Paper : Football(Soccer)

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

Kozue Ando*, Shota Mishio* and Takahiko Nishijima*

*Faculty of Health & Sport Sciences, University of Tsukuba
1-1-1 Tennodai, Tsukuba, 305-8574 Japan
fcrando@googlemail.com
[Received January 9, 2018; Accepted March 8, 2018]

The purpose of this study was to analyze the items and characteristics of a criterion-referenced self-administered test utilizing video images for the development of a computerized adaptive test of tactical skills in soccer (TSS-CAT). In total 141 male soccer players in a University club, age (20.2±0.99 years), athletic career (11.9±2.16 years) participated in the study. The structure and the test items of tactical skills in soccer were constructed using qualitative analysis. The test items regarding tactical skills in soccer were analyzed based upon item response theory (IRT) with the 2-parameter logistic model (2PLM). The test items of tactical skills in soccer consisted of 4 domains of individual and group attacking tactical skills, and individual and group defensive tactical skills. The item uni-dimensionality, goodness-of-fit to the item characteristic curve (ICC), invariance of estimated parameters and ability score, the test reliability, validity and goodness-of-fit to the ICC were examined using IRT with 2PLM. The results confirmed the item characteristics of estimated parameters and ability score applying to TSS-CAT.

Keywords: criterion-referenced measurement, item response theory, item characteristics, test characteristics, causal effect analysis

[Football Science Vol.15, 26-37, 2018]

1. Introduction

Tactical skills employed in soccer have been objectively evaluated by the visual observation of player movement during games (Hughes et al., 1997; Stiehler et al., 1993. The Teaching Committee, published by the Japan Football Association (2000), defines teaching goals as the ability to understand actual game situations, select appropriate measures, and effectively execute them. In addition, it clearly differentiates personal techniques from tactics, and introduces a soccer skill test to examine the level of personal technical skills. While performance tests are effective in measuring technical skills, however, they are less effective in measuring tactical skills associated with the assessment of game situations.

Tactical plays and skills are often evaluated visually by coaches and managers. In addition, it is necessary to employ tactical skills tests that can readily evaluate the effectiveness of training programs. The JFA Teaching Committee (2002) reported that the organizational ability of individual players during games in modern soccer has been increasingly prioritized, and teaching guidelines for youth players in many countries have set the acquisition of basic tactical skills as a target. These developments highlight the need to evaluate tactical skills.

Tactical skills are essential to performance in both the offense and defense phases of play. However, there are as yet no standardized tests that measure these. The functional reason for this is that tactical skill cannot be measured by performance tests due to the fact that performance in games is a continuum of success or failure in multiple tactical plays, making evaluation of the success or failure of individual tactical plays a qualitative rather than quantitative judgment. The level of achievement in a wide variety of tactical plays is the subject of criterion-referenced test. Quantifying the success or failure of tactical plays to use as base data, it is required to develop test methods to evaluate tactical skills utilizing the level of achievement in tactical plays.

Item Response Theory (IRT) applies to criterion-

referenced test. IRT allows the construction of ability scales from correct response pattern data in paper-based questionnaires, the estimation of item characteristic and ability values, and the construction of item pools for computerized adaptive tests (CATs). In addition, IRT solves problems encountered in traditional test theory applied to relative evaluation. IRT provides three major advantages: (1) It enables the calculation of item characteristics without dependence upon subject groups; (2) It enables the measurement of ability values without dependence upon items; and (3) It enables the examination of measurement accuracy. Considering these characteristics, application of IRT to tactical play pattern data would allow construction of criterion-referenced test for tactical skills in both offense and defense.

In addition, the computerized adaptive test (CAT) for criterion-referenced test is easy to use (Watanabe et al., 1999). CAT has been applied broadly in language proficiency and vocational aptitude testing. CAT does not, however, evaluate knowledge on a pass/ fail basis. Instead, it is a maze-type skill test in which players answer individual computerized questions about the level of a tactical play (evaluation of the individual player's ability or inability to perform a given tactical play). CAT is capable of showing video images of the tactical play asked about in each item. This ensures the examinee's understanding of the plays being asked about in each item, something which is often a problem in paper-based questionnaires.

Although it may be possible to evaluate game performance by evaluating the level of achievement of tactical plays, no studies have been conducted on criterion-referenced test and CAT in the evaluation of tactical skills in soccer.

Therefore, the purpose of this study was to clarify the items and characteristics of items in self-administered criterion-referenced test utilizing questions based on video images, and the characteristics of the test to select items for pools of computerized adaptive tests for tactical skills in soccer (TSS-CAT).

2. Method

2.1 Samples

Samples used for the analysis were 141 male university soccer players. Their age (mean±SD) was 20.2±0.99, athletic career was 1.9±2.16 years, height was 174.1±6.26cm, and weight was 66.9±5.77kg. We explained the purpose and content of this study and obtained consent for participation. This study was approved by the Research Ethics Committee at the Graduate School of Human Comprehensive Sciences, University of Tsukuba (Project No. 22-364).

2.2 Item construction

Qualitative analysis was used to construct tactical skill items with content validity in accordance with the classification by the JFA Teaching Committee (2002) and tactical skill terminology by FIFA Coaching (FIFA, 2004). Cause-and-effect analysis using the Delphi method (QC Technique Development Division, 1979; Hosoya, 1982; Sugiyama 1988) was applied to four soccer specialists to construct a hierarchical scheme of domains, tactics, and items of tactical skills in soccer. This is a qualitative analytical method that integrates the opinions of subjects through repeated feedback. Cause-and-effect analysis is also a qualitative analytical method in which elements are placed into a hierarchical structure and shown in a fishbone diagram. The four soccer specialists were a JFA certified Class S coach who had experience managing the Japan National Team, a soccer player of the Japan National Team with a JFA certified Class C coaching license, an technical assistant coach of the Japan National Team, and a soccer researcher with experience coaching the Universiade Japan National Team.

Titled video images consisted of offensive or defensive tactical plays during games. The videos were highlighted with arrows and circles to indicate tactical movements and edited with slow-motion replay and stop-motion functions. Each video segment was approximately 30 seconds long and the total time used for all 82 items was approximately 41 minutes (2,460 seconds).

Subject ability to evaluate tactical plays was measured by questionnaire responses based on a 5-point scale. In the preliminary survey, we confirmed that responses differed by subject, that level of performance influenced evaluation, and that the standard achievement response was indicated to solve the problems. Based on these findings it was deemed necessary to set quantitative rates for each response to prevent subjective influence on evaluation (Dunning-

Kruger effect). For this purpose, oral instructions were provided before the survey. Ability to evaluate the tactical plays in the items was based on data from the preliminary survey. "Completely capable" was set as 90%, "Highly capable" was set as 70%, "moderately capable" was set as 50%, "Slightly capable" was set as 30%, and "Incapable" was set as 0%. Items for which the subject could not evaluate were recorded as "Unknown." To prevent guesswork, subjects were provided a 2-point scale for tactical skill recognition.

2.3 Measurement method

The video-image tactical skill test was given to subject groups, with two 10-minute breaks provided during the test. The video images used for questionnaire were shown on a large screen, and subjects responded on individual answer sheets.

2.4 Statistical analysis

Two measurement values for each item were used to analyze the test-retest reliability of intra-class correlation coefficients. Considering the 5-point scale measurement data, data obtained from items that exhibited low reliability (R<0.30) were excluded while the mean value of two measurements for items that reveled moderate reliability (0.30>=R<0.60) were included along with data obtained from question item that revealed high reliability (0.60=<R). Because the item characteristic and ability values were estimated from the right pattern data in IRT analysis, the test-retest reliability of the items was high, and the response error was exist in the first answer, we used the data from the second for analysis. We also created binary pattern data for correct-answers utilizing three answers "Completely capable," "Highly capable," and "Moderately capable" as the correct answers (1) for IRT analysis. Utilizing the binary data, we calculated bi-serial correlation coefficients among the achievement rates, items, and test scores. The achievement rates were reflected to item difficulty. The bi-serial correlation coefficients were reflected to first factor loading and item discrimination.

In order to analyze the item characteristics and test characteristics of criterion-referenced test for development of a computerized adaptive test was performed using IRT with the 2-parameter logistic model (2PLM). We also analyzed item uni-dimensionality, goodness-of-fit to the item characteristic curve (ICC), invariance of estimated parameters (item difficulty and discrimination) and ability values, test reliability, goodness-of-fit to the test characteristic curve (TCC), and validity of ability values with the total score of the test.

Single factor analysis was performed to analyze item uni-dimensionality. Exploratory factor analysis with principal factor solution was applied to tetrachoric correlation matrix among items, and one factor was extracted. The requirements of item uni-dimensionality were a significantly high contribution rate of the first eigenvalue, which differs significantly from the contribution rate of the second eigenvalue, and a positive 1st factor loading (>0.0). Scree test was applied to confirm that there was a significant difference in contribution rates between the 1st and 2nd eigenvalues. For items with negative loading in the 1st factor, the item with the maximum absolute value was deleted to carry out repeated single factor analysis until the item with negative factor loading was completely deleted. Application software M-plus, ver. 2.11 was used for single factor analysis.

The ICC with 2PLM is described in the following formula:

$$P(\theta) = 1/(1 + \exp(-Da(\theta - b)))$$

Where "P" is percentage correct, "θ" is ability value, "D" is scale component (1.7), "a" is item discrimination parameter, and "b" is item difficulty parameter. P(θ) indicates the percentage of correct answers by subjects with ability (θ). D (scale component) is a constant at 1.7. "a" equals inclination when θ equals "b" in the graph of the model formula. The lager "a" becomes, the capacity to recognize subjects with ability value θ of approximately "b" is high. "b" indicates item difficulty. The percentage correct of subjects with an ability value θ equal to "b" is 50%. With 2PLM, the correlation between ability value (θ) and percentage correct (P) is indicated with the ICC described in the model formula utilizing these parameters. The maximum likelihood method was used to estimate parameters.

To analyze the invariance of estimated item parameters, we used Pearson's correlation coefficients between parameters estimated from the samples divided into two by the systematic random sampling method as invariance coefficients. Similarly, in order to analyze the invariance of estimated ability values, we used Pearson's correlation coefficients between ability values estimated from the item groups divided into two by the systematic random sampling method

as invariance coefficients.

Chi-square goodness-of-fit test was applied to analyze the goodness-of-fit to the items. Significant level of the statistical hypothesis test was set at p> = .05.

Criterion-related validity for estimated ability values using total scores as the validation criterion was analyzed to confirm the validity of the test. Using the scatter diagram plotting test scores and ability values on the test characteristics curve (TCC), we confirmed the goodness-of-fit of the test by sight.

We used the test information function (TIF) and test reliability (TR) to analyze the accuracy of the criterion-referenced tests. The item information function is described in the following formula:

$I_j(\theta) = D^2a_j^2P_jQ_j$

Where "I" is item information amount, "θ" is ability value, "j" is item, "D" is scale component (1.7), "a" is item discrimination parameter, "P" is percentage correct, and "Q" is percentage incorrect (1-P).

TFI is the sum of item information functions (IIF) and described in the following formula:

$T(\theta)= \Sigma I_j(\theta)$

Where "T" is test information amount, "I" is item information amount, "j" is item, and "θ" is ability value.

Reliability coefficient for criterion-referenced test is calculated from the TIF and described in the following formula:

$\rho(\theta) = 1/(1+1/T(\theta))$

Where "ρ" is reliability coefficient, "T" is test information amount, and "θ" is ability value.

In the computerized adaptive test, scores are obtained for each tactical skill. Therefore, we analyzed reliability coefficients for the criterion-referenced test for each tactical skill.

Application software BILOG-MG, ver. 3.0 was used for IRT analysis.

3. Results

3.1 Cause-and-effect analysis by Delphi method

Figure 1 is the fishbone diagram showing the qualitative structure of tactical skills in soccer obtained by cause-and-effect analysis using the Delphi method. Since soccer games consist of offense and defense phases, tactical skills are largely classified into offensive tactical skills and defensive tactical skills, each of which are classified into personal, group, and team tactical skills. Team tactical skills are based on group and personal tactical skills.

Personal offensive tactical skills consisted of on-the-ball and off-the-ball phases. Personal offensive

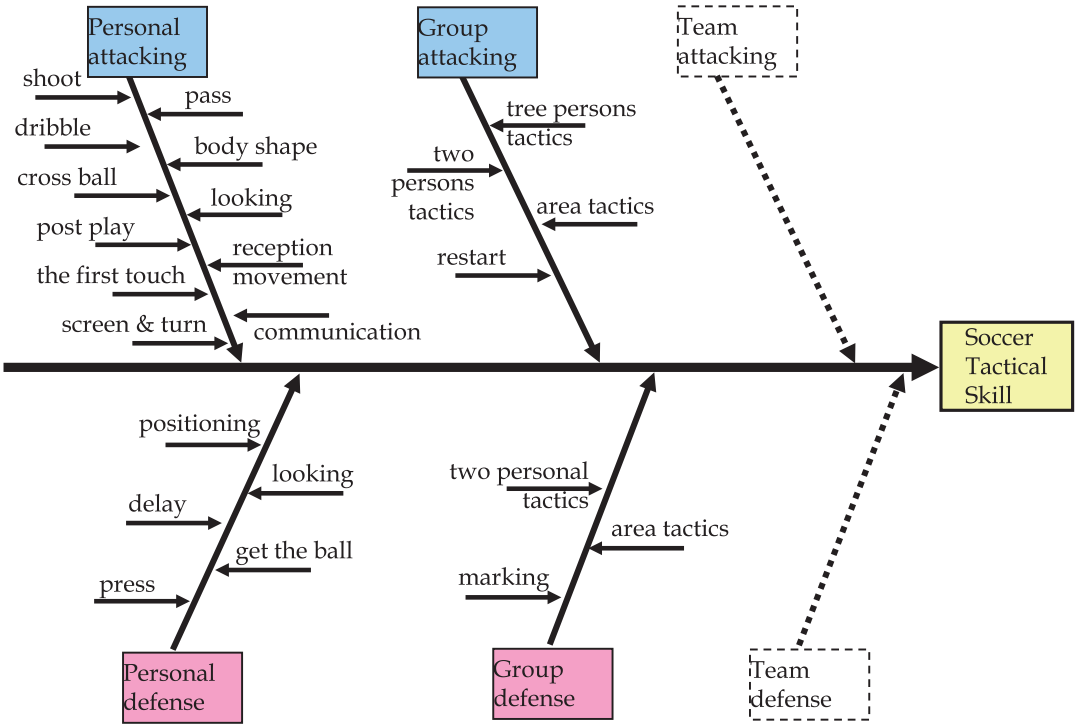


Figure 1 Causal-effect (fish bone) diagram of soccer tactical skill by qualitativeanalysis.

tactical skills in on-the-ball phase were screen & turn, first touch, post play, cross ball, dribbling, shooting, passing, body shape, and looking. Personal offensive tactical skills in off-the-ball phase were communication and receiving the ball. Group offensive tactical skills consisted of two-person tactics, three-person tactics, area tactics, and restarting.

Table 1 shows the definitions of tactical skills in soccer obtained by qualitative analysis. More than

Table 1 Classification and Definition of Tactical Skills in Soccer by Qualitative Analysis

Domain	# Tactical skill	Definition
Personal offense	1 Shooting	The final tactical skill in the offensive phase. It is shooting the ball into the goal by kicking and heading.
	2 Dribbling	A personal offensive tactical skill used to make space for passing and shooting as well as to disengage from opponent defenders and prevent them from participating in the play.
	3 Crossing	The primary purpose of crossing is to pass the ball to the back of the players on the defense side in the prime target area. The length of the prime target area is 7.3 m expanding the 1.8 m of the goal area to the penalty spot, and the width is 18 m, which is equal to the width of the goal area. Crossing means passing with sufficient strength to deny players on the defense side the opportunity to make adjustments. Therefore, accuracy is important to match the timing of the player coming toward the ball.
	4 Post play	Passing the ball to a target player in the front line to start an attack. The target player passes the ball with just a single touch or retains the ball to attract the opponent players and then make the next move in the space created.
	5 The first touch	This first touch is not simply stopping the ball, but is associated with a judgment regarding the next play. To create an opportunity to attack under pressure, it is essential to have the most appropriate first touch in the offense phase. It requires a view of the surrounding situation until the moment the ball is touched. Players capable of using the first touch to threaten an opponent have the tactical skill to wait to ascertain the opponent's situation and exercise good judgment in selecting the best play at the last moment.
	6 Screen & turn	Screen is a personal offensive tactical skill used when the player is marked by the opponent. It is used to block opponents and prevent them from taking possession of the ball by inserting the body between the opponent and the ball. It can be applied to a wide range of situations such as shooting while handling a loose ball or blocking the opponent. Employing the screen alone does not advance the game. The player who possesses the ball applies the turn using a supporting player to create a new phase. The turn skill is used differently when facing the defender (DF) from the front and from the back. When an attacker receives the ball, it is always necessary to be aware of the space to turn. The turn is extremely important in attacking. Losing a chance to execute the turn means to delay the opportunity to attack, it allows the opponent to recover, and it also means losing the chance to send other attackers to attacking positions in the front. When a screening attacker recognizes sufficient space, the attacker can increase the speed and power of the next attack by turning, which increases confusion among defenders.
	7 Passing	An important tactical skill for attacking. High priority passes in offensive tactical skills are, in order of importance, (1) passing to the space behind a defender, (2) passing the ball to the foot of the attacker in the front line, (3) passing beyond at least one defender, (4) cross passing to change the direction of attack, and (5) passing the ball to supporting players in the rear. An important factor in passing is judging what action the receiver will be able to take. Therefore, the tactically important factor in passing is the quality of the pass. The quality of the pass generally means the timing, direction, and weight of the ball.
	8 Body shape	The concept of integrating body direction and posture. Regardless of whether or not the ball is possessed, it is an important tactical skill to maintain effective eyesight to perform creative plays.
	9 Looking	Looking is important in judging the priority of plays by confirming the location of the ball, goal, locations of team and opponent players, and space as well as ensuring effective eyesight utilizing good body shape.
	10 Receiving	Receiving the ball in the off–the–ball phase include communication with the player in possession of the ball, checking, pulling away, waving, diagonal running, support, disappearing from eyesight, appearing from the outside of the DF's eyesight, and running toward the back of DF line.
	11 Communication	A tactical skill between the player who possesses the ball and the player who will receive the ball utilizing eye contact, voice instruction, and body language.
Group offense	12 Two– person tactics	Combination plays by two players for penetration including movement that attracts the opponent, through pass, wall pass, cross over, and overlap.
	13 Restarting	A group tactical skill used when shifting to the attacking side after out–of–play and foul. It includes goal kick, corner kick, free kick and throw in.
	14 Three– person tactics	Combination plays by two players for penetration. It includes the third–person's movement and over.
	15 Area tactics	These are group tactical skills used to aggressively penetrate the space.
Personal defense	16 Pressing	Pressing is to reduce play time and space of the opponent possessing the ball, including slowing the opponent's attack, preventing the opponent's next play, preventing the opponent from turning toward the attacking direction, and reducing the space with the opponent.
	17 Delaying	Preventing the ball from moving toward the attacking direction, and narrowing and limiting the attacking direction.
	18 Positioning	Maintaining the distance from the opponent to prevent the ability to come behind.
	19 Stealing	Taking the ball from an opponent in the defense zone, at the moment the opponent tries to play the ball, or before the ball is passed.
	20 Looking	Keeping the opponent and the ball in view. As with the personal offensive tactical skill, looking ensures effective sight using good body shape, and the ability to check the ball, opponent, team members, and space. It is very important in judging priority of movement.
Group defense	21 Marking	Marking the opponent and defense zone.
	22 Two–person tactics	Combination plays by two persons for pressing and covering including attempting & covering, and collective defense (xxxxx).
	23 Area tactics	Group tactical skills for the aggressive defense of space. These include blocking, chain reaction, balance, retreat, small field, line control, and compact.

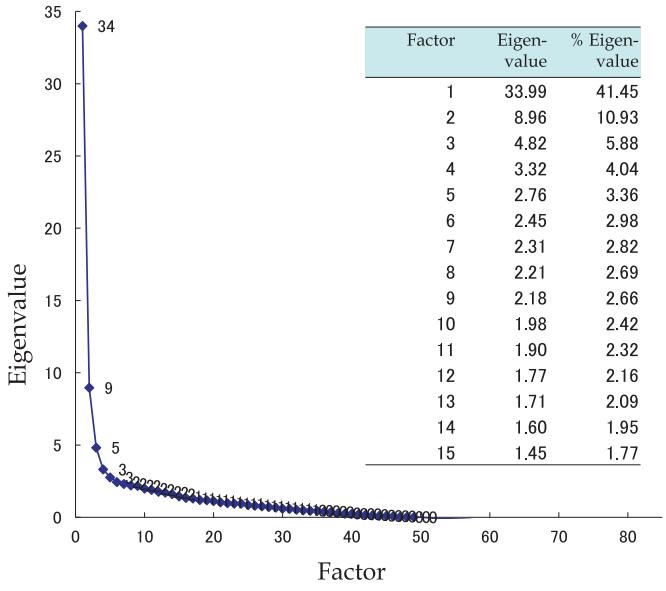
one tactical item was selected for each tactical skill. There were 34 items in personal offensive tactical skills, 22 items in group offensive tactical skills, 11 items in personal defensive tactical skills, and 15 items in group defensive tactical skills.

3.2 Single factor analysis

Figure 2 shows the scree plot of eigenvalues obtained from single factor analysis. The first eigenvalue was 33.99, and the contribution rate to total variance was 41.45%. The contribution rate of the second and subsequent eigenvalues was 11% or less, which revealed a significant difference from the contribution rate of the first eigenvalue. Scree plot showed an L-shape. First factor loading of all items was positive. Mean value and standard deviation was 0.62±0.16, maximum value was 0.99 and minimum value was 0.18 (**Table 2**).

3.3 Item response theory (IRT) analysis

Table 2 and **3** show the level of achievement of tactical skills in soccer, first factor loading, correlation coefficient with total scores, item difficulty, item discrimination, significance of chi-square goodness-of-fit tests, and reliability coefficients. Chi-square values of all 82 items were not significant (p>=0.05). Mean value and standard deviation of level of achievement of all 82 items was 66.8±15.0%, maximum value was 90.1%, and minimum value



Factor	Eigen-value	% Eigen-value
1	33.99	41.45
2	8.96	10.93
3	4.82	5.88
4	3.32	4.04
5	2.76	3.36
6	2.45	2.98
7	2.31	2.82
8	2.21	2.69
9	2.18	2.66
10	1.98	2.42
11	1.90	2.32
12	1.77	2.16
13	1.71	2.09
14	1.60	1.95
15	1.45	1.77

Figure 2 Scree plot of eigenvalue of soccer tactical skill items.

was 11.3%. Mean value and standard deviation of item difficulty was -0.70±0/77, maximum value was 2.35, and minimum value was -2.94. Mean value and standard deviation of item discrimination was 0.86±0.25, maximum value was 1.72, and minimum value was 0.33.

Pearson's correlation coefficient estimated from two randomly selected samples indicated that invariance of difficulty was 0.84, and correlation coefficient indicated that invariance of discrimination was 0.51 ($p<0.05$). Pearson's correlation coefficient estimated from two randomly selected items indicated that the invariance of ability values was 0.90 ($p<0.05$).

Figure 3 shows the scatter diagram of TCC for all 86 items, total scores, and ability values. Goodness-of-fit of data to the characteristic curve was determined to be favorable. Pearson's correlation coefficient indicated that the criterion-related validity coefficient of ability value using the total scores as the validation criterion was 0.96 ($p<0.05$).

Figure 4 shows the reliability coefficients (TR) and information amount (TI) of the test for tactical skills in soccer composed of 82 items. In regard to TR, ability value was -0.8 and the maximum value was 41.3. In regard to TR, the maximum value was 0.98 between -0.5 and -1.0 of the ability values; and it was 0.80 or greater between -3.0 and 2.0 of the ability value.

Figure 5 shows the TR and TI of the test composed of offensive and defensive tactical skills. In regard to TI of 56 offensive tactical skills, the maximum value was 28.5 at -0.5 of the ability value. In regard to TR, the maximum value was 0.98 between -0.5 and -1.0 of the ability values; and it was 0.80 or greater between -0.3 and 2.0 of the ability values. In regard to TI of 26 defensive tactical skills, the maximum value was 12.8 at -1.0 of the ability value. In regard to TR, the maximum value was 0.93 at -1.0 of the ability value; and it was 0.80 or greater between -2.5 and 0.5 of the ability values.

Figure 6 shows TR and TI of the test composed of personal and group offensive tactical skills. In regard to TI of 34 personal tactical skills, the maximum value was 16.0 at -0.5 of the ability value. In regard to TR, the maximum value was 0.94 at -0.5 of the ability value, and it was 0.80 or greater between -2.5 and 1.0 of the ability values. In regard to TI of 22 group offensive tactical skills, the maximum value was 12.7 at -1.0 of the ability value. In regard to TR, the maximum value was 0.93 between -0.5 and -1.0 of

Table 2 Item characteristics of soccer tactical skill test (attaking).

domain	skill	#) item	reliability (a	right rate %	correlation with test score (b	F1 loading (c	difficulty	discrimination	goodness of fit P>=.05
personal attacking skill	the first touch	1) the first touch carrying a ball	0.64	84	0.78	0.73	-1.43	0.94	NS
		2) the first touch fending defense	0.71	70	0.71	0.69	-0.68	1.01	NS
		3) the first touch shoot	0.70	67	0.56	0.55	-0.78	0.67	NS
	screen & turn	4) screen play	0.82	49	0.57	0.59	0.06	0.80	NS
		5) turn with free	0.74	84	0.47	0.46	-2.02	0.56	NS
		6) front turn with defender	0.83	33	0.52	0.53	0.67	0.84	NS
	post play	7) post-play	0.72	45	0.34	0.35	0.28	0.45	NS
	pass	8) direct play pass	0.69	67	0.48	0.46	-0.90	0.55	NS
		9) pass switching the attacking direction	0.70	70	0.69	0.66	-0.74	0.87	NS
		10) timing pass	0.71	62	0.72	0.70	-0.37	1.05	NS
		11) directing pass	0.60	66	0.71	0.67	-0.56	0.94	NS
		12) grading pass	0.60	61	0.63	0.61	-0.40	0.84	NS
		13) dummy pass	0.81	45	0.52	0.56	0.24	0.74	NS
	cross ball	14) corrective cross	0.78	46	0.50	0.52	0.19	0.65	NS
	dribble	15) dribble switching direction	0.76	49	0.44	0.47	0.06	0.62	NS
		16) dribble changing speed	0.81	65	0.43	0.44	-0.77	0.53	NS
		17) dribble with defender	0.78	63	0.72	0.71	-0.43	1.03	NS
		18) dribble penetrating defense line	0.83	38	0.40	0.41	0.62	0.58	NS
		19) dribble penetrating mark-defender	0.76	45	0.59	0.61	0.19	0.84	NS
	shoot	20) direct shoot	0.77	60	0.72	0.71	-0.31	1.10	NS
		21) first control shoot	0.82	77	0.63	0.61	-1.25	0.74	NS
		22) dribble shoot	0.77	65	0.69	0.68	-0.53	0.97	NS
		23) shoot from cross	0.86	55	0.57	0.58	-0.17	0.75	NS
	body-shape	24) body shape	0.78	82	0.81	0.76	-1.28	1.03	NS
	looking	25) look-up	0.67	74	0.66	0.63	-0.97	0.82	NS
	communication	26) communication with ball holder	0.69	87	0.73	0.68	-1.72	0.86	NS
	reception movement	27) checking run	0.90	66	0.58	0.59	-0.69	0.70	NS
		28) pull away	0.84	82	0.72	0.69	-1.39	0.88	NS
		29) waving	0.83	77	0.77	0.74	-1.01	1.03	NS
		30) diagonal runs	0.88	72	0.72	0.71	-0.79	1.02	NS
		31) support	0.66	90	0.68	0.64	-2.19	0.76	NS
		32) pull away from the marker	0.75	60	0.73	0.73	-0.28	1.15	NS
		33) run without the visual field of the mark defender	0.79	69	0.84	0.80	-0.54	1.45	NS
		34) run to the behind defense line	0.78	73	0.68	0.66	-0.91	0.86	NS
group attacking skill	two persons tactics	35) dummy action	0.75	66	0.76	0.73	-0.49	1.21	NS
		36) through-pass(Killer pass)	0.86	60	0.68	0.67	-0.35	0.92	NS
		37) double pass	0.88	81	1.00	0.93	-0.95	1.72	NS
		38) cross-over	0.87	74	0.77	0.74	-0.82	1.11	NS
		39) overlap	0.89	75	0.82	0.79	-0.87	1.14	NS
	three persons tactics	40) Third man Running	0.75	76	0.85	0.81	-0.83	1.33	NS
		41) over	0.75	53	0.59	0.63	-0.07	0.88	NS
	area tactics	42) side change	0.85	55	0.45	0.43	-0.25	0.57	NS
		43) side attack	0.77	75	0.80	0.76	-0.87	1.14	NS
		44) early cross	0.86	64	0.68	0.66	-0.49	0.93	NS
		45) pull back	0.74	55	0.61	0.64	-0.15	0.88	NS
		46) counter-attack	0.74	76	0.67	0.62	-1.12	0.78	NS
		47) direct play	0.75	70	0.70	0.66	-0.77	0.88	NS
		48) possession play	0.71	60	0.66	0.66	-0.30	1.05	NS
	area tactics	49) restart quickly	0.76	89	0.35	0.36	-2.94	0.46	NS
		50) short corner	0.82	78	0.71	0.69	-1.18	0.84	NS
		51) corner kick	0.87	62	0.64	0.64	-0.47	0.81	NS
	restart	52) direct free kick	0.91	32	0.35	0.35	1.07	0.47	NS
		53) penalty kick	0.91	84	0.76	0.72	-1.55	0.87	NS
		54) attack on vital-area	0.67	55	0.67	0.67	-0.15	1.04	NS
		55) attack of fantasists	0.86	11	0.36	0.38	2.35	0.61	NS
		56) switching defense to attack	0.64	90	0.67	0.63	-2.14	0.79	NS

a: intra-class correlation, b: bi-serial correlation, c: the first factor loading.

Table 3 Item characteristics of soccer tactical skill test (defense).

domain	skill	#) item	reliability (a	right rate %	correlation with test score (b	F1 loading (c	difficulty	discrimination	goodness of fit P>=.05
personal defensive skill	press	57) approach	0.77	84	0.72	0.68	−1.57	0.85	NS
		58) pressing	0.86	87	0.96	0.86	−1.37	1.37	NS
		59) keep attacker from turn	0.81	79	0.75	0.71	−1.17	0.95	NS
	delay	60) delay	0.78	85	0.81	0.74	−1.51	0.96	NS
		61) one side cut	0.81	90	0.78	0.72	−1.94	0.94	NS
	get the ball	62) get ready to get the ball	0.75	74	0.67	0.65	−0.93	0.89	NS
		63) get the ball	0.78	76	0.78	0.73	−0.94	1.05	NS
		64) get good wood on the ball	0.73	59	0.57	0.58	−0.35	0.73	NS
		65) intercept	0.83	73	0.74	0.71	−0.85	0.96	NS
	looking	66) watch the mark and ball at the same time	0.74	81	0.31	0.30	−2.17	0.43	NS
	positioning	67) keep attacker from behind defense	0.84	59	0.55	0.55	−0.39	0.65	NS
group defensive skill	marking	68) man to man	0.83	67	0.62	0.62	−0.73	0.73	NS
		69) zone marking	0.86	67	0.61	0.60	−0.68	0.77	NS
		70) exchange mark	0.82	76	0.71	0.67	−1.03	0.89	NS
		71) defense to cross ball	0.77	55	0.18	0.18	−0.35	0.33	NS
	two persons tactics	72) challenge & cover	0.75	81	0.88	0.80	−1.12	1.17	NS
		73) collective defense	0.85	79	0.85	0.80	−1.08	1.10	NS
	area tactics	74) block defense	0.81	47	0.49	0.52	0.16	0.64	NS
		75) chain reaction	0.86	75	0.63	0.61	−1.02	0.81	NS
		76) balance	0.75	76	0.71	0.67	−1.04	0.88	NS
		77) retreating defense	0.88	49	0.47	0.52	0.06	0.58	NS
		78) small field	0.77	52	0.62	0.65	−0.05	0.84	NS
		79) line control	0.81	56	0.51	0.52	−0.25	0.66	NS
		80) compactness	0.74	75	0.85	0.80	−0.79	1.26	NS
		81) vital area defense	0.73	63	0.63	0.63	−0.51	0.78	NS
		82) switching attack to defense	0.72	75	0.31	0.31	−1.69	0.43	NS
		N	82	82	82	82	82	82	
		Mean	0.78	66.8	0.64	0.62	−0.70	0.86	
		SD	0.07	15.0	0.16	0.14	0.77	0.25	
		Med	0.78	68.1	0.67	0.66	−0.75	0.86	
		Max	0.91	90.1	1.00	0.93	2.35	1.72	
		Min	0.60	11.3	0.18	0.18	−2.94	0.33	

a: intra-class correlation, b: bi-serial correlation, c: the first factor loading.

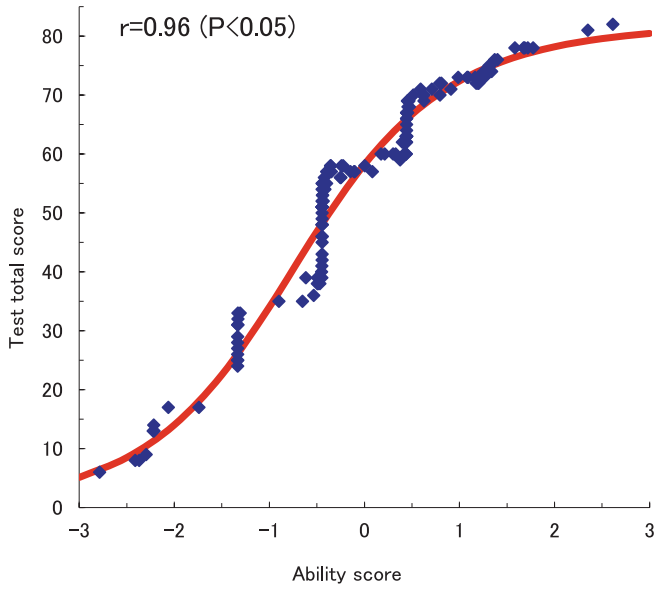


Figure 3 Validity and goodness of fit of soccer tactical skill.

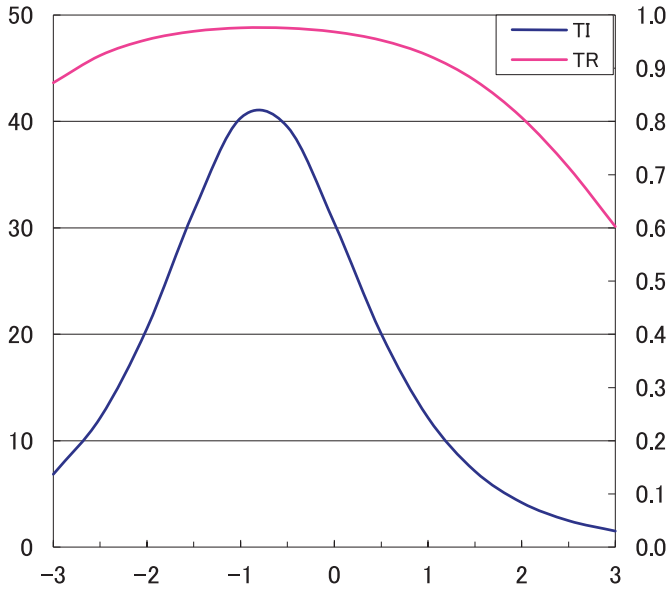


Figure 4 Test reliability (TR) and information (TI) of soccer tactical skill.

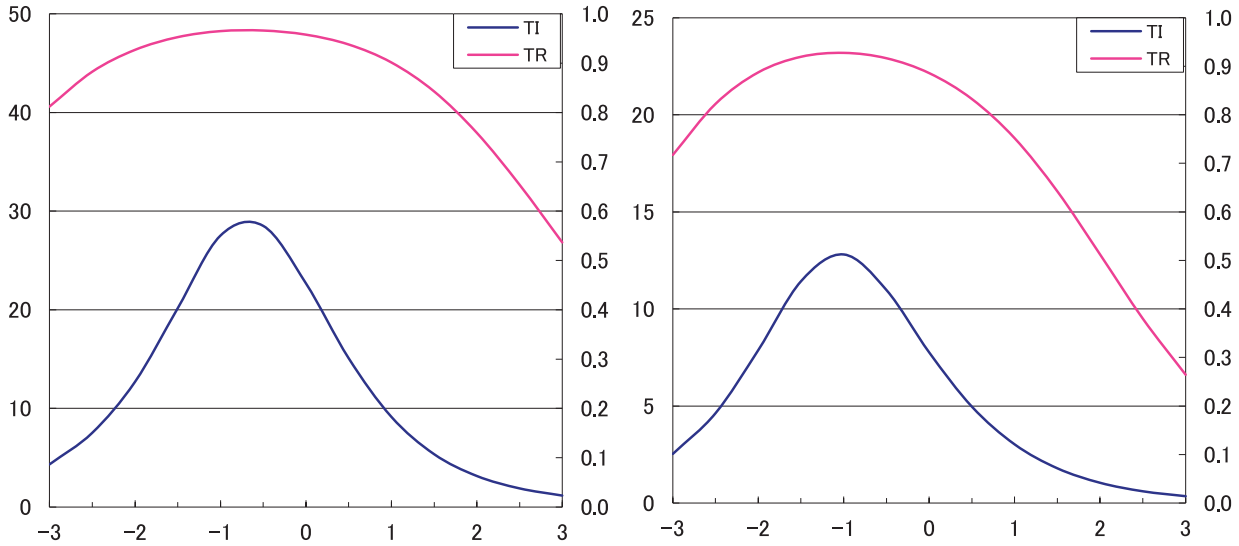


Figure 5 Test reliability (TR) and information (TI) of attacking (left) and defensive (right) tactical skill.

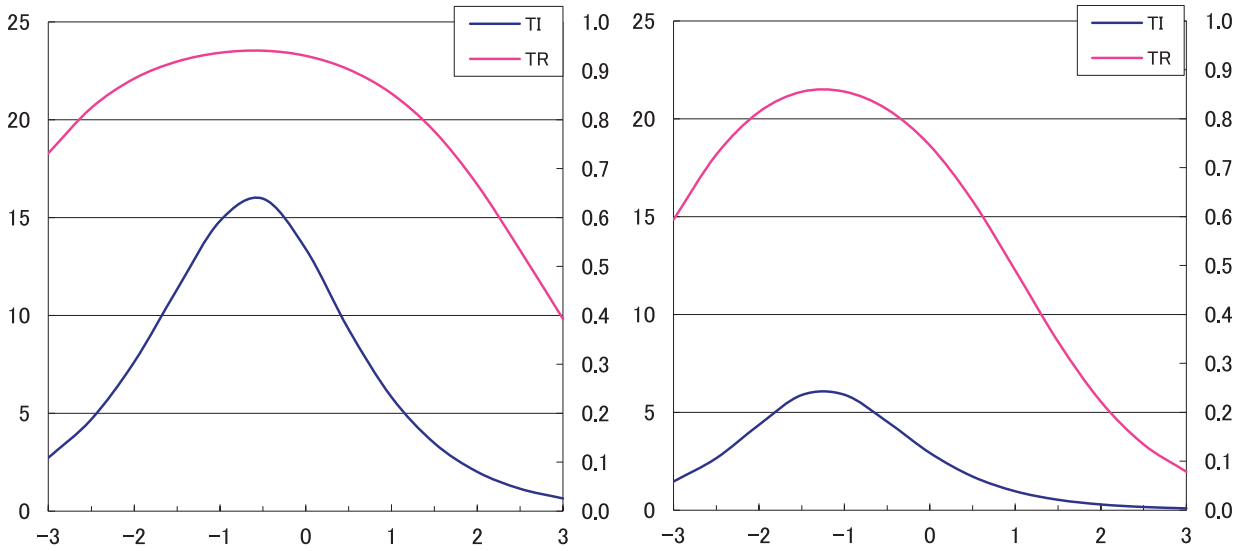


Figure 6 STest reliability (TR) and information (TI) of personal attacking (left) and defensive (right) tactical skill.

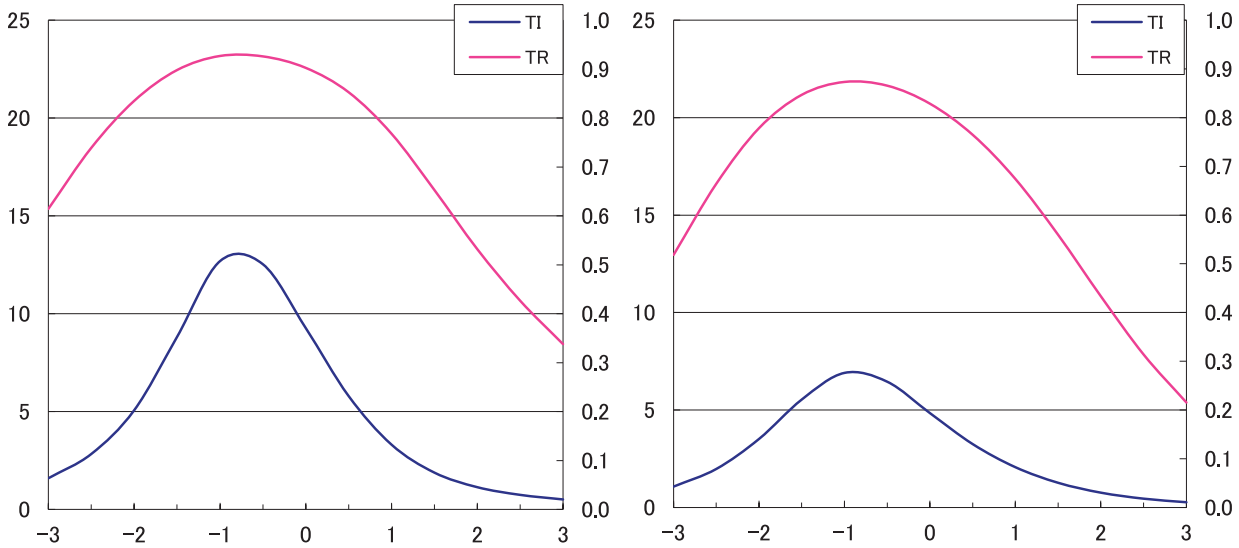


Figure 7 STest reliability (TR) and information (TI) of group attacking (left) and defensive (right) tactical skill. king (left) and defensive (right) tactical skill.

the ability values; and it was 0.80 or greater between -2.0 and 0.5 of the ability values.

Figure 7 shows TR and TI of the test composed of personal and group defensive tactical skills. In regard to TI of 11 personal defensive tactical skills, the maximum value was 5.9 at -1.0 of the ability value. In regard to TR, the maximum value was 0.86 between -1.0 and -1.5 of the ability values; and it was 0.80 or greater between -2.0 and -0.5 of the ability values. In regard to TI of 15 group defensive tactical skills, the maximum value was 6.9 at -1.0 of the ability value. In regard to TR, the maximum value was 0.87 between -0.5 and -1.0 of the ability values; and it was 0.80 or greater between -1.5 and 0.0 of the ability values.

4. Discussion

This study was carried out to clarify the composition of items, the characteristics of criterion-referenced self-administered test and test items to construct an item pool for a computerized adaptive test for tactical skills in soccer.

4.1 Item composition

Qualitative analysis revealed that the tactical skills to be evaluated comprised 34 items in personal offensive tactical skills, 22 in group offensive tactical skills, 11 in personal defensive skills, and 15 group defensive skills for a total of 82 tactical skills (**Table 1**). Since group tactical skills are based on personal tactical skills, it was thought appropriate that the number of personal tactical skills was larger than that of group tactical skills.

Along with the development of soccer, organizational ability during games has improved; and this has made it more difficult for players to break through defense and score. In addition, it is generally thought that players in all positions have improved their comprehensive abilities, which makes it difficult to narrow the gap between the level of individual skills and organizational ability during games. Along with the development of soccer, the change to and from the offensive and defensive phases has become quicker, which limits play time and space. Under such circumstances, each player requires the ability to make judgments on the spot and execute creative plays. The above may have precipitated the classification of test items into shooting skill that determines the win or loss of games, dribbling, passing, and receiving the ball.

Existing studies (González-Víllora et al., 2015; García-López et al., 2013; Memmert, 2010; Kannekens et al., 2009a; 2009b; Elferink-Gemser et al., 2004; Judith et al., 1998) have not analyzed the above-mentioned structures of tactical skills classified into detailed personal and group tactics under an on-and-off-the-ball state in both offensive and defensive phases, and the composition of test items.

Application of the cause-and-effect analysis with Delphi method to structure a scale makes it possible to gather cutting-edge tactical skills that are recognized by soccer specialists, and to construct a tactical-skill structure with logical consistency and test items with content validity. As is shown in the cause-and-effect (Fish bone) diagram in Figure. 1, we constructed a hierarchical structure of tactical skills, which was the composition concept for the evaluation target. In addition, the test items of tactical skill shown in Table 1 were also constructed. Thus, when it is difficult to indicate the qualitative structure of the composition concept for the evaluation target from the studies in existing literature, it may be useful to use apply cause-and-effect analysis with the Delphi method as the first step in structuring a scale for tactical skills.

4.2 Item characteristics

We analyzed item uni-dimensionality and goodness-of-fit, and invariance of item parameters and ability scores utilizing IRT analysis in accordance with the procedures developed by Otomo (1996), Watanabe et al. (1999), Toyoda (2002), Nakano et al. (2004), and Aoyagi (2005) to analyze the item characteristics of the criterion-referenced test for tactical skills in soccer.

In addition to local independence, item uni-dimensionality is an important requirement of IRT and was analyzed using factor analysis results. Factor analysis of 82 items revealed that the contribution rate of the first eigenvalue to total variance was 40%, which was large, and the first factor loading was positive, which confirmed the uni-dimensionality of all 82 items. The mean value of the first factor loading was 0.62, maximum value was 0.93, and minimum value was 0.18. Items with small factor loading were included in the item characteristics analysis after confirming content validity.

Goodness-of-fit test of the items to the ICC with 2PLM revealed that no significance was seen in the chi-square value of all items ($p>=0.05$), which confirmed the goodness-of-fit to the ICC model with 2PLM.

IRT considers invariance between item parameters and ability values an important requirement for construction of items in the criterion-reference test and determines the goodness-of-fit to the mathematical model of 2PLM of the pattern data of the correct answers for each tactical skill question in the test. The goodness-of-fit of the data to the model secured conformity in item parameters regardless of estimations utilizing data obtained from different ability levels and secured conformity in ability values regardless of estimations utilizing any item. Invariance coefficient of item difficulty was significantly high at 0.84 ($p<0.05$), and invariance coefficient of item discrimination was significantly moderate at 0.51 ($p<0.05$).

Comparison between the standard deviations and ranges (maximum value – minimum value) of item difficulty shown in **Tables 2** and **3** revealed a smaller dispersion of the parameters by item discrimination. Pearson's correlation coefficient showed that the correlation between invariances with lower dispersion may become poor. With the goodness-of-fit to each item confirmed by chi-square test, item discrimination was estimated smoothly, as was item difficulty, which suggested that item discrimination of the methods developed in this study was invariable although invariance coefficient (Pearson's correlation coefficient) was not significantly high.

The above results confirmed the invariance of estimated item parameters (difficulty and discrimination). The significant invariance coefficient of ability values (0.90 ($p<0.05$)) confirmed the invariance of ability values.

4.3 Test characteristics

The criterion-related validity coefficient of the ability value estimated by the criterion-reference test for tactical skills in soccer comprising 82 items using the total scores as the validity criterion was significant at 0.96 ($p<0.05$). In addition, TCC favorably matched the scatter diagram of total scores and ability values (**Figure 3**). These results suggested that the criterion-referenced test for tactical skills in soccer comprising 82 items matched the 2PLM and that the test was valid.

We then analyzed the IIF and reliability coefficient of the criterion-referenced test for tactical skills in soccer. In regard to the IIF, maximum value was 41.3 at -0.8 of the ability value. In regard to the test reliability coefficient, the maximum value was 0.98 between -0.5 and -1.0 of the ability values, and was 0.80 or greater between -3.0 and 2.0 of the ability values (**Figure 5**). These results confirmed the high reliability of the criterion-reference test for tactical skills in soccer comprising 83 items.

A computerized adaptive test calculates scores by tactical skill in soccer; therefore, we examined the test reliability coefficient for each skill. The test reliability regarding 56 offensive tactical skills, 26 defensive tactical skills, 34 personal offensive tactical skills, 22 group offensive tactical skills, 11 personal defensive tactical skills, and 15 group defensive tactical skills was high in a broad range of ability values.

These results confirmed local independence, unidimensionality, and goodness-of-fit in 82 items composing the tactical skill test for soccer, invariance in estimated item difficulty, item discrimination, and ability values, which confirmed the test reliability, validity, and goodness-of-fit. The items used in tactical skill test for soccer constructed by qualitative analysis matched ICC and TCC with 2PLM, which revealed the reliability and validity the items in the criterion-referenced test for tactical skills.

5. Conclusion

The purpose of this study was to clarify the items and characteristics of a criterion-referenced self-administered test utilizing video images to construct items for the item pool of a computerized adaptive test for tactical skills in soccer. We applied qualitative analysis to construct tactical skill items, and used IRT to analyze the test and the item characteristics of the criterion-reference test for tactical skills in soccer. The following conclusions were obtained:

1) Tactical skills in soccer were largely classified into offensive and defensive. Each was also classified into personal and group tactical skills. Personal offensive tactical skills include first touch, screen & turn, post play, passing, crossing, dribbling, shooting, body shape, looking, communication, and receiving the ball. Group offensive tactical skills include two-person tactics, three-person tactics, area tactics, and

restarting. Personal defensive tactical skills include pressing, delaying, stealing, looking, and positioning. Group defensive tactical skills include marking, two-person tactics, and area tactics.

2) Tactical skills in soccer include 34 personal offensive tactical skills, 22 group offensive tactical skills, 11 personal defensive tactical skills, and 15 group offensive tactical skills. These 82 skills have the local independence, uni-dimensionality, goodness-of-fit, and invariance of estimated ability score, item difficulty and discrimination. uni-dimensionality, and 2PLM, and invariance with estimated item difficulty, item discrimination, and ability values.

3) The criterion-referenced test for tactical skills in soccer has test reliability, validity, and goodness-of-fit.

Reference

Aoyagi, O. (2005). Measurement of motor ability using item response theory (pp. 1-120). Fukuoka: Touka-shobo (in Japanese).

Elferink-Gemser, M.T., Visscher, C., Richart, H., and Lemmink, K.A.P.M. (2004). Development of the tactical skills inventory for sports. Perceptual and Motor Skills, 99: 883-895.

FIFA (2004). FIFA coaching. (Ch.5) Technical and technical/tactical preparation (pp. 1-29), (Ch.6) Playing style (pp. 1-33).

González-Víllora, S., García-López, L.M., and Contreras-Jordán, O.R. (2015). Decision making and skill development in youth football players. Int J Med Sci Phys Act Sport, 15 (59): 467–487.

García-López, L.M., González-Víllora, S., Diaz-del-Campo, D.G., and Serra-Olivares, J. (2013). Development and validation of the game performance evaluation tool (GPET) in soccer. Revista Euroamericana de Ciencias del Deporte, 2(1): 89–99.

Hosoya, K. (1982). QC seven tools (pp. 18-36). Tokyo: Japan Federation of science and technology, (in Japanese).

Hughes, C. (1999). (ed.) Tsuji, A., and Kyohgoku, S., The winning formula (pp. 1-190). Tokyo: Taishukan-shoten, (in Japanese).

Hughes, M., and Franks, I. (1997). Notational analysis of sport. London: E&FNSPON.

Japan Football Association (2000). Guidelines for teaching. Tokyo: Japan Football Association.

Japan Football Association Teaching Committee (2002). Soccer teaching textbook (pp. 1-88). Tokyo: Japan Football Association.

Japan Football Association Technical Committee (2002). 2002 FIFA World Cup Korea/Japan TM JFA technical report (pp. 31-51). Tokyo: Japan Football Association.

Kannekens, R., Elferink-Gemser, M.T., and Visscher, C. (2009a). Tactical skills of world-class youth soccer teams. Journal of Sports Sciences, 27(8): 807–812.

Kannekens, R., Elferink-Gemser, M.T., Post, W.J., and Visscher, C. (2009b). Self-assessed tactical skills in elite youth soccer player: a longitudinal study. Journal of Sports Sciences, 27(8): 807–812.

Memmert, D. (2010). Testing of tactical performance in youth elite soccer. Journal of Sports Science and Medicine, 9: 199-205.

Nakano, T., Nishijima, T., and Ohtsuka, K. (2004). Feasibility of ADL test construction based on IRT. International J of Health & Sport Sci., 2: 145-155.

Oslin, J.L., Mitchell, S.A., and Griffin, L.L. (1998). The Game Performance Assessment Instrument (GPAI) development and preliminary validation. Journal of Teaching in Physical Education, 17: 231-243.

Otomo, K. (1996). Introduction to Item response theory (pp.1-313). Tokyo: Taishukan-shoten, (in Japanese).

QC technique development division (ed.) (1979). New QC seven tools for admin staff (pp. 19-40). Tokyo: Japan Federation of science and technology, (in Japanese).

Stiehler, G., Konzag, I., and Dobler, H. (ed.) Karaki, K. (1993). Encyclopaedia of ball game coaching (pp. 213-280). Tokyo: Taishukan-shoten, (in Japanese).

Sugiyama, T. (1988). Causal-effect diagram. (ed.) Fujita, K. Ready to use QC technique: problem solution with QC seven tools (pp. 26-39). Tokyo: Japan Federation of science and technology, (in Japanese).

Toyoda, H. (2002). Item response theory [case series] (pp. 1-181). Tokyo: Asakura-shoten, (in Japanese).

Watanabe, N., and Noguchi, H. (1999). Organizational psychology measurement theory (pp. 1-348). Tokyo: Hakuto-shobo, (in Japanese).

Acknowledgement

This work was supported by JSPS KAKENHI (Grant-in-Aid for Challenging Exploratory Research) Grant Number JP15K12629, and JSPS KAKENHI (Grant-in-Aid for Scientific Research (A)) Grant Number JP 23240093.

Name:
Kozue Ando

Affiliation:
Faculty of Health & Sport Sciences,
University of Tsukuba

Address:
Faculty of Health & Sport Science, University of Tsukuba,
1-1-1 Tennodai, Tsukuba, 305-8574 Japan

Brief Biographical History:
2018- Scientist, Faculty of Health & Sport Science,
University of Tsukuba

Main Works:
- KOZUE Method to complete in the World, Kodansha,2012 (in Japanese)

Membership in Learned Societies:
- Japanese Society of Science and football (JSSF)
- The Japan society of coaching studies
- Japan Society of Physical Education, Health and Sport Sciences