Measurement and Research Department Reports 96-2

Evaluating Ability in a Korfball Game

H.H.F.M. Verstralen

Cito

Measurement and Research Department Reports 96-2

Evaluating Ability in a Korfball Game

H.H.F.M. Verstralen

Cito
Arnhem, april 1996

Introduction

The evaluation of the quality of involvement of individual participants in a team sport is a rather unprobed area in psychometrics. In this paper a psychometric analysis of data from a korfball game is proposed that may contribute to the development of similar approaches to other branches of competitive team sports or areas of human endeavor where there is a need to quantify individual contributions. From each of a series of games a special type of sequential data was registered and analyzed with classical and IRT methods. With the IRT methodology some light could be shed on the relative accomplishments of subgroups indicated by variables like gender, and other physical and social characteristics.

The Data

The data were collected as part of the program of the Dutch National Assessment of Educational Progress (DNAEP) conducted at Cito. There were 62 games, with 496 participants from group 8 (the last year in elementary education). In each game two teams of four players each strived to defeat one another in the number of goals. A team scores a goal when they throw the ball into the basket. In this particular basket game there is just one basket for both teams. To stimulate more field activity, a rule prescribes that the ball has to be passed within the same team twice before an attempt to score can be waged. Each participant played in exactly one game. To identify a player they were provided with a clearly visible identification number from 1 through 8. As data the series of players that occupied the ball were registered, interspersed with symbols for other actions than passing the ball to one another. These other actions and their labels are presented in Table 1. The symbol # is used as a shorthand for 'the number of'. The mentioned mnemonics in the column 'Mnem' are sometimes used for lack of space. Appendix A contains the data of the first game as an example.

Table 1

Actions Other than a Pass

Label	Mnem	
Goals	Gl	: #goals
Miss	Ms	: #unsuccessful goal aims
Outs	Out	: #times the ball was thrown beyond the confines of the play ground
Viols	Vio	: #rule violations

These data are transformed into a transition table, that furnishes the basis for all analyses. An example of a transition table is given in Appendix B. The table starts with a list of player identifications, and some general data on the game. These general data are already mentioned except for the first and the last three. 'Actions' gives the number of actions of each of the teams during the game. It is simply the number of recorded data that can be attributed to the members of a team. One may interpret this number as a measure of supremacy of the team. If the number of Actions of Team1 outweighs that of Team2, Team1 clearly has the upper hand. 'SerMn*' and 'SerSd*' give the mean and standard deviation times 10 of the number of contiguous actions within the team before the team lost the ball to its competitors. 'TeamChgs' is the number of times that control of the ball changed teams. The left upper subtable of the transition table contains the number of times the row subject passed the ball to his column team mate, with row and column totals behind and below 'Sub'. The second upper subtable contains the number of times the row subject lost the ball to the column member of the competing team. The right upper subtable records the number of alternative actions (column) of the row subject. The numbers in the column '1Ps' record the number of times the row subject delivered a pass that was directly followed by a goal aim. The column '2Ps' gives the number of times the row subject delivered a pass that was followed by a 1Ps. This explanation suffices also for the rest of the table.

Classical Analysis

Eight 'items' were constructed from the transition tables. They are listed in Table 2.

Table 2

Eight Items for Quality of Involvement

FThro	:	#times the ball was lost to an adversary
SThro	:	#successful passes
GAims	:	#goal aims
Goals	:	#goals
Catch	:	#times the ball was received from a team mate
Sntch	:	#interceptions (the number of times the ball from an adversary was caught)
Amtch	:	#times the ball was gained after a goal aim
APass	:	#1Ps's and 2Ps's

It appeared from preliminary analyses that the first item 'FThro' did not behave properly. A probable cause is the ambivalent character of the item. On the one hand one does not ably contribute to the game when a ball is lost to an adversary. On the other hand, one has to be actively involved in the game to be able to loose many passes. Therefore, the variable is neglected in the sequel.

For each player, the scores on the seven remaining items were computed, and saved together with two other data. The first is the total number of actions in the game, the second the total number of actions of the team of the player. We considered that the raw item scores perhaps did not justifiably reflect the contribution of the player, if compared with all players from all other games. One cannot expect that a player in a less motivated easy going game has the same opportunity to score on the items as an equally involved player in a hectic and fast game. Therefore, a second scale was created by dividing the raw item scores by the total number of actions in the Game. The unweighted scale is labeled 'NoRef', the second 'GActs'. However, one could also reason that a player in a team that has difficulty in gaining control of the ball, has equal scoring opportunity as an equally able member of a superior team. Therefore, the scale 'TActs' was obtained by dividing the item scores by the number of actions of the team of a player.

For each of the three item scales an aggregate score has to be devised that is optimal in some sense. For this analysis a weighted score was chosen that maximizes Cronbach's α reliability coefficient. It is easily derived that the optimal item weights are obtained as follows. Calculate the correlation matrix of the seven item scores and their standard deviations. Take the dominant eigenvector $\underline{e}$ of the correlation matrix, then the optimal weight of item i is given by

$$o_i = \frac{e_i}{\sigma_i},$$

where e_i denotes component i of $\underline{e}$, and σ_i the standard deviation of item i. The optimal score is now given by

$$r_o = \sum_i o_i r_i,$$

with r_i the score on item i.

Results

The correlation matrices of the item scores of the three scales are shown in Table 3.

Table 3

Correlations of Item Scores

Scores NoRef

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
SThro	1.0000						
GAims	0.5365	1.0000					
Goals	0.3355	0.6535	1.0000				
Catch	0.8602	0.6402	0.3986	1.0000			
Sntch	0.5433	0.4194	0.3052	0.3345	1.0000		
Amtch	0.3865	0.2309	0.1128	0.1568	0.2139	1.0000	
APass	0.6968	0.3488	0.2341	0.7026	0.1766	0.1205	1.0000

Scores TActs

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
SThro	1.0000						
GAims	0.2807	1.0000					
Goals	0.1967	0.6030	1.0000				
Catch	0.5453	0.4301	0.2717	1.0000			
Sntch	0.4102	0.2685	0.2004	0.0408	1.0000		
Amtch	0.2204	0.0351	0.0101	-0.2042	0.0669	1.0000	
APass	0.3414	0.1621	0.1634	0.3697	-0.0943	-0.1435	1.0000

Scores GActs

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
SThro	1.0000						
GAims	0.5054	1.0000					
Goals	0.3566	0.6596	1.0000				
Catch	0.8224	0.6173	0.4171	1.0000			
Sntch	0.4993	0.3915	0.3193	0.2577	1.0000		
Amtch	0.3048	0.1823	0.1212	0.0337	0.1553	1.0000	
APass	0.6664	0.3205	0.2524	0.6808	0.1401	0.0368	1.0000

Table 4 presents the means and standard deviations.

Table 4

Means and Standard Deviations of the Item Scores of the Three Scales

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
Means of Item Scores							
NoRef	15.1230	2.2601	0.4456	15.1230	3.0625	1.8044	4.4778
TActs	0.1708	0.0238	0.0046	0.1708	0.0355	0.0210	0.0473
GActs	0.0870	0.0130	0.0027	0.0870	0.0176	0.0103	0.0258
Standard Deviations of Item Scores							
	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
NoRef	6.7650	2.4103	0.9273	6.7291	2.4093	1.6239	3.5881
TActs	0.0474	0.0226	0.0098	0.0455	0.0269	0.0187	0.0310
GActs	0.0350	0.0134	0.0058	0.0347	0.0137	0.0089	0.0201

The optimal weights for the three scales are given in Table 5.

Table 5

Optimal Weights

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
NoRef	0.071	0.173	0.345	0.070	0.129	0.120	0.104
TActs	9.805	21.074	42.221	10.349	9.871	0.033	9.877
GActs	13.734	31.493	60.576	13.533	21.525	15.750	18.646

Because the contribution of an item to the optimal score not only depends on its weight but also on the typical size of a score, the product of weight and mean score is shown in Table 6.

Table 6

Products of Mean Item Scores and Optimal Weights

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
NoRef	1.075	0.391	0.154	1.052	0.394	0.217	0.466
TActs	1.674	0.502	0.194	1.767	0.350	0.001	0.467
GActs	1.195	0.410	0.162	1.178	0.380	0.162	0.481

It can be inferred from Table 6 that the main contribution to the optimal score comes from the successful delivery (SThro) and reception (Catch) of passes, followed by goal aims (GAims) and goal passes (APass). The actual scoring of a goal itself is of less importance, probably because its occurrence is relatively sparse. These findings support that the scale validly measures active involvement in the game.

The correlations of the item score with the optimal score are reported in Table 7.

Table 7

Item Score Correlations (RIT)

	SThro	GAims	Goals	Catch	Sntch	Amtch	APass
NoRef	0.9089	0.7874	0.6052	0.8847	0.5860	0.3692	0.7067
TActs	0.7310	0.7498	0.6504	0.7407	0.4177	0.0010	0.4824
GActs	0.8888	0.7823	0.6499	0.8687	0.5440	0.2604	0.6945

Tables 6 and 7 show that item Amtch (gain of ball control after a goal aim) especially in scale TActs hardly contributes to the optimal score. Although it can, therefore, be neglected in the classical analysis, it is kept for the IRT analysis discussed below.

The α reliability coefficients of the three scales are shown in Table 8.

Table 8

Reliability Coefficients, α, for the Three Scales

NoRef	0.84
TActs	0.70
GActs	0.83

Curiously, Table 8 leads to trust the second scale TActs, with the lowest reliability, as the most valid. Almost certainly, the higher reliabilities of the other two scales are induced by differences in levels of team activity and ball control. Therefore, the IRT analysis is confined to the scale TActs alone.

IRT Analysis

The IRT model selected for IRT analysis is OPLM (Verhelst, a.o.,1995, Verhelst and Glas, 1995), a special case of the GPCM (Muraki, 1992). In both models, the probability that person v obtains score j on item i is given by

$$f_{vij} = P(X_{vi} = j) \propto \exp[a_i(j\theta_v - \eta_{ij})], \tag{1}$$

where $\eta_{i0} = 0$. In the GPCM, the item discrimination is estimated whereas in OPLM it is considered part of the model hypothesis. With correct item discriminations OPLM offers the advantage of consistent estimation of item parameters using the CML approach. Moreover, OPLM supplies statistics that indicate whether the discrimination of an item was well chosen, or point to the direction of a desired change.

The preparation of an OPLM data set from the TActs item scores proceeds in two steps. First, raw item category weights are constructed, strictly increasing with the TActs item scores. Next, optimal category weights are estimated (Verstralen, 1996), under the restriction that they have to be contiguous, because the current implementation of OPLM cannot handle missing category scores.

The raw category weights for item i are obtained as follows. Let M_i denote the maximum raw category weight for item i with $M_i = 9$ for all i. Because the minimum raw category weight equals 0, the upperbound on the number of raw categories for item i equals $M_i + 1$. Let V denote the number of subjects ($V = 496$). Denote integer division with $\backslash$. Let $d = V\backslash(M_i + 1)$. If $d(M_i + 1) < V$, increase d by 1. Assume that the V scores on item i are ordered monotonously increasing with their index v. Denote the score of the individual with index d by s_0. All scores lower or equal to s_0 are subsumed under category 0. Let v_0 be the highest index with score equal to s_0. Next let s_1 be the score of the subject indexed by $v_0 + d$. All scores s with $s_0 < s \leq s_1$ are subsumed under category 1, etc.

The integer category weights for the OPLM model are estimated with the multiple correspondence analysis method described in Verstralen (1996). This method compares favorably with three other methods with respect to performance, computational cost, and implementational simplicity. A description of the method is included in Appendix C.

Results

The upper bounds of the item scores for the raw category weights are shown in Table 9.

Table 9

Item Score Upper Bounds for the Raw Category Weights

	0	1	2	3	4	5	6	7	8
SThro	0.1304	0.1512	0.1667	0.1774	0.1892	0.2000	0.2148	0.2323	0.2923
GAims	0.0132	0.0198	0.0254	0.0336	0.0432	0.0575	0.1148		
Goals	0.0115	0.0250	0.0741						
Catch	0.1340	0.1522	0.1630	0.1754	0.1875	0.1980	0.2125	0.2308	0.2899
Sntch	0.0135	0.0206	0.0259	0.0342	0.0417	0.0495	0.0617	0.0901	0.1455
Amtch	0.0108	0.0135	0.0187	0.0235	0.0286	0.0366	0.0476	0.1231	
APass	0.0198	0.0275	0.0377	0.0441	0.0543	0.0649	0.0759	0.0957	0.1852

The estimates of the integer category weights for OPLM are contained in Table 10. The number between [] shows the maximum category weight for the item.

Table 10

Estimates of Contiguous Integer Category Weights for OPLM

SThro	[3]	0122222333
GAims	[4]	02132344
Goals	[2]	0112
Catch	[4]	0123233334
Sntch	[2]	0112222222
Amtch	[7]	075156342
APass	[4]	0121233444

For instance, the optimal integer weights of categories 0 through 2 of item SThro are equal to their raw weights. But categories 3 through 6 obtain the same optimal integer weight as category 2. Categories 7 through 9 share the same optimal integer weight 3. Table 10 shows that a large number of raw categories can be collapsed into one category. Moreover, for some items, especially for Amtch, the original order of the raw categories has been changed.

Whereas the OPLM analysis of the raw category weights shows an unacceptable model misfit, with the estimated category weights an acceptable model fit has been obtained. This can be judged by the $R1c$-test and the distribution of p-values of the s-statistics as given in Table 11.

Table 11

Global OPLM-fit

$R1c$ = 69.059; df = 75; p = .6714

```
0.--/---/---.1-----.2-----.3-----.4-----.5-----.6-----.7-----.8-----.9-----1.
   1/  2/  1     7     2     1     1     1     1     2     3     3
```

Total: 25

p-values out of range: 1

The distribution of p-values shows a centrifugal tendency, which may be attributed to the fact that the S-statistics are not independent, certainly not within one item. However, it may be indicative of some chance capitalization as well. As a matter of fact the number of records (496) is rather low to give trustworthy information on differential discriminatory power of the items. The results of the classical analysis are reported in Table 12.

Table 12

Classical Analysis of the Data Set with Estimated Category Weights

item	p-value	rit(u)	rit(w)	item	p-value	rit(u)	rit(w)
2	.655	.648	.736	5	.588	.603	.708
3	.480	.710	.783	6	.774	.391	.418
4	.154	.523	.621	7	.480	.639	.355
				8	.585	.582	.576

Results Based on Weighted Scores

Mean	=	46.067
S.D.	=	18.235
Alpha	=	.726

Distribution of Responses

item	label	a	N	P	0	1	2	3	4	5	6	7
2	SThro	6	496	.655	51	52	256	137				
3	GAims	4	496	.480	144	48	105	102	97			
4	Goal	7	496	.154	368	103	25					
5	Catch	4	496	.588	51	51	103	254	37			
6	Sntch	3	496	.774	61	102	333					
7	Amtch	1	496	.480	102	50	35	50	53	104	51	51
8	APass	3	496	.585	55	99	103	100	139			

Table 12 shows that the reliability coefficient increased from 0.70 to 0.73 in comparison with the optimal weighted score scale TActs (Table 8). Another fact worth mentioning is the increase in item test score correlation of the item Amtch from 0 to 0.36 (compare with scale TActs in Table 7). Largely because of the reordering of the categories by the estimated integer weights this item now contributes sizably to the scale, whereas its contribution in the classical analysis was negligible.

The results of the calibration are shown in Table 13. The meaning of the column labels is:

A : discrimination index.

B : $\beta_j = \eta_j - \eta_{j-1}$, $j = 1...J_i$, with J_i the maximum score of item i .

SE(B) : the asymptotic standard error of estimation of β

S : the S-test, a $\chi^2(DF)$ distributed test statistic for model fit.

DF : degrees of freedom of the S-test

P : p-value of the S-test

M(i) : standard normally distributed statistics to indicate the appropriateness of A

Table 13

OPLM Calibration Results

nr	label		A	B	SE(B)	S	DF	P	M	M2	M3
2	SThro		6	-.212	.038	.000	1	.983	-1.578	-.701	-.602
		[:2]		-.334	.028	6.233	4	.182	-1.084	-.745	-.841
		[:3]		.181	.021	3.471	5	.628	1.018	.473	.309
3	GAims		4	.154	.043	6.491	4	.165	-1.300	-.637	-2.063
		[:2]		-.230	.045	2.118	5	.833	-.236	-.059	-.864
		[:3]		.058	.037	1.021	5	.961	-.512	.218	-1.117
		[:4]		.152	.038	1.876	4	.758	-1.651	-.717	-.656
4	Goal		7	.225	.020	4.792	5	.442	-1.837	-1.467	-1.589
		[:2]		.432	.040	.000	0	99.999	.713	-.431	-.844
5	Catch		4	-.180	.052	3.505	2	.173	-.768	-1.268	-1.045
		[:2]		-.267	.045	7.782	4	.100	-1.366	-.454	-1.422
		[:3]		-.211	.031	2.537	6	.864	-.972	-.493	-.493
		[:4]		.609	.046	4.904	3	.179	-2.412	-.012	-.569
6	Sntch		3	-.271	.055	2.351	5	.799	-1.303	-1.442	-1.591
		[:2]		-.398	.039	9.955	7	.191	.817	-.571	.865
7	Amtch		1	.616	.173	12.830	6	.046	1.445	1.924	2.170
		[:2]		.324	.222	15.707	6	.015	2.707	1.535	2.453
		[:3]		-.416	.222	11.594	7	.115	2.051	.970	2.059
		[:4]		-.068	.199	10.802	7	.148	2.211	1.577	2.561
		[:5]		-.579	.172	9.343	7	.229	1.418	.884	1.942
		[:6]		.782	.175	4.918	6	.554	2.598	.904	1.879
		[:7]		.047	.201	1.123	5	.952	3.144	.837	.564
8	APass		3	-.341	.057	3.651	3	.302	1.950	1.890	1.568
		[:2]		-.083	.048	7.436	6	.282	1.763	1.874	1.980
		[:3]		.021	.048	24.252	7	.001	2.487	1.319	2.865
		[:4]		-.011	.045	2.574	6	.860	1.286	1.072	1.257

Differences in Active Involvement Between Subgroups

Differences in latent ability between subgroups can be investigated with SAUL (Verhelst and Eggen, 1989). SAUL directly estimates the mean differences in ability between subgroups, without the intermediate ability estimation of the subgroup members, under the assumption of equal variance of latent abilities within subgroups. The present analyses are reported on a latent ability scale where the mean latent ability of all subjects is arbitrarily equated to 250, and the standard deviation to 50. The group indicators investigated are described in Table 14. The first column contains the numerical indicator of the variable, the second column a short name. Next follows a short description of the variable and, between (), the classes that are distinguished in the analysis. First, a univariate analysis is performed for each of these group indicators. Table 15 gives a short summary of the results, and shows that only the variables 5, 6 and 8 (Body, Freq, and Gender) exhibit significant effects on active involvement in the game. As a result of the moderate amount of observations (496), the size of the significant effects is considerable. The next analysis combines these three variables in a single threevariate model with only main effects. The results are reported in Table 16. It is clear that in a combined model only the variables 6 and 8 (Freq and Gender) remain as variables of interest. Although the effect of length is not significant, the effect of Gender may be confounded with length. Therefore, the interaction effects of Gender × Length were investigated. The results are contained in Table 17.

Table 14

Investigated Group Indicators

1	Stratum	: School disadvantage level (1,2,3)
2	Cert	: Certified teacher for physical education (yes,no)
3	Form	: Student disadvantage level (1,2,3)
4	Ethn	: Ethnic group (Autochthone, Allochthone)
5	Body	: Bodily stature, let x = 0.83 × length-weight-83.7, ($x \leq$ -5 (heavy), $x >$ -5 (Not heavy)).
6	Freq	: Clubsport frequency (Never, 1-7, 8-10, 11.. times per month)
7	Time	: Time (minutes) devoted to sports in general the previous week (0, 1-106, 107-250, 251..)
8	Gender	: (boys, girls)
9	Age	: Let x = (95-yy) × 12-mm + 1, (yy, mm year and month of birth) (x : 144-152, 153-162, 163-181)
10	Length	: cm (131-147, 148-163, 164-178)

Table 15

Summary of the Univariate Saul Analyses

	effect	SE	n	z	size
1. Stratum					
Stratum1	set to zero		146		
Stratum2	.458	8.122	79	.056	.009
Stratum3	-10.888	7.144	120	-1.524	-.219
2. Cert					
NoCert	set to zero		239		
Cert	-9.653	6.750	106	-1.430	-.194
3. Form					
Form1	set to zero		170		
Form2	-8.779	7.104	110	-1.236	-.176
Form3	-8.973	9.612	47	-.934	-.180
4. Ethn					
Autoch	set to zero		276		
Alloch	-4.961	9.161	48	-.542	-.099
5. Body					
Heavy	set to zero		60		
NotHeavy	16.982	8.308	250	2.044	.343
6. Freq					
Never	set to zero		87		
1 - 7pm	14.506	8.069	110	1.798	.302
8 - 10pm	21.462	9.485	59	2.263	.447
11pm	39.526	8.892	76	4.445	.823
7. Time					
NoTime	set to zero		156		
Sometime	-11.127	12.005	27	-.927	-.225
Regularly	1.138	8.697	61	.131	.023
Often	13.825	7.381	101	1.873	.280
8. Gender					
Boy	set to zero		185		
Girl	-28.716	6.086	160	-4.719	-.600
9. Age					
Young	set to zero		112		
Medium	11.435	6.922	185	1.652	.230
Old	-2.704	10.047	47	-.269	-.054
10. Length					
Short	set to zero		39		
Medium	3.417	10.102	226	.338	.069
Long	13.156	12.495	49	1.053	.264

Table 16

Results for the Model with Main Effects of Body, Freq, and Gender

5. Body	effect	SE	n	z	size
Heavy	set to zero		59		
NotHeavy	5.982	8.214	238	.728	.129
6. Freq					
Never	set to zero		83		
1-7pm	17.943	8.378	96	2.142	.387
8-10pm	22.023	9.946	50	2.214	.474
11pm	36.663	9.174	68	3.996	.790
8. Gender					
boy	set to zero		158		
girl	-23.946	6.489	139	-3.690	-.516

Table 17

Interaction Gender $\times$ Length

All main effects and interactions involving one or more of the following ategories are set to zero:
boy short

Main effects

	effect	SE	z	size
medium	-9.191	13.143	-.699	-.193
long	-4.682	15.984	-.293	-.098
girl	-56.115	18.228	-3.079	-1.179

First order interactions

	effect	SE	z	size
girl $\times$ medium	30.722	19.704	1.559	.645
girl $\times$ long	39.398	24.475	1.610	.828

Although the significance of the main effect of Gender decreases a little in the interaction model, still it is significant and sizable. Especially the short girls perform much worse than anybody else. Notice that effects of length within the two genders seem to differ markedly. Whereas length within the boys hardly matters, within the girls there is a nonsignificant tendency for ability to increase with length. The effect of length within the girls remains not significant if medium and long are collapsed into one category.

Discussion

The above analysis aims to measure individual accomplishment and involvement in a series of ball games each played by two different teams. The raw measures on seven variables were constructed from a transition table. The rows in this table can be interpreted as 'actors', the columns 'receivers' or 'actions'. Clearly the raw scores distilled from the transition table do not only reflect the accomplishments of the individual player but also the overall activity level of his team and of the game as such. In the present analysis it is assumed that the raw item scores can be made comparable over games by dividing them by the 'activity level' of the team. This choice was mainly inspired by considering that a scale may obtain a higher reliability by some team-attached halo effect, mediated by activity level. Therefore, the scale, among the three considered, with the lowest reliability seemed the most trustworthy. However, if the activity level across teams varies in a sizeable way, no compensation can be expected to make the scores comparable over teams. Although the line between acceptable and too much variability is not known, it is perhaps of interest to give the mean and standard deviation of the Actions over the teams: (86.77, 26.86). Still, the results of the SAUL analyses do seem to make sense, although, the absence of effects of length within boys is rather unexpected. An after the fact explanation of this finding may be that longer boys of this age are loosing some motor control and agility after the onset of a growth spur.

References

Bock, D.R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. Psychometrika, 37, 29-51.

Greenacre, M.J. (1989). *Theory and applications of correspondence analysis.* London: Academic Press.

Israëls, A. (1987). *Eigenvalue techniques for qualitative data.* Leiden: DSWO Press.

Muraki, E. (1992). A generalized partial credit model: application of an EM algorithm. *Applied Psychological Measurement, 16*,159-176.

Verhelst, N.D., & Eggen, Th.J.H.M. (1989). *Psychometrische en statistische aspecten van peilingsonderzoek.* PPON-rapport nr.4. Arnhem: Cito.

Verhelst, N.D., Glas, C.A.W., & Verstralen, H.H.F.M. (1995). *One-parameter logistic model, OPLM.* Arnhem: Cito.

Verhelst, N.D., & Glas, C.A.W. (1995). The one parameter logistic model. In G.H.

Fischer, G.H., & I.W. Molenaar (Eds.). *Rasch models; Foundations, recent developments, and applications.* New York: Springer-Verlag.

Verstralen, H.H.F.M. (1996). *Estimating integer category weights in two IRT models for polytomous items.* Measurement and Research Department Reports, 96-1. Arnhem: Cito.

Appendix A

The Registered Data of Game 1

The first eight rows show the identification codes of the players, followed by a line with ##8 to indicate the end of identifications. The next lines contain the data proper. The numbers 1 through 8 indicate ball control by the player with that number. 'M' represents an unsuccessful goal aim, 'G' a goal, 'O' ball out, 'V' a rule violation, and 'E' end of game.

```
  1  1 20675  1
  1  2 20675  6
  1  3 20675 21
  1  4 20765 10
  1  5 25340 28
  1  6 25340 59
  1  7 25340  2
  1  8 25340 57
##8
42321M5286O42128O26712342M3241242M743M1M57243M21423M1424O61O686
O65652341M57424573124M767V1432G5642M68253132341M4212V723413M21342
M8671243M68568M24312312M4132M41656578MO21212342G62483M4231MO6O
16231G856M412M56854312G6586O3421M67816M3156872432M8587O86M57E
```

APPENDIX B

Transition Table of Game 1

Game

1

Player Identifications			Type	Team1	Team2	Total
1.	20675	1	Actions	132	75	207
2.	20675	6	Goals	4	0	4
3.	20675	21	Miss	21	5	26
4.	20765	10	Outs	3	7	10
5.	25340	28	Viols	1	1	2
6.	25340	59	SerMn*	46	26	
7.	25340	2	SerSd*	37	19	
8.	25340	57	TeamChgs			
			*: Miltiplied by 10		57	

To →

From	1.	2.	3.	4.	Sub	5.	6.	7.	8.	Sub	Tot	Gl	Ms	Out	Vio	1Ps	2Ps	Sub	Tot
1.	0	12	4	3	19	1	3	0	0	4	23	1	6	1	0	5	2	15	38
2.	8	0	10	10	28	1	1	0	2	4	32	3	8	0	1	4	5	21	53
3.	8	6	0	7	21	0	0	0	0	0	21	0	6	0	0	5	9	20	41
4.	7	13	7	0	27	1	0	0	1	2	29	0	1	1	0	10	6	18	47
Sub	23	31	21	20	95	3	4	0	3	10	105	4	21	2	1	24	22	74	179
5.	0	2	1	1	4	0	7	5	2	14	18	0	0	0	0	1	2	3	21
6.	1	2	0	1	4	5	0	4	6	15	19	0	3	4	0	1	1	9	28
7.	2	3	1	2	8	0	1	0	2	3	11	0	0	1	1	1	1	4	15
8.	1	1	1	0	3	4	5	2	0	11	14	0	2	1	0	2	1	6	20
Sub	4	8	3	4	19	9	13	11	10	43	62	0	5	6	1	5	5	22	84
Tot	27	39	24	24	114	12	17	11	13	53	167	4	26	8	2	29	27	96	263
Gl	0	0	0	0	0	1	2	0	1	4	4	0	0	0	0	0	0	0	4
Ms	2	3	2	5	12	5	3	2	2	12	24	0	0	2	0	0	0	2	26
Out	1	2	1	1	5	0	4	0	1	5	10	0	0	0	0	0	0	0	10
Vio	1	0	0	0	1	0	0	1	0	1	2	0	0	0	0	0	0	0	2
Sub	4	5	3	6	18	6	9	3	4	22	40	0	0	2	0	0	0	2	42
Tot	31	44	27	30	132	18	26	14	17	75	207	4	26	10	2	29	27	98	305

Appendix C

Optimal Integer Category Weights Estimation in the OPLM
with Multiple Correspondence Analysis

Let γ_{ij} be the optimal category weight for category j of item i obtained from multiple correspondence analysis (Greenacre, 1989, Israëls, 1987) of the data set with raw OPLM category weights (obtained from the first step in the creation of the OPLM data set). Denote the vector of weights for item i with $\boldsymbol{\gamma}_i$. Assume that the elements of $\boldsymbol{\gamma}_i$ are monotonously increasing with the category indices, if not reindex the categories. It is discussed in Verstralen (1996) that the elements of $\boldsymbol{\gamma}_i$ are more or less linear in the integer category weights $\boldsymbol{s}_i$ for OPLM, that is $\gamma_{ij} \approx c_i s_{ij} + d_i$, where the probability that person v obtains score j on item i is given by

$$f_{vij} = P(X_{vi} = j) \propto \exp[a_i(s_{ij}\theta_v - \eta_{ij})], \tag{2}$$

Observe that the category index multiplier of θ in formula (1) has been replaced by a integer weight s_{ij} in formula (2) that has to be estimated. Because the following lines apply to each item separately the item index is omitted. Let W be a positive definite symmetric weight matrix of order $J+1$ with $1'W1=1$. Let $\bar{X} = \mathbf{1}'W\boldsymbol{x}$, and $Cov(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}'W\boldsymbol{y} - \bar{x}\bar{y}$. Define a set S of all integer category weight vectors $\boldsymbol{s}$ to be considered, let $\boldsymbol{v}$ be a vector with elements $v_j = cs_j + d - \gamma_j$ ($j=0,\ldots,J$), and minimize the GLS loss function $L(\boldsymbol{s},\boldsymbol{\gamma}) = \boldsymbol{v}'W\boldsymbol{v}$ over S, where the values c and d are the GLS regression coefficients $\boldsymbol{c} = \dfrac{Cov(\boldsymbol{s},\boldsymbol{\gamma})}{Cov(\boldsymbol{s},\boldsymbol{cs})}$, and $\boldsymbol{d} = \bar{\sigma} - c\bar{s}$. In the present application W is equal to $I/(J_i+1)$, with I the identity matrix of order J_i+1.

The set S is defined by all possible collapses of adjacent categories (increasing order of the elements of $\boldsymbol{\gamma}_i$ assumed). Indicate collapsed categories by putting them between {}. For an item with four categories some elements of S are (0,1,2,3), ({0,1},2,3), {0,{1,2,3}}), etc.

Recent Measurement and Research Department Reports:

96-1 H.H.F.M. Verstralen. Estimating Integer Parameters in IRT Models for Polytomous Items.