A CONSTRUCTION OF CRITERION-REFERENCED FOR BADMINTON TEST BATTERY

Han-Dau Yau,
Dept. of Sports Training Science, National College of Physical Education and Sports,
Taoyuan, Taiwan, ROC.

Minia University 1999 ICHPER-SD 42nd World Congress.
International Council for Health, Physical Education, Recreation,
Sport, and Dance
A construction of criterion-referenced for badminton test battery

Han-Dau Yau, Dept. of Sports Training Science, National College of Physical Education and Sports, Taoyuan, Taiwan, ROC.

ABSTRACT

The purpose of the study was to construct a criterion-referenced of badminton test battery. The subjects were the 223 players of first and second level athletes in the Badminton Association of Republic of China. The approach of logistic regression model was used as selecting the best test battery of criterion-referenced test in badminton. The logistic regression equation is logit (p) = \(-22.22 + 0.05 \times \text{push shot test} + 0.03 \times \text{drop test} + 0.06 \times \text{standing long jump test} + 0.12 \times \text{muscle strength test}\). The best of cutoff score (criterion) is 0.52. Empirical approaches to create validations are the phi(\(\phi\)) coefficient 0.53 and cross validation 0.80 in criterion-referenced measurement, as well as to create it's reliabilities of the agreement proportion 0.83, Cohen kappa 0.53 and modified kappa 0.65 in criterion-referenced measurement. The results make it possible to take into "decision strategy", such as the selection of national badminton team, the college admission of badminton placement, and the training of national badminton team. The conclusion was that the test battery had good characteristics to construct a criterion-referenced of badminton test battery.

Keywords : criterion-referenced measurement, test battery, badminton, logistic regression, cross validation, test construction.

This study (NSC 89-2914-I-179-001-A1) was sponsored by The National Security Council, R.O.C. (Taiwan).
Introduction

Over the past three decades there had been a little number of immensely important advances and changes in the field of psychometrics. The criterion-referenced (CR) measurement is one of the domains in the important advances. In 1963, Robert Glaser presented one of the first references to CR testing in a three-page paper in American Psychologist. The CR testing has developed since the 1970s. Hambleton and Rogers (1991) reported: “One of the major changes in the testing over the last 20 years has been the increased interest in and use of CR tests.” Because the purpose of CR testing was making a mastery or non-mastery decision for examinee’s ability on the objective in domain with the corresponding standard. The most important function of CR testing was discriminating the level of examinee’s ability or that had a motor skill. It involved the diagnostic tests, the basic skill tests, the competency tests, the master learning and the competency-based education. Overview its content, the application of CR testing has been selected athletes, screening of college admission and classification of decision accuracy in physical education and sport.

Based on the research analysis of CR testing by Yau (1995), he suggested: “the topic have been presented on the means of twenty papers from 1982 to 1987, and the means of eleven papers from 1988 to 1992 on Quest Dissertation Abstracts on disc in every years. In SPORT Discus (from 1975 to March 1994), only eight papers were related research. It included reliability, validity and cutoff score in testing. The test construction of CR testing was three papers only.” In fact, the application of testing is little on the domain of CR measurement. A CR testing is not found, and the screening test of latent trait for motor skill is very important in sports now. It is not norm-referenced tests. Norm-referenced performance is interpreted best in relation to a “norm” population because performance on a latent trait of athlete is rarely of interest and because of the fact that CR testing can be matched to object of domain.

In the physical fitness, mental and skill of relationship of badminton study, Whetnall and Morris (1981) presented: “badminton is a game of skill, speed, power and
Top international players clearly demonstrate the essential stroke techniques and tactics. Advice is given on equipment, match play and fitness training.”

Abernethy (1988) tested the effects of age and expertise upon perceptual skill development in a racquet sport. The results showed that these data indicated the presence of systematic differences between the perceptual skills of experts and novices, which transcend development age. Chen-Qinghong (1990) studied kinematics analysis on the techniques of heavy smash in the badminton competitions. He demonstrated the velocity of wielding bat depending on the correct arm movements, and the accurate and tricky falling points of the heavy smash depends on the height and angle of batting.

Chen-Fushou (1990) discussion on the ‘sudden killing’ and ‘sliding plate killing’ in the women's singles —discussing the developing tendency of badminton simultaneously. Controlling opponents will be easier. Abernethy (1991) processed visual search strategies and decision-making in sport. Implications of the research series for models of information processing and decision-making, for models of expertise and for the theory and practice of perceptual skill development are briefly considered. Chi etc., (1994) studied the relationship between the double swing of 15 minutes rope skipping and VO₂ peak in badminton players. Their results showed that double swing of 15 minutes rope skipping had significantly correlated to VO₂ peak (r = 0.7867) in elite badminton players. Dong and Kang (1994) compared to study of mental toughness between elite and non-elite female athletes. In the mental toughness of elite female athletes, visualization and imagery control, motivation and attitude control were significantly higher than non-elite female athletes.

Yuan etc., (1995) compared with nine elite male badminton players, eight elite male gymnasts and 20 male students were investigated for hand-eye co-ordination ability and reaction time. The results showed that the badminton players in general had a large number of correct responses than students. But elite male badminton players and elite male gymnast had no significant difference. Chin etc., (1995) investigated sport specific fitness testing of elite badminton players. The subjects are twelve Hong Kong national badminton team players. A low correlation (r = 0.65) was found between the results of the field test and the rank-order list of subjects, based on an objective on-field
physiological assessment and subjective ranking. This may be explained by the requirements of other factors besides physical fitness, which contribute to success in elite level badminton competition. These factors may include, for example, technical skill, mental power, and aesthetic judgements on the court. Louie and Lam (1996) investigated the reliability and validity of badminton wall volley tests. The intraclass reliability coefficients for the wall volley test in the range of .41 to .84. The validity was computed utilizing the interclass correlation coefficients method. It appears that only correlation was found between the wall volley tests and the criterion measures (overhead return test and serve test). It is concluded that the use of wall volley technique to evaluate players' badminton playing abilities is not an adequate method.

Chi (1996) studied a specific badminton physical fitness testing on badminton singles players. The results showed that elite badminton male players must be muscle strength, muscle endurance and agility, and elite badminton female players must be footwork, cardiorespiratory function, power and agility. Huang etc., (1997) studied on the design and administration of badminton information system. They analyzed 1996 badminton Chinese Taipei open. Rasch had the most stroke of clear 41(19.5%), next lob 40(19.0%) and third smash 29(14.3%) in Sun versus Rasch game. The lob, clear, close net shot, cutting, smash and serve were 85% in overall. Chi (1998) studied a specific badminton skill test battery on badminton singles players. The results showed that clear test I, net fight test II, cutting test I, driven fight I and II, and long serve test are significantly related to the ranking of badminton performance.

Inductive previous studies, elite badminton players must have an overall ability of specific fitness, specific skill and mental toughness. This study decided for the tests construction of primary by pilot study in badminton related tests. Physical fitness tests include sidestep test, vertical jump test, sit-and-reach test, muscle strength test, muscle endurance test, footwork I test, footwork II test, footwork III test, cardiorespiratory function test and standing long jump test. Badminton skill tests include clear test, smash test, lob shots test, close net shot test, drop test, drive test and push shot test. Therefore, the purpose of the study was to construct a criterion-referenced of badminton test battery.
Methods

The measurement theory of psychometric is based on the study. It selects Yau's method in "The new development of criterion-referenced for physical education and sport test battery." (Yau, 1998) The application of logistic regression analysis model constructs a criterion-referenced of badminton test battery by the data of testing. Logistic regression analysis models the relationship between a binary or ordinal response variable and one or more explanatory variables. The logistic regression model uses the explanatory variables to predict the probability that the response variable takes on a given value. The key point of response variable can take on binary or ordinal scale.

The first of all, the structure is clearly defined to construct a criterion-referenced of badminton test battery. The response variables are the first and second level athletes in the Badminton Association of Republic of China. The explanatory variables are selected by the primary factor of performance in badminton. Twenty subjects in each skill testing will be approached by the item analysis. Computing expected reliability was used to the program of "Bigsteps" for the fixed-length tests. The formal tests were over 0.80 on expected reliability. Another, the physical fitness tests had moderate reliability and validity in the conventional tests. After item analysis, the seventeen tests were sidestep test, vertical jump test, sit-and-reach test, 1-minute sit-ups test (muscle strength test and muscle endurance test), footwork I test, footwork II test, footwork III test, 10 minutes rope skipping test, standing long jump test, clear test, smash test, lob shots test, close net shot test, drop test, drive test and push shot test.

The first of all, two error data were revised. The data of two rope skipping loss and eighteen random sampling were the samplers of cross validation (approach to ten percent at overall). The data were the 203 players of 54 first level athletes and 149 second level athletes. The indices of difference deviance (DIFDEV) and difference Pearson Chi-square (DIFCHISQ) produced regression diagnostics for logistic regression model. Based on the indices of DIFDEV and DIFCHISQ, extreme observations will be
deleted. The criterion of deletion were the ten times of DIFDEV \( (0.53 \times 10 = 5.3) \) and the five times of DIFCHISQ \( (2.1 \times 5 = 10.5) \). The subjects of number 17 (DIFDEV \( = 8.4 \), DIFCHISQ \( = 33.5 \)) and number 212 (DIFDEV \( = 5.9 \), DIFCHISQ \( = 13.4 \)) were ill-fitted observations. After deletion step, the data were only the 201 players of 53 first level athletes and 148 second level athletes.

The "Stepwise" and "Score" model selection methods of logistic regression analysis were utilized for the best test battery. This study creates a logistic regression equation of criterion-referenced of badminton test battery. After the equation, the best of cutoff score was made decision, and cross validation is computed by a new set of observations to test the predictive accuracy of this model. At final step is establishing the reliability of agreement proportion, Cohen kappa and modified kappa, and the validity of phi(\( \phi \)) coefficient in criterion-referenced measurement.

**Results and Discussion**

**Analysis of Logistic Regression**

1. The four explanatory variables are selected by stepwise option in the model. Use the score option find for the best subset of all models with four explanatory variables. (The best model have the highest score value) The four-variable model contains the push shot test, drop test, standing long jump test and muscle strength test. (Score = 58.21. Score is a score statistic, which has an asymptotic under) The testing global null hypothesis: Bata (\( \beta \)) = 0. The -2 Log Likelihood statistic (-2 LOG L) is 77.41 (\( p < .05 \)). A significant p-value provides evidence that at least one of the regression coefficients for an explanatory variable is nonzero. The Hosmer and Lemeshow test is a goodness-of-fit test for logistic regression model with binary responses. The value of Hosmer and Lemeshow goodness-of-fit test is 8.57 with eight degrees of freedom. (\( \chi^2_{HM} = 8.57 \)) When this statistic is compared to a chi-square distribution, the resulting
probability value is 0.38. Thus, it cannot reject the null hypothesis that the model provides a good fit to the data. The generalized coefficient of determination has a value of 0.32. The adjusted generalized coefficient of determination has a value of 0.47.

Table 1 Summary of the construction of CR for badminton test battery

<table>
<thead>
<tr>
<th>Variable</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Wald Chi-Square</th>
<th>Probability</th>
<th>Standardized Estimate</th>
<th>Odds Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>-22.22</td>
<td>3.71</td>
<td>35.87</td>
<td>0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Push shot</td>
<td>1</td>
<td>0.05</td>
<td>0.01</td>
<td>14.70</td>
<td>0.0001</td>
<td>0.49</td>
<td>1.05</td>
</tr>
<tr>
<td>drop</td>
<td>1</td>
<td>0.03</td>
<td>0.01</td>
<td>7.53</td>
<td>0.0061</td>
<td>0.32</td>
<td>1.03</td>
</tr>
<tr>
<td>Standing long jump</td>
<td>1</td>
<td>0.06</td>
<td>0.01</td>
<td>18.87</td>
<td>0.0001</td>
<td>0.97</td>
<td>1.06</td>
</tr>
<tr>
<td>Muscle strength</td>
<td>1</td>
<td>0.12</td>
<td>0.05</td>
<td>5.44</td>
<td>0.0196</td>
<td>0.26</td>
<td>1.13</td>
</tr>
</tbody>
</table>

2. Parameter estimate gives the estimated coefficients of the fitted logistic regression model. The results are the intercept -22.22, push shot test 0.05, drop test 0.03, standing long jump test 0.06 and muscle strength test 0.12. Here, the logistic regression equation is

$$\text{logit } p = -22.22 + 0.05 \times \text{push shot test} + 0.03 \times \text{drop test} + 0.06 \times \text{standing long jump test} + 0.12 \times \text{muscle strength test}$$

The values of standard error are the intercept 3.71, push shot test 0.01, drop test 0.01, standing long jump test 0.01 and muscle strength test 0.05. The values of Wald Chi-Square are the intercept 35.87, push shot test 14.70, drop test 7.53, standing long jump test 18.87 and muscle strength test 5.44. These $p$-value of Chi-square are significant at the 0.05 level. They show non-zero to any regression coefficients for the explanatory variable. The values of standardized estimate were: push shot test 0.49, drop test 0.32, standing long jump test 0.97 and muscle strength test 0.26. Odds Ratio is computed by exponentiating the parameter estimate for each explanatory variable.
The odds of the event increase by a factor of 1.05 for each unit increase in the value of push shot test. The odds ratio for drop test is 1.03. The odds ratio for standing long jump test is 1.06. The odds ratio for muscle strength test is 1.13. These results of statistic are good enough to conclude that the full model fits data significantly.

**The quality of logistic regression model**

1. Concordant = 87.4%, Discordant = 12.5% and Tied = 0.1%. The concordant of logistic model is 87.4% at this study. It shows good model.

2. The rank correlation of Somer’s D = 0.75 and Gamma = 0.75 are higher values. In a relative sense, it has better predictive ability. The value of c is 0.87. It is equal to eighty-seven percentages into the area under a receiver operation characteristic curve. In the overall, the criterion-referenced of badminton test battery is good.

**The quality of criterion-referenced test**

The best cutoff score decides for the criterion-referenced of badminton test battery by classification table. At 0.52 cut-point produces the largest percentage correct of 81.6 percent and the smallest "false positive rate" and "false negative rate". Therefore, the best cutoff score is 0.52. Correctly gives the probability and the model correctly classifies the sample data for this probability cut-point. The values of actuality and prediction are computed by the best of cut-point. The analysis of test battery has the following result.

1. The proportion of agreement:
   \[ \text{PA} = \frac{(135 + 31)}{201} = 0.826 \]

2. Cohen kappa coefficient:
   \[ \sum P_i P_i = 0.7811 \times 0.7363 + 0.2189 \times 0.2637 = 0.6329 \]
   \[ \kappa = \frac{(0.826 - 0.6329)}{(1 - 0.6329)} = 0.526 \]
3. Modified kappa coefficient:
\[ \kappa_q = \frac{(0.826 - 0.5)}{(1 - 0.5)} = 0.652 \]

4. The validity of criterion-referenced testing (phi):
\[ \varphi = \frac{(135 \times 31 - 22 \times 13)}{\left(148 \times 53 \times 44 \times 157\right)^{0.5}} = 0.530 \]

Table 2  The table of actuality by prediction in badminton test battery

<table>
<thead>
<tr>
<th>Actuality</th>
<th>Frequency</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>second level</td>
<td>first level</td>
</tr>
<tr>
<td>second level</td>
<td>135</td>
<td>13</td>
</tr>
<tr>
<td>first level</td>
<td>22</td>
<td>31</td>
</tr>
<tr>
<td>Total</td>
<td>157</td>
<td>44</td>
</tr>
</tbody>
</table>

The reliability is the proportion of agreement 0.83, Cohen kappa 0.53 and modified kappa 0.65 in criterion-referenced measurement. This result shows the good quality of test battery on study. The phi coefficient 0.53 is the validity of criterion-referenced testing with sample data. It is strength of association between actuality evaluation and prediction evaluation.

Cross validation

Calculation using new data produces the cross validation. The categories of actuality data are the 5 players of first level and the 15 players of second level athletes. Using the logistic regression equation computes predictions. The cross validation is \( (14 + 2)/20 = 0.80 \). The reduction of coefficient is 0.026 to compare with the cross validation (0.80) and the proportion of agreement (0.826). To estimate the accuracy and stability of prediction in predictive validity designs cross validation procedures should be employed. Cross validation is an external validation of testing.
Table 3  The table of actuality by prediction with new data in badminton test battery

<table>
<thead>
<tr>
<th>Actuality</th>
<th>Frequency</th>
<th>Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>second level</td>
</tr>
<tr>
<td>second level</td>
<td>1 4</td>
<td>1</td>
</tr>
<tr>
<td>first level</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Total</td>
<td>1 7</td>
<td>3</td>
</tr>
</tbody>
</table>

Conclusions

The purpose of the study was to construct a criterion-referenced of badminton test battery. Using a multiple logistic regression model, the subjects were the 223 players of first and second level athletes in the Badminton Association of Republic of China. A set of four basic data, ten specific badminton physical fitness tests and seven specific badminton skill tests were administered in 1999. It were name, height, weight, performance level, sidestep test, vertical jump test, sit-and-reach test, 1-minute sit-ups test (muscle strength test and muscle endurance test), footwork I test, footwork II test, footwork III test, 10 minutes rope skipping test, standing long jump test, clear test, smash test, lob shots test, close net shot test, drop test, drive test and push shot test. The first of all, two errors of data were revised. The data of two rope skipping loss and eighteen random sampling were the samplers of cross validation (approach to ten percent at overall). The data were the 54 players of first level and the 149 players of second level for estimation of regression equation. Next step deleted extreme observations by the indices of DIFDEV and DIFCHISQ. It found two subjects with ill-fitted observations. After deletion step, data were the 201 players of 53 first level athletes and 148 second level athletes. The results were analyzed as the following:

1. The criterion-referenced of badminton test battery of regression equation is:
   \[
   \text{logit}(p) = -22.22 + 0.05 \times \text{push shot test} + 0.03 \times \text{drop test} + 0.06 \times \text{standing long jump test} + 0.12 \times \text{muscle strength test}
   \]

2. The best cutoff score is 0.52.
3. Establishing reliability is the proportion of agreement 0.83, Cohen kappa 0.53 and modified kappa 0.65 in criterion-referenced measurement.

4. The test battery are validity of the phi (φ) coefficient 0.53 and cross validation 0.80 in criterion-referenced measurement.

Hence, the conclusion shows that this test battery has good characteristics to construct a criterion-referenced of badminton test battery.

Reference


