Original Research Paper

Which Model is the Best Predictor of Learning Achievement: Raw Score, Relative Growth or Knowledge Retention Score?

Phongthara Vichitvejpaisal[1], Tippawan Panjamawat[2], Pinda Varasunun[3]

Abstract

Objective: To verify the student learning performance using the raw score, relative growth score and knowledge retention score.

Methods: Thirty-two nurse anesthetist students in academic years 2011-12, Faculty of Medicine, Siriraj Hospital, Mahidol University, volunteered to participate. After pretest, they studied the designated subject via a website. After 3 weeks, the system was locked and students underwent the post-test. The final1 and final2 test were held in a classroom without prior notice. The post and final1 test as well as the final1 and final2 test took place exactly 4 weeks apart. The difference scores between pre and post-test, pre and final1 test, as well as pre and final2 test were calculated for relative growth score G1, G2 and G3 respectively. Thus the differences between G1 and G2 as well as G1 and G3 were determined as knowledge retention score R1 and R2 respectively.

Results: The post, final1 and final2 test scores were significantly higher than the pretest one significantly. However, the post, final1 and final2 test scores showed no statistical difference. Though G2 and G3 appeared to decrease as compared to G1, they were not significant. The R2 showed higher than R1 without significant difference; however, they showed a strong correlation to each other (r = .69)

Conclusion: The knowledge retention score was the best prediction on academic gains.

Keywords: raw score, relative growth score, knowledge retention score, self-directed learning

Introduction

At present, the diversity and complexity of current medical information requires students to be on the alert for progress in their knowledge. In addition, it is claimed that memory tests are powerful vehicles for improving long-term retention.

[1]*Department of Anesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700*

[2]*17 Soi Therdthai 66 Banghwa Phaseechareon, Bangkok 10160*

[3]*Department of Education, Faculty of Education and Development Sciences, Kasetsart University_Kamphaeng Saen Campus 73140*

Corresponding author:
Phongthara Vichitvejpaisal,
Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand.

E-mail: Phongthara@gmail.com

As a result, the use of progress tests in rounds and clinics, to be administered at regular intervals, should be encouraged once formal instruction has ended (Bangert-Drowns *et al*., 1991; Roediger & Karpicke, 2006; Landauer & Ainsli, 1975; Landauer & Bjork, 1978).

The use of raw score, relative growth score or gain score and knowledge retention score have long been implemented in educational systems. Though there are different applications of the variety of ways to measure student achievement, it is by no means clear that they are the best choice to account for changes in student learning achievement. In addition, it is possible, even likely, that the application of one model over another will lead to different conclusions regarding the growth in achievement of the same group of students.

In order to prove this hypothesis, the raw score, relative growth score and knowledge

retention score are used to verify the student learning performance-regarding analyses of the data for the longitudinal experimental pre-test/post-test control group design.

Methods

The study was approved by the Institutional Review Board. Thirty-two out of forty nurse anaesthetist students in academic years 2011-12, Department of Anaesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University, volunteered to join the three-week study. Participants logged on to a website for registration before performing the pre-test. Then they spent free time at their own pace to study the selected subject and performed the post-test over the next three weeks. The program identified learning weaknesses while avoiding giving those direct answers immediately after the tests. In other words, to get the proper answer for each item of the test, students needed to study the content appearing on the website. The method promoted the learner's role as the decision-maker and planner, self-assessment designer, and the implementer of the discovered information. No efforts were made to evaluate whether students completed their assignments, as the computer program recorded students' profiles in a real-time fashion (name, ID and password, frequency, date and time of access, number of exercises, and score earned).

After three weeks, the website was locked. As usual, participants followed the educational training program in rotation-patients care in the wards, operating rooms and intensive care unit. Besides their ordinary routine, they were able to gain subject contents through available textbooks, bedside teaching staff, daily patients' visit, discussion of the topic among friends and residents etc. which provided them with current and relevant information. One month later, the final1 test was held in a classroom without prior notice, which students had to complete a paper-pencil test of 40 short answers written examination within one hour. In addition, the final2 test took place exactly 4 weeks apart in the same fashion. All tests were developed by using a concept and knowledge map of the selected subject to determine the table of specifications and were under the same behavioral objectives (Novak, 1996).

After the final2 test, students were also interviewed regarding application of the system to self-learning, perceived problems and obstacles, potential weaknesses, risk prevention, as well as opinions on the achievement arising from learning through the system, and the tendency for system application to improve medical study.

Available test scores in the electronic database with unique identification numbers for students were calculated to determine individual students' relative growth and knowledge retention score over time. The difference scores between pre-test and post-test, pre-test and final1 test, as well as pre-test and final2 test were calculated for the relative growth score1 (G1), relative growth score2 (G2) and relative growth score3 (G3) respectively (Kanjanawasee, 1989).

In addition, the relative growth score difference was determined as knowledge retention score. Therefore, the difference between G1 and G2, as well as G1 and G3 were worked out for the knowledge retention score1 (R1) and knowledge retention score2 (R2) respectively (Figure 1).

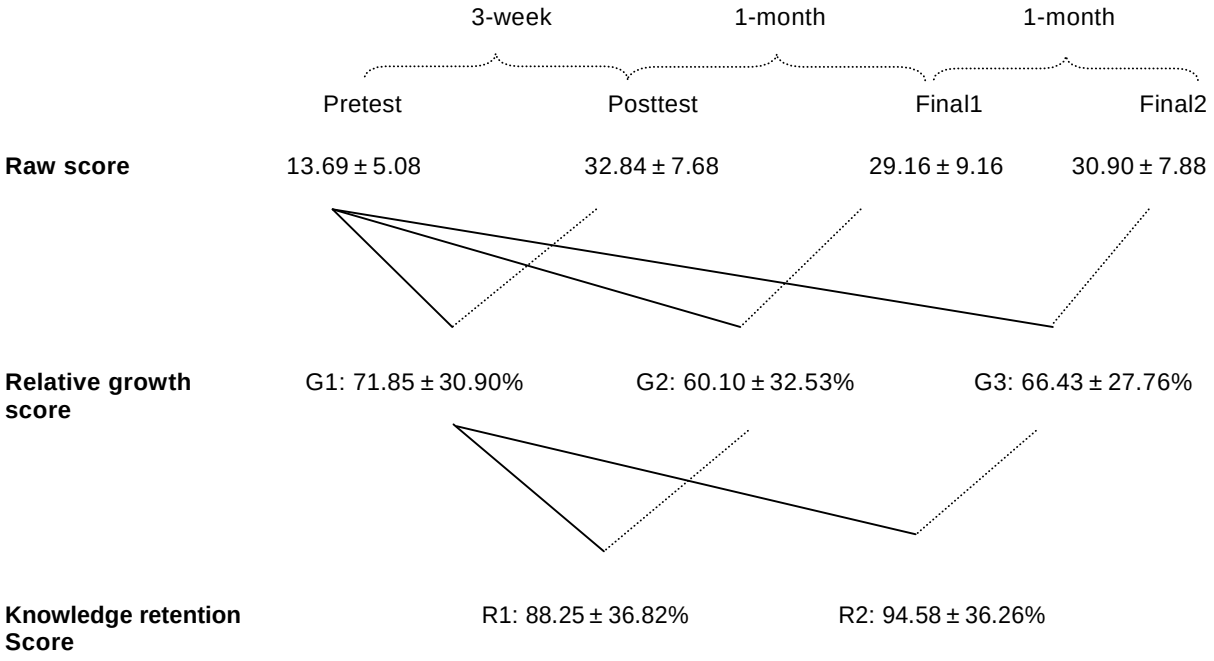
Validation and reliability of the test

The correctness and suitability of all tests (content validity and index of item objective congruence, IOC) were determined by four anaesthesiologists not involved in the project and with at least 10 years of experience in medical science education. For additional review, 16 second-year residents in anaesthesiology performed all tests to verify the assessment of criterion-referenced test item difficulty, discrimination and internal consistency (Cronbach, 1951). Scores graded by four non-physician evaluators measured the outcome.

Statistics

The pre, post, final1 and final 2 test raw scores as well as relative growth scores were analysed by ANOVA and Dunette T3 following the Levine's test. Retention of knowledge scores were compared by t-test dependent using the Statistical Package for Social Sciences for Windows, release 17. Statistically significant differences were noted when p value < 0.05 with a 95% confidence interval.

Figure : Available test scores in an electronic database with unique identification numbers for students were calculated to determine individual students' relative growth and knowledge retention score over time.



$$G1 = \frac{100(Y - X)\%}{F - X} \qquad G2 = \frac{100(Y1 - X)\%}{F - X} \qquad G3 = \frac{100(Y2 - X)\%}{F - X}$$

*G1 = relative growth score1, G2 = relative growth score2, G3 = relative growth score3,
F = full test score, X = pretest score, Y = post-test score, Y1 = final1 test score,
Y2 = final2 test score

$$R1 = 100 - (G1 - G2)\% \qquad R2 = 100 - (G1 - G3)\%$$

*R1 = knowledge retention score1, R2 = knowledge retention score2

Results

For all tests on the website, the IOC of the test was equal to 0.89, 0.80 and 0.95. The assessment of criterion-referenced test item difficulty, discrimination and internal consistency was 0.78, 0.89 and 0.87; 0.17, 0.11 and 0.20; and 0.85, 0.86 and 0.95.

For pre/post/final1 and final2 tests, the IOC was 0.88. The assessment of criterion-referenced test item difficulty, discrimination and internal reliability was 0.59, 0.38 and 0.91.

The pre, post, final1 and final2 test scores were 13.69 ± 5.08, 32.84 ± 7.68, 29.16 ± 9.16 and 30.90 ± 7.88 respectively (Table 1 & 2). The post, final1 and final2 test scores were much higher than the pre-test one significantly. However, the post, final1 and final2 test scores showed no statistically significant difference.

The G1, G2 and G3 were 71.85 ± 30.90%, 60.10 ± 32.53% and 66.43 ± 27.76% respectively (Table 3).Though G2 and G3 appeared to decrease as compared to G1, they were not significantly different.

Table 1: The variance of raw scores between groups and within groups

	ANOVA					Levene's Test	
	Sum of Squares	df	Mean square	F	Sig.	Statistic	Sig.
Between Groups	7393.688	3	2464.563	42.735	.000	4.223	.007
Within Groups	7151.188	124	57.671				
Total	14544.875	127					

** p < .01

Table 2: Comparison of the pre, post, final1 and final2-test score.

	Pre-test	Post-test	Final1 test	Final2 test
Pretest	-			
Post-test	19.16**	-		
Final1 test	15.47**	- 3.69	-	
Final1 test	17.25**	- 1.91	1.78	-

** p < .01

Table 3: The variance of relative growth score 1, 2 and 3 (G1, G2, G3)

				ANOVA				Levene's Test		
	Mean	S.D.		Sum of Squares	df	Mean square	F	Sig.	Statistic	Sig.
G1	71.850	30.901								
G2	60.104	32.534	Between Groups	2212.171	2	1106.085	1.192	.308	.544	.538
G3	66.434	27.761	Within Groups	86304.655	93	928.007				
Total	66.129	30.525	**Total**	88516.825	95					

** p < .01

The average retention of knowledge scores expressed as a percentage of the students' performance in the final1 test (R1) was 88.25 ± 36.82%. Knowledge retention in the final2 test (R2) was 94.58 ± 36.26% (Table 4). The R2 showed higher than R1 without statistically significant difference; however, they showed a strong correlation to each other (r = .69).

Table 4: Comparison of knowledge retention score 1 and 2 (R1, R2).

	Mean	S.D.	Cor-relation	Paired Differences		95% Confidence Interval of the Difference		t	df	Sig. (2-tailed)
				Mean	S.D.	Lower	Upper			
R1	88.253	36.818								
R2	94.583	36.262								
R1 & R2			.694**	-6.3298	28.589	-16.637	3.978	-1.252	31	.220

** p < .01

Discussion

The significantly high raw scores after the pretest implied that the students' learning achievement was satisfactory and the criterion-referenced assessment under this study was able to categorize student performance levels for the selected subjects (Jacobson, 2008).

After the final1 test, the data revealed insignificantly but clinically, a relatively modest decline in both raw score and relative growth score. After the final2 test, the obvious increases of relative growth score and knowledge retention score reflected on student's achievement. This finding seemed to follow what some authors claimed that the growth models had good prediction on academic gains rather than the raw score (Barton *et al.,* 198; Bryk & Raudenbush, 1988; Linn & Haug, 2002; Stevens *et al.,* 2000). However, the knowledge retention score showed a strong correlation to each other as compared to the relative growth score. As a result, the retention of knowledge score could be issued as the best prediction on the assessment of students learning performance.

The features and relationships of raw score, relative growth score and knowledge retention score should be mentioned. Though the raw score has been widely used in many educational fields due to its simplicity, it is not good enough to assess the learner's professional achievement. Based upon the assumption of formative assessment, the use of knowledge retention score seems to play a crucial role in higher academic learning. It implies students' enthusiasm in self-directed learning, problem-solving ability, individual help-seeking strategies, and developing critical thinking with discretion for self-assessment. However, it is a complex and time-consuming process to set up the longitudinal tests and calculate on the knowledge retention score for any selected subjects.

The reasons behind students with high knowledge retention scores by the end of two months, where they would be expected to forget some information may be varied. First, since these students were on a rotation in the patient's ward, operative theatre, recovery room and intensive care unit; they could gain more experience and practice with patients from attending staff to master core knowledge. Second, participants should be convinced that only they are accountable for what they have learned while being committed to the patient care. Therefore, the instructional methods that include understandable real materials rather than struggling with only theoretical concepts and artificial problems develop long-term memory better (Kvan, 2000; Robbs & Meredith, 1994; Kripalani *et al*., 2006; Nadir *et al.,* 2004; Branch & Paranjape, 2002; Paukert *et al.,* 2002). This finding echoed the opinion suggested by Valdez & Paulson (2007) and Euliano *et al.* (2003) regarding the application of technology that appeared to be an effective stepping stone for students to begin developing their higher-level learning and problem-solving skills. Finally, participants had learning group tuition with their friends for rehearsal or relearning the materials after daily rounds. Cooperation resulted in more interaction as group members encouraged and facilitated each other's learning which in consequence affected the outcome achievements and thus enhanced knowledge retention (Bligh, 1972; Johnston *et al.,* 2000; Reinaldo & Rahn, 2006).

Conclusion

The knowledge retention score was the best predictor of academic gains. Frequent testing by interim tests might be the explicit aim of many medical curricula nowadays to train students to become self-directed and lifelong learners.

Acknowledgment

We acknowledge the voluntary help of nurse anaesthetist students in academic years 2011-12, Department of Anaesthesiology, Faculty of Medicine Siriraj Hospital, Mahidol University who fully participated in the project.

Conflicts of interest: None

References

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

Barton, P., Coley, R., & Wenglinsky, H. (1998) *Order in the classroom: Violence, discipline, and student achievement*, Princeton, NJ: Educational Testing Service.

Bligh, D.A., (1972) *What's the use of lectures?* Karmondsworth, England: Penguin.

Branch, W.T. & Paranjape, A. (2002) Feedback and reflection: Teaching methods for clinical settings, *Academic Medicine,* 77, 12, pp. 1185-1188.

Bryk, A.S., & Raudenbush, S.W. (1988) Toward a more appropriate conceptualization of research on school effects: A three-level hierarchical linear model. In R.D. Bock (Ed.), *Multilevel analysis of educational data,* San Diego, CA: Academic Press, pp. 159-204.

Cronbach, L.J. (1951) Coefficient alpha and the internal structure of tests, *Psychometrika,* 16, p. 297.

Euliano, T.Y., Lee, A.I., Melker, J. & Schmalfuss, I.M. (2003) Development and evaluation of an internet-based airway evaluation tutorial, *Medical Education Online* 8, 18, Available at*http://med-ed-online.net/index.php/meo/article/view/4343.*

Jacobson, R.Y. (2008) Examination of the potential of selected norm-referenced tests and selected locally developed criterion-referenced tests to classify students into performance categories, Ed.D. Dissertation, The University of Nebraska – Lincoln, Nebraska: USA.

Johnston, C.G., James, R.H., Lye, J.N. & McDonald I.M., (2000) An evaluation of collaborative problem solving for learning economics, *Journal of Economic Education* 31, 1, pp.13-29.

Kanjanawasee, S. (1989) Alternative strategies for policy analysis: an assessment of school effects on students' cognitive and affective mathematics outcomes in lower secondary school in Thailand, Doctoral Dissertation, Los Angeles: University of California: USA.

Kripalani, S., Jacobson, K.L., Brown, S., Manning, K., Rask, K.J. & Jacobson, T.A. (2006) Development and implementation of a health literacy training program for medical residents, *Medical Education Online,* 11, 13. Available from *http://med-ed-online.net/index.php/meo/article/view/4612.*

Kvam, P.H. (2000) The effect of active learning methods on student retention in engineering statistic, *The American Statistician,* 54, 2, pp.136-140.

Landauer, T. K., & Ainslie, K. I. (1975) Exams and use as preservatives of course-acquired knowledge, *Journal of Educational Research,* 69, pp. 99–105.

Landauer, T. K., & Bjork, R. A. (1978) Optimum rehearsal patterns and name learning. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical aspects of memory*, London: Academic Press, pp. 625-632.

Linn, R. L., & Haug, C. (2002) Stability of school building accountability scores and gains *Educational Evaluation and Policy Analysis*, 24, 1, pp. 29-36.

Naidr, J.P., Adla, T., Janda, A., Feberova, J., Kasal, P. & Hladikova, M. (2004) Long-term retention of knowledge after a distance course in medical informatics at Charles University, Prague, *Teaching and Learning in Medicine,* 16, pp. 255-259.

Novak, J.D. (1996) Concept mapping: a tool for improving science teaching and learning. In Treagust, D.F., Duit, R. & Fraser, B. J. (Eds.) *Improving teaching and learning in science and mathematics,* New York: Teachers College Press, pp. 32-43.

Paukert, J.L., Richards, M.L. & Olney, C. (2002) An encounter card system for increasing feedback to students, *American Journal of Surgery,* 183, 3, pp. 300-304.

Reinaldo, M.R. & Rahn, R. (2006) *Studying knowledge retention through cooperative learning in an operations research course*, College of Engineering and Engineering Technology Northern Illinois University DeKalb, IL 60115, USA.

Robbs, J. & Meredith, S. (1994) The problem-based learning curriculum at Southern Illinois University School of Medicine [Online]. Available from: *http://www.suimed.edu/pblc/ pblcur.html.*

Roediger, H. L. III, & Karpicke, J. D. (2006) The power of testing memory. Basic research and implications for educational practice, *Perspectives on Psychological Science*, 1, 3, pp. 181–210.

Stevens, J. J., Estrada, S. & Parkes, J. (2000) Measurement issues in the design of state accountability systems. Roundtable presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Valdez, C.A. & Paulsen, S. (2007) Design and implementation of a video-based clinical examination, *Medical Education Online,* 12, 9, Available at *http://med-ed-online.net/index.php/meo/article/view/4468.*