Paper : Soccer

Development of Criterion-referenced Measurement Items for Soccer Defensive Tactical Play from Tracking Data

Hiroki Matsuoka*, Yasuhiro Tahara*, Kozue Ando and Takahiko Nishijima****

*Doctoral Program in Physical Education, Health and Sport Sciences, University of Tsukuba
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574 Japan
s1730500@s.tsukuba.ac.jp
** Faculty of Health and Sport Sciences, University of Tsukuba
[Received May 23, 2019; Accepted October 31, 2019]

The purpose of this study was to construct criterion-referenced measurement items for soccer defensive tactical play from soccer game tracking data. Scaling procedure included qualitative analysis of the causal structure of soccer defensive tactical play using the Delphi method and a causal-effect analysis, processing items to measure soccer defensive tactical play from tracking data, an analysis of construct validity of soccer defensive tactical play by structural equation modelling, an analysis of success criteria of defensive plays by decision tree analysis, an analysis of test and item characteristics of the scale of soccer defensive tactical play by using item response theory (IRT), and an analysis of validity of the criterion-referenced measurement test. Qualitative analysis revealed that soccer defensive tactical play consisted of the press defense and block defense. Twenty-five soccer defensive tactical play items were created from the tracking data. Seventeen soccer defensive tactical play items showed construct validity. The success criterion of the soccer defensive tactical play items was revealed by decision tree analysis. Test and item characteristics, which consisted of fourteen items, were shown by IRT. In conclusion, it was statistically valid to construct the criterion-referenced measurement items for soccer defensive tactical play from soccer game tracking data.

Keywords: criterion-referenced evaluation, item response theory, decision tree analysis, structural equation modelling

[Football Science Vol.17, 29-40, 2020]

1. Introduction

Soccer game performance analysis has revealed the construct of soccer defence performance or causal relationship between attacking result and key measurement items (Suzuki and Nishijima, 2005; Tenga et al., 2009; Tenga et al., 2010a; Tenga et al., 2010b). Suzuki and Nishijima (2005) constructed measurement items for defence performance and showed the construct of defence performance by structural equation modelling (SEM). Tenga et al. (2009) constructed measurement items for attacking performance and found key items related to goal scoring (Tenga et al., 2010a) and score box entry (Tenga et al., 2010b) by logistic regression analysis. The purpose of performance analysis is the improvement of team performance; and improving performance requires the analysis of both causal relationships and criterion-referenced measurement of game performance.

In language education research, item response theory (IRT) is applied to construct criterion-referenced measurement of English skill, and it is also applied to the measurement of motor skills and tactical skill in sports (Zhu and Cole, 1996; Looney, 1997; Nakano et al., 2004; Ando et al., 2018a; Ando et al., 2018b). IRT was used to construct tests to measure motor skills in children (Zhu and Cole, 1996), activities of daily living in elderly persons (Nakano et al., 2004), and figure skating skills in athletes (Looney, 1997). Ando et al. (2018) developed a criterion-referenced test of soccer tactical skills. A self-rating style of video-based questionnaire which consisted of a 5-point scale was constructed. In order to apply IRT, the 5-point items were converted to a binary scale (completely capable = 1, incapable = 0) by converting positive answers (over 3 points) to “Completely capable” and negative answer (less than 2 points) to “Incapable”. The constructed criterion-referenced test showed high reliability and validity.

Then, the test characteristic was used to develop a computer adaptive test of soccer tactical skills (Ando et al., 2018b). In the next stage of research, the criterion-referenced test of soccer game performance to provide real-time feedback is necessary because coaches require the analysis to improve soccer game performance (Wright et al., 2012).

Previous research failed to analyse defence performance during the game because most studies were limited to analysing only game event statistics (Barreia et al., 2014a; Barreira et al., 2014b; Maleki et al., 2016; Casal et al., 2016). However, Suzuki and Nishijima (2005) collected data representing the performance of defence players from video image of soccer game and analysed the causal structure of defence performance by SEM. Their study constructed the items with positional information to measure defence performance because they expected that the development of the technology such as tracking data in the future. The top soccer leagues in Europe and Japan collect tracking data from all league games, which allows the measurement of the positions of all players in games; thus, it is possible to construct defence tactical play items, which can be processed from tracking data in real-time.

A criterion-referenced test of soccer defence tactical play is possible to construct using IRT and items which are processed from tracking data by converting the items in ratio scale to a binary scale. A number of measurement items in sports were shown in ratio scale, making it difficult to apply IRT (Zhu and Yang, 2016). This means that ratio-scale items processed from tracking data must be converted to binary data. The value of branch to classify the two classes can be analysed by decision tree analysis, data mining or machine learning technique. Decision tree analysis is normally used to classify or predict. By defining successful and unsuccessful defence play, it is possible to obtain the split value of the items to classify the result of defence play. The value of branch of the items is assumed as "success criteria", and the items in ratio-scale could be converted to binary scale with the success criteria; thus, it is possible to apply IRT to the items processed by tracking data.

The purpose of this study was to develop criterion-referenced measurement items for soccer defensive tactical play from tracking data.

2. Methods

2.1. Samples

Defence plays (n = 1,711) were extracted from tracking data and ball-touch data of two games in 2016 J-League. All data was borrowed by the university from Data Stadium Inc. (Tokyo, Japan) under academic-industrial alliance. This study was approved by the Research Ethics Committee at the Faculty of Health and Sport Sciences, University of Tsukuba (Project No. 30-54).

2.2. Item construction

The causal structure of soccer defensive tactical plays and the items were constructed via qualitative analysis. The causal structure was extracted by soccer specialists (coaches, players, and analysts) using the Delphi method and the causal-effect analysis. Then the fish-bone diagram (the causal-effect diagram) was used to show causal structure. Soccer defensive tactical play items, which corresponded to the fish-bone diagram, were developed from tracking data. In ball-touch data, ball related action such as passing and shooting were recorded in time series order. On the other hand, player position coordinates (x, y) on the pitch were recorded in tracking data. Both data had frame ID, which allowed concatenation of two data points showing ball event and player positions. To develop soccer defensive tactical play items, a dataset, which included ball event and player position, was processed by the concatenation of two types of data (**Figure 1**). The items, which represented defence play distance between players or teams, distance moved, speed, and number of players were processed by tracking data of defence players (**Table1**).

Then, success criteria of soccer defensive tactical play items were analysed. Definition of successful defence play was set as 'ball gain' (Casal et al., 2016; Barreira et al., 2014b; Worthington, 1977) while unsuccessful defence was defined as a play that failed to gain the ball.

2.3. Statistical analysis

To analyse construct validity of soccer defensive tactical play items, confirmatory factor analysis using SEM was completed. Normality of the items were analysed. Absolute value of correlation between

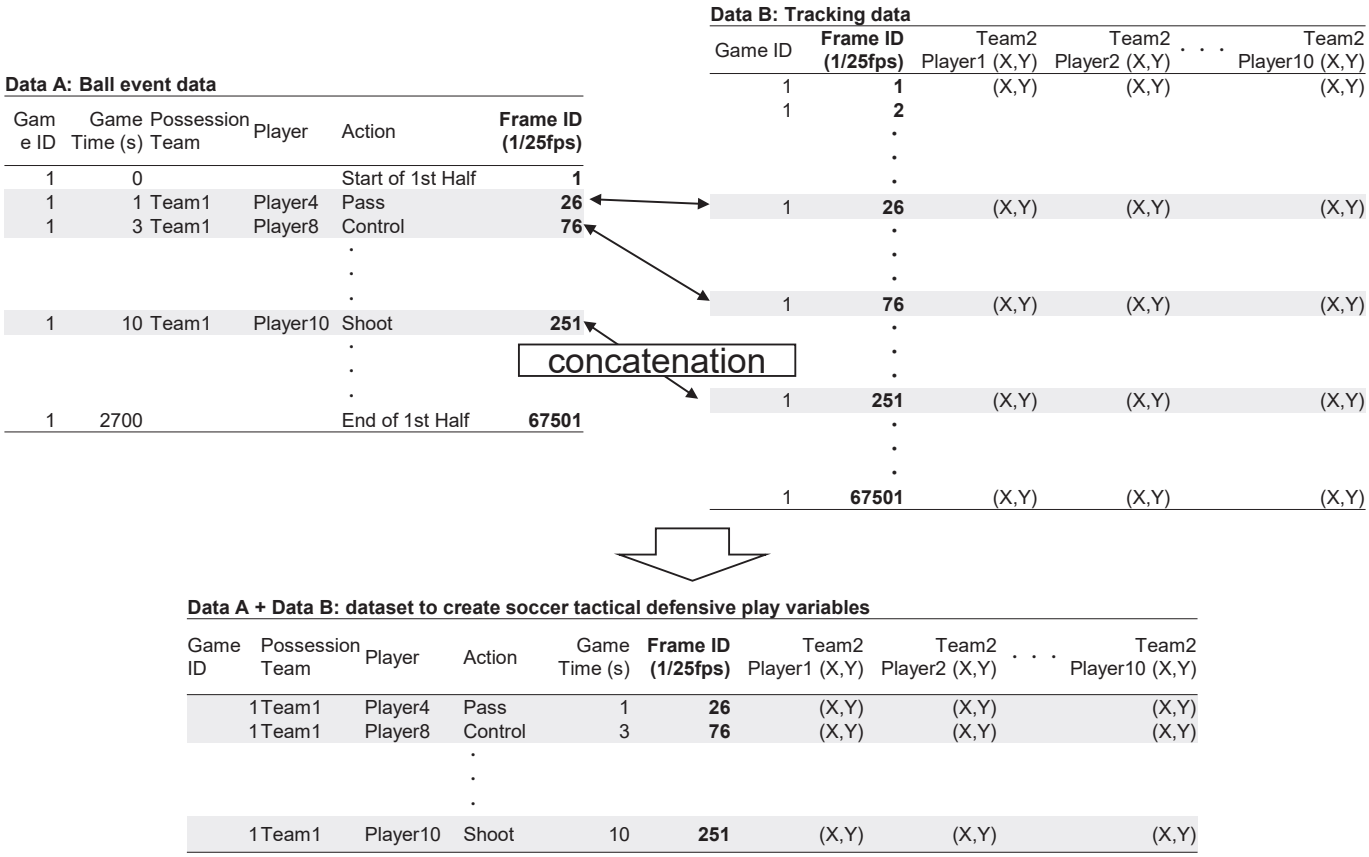


Figure 1 Processing of datasets to create soccer tactical defensive play variables by concatenating ball event data and tracking data

each item over 0.9 or less than 0.3 were then deleted from the analysis (Field, 2013). Next, Confirmatory factor structure (CFS) model and second order model (second-order CFS model) of defensive tactical play items were developed. Maximum likelihood estimation was used to estimate parameter. For analysis of model fit, GFI, AGFI, NFI, CFI, RMSEA, AIC, Chi-square value and p value were analysed. It is assumed that GFI, NFI, and CFI of greater than 0.9 is a good model fit. Less than 0.05 in RMSEA is assumed to be a good fit, and 0.06-0.08 in RMSEA is assumed to be a moderate fit (Browne and Cudeck, 1993; Kline, 2016).

To find success criteria of defensive tactical play items, a split value to classify the successful defence was analysed by classification binary tree, the classification and regression tree (CART) algorithm. To split the branch, Gini index was used. Using a split value as success criteria, defensive tactical play items in ratio-scale were converted to binary data (Achieved or Not Achieved).

Unidimensionality, item characteristics, invariance of items, test fit, reliability, and validity of the test were analysed following previous research procedures (Ando et al., 2018a). To analyse unidimensionality, single factor analysis with principal factor solution was completed. Tetrachoric correlation matrix was used for the analysis. Scree test was used to determine proportion of variance of 1st eigenvalue, and items for which factor loading was positive (>0.0) were selected for subsequent analysis.

IRT with 2 parameter logistic model was applied to estimate item characteristics and ability score (θ) of defensive tactical plays. Maximum likelihood was used to estimate parameters. Only items which met the selection criteria, item difficulty parameter within -5.0 to 5.0 and item discrimination parameter over 2.0 were used. Chi-square goodness of fit test was used to analyse whether the item fit the model. Significant level was set as 0.05.

We followed the procedure reported by Hambleton et al. (1991) and invariance coefficients (Pearson's correlation coefficients) of item parameters and ability were calculated. To analyse invariance of item parameters and ability, the estimated parameters (item difficulty, item discrimination, ability, respectively) were randomly divided into two groups. In IRT, test reliability is shown with Item information (I(θ)). Test

reliability coefficients were calculated from item information in the equation shown below (Wainer et al., 1990; Cheng et al., 2012). Usually, coefficients over 0.7 are considered fair (Nunnally, 1978).

$\rho = 1/(1+I(\theta)^{-1})$

Finally, criterion-related validity for ability and the goodness of fit to the test was analysed. Correlation between the number of achieved items of the defensive tactical play and ability (θ) of the play was used to analyse validity. The test score of defensive tactical play was then plotted on a test characteristic curve and the goodness of fit to the test was confirmed.

IBM SPSS ver.23 and IBM SPSS Amos ver.23 were used for statistical analysis and construct validity. Mplus ver.5.0 was used for analysis of unidimensionality, and Bilog-Mg ver.3.0 was used for IRT. Significance value was set at 0.05 in this study.

3. Results

3.1. Construction of soccer defensive tactical play items

Figure 2 shows the causal effect diagram of soccer defensive tactical play. Soccer defensive tactical play consisted of block and press defence. Block defence consisted of set defence and control defence as sub-defence plays while press defence consisted of concentration defence as lower defence play. A total of 25 soccer defensive tactical play items were constructed based on the causal effect diagram (**Table1**).

3.2. The causal structure model of soccer defensive tactical play items

Seventeenth items, which showed construct validity, were selected from the analysis of confirmatory factor structure model (CFS model) and second-order confirmatory factor structure model (second-order CFS model) of soccer defensive tactical play items. Model fit index of CFS model showed in GFI = 0.917, AGFI = 0.876, NFI = 0.949, CFI = 0.953, RMSEA = 0.085, AIC = 1452.367, CMIN = 1350.367, df = 102, $p < 0.000$.

Correlation coefficients between factors in CFS model were 0.50 between control defence and set defence, 0.13 between control defence and concentration defence, and -0.19 between set defence and concentration defence. Pass coefficients of six of seven items in control defence to each item and four of five items in defence to each item showed over moderate values. All pass coefficients between concentration defence to the items were high.

Model fit index of second-order CFS model showed in GFI = 0.924, AGFI = 0.882, NFI = 0.953, CFI =

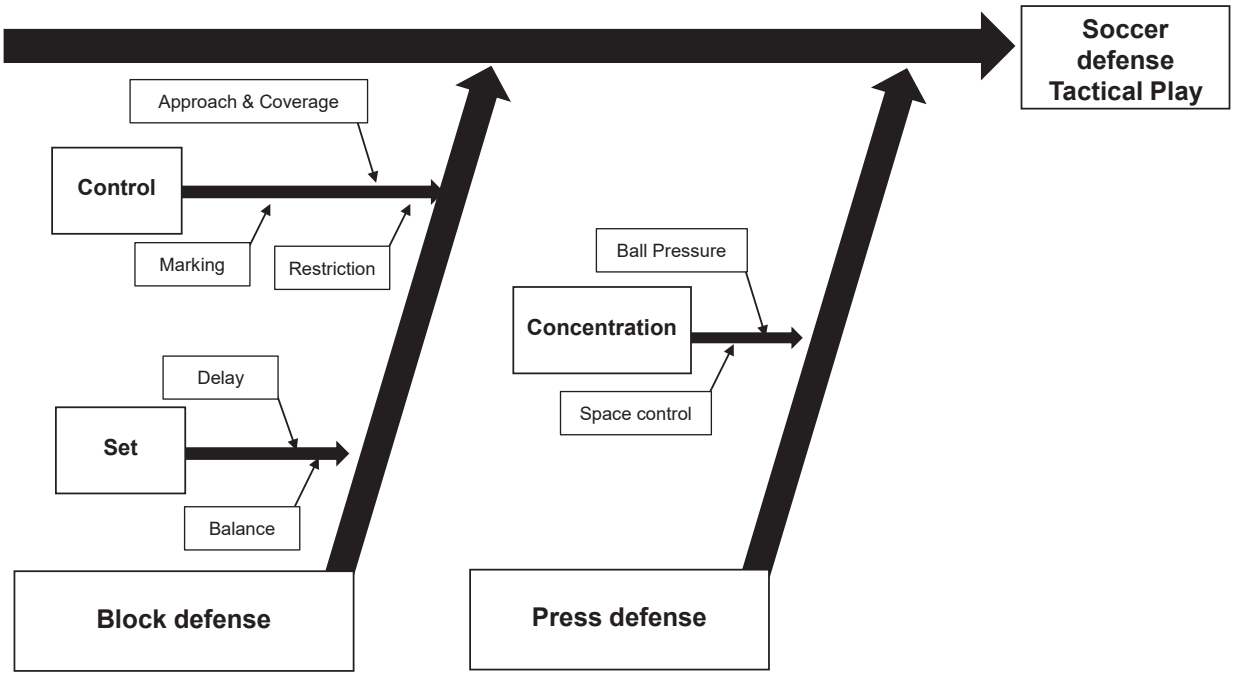


Figure 2 Cause and effect model (fishbone diagram) of soccer defensive tactical play

Table 1 Measurement of soccer defensive tactical play items using tracking data

Tactics	Lower Tactics	Objectives	ID	Items	Description
Block	Set	Delay	v01	Speed of the attacker with the ball	Running speed of the player with the ball
Defense	Set	Delay	v02	Marking of attacking players	Average distance between players on possession team and the nearest player in non-possession team
	Set	Balance	v03	Number of attacking players	Number of players on possession team located between the ball to opponent goal line
	Set	Balance	v04	Defensive organisation	Area of the defensive organization which is multiplying length of ball-the defense line by width
	Set	Balance	v05	Balance of defenders	Average distance from non-possession team centroid to players on same team
	Set	Balance	v06	Number of defenders	Number of players on non-possession team located between the ball to own goal line.
	Set	Balance	v07	Distance from ball to the DFL	Vertical distance from the ball to the defense line (DFL)
	Set	Balance	v08	Defending position	Vertical distance from the defense line to the defense goal
	Control	Approach & Coverage	v09	Defending area	Area of three nearest players on non-possession team to the ball
	Control	Approach & Coverage	v10	Disturbance of the attacker	Average distance of two nearest players on non-possession team to the ball
	Control	Approach & Coverage	v11	Number of ball defense players	Number of players on non-possession team within 9.15m from the ball
	Control	Approach & Coverage	v12	Progression of the attacker with the ball	Vertical distance covered by the player with the ball
	Control	Approach & Coverage	v13	Coverage of the 2nd DF	Distance between the nearest player to the ball and the second nearest player to the ball on non-possession team
	Control	Approach & Coverage	v14	Challenge of the 1st DF	Distance from the nearest player on non-possession team to the ball
	Control	Marking	v15	Marking of the supporting attacker1	Distance between the nearest player on possession team to the ball and the closest marking player on non-possession team
	Control	Marking	v16	Marking of the supporting attacker2	Distance between the second nearest player on possession team to the ball and the closest marking player on non-possession team
	Control	Restriction	v17	Defending against progression	Degree of defending the goal, the player with the ball, and the nearest player on non-possession team to the ball
	Control	Restriction	v18	Pass course cut1	Degree of supporting attacker1, the ball, and the closest marking player on non-possession team to supporting attacker1
	Control	Restriction	v19	Pass course cut2	Degree of supporting attacker2, the ball, and the closest marking player on non-possession team to supporting attacker2
Press	Concentration	Ball pressure	v20	Running speed of defenders	Average running speed of players on non-possession team
Defense	Concentration	Ball pressure	v21	Approaching of the defender	Running speed of the nearest player on non-possession team to the ball
	Concentration	Ball pressure	v22	Pressure of three defenders	Average running speed of three nearest players on non-possession team to the ball
	Concentration	Space control	v23	Coverage of DFL	Vertical movement of the defense line on the pitch
	Concentration	Space control	v24	Stability of defender's movement	Variability of movement by the players on non-possession team
	Concentration	Space control	v25	Stability of defender's movement speed	Variability of running speed of the players on non-possession team

0.956, RMSEA = 0.083, AIC = 1371.043, CMIN = 1263.043, df = 99, $p < 0.000$. Pass coefficients between soccer defensive tactical play and control defence, set defence and concentration defence were 0.58, 0.83 and -0.22, respectively. Pass coefficients of six of seven items to control defence and four of five items to set defence showed over moderate value. All pass coefficients between concentration defence to the items were high. All pass coefficients in the models showed significance ($p < 0.05$) (**Figure 3**).

CFS model and second-order CFS model showed good model fit greater than 0.9 of GFI, NFI, and CFI values and a moderate level of model fit in RMSEA. Model fit of CFS model and second-order CFS model of soccer defensive tactical play items was fair.

3.3. Success criteria of soccer defensive tactical play items

Success criteria of soccer defensive tactical play items were analysed (**Table 2**). Items (Approaching of defender and Speed of attacker with the ball) were excluded from subsequent analysis because they could not be used to classify independent variables.

3.4. Unidimensionality of soccer defensive tactical play items

Figure 4 shows the result of exploratory factor analysis to analysis of unidimensionality of 14 soccer defensive tactical play items. Scree plot shows an L-shape, and the contribution rate of the first eigenvalue differed significantly from the contribution rate of the second eigenvalue. The contribution rate of the first eigenvalue was 6.33, which explained 45.22 % of total variance while the contribution rate of the second eigenvalue only explained 19.30%. Max value of 1st factor loading in 14 soccer defensive tactical play items was 0.90 while minimum value was 0.29. Mean±SD was 0.61±0.23 and all values were positive (**Table 2**).

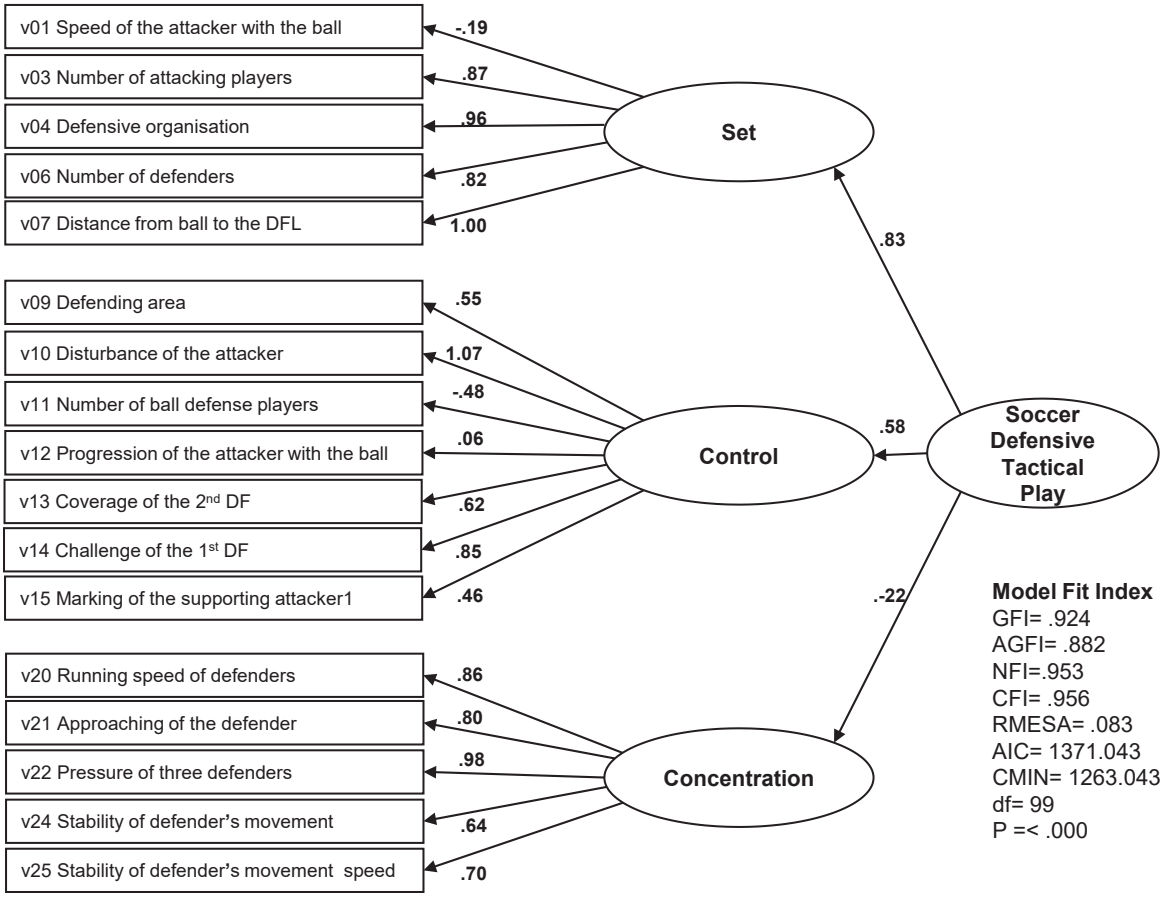
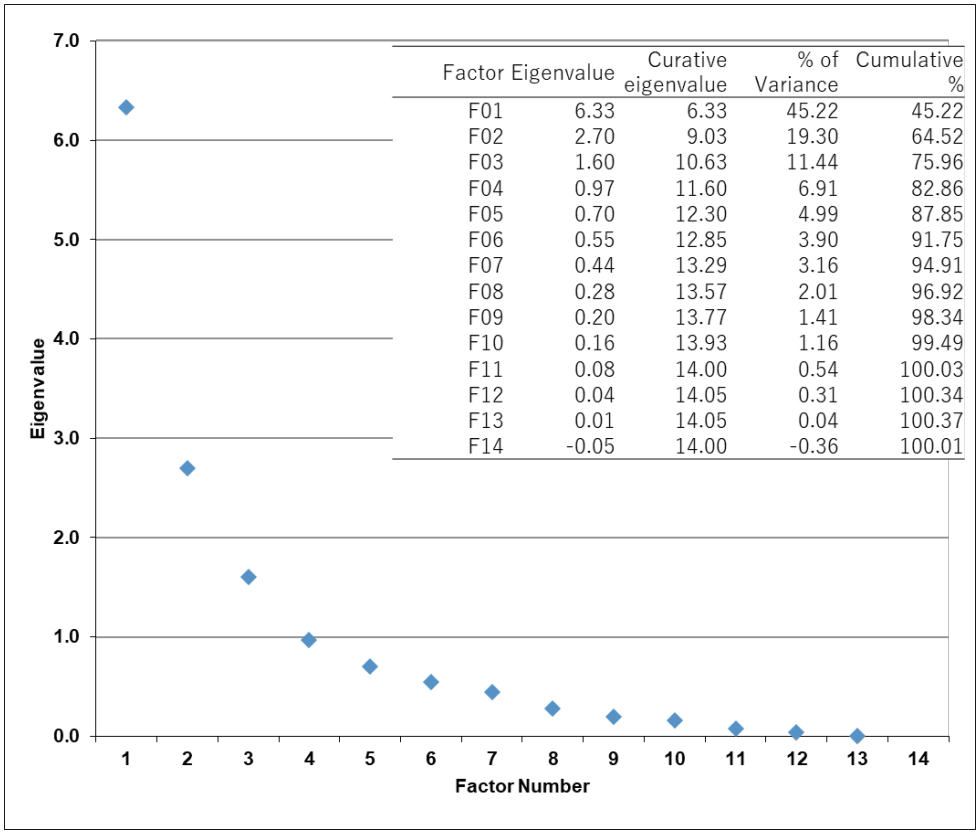


Figure 3 Second-order confirmatory factor structure of soccer defensive tactical play items



Factor	Eigenvalue	Curative eigenvalue	% of Variance	Cumulative %
F01	6.33	6.33	45.22	45.22
F02	2.70	9.03	19.30	64.52
F03	1.60	10.63	11.44	75.96
F04	0.97	11.60	6.91	82.86
F05	0.70	12.30	4.99	87.85
F06	0.55	12.85	3.90	91.75
F07	0.44	13.29	3.16	94.91
F08	0.28	13.57	2.01	96.92
F09	0.20	13.77	1.41	98.34
F10	0.16	13.93	1.16	99.49
F11	0.08	14.00	0.54	100.03
F12	0.04	14.05	0.31	100.34
F13	0.01	14.05	0.04	100.37
F14	-0.05	14.00	-0.36	100.01

Figure 4 Scree-plot of eigenvalues of soccer defensive tactical play items

3.5. Item characteristics and test characteristics of soccer defensive tactical play items

Item characteristics of 14 soccer defensive tactical play items are shown in **Table 2**. The items consisted of 4 items for set defence, 6 items for control defence, and 4 items for concentration defence. One item (Progression of attacker with the ball) was excluded because the discrimination parameter did not satisfy the selection criteria. Chi-suqare values for itmes was not significance ($p \geq 0.05$). Mean±SD of item difficulty parameter were 0.61±.23, and Mean±SD of item discrimination parameter was 0.61±0.23. Coefficient of invariance was $r = 0.99$ ($p < 0.05$) for item difficulty parameter, $r = 0.99$ ($p < 0.05$) for item discrimination parameter, and $r = 0.85$ ($p < 0.05$) for ability parameter (θ)

Figure 5 shows goodness-of-fit of data to the test characteristic curve and the criterion-related validity of the criterion-referenced measurement test of soccer defensive tactical play. It shows a scatter plot of ability score and test score with test characteristic curve. Coefficient of the criterion-related validity, the correlation between the ability score and test score, was 0.93 ($p < 0.05$). Moreover, ability score was visually fitted to the test characteristic curve.

Figure 6 shows reliability of the criterion-referenced measurement test of soccer defensive tactical play. Maximum reliability coefficient was 0.98 at the ability score 0.5. Reliability coefficient was greater than 0.7 in the range between ability score -1.0 to 1.0.

3.6. Validity of criterion-referenced test for soccer defensive tactical play

Ability score of soccer defensive tactical play with ball gain (0.42±0.81) was significantly higher ($t(272.98) = -7.85$, $p < 0.05$) than ability score with non-ball gain (-0.06±0.91).

Figure 7 shows the change of ability score of soccer defensive tactical play. Horizontal axis of **Figure 7** shows attacking plays of the opponents in time order. This example shows that the defence play started from pass by the opponent (offence) team and finished with ball gain by interception of the defence team. The vertical axis of **Figure 7** shows the ability score of soccer defensive tactical play corresponding to the opponents' attacking play. The ability score of soccer defensive tactical play gradually increased at

Table 2 Criterion of success and item characteristics of soccer defensive tactical play items

ID	Items		Criteria	Factor loading	Achievement rate	Item difficulty	Item discrimination	Chi-SQ	p
v03	Number of attacking players	≤	6.50	0.89	63.20	-0.30	3.04	12.48	0.19
v04	Defensive organisation	≤	594.19	0.80	36.10	0.40	3.18	2.24	0.95
v06	Number of defenders	≤	7.50	0.90	51.30	0.02	2.96	0.00	0.99
v07	Distance from the ball to the DFL	≤	13.57	0.84	27.90	0.58	6.89	11.89	0.06
v09	Defending area	≤	22.82	0.63	41.10	0.42	0.64	6.80	0.15
v10	Disturbance of the attacker	≤	9.61	0.88	74.60	-0.79	1.43	0.00	0.96
v11	Number of ball defense players	>	1.50	0.40	58.20	-0.56	0.38	0.67	0.41
v13	Coverage of the 2nd DF	≤	5.03	0.63	23.60	1.27	0.69	1.48	0.69
v14	Challenge of the 1st DF	≤	5.58	0.72	66.70	-0.62	0.91	0.20	0.66
v15	Marking of the supporting attacker1	≤	8.79	0.52	66.00	-0.88	0.51	1.05	0.31
v20	Running speed of defenders	>	5.03	0.29	89.20	-4.35	0.30	3.16	1.00
v22	Pressure of three defenders	>	15.89	0.31	14.40	3.24	0.35	0.37	0.54
v24	Stability of defender's movement	>	6.59	0.36	3.40	4.16	0.52	15.00	0.09
v25	Stability of defender's movement speed	>	4.17	0.41	48.70	0.10	0.33	0.02	0.89
	mean			0.61	47.46	0.19	1.58	3.95	0.56
	sd			0.23	24.38	1.98	1.87	5.32	0.36
	max			0.90	89.20	4.16	6.89	15.00	1.00
	median			0.63	50.00	0.06	0.66	1.27	0.60
	min			0.29	3.40	-4.35	0.30	0.00	0.06

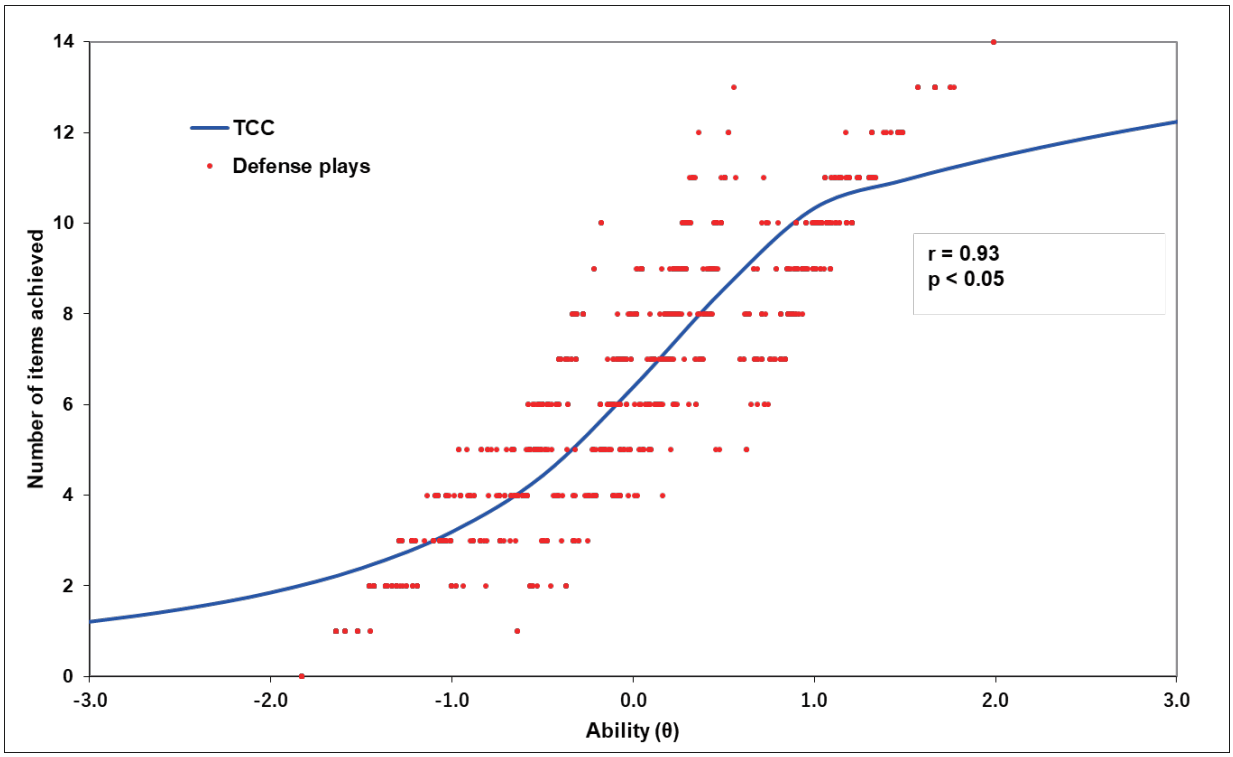


Figure 5 Validity and goodness of fit of criterion-referenced test for soccer defensive tactical play

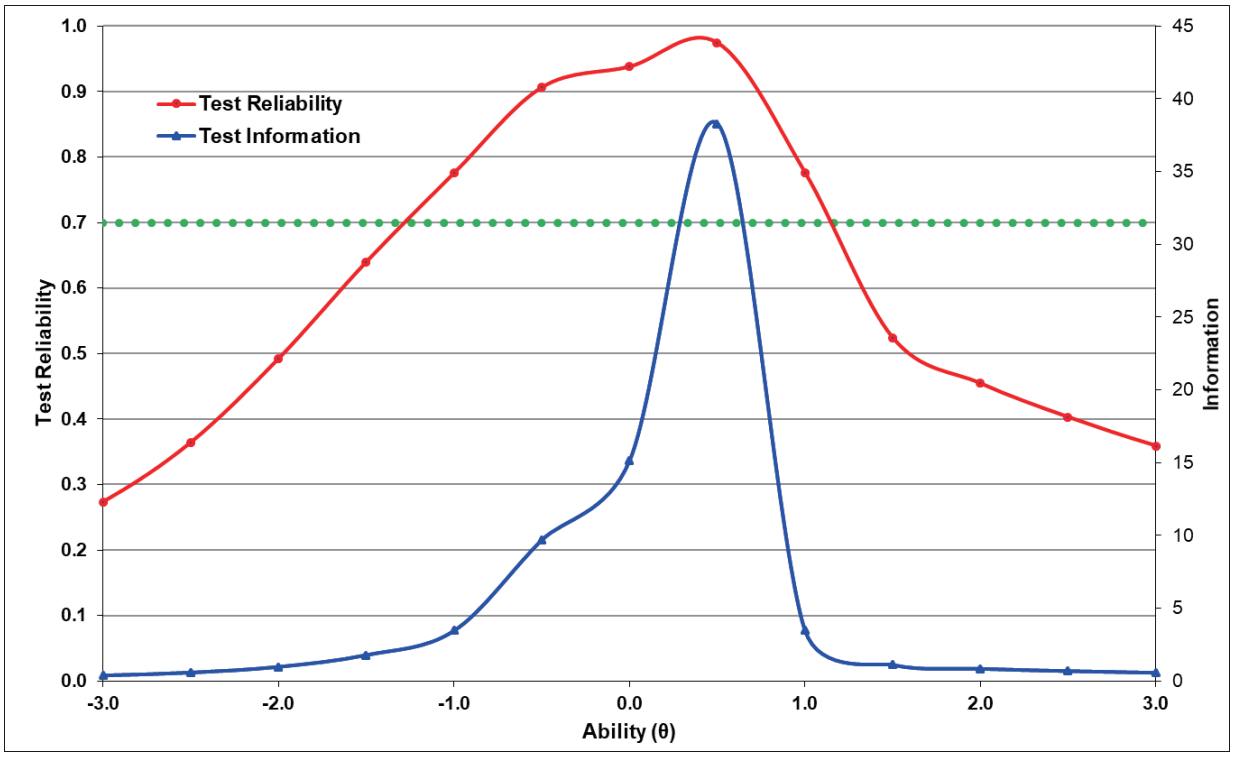


Figure 6 Reliability and information function of criterion-referenced test for soccer defensive tactical play

four plays before interception, and the score achieved the maximum value at interception. In this play, the number of items achieved in each lower defensive tactical play was 4 of 4 items in set defence, 6 of 6 items in control defence, and 1 of 4 items in concentration defence.

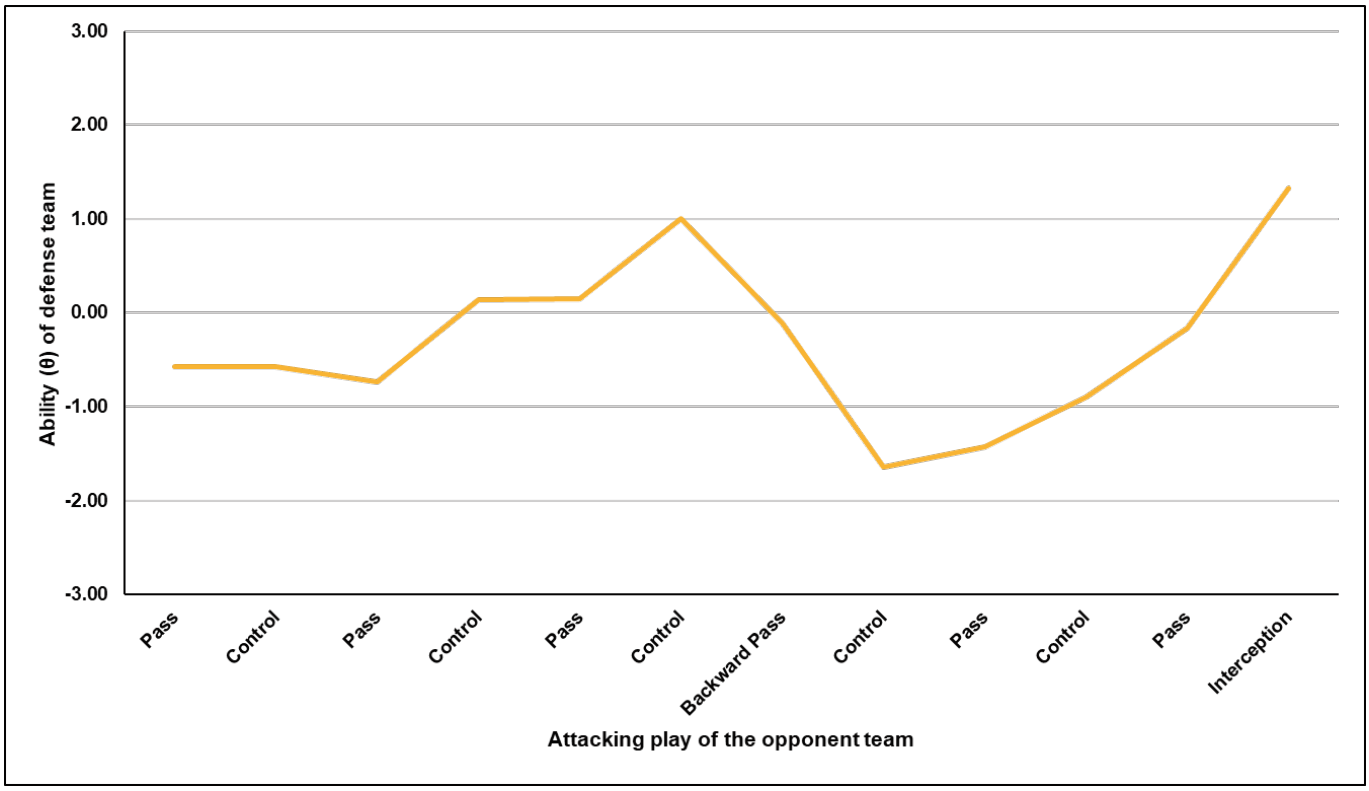


Figure 7 Change of ability scores of soccer defensive tactical play corresponding to the attacking play of the opponent team

4. Discussion

4.1. Construct validity of soccer defensive tactical play

In previous research, soccer tactical principles and construct of soccer game performance were studied through literature review or SEM with 5-points scales (da Costa et al., 2009; Hewitt et al., 2016; Suzuki and Nishijima, 2005; Wade, 1998). The purpose of defence is to protect the goal and regain the ball (Worthington, 1977). Defence principles are delay, coverage, balance, and concentration (da Costa et al., 2009; Hewitt et al., 2016; Suzuki and Nishijima, 2005; Wade, 1998). Delay is aimed at disturbing the offence by approaching the ball holder and restricting passing opportunities. In the coverage principle, the players on the defence team support the first defender and strengthen marking to the opponents. The defence principle of balance is aimed at organising a defensive formation that interrupts the offence and forces the opponents into a smaller space. In concentration, the defence decreases the ball holder's space and time with the aim of gaining the ball.

In this study, soccer defensive tactical play items were processed by tracking data, and the construct of soccer defensive tactical play was analysed by SEM. Model fit index and pass coefficients on SEM were statistically valid values. As a result of this study, sub-defensive tactical plays were set defence, control defence, and concentration defence. Set defence represents the defence play to maintain defensive organisation. Control defence represents the defence play to restrict and guide the opponent's offence move while concentration defence represents the defence play to increase player density around the ball. The tactical plays in this study are supported by defensive principles (delay, coverage, balance, concentration), and the sub-defensive tactical plays are a valid construct (da Costa et al., 2009; Hewitt et al., 2016; Suzuki and Nishijima, 2005; Wade, 1998).

Because Suzuki and Nishijima (2005) expected that the items in units with distance or degree of defensive players would be collected automatically by the development of the tracking technology in the future, they constructed the items in those units and converted their units to a 5-point scale to measure game performance. In this study, construct validity of soccer defensive tactical plays was investigated by items in ratio scales. In recent years, data technology for the measurement of soccer games is continuously developing, and it will allow real-time feedback

during the game because the items can be processed by video.

4.2. Success criterion of soccer defensive tactical plays

IRT was successfully applied to the items that were processed from tracking data by converting the ratio scale of items to a binary scale with decision tree analysis. It was problematic that IRT could not be applied to ratio scale items. Previous research successfully applied IRT analysis to a 5-point scale by converting to binary data of positive response (5 and 4) and negative response (3, 2, 1) (Ando et al., 2018). By decision tree analysis, the success criteria of the item for successful defensive tactical play were obtained. One common use of decision tree analysis is rule classification. For example, Nakano et al. (2007) showed the effectiveness of decision tree analysis in exploring optimal rules of lifestyle of subjective conditioning. However, this study used desicion tree analysis to find a branch value of classification of the item success and solved the problem that converted the ratio scale to binary scale by using a branch value as a cut-off point of success criteria.

Successful defensive tactical play was defined as ball gain (Casal et al., 2016; Barreira et al., 2014b; Worthington, 1977). The success criterion related to defensive principle, such as decreasing time and space for ball gain, were found. The items of block defence showed success criteria expressing compact defence (e.g. success criteria of 'v07 Distance from the ball to the DFL' was ≦13.5m). FIFA (2018) reported the length of defensive team (the distance between the player in the highest position and the deepest position except GK) was approximately 26m. On the other hand, the items of press defence showed success criteria expressing decreased time and space (e.g. success criteria of 'v22 Pressure of three defenders' was ≧15.9km/h). The success criteria show faster movement of defence players to close with the ball holder (da Costa et al., 2009; Hewitt et al., 2016; Wade, 1998). This study excluded two items of 'v01 Speed of attacker with the ball' and 'v21 Approaching of defenders' from subsequent analysis because it was not possible to identify success criteria from the analysis.

4.3. Item characteristics of soccer defensive tactical play items

Fourteen valid items were selected by evaluating item characteristics of soccer defensive tactical play items which were converted to binary data using success criteria. The selected items showed unidimensionality, which means the items measure only one latent ability. Result of Chi-square test showed the selected items fit to the model. Item difficulty parameters in set defence items showed relatively moderate values while item discrimination parameters showed high values. Those items were required to be successful because they were premise to defence against opponents. Items such as defensive organisation and distance from the ball to the DFL showed high discrimination parameters. These were important items for the measurement of compact defence against opponent offence (da Costa et al., 2009; Hewitt et al., 2016). It was interpreted that the items were KPI to measure the achievement of soccer defensive tactical play because of its higher standard of item difficulty parameters. Item difficulty parameters in control defence items showed relatively easy items. The level of item difficulty parameters should be fair because defensive tactics of 'challenge and coverage' and 'mark defence to opponent players' were basic disciplines of defensive tactics (da Costa et al., 2009; Hewitt et al., 2016). Item difficulty parameters in concentration defence items were high. Items such as 'v22 Pressure of three defenders' were considered important and difficult items in defensive play to reduce the time and space of ball holders. In this analysis, 'v12 Progression of attacker with the ball' in control defensive tactical items were excluded as items that did not satisfy the criteria of item characteristics. It was assumed that the item has content validity and the pass coefficient was low in the analysis of construct validity. However, the item characteristics were also insufficient to include the analysis in IRT. Thus, they were invalid and excluded the analysis.

4.4. Test characteristics of the criterion-referenced test of soccer defensive tactical play

In IRT, reliability of the test is analysed by item information and test information. Reliability coefficient is calculated from test information (Wainer et al., 1990 ; Cheng et al., 2012). Reliability

coefficient of the test with 14 items was over 0.7 in the range between -1.0 to 1.0 of ability score. Unfortunately, reliability coefficient was below 0.7 below -1.0 and above 1.0 of ability score. In IRT, ability score is standardized in the estimation with a mean value of 0 and a standard deviation of 1. Thus, reliability coefficient of the test in this investigation shows that the test covers 68% of data range with high reliability. Moreover, criterion-related validity of the test was found because correlation coefficient between test score and ability was significantly high ($r = 0.93$, $p < 0.05$). The test score, the ability score, and the test characteristic curve were shown on scatter plot, and the ability score was visually fitted on the test characteristic curve (Hambleton et al., 1991). The criterion-referenced test of soccer tactical defensive play showed reliability, validity, and model fit of the data.

4.5. Validity of the criterion-referenced test of soccer defensive tactical play

Ability score at ball gain was significantly higher than ability score when the ball was not gained. This result shows the validity of soccer defensive tactical play items as criterion-referenced measurement items. Moreover, ball gain is an important purpose of defence (Worthington, 1977) and it seemed a valid result. **Figure 7** shows the change of ability score of soccer defensive tactical play corresponding to the attacking play of the opponent team. The scores are different in per opponent action. At opponent pass, which was four plays before interception, the ability score increased and achieved the highest value at ball gain. At interception of the defensive team, all items of set and control defensive tactical play were achieved; therefore, it was assumed that the defensive team organised a defensive block to force the opponents to move the ball to where the defensive team wanted to gain it. The achievement of soccer defensive tactical play can be measured by both the ability score and the number of items achieved.

5. Conclusions

The purpose of this study was to develop criterion-referenced measurement items for soccer defensive tactical play from tracking data in soccer games. Investigation yielded the following results:

1) Soccer defensive tactical plays consisted of press defensive tactical play and block defence tactical play.
2) Seventeen items of soccer defensive tactical items from tracking data, which consist of 5 items of set defensive tactical play items, 7 items of control defensive tactical items, and 5 items of concentration defensive tactical play items, have construct validity.
3) Fourteen items of the criterion-referenced items of soccer defensive tactical play have unidimensionality of items, model fit to 2 parameter logistic model, invariance of item parameters and ability parameters.
4) Fourteen items of the criterion-referenced test of soccer defensive tactical play have reliability, validity, and model fit to 2 parameter logistic model.

Acknowledgement

We appreciate Data Stadium Inc. for permission to use the soccer game data for this study.

References

Ando, K., Mishio, S., and Nishijima, T. (2018a). Analysis of items and characteristics of a criterion-referenced self-administered test utilizing video images for the development of a computerized adaptive test for tactical skills in soccer. Football Science, 15: 26-37.

Ando, K., Mishio, S., and Nishijima, T. (2018b). Validity and reliability of computerized adaptive test of soccer tactical skill. Football Science, 15: 38-51.

Barreira, D., Garganta, J., Guimarães, P., Machado, J., and Anguera, M. T. (2014a). Ball recovery patterns as a performance indicator in elite soccer. Proc. Inst. Mech. Eng. P. J. Sport Eng. Technol., 228: 61-72.

Barreira, D., Garganta, J., Machado, J., and Anguera, M. T. (2014b). Effects of ball recovery on top-level soccer attacking patterns of play. Rev. Bras. de Cineantropometria e Desempenho Hum., 16: 36-46.

Browne, M. W., and Cudeck, R. (1993). Alternative ways of assessing model fit. In: Bollen. K.A. and Long. J. S. (eds.), Testing structural equation models. Newbury Park, Sage.

Casal, C., Andujar, M., Losada, J., Ardá, T., and Maneiro, R. (2016). Identification of defensive performance factors in the 2010 FIFA World Cup South Africa. Sports., 4: 54.

Cheng, Y., Yuan, K., and Liu, C. (2012). Comparison of reliability measures under factor analysis and item response theory. Educ. Psychol. Meas., 72: 52-67.

da Costa, I. T., da Silva, J. M. G., Greco, P. J., and Mesquita, I. (2009). Tactical principles of soccer: Concepts and application. Motriz., 15: 657-668.

Field, A. (2013). Discovering statistics using IBM SPSS statistics (4th ed.), Germany, Sage.

FIFA. (2018). Technical report 2018 FIFA World Cup Russia.

Hambleton, R. K., Swaminathan, H., and Rogers, H. J. (1991). Fundamentals of item response theory, USA, Sage.

Hewitt, A., Greenham, G., and Norton, K. (2016). Game style in soccer: What is it and can we quantify it?. Int. J. Perf. Anal. Spor., 16: 355-372.
Kline, R. B. (2016). Principles and practice of structural equation modelling. Guilford Publications.
Looney, M. A. (1997). Objective measurement of figure skating performance. J. Outcome Meas., 1: 143-163.
Maleki, M., Dadkhah, K., and Alahvisi, F. (2016). Ball recovery consistency as a performance indicator in elite soccer. Rev. Bras. de Cineantropometria e Desempenho Hum., 18: 72–81.
Nakano, T., Nishijima, T., and Ohtsuka, K. (2004). Feasibility of ADL test construction based on IRT. Int. J. Sport Health Sci., 2: 145-155.
Nakano, T., Nishijima, T., and Suzuki, T. (2007). The effectiveness of decision tree analysis on routine athletic conditioning data. Int. J. Sport Health Sci., 5: 12-20.
Nunnally, J. C. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.
Suzuki, K., and Nishijima, T. (2005). Measurement of a soccer defending skill using game performances. In T. Reilly, J. Cabri, and D. Araújo (eds.), Science and football V. The proceedings of the fifth world congress on science and football (pp. 253–261). Lisbon: Routledge.
Tenga, A., Kanstad, D., Ronglan, L. T., and Bahr, R. (2009). Developing a new method for team match performance analysis in professional soccer and testing its reliability. Int. J. Perf. Anal. Spor., 9: 8-25.
Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010a). Effect of playing tactics on goal scoring in Norwegian professional soccer. J. Sports Sci., 28: 237-244.
Tenga, A., Holme, I., Ronglan, L. T., and Bahr, R. (2010b). Effect of playing tactics on achieving score-box possessions in a random series of team possessions from Norwegian professional soccer matches. J. Sports Sci., 28: 245-255.
Wade, A. (1998). Principles of team play. USA :Reedswain.
Wainer, H., Dorans, N. J., Flaugher, R., Green, B. F., and Mislevy, R. J. (2000). Computerized adaptive testing: A primer. New Jersey:Routledge.
Wright, C., Atkins, S., and Jones, B. (2012). An analysis of elite coaches' engagement with performance analysis services (Match, notational analysis and technique analysis). Int. J. Perf. Anal. Spor., 12: 436-451.
Worthington, E. (1977). Learning and teaching soccer skills. Wilshire Book Company.
Zhu, W., and Cole, E. L. (1996). Many-faceted rasch calibration of a gross motor instrument. Res. Q. Exerc. Sport, 67: 24-34.
Zhu, W., and Yang, Y. (2016). Item response theory and its applications in kinesiology. In N. Ntoumanis, and N. D. Myers. (eds.), An introduction to intermediate and advanced statistical analyses for sport and exercise scientists (pp.57-77). Singapore: Wiley.

Name:
Hiroki Matsuoka, MS

Affiliation:
Doctoral Program in Physical Education, Health and Sport Sciences, Graduate School of Comprehensive Human Sciences, University of Tsukuba

Address:
1-1-1 Tennodai, Tsukuba, Ibaraki 305-8574 Japan

Brief Biographical History:
2017- Doctoral Program in Physical Education, Health and Sport Sciences, University of Tsukuba

Main Works:
• Matsuoka, H., Tahara, Y., Yamamori, A., Honda, T., Ando, K., and Nishijima, T. (2019). Prediction of soccer attacking play patterns by natural language processing. The Institute of Statistical Mathematics Cooperative Research Report 423 Research on Sports Data Analytics: Theory, Methodology, and Applications., 6, 77-82. (in Japanese)

Membership in Learned Societies:
• Japanese Society of Science and Football
• Japan Society of Physical Education, Health and Sport Sciences