- 1 Section: Training and Testing
- 2 Title: The reliability and validity of a field hockey skill test.
- 3
- 4 Running title: Reliability of a field hockey skill test.
- 5
- 6

## 7 Abstract

- 8 High test retest reliability is essential in tests used for both scientific research and to monitor
- 2 9 athletic performance. Thirty-nine (20 male and 19 female) well-trained university field
- 3 10 hockey players volunteered to participate in the study. The reliability of the <u>in house designed</u>
- 4 11 test was determined by repeating the test (3-14 days later) following full familiarisation. The
- 5 12 validity was assessed by comparing coaches ranks of players with ranked performance on the
- 6 13 skill test. <u>The mean difference and confidence limits in overall skill test performance was 0.0</u>
- 7 14  $\pm 1.0\%$  and the standard error (confidence limits) was 2.1% (1.7 to 2.8%). The mean
- 8 15 difference and confidence limits for the 'decision making' time was  $0.0 \pm 1.0\%$  and the
- 9 16 <u>standard error (confidence limits) was 4.5% (3.6 to 6.2%). The validity correlation (Pearson)</u>
- 10 17 was r = 0.83 and r = 0.73 for female players and r = 0.61 and r = 0.70 for male players for
- 11 18 overall time and 'decision making' time respectively. We conclude that the field hockey skill
- 12 19 test is a reliable measure of skill performance and that it is valid as a predictor of coach
- 13 20 assessed hockey performance, but the validity is greater for female players.
- 14 21 **Keywords:** Intermittent exercise, team sports
- 15
- 16
- 17
- .
- 18
- 19
- 20
- 21

- 22
- 23
- 23
- 24
- 25

#### 22 Introduction

1 23 To undertake research into field hockey in a controlled setting, it is necessary to employ a 2 24 skill test that can be completed in the laboratory environment. However, there are only a 3 25 limited number of field hockey skill tests and very little has been done scientifically to 4 26 formulate tests that measure playing ability (14). Two decades later, further developments of 5 27 hockey tests had not advanced. Reilly and Borrie (10) noted that it was surprising that even 6 28 though field hockey had been part of the Physical Education curriculum in Europe and North 7 America since the beginning of the 20<sup>th</sup> Century, there had been little attention given to the 29 8 30 design of field tests for the game. 9 31 32 Thus, at present the number of published tests of field hockey skill is limited and no skill tests 10 33 have been published during the last fifteen years. With the advent of synthetic sportsturfs as 11 34 the major playing surface over that period, it is apparent that the skills have changed 12 13 35 significantly and thus there is a need to develop a skill test that is appropriate to modern field 14 36 hockey. Furthermore, the skill tests were designed to determine differences in skill 15 37 performance between players, rather than to monitor improvements or changes for a particular 16 38 player, and thus were not stringently tested for reliability.

17 39

18 40 In the formulation of a skill test, it is important that technique is differentiated from skill.

<sup>19</sup> 41 Technique is the production of some pattern of movements which are technically sound (7).

20 42 The following definition of skill will be used for the purpose of the design of this study: "Skill

<sup>21</sup> 43 is the learned ability to bring about predetermined results with the maximum certainty, often

<sup>22</sup> 44 with the minimum outlay of energy, or of time and energy," (7). This encompasses the idea

<sup>23</sup> 45 that a skilled athlete must take an action that is appropriate and therefore the skill involves

<sup>24</sup> 46 interpreting the needs of the situation and making the correct decision as well as carrying out

25

3 49 skill.

4 50

Over the past decade there has been an increase in the literature regarding the importance of reliability and validity studies and the statistics that should be employed and interpreted. In terms of reliability, it has been advocated that a number of statistical methods be cited and interpreted (1). Reliability has been partially defined to include the "consistency of an individual's performance on a test" (1). It should be recognised that tests will always include some form of measurement error and therefore reliability needs to be considered as the amount of measurement error that has been deemed acceptable for the effective practical use of a measurement tool. When the tool is to be used for scientific research, the acceptable level is of paramount importance. To conclude that a measuring tool is valid, it must show logical, construct and criterion validity (13). Logical validity means that the tool is appropriate to want you want to measure, construct validity refers to a measuring tool that can discriminate between standards and criterion validity refers to how well the measuring tool correlates to previous tools used to measure the same variable (13).

18 64

19 65 The aim of this study was to design a field hockey skill test, which is both reliable and valid
20 66 for the modern game of hockey and determine the acceptable levels to make it a suitable tool
21 67 to use for research in a laboratory environment.

#### 68 Materials and methods

2 69 Participants

Thirty-nine university hockey players volunteered to take part in the study. Twenty males and
19 females completed the validity study, whereas only 14 males and 17 females completed the
reliability of the skill test. The study had Loughborough University Ethical Committee
Approval and informed consent was obtained.

7 74

8 75 Skill Test Design

9 76 The test was designed to include numerous elements of the game of hockey, incorporating

10 77 dribbling, passing and shooting, whilst controlling as many variables as possible. For example

11 78 a field hockey rebound board was used to pass off and the surface for the test was a water-

<sup>12</sup> 79 based sportsturf (Desso), the type of surface all the players regularly play and train on. The

13 80 goal is the width of a normal field hockey goal and the target area for the skill test is 18 inches

14 81 high, which is the height of a backboard in hockey.

15 82

16 83 The objectivity of the skill test was paramount in the design and therefore participants were 17 84 only given instructions about the penalty timing system and completing the test as quickly and 18 85 as accurately as possible. No information on how to approach the test was provided. This 19 86 allowed the participants to use techniques, make decisions and react to the different elements 20 87 as they would in a game.

21 88

22 89 The skill test requires the participants to start from a line 16 yards from the goal. The player
23 90 then runs to a hockey ball and then dribbles round the cones in a specific sequence (Figure 1).

24 91 The completion of the dribbling phase requires the player's foot or ball to break an infra-red

25 92 beam which triggers a light on either side of the goal and starts a computer

1 timing system (BBC microcomputer). The player then makes a pass against the rebound board 93 2 94 (Figure 1) and shoots at either the right side or left side target on the goal. The player 3 must shoot at the opposite target to where the light is on, for example if the light is on above 95 4 96 the right side target, the player must shoot at the left side of the goal. The player must always 5 97 shoot straight at the target and not diagonally. For the previous example to shoot at the left 6 98 side of the goal, the player must bring the ball round the left hand side of the five cones to 7 99 shoot in a straight direction (Figure 1). When the player has shot, the ball will hit the goal and 8 100 stop the timing system, which is triggered by the sound of the ball. The time taken between 9 101 crossing the infra-red beam and the ball hitting the backboard was termed the 'decision 10 102 making' time as it incorporates the decision making elements of how and when to pass 11 103 against the rebound board or shoot and determining which side of the goal to shoot. After the 12 104 completion of the shot the player then runs back to the start line.

13 105

14 106 The player repeats the dribble, pass and shot pattern six times; each time the player has to 15 107 touch the yellow line with a foot. The total time is recorded for the six continuous runs. In 16 108 addition, a penalty time of 2 s per error is added, if the player misses the target area on the 17 109 goal, touches a cone with the ball or the ball touches the player's feet. The total time for the 18 110 six runs and any error time is termed the 'overall time' and is used as the measure of 19 111 performance for the field hockey skill test. The 'decision time' is taken as the average of the 20 112 six decision timings, which incorporates three shots at the right target and three shots at the 21 113 left target, in a randomised order. Three shots at each target controls for the different distance 22 114 that is covered by the player depending on the side of the goal that he/she is shooting at. 23 115

24 116 The players were verbally encouraged to perform maximally and informed about the number25 117 of repetitions remaining. If the players lost control of the ball, they had to continue

1 118 from wherever the ball went in an enclosed $55.5m^2$ are	rea.
--	------

- 3 120 Familiarisation
- 4 121 Subjects were familiarised with the skill test on two occasions. During the first session they
- 5 122 were instructed about how to complete the skill test and the timing and penalty system. They
- 6 123 then completed 10 repetitions of the test, resting between each repetition. The pattern was
- 7 124 randomised, but five shots were completed to each side of the goal. The second
- 8 125 familiarisation session required the subjects to perform the skill test in its entirety. Thus, they
- <sup>9</sup> 126 completed the six repetitions as fast as they could, and the overall time and decision time were
- 10 127 recorded. The mean difference ± confidence interval for the familiarisation and first trial data
- 11 128 was  $-4.2 \pm 2.6\%$  and the typical error (confidence interval) was 4.8% (3.9 to 7.0%).
- 12 129
- 13 130 Reliability Trials
- 14 131 After being fully familiarised, 31, of the 39 subjects who completed the validity study,
- 15 132 completed the skill test on two occasions on separate days, <u>3 to 14 days apart</u>. The subjects
- 16 133 were asked to refrain from vigorous exercise on the day of the skill test. To account for
- 17 134 circadian rhythms, the skill tests were completed at the same time of day.
- 18 135
- 19 136 Validity Trials
- 20 137 Thirty-nine subjects completed the skill test after refraining from vigorous exercise on that
- 21 138 day, but were not informed about their performance. The male players who completed the test
- 22 139 were then ranked for performance and skill on their normal game play by one international-
- 23 140 standard coach (coach 1) and one National League coach (coach 2). Similarly, an
- 24 141 international-standard coach (coach 3) and one National League coach (coach 4) ranked the
- 25 142 female players' who completed the test. The coaches were provided with a definition of skill

1 143 and performance, which they could use to rank the players, so that all the coaches were 2 1 4 4 working to the same criteria. Performance was defined as overall match performance and 3 1 4 5 contribution to a match and skill defined as "the learned ability to bring about predetermined 4 146 results with the maximum certainty, often with the minimum outlay of energy, or of time and 5 147 energy," (7). The coaches were provided with the 6 1 4 8 names of the players, but were not given any information about the performance of the 7 149 players on the field hockey skill test. All the coaches regularly coached and watched the 8 1 5 0 players who they ranked, so were fully aware of their abilities. The performance ranks were 9 151 compared with the overall time for the skill test, whereas the skill ranks were compared with

- 10152 the decision time.
- 11153
- 12<sub>154</sub> Statistical Analyses

<sup>13</sup>155 The reproducibility of the skill test was determined using numerous statistical techniques.

14156 These were mean difference, Bland and Altman limits of agreement, correlations and typical

15157 error (1, 2, 5). The coaches' ranks and skill test scores were compared using a Pearson

16158 correlation. Data were checked for non-uniformity, so that the appropriate statistical

17159 techniques could be employed.

18160

19161 Results

20162 *Reliability* 

21163 The mean ( $\pm$  SD) for the overall performance time for trial 1 and trial 2 was 83.93  $\pm$  6.60 and

- 22164 84.36  $\pm$  7.44 s for men and 96.56  $\pm$  6.68 and 96.26  $\pm$  6.12 s. The mean difference and
- 23165 <u>confidence limits in overall skill test performance was  $0.0 \pm 1.0\%$  and the standard error</u>

24166 (confidence limits) was 2.1% (1.7 to 2.8%). The mean difference and confidence limits for the

25167 <u>'decision making' time was  $0.0 \pm 1.0\%$  and the standard error (confidence limits) was 4.5\%</u>

- 169 overall performance and 'decision making' time reliability of the skill test respectively. There
- 170 is a strong relationship for overall skill test performance as indicated by a Pearson and
- 171 intraclass correlation above 0.90 (Table 1). The relationship for decision time was also good,
- being above 0.70 (Table 2). Figure 2 shows the Bland and Altman plot for overall
- 173 performance time for trial 2-1, and gives a mean difference and limits of agreement of  $0.03 \pm$
- 174 5.11 s. The Bland and Altman plot for decision time shows a mean difference and limits of
- 175 agreement of  $0.01 \pm 0.52$  s (Figure 3).
- 176
- 177 Validity
- 178 The Pearson correlation for the mean women's' coaches rank and overall time was r = 0.83
- 179 (P<0.01) and decision time was r = 0.73 (P<0.01). The Pearson correlation for the mean
- 180 men's coaches rank and overall time was r = 0.61 (P<0.01) and decision time was r = 0.70
- 181 (P<0.01). Figure 4 shows a plot of the z-score from the mean coaches rank (residual) versus
- 182 the overall time for the skill test. The figure shows good uniformity of the data. The standard
- 183 error of the estimate for the overall time is 0.58 for the women and 0.81 for the men. The
- 184 standard error of the estimate for the decision time is 0.70 for the women and 0.74 for the
- 185 <u>men.</u>
- 186
- 187
- 188 Discussion
- 189 The main finding from the present study was that the reproducibility of the skill test was
- 190 good. Correlations between the two trials were high to very high (4). The correlations
- 191 between coaches' rankings and player performance were also high for the high standard
- 192 players used.

193 To assess test reliability, Hopkins (5) has advocated the use of typical error rather than 194 the limits of agreement approach that is recommended by Atkinson and Nevill (1). Hopkins 195 (5) suggested that the value of the limits of agreement approach is dependent upon the sample 196 size of the reliability study. The bias of the limits of agreement are <5% if there are >25197 subjects; however if there is only 8 subjects this bias is 21%. In the current study, there are 198 over 30 subjects for the data for men and women combined and therefore the bias will be low. 199 The Bland and Altman (2) limits of agreement provide a confidence interval for the 200 differences between two trials and it is up to the experimenter to determine whether this range 201 is acceptable. Hopkins (5) suggested that a 95% confidence interval used for the limits of 202 agreement approach is too stringent a measure if used for looking at an athlete's improvement 203 in performance and that half the limits of agreement would still leave approximate odds of 5-1 204 that performance had actually improved. Thus, the limits of agreement allows for an 205 underestimation of the reliability of the protocol as it takes into account 2 standard deviations 206 rather than the usual one that is used as an indicator of variation. The mean difference and 207 limits of agreement for overall skill test performance was  $0.03 \pm 5.11$  s and for 'decision 208 making' was  $0.01 \pm 0.52$  s. In contrast the typical error of overall performance, as advocated 209 by Hopkins (5) of the test was 2.0 s for men and 1.7 s for women. For the 'decision making' 210 data the typical error was 0.20 s for men and 0.18 s for women. The typical error is the 211 within-subject standard deviation and represents the variation we could expect to see from 212 trial to trial for each subject (5).

213

A key aspect of determining whether the reliability of the test is appropriate to the tests use is
to assess the minimum worthwhile change that matters to the coach, player or scientist. This
worthwhile change value may vary between the player, coach and scientist. For team sports
particularly this is a very difficult value to determine as the performance on a skill test, such

1	218	as that presented in this paper may not directly reflect performance on the pitch due to
2	219	numerous and complex interactions that occur during team sports. To try to overcome this
3	220	issue the validity for this test was assessed by ranking players on their overall pitch
4	221	performance and correlating this with the performance on the skill test. In terms of the
5	222	minimum worthwhile change for reliability of tests associated with team sports, Hopkins (6)
6	223	outlined that the smallest worthwhile change should be <0.2 of the between athlete SD. For
7	224	the current test the error of the measurement or typical error related to the between athlete SD
8	225	is 0.26 for the women and 0.28 for the men respectively. Therefore this is slightly higher than
9	226	the value indicated. Based on this data, in practical terms the smallest worthwhile change for
10	227	the field hockey skill test should be 2s. If the performance of a player on the test improves by
11	228	2s, you can be confident that the player has improved his or her performance on the skill test.
12	229	
13	230	These data show that the reliability of the skill test is considerably better than any previously
14	231	published data for field hockey (3, 11, 12, 14). The tests that have been formulated during the
15	232	last 20 years have been designed as field tests to determine differences between players rather
16	233	than for repeated measures on the same player. The Chapman Ball Control Test (3) isolates
17	234	the ability of an individual to control the ball manipulatively by arm, wrist and hand action
18	235	within a 9.5" (24 cm) diameter circle. This could be described as measuring dribbling
19	236	technique rather than field hockey skill per se. Thus, it cannot be a measure of playing ability
20	237	since this is not what constitutes the entire domain of field hockey skill. The testing took
21	238	place on a gymnasium floor, which is a considerably different surface from the outdoor game.
22	239	While the results from the Chapman test correlate well with subjective opinions of playing
23	240	ability, it does not attempt to measure any other characteristics. Testing of ball control is
	241	obviously important, but analyses of match

play highlight how little time players spend with the ball during a match and the very short 242 243 duration of each period with the ball. The validity of the Chapman test would be reasonable if 244 the test scores were compared with subjective ratings of ball control and not overall playing 245 ability. Reilly and Bretherton (11) developed a field-based skill test, namely the "T"-dribbling 246 test and a dribbling and accuracy test, to help determine the fitness of female hockey players. 247 The T-dribbling test was shown to be correlated with aerobic fitness (r = 0.48; estimated  $\dot{V}O_{\gamma}$  max and physical working capacity) and anaerobic power (r = 0.6; stair run test). The 248 249 accuracy was correlated with ectomorphy (r = -0.63). The skill tests provide useful field tests, 250 but do not provide us with a test that includes a passing aspect and 'decision making' element. 251 Furthermore, the "T"-dribbling test, is restrictive in that the players were unable to use reverse 252 sticks, which is an integral part of the game and therefore would not be a suitable measure of 253 hockey performance per se.

254

In our laboratory, previous soccer skill tests have been developed for use in researching the effects of fatigue on skill performance. The reliability in terms of mean difference ( $\pm$  limits of agreement) for the Loughborough Soccer Passing Test was  $-0.1 \pm 11.2\%$  (9). The limits of agreement are much greater than those in the current study ( $0.0 \pm 5.6\%$ ), suggesting that the reliability of the field hockey skill test is good and acceptable for scientific research.

260

The validity of the field hockey skill test is moderate to good. The term 'validity' used in the current study, refers to both logical validity and construct validity. Logical validity means that the test is appropriate to what you want to measure, whereas construct validity refers to a test which can discriminate between groups of performers (13). A further type of validity, should be tested for, namely criterion validity, which means that the test needs to be compared with an established test. However, as there does not seem to be a

1 267 previous field hockey skill test that is regarded as 'established', this is inapplicable. The skill 2 268 test shows better validity for the women than the men for both the correlations and typical error or standardised error of the estimate values. The typical error values were 0.58 for the 3 269 270 women and 0.81 for the men. Though these values are larger than we would hope for they are 4 271 justifiable from the method used. The coaches ranks were based off performance and are 5 6 272 extremely subjective so the variation will be much greater between players. Further more the 7 273 players used were all of a high standard so the variation between players would have been low 8 274 increasing the difficulty for coaches rankings. The greater validity for women may be due to 9 275 the different demands and styles of play adopted by men and women. In field hockey there are 10 276 "physical and physiological differences between the sexes" that means that the game of 11 277 hockey will be played differently by men and women (8). For elite hockey players, men were <sup>12</sup> 278 found to have a higher  $\dot{V}O_2$  max and haemoglobin content and were faster, taller and heavier <sup>13</sup> 279 than the women (8).

<sup>14</sup> 280

<sup>15</sup> 281 Skill tests need to be objective as well as valid and reliable. Though the objectivity of the test
<sup>16</sup> 282 has not been statistically determined, the test should exhibit good objectivity. The test
<sup>17</sup> 283 performance is determined by timings, which are completed by a computer and stopwatch and
<sup>18</sup> 284 penalty time. The players are only instructed in what order to complete the test and the
<sup>19</sup> 285 penalty system, and thus the inferences of the testers are minimal. The tester is only
<sup>20</sup> 286 responsible for timing and counting the number of penalties so the results should be similar, if
<sup>21</sup> 287 not identical between all testers.

22 288

23 289 The test was performed on a typical sportsturf and is thus easily transferable between pitches.

<sup>24</sup> 290 The field hockey skill test could be easily transferred to the pitch, using the goal and could be

<sup>25</sup> 291 made as realistic as is required. The movement of a goal keeper could determine the side for

1	292	shooting, with another attacker playing the pass and a defender taking the place of the five
2	293	cones to shoot around. Thus, the test could be as scientific or match-like as is required, and
3	294	could range from a coaching aid to a selection aid.
4	295	
5	296	The limits of agreement and typical error indicate the reliability of the skill test is very good
6	297	and that changes in overall performance of greater than 2.1% could be attributed to the
7	298	intervention. In summary, the field hockey skill test provides a reliable, objective and valid
8	299	tool for testing the skills of good to elite field hockey players. The high reliability and validity
9	300	allows it to be used for scientific research as well as determining how the skills of individual
10	301	players are developing.
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		

## 302 **References**

- <sup>1</sup>Atkinson G, Nevill AM. Statistical methods for assessing measurement error (reliability) in
   variables relevant to sports medicine. Sports Med 1998; 26: 217-38.
- <sup>2</sup>Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of
- 306 clinical measurement. Lancet 1986; 1: 307-310.
- <sup>3</sup>Chapman N. Chapman Ball Control Test- Field Hockey. Res Q Exerc Sport 1982; 53: 239242.
- <sup>4</sup>Cohen L, Holliday M. Statistics for Social Scientists. London: Paul Chapman Publishing,
  1982.
- <sup>5</sup>Hopkins WG. Measures of reliability in sports medicine and science. Sports Med 2000; 30:
- 312 1-15.
- 313 <u><sup>6</sup>Hopkins WG. How to interpret changes in an athletic performance test. Sportscience 2004;</u>
  314 8:1-7.
- <sup>7</sup>Knapp B. Skill in Sport: The attainment of proficiency. London: Routledge and Kegan Paul,
  1963.
- <sup>8</sup>Lakomy J, Dabinett J, Nevill ME. Physiological differences between elite male and female
- 318 games players. Faster, Higher, Stronger 2000; 20-21.
- <sup>9</sup>McGregor S. The influence of fluid ingestion on metabolism and soccer skill following
- 320 intermittent high intensity shuttle running. Unpublished doctoral dissertation 1999.
- 321 Loughborough University, UK.
- <sup>10</sup>Reilly T, Borrie A. Physiology applied to field hockey. Sports Med 1992: 14: 10-26.
- <sup>323</sup> <sup>11</sup>Reilly T. Bretherton S. Multivariate analysis of fitness of female field hockey players. In:
- 324 Day JAP (ed). Perspectives in Kinanthropometry Champaign, IL: Human Kinetics, 1986:
- 325 135-142.

- 326 <sup>12</sup>Schmithals M, French E. Achievement tests in field hockey for college women. Res Q Exerc
- 327 Sport 1940; 11: 84-92.
- <sup>13</sup>Strand BN, Wilson R. Assessing Sports Skills. Champaign, IL: Human Kinetics, 1993.
- 329 <sup>14</sup>Wessel JA, Koening FB. Field hockey. In:. Larson LA, (ed). Encyclopaedia of Sports
- 330 Sciences and Medicine. London: Collier-MacMillan, 1971: 679-680

	Men	Women	All
Mean $(\pm SD)$ time trial 1 (s)	$83.93 \pm 6.60$	$96.56\pm6.86$	90.85 ± 9.21
Mean ( $\pm$ SD) time trial 2 (s)	$84.36\pm7.44$	$96.26\pm6.12$	$90.89 \pm 9.18$
Mean difference (s) (confidence interval -, +)	0.4 (-1.2, 2.1)	-0.3 (-1.6, 1.0)	0.0 (-0.9, 1.0)
Typical error (s)	2.0	1.7	1.9
Pearson correlation (r)	0.93 P <0.0001	0.94 P <0.0001	0.96 P<0.0001
Intraclass correlation (r)	0.92	0.94	0.96

Table 1 Statistical summary of the reproducibility for the overall time of the skill test.

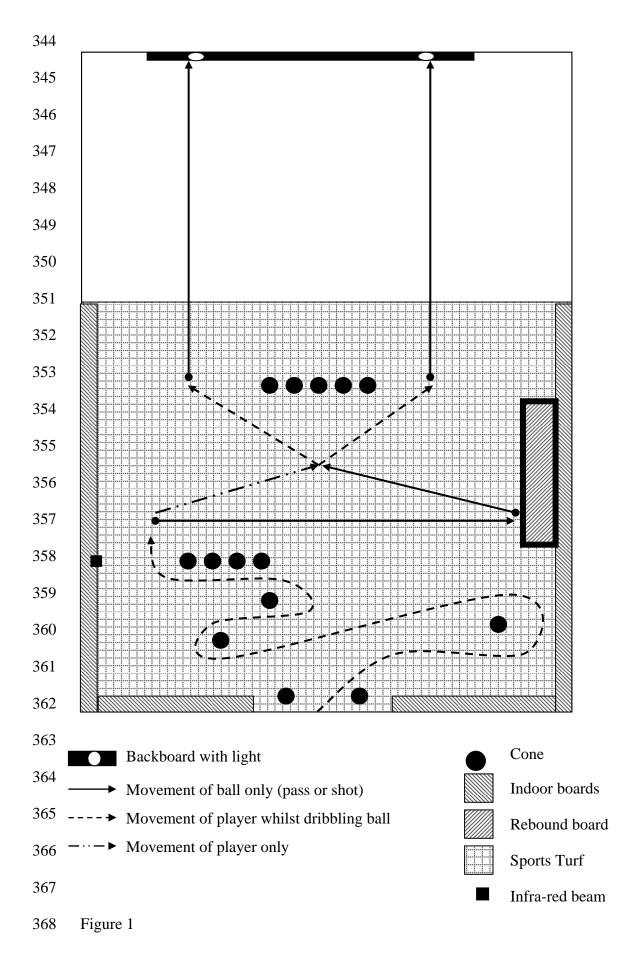
Table 2 Statistical summary of the reproducibility of the 'decision making' time of the skill

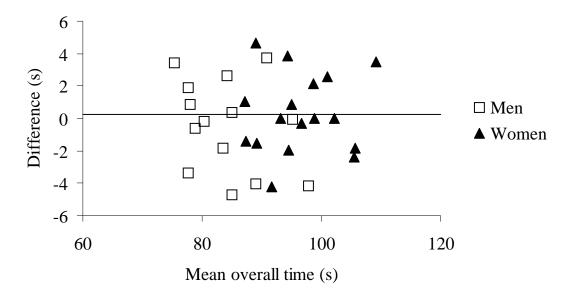
334 test.

	Men	Women	All
Mean ( $\pm$ SD) time trial 1 (s)	$3.82\pm0.37$	$4.43\pm0.41$	$4.17\pm0.49$
Mean ( $\pm$ SD) time trial 2 (s)	$3.78\pm0.35$	$4.47\pm0.53$	$4.17\pm0.57$
Mean difference (s) (confidence interval -, +)	-0.04 (-0.2, 0.13)	0.04 (-0.1, 0.17)	0.0 (-0.1, 0.11)
Typical error (s)	0.20	0.18	0.19
Pearson correlation (r)	0.70 P <0.01	0.89 P <0.0001	0.89 P <0.0001
Intraclass correlation (r)	0.70	0.85	0.88

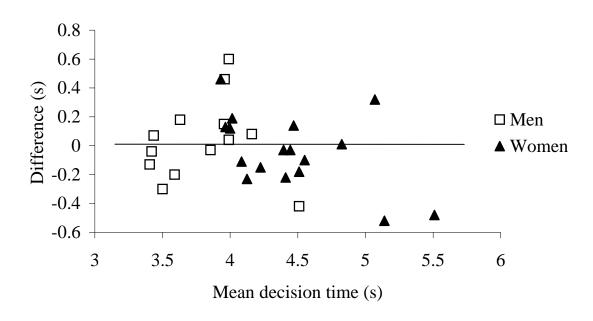
# Figure Captions

- 340 Figure 1. Schematic representation of the field hockey skill test.
- 341 Figure 2. Bland-Altman plot for the overall time raw data.
- 342 Figure 3. Bland-Altman plot for the 'decision making' time raw data.
- 343 Figure 4. A residuals versus predicted plot for the overall time for men and women.





369 Figure 2



371 Figure 3

