4), 382–395.

Roediger, H. L., & Louis, S. (2014). Test-enhanced learning, *17*(3), 249–255.

Rushton, A. (2005). Formative assessment: A key to deep learning? *Medical Teacher*, *27*(6), 509–513.

Segers, M., & Dochy, F. (2006). Introduction Enhancing student learning through assessment: Alignment between levels of assessment and different effects on learning. *Studies in Educational Evaluation*, *32*(3), 171–179.

Taras, M. (2005). Assessment – summative and formative – some theoretical reflectionS. *British Journal of Educational Studies*, *53*(4), 466–478.

Trumbo, M. C., Leiting, K. A., McDaniel, M. A., & Hodge, G. K. (2016). Effects of reinforcement on test-enhanced learning in a large, diverse introductory college psychology course. *Journal of Experimental Psychology: Applied*, *22*(2), 148–160.

Van Etten, S., Freebern, G., & Pressley, M. (1997). College students' beliefs about exam preparation. *Contemporary Educational Psychology*, *22*(2), 192–212.

Wilder, D. A., Flood, W. A., & Stromsnes, W. (2001). The use of random extra credit quizzes to increase student attendance. *Journal of Instructional Psychology*, *28*(2), 117–117.

Wiliam, C. L., Paul, C. H. &, & Black. (2011). What is assessment for learning? *Studies in Educational Evaluation*, *37*(1), 3–14.

William, D., Lee, Harrison, C., & Black, P. (2004). Teachers developing assessment for learning: impact on student achievement. *Assessment in Education: Principles, Policy & Practice*, *11*(1), 49–65.

Wooldridge, C. L., Bugg, J. M., McDaniel, M. A., & Liu, Y. (2014). The testing effect with authentic educational materials: A cautionary note. *Journal of Applied Research in Memory and Cognition*, *3*(3), 214–221.