Validity and Reliability of Computerized Adaptive Test of Soccer Tactical Skill

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

The purpose of this study was to analyze test characteristics of computerized adaptive test (CAT) of soccer tactical skill with movie image questions for constructing criterion-referenced measurement test of soccer specific tactical skill based upon item response theory (IRT) and CAT techniques. Samples for statistical analysis were 141 male college soccer players, age (20.2±0.99 yr), athletic career (11.9±2.16 yr) whose positions were divided into goal keeper (GK), central defenders (CDF), side defenders (SD), defensive midfielders (DMF), attacking and side midfielders (AMF), forward and attackers (FW). The attacking tactical skill score was high in attacking positions of AMF and FW, and the defensive tactical skill score was high in defensive positions of CDF, SDF and DMF. The criterion-related validity of CAT of soccer tactical skill to criterion-referenced measurement test consisted of 82 items, and the test-retest reliability of CAT of soccer tactical skill were high and significant statistically. It is concluded that CAT of soccer tactical skill have testability, reliability and validity.

1. Introduction

While tactical skills in soccer are employed in play and reflected in game performance, there are as yet no standardized tests for the easy measurement of these skills; and because coaches and managers must evaluate tactical plays in games, it is essential to consider criterion-referenced measurement test from a multivariable perspective to develop the most efficacious methods of measurement possible.

Performance test has been employed as a measurement item in studies on specific motor skills; however, this yields low validity for performance in soccer games, which reduces its effectiveness as a test tool. Since 1980, sports skills have been evaluated by game performance rather than test one (Hughes, 2003), while game performance is evaluated for statistics and analysis. The enumeration data index is used to evaluate team performance in games and compare performance among teams in competition; however, this limits us to descriptive evaluations, which is inadequate for the measurement of soccer skills (Suzuki et al., 2006).

The Technical Report (Japan Football Association Technical Committee, 2000) included a similar method for the qualitative measurement of group skills through visual inspection (Stiehler, 1993). Although this method enables the evaluation of performance from a multidimensional perspective, it is impossible to eliminate subjectivity and arbitrariness in the analysists because methods for the evaluation of tactical game performance of individuals, small groups, and teams in both offense and defense sides have not yet been developed.

Memmert et al. (2016) summarized reviews of tactical skills in soccer and noted the lack of previous studies that measure tactical skills. There are no previous studies that apply criterion-referenced measurement test to directly measure personal tactical skills in soccer.

The use of CAT based on the item response theory (IRT) may make it possible to easily conduct continuous measurement of personal tactical skills. CAT allows us to continuously and easily measure the growth of many players within a short period of time. In addition, CAT allows confirmation of compatibility between the tactical training planned by coaches and the growth of tactical skills of individual players. With the application of IRT, which is criterion-referenced measurement test, reference

values for individual items can be analyzed as item characteristic values associated with invariance. This does not require the setup of criteria for achievement in each tactical play, and addresses the difference in base scores.

Meanwhile, computerized technical and tactical skill tests have as yet not been developed in previous soccer researches. The Soccer Instruction Guidebook published by the Japan Football Association Teaching Committee (2000) clearly differentiates personal techniques from tactics, and introduces soccer skills tests to examine the level of personal techniques. While performance tests are effective in measuring technical skills, they are less effective in measuring tactical skills associated with the assessment of game situations. Therefore, tactical skills employed in soccer have been objectively evaluated by visual observation of player movement during games (Hughes et al., 1997; Stiehler et al., 1993).

Tactics in soccer are performed by individuals, groups, and teams; therefore, written questions may cause misunderstanding of tactical plays. To avoid such misunderstanding, it is helpful to show video questions, which can be provided by CAT.

Conditions for a highly useful criterion-referenced measurement test to evaluate the level of achievement for tactical skills in soccer are listed below:
(1) Use of an item pool that analyzes item characteristics based on the item response theory (IRT)
(2) Use of a computerized adaptive test (CAT) method; and
(3) Use of video images for questions regarding each item.

It was needed to construct the criterion-referenced measurement test that can easily measure play performance within a short period of time and with a high degree of accuracy.

The purpose of this study was to analyze test characteristics of computerized adaptive test (CAT) of soccer tactical skill with movie image questions for constructing criterion-referenced measurement test of soccer specific tactical skill based upon item response theory (IRT) and CAT techniques. We also analyzed the testability, reliability and validity of CAT.

2. Method

2.1 Procedures

We analyzed the validity of CAT for criterion-referenced measurement test for tactical skills in soccer by:
(1) Creating CAT utilizing questions with video images;
(2) Giving CAT to estimate the ability value (tactical skill scores);
(3) Analyzing the test-retest reliability of CAT;
(4) Analyzing the criterion-related validity of CAT based on the scores of the test used for criterion-referenced measurement test for tactical skills in soccer (estimated ability values) utilizing all items; and
(5) Analyzing the discriminant validity of CAT for characteristics of tactical skills by position.

2.2 Development of a Computerized Adaptive Test (CAT)

Following the methods described by Watanabe et al. (1999), Nakano (2004), and Aoyagi (2005), we created a computerized adaptive test (CAT), we employed data obtained from a criterion-reference test for tactical skills in soccer to develop CAT utilizing video images for questions (**Figure 1**). The data used are listed below:
(1) Question items and answer options;
(2) Video images for items, an image used to start the test, and an image used to finish the test;
(3) Item difficulty, item discrimination, item characteristic curve (ICC), item information function (IIF); and
(4) Initial items, criteria for termination of the test.

We employed the variable item and multi-stage test system and maximum information system for the algorithm of CAT. The variable item and multi-stage test system was used to extract optical items from the item pool whose item difficulty and discrimination parameters are estimated, and then they were showed to respondents. We used the maximum information system to calculate estimated scale values (ability estimates: θ) based on true-false responses to each item at given time points, selected items to show maximum item information based on the estimated scale value among the items in the item pool that have not yet been given, and then showed them to

respondents. We repeated the procedure from the initial to the final item.

We set three initial items for each tactical skill domain to ensure an even difference between each item from lower to higher item difficulty. According to the results of the pre-test, we deemed the termination of the test to be when the amount of information for the rest of the items that have not yet been answered becomes lower than 0.3, which is the termination of the estimate ability value.

CAT was conducted as described below.

(1) Respondents responded to the initial items.
(2) Estimated scale value (θ) was calculated by each respondent's true-false response pattern for initial items. Item information function (IIF) of the rest of the items at the estimated scale value (θ) is calculated.
(3) Then, the items with the largest IIF was presented to respondents.
(4) The true-false response patterns for the items given in (3) were added to the patterns obtained from the initial items to calculate the estimated scale value (θ) again. Information amount of the items that had already been given was calculated. Then, we decided the item to be presented next.
(5) In the case of reaching the criteria for the termination of the test, we moved to (6). In the case of not reaching the criteria for the termination of the test, we returned to procedure (3).
(6) Necessary information was recorded and retained before finishing the test.
(7) During the test, all the information including the number of items given, IIF, answers, true and false responses, estimated scale value after presenting each item, and test information amount by the end of each item were recorded and retained.
(8) At termination of the test, number of given items, final estimated scale value, and test information amount for given items were recorded and retained.

2.3 Structure of the Item Pool

We followed the tactical skill classification in the coaching guidelines provided by JFA (2002). We also referred to FIFA Coaching (FIFA, 2004) for terminology and concept. Cause-and-effect analysis using the Delphi method was applied to four soccer specialists to construct a hierarchical scheme of domains, tactics, and items of tactical skills in soccer.
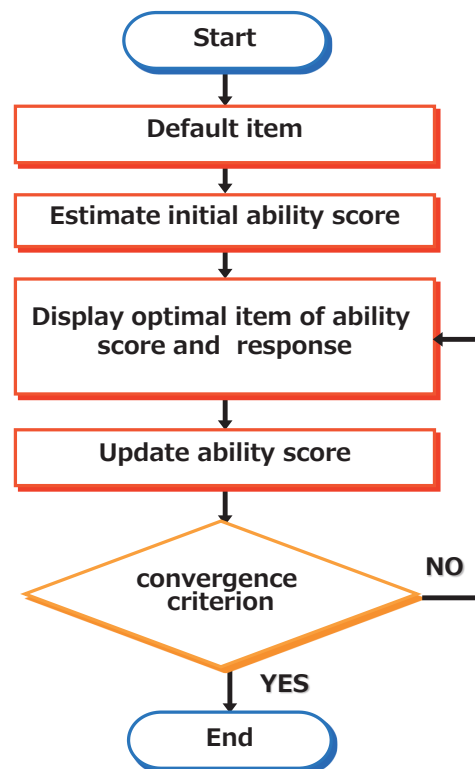

Figure 1 Algorithm of soccer tactical skill-CAT

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, a technical assistant coach of Japan National Team, and a soccer researcher with experience coaching the Universidad Japan National Team.

According to the coaching guidelines provided by JFA (2002), tactical skills in soccer are classified into offense and defense, each of which is classified into personal, group, and team tactics. The tactical skill test for soccer is employed to measure personal skills; therefore, the test focuses on four domains: personal offense, group offense, personal defense, and group defense excluding team tactical skills.

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

defensive tactical skills were marking, two-person tactics, and area tactics.

More than one measurement item was selected for each tactical skill. Personal offensive tactical skills comprised 34 items. Group offensive tactical skills were comprised 22 items. Personal defensive tactical skills were comprised 11 items. Group defensive tactical skills comprised 15 items.

We confirmed the local independence of each item to edit the questions with video images. Video images were plays in the offensive or defensive phases. Tactical plays were highlighted for identification utilizing arrows and circles. We also used slow-motion and stop-motion replay. Each video image was edited to approximately 30 seconds and the total time for all video images was approximately 41 minutes (2,460 seconds).

To analyze item characteristic values used for CAT, tactical skill tests utilizing video images for all items in the item pool were conducted twice with a 10-minute interval in between utilizing a group survey method. Video images were shown on a large screen. Responses are shown in **Figure 1**.

Following the procedures described by Otomo (1996) and Toyoda (2002), we conducted item response theory (IRT) analysis with the 2-parameter logistic model (2PLM). We confirmed the uni-dimensionality, goodness-of-fit to 2PLM, estimated item difficulty, item discrimination, and invariance of ability values, and calculated the ICC, IIF, item standard errors, test characteristic curve (TCC), test information function, test standard errors, and test reliability coefficients (**Table 2** and **3**). Item difficulty and discrimination of all 82 items satisfied the criteria.

2.4 Samples

We used data from 141 male college soccer players. **Table 1** shows the number of samples by position in accordance with the classification of FIFA (2004). We explained to all subjects that positions used in this study were Goalkeeper (#1), Left and right-side defenders (#2+#3), Central defenders (#4+#5), Defensive midfielder (#6), Left and right-side and attacking midfielders (#7+#8), Forward and attackers (#9+#11), Strategist (or Trequartista) (#10) in accordance with the 4-1-3-2 formation system and in reference to the classification of positions specified by the FIFA Coaching (2004).

Subject age was 20.2±0.99 years, athletic career was 11.9±2.16 years, height was 174.1±6.26 cm, and weight was 66.9±5.77 kg. We explained the purpose and content of this study in detail to all subjects and obtained individual consent for participate. This study was approved by the Research Ethical Committee, Graduate School of Human Comprehensive Science, University of Tsukuba (Project No. 22-364).

2.5 Measurement method

The computerized adaptive test for tactical skills in soccer (TSS-CAT) utilizing questions with video images was given to subjects twice with a 10-minute interval in between. Subjects input their position and other attributes in the initial window. Video images were replayed in the question window until respondents selected an answer. An option button was placed in the lower section of the window. There were seven options; namely, a five-point scale for criterion-referenced measurement test for tactical plays and two-point scale for tactical skill recognition. After subjects selected one answer, the window changed to the next question. Subjects chose one from the five-point scale options for the criterion-referenced measurement test for tactical plays in recent soccer games. We measured the achievement of tactical plays that the subjects evaluated from the five-point scale options selected. In the preliminary survey, we confirmed that responses differed by subject, the level of performance influenced self-evaluation, and that the standard achievement response was indicated to address these. 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.

2.6 Statistical analysis

To analyze the convenience of CAT, we calculated the mean value, standard deviation, median value,

Table 1 Item characteristics of soccer tactical skill test (attacking)

domain	skill	#) item	right rate %	F1 loading	difficulty	SE difficulty	discrimination	SE discrimination	Max information	Goodness of fit P>=.05
personal attacking	the first touch	01) the first touch carrying a ball	84	0.73	−1.43	0.23	0.94	0.17	0.64	NS
		02) the first touch　fending defense	70	0.69	−0.68	0.16	1.01	0.20	0.72	NS
		03) the first touch　shoot	67	0.55	−0.78	0.22	0.67	0.14	0.32	NS
	screen&turn	04) screen play	49	0.59	0.06	0.14	0.80	0.14	0.46	NS
		05) turn with free	84	0.46	−2.02	0.41	0.56	0.12	0.22	NS
		06) front turn　with defender	33	0.53	0.67	0.17	0.84	0.16	0.50	NS
	post play	07) post−play	45	0.35	0.28	0.24	0.45	0.09	0.14	NS
	pass	08) direct play pass	67	0.46	−0.90	0.23	0.55	0.10	0.22	NS
		09) pass switching the attacking direction	70	0.66	−0.74	0.17	0.87	0.17	0.53	NS
		10) timing pass	62	0.70	−0.37	0.13	1.05	0.19	0.79	NS
		11) directing pass	66	0.67	−0.56	0.13	0.94	0.16	0.64	NS
		12) grading pass	61	0.61	−0.40	0.14	0.84	0.15	0.50	NS
		13) dummy pass	45	0.56	0.24	0.15	0.74	0.16	0.38	NS
	cross ball	14) corrective cross	46	0.52	0.19	0.17	0.65	0.14	0.30	NS
	dribble	15) dribble switching direction	49	0.47	0.06	0.18	0.62	0.14	0.28	NS
		16) dribble changing speed	65	0.44	−0.77	0.24	0.53	0.11	0.20	NS
		17) dribble　with defender	63	0.71	−0.43	0.14	1.03	0.22	0.76	NS
		18) dribble penetrating defense line	38	0.41	0.62	0.22	0.58	0.13	0.24	NS
		19) dribble penetrating　mark−defender	45	0.61	0.19	0.14	0.84	0.18	0.50	NS
	shoot	20) direct shoot	60	0.71	−0.31	0.12	1.10	0.23	0.85	NS
		21) first control shoot	77	0.61	−1.25	0.25	0.74	0.14	0.38	NS
		22) dribble shoot	65	0.68	−0.53	0.15	0.97	0.20	0.68	NS
		23) shoot from cross	55	0.58	−0.17	0.15	0.75	0.14	0.40	NS
	body−shape	24) body shape	82	0.76	−1.28	0.20	1.03	0.21	0.74	NS
	looking	25) look−up	74	0.63	−0.97	0.20	0.82	0.16	0.49	NS
	communication	26) communication with ball holder	87	0.68	−1.72	0.25	0.86	0.18	0.52	NS
	reception movement	27) checking run	66	0.59	−0.69	0.21	0.70	0.14	0.35	NS
		28) pull away	82	0.69	−1.39	0.27	0.88	0.17	0.56	NS
		29) waving	77	0.74	−1.01	0.20	1.03	0.20	0.77	NS
		30) diagonal runs	72	0.71	−0.79	0.17	1.02	0.22	0.73	NS
		31) support	90	0.64	−2.19	0.36	0.76	0.16	0.41	NS
		32) pull away from the marker	60	0.73	−0.28	0.13	1.15	0.25	0.92	NS
		33) run without the visual field of the mark defender	69	0.80	−0.54	0.11	1.45	0.32	1.51	NS
		34) run to the behind defense line	73	0.66	−0.91	0.21	0.86	0.17	0.53	NS
group attacking	two persons tactics	35) dummy action	66	0.73	−0.49	0.12	1.21	0.25	1.06	NS
		36) through−pass(Killer pass)	60	0.67	−0.35	0.14	0.92	0.16	0.60	NS
		37) double pass	81	0.93	−0.95	0.13	1.72	0.40	2.13	NS
		38) cross−over	74	0.74	−0.82	0.15	1.11	0.22	0.86	NS
		39) overlap	75	0.79	−0.87	0.17	1.14	0.20	0.92	NS
	three persons tactics	40) Third man Running	76	0.81	−0.83	0.17	1.33	0.27	1.23	NS
		41) over	53	0.63	−0.07	0.13	0.88	0.18	0.56	NS
	area tactics	42) side change	55	0.43	−0.25	0.19	0.57	0.12	0.23	NS
		43) side attack	75	0.76	−0.87	0.16	1.14	0.24	0.92	NS
		44) early cross	64	0.66	−0.49	0.15	0.93	0.19	0.62	NS
		45) pull back	55	0.64	−0.15	0.14	0.88	0.18	0.56	NS
		46) counter−attack	76	0.62	−1.12	0.21	0.78	0.17	0.44	NS
		47) direct play	70	0.66	−0.77	0.15	0.88	0.15	0.55	NS
		48) possession play	60	0.66	−0.30	0.12	1.05	0.22	0.77	NS
	area tactics	49) restart　quickly	89	0.36	−2.94	0.69	0.46	0.11	0.15	NS
		50) short corner	78	0.69	−1.18	0.24	0.84	0.15	0.50	NS
		51) corner kick	62	0.64	−0.47	0.17	0.81	0.17	0.47	NS
	restart	52) direct free kick	32	0.35	1.07	0.30	0.47	0.10	0.16	NS
		53) penalty kick	84	0.72	−1.55	0.27	0.87	0.16	0.55	NS
		54) attack on　vital−area	55	0.67	−0.15	0.12	1.04	0.20	0.77	NS
		55) attack of fantasist	11	0.38	2.35	0.53	0.61	0.16	0.26	NS
		56) switching defense to attack	90	0.63	−2.14	0.38	0.79	0.18	0.44	NS

Table 2 Item characteristics of soccer tactical skill test (defense)

domain	skill	#) item	right rate %	F1 loading	difficulty	SE difficulty	discrimination	SE discrimination	Max information	Goodness of fit P>=.05
personal defense	press	67) approach	84	0.68	−1.57	0.26	0.85	0.17	0.52	NS
		57) pressing	87	0.86	−1.37	0.24	1.37	0.35	1.33	NS
		58) keep attacker from turn	79	0.71	−1.17	0.22	0.95	0.20	0.65	NS
	delay	59) delay	85	0.74	−1.51	0.24	0.96	0.22	0.67	NS
		60) one side cut	90	0.72	−1.94	0.31	0.94	0.21	0.63	NS
	get the ball	61) get ready to get the ball	74	0.65	−0.93	0.19	0.89	0.17	0.57	NS
		62) get the ball	76	0.73	−0.94	0.18	1.05	0.25	0.80	NS
		63) get good wood on the ball	59	0.58	−0.35	0.17	0.73	0.15	0.38	NS
		64) intercept	73	0.71	−0.85	0.19	0.96	0.22	0.66	NS
	looking	65) watch the mark and ball at the same time	81	0.30	−2.17	0.52	0.43	0.11	0.14	NS
	positioning	66) keep attacker from behind defense	59	0.55	−0.39	0.17	0.65	0.12	0.30	NS
group defense	marking	68) man to man	67	0.62	−0.73	0.21	0.73	0.15	0.38	NS
		69) zone marking	67	0.60	−0.68	0.19	0.77	0.16	0.42	NS
		70) exchange mark	76	0.67	−1.03	0.20	0.89	0.17	0.57	NS
		71) defense to cross ball	55	0.18	−0.35	0.31	0.33	0.08	0.08	NS
	two persons tactics	72) challenge & cover	81	0.80	−1.12	0.16	1.17	0.24	0.98	NS
		73) collective defense	79	0.80	−1.08	0.18	1.10	0.21	0.87	NS
	area tactics	74) block defense	47	0.52	0.16	0.18	0.64	0.12	0.29	NS
		75) chain reaction	75	0.61	−1.02	0.21	0.81	0.18	0.47	NS
		76) balance	76	0.67	−1.04	0.19	0.88	0.17	0.55	NS
		77) retreating defense	49	0.52	0.06	0.19	0.58	0.12	0.24	NS
		78) small field	52	0.65	−0.05	0.14	0.84	0.18	0.51	NS
		79) line control	56	0.52	−0.25	0.17	0.66	0.14	0.31	NS
		80) compactness	75	0.80	−0.79	0.13	1.26	0.23	1.09	NS
		81) vital area defense	63	0.63	−0.51	0.17	0.78	0.14	0.44	NS
		82) switching attack to defense	75	0.31	−1.69	0.43	0.43	0.10	0.13	NS
		N	82	82	82	82	82	82	82	
		Mean	66.8	0.62	−0.70	0.21	0.86	0.18	0.57	
		SD	15.0	0.14	0.77	0.10	0.25	0.05	0.33	
		Med	68.1	0.66	−0.75	0.18	0.86	0.17	0.53	
		Max	90.1	0.93	2.35	0.69	1.72	0.40	2.13	
		Min	11.3	0.18	−2.94	0.11	0.33	0.08	0.08	

Table 3 Sample size

position	#	N
goalkeeper	1	12
central defenders	4, 5	23
right & left side−defenders	2, 3	21
defensive midfielder	6	15
left & right side and attacking midfielders	7, 8	38
forward & attackers	9, 11	31
the strategist or trequartista (#10)	10	1
total		141

maximum value, and minimum value of the number of response items as a whole and by tactical skill domain. To analyze the reliability of CAT, we used Pearson's correlation coefficient and intra-class correlation coefficient as test-retest reliability coefficients. We used a one-way qualitative response model for the intra-class correlation coefficient. We calculated criterion-related validity coefficients between the ability values estimated from all items in the item pool and the ability values using the total test scores as the criterion of validity as a whole and by tactical skill domain to analyze the criterion-related validity for CAT. We employed one-way ANOVA and Tukey's multiple comparison test as post hoc test to analyze the discriminant validity for CAT and compared ability value means by position. Statistical

significance level was set at p<0.05. IBMSPSS, Ver.23.0 was used for data analysis.

We applied the IRT analysis with 2PLM for CAT. BILOG-MG Ver.3.0 was used as the application software. We also used maximum likelihood estimation method to estimate parameters.

The item information function is described in the following formula:

$I_j(\theta) = D^2 a_j^2 P_j Q_j$

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

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 measurement test is 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.

3. Results

3.1 Number of response items

Table 4 shows the mean, standard deviation, median, maximum, and minimum of the number of response items by domain of tactical skills. Mean and standard deviation of the number of response items in all domains of tactical skills were 47.7±11.8,

median was 49, maximum was 68, and minimum was 27, which was the number set in advance. Mean and standard deviation of the number of response items in offensive skills were 32.7±10.7, median was 34, maximum was 47, and minimum was 15, which was the number set in advance. Mean and standard deviation of the number of response items in defensive skills were 15.03.2, median was 13, maximum was 21, and minimum was 12, which was the number set in advance. The results of other domains are shown in the table.

3.2 Reliability coefficient

Table 5 shows test-retest reliability coefficients of CAT by domain of tactical skills. Reliability coefficients (Pearson's correlation coefficient and intra-class correlation coefficient) of estimated ability values in all domains in the tactical test were 0.91 and 0.89, respectively. Reliability coefficients in offensive skills were 0.90 and 0.88, and those in defensive skills were 0.89 and 0.89. Reliability coefficients in personal offensive skills were 0.86 and 0.85, those in group offensive skills were 0.84 and 0.84, those in personal defensive skills were 0.87 and 0.87, and those in group defensive skills were 0.84 and 0.84.

3.3 Criterion-related validity coefficient

The criterion-related validity coefficient of CAT utilizing test scores as the criterion of validity was 0.97 (p<0.05).

Table 6 shows criterion-related validity coefficients of CAT utilizing scores of test consisting of 82 items as the criterion of validity. Criterion-related validity coefficient of all tactical skills was 0.96, that of the offensive skills was 0.96, that of the defensive skills was 0.87, that of the personal offensive skills was 0.96, that of the group offensive skills was 0.96, that

Table 4 Number of items response of soccer tactical skill-CAT.

domain	N	Mean	SD	Med	Max	Min
total	82	47.7	11.8	49	68	27
attacking	56	32.7	10.7	34	47	15
defense	26	15.0	3.2	13	21	12
personal attacking	34	17.3	6.7	18	27	7
group attacking	22	13.3	4.8	15	19	7
personal defense	11	7.0	1.4	6	10	6
group defense	15	7.9	2.2	6	11	6

Number of samples: 141

Table 5 Reliability of soccer tactical skill-CAT

domain	Pearson correlation	intra-class correlation
total	0.91	0.89
attacking	0.90	0.88
defense	0.89	0.89
personal attacking	0.86	0.85
group attacking	0.84	0.84
personal defense	0.87	0.87
group defense	0.84	0.84

N=141

Table 6 Criterion-related validity of CAT to total test score.

domain	Pearson correlation
total	0.96
attacking	0.96
defense	0.97
personal attacking	0.96
group attacking	0.96
personal defense	0.97
group defense	0.96

N=141

of the personal defensive skills was 0.97, and that of the group defensive skills was 0.96.

3.4 Comparison of tactical skill scores by player position

Table 7 shows one-way ANOVA results for CAT scores by position. Mean values of six positions, goal keeper (GK), center defender (CDF), side defender (SDF), defensive midfielder (DMF), attacking & side midfielder (AMF), and forward (FW) showed significant differences for all items and domains ($p < 0.05$).

Figure 2 shows multiple comparison test results of CAT scores among positions. The target positions were 6 positions of GK, CDF, SDF, DMF, AMF, and FW. As there was a subject for #10 (trequartista), the measurement score was shown instead of the mean value.

Mean values of tactical skills estimated from all items showed a significant difference between GK and other positions ($p < 0.05$). Mean values of tactical skill scores, offensive tactical skill scores, and group offensive tactical skill scored estimated by CAT also showed a significant difference between GK and other positions ($p < 0.05$). Mean values of personal offensive tactical skill scores showed a significant difference

between GK and other positions, and between AMF and CDF/ SDF ($p < 0.05$). Mean values of defensive tactical skill scores, personal and group defensive tactical skill scores showed a significant difference between GK and other positions, between FW and CDF, SDF, and DMF ($p < 0.05$).

4. Discussion

4.1 Convenience of computer adaptive tests

We analyzed the number of response items to examine the convenience of CAT. Mean value and standard deviation of the number of response items in all domains of tactical skills were 47.7±11.8, and the median value was 49. These results showed that 58% of the response items among 82 tactical skill items was sufficient for measurement in CAT. The maximum number of response items was 68 (83%) and the minimum number was 27 (33%), both of which were sufficient for measurement in CAT.

Mean value and standard deviation of the response items in offensive tactical skills were 32.7±10.7, median value was 34, maximum value was 47, and minimum value was 15, which was the number set in advance.

Tactical plays and skills are often evaluated visually by coaches and managers. The results of this study confirmed that CAT utilizing questions with video images of tactical plays is a convenient means of evaluating the level of achievement of tactical skills. It can be used by players, coaches and managers to examine the development of individual tactical skills by training phase.

4.2 Characteristics of a computerized adaptive test

Test-retest reliability coefficient of tactical skill scores estimated by CAT was 0.91. Reliability coefficient of the offensive skills was 0.90 and that of the defensive skills was 0.89. Reliability coefficient of the personal offensive skills was 0.86, that of the group offensive skills was 0.84, that of the personal defensive skills was 0.87, and that of the group defensive skills was 0.84.

Reliability coefficient (intra-class correlation coefficient) of CAT developed by Kramer (2016) to evaluate the level of autism in children was 0.92 for

Table 7 One-way ANOVA of CAT score by positions

Skill domain	Group	Sum SQ	df	Mean SQ	F	P
all items	between	39.39	5	7.88	10.59	0.00
	within	100.40	135	0.74		
	total	139.78	140			
tactical skill	between	39.46	5	7.89	11.97	0.00
	within	89.00	135	0.66		
	total	128.46	140			
attacking	between	42.85	5	8.57	11.99	0.00
tactical skill	within	96.48	135	0.71		
	total	139.33	140			
defensive	between	51.82	5	10.36	13.79	0.00
tactical skill	within	101.48	135	0.75		
	total	153.30	140			
personal	between	38.75	5	7.75	12.53	0.00
attacking	within	83.51	135	0.62		
tactical skill	total	122.25	140			
group	between	30.98	5	6.20	11.29	0.00
attacking	within	74.08	135	0.55		
tactical skill	total	105.06	140			
personal	between	28.77	5	5.75	10.54	0.00
defensive	within	73.68	135	0.55		
tactical skill	total	102.44	140			
group	between	42.48	5	8.50	12.49	0.00
defensive	within	91.86	135	0.68		
tactical skill	total	134.34	140			

daily activities, 0.86 for social/cognitive value, and 0.90 for responsibility. In addition, CAT developed by Dumas (2012) to evaluate the level of disabilities and CAT (a role functioning computerized adaptive test "RF-CAT") developed by Anatchkova (2013) revealed high reliability as well. These results suggest that the criterion-referenced measurement test for tactical skills in soccer utilizing CAT is highly reliable.

CAT is characterized by different questions for individual respondents because it selects items that match the level of the ability of each respondent. Therefore, it is important to confirm criterion-related validity utilizing tactical skill scores estimated from all items in the item pool and test scores as the criterion of validity.

The criterion-related validity coefficient of CAT for tactical skills utilizing the tactical skill scores estimated from all items as criterion of validity was 0.97. Criterion-related validity coefficient utilizing test scores as criterion of validity was 0.96, and criterion-related validity coefficients by tactical skill domain were nearly the same.

These results suggest that criterion-referenced measurement test for tactical skills in soccer utilizing CAT has high validity with tactical skill scores estimated from all items and test scores.

Criterion-referenced measurement test for tactical skills in soccer utilizing CAT can measure play performance with a high degree of accuracy in a short period of time although content and number of response items differ by respondent because items are provided in accordance with the level of ability of each respondent.

We evaluated the effects of positions on tactical skill scores and the difference of mean values among positions to analyze the discriminant validity of CAT. GD, CDF, SDF, DMF, AMF, and FW revealed significant effects on tactical skill scores measured by CAT in all items and tactical skill domains ($p<0.05$).

In regard to the mean values of tactical skill scores of all items, significant difference was observed between GK and other positions ($p<0.05$). Mean values of tactical skill scores estimated by CAT, offensive tactical skill scores, and group offensive tactical skill scores also revealed a significant difference between GK and other positions ($p<0.05$). These results discriminate between highly specialized GK and other field players. In addition, these results also showed that offensive and defensive tactical skill tests are not capable of evaluating the tactical skill of GK.
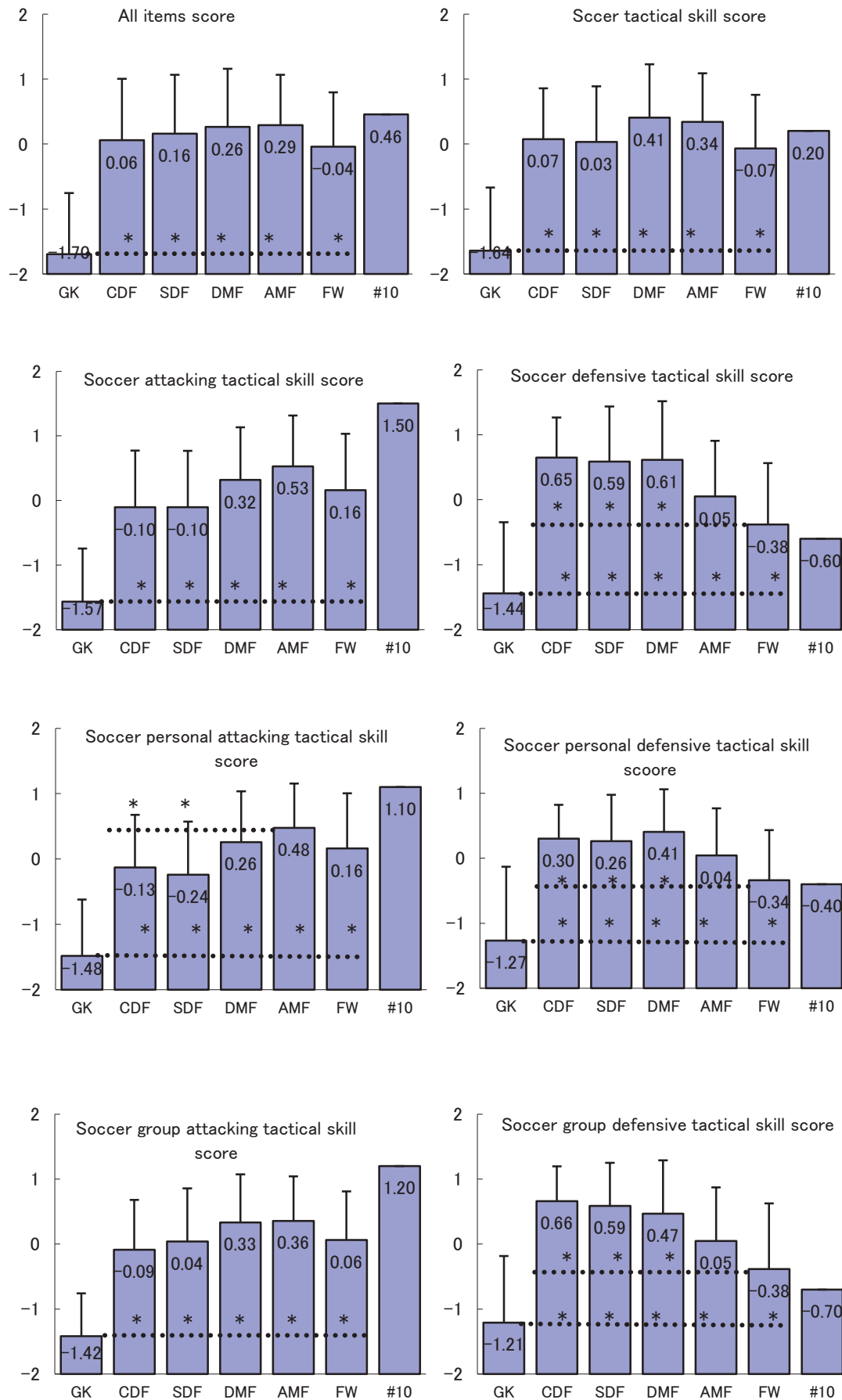
Figure 2 Comparison on CAT score among positions (*:P<0.05)

Mean values of personal offensive tactical skills showed a significant difference between GK and other positions, between AMF, which is an offensive position, and CDF and SDF, which are defensive positions (p<0.05). These results showed that personal offensive tactical skill test items discriminate between tactical skills of offensive positions and defensive positions.

Mean values of defensive tactical skill scores, personal and group defensive skill scores showed a significant difference between GK and other positions, and between FW and CDF, SDF, and DMF (p<0.05). These results showed that personal defensive tactical skill test items discriminate between players in offensive and defensive positions.

DMF showed the highest personal defensive tactical skill scores followed by CDF and SDF. CDF showed the highest group defensive tactical skill scores followed by SDF and DMF. DMF showed higher scores in group defensive tactical skill scores than personal defensive tactical skill scores. The DMF position is characterized by its high group defensive skills because it prioritizes organizational defensive tactics, as is the case with the CDF and SDF positions. As soccer strategy has advanced, DMF players have been required to serve in the CDF position. This characteristic appears to have been clearly discriminated.

There was only one subject who considered himself qualified to serve in the #10 position; however, his tactical skill score was very interesting. Compared with other positions, he rated his personal and group offensive tactical skills higher. On the other hand, his personal and group defensive tactical skills revealed the lowest scores among all positions except GK. These results suggested that the tactical skills displayed by #10, which plays a central role in offense, were clearly discriminated.

Prior to conducting measurements in this study, we explained to all players that midfielder (trequartista) (#10) was not simply a position but was assigned to a creative player capable of instantly determining what play would affect the direction of the game. (FIFA, 2024). Only one player identified himself as midfielder (trequartista) (#10). Because his offensive tactical skill scores were distinctly higher than other samples, fostering players that fit the conditions of #10 is extremely important. On the other hand, the defensive tactical skill scores of #10 were relatively low. Because players are required to have high

defensive skills in modern soccer, the defensive tactical skill of midfielders (trequartista) (#10) needs to be improved.

JFA coaching guidelines (JFA, 2000) sets the fostering of creative players as the primary target. Modern soccer requires proper and instant judgment by players, which means individual players must perform creative plays. The leading soccer countries clearly understand the importance of fostering such creative players and have worked on this from a long-term perspective (Ono, 2000). As a result, they have succeeded in fostering many creative players such as Lionel Messi, Cristiano Ronaldo, and Neymar.

5. Conclusion

In this study, we analyzed the test characteristic of CAT for tactical skills in soccer with video image questions for use in developing a criterion-reference measurement test of tactical skill in soccer based on IRT and CAT techniques, and obtained the following conclusion:

CAT for tactical skills in soccer with video image questions is capable of evaluating performance in a short period of time with testability, reliability, criterion-related validity and discriminant validity. It is highly accurate and appropriate as a criterion-referenced measurement test for tactical skills in soccer.

Reference

Anatchkova, M.D., Rose, M., Ware, J., and Bjorner, J.B. (2013). Evaluation of a role functioning computer adaptive test (RF-CAT). Qual Life Res, 22: 1085–1092.

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

Barthel, D., Fischer, K.I., Nolte, S., Otto, C., XMeyrose, A.K., Reisinger, S., Dabs, M., Thyen, U., Klein, M., Muehlan, H., Ankermann, T., XWalter, O., Rose, M., and Ravens-Sieberer, U. (2016). Implementation of the Kids-CAT in clinical settings: a newly developed computer-adaptive test to facilitate the assessment of patient-reported outcomes of children and adolescents in clinical practice in Germany. Qual., Life Res., 25: 585–594.

Dumas, H.M., Fragala-Pinkham, M.A., Haley, S.M., Ni, P., Coster, W., Kramer, J.M., Kao, Y.C., Moed, R., and Ludlow, L.H. (2012). Computer adaptive test performance in children with and without disabilities: prospective field study of the PEDI-CAT. Disability & Rehabilitation: 34(5): 393–401.

Fédération Internationale de Football Association (FIFA, 2004). FIFA coaching (pp. 1-33).

FIFA (2004). FIFA coaching. (Ch.5) Technical and technical/

tactical preparation (pp. 1-29), (Ch.6) Playing style (pp. 1-33).

Gamper, E.M., Petersen, M.A., Aaronson, N., Costantini, A., Giesinger, J.M., Holzner, B., Kemmler, G., Oberguggenberger, A., Singer, S., Young, T., and Groenvold, M. (2016). Development of an item bank for the EORTC Role Functioning Computer Adaptive Test (EORTC RF-CAT). Health and Quality of Life Outcomes, 14: 72.

Hughes, M. (2003). Notational analysis. In: Reilly, T., and Williams, M. (eds.), science and soccer (2nd ed., pp. 245-264). London: Routledge.

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

Japan Football Association (JFA, 2002). Guidelines for teaching, (in Japanese).

Japan Football Association (JFA, 2000). Guidelines for teaching, (in Japanese).

Japan Football Association Teaching Committee (2000). Soccer teaching textbook (pp. 1-88). Tokyo: Japan Football Association, (in Japanese).

Kramer, J., Liljenquist, K., and Coster, W. (2016). Validity, reliability, and usability of the pediatric evaluation of disability inventory-computer adaptive test for autism spectrum disorders. Developmental Medicine & Child Neurology, 58: 255–261.

Memmert, D., Lemmink, K.A.P.M., and Sampaio, J. (2016). Current approaches to tactical performance analyses in soccer using position data. Sports Medicine, 47(1): 1-10.

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.

Ono, T. (2000). Creative soccer coaching (pp. 1-188). Tokyo: Taishukan-shoten, (in Japanese).

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

Suzuki, K., and Nishijima, T. (2005). Measurement of a soccer defending skill using game performances. In Science and Football V: The Proceedings of the Fifth World Congress on Sports Science and Football (p. 259). Routledge.

Thamsborg, L.H., Petersen, M.A., Aaronson, N., Chie, W.C., Costantini, A., Holzner, B., Leeuw, I.V., Young, T.E., and Groenvold, M. (2015). Development of a lack of appetite item bank for computer-adaptive testing (CAT). Support Care Cancer 23: 1541–1548.

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

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

Appendix 1 Response sheet.

サッカー戦術チェック①

記入年月日：＿＿＿＿＿＿＿＿＿＿＿＿

氏名：＿＿＿＿＿＿＿ 男・女 ID ＿＿＿＿＿＿＿ クラブ・チーム名：＿＿＿＿＿＿＿＿＿＿

競技年数：満　　　　年　　年齢：満　　　　歳　　身長：＿＿＿＿＿＿ cm　　体重：＿＿＿＿＿＿ kg

競技水準： 1. 日本代表(全年齢代表)　2. 地域選抜　3. 県選抜　4. 県内地区選抜　5. その他

ポジション： 1. GK　2. センターDF　3. サイドDF　4. 守備MF　5. 攻撃/サイドMF　6. FW　7. #10

知らない	知っている	問1．次のような戦術は知っていますか．また，試合でプレーできますか．	かなりできる	できる	ややできる	あまりできない	まったくできない	わからない
1	2	1)ボールを運ぶためのファーストタッチ	5	4	3	2	1	0
1	2	2)DFをかわすためのファーストタッチ	5	4	3	2	1	0
1	2	3)シュートのためファーストタッチ	5	4	3	2	1	0
1	2	4)スクリーン	5	4	3	2	1	0
1	2	5)フリーで前を向くターン	5	4	3	2	1	0
1	2	6)DFを背負って前を向くターン	5	4	3	2	1	0
1	2	7)ポストプレー	5	4	3	2	1	0
1	2	8)優先順位の高いパス	5	4	3	2	1	0
1	2	9)攻撃方向を変えるパス	5	4	3	2	1	0
1	2	10)タイミングの良いパス	5	4	3	2	1	0
1	2	11)方向が正確なパス	5	4	3	2	1	0
1	2	12)強さが適当であるパス	5	4	3	2	1	0
1	2	13)見せかけのパス	5	4	3	2	1	0
1	2	14)正確なクロス	5	4	3	2	1	0
1	2	15)方向を変えるドリブル	5	4	3	2	1	0
1	2	16)スピードを変えるドリブル	5	4	3	2	1	0
1	2	17)DFを引きつけるドリブル	5	4	3	2	1	0
1	2	18)DFラインを突破するドリブル	5	4	3	2	1	0
1	2	19)マークDFを突破するドリブル	5	4	3	2	1	0
1	2	20)ワンタッチシュート	5	4	3	2	1	0
1	2	21)コントロールからシュート	5	4	3	2	1	0
1	2	22)ドリブルからシュート	5	4	3	2	1	0
1	2	23)クロスからシュート	5	4	3	2	1	0
1	2	24)体の向き	5	4	3	2	1	0
1	2	25)ルックアップ	5	4	3	2	1	0
1	2	26)ボール保持者とのコミュニケーション	5	4	3	2	1	0
1	2	27)チェックの動き	5	4	3	2	1	0
1	2	28)プルアウェイ	5	4	3	2	1	0
1	2	29)ウェーブ(曲線の動き)	5	4	3	2	1	0
1	2	30)ダイアゴナルラン	5	4	3	2	1	0
1	2	31)サポート	5	4	3	2	1	0
1	2	32)視野から消える動き	5	4	3	2	1	0
1	2	33)DFの視野外から飛び出す動き	5	4	3	2	1	0
1	2	34)DFライン背後への走りこみ	5	4	3	2	1	0
1	2	35)相手を引きつける動き	5	4	3	2	1	0
1	2	36)スルーパス	5	4	3	2	1	0
1	2	37)壁パス	5	4	3	2	1	0
1	2	38)クロスオーバー	5	4	3	2	1	0
1	2	39)オーバーラップ	5	4	3	2	1	0
1	2	40)第3の動き	5	4	3	2	1	0

Appendix 1 Response sheet (continue).

知らない	知っている	問1. 次のような戦術は知っていますか. また, 試合でプレーできますか.	かなり できる	できる	やや できる	あまり できない	まったく できない	わから ない
1	2	41)オーバー	5	4	3	2	1	0
1	2	42)サイドチェンジ	5	4	3	2	1	0
1	2	43)サイドアタック	5	4	3	2	1	0
1	2	44)アーリークロス	5	4	3	2	1	0
1	2	45)プルバック	5	4	3	2	1	0
1	2	46)カウンターアタック	5	4	3	2	1	0
1	2	47)ダイレクトプレー	5	4	3	2	1	0
1	2	48)ポゼッションプレー	5	4	3	2	1	0
1	2	49)素早いリスタート	5	4	3	2	1	0
1	2	50)ショートコーナー	5	4	3	2	1	0
1	2	51)コーナーキック	5	4	3	2	1	0
1	2	52)直接フリーキック	5	4	3	2	1	0
1	2	53)ペナルティキック	5	4	3	2	1	0
1	2	54)バイタルエリアの攻略	5	4	3	2	1	0
1	2	55)ファンタジスタの攻撃	5	4	3	2	1	0
1	2	56)守備から攻撃への切り替え	5	4	3	2	1	0

知らない	知っている	問2. 次のような戦術は知っていますか. また, 試合でプレーできますか.	かなり できる	できる	やや できる	あまり できない	まったく できない	わから ない
1	2	1)プレッシング	5	4	3	2	1	0
1	2	2)振り向かせない	5	4	3	2	1	0
1	2	3)遅らせる	5	4	3	2	1	0
1	2	4)方向の限定	5	4	3	2	1	0
1	2	5)ボールの奪いどころをつくる	5	4	3	2	1	0
1	2	6)ボールを奪う	5	4	3	2	1	0
1	2	7)ジャストミート(ボールを奪うとき)	5	4	3	2	1	0
1	2	8)インターセプト	5	4	3	2	1	0
1	2	9)同一視野	5	4	3	2	1	0
1	2	10)裏をとられない	5	4	3	2	1	0
1	2	11)アプローチ	5	4	3	2	1	0
1	2	12)マンマーキング	5	4	3	2	1	0
1	2	13)ゾーンマーキング	5	4	3	2	1	0
1	2	14)マークの受け渡し	5	4	3	2	1	0
1	2	15)クロスボールへの対応	5	4	3	2	1	0
1	2	16)チャレンジ＆カバー	5	4	3	2	1	0
1	2	17)コレクティブディフェンス(挟み込み)	5	4	3	2	1	0
1	2	18)ブロック(守備の壁, 絞り込み)	5	4	3	2	1	0
1	2	19)チェーンリアクション(連動)	5	4	3	2	1	0
1	2	20)バランス	5	4	3	2	1	0
1	2	21)リトリート	5	4	3	2	1	0
1	2	22)スモールフィールド	5	4	3	2	1	0
1	2	23)ラインコントロール	5	4	3	2	1	0
1	2	24)コンパクト	5	4	3	2	1	0
1	2	25)バイタルエリアの守備	5	4	3	2	1	0
1	2	26)攻撃から守備への切り替え	5	4	3	2	1	0