Asia Pacific Education Review
2005, Vol. 6, No. 2, 179-190.

A Quasi-Experimental Research on the Educational Value of Performance Assessment

Sun-Geun Back
Seoul National University
Korea

Eun-Hui Hwang
Baekhyun Middle School
Korea

The purpose of this study is to demonstrate that performance assessment increases educational value in teaching-learning activities using a quasi-experimental research design. In this research, the three measurement criteria of educational value are suggested as 'improvement & advancement,' 'sincerity & enthusiasm,' and 'individuality & wholeness.' A pre-test was administered to 4 classes (156 students) in 7th grade. Classes were divided into an experimental group (2 classes, 79 students) and a control group (2 classes, 77 students), according to the pre-test results. Only the experimental group was involved in the performance assessment for 9 weeks. The results of this study show that performance assessment has a positive effect on the educational value of teaching-learning activities in schools.

Key Words: performance assessment, educational value, teaching-learning activity, quasi-experimental research

Education can be defined as a specific human activity in order to increase educational value through the interaction between teacher and learner. Educational value can be in turn defined as the characteristics that are desirable and essential to education. The educational value is conceptually independent of other values such as moral value, economic value, aesthetic value, political value, etc. (Baek, 2000a; Broudy, 1952; Spranger, 1930; Taylor, 1961).

The three criteria for evaluating educational value were suggested as follows; 'improvement & advancement,' 'sincerity & enthusiasm,' and 'individuality & wholeness' (Baek, 2000a; Korea Ministry of Education, 1997). In this respect, 'improvement & advancement' was defined as increasing academic achievement in various areas. 'Sincerity & enthusiasm' was defined as attending sincerely and enthusiastically in teaching and learning activities. 'Individuality & wholeness' was defined as accepting unique personality of each other and pursuing to be a whole person. These kinds of measurement criteria are related to each other, but are independent conceptually (Baek, 2000a). Therefore, the sum of educational value of certain activity can be measured as the volume of a hexahedron, which is constructed by three criteria (see [figure 1]).

Educational testing has traditionally focused on the technical aspects of measurement rather than the educational value of certain teaching-learning activity in Korea. Objective multiple-choice item format tests were widely used to examine the student's achievement in schools. Even though the higher-order thinking skills involved in such things as drawing inferences, analyzing text, or demonstrating a deep understanding of a domain can be measured by objective multiple-choice item format tests, it is very difficult to write multiple-choice exam questions that assess the higher-order thinking skills (Glaser, Lesgold, & Lajoie, 1987; McMillan, 2004). As a result, a large proportion of the items on an achievement test measure only factual knowledge in Korea. These tests fail to show understanding or to appraise the

Sun-Geun Baek, Professor at the Department of Education, Seoul National University; **Eun-Hui Hwang**, Science teacher at Baekhyun Middle School.
Correspondence concerning this article should be addressed to Sun-Geun Baek, Department of Education, Seoul National University, 599 Kwanak-Ro, Kwanak-Gu, Seoul, 151-748, Korea. Electronic mail may be sent to E-mail: dr100@snu.ac.kr; Phone: +82-2-880-7645, Fax: +82-2-878-1665

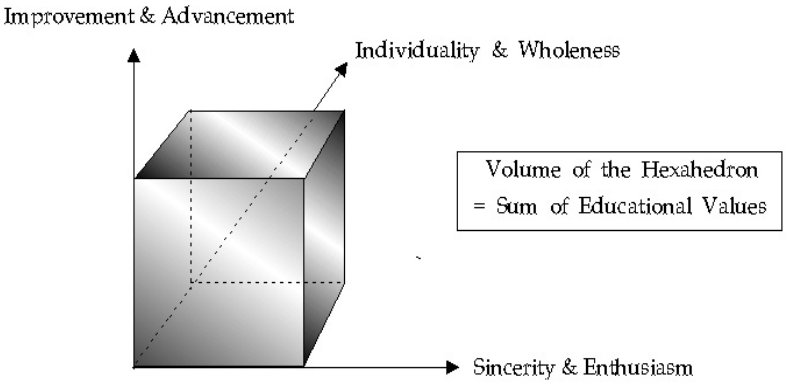


Figure 1. *Three Criteria of Educational Value (Baek, 2000b, p11)*

knowledge structure of the cognitive processes underlying differential performance in specific fields or domains of study (Baek, 1994; KMOE, 1997). In other words, such tests are designed to determine who the biggest information 'container' is, but are not designed to determine how one becomes an expert or how one's competence can be improved. Therefore, multiple-choice exams are inadequate for providing an understanding of the instructional and learning processes. In addition, they are insufficient for prescribing remedies or other instructional interventions. This history of testing has had very little affect on increasing educational value in teaching-learning activities in Korea (Baek, 2000b).

In order to solve these problems, many educators have been interested in performance assessment, rather than multiple-choice tests. They have been interested in performance assessment for a number of reasons: (a) to improve learning and understanding; (b) to develop teaching strategies for individualized instruction; (c) to express the idea that people should learn how to apply what they know; (d) to foster student's acquisition of authentic cognitive performance within a learning domain; (e) to look beyond standardized tests for ways of sampling students' performance that are more closely linked to instruction; (f) to enhance student's self-awareness and self-regulative learning; (g) to integrate teaching, learning, and assessment in the classroom (Bae, 2000; Baek, 2000b; Baron & Boschee, 1995; Darling-Hammond, Ancess, & Falk, 1995; Herman, Aschbacher, & Winters, 1992; McMillan, 2004; Sternberg, 1991; Wiggins, 1989). Many educators expect that if performance assessment is appropriately implemented within teaching-learning activities, it will increase the educational value more than traditional multiple-choice tests do (Baek, 2000a; KMOE, 1997).

The main purpose of this study is to investigate, comprehensively, whether or not performance assessment based teaching-learning activities create more educational value in comparison to the traditional teaching-learning activities that are not performance assessment based. The reason for selecting a performance assessment based teaching-learning activity, as the subject of study, is that there are many theoretical studies that emphasize the settlement and extension of performance assessment, but there are few studies that compare empirically and comprehensively performance assessment based teaching-learning activities to that which are not performance assessment based in Korea.

Performance assessment was introduced into Korean elementary schools, middle schools, and high schools officially in late 1990s (KMOE, 1998; National Institute of Educational Evaluation, 1996). After the introduction, the Korean Ministry of Education, and many other organizations, undertook continuous studies and efforts to establish and expand performance assessment. Additionally, many experts have been studying the understandings, practical uses, problems, and improvement methods regarding performance assessment (Bae, 2000; Baek, 2002b; Baek et al., 1998; Kim, 2000; KSEE, 2000; Lee et al., 1998; Paek, 1999). There are 6 doctorate dissertations, 389 master's degree theses, 44 separate volumes of research, and 254 research papers in Korea regarding this topic (Heo et al., 2001). However, the majority of those studies are theoretical reviews and only a few studies are empirical studies. Those empirical studies show that performance assessment has had positive effects on the improvement of students' intellectual and emotional abilities in areas such as achievement, learning attitude, creativity and inquiry ability, etc. (Bae, 2001; Cho et al., 2001; Han et al., 2000 No, 2000; Park & Baik, 2000). Even though those empirical studies show some evidences about the educational value of performance assessment, many Korean

people still believe that there is no distinct evidence of educational value in performance assessment. Additionally, only a part of educational value, i.e., only one or two criteria of educational value, are used as research variables in those studies.

Again, the purpose of this study is to investigate, comprehensively, through quasi-experimental research, whether or not performance assessment increases educational value in teaching-learning activities. For this study, 'educational value' is defined as the value which can divide certain activities as educational, or not educational, and three measurement scales for the three criteria of educational values were developed: the 'improvement & advancement' scale (i.e. science achievement test), the 'sincerity & enthusiasm' scale, and the 'individuality & wholeness' scale. By using the measurement scales, the question as to whether the performance assessment based teaching-learning activities create more educational value or not will be examined.

Even though this research was carried out under restricted conditions, it will contribute to making the characteristics of performance assessment clear and to further establish the settlement of performance assessment in Korea and elsewhere.

Research Questions

It was hypothesized that there would be significant differences in the educational value, with or without the implementation of performance assessment within teaching-learning activities. The three criteria for evaluating educational value were suggested as 'improvement & advancement', 'sincerity & enthusiasm', and 'individuality & wholeness'. In order to investigate the hypotheses, the following research questions were proposed and investigated.

1) Does performance assessment improve and advance science achievement?
2) Do students have more sincerity and enthusiasm about science after the performance assessment?
3) Does performance assessment increase students' individuality and wholeness?

Methodology

Subjects

For this study, three pre-tests were administered to 4 classes (156 students) in 7th grade. Classes were divided into an experimental group (2 classes, 79 students) and a control group (2 classes, 77 students), according to the pre-test results (see Table 1). The same science teacher taught all students in order to control teacher variability, but only the students in the experimental group participated in the performance assessment based teaching-learning activities for nine weeks. After nine weeks, three post-tests were administered to both the experimental group and control group (More details about the quasi-experimental research procedure may be found within the fourth portion of this section).

Table 1. *Number of Subjects*

Group	Number of Classes	Number of Students
Experimental Group	2	79
Control Group	2	77
Total	4	156

There were no statistically meaningful differences between the experimental group and the control group in pretest scores of the three research variables: 'improvement & advancement', 'sincerity & enthusiasm', and 'individuality & wholeness'. The pretest result for 'improvement & advancement' (i.e. science achievement test) is shown in Table 2. The experimental group and control group had no statistically meaningful difference (p=0.92).

Table 2. *Pretest Result of 'Improvement & Advancement'*

	Number of Students	Average	Standard deviation	t-value
Experimental Group	79	70.03	21.10	-0.11
Control Group	77	70.34	21.92	

The pretest result for 'sincerity & enthusiasm' is shown in Table 3. The experimental group and control group had no statistically meaningful difference (p=0.67).

Table 3. *Pretest Result of 'Sincerity & Enthusiasm'*

	Number of Students	Average	Standard deviation	t-value
Experimental Group	79	56.28	15.22	0.43
Control Group	77	55.31	12.88	

The pretest result for 'individuality & wholeness' is shown in Table 4. The experimental group and control group had no statistically meaningful difference (p=0.07).

Table 4. *Pretest Result of 'Individuality & Wholeness'*

	Number of Students	Average	Standard deviation	t-value
Experimental Group	79	67.16	11.95	1.84
Control Group	77	63.73	11.32	

Development of Measurement Scales

The three criteria for evaluating educational value were suggested as 'improvement & advancement', 'sincerity & enthusiasm', and 'individuality & wholeness'. Three measurement scales were developed: the 'improvement & advancement' scale, the 'sincerity & enthusiasm' scale, and the 'individuality & wholeness' scale. In this research, the procedure for quantitatively developing each measurement scale is as follows.

'Improvement & Advancement' Scale

In order to measure students' 'improvement & advancement', science achievement tests for midterm and final exams were developed, based upon the findings of Baek, So, & Cho (1998), Hibbard (2000), and Kim (2000). To make these test items, a pilot test was administered to 100 7th grade students. Through the analysis of the pilot test results, some items that have high or low correct-answer rates and low discrimination indices were deleted while others were partially revised, if necessary. Two middle school science teachers examined the finalized items. Items for the midterm and final exams were made respectively. Each exam had 30 items that consisted of 10 knowledge items, 10 understanding items and 10 application items, and full marks totaled to 100 points. To check the reliability of these achievement tests, Cronbach's alpha coefficients were used (see Table 5). To establish the validity of items, two science teachers examined test items.

'Sincerity & Enthusiasm' Scale

In order to measure students' sincerity and enthusiasm, the 'sincerity & enthusiasm' scale was developed. By analysis of related studies, the subscale of sincerity and enthusiasm was established as 'attachment to science learning' and 'interest about science'. Items of the 'sincerity & enthusiasm' scale were developed referring to Baek (1986)'s study and Fraser (1981)'s TOSRA (Test of Science-Related Attitudes).

A pilot test was administered to 100 7th grade students. Through analysis of the pilot test results, 20 items were selected among 30 preliminary items (see Appendix 1). The scale's scores ranged between 20 and 100. The reliability coefficient (Cronbach's α coefficient) of the 'sincerity & enthusiasm' scale was 0.95 and those of each subscale were equally 0.92 (see Table 6).

To establish the content validity of the 'sincerity & enthusiasm' scale, a series of interviews with a professor and graduate students in the department of education specializing in educational measurement and evaluation were conducted. In addition, factor analysis (principal component analysis with varimax orthogonal rotation) was used to identify factorial

Table 5. *Reliability of 'Improvement & Advancement' (i.e. Science Achievement Test)*

Achievement Test	α	Sample of Items
Midterm Exam	.94	Make groups of the following elements based upon its status at the present temperature (22℃). : Ethyl alcohol, Air, Hydrogen, Coal, Cupper, Quartz, Mercury, Ether
Final Exam	.95	I observed 0.02mm cell by using an eyepiece (X10) and an object lens (X20). What's the length of the cell that is observed in the microscope?

Table 6. *Reliability of 'Sincerity & Enthusiasm' Scale*

Scale	α	Subscale	α	Sample of Item
Sincerity & Enthusiasm	.95	Attachment to science learning	.92	I wish to have science class more often.
		Interest about science	.92	I like to do scientific experiments at home.

validity (or construct validity) of the instrument (Rice, 1988) (see Table 7).

'Individuality & Wholeness' Scale

In order to measure students' individuality and wholeness, the 'individuality & wholeness' scale was developed. Items for the 'individuality & wholeness' scale were developed referring to Baek (1986)'s self-development scale and Dakedosi (1989)'s books on individuality.

A pilot test was administered to 100 7th grade students. Through analysis of the pilot test results, 20 items were selected among 30 preliminary items (see Appendix 2). The scale's scores ranged between 20 and 100. The reliability coefficient (Cronbach's α coefficient) of the 'sincerity & enthusiasm' scale was 0.94 and those of each subscale were 0.88 (individuality) and 0.91 (wholeness) (see Table 8).

To establish the content validity of the 'sincerity & enthusiasm' scale, a series of interviews with a professor and graduate students in the department of education specializing in educational measurement and evaluation were conducted. In addition, factor analysis (principal component analysis with varimax orthogonal rotation) was used to identify factorial validity (or construct validity) of the instrument (Rice, 1988) (see Table 9).

Table 7. *Factor Analysis of 'Sincerity & Enthusiasm' Scale*

	Subscale	Item Number	Factor I	Factor II	Communality
Sincerity & Enthusiasm	Attachment to Science Learning	1	.85	.22	.77
		3	.84	-.08	.71
		5	.79	.09	.64
		7	.74	.28	.64
		9	.58	.27	.40
		11	.57	.20	.36
		13	.49	.40	.40
		15	.39	.26	.22
		17	.39	.19	.18
		19	.38	.23	.20
	Interest about Science	2	.15	.75	.59
		4	.26	.67	.51
		6	.21	.58	.38
		8	.20	.49	.29
		10	.30	.46	.30
		12	.26	.46	.29
		14	.10	.46	.22
		16	.34	.41	.28
		18	.19	.41	.20
		20	-.05	.30	.09
	Eigen Value		4.45	3.22	7.67
	Variance Explained (Cumulative %)		22.25	16.10	38.35

Table 8. *Reliability of 'Individuality & Wholeness' Scale*

Scale	α	Subscale	α	Sample of Item
Individuality & Wholeness	.94	Individuality	.88	I know that I'm good at making relationships.
		Wholeness	.91	I'm proud of myself when I follow the rules.

Table 9. *Factor Analysis of 'Individuality & Wholeness' Scale*

	Subscale	Item Number	Factor I	Factor II	Communality
Individuality & Wholeness	Individuality	21	.76	.07	.58
		23	.75	.03	.56
		25	.65	.35	.55
		27	.64	.22	.46
		29	.61	.18	.41
		31	.61	-.21	.41
		33	.60	-.27	.43
		35	.54	.30	.38
		37	.52	.45	.47
		39	.49	.24	.29
	Wholeness	22	-.06	.68	.46
		24	.17	.60	.39
		26	.16	.52	.29
		28	-.08	.50	.26
		30	.29	.46	.30
		32	-.02	.44	.19
		34	.26	.42	.25
		36	.04	.41	.17
		38	.34	.39	.27
		40	.23	.33	.16
	Eigen Value		4.27	3.03	7.30
	Variance Explained (Cumulative %)		21.35	15.15	36.50

Table 10. *Pearson Correlations among Measurement Criteria of Educational Value (N=156)*

		Improvement & Advancement	Sincerity & Enthusiasm			Individuality & Wholeness		
			Attachment	Interest	Subtotal	Individuality	Wholeness	Subtotal
S & E	Attachment	.46**						
	Interest	.56**	.67**					
	Subtotal	**.56****	.90**	.93**				
I & W	Individuality	.29**	.38**	.28**	.36**			
	Wholeness	.44**	.58**	.46**	.56**	.69**		
	Subtotal	**.40****	.52**	.41**	**.50****	.92**	.92**	
Multiplication of 3 Criteria (Volume)		**.86****	.73**	.76**	**.82****	.51**	.68**	**.65****

**p < .01

Pearson Correlations among the Three Measurement Criteria of Educational Value

Pearson correlations among the three measurement criteria of educational value were reasonably high. The results of the correlation analysis are shown in Table 10.

Instructional Contents and Performance Assessment Techniques

The instructional contents for both experimental and control group were the same; however, the only difference was whether or not performance assessment in the teaching-learning activities was implemented and utilized. The performance assessment focused on doing, not merely knowing, and on the process or procedure used as well as the product resulting from one's performance of a task. In Table 11 are shown the instructional contents that were taught to both groups and the performance assessment techniques that were implemented only to experimental group.

Quasi-Experimental Research Procedure

A quasi-experimental research exercise was conducted for the duration of nine weeks. The research procedure was as follows (see table 12).

Results

According to the reliability, validity and correlation analysis results, it was revealed that the developed measurement scales had high reliability and acceptable validity. By using those scales, the educational value of performance assessment was examined. ANCOVA analyses were used to reveal information about the differences between the experimental group and the control group in the three criteria of educational value.

First, the performance assessment had a significantly positive effect on 'improvement & advancement,' i.e. students' science achievement (see Table 13). Pre- and post-test mean scores of the experimental group were

Table 11. *Instructional Contents and Performance Assessment Techniques*

Instructional Contents for Both Groups		Performance Assessment Techniques
Three States of Materials	The properties of solids, liquids and gases State change of materials State change and molecule	Evaluation of experiment Portfolio Self evaluation
The Motion of Molecule	The molecule which moves by itself The pressure of gas The pressure and volume of gas The temperature and volume of gas	Conceptual map Self evaluation
Biology	The discovery of cell To use microscope Observation of cell The structure of cell	Evaluation of experiment Project evaluation Self evaluation

Table 12. *Quasi-Experimental Research Procedure*

1. Three pre-tests were administered to 4 classes in the 7th grade. Classes were divided into an experimental group (2 classes) and a control group (2 classes), according to three pre-test results.
↓
2. The experimental group participated in performance assessment based teaching-learning activities for nine weeks. The control group participated in traditional (i.e., not performance assessment based) teaching-learning activities with the same curriculum materials.
↓
3. Both experimental and control group students took three post-tests.
↓
4. Data were analyzed by using the ANCOVA technique.

Table 13. *ANCOVA Analysis on 'Improvement & Advancement'*

	Sum of Squares	df	Mean Square	F
Control variable (Pre-test)	51290.26	1	51290.26	396.01**
Main effects(Performance assessment)	2079.74	1	2079.74	16.06**
Explained	53195.41	2	26957.71	205.36**
Residual	19186.09	153	129.52	
Total	72381.50	155		

(**p < .01)

Table 14. *ANCOVA Analysis on 'Sincerity & Enthusiasm'*

	Sum of Squares	df	Mean Square	F
Control variable(Pre-test)	17058.56	1	17058.56	202.48**
Main effects(Performance assessment)	561.89	1	561.89	6.67**
Explained	17854.91	2	8927.46	105.97**
Residual	12889.68	153	84.25	
Total	30744.59	155		

(**p < .01)

Table 15. *ANCOVA Analysis on the Differences of Individuality and Wholeness*

	Sum of Squares	df	Mean Square	F
Control variable(Pre-test)	7167.60	1	7167.60	101.37**
Main effects(Performance assessment)	333.21	1	333.21	4.71*
Explained	8140.43	2	4065.22	57.50**
Residual	10817.79	153	70.71	
Total	18948.22	155		

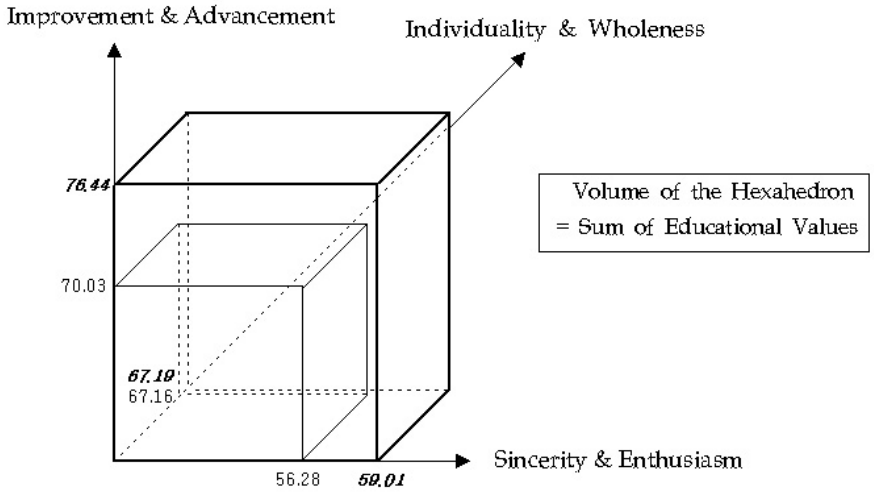
(*p < .05; **p < .01)

70.03(SD=21.10) and 76.44(SD=19.65), respectively. However, pre- and post-test mean scores of the control group were 70.34(SD=21.92) and 69.45(SD=23.22), respectively. The effect-size was 0.30.

Second, the performance assessment increased students' 'sincerity & enthusiasm' for learning science (see <Table 14>). Pre- and post-test mean scores of the experimental group were 56.28(SD=15.22) and 59.01(SD=12.88), respectively. However, pre- and post-test mean scores of the control group were 55.31(SD=12.40) and 54.49(SD=15.37), respectively. The effect-size was 0.29.

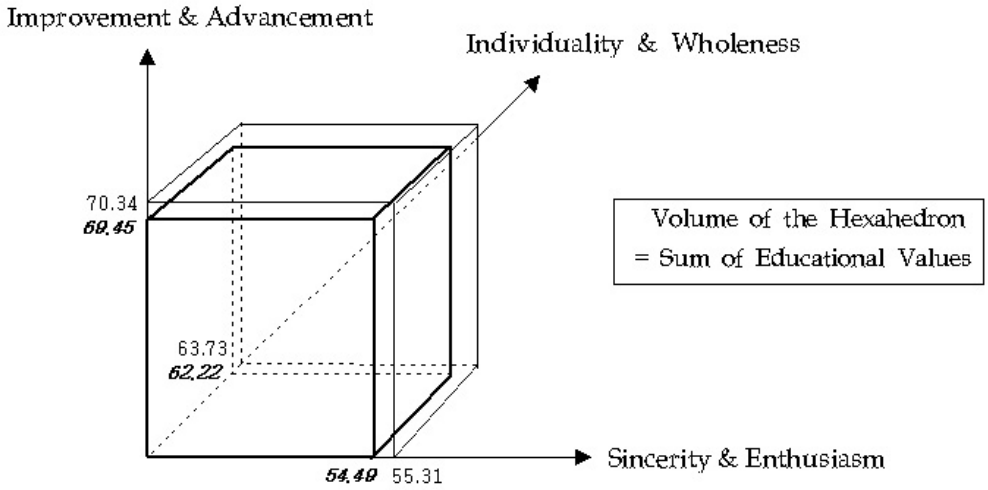
Third, the performance assessment improved students' 'individuality & wholeness' (see <Table 15>). Pre- and post-test mean scores of the experimental group were 67.16(SD=11.93) and 67.19(SD=10.57), respectively. However, pre- and post-test mean scores of the control group were 63.73(SD=11.32) and 62.22(SD=11.05), respectively. The effect-size was 0.45.

As a result, this research shows that performance assessment based teaching-learning activities are more educationally valuable than activities that are not based on performance assessment. This is expressed in figures 2 and 3.



** The bold lines represent post-test results and the thin ones do pre-test results.*

Figure 2. Change of Educational Value (Experimental Group)



** The bold lines represent post-test results and the thin ones do pre-test results.*

Figure 3. Change of Educational Value (Control Group)

In the case of the experimental group, the volume of the hexahedron (sum of educational value) increased (see Figure 2). However, in the case of the control group, the volume of the hexahedron (sum of educational value) decreased (see Figure 3).

Discussion and Conclusion

Performance assessment has been applied in Korea since the 1990s. Even though this research was carried out under restricted conditions, this study confirms how valuable performance assessment is, empirically and comprehensively. Educational value may be defined as characteristics that are desirable and essential to education. For this study, three criteria for evaluating educational value are suggested as 'improvement & advancement', 'sincerity & enthusiasm', and 'individuality & wholeness'. In order to measure each criterion quantitatively, three measurement scales of educational value were developed: the 'improvement & advancement' scale (i.e. science achievement test), the 'sincerity & enthusiasm' scale, and

the ‘individuality & wholeness’ scale.

For this study, three pre-tests were administered to 4 classes (156 students) in the 7th grade. Classes were divided into an experimental group (2 classes, 79 students) and a control group (2 classes, 77 students), according to the pre-test results. Only the experimental group was involved in the performance assessment for 9 weeks. After the duration of the performance assessment based teaching-learning activities, three post-tests were administered to both groups. The results are as follows.

First, each measurement scale had high reliability as well as acceptable validity. The reliability coefficient (Cronbach’s α coefficient) of the ‘improvement & advancement’ scale was 0.95. Those of the 'sincerity & enthusiasm' scale and the 'individuality & wholeness' scale were 0.94 and 0.95, respectively. The results of both correlation analysis and factor analysis showed acceptable validity of those scales.

Second, the performance assessment had a significant effect on students' science achievement ($F=16.06$, $p<.01$). Pre- and post-test mean scores of the experimental group were 70.03(SD=21.10) and 76.44(SD=19.65), respectively. However, pre- and post-test mean scores of the control group were 70.34(SD=21.92) and 69.45(SD=23.22), respectively. The effect-size was 0.30.

Third, the performance assessment increased students' sincerity and enthusiasm in relation to science education ($F=6.67$, $p<.01$). Pre- and post-test mean scores of experimental group were 56.28(SD=15.22) and 59.01 (SD=12.88), respectively. However, pre- and post-test mean scores of the control group were 55.31(SD=12.40) and 54.49(SD=15.37), respectively. The effect-size was 0.29.

Fourth, the performance assessment improved students' individuality and wholeness ($F=4.71$, $p<.05$). Pre- and post-test mean scores of experimental group were 67.16(SD= 11.93) and 67.19(SD=10.57), respectively. However, pre- and post-test mean scores of control group were 63.73(SD=11.32) and 62.22(SD=11.05), respectively. The effect-size was 0.45.

In brief, this study was a trial to explore the educational value of performance assessment empirically and comprehensively. It was empirically confirmed that performance assessment based teaching-learning activities are more educationally valuable than those that were not performance assessment based, even though the effect-sizes were relatively not so high. Therefore, performance assessment should be used widely for students' cognitive and affective development in schools. However, there are many restrictions when generalizing this study's results because of the short research period and restricted study subjects. In addition, only one teacher participated in the teaching and learning activities, which may have affected research results. To solve these restrictions and confirm the educational value of performance assessment systemically, many teachers should pay increased attention to these kinds of studies and such studies should be conducted longitudinally across a wider variety of areas.

References

Anastasi, A. & Urbina, S. (1997). *Psychological testing* (7th ed). EJ: Prentice Hall.

Bae, H. S. (2000). *Theoretical foundation of performance assessment*. Seoul: Hakjisa.

Bae, M. W. (2001). The effect of portfolio assessment based biology class on students’ attitude and achievement. Unpublished master degree’s thesis, Yonsei University, Korea.

Baron, M. K., & Boschee, F. (1995). *Authentic assessment: The key to unlocking student success*. PA: Technomic Publishing Company, Inc.

Baek, S. G. (1986). The effects of evaluation style using change score on school learning. Unpublished master degree’s thesis, Seoul National University, Korea.

Baek, S. G. (1994). Implications of cognitive psychology for educational testing. *Educational Psychology Review, 6*(4), 373-389.

Baek, S. G. (2000a). *Principles of performance assessment*. Seoul: Educational Science.

Baek, S. G. (2000b). Inquiry on the concept of educational evaluation. *Journal of Educational Measurement and Evaluation, 13*(1), 1-20.

Baek, S. G. (2001). A proposal for successful implementation of performance assessment: Focus on the quality control of its methods and instruments. *Journal of the College of Education, 62*, 55-72.

Baek, S. G., So, K. H., & Cho, N. S. (1998). *Development of national curriculum-based assessment standards*. Seoul: Korea Institute Curriculum & Evaluation.

Broudy, H. S. (1952). Building a Philosophy of Education, 169-420. In Beck, L.W. & Holmes, R. L. *Philosophic inquiry*. Englewood Cliffs, N. J.: Prentice-Hall.

Cho, H., et al. (2000). The effect of portfolio-based instruction on socio-psychological environment in elementary school classroom. *Korea Science Education, 20*(3), 529-536.

Dakkuma Dakedosi. (1989). *Psychology of individuality and aptitude*. Seoul: Voice Press.

Darling-Hammond, L., Ancess, J., & Falk, B. (1995). *Authentic assessment in action: Studies of schools and students at work.* NY: Teachers College Press.

Eisner, E. W. (1994). *The educational imagination: On the design and evaluation of school programs* (3rd ed.). NY: MacMillan College Pub.

Fraser, B. J. (1981). *Test of science-related attitudes. Handbook.* The Australian Council for Educational Research.

Glaser, R., Lesgold, A, & Lajoie, S. (1987). Toward a cognitive theory for the measurement of achievement. In Ronning, R. R., Glover, J. A., Conoley, J. C. & Witt, J. C. (Eds.), *The influence of cognitive psychology on testing* (pp.1-8). NJ: Erlbaum.

Han, S., et al. (2000). The effect of portfolio-based instruction on student's creativity and problem-solving ability in elementary school. *Korea Science Education, 20*(3), 421-429.

Heo, K., et al. (2001). *A meta-analysis of researches on performance assessment in elementary and secondary schools*. Seoul: Korea Institute Curriculum & Evaluation.

Herman, J. L., Aschbacher, P. R. & Winters, L. (1992). *A practical guide to alternative assessment*. VA: Association for Supervision and Curriculum Development.

Hibbard, K. M. (2000). *Performance-based learning as assessment in middle school science*. NY: Eye On Education.

Kim, E. J. (2000). Development of science performance assessment scale. Unpublished doctoral dissertation, Seoul National University, Korea.

Korea Ministry of Education (1997). *The 7th revision of national curriculum for elementary and secondary education*. Seoul: Korea Ministry of Education.

Korea Ministry of Education (1998). *New system of university entrance and educational vision: Creation of new school culture*. Seoul: Korea Ministry of Education.

Korean Society for Educational Evaluation (2000). *Practical strategies for implementing and establishing performance assessment in schools*. Seoul: KSEE.

Lee, Y., et al. (1998). *Development of national curriculum-based assessment standards: Focused on high school sciences*. Seoul: Korea Institute Curriculum & Evaluation.

McMillan, J. H. (2004). *Classroom assessment: Principles and practice for effective instruction* (3rd ed.). Boston: Pearson.

National Institute of Educational Evaluation (1996). *Theories and practices of performance assessment*. Seoul: Daehan Textbook Company.

No, K. H. (2000). Research on effect of performance assessment on learning attitude in social studies. Unpublished master degree's thesis, Seoul National University, Korea.

Paek, N. K. (1999). Development of science performance assessment measurement instruments. 178-210. *Elementary Education Research at JinJu College of Education.*

Park, H. & Baik, S. (2001). The effect of portfolio assessment on student's creativity and problem-solving ability in elementary school. *Korea Science Education, 21*(1), 22-29.

Rokeach, M. (1973). *The nature of human values*. New York: Free Publishers.

Rice, J. A. (1988). *Mathematical statistics and data analysis*. CA: Wadsworth & Brooks.

Sowell, T. (1980). *Knowledge and Decisions*. NY: Basic Books.

Spranger, E. (1930). *Lebensformen*. Halle: Max Niemeyer Verlag.

Sternberg, R, J. (1991). Cognitive theory and psychometrics. In Hambleton, R. K., & Zaal, J. N. (Eds.), *Advances in educational and psychological testing* (pp. 367-393). MA: Kluwer Academic Publishers.

Taylor, P. W. (1961). *Normative Discourse*. NJ: Prentice-Hall.

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological process*. MA: Harvard University Press.

Wiggins, G. (1989). A true test: Toward more authentic and equitable assessment. *Phi Delta Kappan, May*, 703-713.

Received October 4, 2004
Revision received November 10, 2005
Accepted December 5, 2005

Appendix 1. ‘Sincerity & Enthusiasm’ Scale

1. Strongly Disagree; 2. Disagree; 3. Not Sure; 4. Agree; 5. Strong Agree

1. During science class, I take good note.
2. I like to do scientific experiments at home.
3. I wish to have science class more often.
4. I like to collect specimens.
5. I look forward to science class.
6. I like to watch animals or plants grow.
7. I think that what I learn in science class is important and fun.
8. I'm interested in Korean science technology.
9. I often think that science class is too short.
10. I often talk about science with my friends.
11. If I had no science class, then school would not be fun.
12. I like science related topics on TV, on radio and in the newspaper.
13. If I don't know something about science, then I make efforts to look it up.
14. I like to make scientific models.
15. I always work hard to get good science grades.
16. I want to participate in science clubs.
17. I want my science grade to be higher than all my other grades.
18. I make efforts to know more about science when I see it in my daily life.
19. Science is more fun than all other subjects.
20. I enjoy visiting science labs and museums during the weekends and on holidays.

Appendix 2. ‘Individuality & Wholeness’ Scale

1. Strongly Disagree; 2. Disagree; 3. Not Sure; 4. Agree; 5. Strong Agree

21. I believe all people are good at, at least one thing, myself included.
22. I'm proud of myself when I follow the rules.
23. I know that I'm good at making relationships.
24. I'm good at doing things that I know that I am supposed to do.
25. I often hear that I'm full of character.
26. I am proud of and enjoy and my leadership abilities.
27. I have many hobbies in order to enjoy life.
28. I accept self-criticism well.
29. I do not fall into peer-pressure.
30. I have the ability to advance my own knowledge and experiences.
31. I plan to choose what I like to do as my career.
32. I use all that I know to help a friend in need.
33. I have goals in my life.
34. I live a more prosperous life as a result of keeping my promises.
35. I know the positive and negative qualities about my personality.
36. I practice the manners that I learn at school everyday.
37. I regularly use my talents.
38. I work toward making myself an upright citizen.
39. I try to accept ideas that are different from mine.
40. I'm more concerned with my own satisfaction than the approval of others.