



KDI SCHOOL

WORKING PAPER SERIES

KDI 국제정책대학원

KDI School of Public Policy and Management

KDI SCHOOL WORKING PAPER SERIES

From Multiple Choices to Performance Assessment: Theory, Practice, and Strategy

Ju Ho Lee

KDI School of Public Policy and Management

Sung-Chang Ryoo

Kookmin University

Sam-Ho Lee

Korea University

December, 2014

Working Paper 14-07



This paper can be downloaded without charge at:

KDI School of Public Policy and Management Working Paper Series Index:

<http://www.kdischool.ac.kr/new/eng/faculty/working.jsp>

The Social Science Network Electronic Paper Collection:

<http://ssrn.com/abstract=2543415>

* We are grateful to the KDI School of Public Policy and Management for providing financial support.

**From Multiple Choices to Performance Assessment:
Theory, Practice, and Strategy**

December, 2014

Ju-Ho Lee, Josh Sung-Chang Ryoo, and Sam-Ho Lee

From Multiple Choices to Performance Assessment: Theory, Practice, and Strategy

Ju-Ho Lee, Josh Sung-Chang Ryoo, and Sam-Ho Lee¹

Abstract

This paper shows that heavy reliance on multiple choice questions for student assessment in Korean education has distorted human capital investment in a way that it centers only on cognitive skills that are testable through multiple choices. We conduct expert meetings with teachers and analyze performance assessment tools that are actually used by teachers to look into how and why the performance assessment, which has been implemented in Korea's education for roughly fifteen years, has failed to take root in Korean classrooms. We found an enormous gap between institution and the classroom concerning performance assessment, which reflects the fact that policy makers, with their top-down approaches, have focused excessively on institutional changes while ignoring actual changes in classrooms. We suggest that Korea should take a bottom-up gradual approach that supports actual changes in the classroom in order to transform its assessment system from multiple choices to performance assessment.

Ju-Ho Lee: KDI School of Public Policy and Management, 85 Hoegiro Dongdaemun Gu, Seoul 130-722, Korea. Tel: 82-2-3299-1043; Fax: 82-2-3299-1240; Email: jhl@kdischool.ac.kr Ryoo: Kookmin University, Email: ryoo@kookmin.ac.kr Sam-Ho Lee: Korea University, Email: samho@korea.ac.kr The Original version of this paper was presented in October 2014 at the KDI Workshop on Human Capital Policy, held in Seoul, Korea. We thank Jung-Hee Choi and Ki-Jung Kwon for excellent research assistance. We received helpful comments from Richard Freeman, Patrick Griffin, Eric Hanushek, Soonhee Kim, Tae-Wan Kim, Sung-Joon Paik, Yoon-Soo Park, Daehee Song, Petra Todd. Shun Wang.

1. Introduction

From primary and secondary education to college entrance, Korean education heavily relies on multiple choice questions for student assessment. The problems that arise due to such strong reliance on multiple choices are severe. Most of all, students are unable to sufficiently develop diverse skills other than cognitive skills tested by multiple choice questions. The 21st century requires an education that develops diverse skills, not one with excessive emphasis on cognitive skills that are measurable by multiple choices. Fullan and Langworthy's (2014) proposal of the 6 C's, which are citizenship, communication, critical thinking and problem solving, collaboration, creativity and imagination, reflects a rising consensus on major skills required in the 21st century among policy makers and scholars around the world. With excessive dependence on multiple choice questions, it would be very difficult for Korea's education to develop the 21st century skills.

Another important problem of Korean education due to its heavy dependence on multiple choices is that even people with great potential in other areas focus only on developing cognitive skills. Horizontal differentiation, which allows people to develop different skills, becomes severely weakened, while the problem of vertical differentiation, with everyone focused solely on multiple choices becomes worse. Such stricter vertical differentiation of education leads to the problem of the education bubble, where increased expenditure in education, such as private education, fails to increase human capital (Lee, Jeong, and Hong, 2014).

Among various assessment methods, measurement of student ability can generally be divided into assessments that use (1) multiple choice questions or (2) performance-based tasks or questions (AERA, APA, and NCME 2014; Livingston, 2009; Lissitz, Hou, and Slater, 2012; DeCarlo, Kim, and Johnson, 2011). The dichotomy in these two forms of assessment methods rely heavily on the particular cognitive or behavioral aspects of students' ability being measured (Dwyer, 2008). While many large-scale assessments use multiple choice questions to assess what students "know" (knowledge) or "knows how" (application of knowledge), assessments that require students to "show how" (demonstration of knowledge) or "does" (translation of knowledge to actual performance) cannot be easily measured using multiple choice questions; they require the examinee to construct their response using a performance assessment, rather than select the correct answer from a list (Miller, 1990).

This paper is constructed to study the following three important questions pertaining

to performance assessment in Korean education. The first is why performance assessment is important for Korea's education. The second is how successful performance assessment has been in Korean classrooms. The third is how we can foster performance assessment.

First, we apply a theoretical model to explain why performance assessment is important. Through the model it will be shown that when student assessment is focused solely on a single aspect of human capital, human capital investment centers only on cognitive skills that are testable through multiple choices. This leads to vertical differentiation of education, which results in inefficiency of human capital investment.

Second, we analyze 67 performance assessment tools that are actually used by teachers to see why the performance assessment, implemented fifteen years ago, has become distorted and unable to fulfill its original purpose, and focus on the fact that nearly none of the major conditions for performance assessment are being met in actual classrooms despite fifteen years of institutional existence. In particular, nearly one-third of the performance assessment tools failed to have any sort of 'performance' element, and more than half of the samples displayed very low openness and evaluative authority. Additionally, approximately 75 percent of the samples were composed of non-performance elements including simple comparison of answers or counting the number of answers.

This paper emphasizes the need to focus on the enormous gap between institution and classroom when it comes to performance assessment. We should reflect on whether such a gap implies that policy makers and analysts have focused excessively on institutional changes while ignoring actual changes in classrooms. The authors of this paper conducted a series of expert meetings with ten school teachers¹ and made field visits to schools to fully comprehend the actual situation, and based on such efforts developed a national strategy for expanding performance assessment. This paper proposes a 'gradual bottom-up approach' that provides active support for bottom-up changes in actual classrooms while pursuing gradual

¹ Ki-Jung Kwon (Ohjung Elementary School), Ji-Hye Kim (Cheongdam Elementary School), Mi-Yeong Na (Seoul National University Middle School for Girls), Jin-Soo Park (Surak High School), Kyung-Pyo Ban (Shinsang Middle School), Joon-Hyun Seong (Miseong Elementary School), Seong-Bum Won (Yangmok Elementary School), Jung Yoon (Neungdong Elementary School), Dae-Ho Lee (Cheongryang High School), Kyung-Ah Jung (Sangkyung Elementary School) participated very actively in the meetings with their valuable ideas and opinions.

yet consistent changes in institutional elements like the national college admissions system and the national education curriculum.

The composition of this paper is as follows. In Chapter 2, through a theoretical model we emphasize why Korea's education needs performance assessment. In Chapter 3, the results of surveys reveal in detail how the performance assessment, implemented fifteen years ago, is failing to be properly conducted. In Chapter 4, we propose five strategies for fostering performance assessment in classrooms through a gradual bottom-up approach. Chapter 5 concludes.

2. Why Does Korea Need to Reform the Assessment System?

Education, as an investment for human capital, would be more effective if it can accommodate various forms of human capital. By various forms of human capital, we mean the distinction between cognitive and non-cognitive skills a la Heckman and his colleagues on one hand, but also the distinction among various forms of skills within either cognitive or non-cognitive classifications. The category of cognitive skills include a wide spectrum; from acquiring and understanding the provided knowledge to retrieving the necessary knowledge in a new context and applying it. Even though they have the same label ‘cognitive skills’, they may require different sets of training and assessment. Korean education, however, seems to focus on a very narrow sense of cognitive skills; rote memory of the information. Multiple choice questions are a typical type of assessment of this skill.

Though many reform agendas were proposed and implemented to induce the diversification of education, the effects of those reforms were limited as they seem to fail to change the teaching and assessment methods in the classrooms. New types of school institutions were introduced such as Innovation Schools and Autonomous Private High Schools. The change in universities’ admissions policies were supported to enhance the role of subjective evaluation of Admissions Officers. With new institutions, however, educational experimentation is relatively scarce. The Admissions Officers, who were expected to utilize other merits than academic achievement, seemed to have difficulty in obtaining the assessment materials to evaluate those merits.

Though institutional change may catalyze the educational reform, the effective reform does not seem to be achievable without the development of new educational content and its assessment method.

In particular, the assessment method seems to play a critical role in attaining educational reform. Education to cultivate various types of human capital cannot be sustained unless those newly cultivated skills are assessed and appreciated. The current assessment system, with its lopsided emphasis on narrowly defined cognitive skills, seems to hinder effective educational reform.

In this section, we propose a simple logical model to explain 1) why the current assessment system seems to induce the concentration of human capital development for a certain type (we call this concentration of investment and severe competition in a single dimension “vertical differentiation” in contrast to “horizontal differentiation” which spreads

competitive pressure to multiple dimensions.) and 2) why the inefficiency in human capital development ensues.

2.1. Model

A. Agents

There are a unit mass of agents who can invest in two types of human capital; K_1 and K_2 .

B. Human Capital

Two types of human capital, K_1 and K_2 typically denotes cognitive and non-cognitive skills. As mentioned before, there are finer classifications of skills under cognitive and non-cognitive skills. The assumed two types are just simplification of various types. Thus this distinction may represent the difference between narrowly defined cognitive skills such as accumulation of information and other dimensions of cognitive skills such as inference and creativity.

We assume there are two steps of investment in human capital. An agent starts with an initial investment, and then the deepening of the skill follows. Initial investment is required before further deepening of the skill.

For simplicity, we assume that an agent will choose one type of human capital at the initial investment stage. Of course, an agent in reality will invest in multiple types at the same time. However, the depth of investment will be different in skills and this specialization is modeled as a choice in the initial investment stage.

The costs of initial investment in each type of human capital is denoted as a pair (c_1, c_2) , and these pairs are uniformly distributed over $[0, 1]^2$. An agent observes his pair of costs and chooses which type he specializes in. We may assume that there will be an additional cost other than $c_1 + c_2$ if an agent invests in both types of human capital and that this additional cost is huge. Then an agent only chooses one type to invest in.

After this initial investment choice, an agent decides whether to deepen the chosen skill further. An agent is assumed to be uncertain about his cost of deepening the skill. Only after initial investment, an agent realizes whether he is apt for the chosen skill. In modeling terms, cost of deepening σ is ex-ante unknown and uniformly distributed over $[0, \bar{\sigma}]$.

C. Productivity and Assessment of Human Capital

We assume, for simplicity, that the productivities of two types of human capital are the same. Thus there is no reason to have different levels of investment in two types as long as their productivities are concerned.

Initial investment will raise the productivity by 1. Since we assumed that the cost of initial investment is less than 1, this means that everyone will choose one type or the other for initial investment. Thus we exclude the possibility that an agent falls out in the initial investment stage.

Additional productivity ν accrues with the deepening of human capital. That is, the productivity is 1 if only initial investment is made, and $1 + \nu$ if the skill is further deepened.

It is observed which human capital an agent specializes in. That is, it is known which type of human capital is chosen in initial investment. However, it is imperfectly observed whether one excels in his chosen skill. We treat this observation of the deepened skill as a role of assessment system. With probability p , it is observed whether the chosen skill is deepened or not. With probability $1 - p$, it is not observed. The parameter p represents how accurate an assessment system is. If an assessment system is more developed along one type of human capital than the other, it will be more accurate for that human capital. Here we assume that assessment of K_1 is more accurate than that of K_2 , or $p_1 > p_2$.

D. Reward for Human Capital Investment

Reward is given to the productivity increase by human capital investment. If the model is applied in the competitive labor market context, wage is given at an expected productivity. If the reward is an educational one such as advancement to higher educational level, it is given in response to educational performance, which is, in our model, productivity increase. The reward is dependent on the result of assessment. If further development is correctly perceived, then either 1 or $1 + \nu$ is given depending on the deepening of the skill. If it is not perceived due to an imperfect assessment system, the reward is the expected productivity.

E. Time line

1. Initial investment is made for either K_1 or K_2 considering the cost pair (c_1, c_2) .
2. When the cost of further development σ is realized after the initial investment, an agent decides whether to deepen the skill or not.

3. The skill is assessed by the accuracy p , and the reward is given as a result of the assessment.

2.2. Analysis

Two decisions are made in the model. First, an agent chooses the type of human capital for initial investment. Second, an agent decides whether to deepen the chosen skill or not. Analysis is done backward from the second decision.

A. Whether to Deepen the Chosen Skill

The deepening of the skill takes place if the expected reward increase is greater than the cost. Thus the skill is deepened when

$$Vp \geq \sigma.$$

Let q be the probability that an average agent deepens the skill or the portion of agents who deepens the skill. Since we assume that σ is uniformly distributed,

$$q = \frac{Vp}{\sigma}$$

That is, a higher p , or a more accurate assessment system induces more development of that skill.

We can evaluate the expected return when each type of human capital is chosen in the initial investment stage. Before σ is realized, the probability that an agent further deepens the chosen skill is q and the expected return in that event is $Vp + (1-p)qV + 1 - E[\sigma | \sigma \leq Vp]$. When the deepened skill is correctly assessed with probability p , the increase in reward is V . When the deepened skill is not observed with probability $1-p$, the expected reward is qV . The cost σ is paid regardless of whether the deepened skill is correctly assessed or not. On the other hand, an agent does not deepen the skill with probability $1-q$, and the expected reward in that event is $(1-p)qV + 1$. Let π be the expected return after initial investment stage. Then,

$$\begin{aligned} \pi &= qVp + (1-p)qV + 1 - qE[\sigma | \sigma \leq Vp] \\ &= qV + 1 - qE[\sigma | \sigma \leq Vp] \end{aligned}$$

Using the uniform distribution of σ , it is simplified as

$$\pi = \frac{V^2 p}{\sigma} \left(1 - \frac{1}{2} p\right) + 1$$

The expected return π is dependent on the accuracy of assessment p , and it is increasing in p . The higher p induces more development of the skill, or higher q . Since an agent deepens the skill if its return is higher than the cost, higher q means better return.

B. Which type of human capital to choose in the initial investment stage

An agent will choose which type of human capital to specialize in weighing the expected return and the cost. Thus an agent chooses K_1 when

$$\pi_1 - c_1 \geq \pi_2 - c_2 \text{ or } c_1 \leq c_2 + (\pi_1 - \pi_2)$$

If both types are assessed with the same accuracy ($p_1 = p_2$), the expected return is the same ($\pi_1 = \pi_2$) and an agent's given suitability for each type of human capital solely determines which human capital he specializes in. That is, agents who are more suitable for K_1 ($c_1 < c_2$) choose K_1 , and other agents choose otherwise. However, difference in assessment accuracies leads to difference in expected returns, and thus the skewed investment decision toward K_1 . Considering the same productivity of two types of human capital, it would be more efficient when both types are equally specialized with the specialization decision being driven by agents' suitability for two types (c_1, c_2) . With assessment system focused on one type of human capital, however, investment is more skewed toward that type.

We can call this phenomenon 'vertical differentiation' à la industrial organization theory; that is, concentrated specialization on one type of human capital and competition to perform better in that dimension. To be competitive in a market, a firm may try to excel in quality of a given dimension or to differentiate its product in a different dimension. We call the former case 'vertical differentiation' and the latter 'horizontal differentiation'. If we acknowledge that there are multiple dimensions of human capital and people are different in their suitability for different dimensions, horizontal differentiation is the more desired type of competition in education. However, if human capital investment is concentrated in a specific type of human capital, competition in that dimension becomes severe.

The degree of vertical differentiation gets higher as the difference between p_1 and p_2 gets larger. If the difference is too large, then every agent will invest in K_1 regardless of their

cost type (c_1, c_2) .

The degree of vertical differentiation also gets higher when the return to educational investment is larger. As V gets larger, $\pi_1 - \pi_2$ gets larger and more concentration on K_1 ensues.

The degree of vertical differentiation gets even higher when tournament competition such as university admissions is involved to determine educational return. If we incorporate a tournament competition in our model, only a fixed portion will be selected based on the assessment. Then agents who are perceived to deepen their skills will be selected first. Among those whose deepening of the skill are not, the ones specializing in K_1 are selected first. When it is not known whether the skill is deepened or not, the expected productivity should be considered. Since more agents deepen their skill with K_1 , they are preferred. This will lead to concentration of investment in K_1 .

2.3. Discussion and Implication

This analysis may give a clue on why the policy efforts for “education diversification” had difficulty in achieving their aims (Lee, 2014). Allowing various school types such as Autonomous Private School or the support to enhance the role of Admissions Officers in university admissions are all efforts to accommodate various types of human capital in education. These policies are aimed at fostering horizontal differentiation. However, it is debatable whether those institutional changes were successful in achieving horizontal differentiation. Our analysis suggests that it is not certain whether institutional change, without the right assessment system in place to allow room for differentiation, may lead to either horizontal or vertical differentiation. The critical element is the assessment system. In order to strengthen horizontal differentiation in education, it is necessary to develop a performance assessment system to improve the accuracy of the assessment for relatively neglected dimensions of human capital.

3. How Successful Has Performance Assessment Been in Classrooms?

Given the aforementioned importance of diversification of educational contents, methods and assessment, it must be pointed out that the Korean government has not been completely ignorant to such needs. In fact, the government has attempted one important reformation in assessment by implementing the Performance Assessment Policy to equip the next generation with new sets of skills since the late 1990s. Then, it is now the rightful time to examine the effectiveness of the policy and diagnose any possible areas of policy revision to further facilitate the reform. With such intent in mind, this section asks the following specific research questions: how successful has it been, what is the current practice of performance assessment in Korean schools, and what are the specific areas in need of improvement for effectively implementing performance assessment in the Korean educational context?

As one important assessment method for qualitative aspects of educational contents, whether they are traditional cognitive materials or newly suggested competency sets, performance assessment implemented in the right way can be an effective tool to spark changes in educational practices. In order to realize successful education reform through utilization of the Washback effect (Messick, 1996), namely, revamping assessment to bring about changes in learning and teaching practices, understanding whether the performance assessment is being correctly implemented in classrooms is crucial. With reference to the key elements of performance assessment that relevant literature discusses as well as the general Korean educational culture in which performance assessment is implemented under the lead of the governmental recommendations as portrayed above, we begin with four particular conditions that are suggested below as essential norms to which we can assert whether performance assessment is well put into practice reflecting its original intent.

3.1. Conditions of Performance Assessment

A. Performance Condition

This first condition is rather tautological as it basically means that ‘there should be performance in performance assessment.’ This condition stipulates that performance assessment must contain observable learning activities or ‘learning performance’ that can demonstrate types, qualities and levels of the performer’s learning. This condition may not be

applicable to all contexts, especially where the basic understanding of the nature of performance assessment is widely established. Yet, after experts' close examination of collected performance assessment tools being used in Korean schools, it has been repeatedly pointed out that in most of the collected cases there is no actual performance of learning, as traditional paper-pencil tests are still in use, now only with a different name, 'performance assessment.' In such cases, even if students, teachers and parents actually use the term 'performance assessment' quite frequently as teachers have to administer it implementing the governmental recommendations, and students and parents are using the term as the teachers are using it, there is found a very low degree of real learning performance in their tools as will be shown below.

B. Validity Condition: Content validity

For any type of assessment, it is critical to assure that an assessment tool examines the items that are *core* to the subject or field rather than the ones that are superficial, trivial or unessential. For developing performance assessment, identifying the core elements, or "key performance indicators," is the most significant step, as a professionally acceptable composition of key performance elements may make it possible to construct a mathematical equation to engender a single digit assessment result that can represent the overall quality and level of a target performance (Lavy, 2014). In addition, it is of utmost importance to consider a proper *balance* or weighted differentiation among the content elements for performance assessment, not allowing certain elements to over-represent the overall quality of learning.

Such balancing among assessment items must give due consideration to learning contexts, for different groups of learners may have different needs, motivations, cultural perspectives or interests of learning even about the same topic. The "contextual authenticity" of performance assessment (Yan, 2006) then requires a performance assessment tool to contain core materials, with the right compositional balance reflecting learners' contextual needs, motivations, perspectives and interests. The content validity of performance assessment can then be determined by how accurately it measures context-free core subject contents and how truthful it reflects relevant contextual demands, with any biased materials minimized (Conley, 2013: 17).

C. Flexibility Condition: Open-endedness

One outstanding characteristic of performance assessment is a wide range of decisions that can be made in the process of assessment even without pre-determined rules, standards and expectations, whenever it is deemed necessary and appropriate by the evaluator. Performance assessment examines not only the end results of learning activities, or namely ‘answers,’ but also each step of the processes in which students’ thinking, exploration and critical evaluation occur. In assessing both processes and answers, it is of pivotal importance to allow ‘scoring divergences’ whenever any element advancing understanding of the materials are found either in the processes or answers. Given the procedural, lively and improvisational nature of all types of performances, ‘open-endedness’ in their assessment is conceptually natural and thus ought to be implemented in practice (Yan, 2006).

D. Evaluative Authority Condition

The aforementioned flexibility condition leads to an additional condition that has a significant practical implication, namely the condition that an evaluative subject, usually a teacher, should be able to exercise a meaningful degree of authority in determining flexible application of assessment standards or summing up the overall results using sub-scores and relevant information such as students’ background and local culture. In other words, when there is no practical authority a teacher can exercise in presenting final assessment results, it cannot be properly called performance assessment, as the assessment results might have been calculated based upon fixed formula and thus the kind of learning that such an assessment would bring about would always occur strictly abiding by fixed norms of evaluation, leaving no room for active performance, creative divergence and critical exploration of learning materials.

It is generally accepted that there can be two possible methods of scoring in performance assessment: the holistic method that gives a single score based upon an overall evaluation and the analytic method that breaks down into a multiple number of sub-items, assigns scores for each sub-item and sums up, if necessary, those sub-scores to produce one assessment result for the overall performance. While the choice between the two methods should be left as the teacher’s own practical authority, as both show similar levels of statistical validity and reliability, it is generally suggested that the analytic method is slightly

more useful in distinguishing students' levels (Chi, 2000). It is then suggested that practical decisions regarding flexible summing up of assessment results or the choice of scoring methods such as the holistic and analytics methods ought to be left to teachers' own professional expertise and authority.

3.2. Research Methods

In order to examine the current usage of performance assessment tools, a total of 67 performance assessment tools (problem sheets, scoring rubrics, assessment plans, etc.) have been collected from five primary school teachers and five secondary school teachers. Though a multiple number of research meetings have been held with the same numbers of primary and secondary teachers, more primary school cases were collected than those of secondary schools (See Table 1). The cases were from diverse subject matters including Korean, Mathematics, English, Social Studies and Sciences, and from all grade levels.

Table 1. Numbers of Collected Performance Assessment Tools

	Primary School (1 st -6 th grades)	Middles School (7 th -9 th grades)	High School (10 th -12 th grades)
Korean	13		2
Math	9	2	
English	3	8	4*
Social Studies (including History and Ethics)	15		3
Science (Physics, Chemistry, Biology and Geology)	7		1
Total	47	10	10

* One Chinese subject is included.

Once the cases were collected, a research team comprised of six experts (2 in educational assessment, 2 in curriculum studies and 2 in economics) reviewed collected cases and relevant literature in order to establish standards to which each case was to be categorized. As four meetings between the research group and the teachers who collected the tools had already been held before the research team's review, each researcher was quite familiar with the types of tools and related educational issues. As a result of researchers' review of the collected cases with reference to teachers' opinions, four specific conditions for a good performance assessment tool have been set up as discussed above: the performance condition, the validity condition, the flexibility condition, and the evaluative authority condition. In addition, researchers further analyzed the scoring patterns that appeared in some of the collected cases as they could represent additional sort of limitation in current usage of performance assessment.

3.3. Research Findings

The analysis on the collected 67 performance assessment tools reveals two general trends. First, the overall quality of performance assessment tools being used in primary schools is low compared to that of middle and high school samples. And second, the estimated quality of assessment tools is higher in humanistic subject matters such as Korean, English, and Social Studies than hard scientific subject matters such as Mathematics and Science (See Table 2). It is not readily recommendable, however, to generalize with this small size of samples, particularly because most of the collected tools from secondary schools are linguistic subject matters such as Korean and English, of which tools also show better quality in primary school samples. And yet, it is quite conspicuous even with this small collection that many of the tools being used in Korean schools could be improved further.

Although the sample size of performance assessment tools may be lacking in number, the results displayed a level of consistency high enough to provide implications for ways to improve the performance assessment. In particular, by focusing our analysis on how well the samples fit the four conditions for performance assessment mentioned above, we attempt to provide a basis for further in-depth discussions to become possible.

Table 2. Analysis of Collected Performance Assessment Tools

		Primary School (1 st -6 th grades)		Middles School (7 th -9 th grades)		High School (10 th -12 th grades)	
Performance Condition	High	K6, C9, E1	34%	M2, E7	90%	K2, E2, C3, S1	80%
	Medium	K2, M5, C4, S3	30%				
	Low	K5, M4, E2, C2, S4	36%	E1	10%	E2	20%
Validity Condition	High	K5, C9, E1	32%	M2, E7	90%	K2, E2, C3, S1	80%
	Medium	K4, M9, C4, S7	51%	E1	10%	E2	20%
	Low	K4, E2, C2	17%				
Flexibility Condition	High	K4, C4, E1	19%	M2, E7	90%	K2, E2, C3, S1	80%
	Medium	K3, M2, C7	26%			E1	10%
	Low	K6, M7, E2, C4, S7	55%	E1	10%	E1	10%
Evaluative Authority Condition	High	K5, C4, E1	21%	M2, E7	90%	K2, E2, C3, S1	80%
	Medium	K2, M2, C6	21%			E1	10%
	Low	K6, M7, E2, C5, S7	58%	E1	10%	E1	10%

* Alphabets stand for subject matters with K for Korean, M for Mathematics, E for English, C for Social Studies and S for Science. For instance, K5 means five cases in Korean belong to the box.

A. The Performance Condition

The government has been asking schools to increase the proportion of performance assessment in overall assessment composition for school grades, but it seems that in-service teachers do not accurately grasp the nature of performance assessment, as approximately one-third of all samples collected from primary schools involve a low degree of performance (See Table 2), even if it is ‘performance’ assessment. In fact, most of the cases in the ‘low’ category did not have any element of learning performance at all. For instance, in a science assessment sheet (See Figure 1), the item asks students to fill out the correct names of each part of a spring balance, and there are neither learning activities nor performance requested for the problem in spite of the title on top of the sheet being ‘Science Performance Assessment Sheet.’

Figure 1. Performance Assessment Tool in Use: 4th grade Science

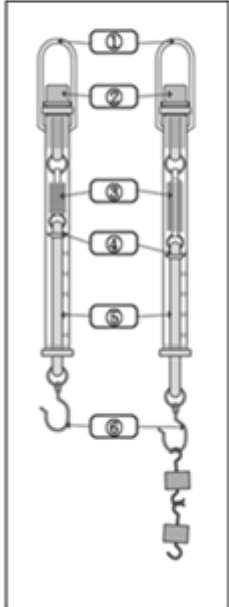
Science Performance Assessment Sheet

Grade: Class: Number: Name:

Chapter	1. Measuring Weight	Lesson		Date	Month: Day: Yr:
Task				Evaluator	(Signature)

Task

1. Write the name of the part of the spring inside the parentheses according to the number



①()
②()
③()
④()
⑤()
⑥()

This is a critical finding because performance assessment is sometimes defined as all kinds of assessment excluding standardized quantitative tests (Gripps, 1994), as standardization of a test means that learning performance is excluded and only final responses to problems are counted and statistically treated; thus, a performance assessment tool with no performance involved is not an actual performance assessment tool. What is not performance assessment is then being used for approximately one-third of alleged performance assessment in Korean schools.

B. The Validity Condition

The content validity of collected samples was examined to be high with only 17% of primary school samples in the low category. This can be perhaps accounted for by the high academic quality of Korean school teachers compared to the counterparts in the international community. As the validity condition stipulates that an assessment tool should assess the genuine capability a subject matter is supposed to foster, however, the percentage should be ultimately further lowered.

As an example of a low validity performance assessment tool, one interesting English case is worth our attention here. What an English language class is supposed to teach is not only the memorization of numerous words but also other linguistic capacities such as fluency, pronunciation, intonation, speech attitude and cultural understanding of English speaking countries; yet, in this example (Figure 2), the scoring is done solely by the number of words a student is using in the speech (score 35 for 120 words and more, score 31 for 110~119 words, and so on). Besides the practical difficulty in counting the number of words being spoken, this type of assessment and scoring has a greatly low content validity for the English subject matter, as it does not assess what it is supposed to assess for an English class. If students are learning English to excel in this type of performance assessment, what they should be concerned about is the memorization and reproduction of as many words as possible with little attention to other important aspects of linguistic skills.

Figure 2. Performance Assessment Tool in Use: Middle School English

Speaking Assessment Scoring Criteria (Revised)

1. Voice, Attitude Deduction (-1 point)

2. Scoring Distribution

120 words or more ~	35 points
110 words or more ~	33
100 words or more ~	31
90 words or more ~	29
80 words or more ~ less than 90 words	25
60 words or more ~ less than 79 words	20
40 words or more ~ less than 59 words	15
20 words or more ~ less than 39 words	10
10 words or more ~ less than 19 words	5
1 word or more ~ less than 9 words	3

3. Non-participation 0 points

C. The Flexibility Condition.

More than 50% of the samples collected from primary schools are diagnosed with low flexibility in assessment items. That is, students did not have enough space or opportunity to think freely with a wider range of ideas and interpretations, but were simply asked to reproduce what they were supposed to remember or come up with correct answers based upon what they had learned in class. In these cases, there were definitely right or wrong answers and there was no room for divergent thinking, ideas or challenges. On a third grade Korean performance assessment sheet (Figure 3), for example, which classifies itself as an essay-type assessment at the top, there is no space where students can write their own words, thoughts, or essays; what students find there is only space for drawing correct lines and circles. Performance assessment is not a panacea for all types of learning, and that is why performance assessment is recommended by the government to take up only about 30-40% of all assessment, with the remaining 70-60% filled with traditional paper-pencil tests to examine basic information and hard knowledge. Perhaps in this transitional period, however, due to the low level of understanding of the intent and nature of performance assessment, performance assessment assesses what traditional assessment can do while leaving what only it can do untouched.

Figure 3. Performance Assessment Tool in Use: 3rd grade K

(Korean) Essay Assessment		Grade 3, Class 2,									
		Student No. ()									
		Name:									
Chapter	4. Correct Usage of Honorifics										
Content	Knowing when to use honorifics and using them correctly										
<p>1. Circle the correct honorific inside the parantheses and draw a line to connect it with the correct method on the right</p> <table border="1"> <tr> <td>(1) "어머님, 간식 잘 (먹었다 , 먹었습니다)."</td> <td rowspan="8"> <input checked="" type="radio"/> ㉠ "-시-"를 넣었다. <input type="radio"/> ㉡ "-계"나 "-게서"를 붙였다. <input type="radio"/> ㉢ 존칭을 "-습니다"로 끝냈다. <input type="radio"/> ㉣ 높임의 뜻이 있는 낱말을 사용하였다. </td> </tr> <tr> <td>(2) 선생님(ㅁ , 어제) 정찬을 물었다.</td> </tr> <tr> <td>(3) 만년필의 주인은 할아버지(였다 , 셤다).</td> </tr> <tr> <td>(4) "아버지, (물어볼 , 여쭙어 볼) 제 있어요."</td> </tr> <tr> <td>(5) "삼촌, (고맙다 , 고맙습니다)."</td> </tr> <tr> <td>(6) "아버님, 별리 (보고 , 뵈고) 싶어요."</td> </tr> <tr> <td>(7) 이요, (만촌 , 만도신) 제 뽕뽕인가요?</td> </tr> <tr> <td>(8) 교통경찰 아버지(ㅁ , 어제) 편지를 썼다.</td> </tr> </table> <p>2. Mark the correct phrase with a circle</p> <p>㉠ 존경 : 아주머니, 저희 어머니께서 꽃갑 통 빌려 달라고 (했니다 , 하십니다). ㉡ 아주머니 : 그라, 송년회 (사용한 , 사용하신) 뒤에 붙여 달라고 전해 드림.</p>			(1) "어머님, 간식 잘 (먹었다 , 먹었습니다)."	<input checked="" type="radio"/> ㉠ "-시-"를 넣었다. <input type="radio"/> ㉡ "-계"나 "-게서"를 붙였다. <input type="radio"/> ㉢ 존칭을 "-습니다"로 끝냈다. <input type="radio"/> ㉣ 높임의 뜻이 있는 낱말을 사용하였다.	(2) 선생님(ㅁ , 어제) 정찬을 물었다.	(3) 만년필의 주인은 할아버지(였다 , 셤다).	(4) "아버지, (물어볼 , 여쭙어 볼) 제 있어요."	(5) "삼촌, (고맙다 , 고맙습니다)."	(6) "아버님, 별리 (보고 , 뵈고) 싶어요."	(7) 이요, (만촌 , 만도신) 제 뽕뽕인가요?	(8) 교통경찰 아버지(ㅁ , 어제) 편지를 썼다.
(1) "어머님, 간식 잘 (먹었다 , 먹었습니다)."	<input checked="" type="radio"/> ㉠ "-시-"를 넣었다. <input type="radio"/> ㉡ "-계"나 "-게서"를 붙였다. <input type="radio"/> ㉢ 존칭을 "-습니다"로 끝냈다. <input type="radio"/> ㉣ 높임의 뜻이 있는 낱말을 사용하였다.										
(2) 선생님(ㅁ , 어제) 정찬을 물었다.											
(3) 만년필의 주인은 할아버지(였다 , 셤다).											
(4) "아버지, (물어볼 , 여쭙어 볼) 제 있어요."											
(5) "삼촌, (고맙다 , 고맙습니다)."											
(6) "아버님, 별리 (보고 , 뵈고) 싶어요."											
(7) 이요, (만촌 , 만도신) 제 뽕뽕인가요?											
(8) 교통경찰 아버지(ㅁ , 어제) 편지를 썼다.											

D. The Evaluative Authority Condition

While the flexibility condition aims at preserving space for students' free, creative and critical learning, the evaluative authority condition purports ultimately to increase teachers' professional authority. As mentioned earlier, there is a close conceptual link between the flexibility of performance assessment and evaluative authority, for flexibility is possible only when a teacher has educational expertise to be a judge of various processes and answers; where there are only mechanically pre-fixed norms, there is no need of a judge. Among the collected samples, more than 50% of primary schools' performance assessment tools were low in evaluative authority, leaving no other options for teachers to evaluate otherwise than pre-determined standards, mostly simple answers. In addition, peculiar patterns of *scoring* must be discussed in regard to this condition, because it is likely that non-performance scoring practice is the cause of lack in evaluative authority.

Four general types of scoring patterns have been witnessed in the analysis of the collected performance assessment tools. The first type is the ‘correct answer’ type, for which a teacher needs to compare students’ responses with pre-recorded answers. There is neither learners’ performance, nor room for a teacher’s exercise of educational evaluative expertise possible in this type of scoring. The second is a type of ‘counting,’ in which scores are assigned based upon the *number* of answers, with little reference to the quality, creativity and diversity of answers. For instance, Figure 4 shows a fourth grade Social Studies case in which a score of 8 is assigned for four correct answers, 6 for three correct answers, etc. For another example, Figure 5 is a high school Social Studies case (Economics) in which deduction in scores occurs as the number of concepts presented in answers decreases.

Figure 4. Scoring Pattern: Counting

※ Evaluation Criteria

Problem Number	Evaluation Criteria		
3-(1)	Assessment Element	Drawing a map according to the given conditions (8 points)	
	Score Distribution	8	All four answers match the example answers
		6	Three answers match the example answers
		4	Two answers match the example answers
		2	One answer matches the example answer
		0	No answer matches the example answer or no questions were answered

Figure 5. Scoring Pattern: Counting

Detailed content and criteria for Performance Assessment of 'Economics'				
Assessment Area	Score Distribution	Detailed Assessment Content	Assessment Criteria and Method	
Development Assessment	100 points	A written form of development assessment is conducted on the major economic themes and concepts studied in class	Under the agreement of the teacher in charge, decide the detailed assessment criteria for each theme and concept	
			After announcing the 20-30 major economic theme and concept selected (5 themes or concepts per paper, 4-5 papers), randomly select one paper and conduct development assessment for all classrooms under identical conditions	
			For writing 3 concepts	
			A(20 points)	All essentials conditions for each concept are provided
			B(15 points)	1 essential condition is missing from the answer provided
			C(10 points)	2 essential condition is missing from the answer provided
			D(5 points)	The wrong answer is provided
			For writing 2 concepts	
			A(20 points)	All essentials conditions for each concept are provided
			B(15 points)	1 essential condition is missing from the answer provided
C(10 points)	The wrong answer is provided			
			In the case a limit is put on the required answers, assessment is made for only the number of answers required (if the limit is 3, the answers in excess of 3 are not assessed)	
			If a student is unable to partake in the performance assessment due to absence from class for justifiable reasons, an opportunity for re-assessment is provided.	
			If a student is unable to partake in the performance assessment due to absence from class for unjustifiable reasons, a single opportunity for re-assessment is provided, but if the student rejects this opportunity a score of '0' is given.	
			In the case of grammatical errors, if the error is essential to the answer provided, 1 point will be deducted	

The third type of scoring found in the samples is assigning scores based upon very simple and abstract norms. The scoring rubrics collected from primary schools, for instance, states in the following pattern; if a student responded *very well*, then a score of 3; if she responded *properly*, then a score of 2; if a student did *not* respond *well*, then a score of 1; if there was no response, then a score of 0. There are no examples of responses, no detailed explanation or notes for the teachers themselves, but the stated rubrics are so simple that reliability (coherence in assessment and scoring) cannot be expected. The fourth type, which is the most acceptable type for performance assessment, has a set of complex statements, embodied in the rubric, enough to embrace unexpected but cognitively and non-cognitively sound responses under a teacher's own evaluative authority. It was only about 25% (See Table

3) but certainly existent among the collected samples, which indicates that Korean teachers are capable of creating and administering performance assessment in the right way.

Table 3. Scoring Types of Performance Assessment Tools in Use

		Primary School (1 st -6 th grades)		Middles School (7 th -9 th grades)		High School (10 th -12 th grades)	
Scoring Types (only the cases with scoring information)	Correct Answer Type	K3, M8, E2, C3, S4	45%	E1	10%	E1	8%
	Counting Type	K4, C8, S1	28%	E1	10%	E1, C2	25%
	Simple Quality Check Type	K1	2%				
	Complex Quality Check Type	K5, M1, E1, C3, S1	25%	M2, E6	80%	K2, E2, C3, S1	67%

The critical shortcoming of the former three types of non-performance scoring (the correct answer type, the counting type and the simple quality check type) is not only the lack of performance considered in scoring, but such a scoring can be conducted by anybody who does not have any expertise in teaching or evaluation, because what needs to be done is simple comparison of students' answers with correct answers or just counting the number of presented answers. This type of scoring is not recommendable for the advancement of educational practices, and a reform should be pursued in a way in which people involved in the field can develop their expertise and professionalism, which will ultimately translate into the betterment of the practice itself. Thus, the most desirable scoring practice for performance assessment should enable teachers to utilize their educational wisdom and experiences to the extent that rightfully high scores can be assigned to students' unexpected but creative, diverse,

and positively critical responses. With only simplistic rubrics that anyone on the streets can use to evaluate students' works, assessment, thus teaching and learning, of various cognitive and non-cognitive competencies required for the future society is out of reach.

3.4. Discussion and Implication

It is the rightful time to explore new educational practices for a new era as the world changes rapidly and global changes have now grave influence upon local communities. For implementing new practices of learning and teaching, assessment reform is considered to be one of the most effective means. The Korean education has in fact not been neglectful about such changes and demands, and has pushed forward to the new direction, adopting a number of strategies including an assessment reform. It has been almost fifteen years since the Performance Assessment Policy has been put into practice, mostly in the form of governmental administrative guidelines. And nowadays 'performance assessment' has become a jargon in the field as teachers, students and parents frequently use the term to denote a type of assessment; yet, in many cases they use it in the wrong way. It has been seen through experts' close examination and analysis of 67 actual performance assessment tools that are being used in Korean schools that performance assessment has in general serious weaknesses in the performance condition (that there should be performance involved in performance assessment), the validity condition (that performance assessment should assess core contents of the subject), the flexibility condition (that performance assessment should be open-ended for students' unexpected but positive processes and responses) and the evaluative authority condition (that teachers should be able to exercise evaluative expertise to acknowledge students' positive learning performances). More in particular, about one third of the collected samples did not have any performance involved in the assessment, and more than half of the collected cases had very low flexibility and evaluative authority. Scoring patterns teachers are using are found problematic as well, because approximately 75% of scoring practices are non-performance type, only examining correct answers or the number of answers.

In sum, we found huge gaps in performance assessment between what should be done as defined by government policy and what is actually being done in classroom practice. However, we cannot simply blame teachers for not making any changes. There are two

fundamental reasons why performance assessment in the classroom is not conducted in the way it should be.

First, the administrators of either the Ministry of Education or local education offices do not appear to fully understand the nature and strength of performance assessment. Among the collected cases were found actual examples published by a metropolitan office of education, to show how performance assessment can be conducted in the field. As teachers who had been raised in traditional education are accustomed to the old ways, they needed a new model to imitate and follow when the government first implemented the initiative. Yet the models that the local administrators distributed to schools were in fact far from even adequacy. Inflexible, short-answer-seeking, stringent formats of assessment as well as the scoring guides that are criticized above such as counting and simple quality check, for instance, were in fact also found in some of the governmental models that teachers were supposed to make reference to.

Figure 6. Governmental Models of Performance Assessment and Scoring

Example Questions for Science

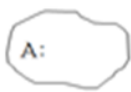
[Activity Sheet]


Performance Assessment Activity Sheet


Grade: 1 Class: No.: Name:


1. Write the number of my basket

2. Write the name of each rock in the basket

A:


B:


C:


D:


3. Classify the rocks into 2 groups based on cause of creation

4. Write the criteria for classification

Middle School Performance Assessment Example

Assessment element	Score	Assessment content
Level of Interest	15	Method was creative and interesting
	7	Method was formulaic, straight from textbook
	4	Neither from textbook nor creative
Logic	10	Both hierarchy and consistency are evident
	7	For hierarchy and consistency, only one is evident
	4	Neither hierarchy nor consistency are evident

The second fundamental reason why performance assessment is not in place in Korean schools is the general reluctance or resistance among the public believing that performance assessment is not ‘objective’ and thus unfair. In Korea, the practice of sorting out youths by their ‘assessment performances’, or simply, ‘test scores’, into various levels of colleges and government positions has become deeply entrenched in the culture since as far back as the medieval times (Cheng, 1998). Even the ordinary terms reflect people’s prejudice toward quantitative and performance assessment as one of the Korean terms for multiple choice tests is “Gaek-kwan-shik”, which means ‘the objective style,’ and a term for short or long essay tests is “Ju-kwan-shik”, which means ‘the subjective style.’ Concerned about probable parents’ complaints about their children’s scores, whether low or not high enough, teachers are not willing to take risks to adamantly go for the new so-called ‘subjective style’ of performance assessment. Furthermore, test administering organizations and school teachers who make mid-term and final examinations sometimes capitalize upon the advantage of multiple choice tests to statistically render a rather perfectly vertical line-up of students by inserting extremely tricky problems that has low content validity but a technical power to sift out low scorers. For those test-makers, such ‘stratification efficiency’ matters more than the educational validity that education is supposed to do what education should do. Unbelievably difficult problems often appear even on the college-entrance examination to serve a special purpose of effectively lining up students and yet the public is satisfied with such imbalance between heavy stratification efficiency and weak validity.

Even administrators are not free from the cultural belief, or bias, about the objective style (multiple choices) versus the subjective style (performance assessments). Well aware of this type of classroom practice obsessed with stratification efficiency, they might distribute less-qualitative (thus less-subjective) performance assessment guidelines in order to minimize public resistance.

Yet such a classroom practice obsessed with stratification efficiency needs re-examination, particularly for an economically advancing country like Korea. The

simplification of complex human capabilities based on stratification efficiency in contemporary Korean educational evaluation can bring about a serious challenge to the whole society. The skill mismatch problem occurs, for instance, where college graduates graduate from colleges without acquiring proper sets of skills for jobs and life because they have been studying for the tests that assess what is less valid for economic, social and civic life (Kochen, 2014). This should be then the transitional time Koreans come to pay more attention to educational validity than stratification efficiency. Moreover, it must be reminded that performance assessment put into practice the right way can ultimately lead to higher 'efficiency,' for more correct measurement of core activities and performances will render reduction in human resource waste and thus the saving of cost and time (Alaskari, 2014).

4. How Can We Foster Performance Assessment?

As shown before, there is an enormous gap in performance assessments between institution and classroom. Korean policy makers tend to have focused excessively on institutional changes such as the college entrance system and the national education curriculum. It needs to be questioned whether actual practice in the classroom became ignored amid such fierce dispute and conflict over institutional reforms.

This chapter emphasizes that the top-down approach the government has taken thus far, including changing the college admissions system and national education curriculum, is clearly limited in its capacity to overcome the deeply rooted Korean culture that sacrifices assessment validity for stratification efficiency.

First, the college entrance examination has always been the center of public attention. In the 1990's, the previous essay examinations held independently by each college was changed into the CSAT(College Scholastic Ability Test), a single test of multiple choice questions held by the government. This change, however, brought about important side effects of quantitative assessment becoming reinforced in Korea's classrooms. The stated goal of the difficult multiple choice test of CSAT was to strengthen the ability of comprehensive thinking, but in reality the limitations of multiple choices could not be overcome and students ended up focusing all of their energy on solving test problems (Kim et al., 2010). As a solution, the ministry of education emphasized that school grades assessed by teachers should be given more weight in the college entrance process.

However, in order to do this, very strict relative grading scale composed of nine different ratings was imposed for every class in every high school. This strong regulation on “nine relative grades system” was introduced because of the concerns of grades inflation that every school could higher the average grades so that their students could be in a better position in the university entrance procedures. However, the nine relative grades system has serious side effects in that it induces teachers to focus more on stratification efficiency rather than assessment validity, which makes performance assessment hard to take root in the classroom.

Such a college entrance process that emphasizes the difficult CSAT and nine relative grades system cannot but lead teachers to rely on objective methods like multiple choice questions, and in turn students only focus on developing skills to solve such problems in order to receive good grades.

In order to resolve such problems, the Admissions Officer System was implemented in 2007. As of 2013, 13.6% of newly admitted college students were admitted through this system. Additionally, there have been efforts to change the nine-rating relative evaluation system into one based on an absolute grading scale known as the “achievement assessment system”. Also, in order to lessen the burden on students, the number of CSAT test subjects was reduced and test content was connected to lectures provided by the national broadcast company EBS (Education Broadcast System).

Despite such efforts, although the Admissions Officers are incorporating records of creative activities and student interviews into the admissions process, the process does not sufficiently consider performance assessments of students. Also, the achievement assessment system is still in the preparation stage before actual implementation. Upgrading the CSAT has lacked consistency and momentum as the newly planned NEAT(National English Ability Test), designed to reduce student burden and more effectively assess students’ expressive English capacity, has been canceled (Sung et al., 2013). Nonetheless, although there is significant room for improvement, the direction of college admissions has shifted toward reducing CSAT burden and emphasizing teachers’ assessment and the role of Admissions Officers, and thus the overall direction is headed towards fostering performance assessment.

In addition to the university entrance examination, the national education curriculum is another focus of public attention when it comes to education reform. In fact, the teaching material required by the current national education curriculum forces teachers to teach an excessive amount of knowledge content, and this has been criticized as a hindrance to

implementing newly designed teaching methods and performance assessment. For this reason, the revised national curriculum of 2009 has given each school the freedom to reduce 20% of the content while increasing the time devoted to creative activities. In pursuit of changing the national curriculum, the Lee Myung Bak administration has emphasized “creativity and character education”, STEAM education, smart education, and the Park Geun Hye administration has focused on “talent and dream building education” and the Free-Semester Policy. In such ways different administrations have made efforts to make positive changes to the education curriculum and teaching pedagogies.

However, it has been noted that such government efforts have been insufficient for bringing actual changes in the classrooms. In particular, there is still an excess amount of material to be taught at the more advanced levels (Choi et al., 2011). Rather than assisting teachers to bring about actual changes in the classroom in terms of teaching and assessment methods, the government has been heavily criticized for merely making orders and penalizing teachers who fail to follow them. Also, consistency has been broken every time a new administration or new minister comes into office, and the government has failed to present the direction of change while inducing schools and teachers to make sufficient long-term preparations. Furthermore, policies on college admission, national curriculum, and teaching and assessment methods have not been presented under an integrated framework (Jeong et al., 2013).

Korea’s approach of focusing more on institutional reform rather than actual changes in the classroom needs to be reconsidered, and in the future the focus needs to be decisively shifted to actual changes in classrooms. Under this context, this paper proposes a gradual bottom-up approach that supports changes in classrooms as a national strategy for fostering performance assessment.

Above all, a bottom-up approach that supports the capacity development and active participation of teachers, as well as continuous expansion of changes in schools needs to be emphasized. Also, policies pertaining to college admission, national curriculum, and teaching and assessment methods need to be pursued in a gradual and consistent manner under an integrated national strategy, and not be altered every time a new administration or minister comes into office.

This paper suggests five major strategies for pursuing a gradual bottom-up approach for the change from multiple choices to performance assessment. Based on field visits and expert meetings with ten school teachers, these strategies were constructed by listening to the

voices from the actual scenes of education. The explanations of the five strategies include original quotes from the teachers. The detailed five major strategies are as follows.

4.1. Strategy 1: Establish a Framework for Putting Performance Assessment into Practice

Because the legal and institutional basis for performance assessment has already been set up for more than fifteen years, making sure that performance assessment is effectively put into practice in classrooms will bring about important positive changes without unnecessary controversies and conflicts that could occur for implementing a completely new institution.

Fifteen years have passed since the performance assessment was first implemented, but it has still yet to become properly rooted in Korean classrooms. This is fundamentally due to the government's top-down approach of focusing excessively on changing college admission procedures and the national curriculum, while ignoring bottom-up approaches like establishing a well-functioning framework that brings about actual changes in classrooms. The specific problems mentioned by teachers on the current performance assessment included those that could not be solved by teachers alone.

Teacher A stated that teachers were not given the proper information and training on how to adequately provide feedback to students on performance assessment, and therefore teachers are unable to pursue them with certainty.

“When I do performance assessment, there are times I feel that students do not show a strong desire to study. I believe that performance assessment is important, but it is also true that the lack of objectively observable results encompass a sense of ambiguity. Detailed records of teacher observations need to be made for performance assessment. Also, it is easy to record positive attributes for a student, such as creativity, logical thinking, and the capacity to cooperate with others, but it feels uneasy to write negative attributes about students, such as inability to work with others or lack of capacity or will to complete a task. Teachers try to avoid writing such things.”

Teacher B said that since there is no national support system for conducting performance assessment, rather than going through sufficient research or receiving the proper training, teachers defer to private publishing companies that produce workbooks, download question sheets from the Internet, and develop tasks that are relatively easier to evaluate.

“Although performance assessment is in the form of essay writing, many portions are comprised of problems that require memorization of answers. Teachers often refer to

publishing companies or the Internet for developing questions, and rarely conduct thorough research on their own. For essay questions it becomes difficult for teachers to answer to parents who demand to know why their child received the score that was given. In order to avoid such complaints from parents, questions are naturally designed to test memorized knowledge for the convenience of evaluation.”

Teacher C said that the government, with the help of teachers and experts, needs to develop a more detailed guideline for performance assessment, and suggested the following.

“Providing individual evaluations based on performance assessment for a class of thirty students can be a large burden on teachers. In the national curriculum there is no guideline provided on performance assessment. Regular teachers are limited in their capacity for developing questions designed for performance assessment which includes the core achievement elements, and thus this should be done by a group of experts. To lessen the burden on teachers and reduce the differences among teachers, a guideline must be provided.”

According to the opinions of teachers, it becomes evident that what is most urgent in order for the performance assessment to become rooted in Korea’s education is more information and training. It is the lack of a national support system that leads teachers to refer to publishing companies and the Internet. Therefore it is necessary for the government to work with teachers and experts to develop a sufficient guideline concerning performance assessment and to provide training programs for teachers.

Next, teachers state that the reason why performance assessment had unsatisfactory results is not the teachers’ unwillingness to change, but due to the government’s unilateral way of ordering teachers. Teachers say that before requiring additional work from teachers, reducing some of the current workload to lessen teachers’ burdens is needed to increase the chance of success. For performance assessment to function properly, it appears that the workload of teachers should be taken into consideration and any unnecessary paperwork and procedures should be reduced. Also, it was noted that a major reason why teachers feel burdened by performance assessment is the unreasonable complaints made by students and parents, and thus efforts to protect teachers should be made by schools and the education offices. Teacher D stated the following regarding teachers’ burdens from performance assessment.

“Teachers do not dislike change, but what they find difficult is that they are put in a tough situation and are constantly required to do more and more work. Pushing teachers in such a unilateral way increases the burden on them. If the majority of teachers feel this way,

there is bound to be negativity toward change. Rather than increasing the burden, change can be successful if it is approached in a way that lessens teachers' burdens. A system needs to be put in place to promote the validity and trust for performance assessment, and protect the teachers from the complaints of parents and students. It is very difficult for a teacher that has been hurt to recover."

Teacher E mentioned the following as some problems of performance assessment that require long-term solutions.

"For English, performance assessment requires students to create scripts to show their English speaking abilities, but there is a wide gap in skills among students. Because the level of English skills vary greatly, it is almost impossible to conduct individual performance assessment on each student. Even having each student speak once is not easy in a controlled environment if the willingness of the teacher is not very high. I believe that revising essays and providing constant feedback is an effective teaching method, but this too is very difficult. After going to Canada on a training program, I noticed that teaching and administrative tasks are completely separated. Korea also needs to create for teachers an environment in which we can focus our energies solely on teaching."

Taking a comprehensive look at the opinions of teachers, it appears that in order to establish a system to support the proper practice of performance assessment in classrooms, there needs to be a performance assessment guideline with enough information for teachers, and also enough training programs should be provided. Additionally, excessively complex paperwork and procedures pertaining to performance assessment need to be diminished, and a way to protect teachers from unreasonable complaints from students and parents needs to be implemented. A campaign should be run to actively promote to parents the importance and purpose of performance assessment. Also, long-term measures for reducing the gap between students in classrooms and eliminating the excessive administrative workload of teachers need to be pursued.

4.2. Strategy 2: Establish a System to Develop the Performance Assessment Capacity of Teachers

Changes in assessment method and changes in teaching method in the classroom are like two sides of a coin. If the teacher does not have the capacity to carry out such changes, the changes will never be successful. However, a long-term approach for developing teachers'

capacities for changing assessment and teaching methods has been lacking.

Teacher F said that a majority of teachers lack confidence in performance assessment because they have not received the necessary training.

“Unlike multiple choices, which can be done with swift, objective results, performance assessment requires teachers to have sufficient time. Project-based teaching has been first attempted by zealous teachers fifteen years ago, but it has failed to expand. The first teachers to attempt it have either given up or moved to rural area. The remaining majority of teachers do not have confidence in conducting performance assessments. The major reason is the lack of necessary training.”

In the same context, teacher G said that a one-time training program was all that was given regarding project-based teaching, and this has led to teachers conducting only what was deemed easy and convenient. However, she also said that a change is occurring where teachers are more actively researching and sharing information on ways to change teaching methods.

“Project-based teaching requires the guidance of an expert. Even when teachers go through training programs they usually take the easy route by simply paying for available material on the Internet. The response from parents and students are positive so teachers see no need to fundamentally change their teaching methods. Many teachers utilize the online community known as “Indischool”. The Internet community provides many useful class materials that have been voluntarily uploaded by teachers. I believe utilizing such communities is very helpful.”

Teacher H said a one-time training program is not enough for a change in teaching and assessment methods, and that an official training program that provides constant feedback from experts is necessary, particularly for teachers with five to ten years of experience.

“What is for sure is that in order to instill the will to change in teachers, constant feedback is necessary. Rather than leaving it up to schools, an expert group needs to consistently provide detailed feedback and present a vision in order for fundamental changes to take place. There are not that many opportunities for teachers to interact with one another and share ideas. The only way to change perception and provide a positive shock is through official training programs. For such e-learning programs, those that are operated remotely make interaction difficult. A forum for debate may be positive, but current training programs are approached merely as nothing more than obligations.

For teachers in their fifties, implementing new teaching methods is not easy. It is

more important to give teachers with around five to ten years of experience a new opportunity for change. These teachers are in the periods of their teaching careers that can bring about decisive changes. Current training programs are mostly composed of lectures. If the direction of teaching becomes more focused on project-based teaching, targeting a group of teachers with an adequate level of experience may be more effective. It will be meaningful to first implement the new method on teachers of a certain experience level.”

Putting together the opinions of teachers, it becomes evident that despite the rising interest from teachers in changing teaching and assessment methods, there are insufficient opportunities for receiving the necessary training. Therefore it is necessary that opportunities for training first focus on teachers with five to ten years of experience. The training programs themselves should be drastically changed so that teachers can actively share opinions on potential problems that may arise during the transition to new teaching methods, and constant feedback from experts becomes available.

Additionally, the government should pay close attention to the recent rise in teachers’ interest in changing teaching and assessment methods, and develop an environment for such efforts to become more active while creating measures for providing support. Also, a long-term approach would be to increase the time devoted to developing new teaching and assessment methods within the curriculum of education universities and colleges of education.

4.3. Strategy 3: Support the Expansion of Performance Assessment among Schools

Changes in teachers are the most important, but changes in schools are just as important when it comes to changes in teaching and assessment methods. Without the support from the school, there is a great chance the individual teacher’s pursuit for changing the teaching and assessment method will be unsuccessful. Therefore, as a bottom-up approach for changes in teaching and assessment methods, it is important for certain schools to develop leading models for change that can be followed and expanded by other schools.

There are already schools that provide increased autonomy and financial assistance to allow changes in teaching and assessment methods to occur. As part of the policy for high school diversification, the Lee Myung Bak administration created Local Dormitory Schools, Autonomous Private High Schools, and Autonomous Public High Schools, and also there are elementary and middle schools that receive financial assistance after being designated as Creative Management Schools and Innovation Schools. In order to induce fundamental

changes, the Park Geun Hye administration is pursuing the Free-Semester Policy for middle schools. Therefore in order to bring about fundamental changes in teaching and assessment in schools, it is effective to make concentrated efforts to support these schools, and encourage the changes to be expanded to other schools.

Teacher I criticized model schools and research schools for excessively using research grants on short-term facility expansion or increasing school publicity, and said that schools should be supported to bring in external experts. He added that rather than leaving the decision-making to a single individual like the vice principal, the opportunity to participate should be expanded to teachers as well.

“When the ministry of education designates certain schools as research schools to help performance assessment take root, the person that leads the research is usually the vice principal or the head of the research department of the school. The truth is, in many cases such individuals lack the capacity to lead researches. Therefore measures like bringing in an external expert to lead the research or allowing more active participation from teachers is necessary. After a two to three year process of researching and supplementing deficiencies, the next step should be visiting other schools or taking part in training programs. Currently the task is conducted too hastily where research grants are meaninglessly wasted on simple facility expansion and school promotion activities. Unlike the previous research schools, there need to be model schools that conduct genuine research on specific measures, and the positive results that they produce should be expanded to other schools.”

Teacher J said that changing teaching and assessment methods in schools require much time, so rather than a short-term result-based approach, a long-term strategy that acknowledges differences among schools is needed.

“In order to implement positive changes in schools, the number of research schools need to be expanded and individual focus needs to be given to each subject. Desiring short-term results makes no sense, and generalizing the models of a few schools is not feasible as well. Each school is under different circumstances and is composed of different teachers. Under the current system, where public school teachers have to move to a different school every five years, the expansion of research outcomes to other schools cannot but be slow. Nonetheless, despite the slow process, focusing on research schools is the direction that will lead to the expansion of positive changes.

Teacher K approached the issue from a slightly different perspective, and proposed the following on enhancing communication among teachers and developing an education

community that includes participation from parents to expand performance assessment.

“The expansion speed of performance assessment can be increased by operating communication channels such as course and training tours. Rather than focusing on possible incentives for changing teaching methods, efforts should be made to develop an education community that can ask questions and cooperate toward a common goal. After the government first sets the direction, it would be preferable for the teacher to reinforce communication. It would be even better if parents are involved in the process and an atmosphere of agreement is formed. Of course, such efforts will be effective when there is consistency in government policies.”

Considering the opinions given by teachers, it becomes evident that for measures like the Free-Semester Policy, Autonomous Schools, and the establishment and expansion of performance assessment to be successful, the schools designated to lead the change should receive help from external experts and expand the level of participation to regular teachers. At the same time, through methods like course and training tours, channels for communication among teachers should be strengthened to assist in the establishment of an education community that can work together toward a feasible solution.

4.4 Strategy 4: Introduce Performance Assessment in the National Assessment of Educational Achievement (NAEA)

While the three strategies mentioned above are strategies for bottom-up changes for performance assessment, the following two strategies are focused on a more gradual perspective of top-down approach. Until now, efforts for bringing about changes in the classroom were focused on changing the high-stake examinations that are directly related to students' college admission such as CSAT. However, attention should also be given to changes in low-stake examinations such as the NAEA, which do not directly affect college admission, but can induce changes in teachers and principals.

The frequent changes brought on the college admissions process in the past have been criticized as merely having increased the burden on students and contributed to the expansion of private tutoring, rather than leading to actual changes in teachers' teaching and assessment methods. Therefore, rather than drastic changes in the college admissions process, more focus should be put on improving the NAEA, which will be less of a burden on students while inducing changes in teaching and assessment methods of teachers. The introduction of

the NAEA by the Lee Myung Bak administration has been effective in helping improve the academic performance of students whose achievement levels are below the basic minimum. However, because the current NAEA is conducted in multiple choices, the next policy agenda is to shift it into one that includes performance assessment.

In particular, for schools that have implemented the Free-Semester Policy before others, the quantitative NAEA can be a burden for teachers and students after spending a full semester without any sort of quantitative assessment.

Teacher L has said the following regarding the difficulties of a school designated as a model school for the Free-Semester Policy.

“Bupyeong Middle School has been designated as one of the model schools for the Free-Semester Policy. We have decided to make the second semester of first-year students a Free-Semester, and in the afternoons of the semester we operate project-based classes or allow students to take part in various activities of their interest. One thing that worries me is the fact that such efforts from our school will not be well-reflected in the NAEA. As is well known, the results of the NAEA are used for comparison among schools, rather than students, so it is quite burdensome for principals, vice-principals, and teachers. Performance assessment should be reinforced in the NAEA so that schools implementing new systems will be acknowledged for their efforts rather than be at a disadvantage. In this way, changes of the NAEA to include performance assessment will contribute to the Free-Semester Policy taking root quickly in schools.”

In the U.S., after the NCLB (No Child Left Behind) policy was implemented, there was strong criticism on academic achievement evaluations conducted on all schools nationwide, and this resulted in the “Race to the Top Assessment Program”, which encourages statewide efforts to improve academic achievement evaluation in the NCLB by putting more emphasis on performance assessment. In addition, the OECD is planning to assess students’ cooperative capacity on the 2015 PISA. Also, researchers abroad are actively conducting researches through the TAC21S to develop a method to assess the 21st century skills in an online platform (Ryu, 2012).

As Korea also needs to consistently supplement the NAEA to make performance assessment possible, Korean researchers should take part in efforts for innovating assessment methods and attempt to include them into Korea’s nationwide evaluation process. At the same time, efforts should be made to make use of the already developed NEAT, and include it in the English section of the NAEA. To change the current English assessment toward

performance assessment, the evaluation criteria should include speaking and writing in addition to the pre-existing reading and listening.

4.5. Strategy 5: Pursue Gradual Changes with Consistency in the College Admissions System

If pursued in a consistent and long-term manner, changes in the college admissions system could be very powerful in inducing changes in the classroom. Therefore, in addition to the changes in low-stake examinations like the NAEA, this paper proposes that gradual changes in the college admissions system be pursued consistently with a more long-term perspective.

Due to its nature, reforming the college admissions system requires a very timely process. Nonetheless, policies initiated during one administration will most likely be reversed during the next, if an agreement strong enough to overcome administration changes that occur every five years is not formed. Furthermore, colleges, the most important party of the college admissions system, will focus more on simply selecting students with the highest academic grades, rather than considering the impact college admissions system will have on school education or national investment in human capital (Lee, 2013). Therefore, in order to pursue long-term gradual changes in the college admissions procedure with consistency, representatives from colleges, including the KCUE (Korean Council for University Education) and college presidents, need to come together with teachers and experts to establish a national consultative group designed to develop a long-term vision for changes in the college admissions system. The political circle and education organizations need to provide strong support and help develop a long-term measure for change. The only way to break the vicious cycle of policy inconsistency in the college admissions system is to develop a reform measure that sufficiently includes the opinions of both colleges and teachers as well as major interest groups in politics and education.

Teacher M said that in order for gradual and long-term changes to take place in the college admissions system, a national consultative group needs to form a measure that leads to self-constraint on the part of colleges.

“Rather than getting rid of the CSAT, I believe it is more preferable to gradually diminish it. The weight given to school grades and teacher evaluation in college admissions should be gradually increased while colleges increase Admissions Officers in both quantity and quality to create an environment in which the school records of applicants can be

evaluated meticulously. In addition, rather than merely preparing students for multiple choices of CSAT, schools should try to better meet the purpose of education by implementing performance assessment. I believe this will be beneficial for both schools and colleges. I think the government's efforts to change the college admissions procedures has failed each time. I heard that in the U.S. colleges take a leading role in changing the college admissions system in a direction that helps normalize school education. Rather than hastily changing the college admissions system, Korea should create a consultative group to develop a measure that allows colleges to constrain themselves. Teachers, scholars, college presidents, Admissions Officers should all come together to create a measure that can replace the current CSAT system."

In order for measures like the Admissions Officer System to become rooted in Korea's education, opinions from the actual scene of education need to be collected and consistent efforts to maintain the original purpose of institutional changes are necessary. The approach of simply replacing any policy that appears to show problems must be avoided. What is currently needed is efforts on helping policy measures sufficiently take root in Korea's education.

Under this context, teacher N suggested in more detail that Admissions Officers should first focus on assessing students reading activities.

"For changes in teaching methods and diverse forms of performance assessment to take place, including the Admissions Officer System, things have to be reflected in the college admissions system. Unless the precondition of inclusion into the college admissions system is made, all efforts can be unsuccessful. A system where the reading or debating activities in schools is acknowledged and evaluated by Admissions Officers must be established. It is for certain the current Admissions Officer System carries the limitation of monotonously assessing students based on grades and other objectively assessable characteristics. If you look at the current academic records of students, reading activities take the largest portion, and so I believe that with additional effort diverse ways for evaluation can be developed."

Teacher O suggested that subjects that are not included in the CSAT should be evaluated based 100 percent on performance assessment.

"Overcoming elitist education is another purpose of performance assessment. The enhancement of character and passion should also be reflected. Ultimately, in order to normalize Korea's education, grades for Physical Education need to be given based 100 percent on performance assessment. The problem is that on one side we have people stressing higher test scores in multiple choices while on the other we have voices for performance

assessment. It may seem like a drastic measure, but we should consider all non-CSAT subjects to be evaluated solely through performance assessment.”

In regards to long-term changes to the college admissions system, this paper has listened mostly to the voices of teachers, but the ideas and opinions of colleges, including the KUCE and college presidents, as well the perspectives of students and parents need to be sufficiently taken into account to develop feasible, gradual, long-term policy measures through a national consultative group. Furthermore, when putting into practice such policy measures, it should be pursued gradually while the opinions and feedbacks coming from actual classrooms are sufficiently accepted.

5. Conclusion

The current education scene seems to be at an inflection point where the speed of change will gradually accelerate. Until now, there was much controversy and conflict over government efforts for institutional reform, and the lack of integration and consistency was heavily criticized. Nonetheless, the changes in the college admissions system and the national curriculum somehow have made enough room for teachers to shift assessment methods from multiple choices to performance assessment. The problems of rote learning and multiple choices have reached a crisis level, and the environment for turning a crisis into an opportunity has been set.

Most of all, Korean students are not happy in classroom. The question of happiness in the 2012 PISA revealed that Korean students had the lowest level of happiness, and Korean teachers had drastically low levels of self-efficacy (Lee, Jeong, and Hong, 2014). It is highly likely that students are unable to feel the joy of learning at school and feel excessively pressured to outcompete other classmates in multiple choice tests. Such low levels of teachers' self-efficacy are likely due to their inability to fulfill the role of learning partners to motivate students to gain interest in the material that is taught and study on their own. Rather, teachers remain as mere deliverers of book knowledge.

The problems of Korea's education, where students feel unhappy and teachers lack self-efficacy, present strong evidence for the necessity of shifting to performance assessment. At the same time, Korea's education carries the potential ability to bring about the necessary changes. Most of all, hope can be found in face-to-face interviews with teachers that have revealed teachers' desire for change. Additionally, due to the advanced level of ICT

(information communication technology), Korean students and classrooms have access to an immense amount of digital knowledge and information. Such technology can be a powerful partner for assisting the changes in teaching and assessment methods.

Therefore, by overcoming the problems of the top-down approach taken until now, while meeting students' and teachers' demands for change and consistently pursuing a gradual bottom-up approach for changes in teaching and assessment, Korea's education can be successful in shifting from multiple choices to performance assessment.

Lastly, the reason why we are emphasizing a gradual bottom-up approach is certainly not to slow down or reverse education reform, but because it can work to speed up changes in the actual scene of education. It is evident from Korea's past experiences that changes in the college admissions system or the national curriculum, when pursued without agreement within the education field, will lead to changes in reform policies every time a new administration takes office and ultimately slow down the process of education reform. Therefore, when dealing with such institutional changes, pursuing a gradual, yet coherent and consistent approach is a way to achieve faster changes in actual classrooms. On the other hand, policies to support actual changes in classrooms, including reinforcement of performance assessment or project-based teaching for future teachers in the curriculum of education universities and colleges of education, are policies that the government needs to more actively pursue.

References

- Alaskari, O. (2014). Development of assessment methodology for improving performance in SME's. *International Journal of Productivity & Performance Management*, 63(4), 477-498.
- American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME). 2014. *Standards for Educational and Psychological Testing*. Washington, DC: AERA.
- Cheng, K. (1998). Can education values be borrowed? Looking into cultural differences. *Peabody Journal of Education*, 73(2), 11-30.
- Chi, E.-L. (2000). Comparing scoring methods for essay type performance assessment. *The Journal of Educational Research*, 16, 235-246.
- Choi, Sunghyun, Youngsun Kwak, Eunhee Noh (2011). *Research on Teaching and Learning and Teacher Education to improve Learners' Key Competencies: Centering on Mother Tongue, Mathematics and Science*. Korea Institute for Curriculum and Evaluation RRI 2011-1. (In Korean)
- Chung, Mi-kyung et al. (2013). *Research on Improvement Measures of Primary-secondary Education System: Focused on Education Contents, Methods, and Evaluation Systems*. Korean Educational Development Institute. (In Korean)
- Conley, D., Darling-Hammond, L. (2013). *Creating Systems of Assessment for Deeper Learning*. Stanford, CA: Stanford Center for Opportunity Policy in Education.
- DeCarlo, Lawrence, Youngkoun Kim, Matthew Johnson. 2011. A hierarchical rater model for constructed responses, with a signal detection rater model. *Journal of Educational Measurement*. 48 (3): 333-356.
- Dwyer, Carol Anne. 2008. *The future of assessment*. New York, NY: Taylor & Francis Group, LLC.
- Darling-Hammond, Linda (2012). *Policy Framework for New Assessments*, in Patrick Griffin, Barry McGaw, Esther Care ed. *Assessment and Teaching of 21st Century*, Springer.
- Elias, Maurice, (2009). Social-emotional and character development and academics as a dual focus of education policy. *Educational Policy* 23(6): 831-846.
- Fullan, Michael and Maeia Langworthy (2014). *A Rich Seam: How New Pedagogies Find Deep Learning*, Pearson.
- Gripps, C. V. (1994). *Beyond Testing: towards a theory of educational assessment*. London: Falmer Press.

- Kim, Hui-Kyung et al. (2013). *Measures to Improve Assessment Methods for the Realization of Creative Character Education*. Korea Institute for Curriculum and Evaluation. (In Korean)
- Kim, Jae-woong, Kyung-chul Huh, Eun-soon Oh, Hui-jeong Yun (2010). *Reflection and Prospect of Teaching and Learning Reform in Korea*, Korea Institute for Curriculum and Evaluation RRI 2010-3. (In Korean)
- Kochen, M. (2014, March 17). Seven reasons why Korea has the worst productivity in the OECD. *Business Korea*. Retrieved from <http://www.businesskorea.co.kr/article/3698/insider-perspective-seven-reasons-why-korea-has-worst-productivity-oecd>
- Lavy, S., Garcia, John., Dixit, Manish. (2014). KPIs for facility's performance assessment, Part II: identification of variables and deriving expressions for core indicators. *Facilities*, 32(5/6), 275-294.
- Lee, Ju-Ho, Hyeok Jeong and Song-Chang Hong (2014) Is Korea number one in human capital accumulation? Education bubble formation and its labor market evidence. *KDI School Working Paper 2014-03*.
- Lee, Ju-Ho, Making education reform happen: removal of education bubble through education diversification. *KDI School Working Paper 2014-05*.
- Lee, Sam-Ho (2013). The demand for shadow education: a theoretical approach. *Korea Association of Applied Economics*. (In Korean)
- Lissitz, Robert, Xiadong Hou, Sharon Slater. 2012. The contribution of constructed response items to large scale assessment: Measuring and understanding their impact. *Journal of Applied Testing Technology*. 13 (3): 1-50.
- Livingston, Samuel A. 2009. Constructed-response test questions: Why we use them; how we score them. *ETS R&D Connections*. 11: 1-7.
- Messick, S. (1996). Validity and washback in language testing. *Language Testing*, 13(3), 241-256.
- Miller, George. 1990. The assessment of clinical skills / competence/ performance. *Academic Medicine*. 65 (9): S63-S67.
- OECD (2003). *The Definition and Selection of Key Competencies*
- OECD (2007). *Understanding the Brain: the Birth of a Learning Science*
- Ryoo, Sung-Chang (2012). *Research on the possibility of capacity-centered reform for college admissions system*. Korea Educational Development Institute. (In Korean)
- Seong, Tae-Jae et al. (2013). *2020 The Direction and Challenge of Korea's Elementary and Middle School Education: Curriculum, Teaching, and Assessment*. Hakjisa. (In

Korean)

Trilling Bernie and Charles Fadel (2009). *21st Century Skills: Learning for Life in Our Times*. Jossey-Bass A Wiley Imprint. Chapters 2-5.

Yan, Z., Liang, F. (2006). *The effects of using performance assessment tasks on Singapore students' learning of mathematics*. Paper presented at the American Educational Research Association.