Free throw shot proficiency in basketball – a probability model

Czyż S.H.[1], Marzec J.[2], Styrkowiec P.[3],

[1]Department of Sport Didactics, University School of Physical Education in Wrocław, Poland and Department of Cognitive Psychology and Individual Differences, Warsaw School of Social Sciences and Humanities, Campus in Wrocław, Poland

[2]Department of Econometrics and Operations Research, Cracow University of Economics, Poland

[3]Institute of Psychology, University of Wroclaw, Poland

Corresponding author:

Stanisław Czyż

stanislaw.czyz@awf.wroc.pl

Submission date: …………………… 2012

Abstract

It has been claimed that extensive amounts of basketball practice results in the emergence of especial skills. We aimed to determine whether especial skill emerges as a result of other predictors such as age, years of experience, efficiency in shot performance, etc. We also tested a hypothesis regarding visual dependency in especial skills effect.

Ten male basketball players took part in two different experimental conditions: in normal and blurred vision conditions. Players performed free throw shots from 7 distances including shots from the free throw line (4.57 m). We detected especial skills using previously reported statistical methods, however this was not the case when analyzing data with the probit model.

Introduction

The relationship between extensive practice and the emergence of especial skills has been addressed in a number of recent studies (Breslin, Hodges, Kennedy, Hanlon, & Williams, 2010; Keetch, Lee, & Schmidt, 2008; Keetch, Schmidt, Lee, & Young; 2005; Simons, Wilson, Wilson, & Theall, 2009). Keetch et al. (2005) for the very first time showed that massive amount of practice may lead to the development of especial skills, a skill that has a unique place within the class of movements executed by one General Motor Program (GMP). Studying basketball players who were estimated to have already performed thousands of shots from the three throw line, Keetch et al. observed that shot proficiency from this standard distance, i.e. 4.57 m (15 ft), was higher than predicted by regression line computed for several shooting distances excluding 4. 57 m.

Keetch et al. (2005) proposed a number of possible explanations for the phenomenon of especial skills (for review see: Breslin, Schmidt & Lee, 2012). One explanation relates to Thorndike's identical elements theory (Thorndike, 1911, 1914, 1932) and refers to specificity of practice whereas another, reflects generality as one of the assumption of Schmidt's schema theory (Schmidt, 1975). They verified also hypothesis about context dependency, and having founded none concluded that the most suitable explanation is parameterization, in which constant perceptual cues are linked with specific parameters, and while executing movement, automatic mechanism select appropriate ones. Keetch et al. (2005, Exp. 2) studied context dependency by covering the shooting area of the floor. It could be assumed that more visual cues are available from the basket hoop and the backboard than from the floor of the court. Parameterization was also supported in studies by Breslin et al. (2010) and Simons et al. (2009).

All previous papers focused on very experienced players and in all of them, authors used linear regressions, either straight lines (Breslin et al, 2010; Keetch et al, 2005) or parabola

(Simons et al., 2009), to explain the phenomena of especial skills. Therefore, it was of interest to look at a broader context of especial skills including other possible predictors such as general shot proficiency, commonly referred as "hot hand" (Burns, 2004; Gillovitch, Villon, and Tversky, 1985; Rao, 2009), playing position, number of games played, time spent on the court, or simply the height of basketball players, experience and age of the participants into the probability model, based on Bernoulli distribution. Probit or logit models appear to be useful in such applications, therefore our manuscript focuses on the methodological context of the especial skills phenomenon.

Statistical model – form, estimation and inference

Probabilistic model

We present statistical analysis of individual-level data on the behavior of individuals. Observed variable has binomial distribution with just one trial: miss or successful shot. Its value is y_{ijgt}, where i indicates the number of experiment (i=1,…,75), j=1 or 0 indicates shots performed with or without glasses, g indicates the distance of shots (g=1,…,7) and l is the number of a player (l=1,…,10). Total number of observation y_{ijgt} was 10500 (75 shots*2conditions*7 distances*10 players). Each of player was described by other predictors (independent variables): number of years of training, number of games played last season, mean time (min) played, mean number of points scored in games in the last season. Additionally, we included shot proficiency (%) for 2 and 3 points as well as free throw proficiency for the last season (total percentage for last season). It is worth mentioning that all of these factors only characterize players and are constant while shooting with or without glasses from each of the seven shooting distances. Observed data set is a group data (Gourieroux, 2000), i.e. for different values of endogenous variable identical combination of exogenous variable is noticed. Grouped data are obtained by observing the response of N_t

individuals all of whom have the same set of explanatory variables. Such data may be obtained by aggregation of individual observation. This type of grouped data is called many observations per cell (Cameron and Trivedi, 2005). The objective of our study was to determine the effect of a shot (miss or success) performed by a player from a certain distance with or without glasses. Our statistical model based on observation y_{it} obtained for T=140 objects in experiments i=1,…,75 (then N_t=75 for every t). Dependant variable y_{it} may have two values, y_{it}=1 means successful shot from a certain distance with or without glasses whereas y_{it} =0 means miss shot, for t=1,…,T. Therefore each observation y_{it} is treated as a single draw from a Bernoulli distribution. The data are discrete, so in which case nonlinear methods as probit and logit models are used.

Probabilistic assumption and model

In the literature (e.g. Greene, 2003, Cameron and Trivedi 2005) binary outcome models have hierarchical representation and one of the equations is regression equation for latent (or unobserved) variable z_{it} (i=1,…,75, t=1,…,140) which is interpreted as individual's utility of two choices between 1 and 0. In our study, this variable represents shooting skills from different distances with or without glasses. If these skills are higher than the threshold value, a successful shot will be recorded, otherwise a shot will be unsuccessful. In an individual observation, binary choice model adequate to our study will have following form (see: Amemiya, 1981, 1985; Gourieroux, 2000):

$$z_{it} = x_t \cdot \beta + \varepsilon_{it}$$
$$y_{it} = \begin{cases} 1 & \text{gdy} \quad z_{it} \geq 0 \\ 0 & \text{gdy} \quad z_{it} < 0 \end{cases} \tag{1}$$

where x_t is a vector of k-variable values for an object t, and a vector of k×1 unknown parameters β. In this case, x_t represents player's characteristic and information about distance and vision conditions (normal – without glasses and blurred – with glasses). Additionally, it includes intercept which represents a constant term in regression for z_{it}.

Random components ε_{it} are independent on both indexes and considering an identifiability of model parameters (1) and convenience of their estimation it is assumed that has mean zero and fixed (known) variance.

Different distributions for ε_{it} lead to different binary outcome models. The most common models are logit and probit, that assume, respectively, disturbances are standardized logistic distributed and standard normal distributed.

For symmetric distributions probability of success is equal $\Pr(y_{it} = 1) \equiv p_{it} = F(x_t \cdot \beta)$, where $F(a)$ is an cumulative distribution function in point a, one of the probability distribution for assumed ε_{it}. Estimation of such a model (1) is usually done by a maximum likelihood method (ML), a one-step procedure. The logit and probit log likelihoods are globally concave and hence relatively easy to maximize using the Newton-Raphson algorithm. The estimator of an asymptotic covariance matrix of the maximum likelihood estimator is quite simple:

$$\hat{V}\left(\hat{\beta}_{MNW}\right) = \left[\sum_{t=1}^{T} \frac{N_t \cdot \hat{f}_t^{\,2}}{\hat{F}_t \cdot \left(1 - \hat{F}_t\right)} x_t' x_t\right]^{-1}, \tag{2}$$

where $\hat{F}_t = F\left(x_t \hat{\beta}\right)$, a $f_t = f\left(x_t \hat{\beta}\right)$ is the probability density function of ε_{it} and $\hat{\beta}$ is the ML estimator,

If the experiment was conducted only once (N_t =1) then for each t may be obtained a binary choice model for individual data. It is worth noting that having group data (N_t >1) causes better precision of parameter estimation compared to the lower number of individual data (see equation (2)) for N_t =1.

In our study we use model (1) i.e. probit model. Parameter estimation was undertaken using maximum likelihood method.

Model description

Hypotheses

We advanced six hypotheses which were verified in our probabilistic model:

1. Increased throwing distance decreases the probability of a successful shot (H1). This hypothesis was based on Schmidt, Zelaznik, and Frank's (1978) conclusion that increase in throwing distance results in a linear increase in the variability of the aimed movement.
2. Players shooting in glasses, which deteriorate player's vision, will decrease their shot proficiency (H2). There is a continuous and integrated relationship between perception and movement outcome (see e.g. Montagne, Laurent, Durey, and Bootsma, 1999) therefore, we may expect a decrease in shot proficiency while shooting in glasses.
3. There are some predictors which substantially influence shot proficiency, for example years of training, shot proficiency in played games, etc. all of which positively influence probability of successful shots (H3). As noted in previous studies (Keetch et al, 2005; Simons et al., 2009; Breslin et al., 2010), a significant amount of practice may be the crucial predictor of especial skills. However, more predictors may be included into the model.
4. Probability of a successful shot performed from the free throw line (4.57 m) should be higher compared throws from other distances (H4). This should hold true for both conditions – throws performed by players with or without glasses. This hypothesis is based on parameterization explanation (see Breslin et al., 2012).
5. There is an unknown distance (*ex ante*) from which the probability of successful shots is the highest and that distance is not necessarily the closest distance to the basket rim (2.74 m) or free throw line (4.57 m) (H5). There may be at least two possible distances from which the probability of successful shots will be the highest: the closest distance

(Schmidt et al., 1978), or free throw distance 4.57 m, which is the highly trained distance, and is related to especial skills (Keetch et al., 2005). However, we assumed that there might be a different distance, unknown before analysis, neither 4.57 m nor 2.74 m.

6. If the parameterization hypothesis about emergence of special skills advanced by Keetch et al. (2005) is correct, then in conditions of throwing in glasses deteriorating visual acuity shot proficiency at 4.57 m should decrease less than at other distances (H6).

Predictors

In a regression, omission of a variable is often the first example of inconsistency of many estimators. Too many regressors cause little harm, but too few regressors can lead to inconsistency. Thus, a long list of potential explanatory variables was examined. We assumed that the following predictors (independent variables) may characterize each player: number of years of training (w_3), number of games played in last season (w_4), mean time spent per game in the last season (in minutes) (w_5), mean points per game in the last season (w_6), shot proficiency for 2-points (percent) (w_7), shot proficiency for 3 points (percent) (w_8), shot proficiency for three throws for the last season (percent) (w_9); see (Table 1). Other regressors were considered, such as height of a player and playing position. However, both predictors caused co-linearity problems and had small variability. Moreover, playing position predictor had a nominal scale (5 values representing 5 different playing positions) what required 4 more free parameters and additional binary variables. Therefore, these two variables were excluded from the further analysis.

Two of the important determinants that may explain variability of shot proficiency may be the shooting distance and glasses worn or not while shooting. Following Keetch et al.'s procedure

(2005) we set 7 shooting distances: 2.74 m (9 ft), 3.35 m (11 ft), 3.96 m (13 ft), 4.57 m (15 ft), 5.18 m (17 ft), 5.79 m (19 ft) and 6.4 m (21 ft).

Model specification

To verify advanced hypotheses we constructed the following regression model for a latent variable z_{it}:

$$z_{it} = \beta_0 + \sum_{j=1}^{9} \beta_j \cdot w_{tj} + \beta_{10} \cdot I_{(4,57)}(w_{t1}) + \beta_{11} \cdot (w_{t1})^2 + \beta_{12} \cdot I_{(1)}(w_{t2}) \cdot I_{(4,57)}(w_{t1}) + \varepsilon_{it}, \quad (3)$$

where $I_\Omega(\omega)$ is a indicator function of a data set i.e. if $\omega \in \Omega$, than $I_\Omega(\omega)=1$, otherwise $I_\Omega(\omega)=0$. Descriptive variables are included in Table 1. We verified our hypothesis using equation (3).

Table 1. Variables description.

Variables description	Variables	Parameters	Hypotheses	Parameters sign
Shot proficiency at 4.57m ($\bar{p}_t$)	Dependent	———	———	———
Shooting distance (w_{t1})	Independent	β_1	H1	–
Wear glasses (yes: w_{t2}=1 or no: w_{t2}=0)	Independent	β_2	H2	–
Years of training (w_{t3})	Independent	β_3	H3	+
Number of games player in last season (w_{t4})	Independent	β_4	H3	+
Mean time played in last season (minutes per game) (w_{t5})	Independent	β_5	H3	+
Mean points per game (w_{t6})	Independent	β_6	H3	+
2-points shot proficiency (w_{t7})	Independent	β_7	H3	+
3-points shot proficiency (w_{t8})	Independent	β_8	H3	+
Free throw shot proficiency (w_{t9})	Independent	β_9	H3	+
Shot from 4.57m? (yes: w_{t10}=1, if w_{t1} =4.57m or no: w_{t10}=0)	Independent	β_{10}	H1 and H4	H4: +
(Distance)^2 $w_{t11} = (w_{t1})^2$	Independent	β_{11}	H1 and H5	H5: –
Shot from 4.57m in glasses? (yes: w_{t12}=1, if w_{t2}=1 and w_{t10}=1, or no: w_{t12}=0)	Independent	β_{12}	H1, H2, andH6	H6: +

In equation (3) it was assumed that most of the predictors affect linearly for z_{it}. Therefore, parameters β_3, β_4, β_5, β_6, β_7, β_8, and β_9 inform how strong the influence of predictors are in characterizing each player for shot proficiency (hypothesis H3). We may assume that these parameters should be positive if statistically significant, and furthermore, two predictors (w_{t1}) and (w_{t2}) are not linear.

If we assumed that the dependence of probability of successful shots and throwing distance is unknown, we could expect that this probability decreases while distance increases. [Although, it would be necessary to estimate whether this probability decreases proportionally (linearly), decreases less than proportionally or more than proportionally]. In the general case this dependence can be described by a function that is neither convex nor concave. Sigmoid function is an excellent example of this form, because it has a concave part and a convex part (see e.g. S-shaped curve in Rao, 2009).

Individual-level data are at a low level of aggregation. Cameron and Trivedi (2005, p. 5) remarked that in this case "in many, if not most, cases linear function forms turn out to be simply inappropriate". A disaggregation brings to the heterogeneity of individuals that should be properly controlled (modeled). Many variables reflect inter-individual heterogeneity, for example, number of years of training. There are also differences in individual motivation and ability and so forth, which are not observed. Heterogeneity plays a very importer role in this analysis differences in individual motivation and ability are not observed. This unobserved heterogeneity is controlled for in probit regression model.

From a mathematical point of view, if we don't know the analytical form of the regression model, then under certain assumptions, it can be approximated by a polynomial of degree n over a small interval. The higher the n, the better the approximation. Another distinct statistical problem is related to the method of approximation and its accuracy. In this case, Taylor series expansions and approximations have several advantages. One of them is that it is linear in their parameters, what make the estimation easier. On the other hand, non-linearity regarding predictors (independent variables) enriches interpretation characteristic of the model. The simplest universal statistical model used for description of dependency between variables is multiple regression. However, more complex non-linear models are used, especially so-called Flexible Functional Forms. The most frequently used functional form is translog function, which is a second order (all cross-terms included) log-linear form. This function fit data much better than linear function (e.g. Cobb-Douglas specification).

Methodological and empirical arguments for using such a function have been aroused in 70. in the XX. century (see: Christensen, Jorgenson, and Lau, 1973; Fare and Sung,1986; Diewert, 2002). Approximations of higher order are not very useful. Occasionally, cubic function is used but only for one predictor. Advantages of non-linear function were also noted in medical and sport sciences. For example, Hosmer and Lemeshow (2000) discussed a criteria for including a variable in a logit model and noted that (p. 97) "(…) for continues variables we should check the assumption of linearity in the logit model (…)", because "(…) assuming linearity in the logit at the variable selection stage is common practice". As an alternative they proposed methods of fractional polynomials, developed by Royston and Altman (1994); see also Royston and Altman (1997). We wish to determine what value of w^p (w>0) yields the best model for covariates. Royston, Altman (1994) proposed restricting the power to be among those in the set {-2, -1, -0.5, 0, 0.5, 2, 3}, where p =0 denotes the log of the variable. This Box–Cox power transformation a variable is a useful method, because including it alleviate heteroscedasticity. It is a device for generalizing the linear model, used in many sport researches (e.g. Hamrick and Rasp, 2011; Wimmer, Fenske, Pyrka, and Fahrmeir, 2011).

Considering the model presented in his paper, it is worth noting that quadratic function is a natural and economical generalization of a linear model. Of course, it does not exclude the situation that statistical data can be well explained by a linear model. In such case, estimation of parameter β_{11} will be statistically insignificant or estimations of parameters β_1 and β_{11} will result in values on the right arms of the parabola, approximation linear function. The relevant question is whether this quadratic model is significantly better than a linear model. We can answer this question by verifying the hypothesis β_{11}=0. In our model we used also hyperbolic function or sigmoid function, which is nonlinear in their parameters, what caused numerical complication. Ex post analysis showed that according to the Akaike information criterion (AIC) quadratic approximation is better than hyperbolic or sigmoid function (see equation (1)). Therefore, both were excluded from further analysis.

Considering probability of a successful shot as a function of these two predictors (distance and wearing glasses) while all of the other predictors are held constant, we get that:

$$p_t = F\left(\beta_{11}(w_{t1})^2 + \beta_1 w_{t1} + \beta_{10} \cdot I_{(4,57)}(w_{t1}) + \beta_{12} \cdot I_{(1)}(w_{t2}) \cdot I_{(4,57)}(w_{t1}) + \beta_2 w_{t2} + c_0\right), \qquad (4)$$

where c_0 is constant.

Negative parameter β_1 and restriction $\beta_{10}=\beta_{11}=\beta_{12}=0$ makes hypothesis H1 more probable i.e. increasing distance decreases probability of a successful shot. Additional restriction $\beta_{10}=\beta_{11}=\beta_{12}=0$ are sufficient to keep strong negative correlation, this is a situation in which a probability of successful shot at 4.57m is higher than from closer to the basket distances. Therefore restriction β_{11}=0 is weaker and necessary, if H1 is true. Such a definition of variables, where w_{t2}=1 means blurred vision conditions (wearing glasses) implies that negative β_2 is a necessary stipulation if H2 is true – wearing glasses while shooting substantially decrease probability of successful shot. On the other hand, positive β_{10} reflect the assumption that throws fromthe free throw distance increases the probability of success.

In H5 a decisive parameter is β_{11}, if it is positive it means that there is a value of w_{t1} which minimalize p_t, whereas negative maximalize p_t. It was expected that β_{11} will be negative and as a result, the highest shot proficiency will be at the closest to the basket distance 2.74 m or the 4.57 m. Hypothesis H5 will be therefore accepted, if β_{11} is negative. It has to be noticed that hypotheses H1 and H4 are specific cases of H5, and of interest would be finding out which effect is dominant.

Hypothesis H6 is a conjunction of hypotheses H2 and H4. There are three separate situations which fully describe H6 as a product of variables w_{t1} and w_{t2}.

Verification of hypothesis H6 is a) a shot from a distance without glasses (w_{t1} – any, w_{t2}=0), b) a shot in glasses at 4.57 m (w_{t1}=4.57, w_{t2}=1), c) a shot in glasses from any other distances than 4.57 m ($w_{t1}\neq$4.57 m, w_{t2}=1). As a result, we have a nominal variable with three possible values. To identify parameters in regression equation, we recognize two variables w_{t2} and w_{t12} which reflect situations a) and b), and situation c) is a reference category. If a player in glasses shoot from the 4.57m distance, his shot proficiency should decrease less than while shooting in glasses from other distances.

A positive parameter β_{12} is crucial and sufficient for confirming hypothesis H6. This parameter provides information about an increase of shot proficiency at 4.57 m distance comparing to shot proficiencies from other distance while shooting in glasses. Moreover, parameters β_2 and β_{12} are unequally related. Parameter β_2 reflects an average effect of shot proficiency at all distance. However, one of these distances is special, because it is expected that shot proficiency at this distance should be higher than from the other distance. In other words, parameter β_2 measures the effect c), that is resultant in situations b) and a). It is worth mentioning, that our model tries to verify different hypotheses, especially H4 and H6, which are a combination of main hypotheses H1 and H2. As a result, we may expect that effects of H1 and H2 will be dominant, so statistical results for this two hypotheses should be convincing.

Furthermore, in hypothesis H6 a crucial binary variable w_{t12} shows rather small variation, since only in 10 cases for 140 (about 7%) are present conditions confirming hypothesis (w_{t12}=1). Mean experimental probability of a successful shot in glasses at 4.57 m distance is 0.60 (±0.06), whereas from other distances is 0.56 (±0.18).

Verification of hypotheses H1-H6 consisted of estimation of unrestricted probit model and six additional restricted models (Table 2). The most general model M0 included all predictors, whereas models M2-M6 were computed including restrictions β=0, that is reflecting falsification of particular hypothesis. For example, model M2 assumes that shooting in blurred vision conditions (with glasses) does not impact the probability of success; model M1 assumes that increasing the shooting distance, the probability of a successful shot decreases. If model M1 is better than M0, it would mean that there are no sufficient arguments to doubt monotonic character of relation between shooting distance and shot proficiency. Otherwise, the hypothesis about existence of relation between these variables should be accepted. This relation is represented by model M1. Model M0 represents all effects simultaneously.

Table 2. Model description.

Model	Model interpretation	Restriction in relation to model M0	Number of parameters in the model
M0	Model with all effects	-	13
M1	Model with linear distance effect	$\beta_{10}=\beta_{11}=\beta_{12}=0$	10
M2	Model without glasses effect	$\beta_2=0$	12
M3	Model without players characteristic	$\beta_3=\beta_4=\beta_5=\beta_6=\beta_7=\beta_8=\beta_9=0$	6
M4	Model without especial skills effect	$\beta_{10}=0$	12
M5	Model without optimal shooting distance effect	$\beta_{11}=0$	12
M6	Model without especial skills effect in blurred conditions effect	$\beta_{12}=0$	12

The econometric methodology of general-to-specific modeling was employed for model building. The general model M0 expresses the least restrictive conviction about potential relations between predictors and dependent variable. Model reductions were done thereafter, using t-test and likelihood ratio test (LR) (Greene, 2003). Additionally, the AIC was used to compare strength of the models, which are nonnested: M2, M4 and M5 as well as M3 and

M5. Model M2, M4, M5 and M6 have same number of parameters, and therefore AIC gives the same conclusions as LR test. A lower rank model Mm (m=1,…,6) in models ranking in relation to M0 will make particular specification (and the related hypothesis) more probable. On the other hand, higher position in this ranking will reflect a lack of effect, that is the advanced hypothesis should be rejected.

Experiment 1

Participants

Ten male players (mean age 17.7, SD 2.16) participated in the experiment. Two players belonged to the club Śląsk Wrocław and the others were members of the WKK Wrocław club. Participants were seniors playing in the 2nd league at the time of the study and represented various positions on the team (e.g. point guard, shooting guard, center) and each had at least 7 years of experience in basketball shooting (mean 7.75, SD 1.9). Statistics including number of games played in the last season, mean time spent per game in the last season (in minutes), mean points per game in the last season, shot proficiency for 2-points (percent), shot proficiency for 3 points (percent), shot proficiency for three throws for the last season (percent) were collected by experimenters in an interviews and from official basketball web pages (Table 3). All participants exhibited normal visual acuity.

Table 3. Descriptive statistics of players.

i	Years of training (w_{t3})	Played games (w_{t4})	Mean time spent in games (minutes) (w_{t5})	Mean points per game (w_{t6})	Shot proficiency		
					2 points (w_{t7})	3 points (w_{t8})	Free throws (w_{t9})
1	6	8	17	4.4	42	27	50
2	4.5	16	10	2.3	44	28	60
3	11	11	6	1.3	40	25	57
4	7	24	12	1.9	50	23	44
5	8	23	25	13.6	61	30	63
6	5	14	9	2.9	58	0	63
7	10	23	21	9.7	48	27	67

8	7	15	7	0.9	43	9	56
9	9	16	8	2.8	53	30	67
10	8	24	26	6.1	47	15	71

All participants gave written informed consent to take part in the experiment and the study was approved by Wroclaw's School of Physical Education Research Ethics Committee.

Apparatus and Procedure

Procedure was based on methods reported in Keetch et al. (2005, *Exp*. 1). Participants performed set shots from seven different distances: 2.74 m (9 ft), 3.35 m (11 ft), 3.96 m (13 ft), 4.57 m (15 ft), 5.18 m (17 ft), 5.79 m (19 ft), 6.4 m (21 ft) from the spot on the floor directly under the front edge of the backboard. According to FIBA regulations, the 4.57 m line on the basketball court is the foul line (free throw line). The throwing distances were marked on the floor with a strip of masking tape 3 cm wide and 5 m long. All throwing positions were placed on the straight line going from the backboard toward the centre of the court. Set shots were taken with an official basketball, normally used in matches of the 2nd league. The rim to which players were aiming was mounted at the regular height of 3.05 m (10 ft).

Participants performed their shots like regular free throws in basketball and without any pre-shot routine (e.g. dribbling the ball). Shots were taken by players with their preferred limb and with their feet maintaining contact with the floor at all times. Participants were asked to keep their feet as close as possible to, but not on, the marking tape while performing the shot.

Each participant performed 1050 shots in total on four consecutive days of testing (approximately 250 – 275 – 250 – 275 shots). There were 75 shots for each distance performed in two throwing conditions – in the first one players were throwing without eyeglasses (normal vision), in the second (blurred vision) participants while throwing wore eyeglasses used by people suffering from hyperopia (eyeglasses with dioptric value +4.00). It is expected that the average person's visual acuity will drop to approximately 6/60 (Snellen fraction), which constitutes legal blindness (in Australia, Great Britain, USA) if a +3.00 over-refraction lenses or glasses are used (Mann, Ho, De Souza, Watson, and Taylor, 2007). Mann et al. (2007) reported that only a +3.00D over-refraction is required before any significant decrease in movement performance is achieved, thus using glasses +4.00 assured that players would have limited accessibility to visual cues.

All participants performed set shots without eyeglasses on the first two days of testing and with eyeglasses on the subsequent two days. Before each throw an experimenter announced the shooting distance and the shot number. The shots were taken with 5-second rest intervals between trials, with a quasi-random order such that no more than two shots were performed from the same distance on consecutive trials. After a shot was taken, the first experimenter retrieved the ball and handed it back to the participant. All shots were recorded with the use of Sony Handycam DCR SR50 camera. Participants were encouraged to perform each throw with the same level of effort and desire to score the shot.

The second experimenter assessed scoring shot accuracy and recorded the results on the mark sheets. Outcome scores were assessed on each trial using a 2-point scoring system. This type of coding system was also reported by Keetch et al. (2005) in experiment 1.

One point was given for a successful shot whereas zero point was awarded for unsuccessful shot.

Preliminary analyses

Following the previous study, linear regression was computed for shots performed in normal visual context and in blurred vision (shooting in glasses). For computing linear regression we used shot proficiency at all shooting distances but 4.57 m. The special skills were noted while players performed shooting in normal vision conditions (empirical mean at 4.57 m=71.87±5.74; predicted 4.57 m=65.53±5.23; one tailed $t=4.81$, $df=9$, $p<0.0001$; **Figure 1**).

The special skills were also present when we compared empirical (mean=60.39±6.53) and predicted (mean=55.79±7.59) shot proficiency at 4.57 m in blurred vision conditions (one-tailed t=2.05; df=9, p=0.035; **Figure 2**).

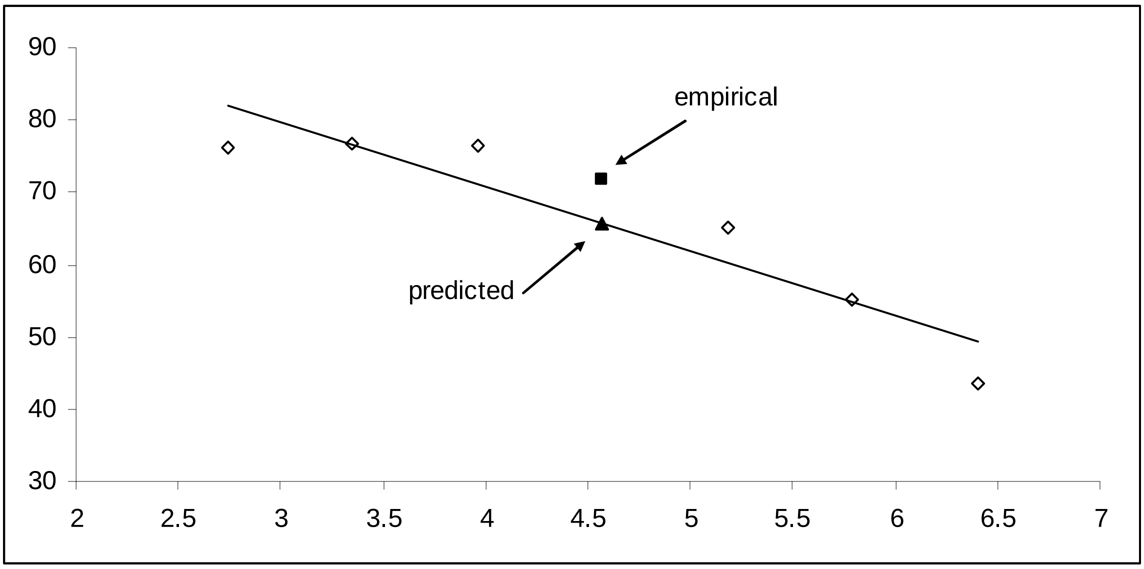

Figure 1 Linear regression computed for shot proficiency in normal vision conditions.

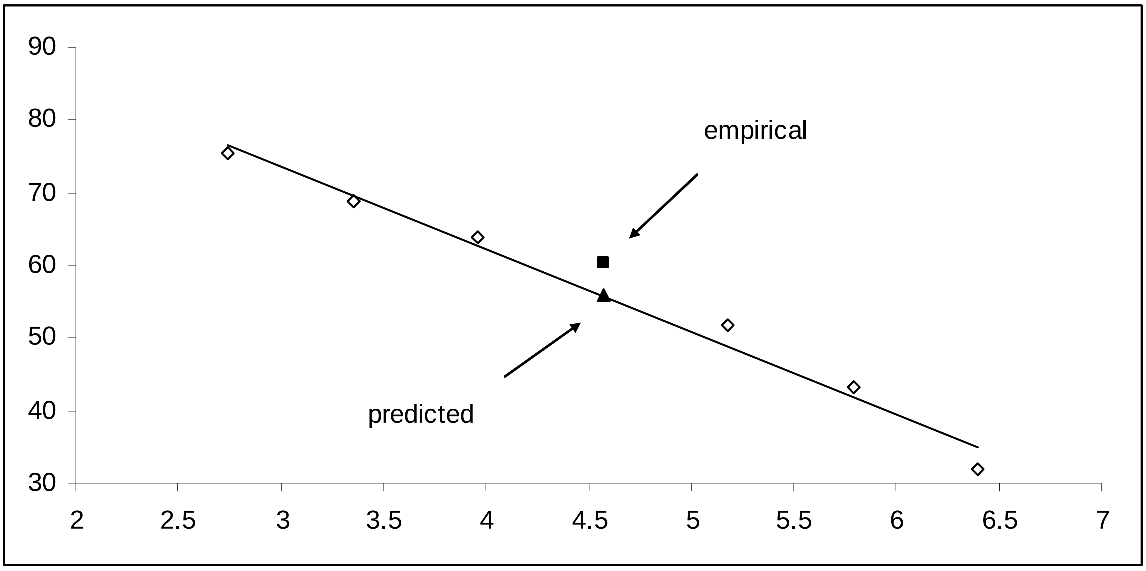

Figure 2 Linear regression computed for shot proficiency in blurred vision conditions.

Table 4. Probability and SD of successful shots for individual players in blurred and in normal vision conditions.

i	Pr (y_{it}=1)	herein	
		Without glasses (w_{t2}=0)	With glasses (w_{t2}=1)
1	0.57 ±0.11	0.61 ±0.10	0.52 ±0.09
2	0.70 ± 0.08	0.72 ±0.08	0.68 ±0.07
3	0.60 ± 0.11	0.66 ±0.09	0.53 ±0.09
4	0.53 ±0.18	0.61 ±0.14	0.45 ±0.19
5	0.62 ±0.14	0.65 ±0.12	0.60 ±0.16
6	0.54 ±0.21	0.59 ±0.18	0.48 ±0.22
7	0.64 ±0.21	0.70 ±0.19	0.59 ±0.21
8	0.62 ±0.17	0.68 ±0.13	0.56 ±0.18
9	0.64 ±0.14	0.66 ±0.12	0.61 ±0.15
10	0.69 ± 0.16	0.76 ±0.14	0.62 ±0.16

Table 5. Mean values and SD of probability of a successful shot from different distances

Distance in meters w_{t1}	Pr(y_{it}=1)	herein	
		Without glasses w_{t2}=0	With glasses w_{t2}=1
2.74	0.76 ±0.08	0.76 ±0.09	0.75 ±0.07
3.35	0.73 ±0.10	0.77 ±0.08	0.69 ±0.10
3.96	0.70 ±0.11	0.77 ±0.06	0.64 ±0.10
4.57	0.66 ±0.08	0.72 ±0.05	0.60 ±0.06
5.18	0.58 ±0.11	0.65 ±0.08	0.52 ±0.10
5.79	0.49 ±0.12	0.55 ±0.07	0.43 ±0.12
6.40	0.38 ±0.13	0.43 ±0.10	0.32 ±0.13

Probit model analysis

Results estimation and testing from the discrete choice model perspective

Parameter estimation and estimation errors were stable while comparing different models (detailed estimation information is presented in 6). Only absolute terms showed some co-linearity between an intercept and the variables w_{t10} and w_{t12}. Supplementary, tables 7 and 8 include results of likelihood ratio tests and ranking according to AIC. Tests results are satisfactory and the ranking reliable. Logit model yields equivalent results to probit.

It was advanced in hypothesis H1 a strong monotonic relation between shot proficiency and a shooting distance. We assumed that increasing distance from the basket would have resulted in decrease of shot proficiency. Falsification of this hypothesis was determined in models M0 and M1. In M0, estimation of parameter β_1 is significantly negative, since there is a small standard error of estimation in relation to estimated coefficients. Hypothesis that β_1<0 was confirmed in analysis, however, it was confirmed that β_{10}=β_{11}=β_{12}=0, thus an advantage of model M1 over M0 was not shown. The likelihood ratio statistic for both models was 44.08 with the critical value of 7.81 and p-value around 10^{-8}. The AIC also prefers M0 over M1, since M1 is only 5th in the ranking of all models (Table 8). As the result hypothesis H1 failed with hypothesis H5, which means that the effect of “optimal distance” is stronger than “distance” effect.

Hypothesis H2 claimed that blurred vision conditions (wearing glasses) should decrease shot proficiency and it was supported in all models. The value of parameter β_2 is negative, stable and very precisely estimated. Model M2 which is obtained by restriction β_2=0 in model M0, When removed from the general model M0, we received model, took the 6th place in the ranking of all 7 models (see table 8).

The LR statistic of parameters β_3, β_4, β_5, β_6, β_7, β_8 , and β_9 confirm hypothesis H3. Testing model M3 against M0 LR is 134, critical value 12.6 at p-value is lower than 10^{-25}. Model M3, in which was assumed that aggregate influence of these variables is not significant is the last model in the AIC ranking. Moreover, in general model M0, t-test values for individual parameters are higher than 2. It may be concluded therefore, that the influence of these variables are significant and hypothesis H3 was confirmed.

Hypothesis H4 about special skills at 4.57 m was not confirmed. Although, estimation of parameter β_{10} is positive but statistically not significant in models from first (M6) and third (M0) place in the AIC ranking. Reduction of the M0 models to model M4 is therefore legitimate. The p-value (≈0.38) for LR test indicates no sufficient evidence for existence of „special skills”. The “special skill” effect is present provided that in analysis will be omitted effects of distance (model M2) or “optimal distance” (model M5). However, both of these models (M2 and M5) are not very probable, they were ranked 6th and 4th in the AIC ranking accordingly.

Hypothesis H5 advances that there is a certain optimal distance from which the probability of a successful shot is the highest. It is worth noticing, that hypotheses H1 and H4 were not strongly supported by data, it was reasonable to expect that H5 may be confirmed.

The estimation of parameter β_{11} is negative in all models. Call in question the effect optimal distance present in estimation results in model M5, is not is not authorized in the data. LR test for model M5 against M0 strongly support model M0. Our data strongly confirmed hypothesis H5. One could therefore set an optimal shooting distance which maximizes the probability of a successful shot. However, falsification H5, the distance of 4.57 m was excluded as the "optimal distance" and it was necessary to compute additional equation:

$$\hat{z}_{it} = \hat{\beta}_{11}(w_{t1})^2 + \hat{\beta}_1 w_{t1} + \text{constant}, \qquad (5)$$

where $\hat{\beta}_{11}$ and $\hat{\beta}_1$ represent estimation of the values and $\hat{z}_{it}$ represents theoretical ability to shot successfully computed on estimated model. Based on results of estimation of model M0, optimal distance maximizing the probability of successful shot was computed. Considering models M0, M4, and M6, the optimal distance was 2.37 m, 2.49 m, and 2.37 m, respectively, and it was shorter than the closest shooting distance in our experiment (2.74 m).

Verification of hypothesis H6 consisted of examining whether $\beta_{12}>0$. In all five models, in which the effect of "especial skills in blurred vision conditions" (shot proficiency at 4.57 m in glasses) was included, this parameter was negative, and not different from zero when performing t-test. The model in which variable w_{t12} was excluded, turned out to be the best, according to the AIC. Testing models M6 over M0 (LR test) showed that it is better to remove this variable (high p-value>0.53). Hypothesis H6 was not therefore confirmed. The effect of "special skills in blurred vision conditions" could have been confirmed, provided that two other effects had been excluded: optimal distance (H5) and special skills, i.e. effect of 4.57 m (H4). In this case, parameter β_1 was negative, which is discrepant with our intuition and earlier study (e.g. Keetch et al., 2005).

Table 6. Estimation values of profit models. Symbol „-” means that variable was excluded from the model. Critical value of t-test ≈1.98, at p=0.05 (two tailed test). Symbol „*” means that test was not significant at p=0.05, Ln L () – log likelihood value.

		Model 0		Model 1		Model 2		Model 3		Model 4		Model 5		Model
Variable	Parameter	Estimate ± error	t ratio	Estimate ± error	t ratio	Estimate ± error	t ratio	Estimate ± error	t ratio	Estimate ± error	t ratio	Estimate ± error	t ratio	Estimate ± error
„1”	β_0	0.277* ±0.169	1.64	1.513 ±0.115	13.15	0.153*±0 .175	0.88	0.482 ±0.119	4.04	0.217* ±0.165	1.32	1.486 ±0.116	12.81	0.280* ±0.169
Shooting distance (m) (w_{t1})	β_1	0.296 ±0.057	5.17	-0.278 ±0.006	49.62	0.287 ±0.059	4.84	0.292 ±0.056	5.26	0.328 ±0.054	6.12	-0.276 ±0.006	49.04	0.296 ±0.057
Wear glasses (w_{t2})	β_2	-0.273 ±0.015	17.82	-0.278 ±0.014	20.33	-	-	-0.270 ±0.015	18.15	-0.279 ±0.015	18.83	-0.271 ±0.015	18.25	-0.280 ±0.014
Years of training (w_{t3})	β_3	-0.039 ±0.005	8.36	-0.039 ±0.004	8.65	-0.039 ±0.005	8.06	-	-	-0.039 ±0.005	8.36	-0.039 ±0.005	8.60	-0.039 ±0.005
Number of games played in last season (w_{t4})	β_4	0.014 ±0.002	7.10	0.014 ±0.002	7.40	0.013 ±0.002	6.83	-	-	0.014 ±0.002	7.10	0.014 ±0.002	7.34	0.014 ±0.002
Mean time played in last season (minutes per game) (w_{t5})	β_5	-0.008 ±0.003	3.10	-0.008 ±0.003	3.23	-0.008 ±0.003	3.03	-	-	-0.008 ±0.003	3.10	-0.008 ±0.003	3.22	-0.008 ±0.003
Mean points per game(w_{t6})	β_6	0.013 ±0.006	2.33	0.013 ±0.005	2.45	0.013 ±0.006	2.24	-	-	0.013 ±0.006	2.32	0.013 ±0.005	2.43	0.013 ±0.005
2-points shot proficiency (w_{t7})	β_7	-0.019 ±0.002	9.92	-0.019 ±0.002	10.27	-0.019 ±0.002	9.49	-	-	-0.019 ±0.002	9.92	-0.019 ±0.002	10.21	-0.019 ±0.002
3-points shot proficiency (w_{t8})	β_8	0.006 ±0.001	6.57	0.006 ±0.001	6.74	0.006 ±0.001	6.35	-	-	0.006 ±0.001	6.57	0.006 ±0.001	6.71	0.006 ±0.001
Shot proficiency at 4.57m (w_{t9})	β_9	0.018 ±0.001	16.86	0.018 ±0.001	17.44	0.018 ±0.001	16.19	-	-	0.018 ±0.001	16.87	0.018 ±0.001	17.32	0.018 ±0.001
Shot from 4.57m? (w_{t10})	β_{10}	0.050* ±0.031	1.62	-	-	0.190 ±0.031	6.21	0.049 ±0.030	1.66	-		0.157 ±0.028	5.65	0.025* ±0.023
(Shooting distance)^2 w_{t11}	β_{11}	-0.062 ±0.006	10.01	-	-	-0.061 ±0.006	9.48	-0.062 ±0.006	10.19	-0.066 ±0.006	11.30	-	-	-0.062 ±0.006
Shot from 4.57m in glasses? (w_{t12})	β_{12}	-0.046* ±0.041	1.14	-	-	-0.319 ±0.039	8.23	-0.045 ±0.039	1.16	-0.003* ±0.030	0.09	-0.048* ±0.039	1.22	-
Ln L(β; y)	-	-6 509.62		-6 531.66		-6 558.33		-6 576.67		-6 510.01		-6 524.83		-6 509.82

Table 7. LR test results.

Hypotheses	LR statistic	Number of restriction	Critical value for χ^2 - test (α=0.05)	p-value
M1 over M0	44.08	3	7.81	$<10^{-8}$
M2 over M0	97.41	1	3.84	$<10^{-22}$
M3 over M0	134.10	6	12.59	$<10^{-25}$
M4 over M0	0.78	1	3.84	0.377
M5 over M0	30.42	1	3.84	$<10^{-7}$
M6 over M0	0.40	1	3.84	0.525

Table 8. Ranking of models according to AIC.

Model	Model interpretation	Number of parameters	AIC	Rank
M0	Model with all effects	13	13 045.24	3
M1	Model with linear distance effect	10	13 083.32	5
M2	Model without glasses effect	12	13 140.66	6
M3	Model without players characteristic	7	13 167.35	7
M4	Model without especial skills effect	12	13 044.02	2
M5	Model without optimal shooting distance effect	12	13 073.66	4
M6	Model without especial skills effect in blurred conditions effect	12	13 043.65	1

Discussion

We did not confirm hypothesis H1 regarding decreased shot performance when the shooting distance was increased. Hypothesis H2 regarding the detrimental influence of blurred vision conditions on shot performance was confirmed. Similarly, the influence of player's characteristics on shot performance was also confirmed (H3). Probit analysis did not confirm that shot proficiency at 4.57 m is significantly better than at other distances, either with or without glasses (H4). However, the optimal shooting distance (H5), the distance from which the shot proficiency should be the highest was computed at 2.37, 2.49 and 2.37 m. These distances were closer to the basket rim than any shooting distances recognized in our experiment. Hypothesis H6 about emergence of special skills in blurred vision conditions was not confirmed.

Analysis of the probit model confirmed the parameterization hypothesis (Ketch et al, 2005; Breslin et al., 2010). It has to be pointed that in blurred vision conditions special skills were noted using Keetch et al.'s (2005) method however this effect was rather weak (p=0.035 for one-tailed t-test). The lack of special skills in blurred vision conditions either in preliminary analysis (line plus one data point) or in the probit model confirmed that constant perceptual cues are linked with specific parameters. It should however be underlined that this result is a little confusing. If we claim that there were no special skills in blurred vision conditions noted while using Keetch' et al. statistical method, we may assume that the conclusion about the lack of context dependency (Keetch et al. 2005) is not necessarily applicable in our case.

However, it may be very interesting why there especial skills were absent when analyzing the probit model, whereas they were noticed when using Keetch et al.'s analysis. Both approaches have quite different foundations. The goal of this analysis was to identify and estimate fundamental parameters that characterize the relationship between basketball shooting efficiency and many explanatory variables, often called the structural approach. Structure was defined as the set of relations describing human ability or behavior. In the above model, many hypotheses were tested in an attempt to explain the mechanism behind these effects. Formally, this approach is consistent with the principles of statistical inference.

We use a microdata which may be quite noisy. On the other hand, the process of aggregation leads to smoothing with many movements in opposite directions canceling in the course of summation. Cameron and Trivedi (2005, p. 5) remarked that "A relationship between two variables a micro level may be piecewise linear with many nodes. After aggregation the relationship is likely to be will approximated by smooth function." Here you can look for the causes of negative conclusions.

Some limitations of our study have to be emphasized: a small group of participants and a different sample than in the studies by Keetch et al. (2005) and Breslin et al. (2010).

Basketball players in these studies were recruited in North America while in our experiment Polish players were used. It is obvious that the training system may be different in both cases.

In summary, our study has shown that different statistical approach may result in quite contradictory conclusions about the absence or presence of special skills and its causes. However, we encourage future research to apply different statistical models to assess its applicability.

References

Amemiya, T. (1981). Qualitative Response Models: A Survey. *Journal of Economic Literature*, 19, 1483-1536.

Amemiya, T. (1985). *Advanced Econometrics*. Harvard University Press: Cambridge (Massachusetts).

Breslin, G., Hodges, N., J., Kennedy, R., Hanlon, M., & Williams, A., M. (2010). An especial skill: Support for a learned parameters hypothesis. *Acta Psychologica*, 134, 55–60.

Breslin, G., Schmidt, R.A., Lee, T. (2012). Especial Skills: Generality and specificity in motor learning, In Skill Acquisition in Sport, Research Theory and Practice (2nd Edition) Edited by Williams, A, M., & Hodges, N, J. (2012) Routledge, London.

Burns, B. (2004). Heuristics as beliefs and as behaviors: The adaptiveness of the hot hand. *Cognitive Psychology*, 48, 295-331.

Cameron A.C., Trivedi P.K. (2005), *Microeconometrics*: *Methods and Application*, Cambridge University Press, New York.

Christensen L., D. Jorgenson L., and L. Lau (1973). Transcendental logarithmic production frontiers. *The Review of Economics and Statistics*, 55(1), 28-45.

Diewert W.E (2002). The quadratic approximation lemma and decompositions of superlative indexes. *Journal of Economic and Social Measurement*, 28, 63-88.

Fare R., Sung, K.J. (1986). On second order taylor's series approximation and linear homogeneity. *Aequationes Mathematicae*, 30, 180–186.

Gilovitch, T., Villon, R, and Tversky, A. (1985). The hot hand in basketball: On the misperception of random sequences. *Cognitive Psychology,* 17, 295-314.

Gourieroux C. (2000). *Econometrics of Qualitative Dependent Variables,* Cambridge University Press: Cambridge.

Greene W.H. (2003). Econometric Analysis (5th editon), Macmillan Publishing Company, New York.

Hamrick J., J. Rasp (2011). Using Local Correlation to Explain Success in Baseball. *Journal of Quantitative Analysis in Sports*, 7(4), 1–48.

Hosmer D., S. Lemeshow (2000). *Applied logistic regression*, Wiley, New York.

Keetch, K.M., Lee, T.D., Schmidt, R. A. (2008). Especial Skills: Specificity Embedded Within Generality. *Journal of Sport & Exercise Psychology,* 30 (6), pp. 723-736.

Keetch, K.M., Schmidt, R.A., Lee, T.D., Young, D.E. (2005) Especial skills: their emergence with massive amounts of practice. *Journal of Experimental Psychology: Human Perception and Performance*, 31, 970-978.

Mann, D.L., Ho, N.Y., De Souza, N.J., Watson, D.R., Taylor, S.J. (2007). Is optimal vision required for the successful execution of an interceptive task? *Human Movement Science*, 26, 343-356.

Montagne, G., Laurent, M., Durey, A., & Bootsma, R. J. (1999). Movement rehearsals in ball catching. *Experimental Brain Research*, 129, 87–92.

Rao J. (2009). Experts Perceptions of Autocorrelation: The Hot Hand Fallacy Among Professional Basketball Players, Working Paper, www.justinmrao.com/playersbeliefs.pdf.

Royston P., D.G. Altman (1994). Regression using fractional polynomials of continuous covariates: parsimonious parametric modeling. *Applied Statistics*, 43, 429-467.

Royston P., D.G. Altman (1997). Approximating statistical functions by using fractional polynomial regression. *The Statistician*, 46, 1-12.

Schmidt, R. A. (1975). A schema theory of discrete motor skill learning. *Psychological Review,* 82(4), 225-260.

Schmidt, R. A., Zelaznik, H. N., & Frank, J. S. (1978). Sources of inaccuracy in rapid movement. In G. E. Stelmach (Ed.), *Information processing in motor control and learning* (pp. 183–203). New York: Academic Press.

Simons J.P., Wilson, J.M., Wilson, G.J., Theall, S. (2009). Challenges to cognitive bases for an especial motor skill at the regulation baseball pitching distance. *Research Quarterly for Exercise and Sport*, 80 (3), 469-479.

Thorndike, E.L. (1911) Animal intelligence. New York: MacMillan.

Thorndike, E. L. (1914). Educational psychology: Briefer course. New York: Columbia University Press.

Thorndike, E.L. (1932). The fundamentals of learning. New York: Teachers College.

Wimmer V., N. Fenske, P. Pyrka, L. Fahrmeir (2011). Exploring competition performance in decathlon using semi-parametric latent variable models. *Journal of Quantitative Analysis in Sports*, 7 (4),1–19.

Acknowledgment:

Experiments described in the manuscript were conducted according to the ethical principles of APA. All participants took part in the study on a voluntary basis and could discontinue their participation at any time without any consequences. Permission for the subjects to participate in the study was granted by the Committee for Ethics of the University School of Physical Education in Wroclaw, Poland.

We would like to thank the authorities of WKK Wroclaw for allowing us to conduct this study.

We would like to thank prof. Zygmunt Pizlo for his effort and useful remarks which helped us to final version of the manuscript and prof. Andrew Hills for his help in English editing.