*영어어문교육, 15*권 4호 2009년 겨울

A Review of Media Argumentation: Roles of Background Knowledge in Critical Reading

Jong-Hee Lee
(Kangwon National University)

Lee, Jong-Hee. (2009). A review of media argumentation: Roles of background knowledge in critical reading. *English Language & Literature Teaching, 15*(4), 157-175.

This paper offers a critical review of a newspaper argument regarding the problems of high school education assessment for university entrance examination system in the United Kingdom. The media account raises three sets of questioning to hold that the nation's long-standing A-levels have failed and is no longer viable as a high-stakes test. However, it is found that the writer's argumentations involving misleading conceptions can be deconstructed because of invalid reasoning and unreliable evidence. So, it is proposed that a reasonable solution to replace the discredited A-level exams should be to adopt an eclectic approach for assessing candidates' multiple capabilities; performance, potentiality and critical thinking skills. These criteria for component-oriented assessments are designed to measure their high school academic achievements and intellectual capacity for tertiary education; in the process of such measurement, critical-logical reasoning abilities for sound judgment and problem-solving tasks should be incorporated with the basic precondition that each university possesses its own discretion for the determination of adequate proportions to reflect each of the assessment outcomes. It is, therefore, expected that this critical review will inspire the readers to understand aspects of assessment as an educational field and to confirm how seriously they may be misguided by a distorted media argumentation without substantive background knowledge.

[assessment/measurement/validity/reliability/A-level exams/high-stakes test]

I. INTRODUCTION

The main purposes of this paper are three-fold: (a) to subject a media discourse entitled "Filling A-level Vacuum" (see Appendix A) to critical examination; (b) to suggest a set of new approaches for the improvement of discredited A-level exams; and, as a natural output

of (a) and (b) above, (c) to inspire the readers to recognize the extent to which they may be manipulated to have certain lines of thought by a distorted media argument. The final objective specified here is associated directly with the essential roles of background knowledge in facilitating the processes of reading comprehension in Korea's EFL educational contexts (Lee, 2003; Im & Lee, 2004).

The critical review is based on the assumption that the writer's arguments focused on abolishing the U.K. system of 'A-level exams' (see Appendix B) can be challenged and deconstructed. In this criticism I will address a range of technical and conceptual issues related to his controversial viewpoints, presenting what lines of media arguments are fallacious and how such misconceptions could have been duly repaired in the frame of educational assessment. Thus, through an in-depth review of the said media argument, I intend to carry out a conceptual survey of 'assessment' as an educational field in order to provide useful guidelines for the substitution of current A-level examination system.

It is widely recognized that mass-media, including newspapers, desire to guide the public opinion to work out their individual goals. As a vital pressure group in a nation, they often attempt to even manipulate it with the assistance of their experts and supporters. Quite understandably, in the case of newspaper, it is not surprising that some of the articles are not well grounded on valid foundations and logical reasoning processes, and so readers may be persuaded by certain allegations. This tendency seems to me, in a crude sense, to justify my attempt for the critical examination of the media account stated above.

To gain a clear picture of the arguments set forth in the said newspaper article, an analytical approach to their disputable elements would be required. In doing this, I will begin with an overview of assessment and measurement concepts in an effort to figure out the key components of the writer's opinion, addressing a few sets of technical terms such as formative and summative tests, correlation and regression, validity and reliability. Then, by exploring a number of issues raised in the said article, I would like to clarify the requirements for decent assessment systems in terms of college entrance examination.

Much attention will be given to the systematic analysis of technical issues embedded in the media account which contains misguiding perspectives on allegedly discredited A-level exams as the needs for reforming the current assessment frameworks. By drawing on all viewpoints considered, I will also try to help the reader confirm the extent to which a media discourse may be professionally distorted so as to disguise the valid and reliable picture of a relevant educational issue in terms of the roles and responsibilities of media.

II. CONCEPTS OF ASSESSMENT AND MEASUREMENT

1. Assessment

To define the term *assessment* is considered necessary before discussing the purpose and practical aspects of assessment, which may well lead to a critical inquiry into the media account. As an inclusive term, assessment has a plurality of definitions. In this respect it is useful to illustrate some of them here. Linn & Miller (2005) define assessment as "the full range of procedures used to gain information about student learning (observations, ratings of performances or projects, paper-and-pencil tests) and the formation of value judgments about learning progress" (p. 26). Also, stating that this term is a broader process than a test, they refer to a test as 'a particular type of assessment that typically consists of a set of questions administered during a fixed period of time under reasonably comparable conditions for all students'. From a similar point of view, Reynolds, Livingston & Wilson (2006), distinguishing assessment from a test, introduce the definitions of the two terms: (a) assessment is any systematic procedure for collecting information that can be used to make inferences about the characteristics of people or objects; and (b) a test is a procedure in which a sample of an individual's behavior is obtained, evaluated, and scored using standardized procedures (p. 3).

In a slightly different vein from the above, focusing on the roles of assessment to support learning, Gardner (2006) presents the conception that assessment is "the process of seeking and interpreting evidence for use by learners and their teachers, to identify where the learners are in their learning, where they need to go and how best to get there" (p. 2). And, with due weight on the complete process of assessment, McMillan, (2001, p. 4) provides its concise definition with "the process of collecting, synthesizing, and interpreting information to aid in decision making". With all the concepts above in mind, I intend to take this viewpoint by recognizing the essential needs for simplicity and comprehensiveness in defining a technical term. What is understood here in effect is that *A-levels* are, as the writer puts it, a *test* having standardized procedures within the frameworks of assessment.

2. Measurement

With the help of the overall definitions quoted above, it is understandable that assessment incorporates a test which is viewed as "an instrument or systematic procedure for *measuring* a sample of behavior by posing a set of question in a uniform manner" (Linn & Miller, 2005, p. 26). The term *measurement* at this point, often referred to as "a set of rules for assigning numbers to represent objects, traits, attributes, or behaviors" (Reynolds,

Livingston & Wilson, 2006, p. 3), is concerned with the quantification of assessment procedures (Hopkins, 1998). Such numerical criteria that represent the degree of test-taker's performance are a specific device to attain the adequate precision required in a measurement, thus associated closely with the purpose served by the assessment (Hopkins, 1998; Kubiszyn & Borich, 2003; Linn & Miller, 2005). Taking due note of this, it may be fallacious to argue that only the 'number' – i.e., any mention of *wider participation*; when *only a minority* of students ever took them, as specified in the writer's media arguments – of individuals to participate in A-levels provides a significant variable for evaluating the validity and reliability of the exams. This crucial issue will be further explored in the following section.

III. CRITICAL REVIEW OF KEY ARGUMENTS

To facilitate a critical investigation here, it may be helpful to look at a summary of key elements described in the writer's media discourse. In the introductory part (see Appendix A: paragraphs 1~2), the writer makes clear that elitism, punishing failure by ranking for excluding candidates, loses educational justification. Further he stresses that this meritocracy has a political nature, offering benefits to specific universities in the U.K. Based on this, in the main body (paragraphs 3~7) he draws attention to a small minority of students who successfully pass GCSEs (see Appendix C) and A-levels, presenting three sets of arguments leading to a few important technical and conceptual issues in terms of educational assessment: (1) correlation, validity and reliability; (2) accountability and certification; and (3) diagnosis and reform of high-stakes testing system. To illuminate these points, the full wording of his key propositions is underlined in Appendix A. In conclusion (paragraph 8), he argues for terminating A-level system, and adds that university tests are not decent replacement.

It is required, then, to examine his major lines of reasoning, from the technical and conceptual perspectives, to provide a critical inquiry into each component with respect to its cogency in the following:

1. Technical Issues

[The Writer's First Arguments: 5th Paragraph of Appendix A]

The writer's critical views on A-level exams here are concerned largely with the generic purposes and quality of the assessment system with its key criteria – correlation, validity and reliability – as the standardized achievement testing that has long regulated entry to

higher education and high-status professions. The following subsections will address each of these technical concepts in detail to identify the faulty elements of his argumentation.

1) Components of Assessment

In general, assessment is conceived as having four major and sequential components; (a) goal or purpose, (b) method of measurement or description, (c) evaluation, and (d) use or applications (Hopkins, 1998; Kubiszyn and Borich, 2003; McMillan, 2001; Pellegrino, Chudowsky and Glaser, 2001). To work out the prime intentions of this critical review – attempting to challenge and deconstruct the writer's arguments, it would be appropriate to explore the purposes of assessment and methods of measurement.

(1) Purposes and Methods

With respect to the purposes of assessment, Pellegrino, Chudowsky & Glaser (2001, p. 37-40) classify them into three main categories – (a) assessment to assist learning, (b) assessment of individual achievement, and (c) assessment to evaluate programs. They further mention that the first type of assessment here, also referred to as *formative assessment*, provides necessary information about students' academic progress in the course of learning. So this indicates any of evaluative activities aimed to provide feedback to students during instruction (Reynolds, Livingston & Wilson, 2006). Also, the second type of assessment helps determine the level of performance that a student has attained at the end of instruction, which is often called *summative assessment*. Such a type means the formal confirmation of the value or quality of a student's learning outcome (ibid.). The last goal of assessment noted above is to assist policy-makers in establishing judgments about the effects and practicality of educational systems.

As a result, I can easily confirm two interrelated points from these generic purposes of assessment: (a) any single measurement may be used for multiple objectives; and (b) evaluative activities are often employed to make high-stakes decisions not only about individual students but also about relevant institutions (Pellegrino, Chudowsky & Glaser, 2001). With these goals in mind, I can recognize that assessment normally begins with identifying its particular objective, and more importantly, the intended purpose helps select one or more appropriate methods of measurement (McMillan, 2001). It may be useful to draw a distinction between classroom assessment and large-scale standardized testing administered at a state and/or national level. However, a contemporary society tends to place greater value on large-scale than on classroom assessment (Pellegrino, Chudowsky & Glaser, 2001). In the light of this, I find it helpful to look at the following statement:

> Standardized measures provide the basis for most of the guidance and

> administrative test roles, whereas locally constructed measures are used principally for instructional functions (Hopkins, 1998, p. 10).

A-level exams administered in the U.K. are well known to be high-stakes achievement testing designed to use *standardized measures*. Hence, it is apparent that A-levels have been conducted to serve for the nation's wide-ranged administrative functions: Hopkins (Hopkins, 1998., p. 12-14) introduces their four principal categories; (a) to provide a mechanism of quality control; (b) to facilitate better classification and placement decisions; (c) to increase the quality of selection decisions; and (d) to be a useful means of accreditation, mastery, or certification.

(2) Evaluation of Arguments in Purposes and Methods

At this point, I need to get back to the writer's first set of arguments given in the media discourse and to paraphrase them for elucidating his points. Then what is required to inquire into here would be:

> (a) whether permitting wider participation than ever in A-levels may have resulted in "dumbing down" the candidates' academic standards for entry to higher education; (b) whether it was impossible to know A-level exams were sufficiently difficult when only a minority of students ever took them; and (c) whether A-levels have been an inadequate test from the very beginning if the exams turn out to be so at present.

Basically I can see that the questions (a) and (b) share almost the same issues as to whether the students' level of competency is dependent on the controlled scale of participants in A-levels. Question (c) asks about the possibility that the inadequacy of A-level test system, if any now, may be traced back to its past. The first issue here, given the administrative functions of achievement testing, is related to the rationale for *norm-referenced tests* (NRT) and *criterion-referenced tests* (CRT), which are widely accepted as two main types of collecting objective data in educational assessment. In general the nature of a test should reflect its primary purpose (Hopkins, 1998). So when a specific purpose is determined, it would be important to use the best method of measurement in order to achieve it. As one of the basic ways to interpret student performance, NRT describes a student's place or rank by comparing the student's performance to a norm or average of performance made by other students in a known group; but CRT, as the other type, describes a student's specific level of performance by comparing the student's performance to a standard or criterion of mastery called a criterion (Kubiszyn & Borich, 2003; Linn & Miller, 2005; McMillan, 2001; Reynolds, Livingston & Wilson, 2006).

From these purpose-oriented tests and their differentiated nature, I can infer that the writer's criticism on the inadequacy of A-levels loses its foundations simply because A-levels based on *Curriculum 2000* are currently a *soft criterion referencing*, neither fully NRT nor fully CRT (House of Commons, 2003). The current system stepped out of its past practices which had begun with NRT and then moved on to CRT. It is also obvious that A-levels, which follow the GCSEs, have taken the form of summative tests designed to certify student mastery of the intended learning outcomes (Hopkins, 1998; Linn & Miller, 2005). This enables us to refute the writer's argument that the small number of participants in A-levels has been a critical factor to impair the validity of the exams. To put it correctly, the writer should have held that the inadequacy of A-levels originates mainly from their inappropriate measurement methods, and verified it with the systematic reasoning from evidence. As a result I am secure in positing that the writer's first arguments may have narrowly definable cogency only during the twenty years 1963~83 when *norm-referenced A-levels* were being administered, which is also enough to deconstruct his additional claim that "if A-levels are an inadequate test now, this might well have been the case all along".

In the light of these observations, the next section will take a look at an essential process of reasoning from evidence in order to address aspects of confidence in assessment which are related to correlation, validity and reliability issues.

2) Correlation, Validity and Reliability

In the first set of arguments noted in the Appendix-A, by stating "How do we know exams were sufficiently difficult when only a minority of students ever took them?" the writer attempts to indicate an insignificant *correlation* between A-level test results and candidates' academic capabilities required to undertake their college degree courses, thus casting a strong doubt mainly on the *validity* and *reliability* of A-levels. In other words, the statement implies that the various types of validity – content, criterion, construct, consequential, and predictive validity – can be contested on the ground that the performance data of the A-level exams only a minority of students have taken are not validated any longer. Then I may expect that an inquiry into some major technical terms noted here will enable us to grasp the underlying rationale for the critical evaluation of his opinion.

(1) Concepts of Terminology

① Correlation and Regression

The major technical terms, as specified above, are correlation, validity and reliability. In educational assessment, *correlation* refers to a statistical procedure that shows how closely

students' test scores on one type of measure correspond in rank with those on another (Chase, 1999). Precisely the degree of correlation is represented by a *correlation coefficient* which is a quantitative measure of the relationship between two variables (Reynolds, Livingston & Wilson, 2006). In this connection, another concept, the *coefficient of determination*, is used to make comparative decisions concerning the relative degree of correlation coefficients, which is indicated by squaring the coefficient and multiply the result times 100 (Kubiszyn & Borich, 2003; Reynolds, Livingston & Wilson, 2006).

Thus the coefficient of determination is a more effective statistical device than correlation coefficient by converting the latter into the *percentage of variability* in one variable that is predictable from the other variable (Kubiszyn & Borich, 2003, p. 295). From this consideration it may be confirmed that correlation coefficient and, more importantly, the coefficient of determination provide the quantitative strength of the relationship between the two sets of scores and the amount of variance shared by the two variables.

In real situations, however, I may get interested in applying students' grades on achievement tests to predict the performance that they will demonstrate on another stage of assessment. So *linear regression* is used to make such useful predictions by drawing a straight line on a diagram indicating the distributed scores of two variables 'to represent the overall tendency of the relationship between the scores, and ignores the scatter of the points which makes the relationship less than perfect' (Black, 1998, p. 166). Clearly it is possible to draw such regression lines among students' GCSE scores, A-level grades and GPAs (Grade Point Averages) obtained in college degree courses. At this point I will not go into details about a *regression equation* and its computation to focus on the main goal of this critical review. Hence, the writer's critical stance may be undermined by his superficial remarks which do not specify a low, inadequate or negative correlation among students' GCSE scores, A-level grades and college GPAs. With this in mind, let us move on to the concepts of validity and reliability in educational assessment.

② Validity and Reliability

It is widely recognized that validity and reliability are central and fundamental psychometric components in terms of securing confidence through reasoning from adequate evidence in test results (Black, 1998; Reynolds, Livingston & Wilson, 2006). We may well begin with a plain definition of validity. Kubiszyn & Borich (2003, p. 299) mention that "a test possesses validity evidence if we can demonstrate that it measures what it is supposed to measure". Hopkins (1998, p. 72) also states that "the validity of a measure is how well it fulfills the function for which it is being used". In some technical sense, McMillan (2001, p. 17) defines validity as "an overall evaluation that supports the intended interpretation, use, and consequences of the obtained scores". And there is, in a

similar vein to this, another definition of validity which is "an evaluation of the accuracy, adequacy and appropriateness of the interpretations and uses of test scores or assessment results" (Linn & Miller, 2005, p. 68; Reynolds, Livingston & Wilson, 2006, p. 118).

Based on these concepts, I am able to realize that tests or instruments cannot be valid; so only inferences or interpretations can be valid simply because it is more accurate to say "the validity of the inference from test scores" than "the validity of the test" (McMillan, 2001, p. 17). There is also a need to be aware that there are a few varieties of validity; (a) performance on a universe of items (*content validity*); (b) performance on some criterion (*criterion-related validity*); and (c) the degree to which certain psychological traits are actually represented by test performance (*construct validity*) (Hopkins, 1998, p. 72). Considering the last type of validity here, it is evident that 'validity may be threatened when a test measures either less or more than the (exact scope of) construct it is designed to measure' (Reynolds, Livingston & Wilson, 2006, p. 118). However, these classifications of validity evidence are not well acknowledged by Messick (1995) for the reason that they cannot represent the overall interpretation of score meaning and social values of score applications. Instead he argues that the newly unified theory of construct validity, of which six general standards are (a) content, (b) substantive, (c) structural, (d) generalizability, (e) external, and (f) consequential criteria, should replace such degenerated and incomplete conventional validities (ibid., p. 741).

Compared with the complexity of validity evidence, reliability is conceived of as a straightforward concept. Initially it would be helpful to look at a succinct definition. In the context of psychological measurement, reliability means the consistency or stability of assessment results (Reynolds, Livingston & Wilson, 2006, p. 86). In addition to this, Linn and Miller (2005, p. 104) state that reliability means the degree to which various kinds of generalizations are justifiable. Kubiszyn & Borich (2003, p. 311) also observe that a test is reliable if it continuously generates the same, or nearly the same, ranks over repeated administrations, all other factors being equal. So it is recognizable that the level of reliability in a test is in proportion to that of consistency of the test scores. Here it is necessary to notice that as is the case of validity evidence, 'the scores or ranks, not tests or other instruments, are reliable' (McMillan, 2001, p. 36).

To elucidate the different concepts of validity and reliability, it may be useful to make a distinction between the two terms. Through the two sets of definitions specified above, it has been confirmed that validity deals with the essence and acceptability of the test scores; whereas reliability questions whether the scores are stable and dependable. So reliability is a *necessary* but *insufficient* condition for validity (Reynolds, Livingston & Wilson, 2006, p. 119). Obviously, this means that even though low reliability constrains validity, high reliability does not guarantee validity. What is substantive here would be how to improve the degree of reliability and how to develop a validity argument.

There are a few separate ways to estimate the reliability of measurement. Linn and Miller (2005, p. 121-124) introduce three factors to affect reliability measures: (a) the number of test items; (b) the spread of scores; and (c) objectivity of scoring. From a different angle, Kubiszyn & Borich (2003, p. 311-316) present four methods: (a) test-retest or stability; (b) alternate forms or equivalence; (c) split-half methods; and (d) Kuder-Richardson methods. Here one of the most natural solutions to enhance reliability may be simply to increase the number of tests and test items on an examination while maintaining the same quality as the original tests and the items. This approach is supported by the alternate-forms estimates and the *Spearman-Brown formula* which are used to predict the effects on reliability upgraded by adding the tests and test items (Linn and Miller, 2005; Reynolds, Livingston & Wilson, 2006). In the development of a validity argument, then, there is a general process of validation to support the intended inferences of test results and their values to the relevant applications. To investigate the validity of standardized achievement tests, i.e., the GCSE and A-levels, it is recommended to implement largely three validity arguments in which major lines of evidence based on *test content*, *relations to other variables*, and *consequences of testing* is integrated into a coherent commentary (Reynolds, Livingston & Wilson, 2006, p. 135-136). For these reasons, going back to the writer's first set of arguments – centered on the limited number of participants in A-levels, we are able to realize how seriously those are invalid and deviated.

(2) Evaluation of Arguments in Correlation, Validity and Reliability

The writer's questioning in this respect can be challenged and deconstructed as follows; in terms of correlation and regression, to make cogent his reasoning on the inadequacy of A-levels he should have explicated *insignificant correlation* first between students' GCSE scores and A-level grades, and then between their A-level grades and college GPAs with the aid of linear regression or regression equation. By doing so, empirical evidence showing flaws in the placement and selection processes underpinned by A-levels could have obtained. To nullify A-level exams, then, the first thing to do would be to demonstrate the *insufficient coefficient of determination* on the interdependency of the GCSEs and A-levels; a range of measurement criteria or standards by which students are transferred from 'mandatory and comprehensive' studies (GCSE) to 'preconditioned and specialized' courses (A-levels). And, the follow-up task would be to present the *low or negative validity coefficient*, which indicates the degree to which A-level measures predict or estimate academic performance on criteria-related measure during the college degree courses. Further, to make his propositions acceptable, he should also have verified a 'low degree of reliability' of A-level test results. For instance, it had to be shown that A-levels have had the insufficient number of tests, test items and alternate-forms, coupled with inherent difficulties in assuring the objectivity of scoring and excessive spreads of test scores, if any,

in the entire assessment processes.

These considerations should have focused on the probability that A-levels have failed to precisely measure the *potentiality* of all candidates for their successful university education, which can provide the examination board with a reliable barometer for such academic qualification. As described above, to work this out, the writer should have proved that the two sets of measurements, A-level test results and the actual performance of those admitted to higher education, have not indicated significant validity coefficient. In this connection the writer might well have instantiated that A-levels have not been an adequate testing device to reflect the generic combinatory effects of placement, formative, diagnostic and summative assessments. This could have supported the assumption that A-levels have been used to assess only the candidates' *performance level* at the final stage of secondary education, instead of measuring their *real potentiality* for undertaking rigorous college degree courses. These critical observations above, within the scope of correlation, validity and reliability, enable us to discredit the writer's key arguments noted in the first components. In the next section, some of the major conceptual issues related to his second set of arguments will be addressed.

2. Conceptual Issues

1) Accountability and Certification of High-Stakes Testing

[The Writer's Second Arguments: 6th Paragraph of Appendix A]

Unlike the technical issues I have so far investigated, this section will explore a few conceptual problems regarding the writer's lines of proposition in terms of educational assessment. Quite understandably these second arguments flatly deny the basic nature, roles and fairness of measurement tools in general, and imply that assessment is interwoven with social, economic and ethical issues. In this context, I find it necessary to take a look at the accountability and certification of high-stakes achievement tests.

① Accountability

A common sense tells us that educators might be influenced by their subjective judgments without relying on objective measurement data in the educational decision-making processes. So as the need for concrete evidence and assessment purposes, the issues of fairness and accountability mechanisms come in and are related to technical integrity of the high-stakes testing (Khattri, Reeve & Kane, 1998). In public recognitions of fairness, they provide two dimensions for its factor: (a) fairness of assessment systems whose purpose is to hold schools and teachers accountable for student performance; and

(b) fairness of treatment of individual students in assessment situations (ibid., p. 71). These two factors here seem to be compatible directly with the writer's second arguments: (a) the testing system is merely an artificial stepping-stone for advancing to the next phase of education; and (b) successful students have strong family backgrounds. These viewpoints demonstrate the fact that school administrators, teachers, and parents, including the government authorities, are all involved in the soundness of assessment system and high-stakes accountability mechanisms to support expenditures with valid evidence of student achievement.

② Certification

Educational decisions are normally based on the objective measurement data which provide the foundation for certification of students at all levels of school education (Black, 1998; Haladyna, 2002; Kubiszyn & Borich, 2003). In most cases, assessment for the certification takes the form of high-stakes achievement testing: e.g., the GCSEs and A-levels (a combined form of teachers' assessment and external test results) in the U.K.; and SAT (Scholastic Aptitude Test) and GRE (Graduation Record Examination) in the U.S. (Black, 1998; Linn & Miller, 2005). These tests for monitoring and selection decision-making are designed to measure the degree to which candidates have mastered their range of skills and knowledge required for a certain stage of education (Chatterji, 2003; Hopkins, 1998). From this viewpoint, it may be, in a broad sense, included in the writer's second set of arguments that A-levels have failed to play such an important role as certification in assessment.

As described earlier, aspects of accountability and certification may be considered to shed light on the social, economic and even ethical implications for educational measurement. Such issues, beyond the scope of technical concepts, seem to be fuzzy and vague by nature. In spite of this, the writer's disapproval of A-levels as inadequate testing from the far-reaching perspectives should have grounded on the inherent defects of the exams, if any, in accountability and certification.

2) Diagnosis and Reform of High-Stakes Testing System

[The Writer's Third Arguments: 7th Paragraph of Appendix A]

Undoubtedly, at this point the writer argues that A-levels are obsolete simply because of their outdated functions, offering the strong implications that their narrowly defined mechanisms cannot meet the needs for massive well-qualified manpower development in contemporary higher education. So I may well review the shortcomings of A-levels using relevant statistical findings in order to explore a reasonable solution to their current

problems.

① Diagnostic Review

Since A-levels were launched by the U.K. government in 1951, their test results have steadily increased over the past two decades with a 2005 pass-rate (A-E) of 96.2% (BBC News, 2006a). According to the media reports (BBC News, 2006b; Education Guardian, 2006a), for the year 2005 (June) series, a total of 783,878 (554,594 male, 229,284 female) candidates obtained their full A-level grades; for the AS-level, it was 1,079,566 (492,248 male, 587,318 female), and 22.8% of A-level final results were graded A; 23.8%, B; 23.3%, C; 23.3%, D; 17.2%, E; 9.1%, and 3.8% were not graded (U; Unclassified). Statistical surveys also indicate that mathematics was the subject with most A grades as a percentage of 58%, and language, science and math subjects tended to produce the highest proportion of A grades, and more seriously, over the last few years full A-level grades have been higher than AS grades. Namely, 22.8% of A-level grades are graded A, compared with 17.9% at AS-level (BBC News, 2006c; Education Guardian, op cit.). These numerical findings support that the *grade inflation* of A-levels makes it hard for universities to identify the best-qualified students, leaving something essential to be desired. Hence it is known that the U.K. government is still reexamining the current testing processes for further reforms.

② Reform Pressures

The most recent changes to A-levels, first introduced more than a half-century ago, date back to the year 2001, when the U.K. government adopted *Curriculum 2000* which divided the A-levels into two components; the AS and A2 exams (House of Commons, 2003). The reform guidelines provide that candidates are awarded an A-level qualification by achieving AS and A2 units satisfactorily; the former is generally taken in Year 12, and the latter in Year 13. Both of the modular courses are often the prerequisites for college-level study, viewed as college entrance examination, of which the policy objective is to offer students greater flexibility through the broadened range of elective subjects and learning types (ibid.).

In spite of this innovation, the public debates and concerns about the A-level grade inflation focus on the matter of distinguishing among a large number of students receiving A grades. Against social pressures on reforming devalued A-levels, the government and school authorities claim that the increased grades derive from students' improved achievement and teachers' advanced teaching methods (BBC News, 2006d). Students are also guided to select easier subjects in which they can receive top grades, specially trained to pass high-stakes tests at the cost of expanding their subject knowledge (Education Guardian, 2006b; House of Commons, op cit.). These tendencies indicate that any past

reforms focused on the internal mechanisms of A-levels have possessed their inherent limitations, and accordingly, combined types of measurement processes would be required to fill the gap.

IV. CONCLUSION

As given in Appendix A, the writer raises three sets of questioning in an effort to argue that the long-standing A-level system has failed and is no longer viable as the nation's high-stakes test. It is, however, my belief that his first arguments containing most critical misconceptions have been deconstructed because the statements are not supported by valid reasoning from reliable evidence.

In this connection, I would like to propose that a reasonable solution to replace the discredited A-levels should be to adopt an eclectic approach for assessing candidates' multiple capabilities; performance, potentiality and critical thinking skills (Lee, 2004). The following three sets of component-oriented assessments are designed to measure their high school academic achievements and capacity for tertiary education, with critical-logical reasoning abilities required for sound judgment and problem-solving tasks, under the condition that each university possesses its own discretion for determining the proportions of reflecting each of these assessment results:

(1) HSAP (High School Academic Performance) measured by school teachers with both formative and summative tests conducted through its entire periods of studies;

(2) CSAT (College Scholastic Abilities Test) measured by a national evaluation institute with standardized achievement testing at the end of secondary education; and

(3) CLRS (Critical-Logical Reasoning Skills) measured by each university with critical essay writing on a few philosophical, thought-provoking, and controversial topics.

It is, therefore, anticipated that this combinatory modality will be recognized as the optimal alternative to the controversial A-level system, so that students and parents may apply to their preferred institution in view of individual needs and strength, each university may evaluate applicants with its autonomous criteria, and the U.K. government and school authorities may use collected objective measurement data for developing the nation's future assessment processes.

The major implication deriving from this suggestion above is that three proposed sets of assessment systems are analogous largely to those of current university entrance examination administered in Korea, which may be recognized as the nation's well-integrated high-stakes testing tool for higher education.

REFERENCES

BBC News. (2006a). *'Modest' rise in A-level passes.* Retrieved June 11, 2006, from the World Wide Web: http://news.bbc.co.uk/1/hi/education/4160594.stm.

BBC News. (2006b). *All subjects A-level grades, 2005.* Retrieved June 11, 2006, from the World Wide Web: http://news.bbc.co.uk/1/shared/bsp/hi/education/05/exam_results /a_levels/html/all_subjects.stm.

BBC News. (2006c). *Mathematics (Further) A-level grades, 2005.* Retrieved June 11, 2006, from the World Wide Web: http://news.bbc.co.uk/1/shared/bsp/hi/education/05/ exam_results/a_levels/html/mathematics_further.stm.

BBC News. (2006d). *Minister attacks A-level critics.* Retrieved June 11, 2006, from the World Wide Web: http://news.bbc.co.uk/2/hi/ uk_news/education/3571138.stm.

Black, P. J. (1997). *Testing: Friend or foe?: The theory and practice of assessment and testing.* London: Falmer.

Chase, C. I. (1999). *Contemporary assessment for educators.* New York: Addison-Wesley Educational Publishers, Inc.

Chatterji, M. (2003). *Designing and using tools for educational assessment.* London: Pearson Education, Inc.

Education Guardian. (2006a). *AS-level results by subject 2005*. Retrieved June 11, 2006, from the World Wide Web: http://education.guardian.co.uk/alevel/page/ 0,,1551645,00.html.

Education Guardian. (2006b). *Pupils favouring 'easier' A-level subjects.* Retrieved June 11, 2006, from the World Wide Web: http://education.guardian.co.uk/alevels/story/ 0,,1707027,00.html.

Gardner, J. (2006). Assessment and learning: an introduction. In J. Gardner (Ed.), *Assessment and learning* (pp. 1-5). London: SAGE Publications LTD.

Haladyna, T. M. (2002). *Essentials of standardized achievement testing: Validity and accountability*. London: Allyn & Bacon.

Hopkins, K. D. (1998). *Educational and psychological measurement and evaluation* (8th ed.). London: Allyn & Bacon.

House of Commons. (2003). *Education and Skills – Third Report*. Retrieved June 12, 2006, from the World Wide Web: http://www.publications.parliament.uk/pa/cm200203/ cmselect/cmeduski/153/15303.htm.

Im, B. B., & Lee, J. H. (2004). The applicability of schema theory to scientific texts. *English Language & Literature Teaching, 10*(1) 1-22.

Khattri, N., Reeve, A., & Kane, M. (1998). *Principles and practices of performance assessment*. London: Lawrence Erlbaum Associates, Publishers.

Kubiszyn, T., & Borich, G. (2003). *Educational testing and measurement: Classroom*

application and practice (7th ed.). New York: John Wiley & Sons, Inc.

Lee, J. H. (2003). The implications of content schemata for standardized reading tests. *English Language & Literature Teaching*, *9*(special ed.), 79-98.

Lee, J. H. (2004). Analytical tools for ideological texts in critical reading instruction. *English Language & Literature Teaching, 10*(3), 89-112.

Linn, R. L., & Miller, M. D. (2005). *Measurement and assessment in teaching* (9th ed.). Upper Saddle River, NJ: Pearson Education, Inc.

McMillan, J. H. (2001). *Essential assessment concepts for teachers and administrators.* London: Corwin Press, Inc.

Messick, S. (1995). Validity of psychological assessment. *American Psychologist*, *50*(9), 741-749.

Pellegrino, J., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment.* Washington, D.C.: Center for Education, National Research Council.

Reynolds, C. R., Livingston, R. B., & Wilson, V. (2006). *Measurement and assessment in education*. Upper Saddle River, NJ: Pearson Education, Inc.

APPENDIX A

Source: The Times Educational Supplement (TES), www.tes.co.uk, P. 23, July 7, 2006
Writer: Paul Nightingale, Sociology Teacher at Mellow Lane School in Hillingdon, UK

Title: Filling A-level Vacuum

It is good to recognize that our exam system has failed, but university tests are no decent substitute, writes Paul Nightingale

(1) At the height of the summer exam season we do well to bear in mind the prime purpose of our education system. From the moment children begin school, this elitist system punishes failure by ranking in order to exclude. Such elitism is based on the view that any qualification is only important and valuable when possessed by a small minority. Now it seems that proposed university tests are needed because A-levels have failed to identify the most able. There might be sympathy for students who have to endure far too many tests; but how else to choose from so many well-qualified applicants?

(2) One should, however, resist the temptation to generalize here. Medicine is cited as an obvious case where admissions officers have an impossible task; and UCAS statistics for 2005 indicate that two-thirds of all successful candidates for medicine have 420-plus points (roughly, the equivalent of three As at A-level and a fourth A at AS, or better). Yet the 5,500 students in question are just 1.5 per cent of all students accepted for entry to all degree courses last year. It is hard to justify the widespread adoption of such tests, then, even if, for the sake of argument, we accept they might serve a purpose in that one subject

area. Rather, this is a political matter, since the introduction of tests designed to exclude candidates will, effectively, confirm the high status of the university in question; as well as reconfirming the "natural" link with private schools. None of this offers justification on educational grounds.

(3) It is a powerful myth that too many students are passing too many exams. Let us, rather, consider what actually does happen. GCSEs regulate entry, at 16-plus, to a range of advanced-level courses: and a little over half of all students, no more, gain the GCSE grades needed to even begin such courses.

(4) Furthermore, of those successful GCSE students, only two-thirds are still in full-time education at 19, their perseverance being, of course, strongly related to parental background. Successful students have successful parents. None of this can be contested; indeed, a year ago, the Government's own white paper acknowledged that less than 4 per cent of the age cohort gain three grade As or better at A-level. That figure represents two things, a small minority of students and a system that rigorously differentiates between candidates without any need for university tests.

(5) Several points can be made in the interests of clarity. First, any mention of wider participation has always been met by charges of "dumbing down"; but how do we know exams were sufficiently difficult when only a minority of students ever took them? If A-levels are an inadequate test now, this might well have been the case all along. Admissions tutors will use the proposed university tests to differentiate between well-qualified students; that they never had to do so before is hardly a criticism of A-levels as a mark of achievement.

(6) Second, we should acknowledge that passing exams only proves that students have done what they were asked in the exam room. One can only infer that they have earned the opportunity to move on to the next stage, where they will have to go on learning to be successful. Discussion of university tests has highlighted their inherent unfairness: students from advantaged backgrounds will benefit from specialist coaching and succeed by default, just as they always have. It is difficult, then, to see why anyone would think the test more valid than anything that has gone before.

(7) Thirdly, we might ask the purpose of education. Historically, A-levels have regulated entry to higher education and high-status occupations, which means they have now been exposed as an anachronism. Underpinning social and cultural reproduction, the system was effective when catering to a minority of students, precisely because it was only a minority. Yes, A-levels have been devalued as currency in the marketplace, but why should it matter if you now need better grades than would have been the case 20 years ago?

(8) There is much about our education system that needs reform; by all means do away with A-levels because their cover has been blown (and we have woken up to the fact that we no longer live in the 1950s). But don't assume that university tests are an adequate solution. **[The End]**

APPENDIX B

The "**A-level**", short for "**Advanced Level**", is a General Certificate of Education (GCE) qualification in Education in England, Northern Ireland and Wales, usually taken by

students during the optional final two years of secondary school (years 12 & 13, usually ages 16-18), commonly called the Sixth Form, or at a separate sixth form college or further education college, after they have completed General Certificate of Secondary Education (GCSE) or International General Certificate of Secondary Education (IGCSE) exams. The qualification is recognized around the world and is used as a sort of entrance exam for some universities. It is a non-compulsory qualification taken by students in England, Wales, and Northern Ireland. In Scotland, students usually take the Scottish Highers and Advanced Highers of the Scottish Qualifications Certificate. However, a small minority of schools offer the A-level as an alternative, typically private schools.

A-levels are also taken in some Commonwealth of Nations' countries and British Overseas Territories, including Bangladesh, Pakistan, Commonwealth Caribbean, Cyprus, Hong Kong, Malaysia, Mauritius, Sri Lanka, Singapore, Zimbabwe, Malawi, Gibraltar, Brunei, New Zealand, Malta, Zambia, South Africa and India. Due to respective changes in the systems, these examinations differ both in terms of content and style from the A-levels taken in the United Kingdom. The most extreme case is observed in Hong Kong. The Hong Kong A-levels (HKAL) maintained a high standard while the British A-level is accused of grade inflation, and over time became more and stricter compared to its British counterpart, as shown by NARIC. A research study into comparison of grades achieved in the Hong Kong HKCEE and HKALE with the GCSE and British GCE A-levels, conducted by NARIC UK. The British GCE A-levels are taken all around the world, as many international schools choose to use the British system as the examinations are widely recognized. Furthermore, students may choose to sit the papers of British examination bodies at education centers such as British Councils around the world (cited from www.wikipedia.org).

APPENDIX C

The **General Certificate of Secondary Education** (**GCSE**) is an academic qualification awarded in a specified subject, generally taken in a number of subjects by students aged 15–16 in secondary education in England, Wales, and Northern Ireland. (In Scotland, the equivalent is the Standard Grade.) Some students may decide to take one or more GCSEs before or afterwards; people may apply for GCSEs at any point either internally through an institution or externally.

The education systems of other British territories, such as Gibraltar, and the former British dominion of South Africa, also use the qualifications, as supplied by the same examination boards. The International version of the GCSE is the IGCSE, which can be taken anywhere in the world, and which includes additional options, for example relating to coursework and the language used. When GCSEs are taken by students in secondary education, they can often be combined with other qualifications, such as BTECs, the DiDA, or diplomas.

Education to GCSE level is often required of students who study for the International Baccalaureate or to GCE Advanced Level (A-level). GCSE exams were introduced as the compulsory school-leavers' examinations in the late 1980s by the then Conservative Party government, replacing the Certificate of Secondary Education (CSE) and GCE Ordinary

Level (O-Level) examinations (cited from www.wikipedia.org).

Examples in: English
Applicable Languages: English
Applicable Levels: College

Jong-Hee Lee
Department of English Language & Cultural Studies
College of Humanities & Social Sciences
Kangwon National University-Samcheok Campus
1 Joongang-ro, Samcheok-si, Kangwon-do 245-711 Korea
Tel: 033-570-6652
Email: freshfields@kangwon.ac.kr

Received in October, 2009
Reviewed in November, 2009
Revised version received in December, 2009