Proposed Skill Assessment Models for College Admissions to the Golf Departments in Korea: An Application of the Rasch Partial Credit Model

Lee Jun-Woo¹, Jeong Tchae-Won² and Yang Chun-Ho³*

¹Department Golf, Hoseo University, 20 Hoseo-ro 79, Baeon-gil Baebang-eup, Asan-si, Chungcheongnam-do, 31499, Korea; ljw75@hoseo.edu
²Industry Academic Research, Konkuk University, 268, Chungwon-daero, Chungju-si, Chungcheongbuk-do, 27478, Korea; sportms1@hanmail.net
³Department Marine Sports, Hanseo University, 46 Hanseo 1-ro, Haemi-myeon, Seosan-si, Chungcheongnam-do, 31962, Korea; healthyang@hanseo.ac.kr

Abstract

Objectives: Currently, for the assessment of skill in drives, iron shots, and approach shots for admission to golf departments in Korean Colleges, an applicant’s golf ability is evaluated based on the sum of successful shots made regardless of the difficulty level. Methods/Statistical Analysis: To address this problem, the lowest weight and the highest weight were determined based on face validity, and then weights were determined in proportion to the difficulty levels estimated from applying the Rasch partial credit model statistically. Findings: Results showed that, the lowest weight (1.00) was assigned to drives, the highest weight (1.50) to approach shots, and the interpolated weight (1.42) to iron shots. For female students, the lowest weight (1.00) was also assigned to drives, the highest weight (1.50) to approach shots, and the interpolated weight (1.47) to iron shots. Application/Improvement: Accordingly, for male students, the difficulty level of approach shots was the highest, followed by those of iron shots and drives. For female students, the difficulty level of approach shots was also highest, followed by those of iron shots and drives.

Keywords: Golf Evaluation Model, Rasch Partial Credit Model, University Admissions, Weight

1. Introduction

South Korea has been reported to have the highest level of education among OECD member countries. This demonstrates that Korea has a globally recognized fervor for education. In fact, Korea’s education fervor has increased over the last few decades, as the country’s college enrollment rate surged from 33.2% in 1990 to 79.0% in 2010. College entrance competition grew accordingly. This trend also applied to high school athletes.

After a smooth progression to high school, Korea’s student-athletes face intense competition for college admission and joining professional teams. In particular, college admission in physical education is highly competitive, especially within each sport type; this is natural given the limited availability of programs, small student pool, and stringent qualification requirements. Accordingly, university authorities strive to select student-athletes in a fair process using skills assessment by sport type, aiming primarily at selecting students that demonstrate abilities adequate for college admissions. Thus, assessment indicators are used as important indicators in student admission.

As such, skills assessment indicators are bound to be also used as important criteria for admissions in golf major programs, which is one of the physical education majors. Therefore, although the objective of skills assessment for admissions to golf departments is to select new
students with the appropriate sport skills for college level education and the ability to handle the curriculum, assessment indicators must have standard values established based on scientific evidence. In other words, skills assessment for golf major applicants must use indicators that are developed based on judges' expert knowledge and experiences, and then validated with data, to enable fairness in selection based on the reliability of the results.

However, the standards used currently are established based on the experience, including the authority and expertise, of the faculty in golf departments. This leaves room for disputes over the fairness of new student selection criteria. For example, skills assessment in golf departments evaluates the accuracy and ball flight indirectly based on ability in long game, middle game, and short game. Specifically, skill performance is scored by examinees choosing their own clubs and making 30 m approach shots, 140 m (120 m for females) iron shots, and >220 m (>200 m for female) drives; the number of successful shots are graded and converted into scores by domain, which are then summed up. Then, the total scores of all examinees are sorted in order, and based on the number of students to be admitted, the examinees with the highest total scores are selected.

Korea's college entrance exam is a large-scale exam conducted by the state government, and follows the procedures of a standardized test. In other words, all forms and procedures are strictly controlled to ensure the uniformity of administration, scoring, and interpretation of the test for all examinees. However, because the skills assessment factors related to golf department admissions are skills that do not require the same abilities, and, accordingly, the score proportion among factors (weight) cannot be considered the same, problems may occur in scoring.

In addition, as the score for each grade in each factor in skills assessment is created based on experience and therefore, the standard is unclear, scoring and interpretations are difficult. Moreover, despite clear gender differences in physical ability for which males and females hold separate competitions, the entrance exam for golf departments does not provide gender-specific estimates, leaving room for ambiguous interpretation. Therefore, the current skills assessment for student admissions to golf departments is unfit as a tool for appropriately estimating the performance level of high school golf athletes.

The colleges of physical education in Korea that conduct skills assessment endeavor to develop assessment items and provide valid procedures. Although the development of the assessment items for selecting the types of students desired by colleges is ongoing, there have been limited efforts to raise questions from a methodological standpoint regarding the interpretation of results, and few measures have been proposed to improve on the issue. Therefore, physical education scholars need to engage in discussions on the validity of interpretation based on valid and reliable skills assessment. In particular, considering the relatively short period of skills assessment in golf departments, multifaceted rethinking of its validity is needed.

As shown above, College golf departments endeavor to produce golf athletes with international level ability. To achieve this objective, it is crucial to resolve the problems in the existing method for selecting new students. Accordingly, this study proposes a skills assessment model differentiated for male and female applicants and validated based on the current skills assessment method for admissions to golf departments in Korea.

In particular, the ultimate goal of this study is to improve shortcomings in the results interpretation of the skills assessment for college admissions to golf departments by applying the Rasch partial credit model, which allows the determination of an applicant's ability using differential weights in the assessment of sports performance.

2. Research Methods

2.1 Ethics Statements

The data for this study were obtained from HRCGCSI (Hoseo Research Center of Golf Cultural Strategy Institution). Subjects' personal information was anonymized and de-identified prior to analysis for the study. New information was not collected from subjects. The protocol for this study was approved by the Bioethics Institutional Review Board at Konkuk University (Project No.7001355-201509-E-039).

2.2 Participants of Study

This study was conducted using the records for 282 students (139 males; 143 females) from skills assessment performed as part of the college entrance exams to golf departments in H Universities, Korea. The subjects were the student golfer applicants for 2014 admissions to golf
departments with at least 3 years of experience (7.15±3.40 years for males; 6.97±3.18 years for females) competing in middle and high school competitions as amateur athletes. The data were obtained from each student performing five approach shots, iron shots, and drives, and adding up the number of successful shots, which yielded a maximum score of 15 points and a minimum score of 0 points.

2.3 Measuring Instruments

The three domains (approach, iron, and drive) of skills assessment for admissions to golf departments measure shot accuracy and ball flight, which are the most important elements in golf. In particular, approach shots are strokes intended to place a ball close to the hole near the putting green, and therefore measure the skills in the short game. Approach shots were performed around 30 m from the hole. Iron shots are strokes intended to place a ball close to the hole on the fairway, and therefore measure the skills in the middle game, and were performed at 140 m (120 m for female) from the hole. These two domains measure accuracy, and shots are considered successful if the ball lands within a 5 m radius of the hole. On the other hand, the long game domain is ball flight, and because it requires striking a ball >220 m (>200 m for female) using a driver, it was performed with reference lines connected at both ends.

Assessment was conducted under the supervision of three or more supervisors and two or more deputy supervisors because the data were also collected for the purpose of selecting new students. The test length, including practice swings, was within 15 minutes in total; examinees took the test individually, and were assigned to test in the morning or the afternoon based on their test IDs. During the test, each examinee chose a club and performed shots in the order of approach shots, iron shots, and drives while supervisors recorded the results.

2.4 Data Analyses

This study estimates the difficulty levels of approach shots, iron shots, and drives for each gender using the Rasch partial credit model, and proposes a weighted assessment model. In other words, because the measured approach shots, iron shots, and drives were considered 5-point scale items in the analysis, it was the application of the partial credit model that extended the Rasch model.

Here, \( p_{i,j,k} \) denotes the probability that the student, \( s \), obtains the score, \( x \), on the variable, \( i \); \( \exp \) denotes the exponential of 2.718 (\( \exp = 2.718^8 \)), \( \beta_s \) denotes the student's ability, and \( \delta_{ij} \) denotes the difficulty level, \( j \), (category) of the variable 1 (approach shot, iron shot, and drive) \( k \) denotes the beginning point for summation, and \( m_i \) denotes the maximum score, \( x \) points. Therefore, in this study, because the minimum score for each domain is 0, and the maximum score for each domain is 5, \( k=0 \) and \( m_i=5 \). For example, the probability that student A obtains 3 points for an approach shot is determined by the student's approach skills and the difficulty level for achieving 3 points in the approach domain (level). Of course, in this study, student A's skills and the difficulty level of the approach are estimated based on the maximum likelihood estimation for Rasch.

In addition, in the Rasch partial credit model, the difference in difficulty level between a student's skills and the domains (\( \beta-\delta \)) is estimated in a logit. The logit has the advantage of converting the ordinal scale of a measured score (0–5 points) to the interval scale. In this study, a larger positive logit value indicates higher skills in comparison to the difficulty level, and a larger absolute value of negative logit indicates a higher difficulty level in comparison to the student's skills. In other words, a greater positive value for \( \beta-\delta \) (logit) indicates a domain with a relatively high difficulty level.

Based on the estimated domain-specific difficulty level (logit), weights were calculated. For the standard setting for maximum weight, the Angoff method was applied. Based on a discussion with two golf experts and a sport measurement and evaluation expert, a weight of 1 point (multiplied) was assigned to the domain with the lowest estimated difficulty level, and a weight of 1.5 points (multiplied) was assigned to the domain with the highest estimated difficulty level. Based on the determined maximum and minimum weight standards, the value in the middle was calculated through interpolation. Among the three domains, the measurement of the domain with the lowest difficulty level estimate was multiplied by 1.50 as its weight, and the measurement of the domain with the highest difficulty level estimate was multiplied by 1.00 as its weight. Then, the weight of the remaining domain was calculated to be proportional to the estimated difficulty level (logit).

For calculation of the difficulty level (logit) for the three domains applied to the model, Winsteps 3.65.0 software was used. The model in this study was designed
such that a higher logit indicates a higher difficulty level, and model fit was tested using the infit index and outfit index. These indices indicate the degree of consistency for variable-specific difficulty levels, and follow the \( \chi^2 \) distribution with an expected value of 1.00. A fit index value closer to the expected value indicates that the data fit the model. A fit index smaller than 0.50 indicates an overfit variable and one larger than 1.50 indicates a misfit variable\(^{10} \). In other words, a variable with a fit index smaller than 0.50 in this study means that the judges (faculty) score highly skilled students high, and low-skilled students low (i.e., overfit variable), whereas a variable with a fit index larger than 1.50 means that the judges score highly skilled students low and low-skilled students high (i.e., misfit variable). In this study, a fit index above 1.50 for a domain (approach shots, iron shots, and driver shots) indicates a problem in the judges’ assessment.

The overall fit of the data of this study between the Rasch model was tested within a 95% confidence interval (95% CI) based on the item characteristic curve (ICC). The testing of undimensionality, the basic assumption of the Rasch model, was conducted by examining the Point-Measure Correlation (PMC) proposed by the Rasch model. This coefficient is an estimation correlation between the 5-category index for each variable and the total category index (sum of indices). The variable with PMC over 0.30 is considered to suggest unidimensionality\(^{17,18} \).

### Table 1. Difficulty Level by Skills Assessment Domain for Male Students and Model Fit

<table>
<thead>
<tr>
<th>Difficulty level (Logit)</th>
<th>SE</th>
<th>Infit</th>
<th>Outfit</th>
<th>PMC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drives</td>
<td>-2.08</td>
<td>.17</td>
<td>1.16</td>
<td>1.56</td>
</tr>
<tr>
<td>Iron shots</td>
<td>.80</td>
<td>.15</td>
<td>.76</td>
<td>.75</td>
</tr>
<tr>
<td>Approach shots</td>
<td>1.28</td>
<td>.15</td>
<td>1.05</td>
<td>1.03</td>
</tr>
</tbody>
</table>

\( \chi^2 = 246.10, \ p<.001, \ SI=9.06, \ RI=.99, \ MDS=12.41 \)

The model fit for the data was tested with the separation index (SI) to verify whether the students’ skills assessment scores effectively discriminate the levels during the college entrance exam, and the reliability index (RI) was used to verify whether the college entrance exam level effectively discriminates students’ skill levels. The minimum satisfaction criteria were set at 2.00 or higher for SI, and 0.80 or higher for RI\(^{10} \). In addition, to test whether the three variables in this study clearly discriminate students’ performance levels, the distance between variables i.e., the number of distinct strata (NDS, number of scales) - were estimated using SI. Specifically, the mathematical model of NDS estimation is \( \text{NDS} = \left( \frac{4 \times \text{SI}}{\pi^2} + 1 \right) \). The number of estimated NDS may be used as the index to determine the effective number of categories for students in the college entrance exam\(^{16,20} \).

### 3. Research Results

#### 3.1 Difficulty Level by Skills Assessment Domain for Male Students and Model Fit

The Table 1 shows the analysis results for difficulty level according to the skills assessment domain for male students as well as the model fit. The difficulty level (logit) for approach shots was highest (logit=1.28), followed by iron shots (logit=.80) and drives (logit=-2.08). Regarding infit and outfit indices, drives (infit=1.16, outfit=1.56) had an outfit index that did not meet the standard. On the other hand, both the infit and outfit indices of iron shots (infit=.76, outfit=.75) and approach shots (infit=1.05, outfit=1.03) met the standards. All PMCs were found to be 0.30 or higher.

#### Table 2. Skills Assessment Domain-specific Weights for Male Students

<table>
<thead>
<tr>
<th>Difficulty level (Logit)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drives</td>
<td>-2.08</td>
</tr>
<tr>
<td>Iron shots</td>
<td>.80</td>
</tr>
<tr>
<td>Approach shots</td>
<td>1.28</td>
</tr>
</tbody>
</table>

In addition, domain-specific difficulty levels were different at a statistically significant level (\( \chi^2 = 246.10, \ p<.001 \)), and SI and RI were 9.06 and 0.99, respectively, which met the standards. The NDS was 12.41, which is very different from the 5 points applied in actual assessment.

#### 3.2 Skills Assessment Domain-specific Weights for Male Students

The Angoff method was applied using the three domain-specific difficulty levels (logit) estimated based on the Rasch model. The results for the derivation of weight are shown in Table 2. In other words, after assigning the...
weight of 1.50 to the approach shots with the highest difficulty level estimate (1.28), and the weight of 1.00 to the drives with the lowest difficulty level estimate (-2.08), the weight of iron shots with the moderate difficulty level estimate (0.80) was calculated.

The weight of iron shots (1.42) was estimated by obtaining the difference in logit between the highest and the lowest difficulty level estimates (1.28-(-2.08)=3.36), and the difference between the highest and the lowest weights (1.50-1.00=0.50) determined by the Angoff method. This shows that the difficulty level of iron shots (0.80) has a difference of 2.88 (0.80-(-2.08)=2.88) from the lowest difficulty level (-2.08). This difference of 2.88 in difficulty level was interpolated to the corresponding difference (x) in weight (3.36:0.50=2.88:x, x=0.42). When the calculated value (0.42) was added to the lowest weight 1.00, the weight of iron shots, 1.42, was obtained. This means that when an iron shot is successful, 1.42 points rather than 1 point is assigned.

### 3.3 Difficulty Level by Skills Assessment Domain for Female Students and Model Fit

Table 4 shows the analysis results for difficulty level according to the skills assessment domain for female students as well as the model fit. The difficulty level (logit) for approach shots was highest (logit=.82), followed by iron shots (logit=.72) and drives (logit=-1.54). The infit and outfit indices of approach shots (infit=.95, outfit=.95), iron shots (infit=.87, outfit=.87), and drives (infit=1.17, outfit=1.24) all met the standards.

<table>
<thead>
<tr>
<th>Difficulty level (Logit)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drives</td>
<td>-1.54</td>
</tr>
<tr>
<td>Iron shots</td>
<td>.72</td>
</tr>
<tr>
<td>Approach shots</td>
<td>.82</td>
</tr>
</tbody>
</table>

All PMCs were found to be 0.30 or higher. In addition, domain-specific difficulty levels were different at a statistically significant level ($X^2=167.00, p<.001$), and SI and RI were 7.36 and .98, respectively, which met the standards. The NDS was 10.14, which is markedly higher than the 5 points applied in actual assessment.

### 3.4 Skills Assessment Domain-Specific Weights for Female Students

Based on discussions with golf experts, weights of 1.00 and 1.50 were assigned to the domains with the lowest and the highest difficulty levels, respectively. Using these numbers, the weight for the remaining domain (with the moderate difficulty level) was interpolated. The estimated domain-specific weights for female students are shown in Table 5.

<table>
<thead>
<tr>
<th>Difficulty level (Logit)</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drives</td>
<td>-1.54</td>
</tr>
<tr>
<td>Iron shots</td>
<td>.72</td>
</tr>
<tr>
<td>Approach shots</td>
<td>.82</td>
</tr>
</tbody>
</table>

After assigning the weight of 1.50 to the approach shots with the highest difficulty level estimate (0.82), and the weight of 1.00 to the drives with the lowest difficulty level estimate (-1.54), the weight of iron shots with the moderate difficulty level estimate (0.72) was calculated.

The weight of iron shots (1.47) was estimated by obtaining the difference in logit between the highest and the lowest difficulty level estimates (0.82-(-1.54)=2.36), and the difference between the highest and the lowest weights (1.50-1.00=0.50) determined by the Angoff method. This shows that the difficulty level of iron shots (0.72) has a difference of 2.26 (0.72-(-1.54)=2.26) from the lowest difficulty level (-1.54). This difference of 2.26 in difficulty level was interpolated to the corresponding difference (x) in weight (2.36:0.50=2.26:x, x=0.47). When the calculated value (0.47) was added to the lowest weight 1.00, the weight of iron shots, 1.47, was obtained. This means that when an iron shot is successful, 1.47 points
rather than 1 point is assigned. The skills assessment estimation model for female students’(y’) college admissions to golf departments considering weights is shown in Table 6.

**Table 6. The Skills Assessment Estimation Model for Female Students’ College Admission to Golf Departments**

\[ y' = \text{No. of successful drives} + (\text{No. of successful iron shots} \times 1.47) + (\text{No. of successful approach shots} \times 1.50) \]

**4. Discussions**

In judging students’ golf skills using the number of successful drives, iron, and approach shots for college admissions to golf departments, anyone interested in assessment would recognize a problem in the fact that the skills in different domains are considered the same and that gender-specific estimates are absent. Accordingly, the current study presents new assessment models based on a scientific approach for valid interpretation of skills assessment results for college admissions to golf departments.

Regarding the results for the domain-specific difficulty level for male students in the present study, the outfit index for drive shots exceeded the standard. In terms of fit index standards (0.50–1.50), exceeding 1.50 indicates a misfit variable. The misfit means that the skills assessment for male students is not systematic. This means that highly skilled male students are scored low, and low-skilled male students are scored high. Therefore, the difficulty level for drives, which indicates a misfit index, lacks validity. However, because the outfit index becomes larger as residuals for the domain become larger, a large outfit index (over 1.50) suggests a lack of reliability. When a student with extremely high skills fails in a domain with an extremely low difficulty level, residuals increase; accordingly, a few extreme responses lead to a large outfit index. The infit index complements this problem. In the case of male students’ drive shots in this study, the outfit index exceeded the standard. However, because the infit index (1.16) met the standard, the weight was estimated using the estimated difficulty level for drives.

In addition, for male and female students, drives showed the lowest difficulty level (male=-2.08, female=-1.54), and approach shots showed the highest difficulty level (male=1.28, female=-0.82). This is consistent with the results obtained by[18] from testing the all-round ranking model for PGA players, showing that the difficulty level of the putting green area (bunker shots, putting) is higher than that of drives (ball flight, landing rate), and the interpretation that the short-game skills that athletes perceive as most difficult and spend the most time practicing determines performance.

Accordingly, in college golf department entrance exams, the 30 m approach shots were found to have more influence on student admission than outcomes in other domains, and the effect is larger for male students than female students. However, for female students, 140 m iron shots (.72) and 30 m approach shots (.82) showed relatively little difference, which suggests that the skills in approach and iron shots can be important factors in selecting new students.

In addition, to determine the model fit of the data used in this study, SI, RI, and NDS were calculated. Although both SI and RI met the standard, NDS was found to be higher than the 5-point scale of the assessment. This indicates that, in college golf department entrance exam skills assessment, the number of measurement categories (excellent, good, average, fair, and poor) for drives, iron shots, and approach shots needs to be set high to ensure the discriminatory power of the assessment of students’ performance level. This suggests that the current number of five shots per domain should be increased to 7–10 shots. On the other hand, NDS was found to fully meet the assumption of unidimensionality, which indicates that one trial can measure a student’s ability because NDS is proportional to SI. However, an increase in the number of trials in each domain of skills assessment in college golf department entrance exams is proposed to comply with the principle of testing.

The Rasch model is a statistical technique appropriate for deriving weights for comparison on common measures using logit scores, and valid assessment is possible even when the difficulty levels of domains vary. However, the model has substantial limitations in weight estimation—specifically, determining weights specific to different domains. In this study, the approach proposed by to determine the weight as the mean of the standard values that experts decide on after discussion was determined, and the calculation models shown in Table 3(males) and Table 6(females) were proposed.

One crucial finding was that there was difference in ranks between the old and the new models. For example, it is notable that out of 20 students in total, all students except seven students showed changes in rank. Change
in rank indicates a violation of the assessment principles and a potential problem of fairness in assessment. The changes are presumably due to the difference in weights based on domain-specific difficulty level and gender.

In particular, while the old model yielded a number of identical scores, the new model yielded no identical scores, which suggests that the new model can provide clearer standards in selecting student using the assessment with a relative standard. Accordingly, it is suggested that the application of the new model, which addresses the problems of the old model, provides valid and objective information, and avoids unfair selection of students in skills assessments for college golf department entrance exams.

5. Conclusions and Suggestion

Fair selection of athletes and fair assessment of performance persists as a topic that draws attention. In this context, this study aimed to propose models that are fairer in selecting new students by addressing the shortcomings of the skills assessment currently conducted for admissions to college golf departments in Korea.

That is, this study improved upon the problems of conventional assessment of student skills based on the sum of the number of successful shots in different domains without considering domain-specific difficulty levels. Specifically, the lowest weight (1.00) and the highest weight (1.50) were determined based on face validity, and then weights were determined in proportion to the difficulty levels estimated from applying the Rasch partial credit model statistically.

Analysis results showed that for male students, the difficult levels were in the order of approach shots (logit=1.28), iron shots (logit=.80), and drives (logit=-2.08). For female students, the difficulty levels were also in the order of approach shots (logit=1.28), iron shots (logit=.80), and drives (logit=-2.08). Accordingly, for male students, the lowest weight (1.00) was assigned to drives, the highest weight (1.50) to approach shots, and the interpolated weight (1.42) to iron shots. Likewise, for female students, the lowest weight (1.00) was assigned to drives, the highest weight (1.50) to approach shots, and the interpolated weight (1.47) to iron shots.

In conclusion, it is proposed that in the skills assessment for college golf department admission, the model in which cross products of the number of domain-specific successful shots and weight are summed up improves the quality of assessment compared to the conventional method.

6. References

Proposed Skill Assessment Models for College Admissions to the Golf Departments in Korea: An Application of the Rasch Partial Credit Model