DOI: 10.52950/TE.2021.9.2.004

HOW FORMATIVE ARE ASSESSMENTS FOR LEARNING ACTIVITIES TOWARDS SUMMATIVE ASSESSMENT?

TILO LI, MARINE YEUNG, EUGENE LI, BEETO LEUNG

Abstract:

To academics of higher education institutes, motiving, enhancing and improving student learning have never been easy. Different types of course work or continuous assessment or assessment for learning activities are designed to lead students to achieve the intended learning outcomes of respective courses. Depending on the nature of courses, these activities vary and they can either be done individually or in groups. The question is whether these activities can achieve their stated goals or if they can, how much these activities can lead students to achieve the intended learning outcomes as reflected by a summative assessment, normally it is a timed final exam or a take-home final essay. The purpose of this paper is to investigate the correlations between formative assessment activities and the summative assessment. Data of an introductory statistics course of different cohorts of a private degree-granting institute in Hong Kong were used in this study. The results indicate that individual assignments have a strong relationship with the exam score, while group projects do not. More surprisingly, group projects are negatively related to exam scores. The implication is that academics should rethink how to make group work a better and fairer assessment for individual student's contribution.

Keywords:

Higher Education, Analysis of Education, Formative Assessment, Continuous Assessment, Assessment for Learning, Summative Assessment

JEL Classification: I23, I21

Authors:

TILO LI, Tung Wah College, Hong Kong, Email: tiloli@twc.edu.hk
MARINE YEUNG, Tung Wah College, Hong Kong, Email: marineyeung@twc.edu.hk
EUGENE LI, The Hong Kong University of Science and Technology, Hong Kong, Email: lceugene@ust.hk
BEETO LEUNG, Tung Wah College, Hong Kong, Email: beetoleung@twc.edu.hk

Citation:

TILO LI, MARINE YEUNG, EUGENE LI, BEETO LEUNG (2021). How formative are assessments for learning activities towards summative assessment? . International Journal of Teaching and Education, Vol. IX(2), pp. 42-57., 10.52950/TE.2021.9.2.004

Introduction

Assessment is an integral part of teaching and learning. Depending on their purposes, assessment activities are categorized into two types: formative and summative assessment. Formative assessment is also called assessment for learning ("AfL"). This refers to learning activities that primarily promote student learning (Black et al., 2004) and are conducted continuously throughout a course of study. They are therefore also called continuous assessments. To make these activities formative, students need to be aware of the gap in their current capabilities relative to the desired goal and then take action to close the gap (Brown, 2019). Such awareness can be facilitated with feedback, most importantly feedback from teachers. Feedback should provide students with qualitative insights about their own understanding and misconceptions in order to help them improve their learning (Shepard, 2005).

For AfL to be effective, related assessment activities should be placed at the center of course design. Teachers and students should be responsible partners. Teachers need to provide students with clear instructions and criteria for attainment beforehand and useful timely feedback afterwards. Students should use feedback to improve their learning (Boud D. and Associates, 2010). The terms “AfL”, “formative assessment” and “continuous assessment” are used interchangeably in this paper. “Course work” may sometimes be used to refer to the same concept in context.

The other type of assessment, summative assessment, is also called assessment of learning. As the name implies, it refers to the final assessment activities that sum up student learning, such as final projects, final essays, or final examinations. They should be designed to assess students' overall achievement of the course intended learning outcomes (“CILOs”) and accordingly carry the most weight towards the final grade awarded to students.

Assessment for learning activities are meant to motivate, enhance, and improve student learning leading to the achievement of CILOs. In theory, AfL activities are conducted throughout the course of study. These activities should contribute to the overall achievement of learning as reflected by the summative assessment, typically a final examination for lower division undergraduate courses. As both kinds of assessment are aligned to the CILOs, with AfL contributing to the performance in the summative assessment, there should exist a correlation between the two types of assessment. However, previous studies have different views on the effectiveness of formative assessment in improving performance in summative assessment. While some studies provide evidence supporting this assumption in certain contexts (e.g. Carpenter et al., 2017; Coutu et al., 2019; Mitra & Barua, 2015), others fail to establish a significant correlation between formative and summative assessment (e.g. Anziani, 2008) or between a certain type of continuous assessment and the summative assessment (e.g. Bazelais et al., 2017). Formative assessments have been shown not to improve overall learning outcomes in some studies (Brothen & Wambach, 2001; Haberyan, 2003); while some studies yield affirmative findings (Buchanan, 2000; Kibble, 2007; Olson & McDonald, 2004), others show the reverse, indicating that students who have lower formative assessment marks tend to have higher summative assessment marks and vice versa (Ekolu, 2006). The inconclusive findings suggest that the predictive power of continuous assessment in relation to the final grade a student obtains in a course may be affected by the context, such as the nature of the course and the types of AfL activities adopted. In order to gain a better understanding of whether or how various kinds of course work contribute to the final grade, a study was conducted

in a tertiary institution in Hong Kong. The study set out to answer the research questions given below.

1. What are the correlations between various kinds of formative assessment activities and the summative assessment?
2. How much do formative assessment activities contribute to summative assessment?

Assessments Tools:

Assessment tools include, among others, assignments, in-class participation, learning portfolios, journal, practicum, case study/projects, group work and presentation, debate/forum, presentation, essay/report, quiz/test/exam (Hounsell & Zou, 2017). Most of these are formative assessment tools except for the final exam. Quizzes and mid-term tests are both formative and summative. They are formative because they inform both students and instructors the level of learning achievement and the gap between expectation and performance. They are summative because they sum up learning achievement from the last quiz/test to the point where the quiz/test is taken. Dalfen, Fienup and Sturmey (2018) find that more announced short quizzes, but not pop quizzes, lead to higher exam scores, especially for high- and mid-performing students. Furthermore, graded quizzes help students remember concepts better. Non-threatening quizzes with proper feedback enhance student learning and improve summative assessment results (Zhang & Henderson, 2015)

For formative assessment activities to be effective, Carless (2017a) suggests some main strategies for AfL activities as follows:

1. Assessment tasks have to be productively designed and aligned with the CILOs.
2. Assessment criteria must be clearly stated and analyzed so that students understand better the nature of quality.
3. Feedback needs to be given timely and comprehensively.
4. Peer assessment allows students to practice making judgements.

For designing assessment tasks, among higher education institutions in Hong Kong, most courses come with their CILOs. Course designers need to plan course activities that lead students to achieve these CILOs. These activities include lectures and various continuous assessment activities mentioned earlier. They also form part of the pedagogy (Brown, 2019; Carless, 2017a; DeLuca et al. 2018). These activities are assessable and allocated with marks toward the final grade. Without mark allocation, students will not be driven to devote efforts to these activities that will become useless (Jessop, 2017).

Feedback serves two purposes: to support and assess student learning. For feedback to be supportive and effective, it has to be given to students timely and comprehensively. Feedback needs to explain student's performance against criteria and identify areas for improvement (Rowe, 2017). Academics should assess what students have achieved and clearly identify the gap between where they are and where they should be, and indicate to students how they can reach the level where they should be. To support and improve students' learning is the primary purpose of providing feedback. If feedback loses this function, it will be useless. However, whether feedback

to students can improve learning is unsure. Some students do read comments from teachers and improve their future work, but some do not. Pitt (2017) finds that a majority of students who achieve a bad grade and receive negative feedback do not utilize feedback and apply to their future assessment. Feedback can also be interpreted differently due to cultural differences (Zhang & Zheng, 2018).

For continuous assessment to be formative, academics have to return the assessed and commented work to students timely. Unfortunately, at times, work is returned to students towards the end of a semester. Feedback is given to students too late, and students do not benefit from it. As a result, the formative function of AfL activities is lost. As Jessop (2017) indicates feedback given at the end of a semester does not appear as helpful in students' learning because they cannot be bothered to do so while the exam (summative assessment) is around the corner. In other situations, if students find they cannot feedforward (that is to say, feedback cannot be applied to future assignments), they do not bother reading feedback either (Deeley, 2018; Zhang & Zheng, 2018). As a result, a lot of written feedback from teachers is therefore wasted.

Rubrics or standard descriptors (Bridges et al., 2017) make teachers' expectations and criteria explicit. Rubrics state clearly the quality standard of expected achievement. They also make teachers' grading more consistent. Students are strategic performers, and they focus on achieving grades (Jessop, 2017). So, if rubrics are carefully designed and clearly listed out the expected quality standard in line with the CILOs, students will work towards them to achieve higher grades. As a result, student learning can be enhanced, and CILOs can be met.

Academics normally provide students with assessment rubrics or grading criteria for written essays and project assignments. If students are given rubrics and exemplars, they will understand better how quality is achieved (Brown, 2019). As a result, they should do better. However, there is concern about considering the exemplar as a model answer to be imitated. Students will produce a similar output (Carless, 2017b).

Students learn more if they are given the chance to assess their peers using the rubrics (Jonsson & Panadero, 2017). Some students find it stressful and time-consuming to assess their peers. Students benefit from the process because they have to understand better the assessment rubrics to rate their peers. Students learn from other students' mistakes. Moreover, their critical thinking skills are trained. However, students concern about unfair assessment from their fellow students who may be friends or rivals (Li, 2014). As for feedback, students prefer feedback from teachers more than their peers (DeLuca, et al., 2018).

Effects of different types of assessment on learning effectiveness

In general, students' perceptions of the assessment tools that they are tackling affect their learning behavior (Carless et al., 2017). They like assessments that they can relate to their daily life, have flexibility and choice, understand what are expected, and receive productive feedback during the process (Carless, 2017b). If students find feedback will not help their future assessment, they probably will not be bothered (Jessop, 2017).

Academics normally assign exercises for students to train their basic skills, especially for quantitative and writing courses. Relatively weaker students like to work on assignments because they are more student-centered. Students can control their learning, discuss with their peers and ask their teachers. Students find they learn more from working on assignments so that they are more motivated and build confidence (Hernandez-Martinez, et al., 2011). In an earlier study, Conway (1992) finds that student grades of course-work better reflect students' learning because knowledge gained from course-work can be retained for a much longer time. Furthermore, Archer and Olson (2018) find that assignment is correlated with increased student learning as reflected by students' exam scores.

Carless (2017b) finds that there are mixed feelings about collectively assessed group work. Students like group work because it reflects the real-life workplace situation that people need to work with others. Students can learn from and support each other. They can also develop their teamwork and interpersonal communication skills. Meanwhile, some students do not like group work because there are free riders. Other members need to do the work left by free riders. The grade of good-performers may be adversely affected.

A typical summative assessment is a timed final examination. Unfortunately, research has found that the final examination does not necessarily correlate with students' learning (Korpi, 2019). Low-stakes continuous assessments, for example, small assignments engage and enhance student learning, but high-stakes exams do not. Examination grades only reflect knowledge retention in the short-term (Conway, 1992). It may be argued that as a form of summative assessment, examinations are mainly used to assess learning progress rather than foster learning; to that end, examinations should bear a close relationship with the continuous assessment component of the course.

It is common for academics of higher education to award final grades to students based on weighted-average marks from all assessment activities. Unfortunately, the meaning of a final mark in terms of the CILOs is ambiguous. For example, a passing mark or a D grade is set at around 50 percent. Boud (2017) questions the meaning of a passing mark. Given a student has passed a course with 50 percent, has the student achieved 50 percent (for example, 5 out of 10) of the CILOs? Has the student achieved 50 percent of each of the CILOs? On what basis can marks be related to CILOs? Averaging across different CILOs is meaningless. After all, to pass a course with 50 or 60 percent weighted-average marks, it is possible that students over-achieve in some areas without any achievement in other areas.

Data Collection and Methodology

In view of students' mixed feelings about various forms of formative assessment, particularly about group work, the continuous assessment tools used in this study were classified into two types for comparison. The first type included individually assessed tasks. They were written assignments (10 per cent of the final grade) and a 90-minute mid-term test (20 per cent of the final grade). The second type was a collectively assessed group project (20 per cent of the final grade). All these tools were allocated some credits toward the final grade to motivate students to devote efforts to them. All continuous assessments accounted for 50 per cent of the final grade. The summative

assessment was a three-hour final exam accounting for 50 per cent of the final grade. The overall assessment consisted of 80 per cent individual and 20 per cent group. Informed by the literature on the design and use of continuous assessment, the continuous assessment activities incorporated were carefully designed to align with the CILOs with clear rubrics to guide students' learning. Feedback was also provided for students in a timely manner. These helped to optimize the effects of the AfLs and enhance the validity of the results.

Data for this study were collected from the instructor's records of an introductory statistics course for business students taught at a private degree-granting institute in Hong Kong. The participants consisted of 214 students from 9 cohorts. As the breakdown of the continuous assessment marks for some cohorts was unavailable, some parts of the analysis consisted of less than 214 students.

Since the mid-term test was both summative and formative, a simple linear regression was conducted using SPSS taking the mid-term test as the dependent variable and individual written assignments as the independent variable represented by equation (1) below:

$$\text{Test Score} = \text{Constant} + \beta^{*}\text{Assignment} \quad \ldots\ldots(1)$$

Multiple linear regression and correlation analysis were then conducted. The dependent variable was the final exam (the summative assessment), and the independent variables were different continuous assessment or AfL activities including the assignments (individual), mid-term test (individual) and written report (group). This relationship can be represented by equation (2) below:

$$\text{Exam score} = \text{Constant} + \beta 1^{*}\text{Assignment} + \beta 2^{*}\text{Term Test} + \beta 3^{*}\text{Group Project} \ldots\ldots(2)$$

For linear regression analysis to be valid, data must conform to linearity, normality and homoscedasticity. Multicollinearity of independent variables is needed to be checked for multiple regression but not simple regression (Aldrich, 2019).

For linearity and normality, the SPSS residuals scatterplots should have a straight-line relationship with the dependent variable shown in the Normal Probability Plot (P-P) of the Regression Standardized Residual. The Normal P-P is a straight diagonal line (see Appendix 1). If the Observed Cum(mulative) Prob(ability) and the Expected Cum(mulative) Prob(ability) dance around the diagonal line, linearity and normality are confirmed (Pallant, 2016). Normality can also be examined by looking at the Histogram of the Regression Standardized Residuals (see Appendix 2).

Homoscedasticity means variances of residuals should be the same for all predicted values. This can be verified by the Scatterplot diagram (see Appendix 3) of the Regression Standardized Predicted Value (on the X-axis) and Regression Standardized Residual (on the Y-axis). If the residuals are randomly distributed with no pattern, and most of the values are centered around zero, homoscedasticity is confirmed (Aldrich, 2019).

Multicollinearity exists when independent variables are highly correlated with correlation coefficient 0.7 and above (Lind, Marchal, & Wathen, 2018). Some even suggest a correlation coefficient of 0.9 and above (Pallant, 2016). If two independent variables are too highly correlated, it may distort their prediction against the dependent variable. Should this be the case, these independent variables should be combined into one variable.

In addition to the above, data distribution was examined using SPSS Skewness Statistic and Kurtosis Statistics. If the Skewness Statistic is negative (i.e. negatively skewed), it means that the mode is greater than the median, and the median is greater than the mean. If the Kurtosis Statistic is negative, it means that the normal distribution bell curve is relatively flat indicating a lot of values lie in the tails. There is a wider spread of the data. On the contrary, if the Kurtosis Statistic is positive, it means most values are clustered around the center with a few on the tails (Aldrich, 2019).

Results and Discussion:

To check for linearity and normality, the Normal Probability Plot (P-P) of the Regression Standardized Residual was created (Appendix 1). It is obvious that all residual values “dance” around the diagonal line. The residuals appear to have a linear relationship, and they do not deviate much from the diagonal line. Linearity and normality requirements are met (Aldrich, 2019; Pallant, 2016). Furthermore, as shown in the Histogram of the dependent variable Exam score (Appendix 2), it is obvious that the distribution is fairly normal.

For Homoscedasticity, the Scatterplot using Exam score as the dependent variable as shown in Appendix 3 shows no pattern of the values. They are randomly dispersed. As a result, an equal variance can be assumed (Aldrich, 2019). It is common to observe a few outliners in large samples, but the equal variance assumption should not be violated (Pallant, 2016).

To check for the existence of multicollinearity, the Pearson Correlation table among all independent variables is presented in Table 1. Although all the correlations are statistically significant at the 0.01 level (2-tailed), none of the correlations is over 0.7. Multicollinearity is therefore not a concern for multiple linear regression to be applied to the data set (Lind et al., 2018).

Table1: Correlations among independent variables

	Assignment	Term Test	Group Project
Assignment	1.000	0.526	0.507
Term Test	0.526	1.000	0.387
Group Project	0.507	0.387	1.000
N	118	118	118

Source: SPSS output based on author’s raw data

As for the sample data distribution, SPSS Skewness and Kurtosis Statistics were run. The SPSS Skewness Statistic had the following results. All the variables Assignment (N = 198; SS = -1.020), Term Test (N = 198; SS = -0.399), Group Project (N = 118; SS = -0.985), and Exam (N = 214; SS = -0.327) have negative Skewness Statistic indicating the distribution is negatively skewed, meaning the concentration of values at the high end of the horizontal axis. This means that the mode is greater than the median, and the median is greater than the mean. This can be verified by checking the descriptive statistics. For example, for Assignment (N = 198), the mode was 94.44; the median was 76.67; the mean was 73.15. For Term Test (N = 198), the mode was 62; the median was 62; the mean was 59.86. For Group Project (N = 118), the mode was 69; the median was 62; the mean was 59.35.

For the Kurtosis Statistics, both the individual timed test (N = 198; KS = -0.238) and exam (N = 214; KS = -0.283) have negative values indicating a large proportion of cases are in the tails. There is a wide spread of the scores. There are students who have very high and very low scores. However, Assignment (N = 198; KS = 0.655) and Group Project (N = 118; KS = 2.871) have positive values indicating a lot of cases are clustered in the center (Aldrich, 2019). This makes sense because students could discuss with other students in completing the assignments and group project. For individual assignments, especially for a quantitative course such as introductory statistics, there is always one correct answer to a question. If students get the correct answer, their answers are all the same. Some students work on the correct answer by themselves, but some "work together" with others to get the correct answer. It is therefore neither uncommon nor surprising to find some similar or identical mistakes. As a result, assignment and group project marks often lean on the high side.

To answer the first research question concerning the correlations of formative assessment activities and summative assessment, the first correlation to be discussed is that between Assignment and Term Test taken in Week 8 (for a 14-week semester). It is summative in the interim of a semester. For a sample of 198 (N = 198), the correlation coefficient is 0.526 (r = 0.526). The correlation is significant at the 0.01 level (2-tailed). Although the correlation cannot be said as strong, the correlation explains roughly 28 percent ($r^2 = 0.28$) of the variations in the Term Test scores.

Taking the final exam as the summative assessment and all continuous assessments as formative assessment, the paired-sample correlations are presented in Table 2.

Table 2: Correlations of formative and summative assessments

		Assignment	Term Test	Group Project	Weighted CA
Exam	Correlation	0.617	0.719	0.295	0.686
	Sig (2-tailed)	0.000	0.000	0.001	0.000
	N	198	198	118	214

Source: SPSS output based on author's raw data

It is interesting to note that the correlation between Assignment and Exam has increased to 0.617 (r = 0.617) versus the former's correlation with Term Test (r = 0.526). Although the correlation still cannot be said as strong, it has become stronger. Individual assignments, taken separately from other independent variables, now explain 38 percent ($r^2 = 0.38$) of the variation in the exam score, compared with 28 percent in the term test score. This echoes previous research that students learn more from working on assignments (Hernandez-Martinez et al., 2011), knowledge gained from course-work can be retained for a much longer time (Conway, 1992), and assignments are correlated with increased student learning as reflected by their exam scores (Archer & Olson, 2018).

The interim summative assessment term test has become a formative assessment relative to the final exam. Its correlation with the final exam is strong at 0.719. Any correlation coefficient above 0.70 is said to be strong (Lind et al., 2018). The term test explains roughly 52 percent ($r^2 = 0.52$) of the variation in the exam score. This makes sense because both of them are timed and individual assessments.

The group project has a weak correlation (r = 0.295) with the exam. It only explains 8.7 ($r^2 = 0.087$) of the variations in the exam score. The weighted continuous assessment total (Weighted CA) is quite strongly correlated (r = 0.686) with the exam score. All continuous assessments collectively account for about 47 percent ($r^2 = 0.47$) of the variation in the exam score. While there are correlations of varying levels of strength between the individual formative assessments and the summative exam, all correlations are significant at 0.01 level (2-tailed).

Paired Samples Test was conducted between Term Test and Exam. The mean of the term test was 59.864, and the mean of the exam was 64.172. The difference is 4.308, and it is significant at the 0.01 level (2-tailed). This might indicate that students had learned from their mistakes in the term test, thereby reflecting the nature of the term test as an effective AfL activity and providing evidence that graded tests help students learn better (Zhang & Henderson, 2015).

The weighted total of continuous assessment can be a good gauge of the final exam performance. The Paired Samples Statistics show that the Weighted CA mean was 63.876, and the mean Exam score was 63.136. The mean difference was 0.740, but it was insignificant (p = 0.435, 2-tailed). Therefore, the means of the two are believed to be the same.

To answer the second research question concerning how much formative assessment activities contribute to summative assessment, a simple linear regression was conducted using Term Test as the dependent variable and Assignment as the individual variable. The result is given in equation (3):

Test Score = 24.768 + 0.480*Assignment ……………………………………(3)

Both the constant (24.768) and the Beta (0.480) coefficient are significant at 0.01 level. Equation (3) suggests that students would earn 24.768 marks even they received “0” from assignments. As their assignment mark increased by the “1” mark, the term test score would increase by 0.48 mark. For this introductory statistics course, the term test contained some multiple guess, simple and straight-forward questions. This could explain the relatively high constant term.

To further investigate the impacts of other formative assessments on the summative assessment, multiple regression tests were run and the results are presented in Table 3 below.

Table 3: Formative assessments are regressed against exam (all CAs)

	Unstandardized Beta	t	Sig
Constant	15.760	2.992	0.003
Assignment	0.341	5.151	0.000
Term Test	0.519	6.861	0.000
Group Project	-0.112	-1.411	0.161

Source: SPSS output based on author’s raw data

The multiple regression model is, therefore:

Exam score = 15.760 + 0.341*Assignment + 0.519*Term Test – 0.112*Group Project……(4)

Equation (4) suggests that students would earn 15.76 marks even they earned “0s” from all other formative assessment activities. For each mark earned from assignment and term test, students would earn 0.341 and 0.519 marks respectively in the final exam. It is surprising to find that Group Project has a negative coefficient (Beta = - 0.112) although insignificant (p = 0.161), indicating group project results might have a negative effect on the exam score, or they were negatively related. It means that the higher the group project score, the lower the exam score, and vice versa. This perhaps can be explained by the free-rider phenomenon, and grades of good performers can be adversely affected as suggested by Carless (2017b).

The collective relationship of the three formative (or continuous) assessment activities and the exam score are strongly correlated (r = 0.733). They collectively accounted for 52.5 percent (adjusted r^2 = 0.525) of the variation in the exam score.

Another multiple regression was run taking away Group Project because it had a negative relationship with the exam score. The results are shown in Table 4 below:

Table 4: Formative assessments are regressed against exam (without Group Project)

	Unstandardized Beta	t	Sig
Constant	9.119	2.652	0.009
Assignment	0.303	6.166	0.000
Term Test	0.550	10.207	0.000

Source: SPSS output based on author’s raw data

The multiple regression model is presented in Equation (5).

Exam score = 9.119 + 0.303*Assignment + 0.550*Term Test……..(5)

Comparing Equations (4) and (5), it is interesting to note that the constant term has decreased from 15.76 to 9.119 (a decrease of 42.14 percent) without the group project. The free-rider effect can be quantified as 6.641 marks. The effect of assignment marks on the exam score was weakened from 0.341 to 0.303 (a decrease of 11.14%), but the effect of the term test on the exam score was strengthened from 0.519 to 0.550 (an increase of 5.97 percent). The correlation (r = 0.772) between these two formative assessment activities (the assignment and the term test) and the final exam score became stronger. They together accounted for 59.2 percent (adjusted r^2 = 0.592) of the variation in the exam score. This could reflect the result of removing the possible free-rider effect of the group project.

Conclusion

Although formative and continuous assessments serve different purposes, they are supplementary and complementary in nature and should bear a close relationship in theory and practice. As research findings on the correlation between the two types of assessment so far have been rather conflicting and contextualized, this study set out to provide further evidence in the Hong Kong context. Specifically, this study investigated the possible variance of different kinds of continuous assessment in explaining or predicting student formative assessments such as individual assignments and term tests are effective in preparing students for the final exam. However, while

group assessments such as group projects may help low-performance students obtain higher marks, they might have an adverse effect on the relative performance of stronger students. It is therefore suggested that academics find ways to assess individual students' contribution to group assignments and award marks accordingly. In addition to demonstrating an overall positive correlation between formative assessment and the examination, this study highlights the need to factor in the different effects various kinds of formative assessment may have on students' overall performance in a course. This may have important implications for practice and research.

References

Anziani, H., Durham, J., & Moore, U. (2008). The relationship between formative and summative assessment of undergraduates in oral surgery. *European Journal of Dental Education*, *12*(4), 233-238. Aldrich, J. (2019). *Using IBM SPSS Statistics: An Interactive Hands-on Approach.* SAGE Publications, Inc. https://doi.org/10.1111/j.1600-0579.2008.00524.x

Archer, K., & Olson, M. (2018). Practice. Practice. Practice. Do homework management systems work? *International Journal for the Scholarship of Teaching and Learning, 12(2)*, Article 12.

Bazelais, P., Doleck, T., & Lemay, D. J. (2017). Exploring the Association Between Formative and Summative Assessments in a Pre-University Science Program. *Journal of Formative Design in Learning*, *1*(2), 65-72.

Black, P., Harrison, C., Lee, C., Marshall, B., & William, D. (2004). Working inside the black box: Assessment for learning in the classroom. *Phi Delta Kappan, 86(1)*, 8-21.

Boud, D. (2017). Standards-based assessment for an era of increasing transparency. *Scaling up Assessment for Learning in Higher Education* (pp. 19-31). Hong Kong: Springer.

Boud, D. and Associate. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education.* Sydney: Australian Learning and Teaching Council.

Bridges, S., Wyatt-Smith, C., & Botelho, M. (2017). Clinical assessment judgements and 'Connoisseurship': Surfacing curriculum-wide standards through transdisciplinary dialogue. *Scaling up Assessment for Learning in Higher Education* (pp. 81-98). Hong Kong: Springer.

Brothen, T., & Wambach, C. (2001). Effective student use of computerized quizzes. *Teaching of Psychology, 28(4)*, 292-294.

Brown, G. (2019). Is assessment for learning really assessment? *Frontier in Education, V(4)*, 1-7.

Buchanan, T. (2000). The efficacy of a World-Wide Web mediated formative assessment. *Journal of Computer Assisted Learning, 16*, 193-200.

Carless, D. (2017a). Scaling up assessment for learning: Progress and prospects. *Scaling up Assessment for Learning in Higher Education* (pp. 3-17). Hong Kong: Springer.

Carless, D. (2017b). Students' experiences of assessment for learning. *Scaling up Assessment for Learning in Higher Education* (pp. 113-126). Hong Kong: Springer.

Carless, D., Bridges, S., Chan, C., & Glofcheski, R. (2017). *Scaling up Assessment for Learning in Higher Education.* Springer .

Carpenter, S. K., Rahman, S., Lund, T. J., Armstrong, P. I., Lamm, M. H., Reason, R. D., & Coffman, C. R. (2017). Students' use of optional online reviews and its relationship to summative assessment outcomes in introductory biology. *CBE—Life Sciences Education*, *16*(2), ar23.

Conway, M. (1992). Why is it that university grades no not predict very-long-term retention? *Journal of Experimental Psychology: General, 121(3)*, 382-384.

Couto, L. B., Durand, M. T., Wolff, A. C., Restini, C. B., Faria Jr, M., Romão, G. S., & Bestetti, R. B. (2019). Formative assessment scores in tutorial sessions correlates with OSCE and progress testing scores in a PBL medical curriculum. *Medical education online*, *24*(1), 1560862.

Dalfen, S., Fienup, D., & Sturmey, P. (2018). Effects of a contingency for quiz accuracy on exam scores. *Bahavior Analysis in Practice, V11*, 106-113.

Deeley, S. (2018). Using technology to facilitate effective assessment for learning and feedback in higher education. *Assessment & Evaluation in Higher Education, 43(3)*, 439-448.

DeLuca, C., Chapman-Chin, A., LaPointe-McEwan, D., & Klinger, D. (2018). Student perspectives on assessment for learning. *The Curriculum Journal, 29(1)*, 77-94.

Ekolu, S. (2006). Correlation between formative and summative assessment results in engineering studies. *the 6th African Engineering Education Association Conference, CUT, FS.*

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

Hernandez-Martinez, P., Williams, J., Black, L., Davis, P., Pampaka, M., & Wake, G. (2011). Mathematics coursework as facilitator of formative assessment, student-centred activity and understanding. *Research in Mathematics Education, 13(2)*, 197-212.

Hounsell, D., & Zou, T. (2017). Surfacing and sharing advances in assessment: A communities-of-practice approach. *Scaling up Assessment for Learning in Higher Education* (pp. 33-48). Hong Kong: Springer.

Jessop, T. (2017). Inspiring transformation through TESTA's programme approach. *Scaling up Assessment for Learning in Higher Education* (pp. 49-64). Hong Kong: Springer.

Jonsson, A., & Panadero, E. (2017). The use and design of rubrics to support assessment for learning. *Scaling up Assessment for Learning in Higher Education* (pp. 99-111). Hong Kong: Springer.

Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Psysiology Education, 31*, 253-260.

Korpi, S. (2019). Portfolio project as summative language assessment: Engaging learners online. *International Journal of E-Learning & Distance Education, 34(2)*, 1-18.

Li, T. (2014). A case study of peer-rater differences in Hong Kong. *International Journal of Teaching and Education, 2(3)*, 112-126.

Lind, D., Marchal, W., & Wathen, S. (2018). *Statistical Techniques in Business & Economics, 17th Edition.* McGraw-Hill Education International Edition.

Mitra, N. K., & Barua, A. (2015). Effect of online formative assessment on summative performance in integrated musculoskeletal system module. *BMC medical education*, *15*(1), 1-7. Olson, B., & McDonald, J. (2004). Influence of online formative assessment upon student learning in biomedical science courses. *Journal of Dental Education, 68(6)*, 656-659.

Pallant, J. (2016). *SPSS Survival Manual: A Step by Step Guide to Data Analysis Using IBM SPSS.* Allen & Unwin.

Pitt, E. (2017). Student utilisation of feedback: A cyclical model. *Scaling up Assessment for Learning in Higher Education* (pp. 145-158). Hong Kong: Springer.

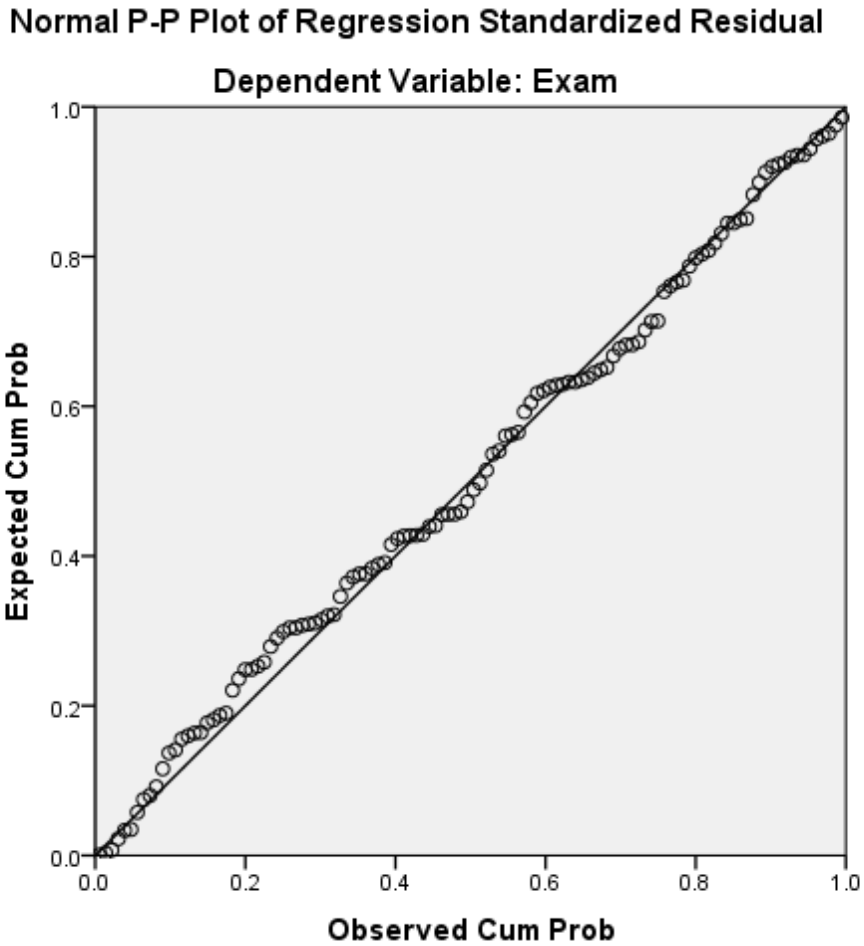
Rowe, A. (2017). Feelings about feedback: The role of emotions in assessment for learning. *Scaling up Assessment for Learning in Higher Education* (pp. 159-172). Hong Kong: Springer.

Shepard, L. (2005). Formative assessment: Caveat emptor. *The Future of Assessment: Shaping Teaching and Learning* (p. Chapter 12). New York: Taylor and Francis Group.

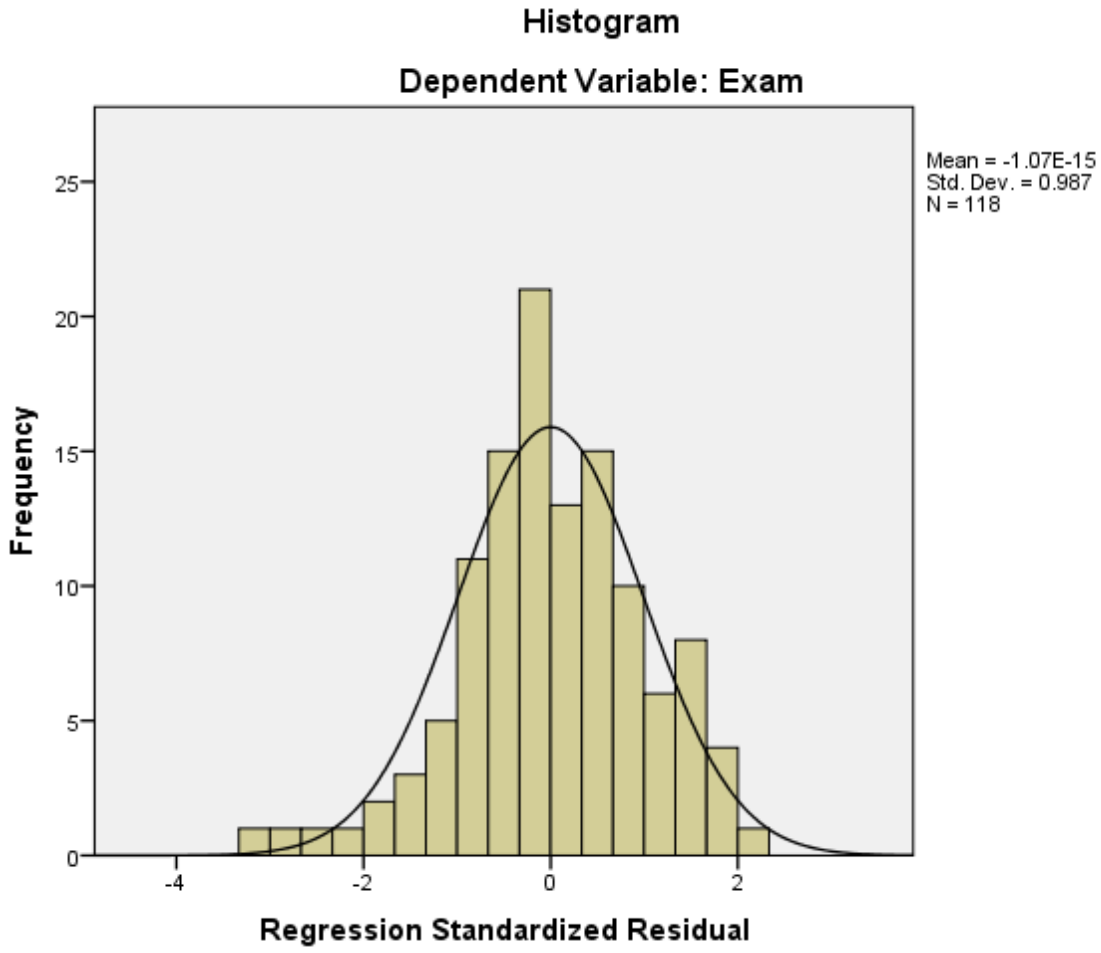
Zhang, L., & Zheng, Y. (2018). Feedback as an assessment for learning tool: How useful can it be? *Assessment & Evaluation in Higher Education, 43(7)*, 1120-1132.

Zhang, N., & Henderson, C. (2015). Can formative quizzes predict or improve summative exam performance. *Journal of Chiropractic Education, 29(1)*, 16-21.

Appendix 1


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Exam
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

Appendix 2


Histogram
Dependent Variable: Exam
Mean = -1.07E-15
Std. Dev. = 0.987
N = 118
Frequency
25
20
15
10
5
0
-4
-2
0
2
Regression Standardized Residual

Appendix 3