DOES PAIRED FORMATIVE ASSESSMENT ENHANCE MATHEMATICS PERFORMANCE?

Jennifer O. Parcutilo
University of Science and Technology of Southern Philippines, Lapasan Highway, Cagayan de Oro City, Philippines
Tel. +639172491907, Email: jennifer.parcutilo@ustp.edu.ph

ABSTRACT: *This study is a sequel of a study conducted in 2008 which find the paired formative assessment effective in increasing calculus performance of students in the science curriculum. This study reports the result of the paired formative assessment as an alternative assessment which aimed to validate its effects on students' mathematics performance among students in the non-science curriculum. The experimental study was conducted for four consecutive school years during the second quarter. Participants were from the last three batches of fourth-year high school in the Revised Basic Education Curriculum (RBEC) and the first batch of Grade 10 of the K-12 Curriculum. Two sections were randomly chosen from each batch as experimental and control groups. This study found that the formative assessment approach have a significant effect on high school students' performance in quarter examinations. Those who were allowed to discuss by pair during frequent short quizzes in mathematics class, consistently performed significantly higher than those who answered quizzes independently. Further analysis also revealed that even for students who are in the K-12 curriculum, the students in the paired assessment performed better than those in the individual assessment. The last batch of RBEC , SY 2014-2015, scored the highest among the four experimental groups. This may be explained by the strong desire of students not to fail and avoid the additional 2 years in senior high school.*

Keywords: paired formative assessment, K-12 curriculum, Revised Basic Education Curriculum

1. INTRODUCTION

Assessment systems, including classroom and large-scale assessment, are organized around the purpose of improving student learning. These provide useful information about whether students have reached important learning goals and about the progress of each student. Assessment can be classified as formative and summative. Formative assessment is defined as a process used during instruction to provide feedback for the adjustment of ongoing teaching and learning for the purposes of improving student achievement related to instructional objectives [1]. Summative assessments on the other hand, are those assessments designed to determine a students' academic development after a set unit of material [2]. Classroom assessment that is integrated with curriculum and instruction is the primary means of evaluation. It provides teachers the full range of information about their students and their classrooms which are gathered and synthesized. Arguments have been constructed regarding the need for new forms of educational assessment [3], for a paradigm shift with a focus on supporting learning rather than on sorting and selecting students [3, 4, 5]. Assessment is part of the learning process by which learners need not fear. More often than not, even in the graduate school, whenever a test is given, students tend to struggle from apprehensions of failing in a test.

In the Philippine setting, accountability testing which are mostly multiple-choice measures conducted in division, regional and national levels, are the assessment systems used by the Department of Education in evaluating the effectiveness of a curriculum and teacher instructions both in Elementary and Secondary Levels. These are high stake assessments anticipated by public school teachers and administrators, for it labels the schools performance and reputation in the academic society. In most cases, the results of these accountability testing particularly in mathematics revealed results which are not favorable and in this case, teachers are always considered the major contributory factor. Even in classroom assessments, mathematics tests almost always show poor results.

For years, educators have already recognized the benefits of students working in pairs or working cooperatively with groups. Most of the research indicates that cooperative learning leads to higher achievement for all students [6]. According to Johnson and Johnson [7], cooperative learning tends to result in higher achievement and a variety of important outcomes that are aims in education. However, there seems to be a problem in aligning assessment with instruction. Students are frightened of being tested because they feel that a test is a way of branding them as failures rather than a way of helping them learn. The idea of not allowing them to talk or even to ask a seatmate while the test is going on somehow increases their anxiety. While it is true that teachers try so hard to be alertly ensuring that nobody cheats in the test, yet teachers cannot be so certain that students really do not cheat. During quizzes, students will always have a tendency to look at other's paper or ask a seatmate about an answer. Seeing the good side of a bad picture, a research conducted in 2008 to fourth-year science curriculum students has revealed a positive effect on the students' performance in high school calculus when students were allowed to answer formative assessments in pairs [8]. The pairs were called quiz-buddies and this assessment approach reduced the mathematics anxiety of the students.

It is in this light that the researcher conducted a series of experiments applying the paired formative assessment to non-science curriculum students to confirm the result of the assessment system for four consecutive years.

2. THEORETICAL FRAMEWORK

The experiment is anchored mainly on the theories of cooperative learning and the Zone of Proximal Development or ZPD. It aimed to validate a theory that the achievement scores or performance of students in Mathematics is influenced by the method of assessment. It particularly posits that learning is better enhanced through pair assessment as opposed to individual assessment. Delandshere [9] reviewed various assessment literature and argued that current educational assessment has limited purposes and methods that generate limited data and is based on poorly articulated, ad hoc theories and assumptions of learning and knowing. She pointed out that assessment is not external to but a part of the act of learning in that they are located in the discourse, actions, and transactions of individuals in participation. The

call for changes in assessment design and practices is a call for a shift to integrate assessment with learning and instruction [10].

Notably, in the theory of social practice, learning, thinking, and knowing are relations among people in activity, with, and arising from the socially and culturally structured world. Learning is a social practice and co-participation mediated by language. Learning is not located in the individual's mind but takes place in interaction with others [11]. Allowing students to communicate even during quizzes and analyzing its effect on the students learning is the focus of this study. This notion, anchored in the sociocultural theories of learning and associated with the Vygotskian school of thought, claims that human thinking is inherently social in its origins [12]. A Vygotskian viewpoint as articulated by Gallimore and Tharp [13] suggests that teaching is beneficial when it "awakens and rouses to life those functions which are in a stage of maturing, which lie in the zone of proximal development" [14]. Vygotsky's definition of the ZPD- zone of proximal development is the distance between a child's problem-solving capability when working alone and with the assistance of a more advanced partner, such as a teacher or peer tutor. From an educational perspective on Vygotsky's notion of the ZPD, there is learning potential in peer groups where students have an incomplete but relatively equal expertise, each partner possessing some knowledge and skill but requiring the others' contribution in order to make progress [14].

Another perspective that framed this study is a constructivist and socio-constructivist view of learning. From a radical constructivist' perspective, learning is about self-organization. The social construction of knowledge asserts that an individual's learning is affected by participating in a wider culture, the classroom, and the outside world [15].

This study delves on the notion that paired formative assessment discourages competition among students allowing them to experience an accumulation of small successes that creates a growing level of self-confidence and motivation leading to greater achievement in summative assessment.

3. OBJECTIVES

This study aimed to determine if the paired formative assessment could enhance students' mathematics performance in the summative test. It sought to validate the result in 2008 [8] by comparing the data for four consecutive years from different batches of fourth-year high school students under the non-science curriculum.

Specifically, the study endeavored to answer the following questions:

1. How do the students' achievement scores compare as influenced by the method of assessment, individual formative assessment and paired formative assessment?
2. How do the assessment results compare to four years?

4. METHODOLOGY

The study made use of the experimental method of research, particularly the Pretest-Posttest Quasi-Experimental- Control Group Design. Since a treatment was tried out, the investigation utilized two intact groups of which, one received the experimental treatment which is assessing students by the pair. The control group received the usual treatment which was the individual assessment of students. The study was conducted for 4 consecutive years, the school year 2012-2013 to 2015-2016. The investigation was conducted at the Misamis Oriental General Comprehensive High School (MOGCHS) during the second grading period. MOGCHS is considered the biggest public school in terms of population in the City of Cagayan de Oro. One of the oldest schools, it holds an average of 8,500 students, a population which is growing every year. It covers an area of 73, 629 square meters with around 128 classrooms. The campus is situated in the main road of Velez Street. The respondents of the study were fourth-year high school students of the Revised Basic Education Curriculum and K-12 Curriculum. Among the sections handled by the researcher, only two sections were randomly chosen as the experimental and control group. The respondents were intact groups predetermined by the fourth year enrolment committee at the start of the school year. The researcher has no control over how the sections were created. The table below shows the no. of respondents.

Table 1. Number of Respondents Per Group, Per Year

School Year	Control Group N	Experimental Group N
2012-2013	57	58
2013-2014	36	43
2014-2015	34	36
2015-2016	44	40

At the onset of the investigation, a pretest in mathematics was given to the experimental and control groups. The test was answered individually by the respondents. Only one teacher handled both groups in teaching the topics in mathematics during the second quarter. Both groups had the same daily lessons and received the same teaching instruction, method, and strategy. Frequent short quizzes were given at the end of instruction as a form of formative assessment. The quizzes were good for an average of 15 minutes of working time. The most number of quizzes in a week was 4 times. The control group answered the given quizzes individually. The experimental group, on the other hand, answered the same quizzes but with their partner called their quiz buddy. At the start of the investigation, the experimental group was divided into two sets as a high performing group and low performing group, the basis of which was their first grading grades in math. Without informing the students how the grouping was done, they were given the freedom to choose a partner coming from the other group. This was done so as to allow the students to have a buddy they are at ease with. They were then given an orientation on the purpose and role of a quiz buddy. Quiz buddies were allowed to freely discuss and process information collaboratively about an item in a quiz. However, they were aware that they need not write exactly the same answer in their individual paper. The researcher also stressed that a mere giving out of answers to a buddy was not allowed. Buddies were to help each other and not to spoil a partner. Their buddy, therefore, would just be their thinking partner and it was emphasized that they would be graded individually based on their output. As seen on their paper upon checking, many times, the pairs did not have the same answer on a test. The students remained with the same partner throughout the treatment period. As to the seating arrangement, the students in the experimental group were seated beside their quiz buddies throughout the class sessions. This was done to avoid waste of time and

movement whenever a quiz was given. The noise was avoided during quizzes by strictly imposing to the class the 1-inch talk rule. The 1-inch talk is a classroom rule that required students to speak in a low voice, presuming that the ear of the listener is 1 inch away from the mouth of the speaker. The respondents from the experimental group faithfully followed this rule. The control group, on the other hand, had their usual seat arrangement and answered the quizzes quietly on their own. At the end of the treatment period, both groups were given a posttest which was answered independently regardless of whether the students had quiz buddies.

The main instrument used in the study is the teacher-made achievement test in mathematics for the second quarter. The 40-item test was guided by a table of specifications and the topics were based on the syllabus provided to public schools for the fourth year. It was validated through the result of the test during the school year 2011-2012, where it was given as a quarter test. 40 items were found to have an acceptable degree of difficulty and discrimination, out of the 50 items. The test has a reliability of 0.83. The items were taken from the west bank of quarter exams for the fourth year level which was used for years. The other instrument was an achievement test for the K-12 curriculum during the SY 2015-2016 on the topics for the second quarter. Some of the items in the test were taken from the DepEd issue, Mathematics for Grade 10 Learner's Module and Teacher's Guide. This was a 30 –item test guided by a table of specifications on the topics in the second quarter. Only face validation was conducted for the test prior to giving it to the respondents since this was the first year of implementation of the K-12 curriculum for Grade 10 and the topics were just newly introduced to grade 10 teachers during the mass training conducted on May 2015, a month before the school year started. When the results came, the test had a reliability index of 0.68 with all the items found to be good items.

For the treatment of data, the mean and standard deviations were used in this investigation to describe the performance of the two groups of respondents in the achievement test in mathematics. Percentages were used instead of the raw score to have a better comparison of results between groups and between school year. The Two-way Analysis of Covariance was applied to test the significant difference between the performance of the experimental and control groups and the performance of students between school years. The school year was considered as factor A and the group as factor B in the 2-factor ANCOVA. Scheffe was used as post hoc test to determine which batch performed better than the rest.

5. RESULTS AND DISCUSSIONS

Table 2 shows the result of the achievement test in mathematics in four consecutive school years. Scores were converted to mean percentage scores (MPS). MPS is derived by dividing the raw score over the total number of items times 100. Among the groups, it can be seen that the pretest mean is consistently lower than the posttest mean. Students did not have prior knowledge about the topics in mathematics at the start of the second quarter. From the table, it can be observed that among the control groups, the 2014-2015 batch obtained the highest mean MPS of 56% in the posttest. Likewise, among the experimental groups, the 2014-2015 batch earned the highest mean MPS in the posttest which is 66%. It can be noted that this 56% mean

MPS in the posttest of SY 2014-2015 among the control groups is higher than the posttest of the experimental groups from other batches. This batch is the last batch of the Basic Education Curriculum before the K-12 curriculum was implemented in 2015-2016.

Table 2. Mean and Standard Deviation of the Mean Percentage Scores in the Achievement Test in Mathematics

Year		Control Group		Experimental Group	
		Pretest	Posttest	Pretest	Posttest
2012-2013	Mean	10%	48%	8%	55%
	Standard Deviation	7%	10%	5%	11%
2013-2014	Mean	28%	44%	26%	50%
	Standard Deviation	11%	13%	10%	11%
2014-2015	Mean	10%	56%	8%	66%
	Standard Deviation	7%	15%	6%	15%
2015-2016	Mean	5%	46%	6%	56%
	Standard Deviation	5%	25%	5%	28%

Consistent with the result in a 2008 study [8], the experimental group consistently performed better in the posttest as compared to the control group. To test for a significant difference in the students' mathematics performance as influenced by the method of assessment and to test if the result significantly differs each year, the Two Way Analysis of Variance was utilized. This analysis also shows the interaction effect between the method of assessment and the year the treatment was administered. Table 3 shows the summary.

In the two way ANCOVA in table 3, the year and group were brought as factors of the analysis and tested at 0.05 level of significance. The interaction effect between the year and group was found not significant ($p=0.785$) hence, we need to see if the year in itself, as well as the group, bear a significant effect on the students' performance in mathematics. The table revealed a significant difference between groups ($p<0.001$). This implies that the experimental group performed better than the control group as can be seen from the mean scores in Table 2. This confirms the previous result in 2008 that those who were allowed to answer formative assessment by the pair, performed significantly higher than those who answered the assessment individually. The paired formative assessment can enhance mathematics performance of students even in a summative test when students no longer answer in pairs but independently.

The table likewise revealed that between batches, there is a significant difference in the scores ($p<0.001$). This means that the students do not perform the same every year. Some batches are really good than the others, which is actually observed by most teachers and educators. To determine how the batches compare, post hoc analysis was performed. The result is presented in table 4.

Table 3. Summary Table of ANCOVA Unequal N's of Students' Mean Percentage Scores in Mathematics Achievement Test

Sources	Sum of Squares	Df	Mean Squares	F	p-value
Factor A (Year)	1.06	3	0.35	12.82	0.000*
Factor B (Group)	0.63	1	0.63	22.81	0.000*
Interaction Effect	0.03	3	0.01	0.36	0.785
Error Within	9.31	339	0.03		

*significant at 0.05 level

Table 4. Scheffe Test Result on Posttest Multiple Comparisons

Year	2012-2013	2013-2014	2014-2015	2015-2016
2012-2013	---	0.316	0.002	0.966
2013-2014	0.316	---	0.000	0.650
2014-2015	0.002	0.000	---	0.001
2015-2016	0.966	0.650	0.001	---

The Scheffe test shows how the posttests scores compare. The values are significant at 0.05 level. The table above displays the simultaneous comparisons of the posttest results in four years. The results in SY 2012-2013 is found comparable with the results in SY 2013-2014 and SY 2015-2016 ($p= 0.316$; $p = 0.966$). However, SY 2014-2015 shows significant result, this batch performed significantly better than the other 3 batches ($p = 0.002$; $p < 0.001$; $p = 0.001$). This batch was the last of the Basic Education Curriculum. The students were well aware of the implementation of the K-12 Curriculum in SY 2015-2016. Those who will fail in the curricular year will be forced to join the first batch of grade 10 in 2015, adding 2 more years in high school for the senior high. This perhaps explains the strong desire of students to perform well in mathematics. Even the students of the control group in this batch performed comparably the same as the experimental group in the other batches. A strong desire to pass in a subject is indeed a strong motivator for higher performance in mathematics. Research demonstrates a relatively consistent relationship between motivation and achievement in mathematics [20, 21, 22]. Fear of failure motivates a person to participate in various types of activities and the utility value which refers to the extent to which task completion is perceived to facilitate current or future goals is what drives a student to perform well [23]. Students with high levels of motivation consistently exhibit higher achievement and class grades than students with low motivation [20]. It can be noted that the rest of the batches performed at a comparable level though, the table shows that those in the K-12 curriculum have the least significant p-value ($p=0.966$). The mathematics performance of the first batch of Grade 10 is the same as those in the Basic Education Curriculum who are 2 to 3 years ahead of them. The curriculum may not be an issue in the performance of students in mathematics in as far as the result of this experiment is a concern.

Current studies have been conducted on formative assessments and similar to the result of this study, the researchers found that formative assessments which involve pairing improve students' achievement in mathematics [16] and frequent quizzes formative assessment showed high influence on students' summative test [17]. Andersson & Palm [18] recognizes that there is an understudy on the impact of professional development in formative assessment and found in their current study that teachers with improved formative classroom practice as a result of professional development input have resulted to higher achievement in students. This emphasizes the need for educators to design formative assessments that result in positive changes in student achievement. As forwarded by Dunn & Mulvenon in 2009 [19], there is potential for the development of sound evaluation practices and statistical methodology that results in formative assessment and evaluation practice that produces powerful and positive changes in student achievement. Over the years educators have continued to explore the strengths of formative assessment.

6. CONCLUSIONS AND RECOMMENDATIONS

Based on the findings of the study, the paired formative assessment can enhance students' mathematics performance. Allowing students to discuss and process quiz answers with a partner give a positive result in summative tests even when students no longer answer the test with a partner or thinking buddy. Students with a strong motivation to pass perform better in mathematics. The curriculum may not be an issue in the students' mathematics performance.

Mathematics teachers may allow students to discuss and process quiz answers by the pair as a form of formative assessment since it can improve the students' performance in mathematics. Teachers handling large classes are encouraged to employ the paired formative assessment because it is a better method than the individual assessment and gives a positive result in summative tests. This formative assessment approach is recommended for heterogeneous groups. The good performing students may help the low performing ones. In this way, the need for remedial classes may no longer be necessary since the good students may serve as a peer tutor on a daily basis. Similar studies may be conducted for varying groups and may consider other factors such as gender, mental ability, age and duration of the treatment period.

Acknowledgments

Special thanks are extended to Dr. Charita Luna and Professor Socorro Ibonia of the University of Science and Technology of Southern Philippines-Cagayan de Oro for the assistance and advice given to the researcher in the data analysis and to the student participants of Misamis Oriental General Comprehensive High School, Cagayan de Oro City, Philippines.

REFERENCES

[1] Melmer, R., Burmaster, E., & James, T. K. (2008). Attributes of effective formative assessment. Washington, DC: Council of Chief State School Officers. Retrieved October 7, 2008, from http://www.ccsso.org/publications/details.cfm?PublicationID=362

[2] Stiggins, R. J. (2002). Assessment Crisis: The absence of assessment FOR learning. Phi Delta Kappan, 83(10), 758-765.

[3] Shepard, L.A. (2000). The Role of Assessment in a Learning Culture. Educational Researcher, 29(7), 2-14

[4] Gipps, C.V. (1999). Socio-Cultural Aspects of Assessment. Review of Research in Education, 24.

[5] Resnick, L.B., & Resnick, D.P. (1992). Assessing the Thinking Curriculum: New tools for education reform. In B. Gifford & M. O' Connor (Eds.) Changing Assessment: Alternative Views of Aptitude, Achievement, and Instruction (pp. 37-76). London : Kluwer Academic Publishers.

[6] Wong, Harry K. & Wong, Rosemary T. (2001). How to be an Effective Teacher: The First Days of School. Harry K. Wong Publications, Inc.

[7] Johnson, D. W., & Johnson, R. T. (1996). The role of cooperative learning in assessing and communicating student learning. In T. R. Gusky (Ed.) *1996 ASCD yearbook: Communicating student learning.* Alexandria, VA: Association for Supervision and Curriculum Development.

[8] Parcutilo, J.O. & Tandog, V.J. (2008). Quiz Buddy: A pair assessment and its influence on students' performance in Calculus and on Mathematics and test anxiety. Mindanao University of Science and Technology, Graduate Journal, Vol.2, July 2011, 53-62.

[9] Delandshere, G. (2002). Assessment as Inquiry. Teachers College Record. Volume 104, Number 7, pp. 1461-1484. Teachers College, Columbia University.

[10] Pellegrino, J.W., Baxter, G.P., & Glaser, R. (1999). Addressing the "Two Disciplines" Problem: Linking Theories of Cognition and Learning with Assessment and Instructional Practices. Review of Research in Education, 24.

[11] Lave, J., & Wenger, E. (1991). Situated Learning : Legitimate Peripheral Participation. Cambridge, UK : Cambridge University Press.

[12] Sfard, A., Forman, E. A., & Kieran, C. (2001). Learning discourse: Sociocultural approaches to research in mathematics education Educational Studies in Mathematics, 46(1/3), 1-11.

[13] Gallimore R & Tharp R 1990. Teaching mind in society: Teaching, schooling, and literate discourse. In LC Moll (ed). Vygotsky and Education: Instructional implications and applications of sociohistorical psychology. Cambridge: Cambridge University Press.

[14] Hufferd-Ackles, Kimberly, Karen C. Fuson, and Miriam Gamoran Sherin. 2004. "Describing Levels and Components of a Math-Talk Learning Community." Journal for Research in Mathematics Education 35 (March): 81–116.

[15] Cobb P. 1994. Constructivism in mathematics and science education. Educational Researcher 23: 4.

[16] Pradana, O., Sujadi & Pramudya (2017). Think Pair Share with Formative Assessment for Junior High School Student.International Conference on Mathematics and Science Education. IOP Conf. Series: Journal of Physics: Conf. Series 895(2017)012032.

[17] Roble, D.B. & Luna, C.A. (2017). Strengths of Formative Assessments (Daily Time-Pressured Quiz, Homework and Portfolio) on Students' Summative Assessment Performance in Integral Calculus. LDCU-RPO. Asian Scientific Journals, Vol. 13, No 1.

[18] 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. ELSEVIER. Learning and Instruction, Vol. 49, 92-102.

[19] Dunn, Karee E and Mulvenon, Sean W. (2009). A Critical Review of Research on Formative Assessments: The Limited Scientific Evidence of the Impact of Formative Assessments in Education. Practical Assessment Research & Evaluation, 14(7).

[20] Broussard, S. C., & Garrison, M. E. B. (2004). The relationship between classroom motivation and academic achievement in elementary school-aged children. Family and Consumer Sciences Research Journal, 33(2), 106–120.

[21] Gottfried, A. E. (1990). Academic intrinsic motivation in young elementary school children. Journal of Educational Psychology, 82(3), 525–538.

[22] Lange, G. W., & Adler, F. (1997, April). Motivation and achievement in elementary children. Paper presented at the biennial meeting of the Society for Research in Child Development, Washington, D.C.

[23] Lai, E.R., (2011). Motivation: A Literature Review. PEARSON, Pearson's Research Report. https://images.pearsonassessments.com/images/tmrs/Motivation_Review_final.pdf