Dr. Karsten Schul & Prof. Dr. Daniel Memmert
(Institute of Training and Computer Science in Sport, German Sport University)

Evaluating Cognitive Training System for Football Players

Abstract

Objective: In this study we have tested the transferability of cognitive skills acquired on a computerized training system to on-field performance of soccer players. Specifically, we tested whether training with the cognitive training system IntelliGym improves footballers' on-field game intelligence.
Design: Two elite football academies (Hamburger SV and DSC Arminia Bielefeld) participated in the research. The study compared the performance of an experimental group (IntelliGym-trained players) and a control-task group (players performing a different computerized task) before and after cognitive intervention. Players were evaluated using the Game Test Situation "Identifying gaps" metrics system, an objective performance analysis, developed by the German Sport University (Memmert, 2010).
Results: The study found that players who trained with IntelliGym significantly improved their on-field decision-making performance more than players in the control-task group.

Introduction

In recent years several types of technology have been developed and implemented in football to enrich practice and foster the transfer of knowledge and acquisition of skills (for a review, see Möller et al., 2015). Hence, in the present research we take the Applied Cognitive Engineering (ACE) product IntelliGym© as an example to investigate potential transfer effects from gaming to decision-making performance in football. The only way to investigate transfer effects from IntelliGym on football performance is to conduct experiments that randomly assign a group of players to training conditions and others to control conditions, and then measuring tactical performance. After reviewing the German version of the cognitive training system IntelliGym developed by ACE and providing general feedback we set up a field experiment (see Methods). Although, no solid knowledge base exists at present to derive straight-forward predictions, we have formulated the following hypotheses.

H1: Concerning transfer effects from IntelliGym to sport performance, we might expect to find positive carry-over effects on tactical soccer performance from playing/training IntelliGym as compared to playing/training non team-sport related digital games.

Due to the fact that the IntelliGym addresses and simulates the "brain environment" of real football situations, two hypotheses were made:

1. Players who trained with the IntelliGym system, will improve their on-field game intelligence more than the players who performed the control task.
2. Players who trained with the IntelliGym system, will improve their on-field game intelligence ranking position inside their team in contrast to players who performed the control task.

Method

The empirical investigation of the aforementioned hypothesis was conducted in a field experiment that was funded by ACE. Two soccer teams (Team 1: the Hamburger SV Under 15 & Team 2: the DSC Arminia Bielefeld U 14) with a total of 26 players participated in this study. The participating clubs were selected by the German Football Association (DFB) while the participating teams were selected by their respective club. Each team was randomly divided to the experimental group (using IntelliGym) and control-task group. The experimental group consisted of 13 players. 7 players from Team 1 (Hamburger SV Under 15) $M_{IntelliGym}$ = 52.2 Min/week; SD = 14.4; $M_{IntelliGym}$ = 8.7 total hours accumulated; SD = 2.4; and 6 players from Team 2 (Arminia Bielefeld Under 14) $M_{IntelliGym}$ = 58.2 Min/week; SD = 12.6; $M_{IntelliGym}$ = 9.7 total hours accumulated; SD = 2.1; received specific digital training intervention using “IntelliGym”. While 8 players from Team 1; $M_{Soccervideo}$ = 49.5 Min/week; SD = 16.8; $M_{Soccervideo}$ = 8.25 total hours accumulated; SD = 2.8; and 5 players from Team 2; $M_{Soccervideo}$ = 58.8 Min/week; SD = 25.2; $M_{Soccervideo}$ = 9.8 total hours accumulated; SD = 4.2; received soccer specific video training intervention and act as task-control group.

Before and after the intervention all participants were tested with the Game Test Situation “identifying gaps in the defense” (Memmert & Harvey, 2010; Memmert & Roth, 2007). The Game Test Situation “identifying gaps in the defense” is one of a few sport-specific tactical creativity tasks that have been constructed and tested for objectivity, reliability, and validity (Memmert et al., 2010). The concept of game test situations acts as a type of compromise between standardized video tests and game observation methods (Memmert, 2013).

Game Test Situations are the testing and diagnostic version of 1-Dimension Games (Memmert, 2015). They contain contextual, real-world representations that provoke valid, creative solutions. In recurring, comparable situations, this competition setting evokes creative behaviour in specific 1-Dimension Games. Game idea, number of players, rules, and environmental conditions are given. The competition settings could involve different kinds of skills (hand, foot, or implement) in a system where the players take turns (two rounds for each person), thus meaning that positions and team players/opponents are systematically varied.

1-Dimension Games for the development of tactical creativity Game Test Situations are simple game forms with clearly defined game ideas, fixed numbers of players, defined rules, and consistent environmental conditions. The athlete's creative behaviour is assessed without standardizing the ball's path and actions of teammates and opponents. Hence, the fundamental idea is basic tactical constellations with clearly allocated roles in order to create recurring and consistent conditions with many repetitions for the participants.

The Game Test Situation "identifying gaps in the defense" evokes in recurring, comparable situations tactical behaviour in the identification of gaps (see Figure 1). In this test situation, the attackers attempted to play the ball past the three defenders and below the upper boundary into the opposite field.

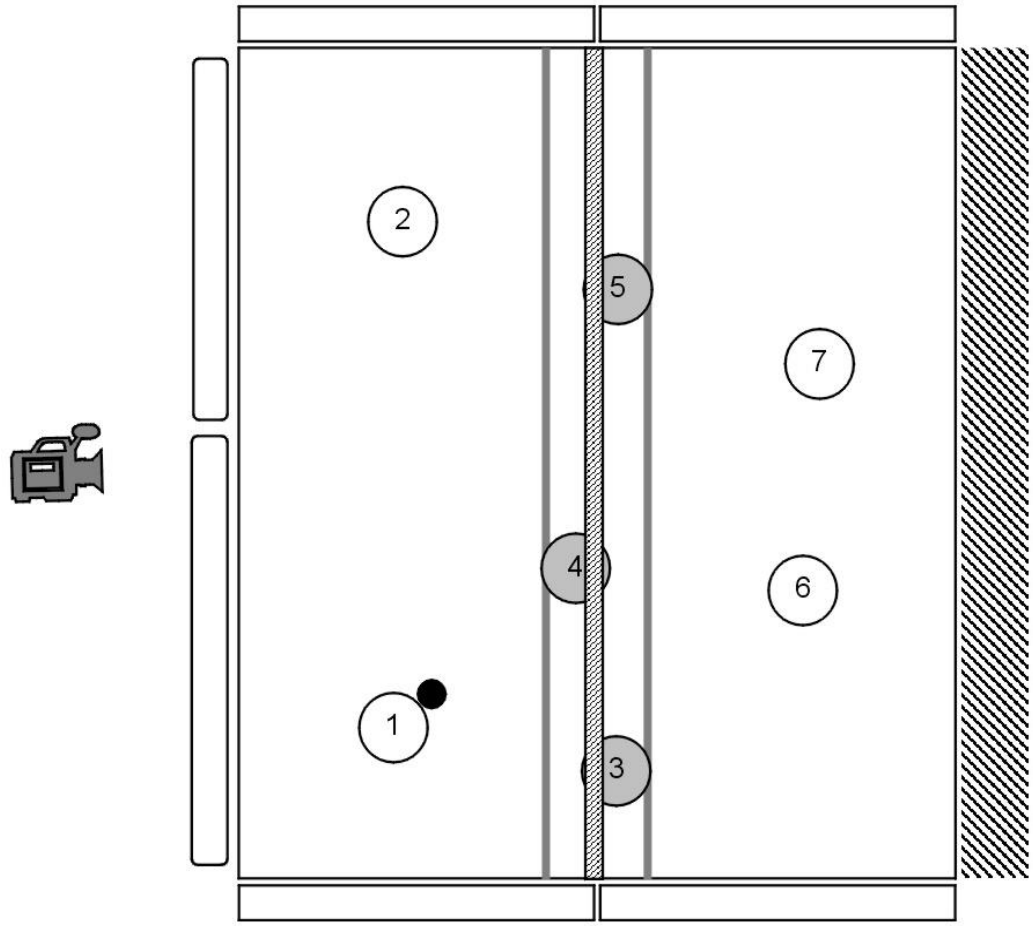

Fig 1: An overview over the Game Test Situation “identifying gaps in the defense”

Performance in the Game Test Situation “identifying gaps in the defense” is always confounded by other performance aspects. For example, motor skills interact with tactical solutions. Of course, the technical skills of a player have an influence on the solution to tactical situations. The evaluation of technical skills must be avoided so that one does not unconsciously assess the observed motor performances of the test person but rather exclusively concentrates on the expected tactical actions of the player. The observation of motor actions for the assessment of tactical actions is, in principal, possible in videos too, but it is rarely used and applicable only to some extent. In game test situations, however, the motor-action coupling is inevitable.

Originality of solutions to the situation (using gaps or passing)	Flexibility in the solutions to the situation (using gaps or passing)	Sca-ling	Anchor examples
Way above average (very unusual)	Two or more (different, original actions)	10	The subject demonstrated different highly unusual solutions to the situations. The gaps and passes found were absolutely unique.
Way above average (unusual)	Two or more (different, original actions)	9	The subject demonstrated different unusual solutions to the situations. Finding of gaps or passes was unique.
Above average (rare)	Two (different, original/rare actions)	8	The subject demonstrated different, still unusual solutions to the situations. The gaps and passes found were very rare.
Average (rather rare)	Two (different, rare actions)	7	The subject demonstrated two different solutions to the situations which were not unusual, but still very rare. The gaps and passes found were very surprising.
Average (quite rare)	Two (different, rare/new actions)	6	The subject demonstrated two different solutions to the situations, which were not unusual, but rare. The gaps and passes found were surprising.
Just below average (still new)	One (rare action)	5	The subject demonstrated one solution to the situations which wasn't the usual standard, but which had already occurred. The gaps and passes found were still innovative.
Just below average (very little new)	One (new action)	4	The subject demonstrated one solution to the situations which wasn't the usual standard, but which had already occurred often. The gaps and passes found were still innovative.
Below average (rather standard)	none	3	The subject generally offered standard solutions to the situations which had been displayed often. The gaps and passes found were rarely innovative.
Way below average (almost all standard)	none	2	The subject almost exclusively offered standard solutions to the situations which had all been displayed already. The gaps and passes found were very rarely innovative.
Way below average (only standard)	none	1	The subject only offered standard solutions to the situations. The gaps and passes found were never new.

Fig. 2: Scaling for the evaluation of tactical creativity in the Game Test Situation "identifying gaps in the defense" (Memmert & Roth, 2007).

In order to analyze creative actions, a video of the recorded behaviour is subsequently rated in regard to specific concepts by several independent experts. With the help of experts, application rules were developed and a scale was anchored (Memmert & Roth, 2007). Figure 2 illustrates the scaling for the Game Test Situation "identifying gaps in the defense".

In the assessment, a global estimation of overall game actions was made. First, the rater was required to rate, on a scale of 1 to 10, each action made. For better orientation, there is information given in the first two columns (aspect of quality and difficulty). The final evaluation result does not have to be an average value. Indications are the anchor examples in the last column. The definitions of attributes and scales have to be provided to the raters and have to be strictly discussed. All regulations and anchors of the scales are explained in detail (cf. Memmert, 2015).

To evaluate usability, game test situations and concept-oriented expert ratings must be measured by quality criteria of classic testing theory. The main focus should be the determination of inter-rater reliability of the rating procedure. “Rater reliability or objectivity is the degree to which the performance of a person is scored the same by two or more raters.” (Safrit & Wood, 1989, p. 46). An important question in this respect is whether it is possible that the rating result is no longer dependent on the judgements of experts. Another important aspect of evaluation objectivity is the intra-rater reliability. This evaluates how consistently the same observer concludes the same performance score in repeated ratings of the same object. These aspects are considered simultaneously with the inter-rater reliability coefficient. By transformation of the inter-rater reliability coefficient, it is possible to calculate the degree of evaluation objectivity that a whole group of future raters will demonstrate in regard to specific attributes.

The standardized Game Test Situation “identifying gaps in the defense” was analyzed by three independent soccer experts in possession of high coaching licenses following the procedure of Memmert and Roth (2007). Inter-rater reliabilities were sufficient (>.75). In addition a questionnaire assessed several socio-demographic variables, control variables, and items assessing the subjective experiences during the training intervention.

Results

On-Field Performance Analysis

To assess the reliability between the 3 evaluators of the performance measurement, 3 Cronbach's alpha tests have been applied. The first one was to compare between the pre-test scores of the 3 evaluators (α=.81), the second was to compare the post-test scores (α =.73) and the third compared the differences between the pre-test scores and post-test scores (the improvement) between the 3 evaluators (α =.78). These figures suggest a relatively high internal consistency. In order to test our first hypothesis, in which players who trained in the IntelliGym will improve more than

players who trained in the control task, the differences between pre- and post-intervention scores were calculated (post-score minus pre-score = difference score). The difference scores were ranked from number 1, the highest score (the player who improved the most) to number 26, the lowest score (the player who improved the least). This ranking list (1-26) was divided to the first 13 players (most improving players, the “Top” group) and last 13 players (less improving players, “bottom” group).

A chi-square test of independence was performed (see Figure 3) to examine the relation between the experiment groups (IntelliGym and control-task) and the improvement groups ("top" and "bottom" groups). In line with the first hypothesis, it was found that the relation of these variables was significant, x^2 (1, N = 26) = 3.84, p <.05. The experimental group players (IntelliGym group) were more likely to be in the "top" category as opposed to the control group players who were more likely to be in the "bottom" category.

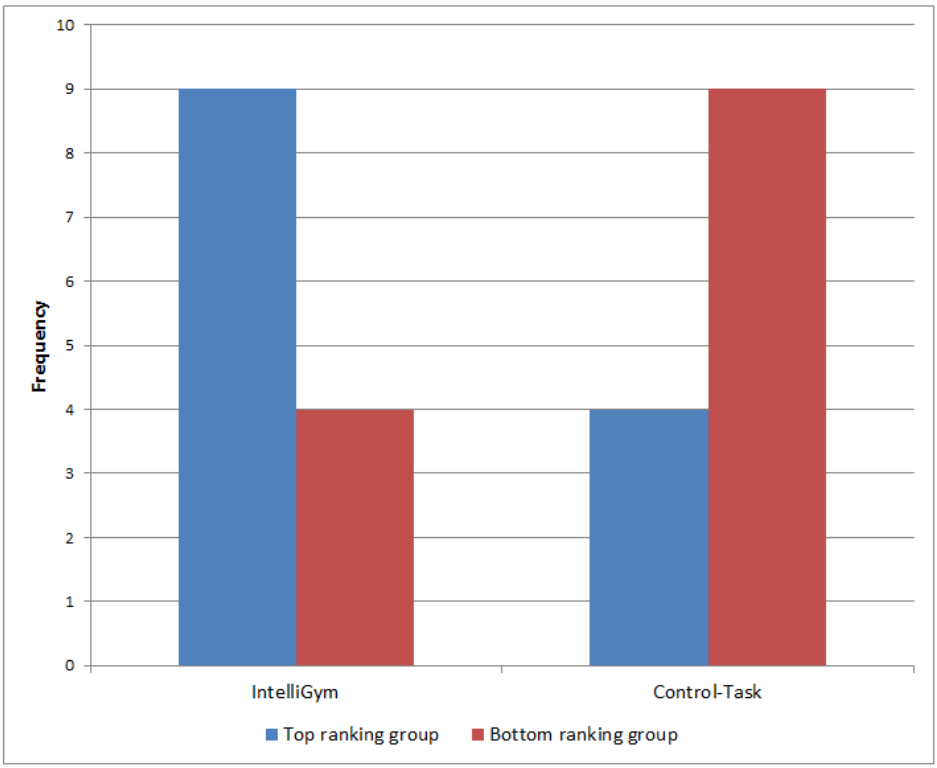

Figure 3. The score bars represent the frequency of each experiment group in the two improvement categories (Blue = “Top”, Red = “Bottom”).

In addition, to test the second hypothesis, in which players who trained with the IntelliGym will improve their ranking position inside the team, we tested the probability of a player from the experimental group as opposed to the control group,

to improve his ranking position within his team. For this purpose, pre-test scores and post-test scores were separately ranked (1-26 each). The differences between pre and post intervention ranking scores were calculated and put into a ranked list. The ranked difference list was divided to two categories (same as above: “top” and “bottom” groups). Then, a second chi-square of independence test was performed, to examine the relation between the experiment groups and the ranking improvement groups ("top", "bottom"). The relation of these variables was significant as well, x^2 (1, N= 26) = 3.84, p <.05 (coincidently, the frequencies were exactly the same as in the first test). That means that the experimental group participants (IntelliGym trained players) were more likely to be in the "ranking improved" (“top”) category as opposed to the control-task group participants who were more likely to be in the "ranking decrease" (“bottom”) category (see Figure 4).

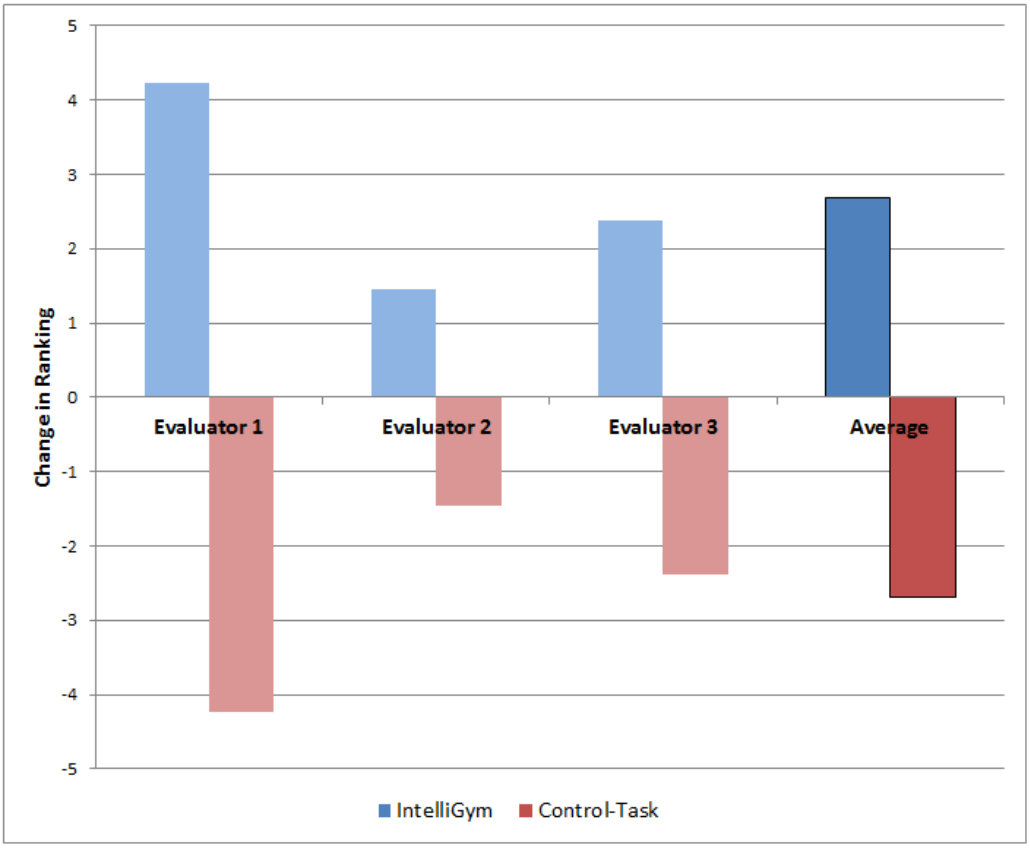

Figure 4. The score bars represent the number of ranking positions the players from each test group moved up (indicating improvement) or down (indicating decline) on average, for each evaluator and an average of the three.

Conclusion

The Project „Evaluating Cognitive Training System for Football Players" demonstrated that within the ranking system players who trained with IntelliGym improved their tactical on-field performance more than players who only watched soccer videos.

Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement number 672930.

References

Memmert, D. & Harvey, S. (2010). Identification of Non-Specific Tactical Problems in Invasion Games. Physical Education and Sport Pedagogy, 15, 287–305.

Memmert, D. (2010). Testing of Tactical Performance in Youth Elite Soccer. Journal of Sports Science and Medicine, 9, 199–205.

Memmert, D. (2013). Tactical Creativity. In T. McGarry, P. O'Donoghue, & J. Sampaio (eds.), *Routledge Handbook of Sports Performance Analysis* (pp. 297-308). Abingdon: Routledge.

Memmert, D. (2015). Teaching Tactical Creativity in Sport: Research and Practice. Abingdon: Routledge.

Memmert, D., & Roth, K. (2007). The Effects of Non-Specific and Specific Concepts on Tactical Creativity in Team Ball Sports. Journal of Sport Science, 25, 1423–1432.

Möller, C., Furley, P. & Memmert, D (2015). More than a waste of time?: Critical Reflection of Digital Games in Sport Contexts. In: uDay XIII.: Digitale Medien in Arbeits- und Lernumgebungen. Pabst Science Publishers.

Safrit, M. J., & Wood, T. M. (1989). Measurement concepts in physical education and exercise science. Champaign: Human Kinetics.