# The Characteristics of Absolute Grading of the College Scholastic Ability Test English Section

**Sang-Keun Shin**

(Ewha Womans University)

**Shin, Sang-Keun. (2018). The characteristics of absolute grading of the College Scholastic Ability Test English section. *English Teaching, 73*(3), 95-113.**

The purpose of this study was to critically examine the absolute grading system applied to the English section of South Korea's College Scholastic Ability Test (CSAT), in a departure from a relative grading system, for the first time in 2017. English language instructors as well as English education researchers have somewhat differing understandings of the concept of absolute evaluation/grading. This paper investigated the nature of absolute evaluation/grading as presented in theses and academic papers as well as research reports, English language assessment textbooks, and press releases from the Ministry of Education (MOE). The study identified four different testing contexts in which the term absolute evaluation/grading is applied: criterion-referenced testing; grading based on fixed cut scores; measurement against external criteria; and norm-referenced testing. While retaining the same test structures as in the prior norm-referenced testing situation, the CSAT English section corresponds to the second context, that is, interpreting results based on fixed cut scores. Achievement standards essential for producing criterion-referenced tests have so far been undefined, with cut scores being arbitrarily set. The examination concluded that the CSAT English section under an absolute grading scale is essentially a norm-referenced test that produces a significantly larger percentage of higher-level test-takers. There is a fair likelihood that, unlike MOE's claim, converting to an absolute evaluation/grading system may not bring about the intended positive change.

**Key words**: criterion-referenced tests, norm-referenced tests, cut score, absolute grading scale

## 1. INTRODUCTION

A variety of government policies intended to improve the level of English education in schools have been introduced in South Korea, but no corresponding increase in the level of

English education in secondary schools in Korea has been achieved (Chung & Choi, 2016). Under the pretext of normalizing public education (Kim, 2015; Kwon, 2015), the English section of the College Scholastic Ability Test (CSAT) has also been the target of numerous policies, including increasing the number of listening items; basing 70 percent of the items on Educational Broadcasting System (EBS) workbooks; developing the National English Ability Test (NEAT) as a possible replacement; and switching from a relative grading system for the CSAT English section to absolute evaluation/grading (in Korean, *jeoldaepyeongga*). The primary rationale for these reforms put forth by the Ministry of Education (hereafter, the Ministry or MOE) was to improve the level of English education in schools.

The most recently adopted initiative is the transition from relative grading in the CSAT English section to absolute grading, first implemented in 2017. The Ministry stated that the switch was imperative to mitigate the adverse effects of the system of relative grading (in Korean, *sangdaepyeongga*) (MOE, 2014). According to the report, under the relative grading that has been in effect until now, students' performance was compared to that of others, creating endless competition to earn even one or two points more than their competitors. The switch to absolute grading is expected to alleviate this undue competition in that classroom instruction is oriented to the improvement of students' communicative competence rather than to preparation for the test.

When it comes to absolute evaluation/grading, views about its capacity to achieve the intended positive effects in the classroom are as likely to be skeptical as they are favorable. As with other policies adopted in South Korea so far, many experts and classroom teachers argue that switching to absolute grading will not bring about the intended changes in the classroom. On the contrary, they claim that it will diminish the level of English education in the schools. Other subjects on the CSAT such as mathematics and language arts continue to be scored through relative grading, and private education may end up simply leaning more toward these subjects. The policy is also likely to increase the percentage of students who receive higher grades, which in turn will reduce the determinative weight of English in college admissions decisions. Once reaching the top-level score band becomes more attainable, students may feel less motivated to study English, leading to an overall decline in their English proficiency (Kang, 2014).

In order to precisely understand the impact of an absolute grading policy on classroom instruction, an accurate understanding of absolute evaluation/grading is requisite. Although terms like absolute grading, absolute scoring, absolute rating, and absolute evaluation are widely used in schools, the concept reflects different meanings in many cases and is hardly dealt with in English language assessment textbooks. Despite much interest in absolute evaluation/grading, research on the inherent characteristics of the corresponding test type itself is scarce. Some scholars claim that tests using an absolute grading system are norm-

referenced tests (Boo, Chae, So, & Kim, 2005; Kwon, 2015; Shin, 2010), while others claim they are criterion-referenced tests (Sung, 2014). Still others consider the absolute grading system as being tantamount to criterion-reference testing, though the characteristics being depicted are quite different.

The goal of the present study is to analyze the characteristics of testing under an absolute grading system in the English education field. Specifically, this paper examined the characteristics of absolute grading depicted in a range of literature, including materials published by MOE and the national curriculum, press releases, academic papers, and graduate theses. It goes on to examine how absolute grading in testing is similar to and different from norm-referenced and criterion-referenced testing.

## 2. THEORETICAL BACKGROUND

### 2.1. Comparison of Norm-referenced Tests and Criterion-referenced Tests

English language tests in South Korea can largely be divided according to their use of either relative or absolute grading systems. However, these terms are not used by experts in educational assessment. Instead, they classify tests into two types depending on how the results are interpreted—namely norm-referenced testing and criterion-referenced testing. Experts agree that the concept of relative evaluation/grading is basically interchangeable with that of norm-referenced testing. Therefore, relative grading will hereafter be referred to as norm-referenced testing. When it comes to absolute evaluation/grading, however, scholars offer wildly conflicting opinions about the relationship between an absolute grading system and criterion-referenced testing. While some scholars argue that it is equivalent to criterion-referenced testing (Sung, 2014), others claim that it is actually a type of norm-referenced testing (Bae & Shin, 2016; Boo et al., 2005; Kang, 2017; Kwon, 2015; Shin, 2010). In order to predict how the transition to the absolute grading system of the CSAT English section will impact classroom instruction, it is necessary to better understand the characteristics of the new test grading format. Unfortunately, most researchers have conducted their studies without accurately defining the nature of the test.

Language tests can fall into the category of either norm-referenced testing or criterion-referenced testing according to how their results are interpreted (Bachman, 1990; Brown & Hudson, 2002). Norm-referenced testing compares a test-taker's performance against the scores of either a selected subset of those who have already taken the same exam, called a "norming group," or other test-takers taking the same test. Therefore, it is crucial in norm-referenced testing to ensure the test's discriminative ability. Norm-referenced tests typically select items of average difficulty and eliminate items that all students are likely to

answer correctly in order to spread test-takers' results out along a continuum of proficiencies to differentiate various performance levels (Linn & Gronlund, 1995). The results are used to "rank-order" test-takers and make decisions such as passing and failing according to their rankings. In schools, this is often referred to as relative grading.

Criterion-referenced tests, on the other hand, "assess student achievement in terms of a certain criterion standard [and] thus provide information as to the degree of competence attained by a particular student" (Glaser, 1963, p. 520). Thus, rather than ranking students in relation to others, criterion-referenced testing aims to simply determine which test-takers have mastered a particular domain of knowledge or skills, according to set performance standards. Thus, they provide an absolute interpretation (i.e., what a student can and cannot do) and not a comparative interpretation (i.e., how a student's performance compares to others) (Brown & Hudson, 2002; Davidson, 2012). If all the students have mastered the goals, then they all receive an 'A'. When score interpretation is based solely on the achievement of specific learning objectives, it is sometimes called objective referenced (Linn & Gronlund, 1995). To make such decisions, it is essential to establish specific learning goals, then employ corresponding test tasks tailored to those goals, and establish criteria to determine mastery of the goals. In order to produce valid criterion-referenced tests, the performance goals that students are expected to achieve must be clearly defined. For example, in developing criterion-referenced tests for a particular unit in a textbook, Shin and Cho (2000) set goals pertaining to the learning objectives of a given unit.

The next important step is to develop test tasks for each learning goal so that testers can determine whether each of the achievement goals has been attained. For example, if the target learning goal is to identify the main idea of a paragraph, teachers ask students to write down the gist of the reading passage or to choose from among multiple options. Another important element in criterion-referenced testing is setting criteria for 'how much is enough' on the performance scale. The minimum performance level at which it can be determined that the learning objectives have been achieved must be defined. Criteria can be established in various ways. For example, testers (1) describe the specific language use tasks students are able to perform; (2) suggest the probability that students who have achieved the specific goals will answer items correctly; (3) indicate the minimum score that successful students will receive; and (4) develop performance level descriptors for each level so that students' performance can be compared to objective standards. As the final step, committees of experts implement standard-setting procedures by determining cut scores for corresponding performance levels.

The score range in criterion-referenced tests may be very narrow, an indicator that the majority of students have achieved the learning goals being tested. That is exactly what we expect and want from an achievement test. An example of a criterion-referenced test is the driver's license exam. Regardless of testing center, driving exam tasks and scoring

standards are the same. Examples of well-known large-scale criterion-referenced tests administered all across the United States include Advanced Placement exams and the National Assessment of Educational Progress.

As we have seen so far, norm referencing and criterion referencing are two distinct concepts in terms of the purpose of an evaluation, the item development process, and interpretation of test results. Given that the term absolute grading is used more than criterion-referenced testing in Korean context, the purpose of this study is to determine whether absolute grading can be regarded as the identical type of criterion-referenced testing.

## 2.2. History of the English Section of the College Scholastic Ability Test

After several experimental evaluations from 1990 to 1992, the College Scholastic Ability Test (CSAT) was first implemented in 1993 and has been in operation until now for 25 years. From 1990 to 1992, a total of 7 experimental evaluations were conducted to establish the nature of English test in CSAT and to verify its enforceability. According to the test structure established through the experimental evaluations, 70 minutes were given to complete the test, which consisted of a total of 40 questions (5- option). From the second experiment evaluation, listening assessment was introduced for the first time in the history of English education in Korea.

For the tests conducted in the school years from 1994 to 1996, the number of questions was increased to 50 and the test time was extended by 10 minutes, from 70 to 80 minutes. Accepting the persistent demand of the English education academia for the assessment of speaking ability, five speaking questions (indirect evaluation) were newly included in the CSAT conducted from the 1997 school year. Thus, from 1997 to 2000, English test in the CSAT consisted of 55 questions, and the given test time was 80 minutes. However, the English test in the CSAT changed again in 2001, consisting of 50 questions which is conducted for 70 minutes. In 2005, the total score of English test was changed to 100 points. For the school year of 2008, there was a temporary change in the score system report. In order to weaken the weight of the CSAT in the admission selection process, a 9-grade system was introduced for each section and subject to indicate only the grade in the CSAT report without the standardized score or the percentile score. This system, however, changed again in 2009, to include the standardized score, percentile score as well as the grade in the CSAT report.

Policy makers made many changes in classes for the 2014 academic school year under the pretext of normalizing public education. In 2013, the 'level-based (type A/B) CSAT' was implemented for the first time according to the 'CSAT Revision Plan for the 2014 Academic Year.' In the same year, however, the Park Geun-hye administration (the Office

for Government Policy Coordination, May 28, 2013) decided to discontinue the level-based (type A/B) CSAT in accordance with the 'College Admission Streamline Policy.' To lessen the test burden of the examinees, the total number of questions for each level (A/B type) of English area in the CSAT (conducted in 2013 and 2014) was reduced by 5 questions to 45 questions, consisting of 22 questions in the listening section and 23 questions in the reading section. The test time was maintained at 70 minutes, although the time allotted for the listening section was extended by 10 minutes to 30 minutes considering the increased number of questions in the section, while the time allotted for the reading section was reduced by 10 minutes to 40 minutes.

The level-based CSAT was no longer implemented for the English area from the 2015 academic year, and the number of questions was adjusted as well, reducing the number of questions for the listening section by 5, while increasing the number of questions for the reading section by 5. The number of indirect speaking questions has been decreased by 3 from the CSAT in 2014. The test time and the total score were maintained at 70 minutes and 100 points, respectively.

The CSAT following a relative evaluation system was found to be problematic because it leads to infinite competition and excessive learning beyond the range and level of the set curriculum. In addition, it was pointed that a balanced improvement of students' English ability is restricted because the lessons in the school settings are oriented to solving questions in preparation of the CSAT rather than to the practical improvement of students' English ability, and that the exam questions prepared with high level of difficulty to distinguish the examinees are leading to unnecessary and excessive learning pressure and private tutoring expenses on the part of students and their families. Accordingly, absolute evaluation was adopted for the English area in the CSAT and implemented for the first time in November 2017. In the CSAT report, therefore, the result of the English test as well as that of Korean History are indicated only in grade according to absolute evaluation. Here, the grade is determined based on a 9-grade system. The total score for the English area is 100 points, and the cut- scores that divide the 9 grades are set at 90, 80, 70, 60, 50, 40, 30, and 20 points.

## 3. METHODS

### 3.1. Data

Multiple databases were used to identify relevant studies for the present meta-analysis: DBpia, KISS (Koreanstudies Information Service System), Kyobo Scholar, RISS (Research Information Service System), and NADL (National Assembly Digital Library).

In the search for relevant literature, the following Korean key words were used: English subject *jeoldaepyeongga* (absolute evaluation/grading), CSAT *jeoldaepyeongga* (absolute evaluation/grading), and English subject criterion-referenced testing. An analysis of the primary data found that Korean researchers used the term 'reference testing' (*gijunpyeongga*) when referring to criterion-referenced testing. In addition, many studies compared absolute evaluation/grading (*jeoldaepyeongga*) with the achievement evaluation system (*seongchwipyeonggaje*) that was adopted in 2012 for grading students' academic performance in schools. These two keywords were then used to perform a subsequent search for additional literature. The final analysis covered 11 graduate theses, six academic journal articles, two media reports from MOE, and two research reports issued by the Korea Institute of Curriculum and Evaluation (KICE). Meanwhile, as it was not feasible to use online searching to determine whether absolute grading was dealt within textbooks on English language assessment, an online search was conducted to locate testing books using the following keywords: English language testing, language testing, and testing in primary English language education. Among the nine books identified, seven of them dealing with the concept were included in the analysis. All the studies included are marked with an asterisk * in the Reference section.

## 3.2. Data Analyses

One researcher and one research assistant initially performed independent search for resources using the specified keywords. They compared their respective findings to select appropriate resources for final analyses. They then worked independently to code each study according to the following criteria: How is terminology used differently to distinguish tests according to the way test results are interpreted; how is the relationship between criterion-referenced testing and absolute evaluation/grading described; how is *jeoldaepyeongga* (absolute evaluation/grading) translated into English; and what tests are named as examples of absolute evaluation? The initial intercoder agreement between the two researchers was 97%. In deriving the result for the final analysis result, the disagreement between the two researchers was resolved through interactive discussion.

## 4. RESULTS AND DISCUSSIONS

This section reports the results of the literature analysis conducted to find out what types of evaluation are referred to by the term *jeoldaepyeongga* (absolute evaluation/grading). It goes on to investigate which characteristics of these types apply to absolute grading of the CSAT English section.

## 4.1. How the Terminology Is Used in the Literature

### 4.1.1. Same concept as criterion-referenced testing

The first type of usage for *jeoldaepyeongga* (absolute evaluation/grading) is to signify the same concept as criterion-referenced testing. Nearly all of the data sources fall into this category. For example, usage of the term in the seven English language assessment textbooks belonged to this first type (Im, 2005; Im, Yeon, & Jang, 1999; Kim, Kim, Mun, & Jeon, 2012; Kim, 2009; Kim, 2014; Lee et al., 2012; Lee, 2012). These books classified English language tests as using either absolute or relative grading and asserted that the term *jeoldaepyeongga* (absolute evaluation/grading) is interchangeable with criterion-referenced (*jungeochamjo*) testing. Yet in describing absolute evaluation/grading, they do not provide examples of pertinent tests. Thus, it was not clear whether the authors believe that the current CSAT English section which rates test-takers according to fixed scores can be regarded as a criterion-referenced test.

Of the 11 graduate theses dealing with absolute grading of the CSAT English section, ten also belong to this type (Im, 2017; Jang, 2017; Kang, 2015; Kim, 2013; Kim, 2018; Kwon, 2018; Lee, 2015; Shin, 2016; Song, 2017; Yi, 2016). Like the assessment textbooks, they also classified English tests as applying either absolute or relative evaluation and stated that absolute evaluation is often referred to as criterion-referenced testing. They also presented the CSAT English section under absolute grading as an example of criterion-referenced (*jungeochamjo*) testing. Finally, among the six research papers, three of them equated absolute evaluation with criterion-referenced testing and claimed that the current CSAT English section under absolute grading is a criterion-referenced test (Choi & Choi, 2016; Kim & Kang, 2015; Kim, 2015).

It is dangerous, however, to equate absolute evaluation with criterion-referenced testing because the application of absolute evaluation/grading differs from criterion-referenced testing in many aspects. As will be explained in the following passage, the concept embodied by criterion-referenced testing is essentially distinct from the CSAT English section, in which grades are based on fixed scores. However, due to the prevalence of claims that the CSAT English section is in the same vein as criterion-referenced testing, there is a great risk that people will have the misconception that these two types of tests are identical. As for the English translation of the term *jeoldaepyeongga* (absolute evaluation/grading), the majority of the studies translated it as 'criterion-referenced testing', which is arguably misleading.

### 4.1.2. Assessment in which grade determination is based on fixed scores

The second definition of *jeoldaepyeongga* (absolute evaluation/grading) found in the data refers to a type of assessment in which students' grades are determined based on fixed scores. Reports issued by MOE and KICE are along this definition (KICE, 2014; MOE, 2014, 2017b). Even though the characteristics of absolute evaluation/grading they depict are very similar to those of criterion-referenced testing, they do not explicitly state that absolute evaluation/grading is the equivalent of criterion-referenced testing.

Meanwhile, a research report issued by KICE offers examples of absolute evaluation/grading, listing tests that reports results based on fixed scores. These examples include domestic tests such as TEPS (Test of English Proficiency), FLEX (Foreign Language Examination), and PELT (Practical English Language Test) as well as official international English tests such as IELTS (International English Language Testing System), GEPT (General English Proficiency Test), and assessment tools of ACTFL (American Council on the Teaching of Foreign Languages). The presented examples indicate that its definition of absolute evaluation/grading refers to assessment tools in which students' performance is interpreted according to levels or bands. It is also important to note that all the above tests, excluding ACTFL Oral Proficiency Interview (OPI), are technically classified as tests that adopt norm-referenced evaluation systems. While these tests provide descriptors for each performance level, details are lacking; it can hardly be said that those descriptors were used to develop the tests. Since absolute evaluation/grading can be sharply different from criterion-referenced testing, it is invalid to argue that various positive changes in classroom instruction that can be expected only through criterion-referenced testing would be induced by adopting absolute evaluation/grading.

### 4.1.3. Tests measuring student performance against external criteria

The third definition describes absolute evaluation as interpreting test results against detailed performance level descriptors. Tests adopting this approach provide performance level descriptors that depict what students can do in English at each level. One of the best examples of criterion-referenced testing is the ACTFL Oral Proficiency Interview (OPI). Having developed highly detailed proficiency guidelines for each level, ACTFL requires test-takers to perform speaking tasks and then assigns them an appropriate speaking level.

Three previous studies belong to this category (Jung, 1999; Kim, Ahn, Oh, & Kim, 1999; Lee, 1999). Specifically, Kim, Ahn, Oh, and Kim (1999) indicated that the achievement goals proposed in South Korea's national curriculum are far too ambiguous to use as the basis for producing textbooks or developing evaluation tools. Bearing this limitation in mind, they proposed English speaking and writing achievement goals for

students who successfully completed the high school curriculum in South Korea, referring to the detailed ACTFL Guidelines, ILR (Interagency Language Roundtable) scale, and ASLPR (Australian Second Language Proficiency Ratings). Jung (1999) developed evaluation criteria and tasks to assess the English speaking ability of Korean university students based on the ACTFL OPI. This approach is considered to be most consistent with criterion-referenced testing in that it purports to develop performance standards for assessing students' performance.

### 4.1.4. Norm-referenced testing

The fourth group of studies assert that absolute evaluation/grading is different from criterion-referenced testing. Although the grade classifications are based on fixed scores rather than comparative rankings, the cut scores are arbitrarily set without reference to external criteria. Only one assessment textbook (Shin, 2010) and one graduate thesis Lee, 2018) fall into this category. Two research papers also claim that absolute evaluation/grading reflects norm-referenced testing (Kang, 2017; Kwon, 2015). Strangely, the studies somewhat contradictorily translated *jeoldaepyeongga* (absolute evaluation/grading) as 'criterion-referenced testing' in English.

In summary, analyzing the characteristics of absolute evaluation in the English education field, this study found that the term entails four different types of testing contexts within the literature. Nearly all of the studies classified English tests as either absolute or relative evaluation, and nearly all also described the current CSAT English section applying absolute grading as a criterion-referenced test.

## 4.2. Comparison of the Characteristics of Absolute Evaluation, Norm-referenced Testing, and Reference Evaluation

So far, we have analyzed various interpretations offered under the heading of *jeoldaepyeongga* (absolute evaluation/grading) in the field of English education. It is generally used in four different contexts. The majority of the studies claimed that absolute evaluation and criterion-referenced testing are interchangeable and that the currently implemented CSAT English section is also a criterion-referenced test. However, there were a few studies that classified it as a norm-referenced test.

As for the absolute evaluation/grading of the CSAT English section, it conforms to the second type of definition in which the test development follows that of the norm-referenced test but the test result is interpreted based on a fixed score. This section analyzes how the absolute evaluation/grading approach of the English section of CSAT is similar to and different from norm-referenced testing and criterion-referenced testing, respectively.

### 4.2.1. Purpose of evaluation

The Ministry of Education maintains that that the purpose of absolute evaluation/grading is to see how well each student has attained achievement goals, as opposed to identifying their rankings relative to other students. As such, the purpose of absolute evaluation appears to be similar to that of criterion-referenced testing. However, towards fulfilling this purpose, there must be clear definitions of the achievement goals and criterion performance levels that form the basic requirements for criterion-referenced testing. No such work has gone into the development of the CSAT English section; it simply divides test-takers into nine levels based on fixed cut scores. Therefore, it should be regarded as norm-referenced testing, not criterion-referenced testing. The only difference between the previous norm-referenced testing (relative evaluation/grading) and the present absolute evaluation/grading of CSAT English section is that, while only the top 4 percent of the test-takers received a ranking of Level 1 in the former assessment, any student who gets 90 points or higher will be placed in the Level 1 score band for the current assessment. The Ministry maintains that the difficulty level of the test will remain similar to that of the relative or comparative grading system. If that is indeed the case, it is highly likely that a much higher ratio of test-takers will successfully achieve the top-level score band.

### 4.2.2. Test development procedures

As norm-referenced testing is conducted to discriminate between test-takers' proficiency levels and rank them relative to each other, item discrimination is crucial; items that are overly easy are modified. Criterion-referenced tests, on the other hand, include items that are directly relevant to the learning objectives to be measured. Thus, item discrimination or item difficulty does not take a distinctive role. The most critical concern is how well items assess mastery of the target learning goals. Thus, test construction is essentially different for these distinct types of tests (Hudson, 2013; Linn & Gronlund, 1995).

According to the Ministry, the CSAT English section intends to focus on whether learning goals are achieved, rather than discriminating between students' performance levels. However, in the interest of a smooth transition to the absolute grading system, there will be no change in the test, test items, or test development procedures. The only difference lies in how the tests results are interpreted. Instead of assigning rankings based on fixed ratios, the new system assigns grades or levels based on fixed cut scores. This approach is completely different from the way in which the results are interpreted in criterion-referenced testing. Criterion-referenced tests interpret results based on established criteria. If every single test-taker were assigned to the Level 1 score band, it would hardly be an issue, but rather a desirable outcome because it demonstrates that the learning goals

have been successfully achieved. The very fact that the same test development procedures are still employed indicates that the test under the new absolute grading system cannot be regarded as criterion-referenced testing.

In proposing the use of fixed cut scores in the CSAT English section grading procedure, KICE (2014) emphasized that it is critical to maintain the same level of test difficulty across all administrations. The fact that the ratio of students attaining each grade or score band is liable to vary according to fluctuating test difficulty levels suggests that it is not a criterion-referenced test. It should be noted that in criterion-referenced tests, such as the driver's license examination and the medical licensing examination, consistency in the difficulty levels must be maintained constantly. To meet the conditions of criterion-referenced testing, appropriate achievement goals have to be established and items based on those goals should be produced. After this process, cut scores able to differentiate nine levels have to be derived by determining performance or achievement standards. One circumstance to be noted is that the achievement standards presented in South Korea's National Curriculum for English are overly broad and sometimes vague, and consequently, it is practically impossible to develop criterion-referenced tests based on them (Kim et al., 1999; Shin, 2017). There is no clearly defined level of English proficiency that should be attained by students who have successfully completed the secondary English curriculum. Nevertheless, the Ministry has stated that if students earn a score of 90 or higher through absolute grading of the CSAT English section, we can conclude that they have successfully achieved the curriculum's learning goals.

### 4.2.3. Criteria for interpretation of test results

In norm-referenced testing, test results are interpreted in terms of each test-taker's relative standing against other test-takers or with reference to the performance of a selected group. The prior grading procedures for the CSAT English section, which conformed to norm-referenced testing system, used a Stanine scale—that is, students' test scores were scaled according to a nine-point standard scale.

When implementing criterion-referenced testing, a minimum performance level must be established in order to interpret the results. For example, if a pass-fail decision is required, the level of performance necessary to pass the test needs to be defined in detail, and cut scores representing the minimum performance level for a passing grade must be set accordingly.

What distinguishes absolute evaluation/grading from criterion-referenced testing in English assessment is whether or not detailed descriptions of test-takers' corresponding English ability for each grade are provided. In norm-referenced testing, the only information revealed by the results is that a test-taker who receives a top-level grade is in

the top 4 percent of test-takers, having performed better than 96 percent of his or her peers. However, in the case of criterion-referenced testing, the results are intended to indicate what test-takers in Level 1 are able to do in English, along with an explanation that describes difference between Level 1 and Level 2 English proficiency.

The only difference between absolute evaluation/grading and the previous norm-referenced format of the test is the method used to differentiate students' grades according to the nine levels: the former Stanine scale vs. the present fixed ratios, such as 90 percent or 80 percent of the total score. In the latter scenario, there is a 10-point difference between grades. Since the absolute grading now being implemented in the CSAT English section uses arbitrary fixed cut scores as the basis for assigning grades, the only piece of information the results can provide is that Level 1 test-takers earned higher scores than L2 test-takers. That is, test-takers classified as Level 1 scored 90 or higher, while those in Level 2 scored in the range of 80 to 89. Thus, the difference between Levels 1 and 2 could be as small as one point. The system is referred to as absolute, not relative, evaluation simply because, as opposed to individuals' rankings varying relative to other individuals' performance, it distinguishes test-takers' grades using fixed scores based solely on the number of points they earned. This does not meet the definition for a criterion-referenced test.

Regarding the implementation of absolute grading for the CSAT English section, the Ministry of Education (2014) clarified that the test is meant to assess how much students have achieved, and that test-takers whose achievement falls within a specified range is to receive the same level, taking no account of their relative rank compared to others. However, what students are expected to achieve has not been defined, nor have achievement criteria been used as the basis for developing test items. In addition, the degree of achievement expected at each grade level has not been defined. That is to say, except for the grading method, the test remains the same as it was under the relative grading system. As mentioned above, scholars have expressed concern over the adoption of an absolute grading system for the CSAT English section, fearing that it is fairly likely that the influence or weight of English in the college entrance examination will be weakened, owing to a larger ratio of students being placed in Level 1.

If the test difficulty level is reduced, the proportion of students receiving higher grade will increase. In other words, increasing number of students will belong to Level 1. Even if the Ministry adheres to their intentions to maintain the difficulty level constant, the proportion of students with higher grades is certain to increase. The following table shows the percentage of students placed in each level according to the results of the CSAT English section under the first implementation of the absolute grading system in the 2018 academic year as compared with the previous year.

**TABLE 1**

**Percentage of Test-takers per Level**

| Level | 2017 CSAT English Results (Norm-referenced) | 2018 CSAT English Results[a] (Absolute Grading) | June 2018 CSAT English Mock Exam Results[b] |
|---|---|---|---|
| 1 | 4 | 10.03 | 4.19 |
| 2 | 7 | 19.65 | 10.02 |
| 3 | 12 | 25.43 | 16.31 |
| 4 | 17 | 17.97 | 18.76 |
| 5 | 20 | 10.49 | 15.71 |
| 6 | 17 | 6.74 | 12.12 |
| 7 | 12 | 4.74 | 8.96 |
| 8 | 7 | 3.49 | 8.32 |
| 9 | 4 | 1.46 | 5.60 |

*Note.* [a] The data are from the Ministry of Education (2017a). [b] The data are from the Korea Institute for Curriculum and Evaluation (2018).

As the table shows, the ratio of test-takers who attained Level 1 under the absolute grading system was 2.5 times higher than in the norm-referenced testing period, increasing to 10.03 percent from 4 percent. What is even more noteworthy is the variation shown between the results for the official 2018 CSAT and the June mock exam. The disparate results seem to indicate that the level of difficulty varied between the two administrations, with the mock exam being more difficult. Such variation in difficulty suggests that the CSAT English section under absolute grading system is not a criterion-referenced test.

**TABLE 2**

**Comparison of Test Characteristics**

| | CSAT English Section (Norm-referenced) | CSAT English Section (Absolute Grading) | Criterion-referenced Testing |
|---|---|---|---|
| Purpose | Discriminating test-takers' English proficiency | Discriminating test-takers' English proficiency | Understanding test-takers' achievement levels |
| Achievement level | Undefined | Undefined | Must be defined |
| Item discrimination | Critical | Critical | Unimportant |
| Cut score | Pre-determined (arbitrary fixed ratios) | Pre-determined (arbitrary fixed cut scores) | Set after test administration (set in consideration of achievement level) |
| Desirable score distributions | Normal distribution | Normal distribution | Negatively skewed distribution |
| Type of interpretation | Relative | Relative | Absolute |
| Implications of grade difference | Score difference | Score difference | English proficiency difference |

Table 2 below summarizes the characteristics of the norm-referenced CSAT English

section, the CSAT English applying absolute grading, and criterion-referenced tests in principle. As can be seen in Table 2, the features of the CSAT English section under the absolute grading system are very similar to the norm-referenced CSAT English section administered until 2016.


## 5. CONCLUSION

The purpose of this study was to investigate the characteristics of CSAT English section after its conversion to an absolute grading format in 2017. Absolute evaluation/grading is a concept widely understood in South Korea as standing in contrast to relative testing. This paper examined the similarities and differences between absolute evaluation/grading and criterion-referenced testing. The analyses of the literature showed that the term *jeoldaepyeongga* (absolute evaluation/grading) is used somewhat differently among researchers. Some experts claim that absolute evaluation is the same as criterion-referenced testing. Other scholars argue that a test that uses fixed cut scores to determine test-takers' levels is absolute evaluation. In addition, several cases occurred in which researchers translated *jeoldaepyeongga* (absolute evaluation/grading) as 'criterion-referenced testing' in English and then described the characteristics of fixed cut scores rather than criterion-referenced testing.

The new grading approach to the CSAT English section is referred to as absolute evaluation/grading since grades are assigned based on fixed cut scores. The highest grade is awarded to test-takers receiving 90 points and above, and the grade levels are divided at 10-point intervals. Since the cut scores themselves are fixed, the current CSAT English section is often misperceived as a criterion-referenced test. However, the test is essentially one variety of relative testing. This is because it places all students who score within a certain point range into the same level. For example, test-takers who score at least 90 will receive the highest grade while those who receive scores ranging from 80 to 89 receive the second highest grade. This scoring system differs from how criterion-reference testing interprets results. Two critical elements for successful implementation of criterion-referenced testing are performance level descriptors and cut scores. In order to produce criterion-referenced testing, the achievement goals, that is, the English proficiency of students at particular level, must be clearly defined, and then test items for each goal should be produced. Criterion levels of performance need to be established for making decisions regarding test-takers' performance through standard-setting procedures. However, such test development stages were not undertaken to construct the CSAT English section. The composition of the test items and the test development procedure remain the same as in the previous norm-referenced CSAT English section. In addition, current cut scores are

established arbitrarily without any reference to external achievement standards.

In conclusion, absolute evaluation/grading of the CSAT English section is quite unlike criterion-referenced testing. Rather, the current absolute grading system takes the form of norm-referenced evaluation, and chances are that the new test format will increase the proportion of test-takers belonging to the upper levels. In fact, a much higher rate of students achieved the first-grade level on English section in November 2017 compared to the test-takers who took the exam in November 2016.

Since the term 'absolute evaluation/grading' is used in reference to different concepts and is especially prone to be misconceived as equivalent to criterion-referenced testing, dividing evaluation tools into either absolute testing or relative testing is an undesirable approach. Rather, tests should be classified as either criterion- or norm-referenced testing, with absolute evaluation being regarded as one variety of relative evaluation. With this in mind, translating *jeoldaepyeongga* (absolute evaluation/grading) into 'criterion-referenced testing' should also be avoided. It is an empirical question whether the adoption of absolute grading has induced positive changes in the classrooms as asserted by the Ministry of Education. One thing that is certain, though, is that we should not expect the changes that would have occurred if criterion-referenced testing had been adopted instead. After all, the CSAT English section under absolute grading is still a norm-referenced test.

## REFERENCES

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bae, W. & Shin, S.-K. (2016). An exploratory study of status quo of standards-based testing system and its impacts on English language instruction in Secondary Schools. *Secondary Education Research, 64*(3), 611–634.

Boo, J. Y., Chae, S. H., So, K. H., & Kim, J. C. (2005). *Educational curriculum and evaluation for pre-service and in-service teachers*. Seoul: Kyoyukkwahaksa.

Brown, J. D. & Hudson, T. (2002). *Criterion-referenced language testing*. Cambridge: Cambridge University Press.

*Choi, S. J., & Choi, J. G. (2016). Perceptions of English teachers on the new criterion-referenced assessment in the CSAT. *Studies on English Language & Literature, 58*(4), 371–402.

Chung, J., & Choi, T. (2016). English education policies in South Korea: Planned and enacted. In R. Kirkpatrick (Ed.), *English language education policy in Asia* (pp. 281-299). Switzerland: Springer International.

Davidson, F. (2012). *Test specifications and criterion referenced assessment*. In G. Fulcher

& F. Davidson (Eds.), *The Routledge handbook of language testing* (pp. 197–207). New York: Routledge.

Glaser, R. (1963). Instructional technology and the measurement of learning outcomes. *American Psychologist, 18*(8), 519–522.

Hudson, T. (2013). Criterion-referenced approach to language assessment. In A. Kunnan (Ed.), *The companion to language assessment* (pp. 561–577). New York: Routledge.

*Im, B.-B. (2005). *How to evaluate English education*. Seoul: Hangukmunhwasa.

*Im, B.-B., Yeon, J.-H., & Jang, G.-S. (1999). *English teaching and learning assessment*. Seoul: Hangukmunhwasa.

*Im, S. O. (2017). *A study of teachers', students', and parents' perceptions of criterion-referenced evaluation in the English section of the CSAT*. Unpublished master's thesis, Chongshin University, Seoul.

*Jang, H. S. (2017). *A study on the perception of English education due to the change of English CSAT to the criterion-referenced system*. Unpublished master's thesis, Korea National University of Education, Chung-Buk.

*Jung, H. (1999). Development of criterion-referenced English oral proficiency interview. *Honam University Research Journal, 20*, 153–165.

Kang, G. H. (2014, May). *Policy considerations for introducing absolute testing of English section of CSAT*. Paper presented at the 26th KICE Curriculum and Evaluation Policy Forum (pp. 17–25). Seoul: Korea Institute for Curriculum and Evaluation.

*Kang, H. E. (2015). *A study of learners' perception on absolute grading system in the English section of CSAT*. Unpublished master's dissertation, Sookmyung Women's University, Seoul.

*Kang, S. (2017). The criterion design and application in the criterion–referenced test of English by the effect of the teaching experience. *Studies in Linguistics 42*, 311–335.

*Kim, D. K., & Ahn, B. K., & Oh, Y. J., & Kim, Y. K. (1999). A proposal for performance assessment criteria with focus on speaking and writing abilities. *English Teaching, 54*(2), 175–200.

*Kim, E. H. (2013). *Trends in test results of norm-referenced and absolute testing of English as a middle school*. Unpublished master's thesis, Korea University, Seoul.

*Kim, J.-L., Kim, J.-G., Mun, T.-H., & Jeon H.-C. (2012). *Assessment guidelines for primary English education*. Seoul: Hangukmunhwasa.

*Kim, J.-S. (2009). *Curriculum and assessment for English subject*. Seoul: Hangukmunhwasa.

*Kim, N. K. (2018). *Absolute testing of CSAT English and some suggestions for secondary English education*. Unpublished master's dissertation, Gangneung–Wonju National University, Wonju.

*Kim, S.-A. (2014). *Understanding English assessment*. Seoul: Globalcontents.

*Kim, Y.-M. & Kang, M-K. (2015). *An issue on introduction of the criterion-referenced evaluation (CRE) into the English domain of CSAT*. Advanced Science and Technology Letters, 115 (Education, 2015), pp. 107–111.

*Kim, Y. M. (2015). The retrospective and the perspective: The 20 years of college scholastic ability test (CSAT). *Studies in British and American Language and Literature, 119*, 275–299.

Korea Institute for Curriculum and Evaluation. (2018). *Scoring result for June trial examination of 2019 College Scholastic Ability Test*. Jincheon: Korea Institute for Curriculum and Evaluation.

*Korea Institute for Curriculum and Evaluation. (2014). *Introduction of absolute testing in English section in CSAT* (CAT 2014-36). Seoul: Korea Institute for Curriculum and Evaluation.

Kwon, O. (2015). A history of policies regarding the English section of Korea's college scholastic ability test. *English Teaching, 70*(5), 3–34.

*Kwon, S. B. (2018). *A text-mining study on public opinion toward criterion-referenced college scholastic ability test*. Unpublished master's dissertation, Korea National University of Education, Chung-Buk.

*Lee, H. W. (2015). *A study of secondary teachers' perception on the criterion-referenced evaluation of CSAT*. Unpublished master's dissertation, Hanyang University, Seoul.

*Lee, J. Y. (2018). *A study about how the application of criterion-referenced evaluation in CSAT affects English studies*. Unpublished master's thesis, Yonsei University. Seoul.

*Lee, J.-H., Kim, M.-K., Shim, K.-N., Shim, C.-Y., Lee, S., Kim, J.-S., Jang, E.-S., Jin, K.-A., Park, T.-J., & Jeong, T.-Y. (2012). *Elementary English education and evaluation*. Seoul: Hangukmunhwasa.

*Lee, S. Y. (1999). A study on the development of national common absolute testing standards for English curriculum. *English Linguistics Science, 1*, 59–62.

*Lee, W.-K. (2012). *English assessment methodology* (3rd ed.). Seoul: MunJinMedia.

Linn, R. L. & Gronlund, N. E. (1995). *Measurement and assessment in teaching*, Seventh edition. New York: Macmillan.

*Ministry of Education. (2014). *Introduction of absolute assessment in the English section of CSAT*. Seoul: MOE.

Ministry of Education. (2017a). *Announcement of grading results and grade notification for the 2018 CSAT.* Seoul: MOE.

*Ministry of Education. (2017b). *Basic plan for 2018 CSAT*. Seoul: MOE.

*Shin, D., & Cho, D. (2010). On the issue of test specifications in the development of a criterion-referenced achievement test. *English Teaching, 55*(2), 309–330.

*Shin, R. B. (2016). *Issues of introduction of absolute testing to English section of CSAT*. Unpublished master's thesis. Korea University, Seoul.

*Shin, S.-K. (2010). *Theory and practice of foreign language assessment*. Seoul: Hangukmunhwasa.

Shin, S.-K. (2017). Korea's English curriculum: Review, issues and future directions. *Journal of Research in Curriculum & Instruction, 21*(3), 206–215.

*Song, S. E. (2017). *A study of high school students' perception on criterion-referenced evaluation of CSAT English*. Unpublished master's thesis, Busan National University of Education, Busan.

Sung, T.-J. (2014). *Foundations of educational evaluation*. Seoul: Hakjisa.

*Yi, J. H. (2016). *A study of students' perceptions on criterion-referenced evaluation of CSAT in 2018*. Unpublished master's thesis, Incheon National University, Incheon.

* indicates studies analyzed in this research.

Applicable levels: Secondary

Sang-Keun Shin
Professor
Department of English Education
Ewha Womans University
52 Ewhayeodae-gil, Seodaemun-gu
Seoul 03760, Korea
Phone: 02-3277-4648
Email: sangshin@ewha.ac.kr