Ger J Exerc Sport Res 2021 · 51:146–155
<https://doi.org/10.1007/s12662-020-00697-0>
 Received: 12 August 2019
 Accepted: 8 December 2020
 Published online: 5 January 2021
 © The Author(s) 2020

Thomas Finkenzeller¹ · Björn Krenn² · Sabine Würth¹ · Günter Amesberger¹

¹ Department of Sport and Exercise Science, University of Salzburg, Hallein/Rif, Austria

² Department of Sport Science, University of Vienna, Vienna, Austria

The design fluency test: a reliable and valid instrument for the assessment of game intelligence?

Introduction

In the last decade, sport scientific research became more and more interested in a construct called executive functions (EF). EF refer to a broad construct (Etnier & Chang, 2009) that encompasses a set of higher-level functions, such as mental set shifting, information updating and monitoring, as well as inhibition of prepotent responses (Miyake et al., 2000). On the one hand, sport science investigated the effect of chronic or acute exercise on EF (Chang, Labban, Gapin, & Etnier, 2012; Lambourne & Tomporowski, 2010; McMorris & Hale, 2012; Verburch, Königs, Scherder, & Oosterlaan, 2013), whereas on the other hand, recent research questioned the role of EF on athletic performances and sport expertise (Jacobson & Matthaeus, 2014; Krenn, Finkenzeller, Würth, & Amesberger, 2018).

In the latter case, seminal findings were published by Vestberg, Gustafson, Maurex, Ingvar, & Petrovic (2012), who found soccer player's EF to be significantly predictive for the number of goals and the number of assists players scored two years later. In addition, they detected higher EF in players of the highest Swedish national league in comparison to soccer players of the second and third Swedish national division and that all soccer performance level groups scored significantly above population norm. As a consequence, EF were considered as key indicator of athletic performances in team sports (Lundgren, Högman, Näs-lund, & Parling, 2016; Vestberg et al., 2012). Subsequent research was able to corroborate this significant role of EF

mainly in soccer: It was found that elite soccer players showed higher EF than subelite players (Vestberg et al., 2012; Vestberg, Jafari, Almeida, Maurex, Ingvar, & Petrovic 2020), that ambitious soccer players showed higher EFs than the population norm (Vestberg et al., 2012) and that player's EF correlated significantly with their scored assists (Vestberg et al., 2012, 2020) and scored goals (Huijgen et al., 2015). Hence, these studies provide alleged evidence that above-average cognitive performance is associated with soccer expertise. Further, EF were suggested to represent a cognitive measure of game intelligence in team sports like ice hockey (Lundgren et al., 2016) and soccer (Vestberg et al., 2020).

The above described findings signifying the role of EF in team sports, predominantly in soccer, mainly are based on the Design Fluency Test (DFT) claiming to measure EF (Huijgen et al., 2015; Lundgren et al., 2016; Vestberg et al., 2012; Vestberg, Reinebo, Maurex, Ingvar, & Petrovic 2017). The DFT was originally developed for the assessment of fundamental skills and higher-level executive functions in clinical populations of children, adolescents, and adults and is one element of the Delis-Kaplan Executive Function System (D-KEFS) (Swanson, 2005). Rows of boxes, consisting of the same array of five dots, are presented. Within each box, different designs have to be generated by connecting the dots using four straight lines. The participants are required to draw as many different designs as possible within 60 s. There are three conditions which differ in the properties of the dots. In condition 1 (C1) all

dots are filled, in condition 2 (C2) the dots are empty, and in condition 3 (C3) the dots are alternately filled and empty. According to the D-KEFS manual (Delis, Kaplan, & Kramer, 2001b), C1 provides a basic test of design fluency. C2 requires design fluency and response inhibition caused by the change from filled to empty dots. C3 is developed to assess design fluency and cognitive flexibility through switching between filled and empty dots. Suchy, Kraybill, and Larson (2010) emphasized that C3 captures a separate construct that needs to be considered when interpreting DFT results. In general, it is stated that the DFT measures the "initiation of problem-solving behavior, fluency in generating visual patterns, creativity in drawing new designs, simultaneous processing in drawing the designs while observing the rules and restrictions of the task, and inhibiting previously drawn responses" (Swanson, 2005, p. 122). It is claimed by many researchers that such skills are crucial for success in several team sports (Furley & Memmert, 2010b; Tillman & Wiens, 2011). In soccer, for instance, "a successful player must constantly assess the situation, compare it to past experiences, create new possibilities, make quick decisions to actions, but also quickly inhibit planned behavior" (Vestberg et al., 2012, p. 4). Huijgen et al. (2015) emphasized that "a soccer player must be able to quickly anticipate and react to fast changing situations that occur during a soccer match" (p. 2). Based on these soccer-specific demands, Vestberg et al. (2012) argued that the DFT is appropriate for the assessment of EFs associated with success in soccer

because it challenges similar EFs as in typical soccer game situations. However, taking DFT's clinical origin into consideration its forthright application in the sample of elite athletes seems challenging and makes the highest demands on its psychometric properties.

In contrast to the findings on the DFT and soccer expertise (Vestberg et al., 2012, 2020, 2017), Furley, Schul, and Memmert (2017) pointed out that findings of improved cognitive performance in expert soccer players is anything but consistent. Several studies failed to provide evidence of superior executive functions in experts (e.g. Furley & Memmert, 2010a, 2015) or showed equivocal findings (e.g. Verburgh, Scherder, van Lange, & Oosterlaan, 2014). These inconsistencies are discussed against the background of confounding variables, sample sizes, expectation of the researchers, and definition of expert (Furley et al., 2017), as well as under the perspective of the need for reliable measurements (Schweizer, Furley, Rost, & Barth, 2020).

In previous studies in sport, the DFT was used based on the test criteria provided by the D-KEFS manual (Delis, Kaplan, & Kramer, 2001a). However, according to Homack, Lee, and Riccio (2005), much research has to be done in order to fully determine the psychometric properties of the DFT. Likewise Shunk, Davis, and Dean (2006) concluded that the psychometric properties of the DFT were not well established. Although the D-KEFS manual provides data on reliability of the DFT, these data refer to a heterogeneous sample in terms of age and other demographic characteristics. Furthermore, the period between test and retest was not kept constant within a time range of 9 to 74 days. Additionally, the D-KEFS manual consists of information on validity that incorporates intercorrelations of measures, and differences between Alzheimer and Huntington disease patients (Delis et al., 2001a). Therefore, there are open questions on short-term and long-term test-retest reliability and practice effects in general, as well as questions concerning the use of the DFT in athletes in particular. Results on the

stability of the rank order of participants in short- and long-term intervals are necessary to be able to evaluate findings on the prospective value of DFT performance. Further questions concern the differential and prospective value of the DFT in team sports in order to expand the existing knowledge on the diagnostic power of the DFT for application in team sports.

As the DFT was designed to detect dysfunctional EF in clinical populations, its application in elite athletes asks for strong evidence for a reliable and valid assessment of EF in this highly skilled sample. So far, research has failed to provide this clear evidence of reliability and validity. The current study aimed to contribute to this target and to enhance the evidence about the assessment of the DFT in sports for practitioners and researchers. We assembled different data sets collected in the applied field of sport psychology to enable a broad and differential analysis of the psychometric properties of the DFT.

The first aim of the present study was to determine reliability of DFT scores. Short-term test-retest reliability of DFT scores were assessed in three samples having different activities of varying duration between test and retest. Thus, reliability was examined in varying contextual situations to enable estimation of the effect of different sources of bias between measurements. Using a correlational coefficient approach, the relationship between test and retest from individual values was evaluated to show how well the rank order of participants in the first test was replicated in the retest (Hopkins, 2000). Additionally, changes between test and retest were examined to assess non-random effects, resulting from i.e. activities between measures, motivational and learning processes (Hopkins, 2000). Long-term test-retest correlation and systematic change of DFT scores between measurements were evaluated in a sample of national volleyball team players who were tested twice within a year.

The second aim was to determine the differential value of the DFT. Previous research (Vestberg et al., 2012, 2017) showed higher DFT scores in the sum of correct designs in soccer players com-

pared to normative data (Delis et al., 2001b). This study focused on differences between adolescent elite soccer players and high-school students. In contrast to previous studies (Huijgen et al., 2015; Vestberg et al., 2012, 2017), all single and composite DFT performance scores were considered in order to reflect DFT performance in a complex manner. Based on the findings by Vestberg et al. (2017), we expected that elite athletes would show a higher total sum of correct designs in the DFT, compared to the student group.

The third aim addressed the prospective value of the DFT in national team volleyball players in order to determine the extent to which the results transfer across different types of ball sports (Vestberg et al., 2012, 2017). Past research suggested that playing soccer attaches high demands towards EF (Vestberg et al., 2012, 2017). However, also volleyball as open-skill sports and strategic sport disciplines seem to make high demands on EF (Alves et al., 2013; Jacobson & Matthaeus, 2014; Krenn et al., 2018; Montuori et al., 2019). Taking several cognitive similarities between both sports into account (e.g. focusing on the ball and movement patterns of team players and opposing players; keeping tactical information and experiences about team members and opponents in mind; adapting to continuously changing situations and creating new ways to solve upcoming problems on the court; cf. Alves et al., 2013), we assumed significant correlations between DFT scores and performance parameters in volleyball. In addition, we expected higher correlations between DFT scores and more compounded and broader performance parameters (e.g. total points scored, attack errors), which should rely more heavily on the EF concepts of inhibition, working memory and cognitive flexibility than more specific performance parameters (e.g. serve aces and serve errors).

Methods

Multiple studies with different designs and different samples were conducted. Students, soccer and volleyball players or their parents/guardians gave informed consent to take part in the study. The lo-

T. Finkenzeller · B. Krenn · S. Würth ·
 G. Amesberger

The design fluency test: a reliable and valid instrument for the assessment of game intelligence?

Abstract

The design fluency test (DFT) has been reported to predict successful sports performance of soccer players and has therefore been in the spotlight of sport psychology research. There is, however, a lack of research regarding the psychometric properties of the DFT in elite sports. Thus, the aim of this research was to provide findings of test–retest reliability, practice effects and the diagnostic power of the DFT. Multiple studies of youth and adult elite athletes, as well as nonathlete students, were conducted in applied settings. Test–retest relationship demonstrated poor to acceptable short-term and long-term correlations. Furthermore, significant changes between test and retest were obtained in some variables that differed among samples. The differential value of the DFT was corroborated by significant differences between adolescent students and adolescent elite soccer players. Regarding the prospective value, significant partial correlation coefficients were found between DFT scores and volleyball performance in adult elite players. Although our research partially confirmed previous findings on the differential and prospective power of the DFT, the findings on test–retest reliability indicate that the DFT cannot be recommended for application in sports. The psychometric properties—in particular the findings on test–retest reliability—of the DFT have to be improved before research can be carried out on the application for the selection of team sport athletes and for the prediction of future success in team sports. Further research is needed to develop a scientific instrument for the assessment of game intelligence.

Keywords

Executive functions · Psychological assessment · Elite athletes · Test criteria · Team sport

cal educational board, the Austrian Football Association (ÖFB) and the Austrian Volleyball Federation (ÖVV) gave their permission to the scientific processing of the data. Ethical approval was obtained from the local university ethics committee. The application of the DFT was done in the context of a longstanding cooperation with the sport associations. Test–retest reliability was assessed in four samples in different contextual situations. Students and soccer players were compared to obtain information on the differential value of the DFT. Elite volleyball players were investigated to gain knowledge about the prospective value of the DFT for future success in volleyball.

Participants

Samples for assessing test–retest reliability and changes of test–retest scores.

Three groups of female adolescents were recruited. The time interval and activity between test and retest, and the familiarity of the participants with the DFT differed across groups. ■ **Table 1** provides information on the sample size, the age, the test–retest interval, the cognitive/physical load between measurements. In sample 1, female students of a vocational high school were examined. The students took the DFT at the beginning and end of a regular school lesson. The cognitive load of the lesson was low in comparison to the demands that were attached towards the female soccer players between the two assessments. Sample 2 was comprised of female soccer players, who completed the DFT during the entrance examination for the Austrian national centre of women's soccer. The DFT was administered at the beginning and at the end of a comprehensive sport psychological battery of tests, which has a high cognitive load. The female soccer players of sample 3 were familiar with the DFT as they previously completed the test during their entrance examination (time span = 0.5 to 3.5 years). The female athletes took the DFT at the beginning and during a comprehensive sport-specific motor test battery for youth national team players. The time interval between the two measure-

ments varied between 60 and 240 min ($M_{\text{time interval}} = 122.11 \pm 47.11$ min).

Long-term test–retest correlation and long-term changes were assessed by administering the DFT to 16 male volleyball players of the Austrian national team and Austrian youth national team (sample 4) twice in a time interval of a year (April 2014–April 2015). All participants of this sample were also included in sample 7 for testing the prospective power of the DFT.

Samples for testing the differential value.

High-school students (sample 5) were compared to adolescent elite soccer players (sample 6). The students sample included 119 individuals (61 females and 58 males) aged between 11 and 19 years ($M_{\text{age}} = 13.99$, $SD = 2.01$), who attended public schools. The students completed the DFT prior to a regular school lesson. The adolescent elite athletes sample ($n = 117$; 69 females and 48 males) consisted of 12- to 18-year-old soccer players ($M_{\text{age}} = 13.90$, $SD = 1.01$), who also attended public schools. Female players were tested during the entrance examination to the Austrian national centre of women's soccer. The data of the boys were collected prior to a training session in Austrian football youth academies.

Sample and design for testing the prospective value.

Volleyball players of the Austrian national team and Under 20 (U20) national team had to perform the DFT at the beginning of a training camp, held in April 2014 and April 2015, respectively. These scores were used to predict their volleyball performance in the following season (season 2014/2015 or season 2015/2016 each lasting from May to April). In total, 36 volleyball players ($M_{\text{age}} = 21.52$, $SD = 2.83$; sample 7) consisting of 13 players from the national team in 2014, 14 national team players in 2015, and 9 Under 20 national team players in 2014 were used to determine the prospective power. Data of each game were obtained from the official game stats sheets provided by the website of the European Volleyball Confederation (CEV). Experienced observers of each game's host country

Table 1 Characteristics of the samples used for assessing test–retest reliability and changes of test–retest scores

			Age (years)		Test–retest	
		<i>n</i>	<i>M</i>	<i>SD</i>	Interval	Load in-between
Sample 1	Female students	56	15.93	0.71	40 min	Low cognitive load
Sample 2	Female soccer players	66	13.76	0.68	180 min	High cognitive load
Sample 3	Female soccer players ^a	38	15.74	1.18	60–240 min	High physical load
Sample 4	Male volleyball players	16	21.45	2.52	1 year	Competition year

^aSoccer players of sample 3 had already conducted the DFT twice in a short-term test–retest interval

documented the data and provided them to the CEV. Past research provided evidence on the reliability of the assessed volleyball performance data (Asterios, Kostantinos, Athanasios, & Dimitrios, 2009; Patsiaouras, Moustakidis, Charitonidis, & Kokaridas, 2011). In total, the data of 39 matches (11 European Championship qualifying matches, 18 matches in the European League, 5 U21 world championship qualifying matches, and 5 friendly matches) were analysed. Each player's scored total points, number of serves, serve errors, aces, conducted attacks, and attack errors were analysed for all played matches of the national team ($k=32$) and U20/21 national team ($k=7$) during the season following the DFT. Past research emphasized the significance of these performance parameters for team's success in volleyball (Patsiaouras et al., 2011; Peña, Guerra, Buscà, & Serra, 2013).

Procedure

The DFT was carried out as a paper–pencil test in a group with supervision of a qualified and specially trained psychologist across all samples. After a standardized introduction which followed the guidelines of the D-KEFS test manual (Delis et al., 2001b), participants completed the three practice trials of C1 (filled dots). After successful completion of the examples, the actual test started. The testing phase required each participant to create as many different designs as possible in 60 s, using only four lines. The experimenter recorded the time using a stopwatch. The procedure was repeated in the same manner for C2 (empty dots) and C3 (switching between filled and empty dots).

Evaluation/scoring

The criteria for scoring were taken from the D-KEFS test manual (Delis et al., 2001a). A qualified psychologist carried out the evaluation and had to check the results for a second time. The primary measure for each of the conditions was the number of correct designs, each different and finished within the 60 s time limit (sum correct). A set-loss design is a design that violates the criterion rule (sum set-loss; e.g. more or less than four lines, at least one free-floating line, etc.). Furthermore, the number of correct, but repeated designs per condition were counted (sum repeated).

Additionally, composite scores which might provide information on higher-level cognitive skills, such as cognitive shifting (Delis et al., 2001a), were calculated. The total sum of all three conditions was computed for correct designs (total_sum correct), for set-loss (total_sum set-loss) and for repeated designs (total_sum repeated), as well as the sum of correct designs of condition 1 and 2 (C1+C2_sum correct). Finally, a contrast measure was determined by using the sum of correct designs in condition 3 minus the mean of the sum of correct designs of condition 1 and 2 (contrast measure). This measure is reported as an indicator of cognitive shifting (Delis et al., 2001b).

Statistical analysis

All statistical analyses were performed using the Statistical Package for Social Sciences (IBM Corp. Released 2013. IBM SPSS Statistics for Windows, Version 22.0., IBM Corp., Armonk, NY, USA). Shapiro–Wilk tests for normality revealed that most of the variables across samples did not follow normal distri-

bution. Therefore, the Spearman's rank correlation was run to test the relationship of test and retest scores. Changes between measures were analysed by the Wilcoxon signed-rank tests. For all correlational and difference analyses, alpha level was Bonferroni adjusted to the number of tested scores ($k=14$), and thus was set at $p<0.004$. According to George and Mallery (2003), test–retest correlation coefficients below $r=0.60$ indicate poor reliability, r values between 0.60 and 0.69 questionable reliability, r values between 0.70 and 0.79 acceptable reliability, r values between 0.80 and 0.89 good reliability, and values of $r=0.90$ and greater excellent reliability.

For assessing the differential power, differences between students and athletes were calculated by a multivariate analysis of variance (MANOVA) with age as a covariate including all single scores. In case of a significant main effect, results of univariate tests for each score are reported. The alpha level of the univariate tests was set at $p<0.006$ corresponding to Bonferroni correction to the number of tested scores ($k=9$).

The prospective power was assessed using partial correlations between the DFT scores and four different performance measures: Player's total points scored, serve errors, serve aces and attack errors. A preselection of these measures was conducted in accordance with Vestberg et al. (2012) in order to replicate previous findings and to reduce the number of partial correlations analyses. Since scoring points differed in their probability of occurrence, we controlled for playing positions (setters/liberoes versus hitters/blockers) and team membership, i.e. national team versus youth national team (cf. Vestberg et al., 2012). In addition, for the separate analyses of the total points, service errors, aces and attack

Table 2 Descriptive statistics of design fluency across samples used for the assessment of test–retest reliability and changes of test–retest scores

Parameter	Sample 1 (n = 56)				Sample 2 (n = 66)				Sample 3 (n = 38)				Sample 4 (n = 16)			
	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2	t1	t2
C1_sum correct	10.09	2.54	14.38	2.75	14.55	3.61	15.03	4.07	15.24	3.48	17.92	3.16	12.25	2.79	15.44	3.43
C1_sum set-loss	0.64	0.82	0.68	0.81	0.73	0.97	0.71	0.87	1.05	1.09	1.00	1.56	0.13	0.34	0.56	0.96
C1_sum repeated	0.25	0.58	1.34	1.91	3.14	3.72	4.80	3.87	5.84	3.15	8.42	4.30	1.00	0.82	2.25	1.88
C2_sum correct	11.64	2.73	14.57	3.20	15.33	3.30	15.61	3.44	15.55	3.03	18.34	2.85	13.88	2.99	16.63	3.14
C2_sum set-loss	0.54	0.54	0.52	0.63	0.68	1.04	0.65	0.87	0.92	1.08	0.66	0.88	0.56	0.89	0.63	0.96
C2_sum repeated	1.04	1.22	1.75	1.92	5.52	3.96	5.80	3.82	6.63	3.51	7.79	3.96	1.44	1.41	1.88	1.63
C3_sum correct	10.38	2.58	11.59	2.78	10.89	2.61	10.86	2.90	11.39	2.99	12.05	2.60	9.13	3.26	10.94	3.43
C3_sum set-loss	2.91	2.94	3.00	2.21	3.48	2.54	3.83	2.71	3.61	2.64	4.34	2.62	3.44	5.05	4.06	4.49
C3_sum repeated	0.64	1.00	0.98	1.36	1.68	1.57	2.47	3.56	1.92	1.62	2.66	2.27	1.50	1.32	0.94	1.18
Total_sum correct	32.11	6.30	40.54	7.26	40.77	7.68	41.50	8.38	42.18	7.11	48.32	6.57	35.25	6.80	43.00	7.55
Total_sum set-loss	4.09	3.41	4.20	2.77	4.89	3.10	5.20	3.30	5.58	3.37	6.00	3.65	4.13	5.38	5.25	4.47
Total_sum repeated	1.93	2.09	4.07	4.12	10.33	7.16	13.08	8.91	14.39	5.99	18.87	9.02	3.94	2.17	5.06	3.47
C1+C2_sum correct	21.73	4.76	28.95	5.41	29.88	6.07	30.64	6.86	30.79	5.47	36.26	5.29	26.13	5.26	32.06	5.97
Contrast measure	-0.49	2.68	-2.88	2.66	-4.05	2.89	-4.45	3.58	-4.00	3.25	-6.08	3.09	-3.94	3.70	-5.09	3.99

Sample 1 = female students; Sample 2 = female soccer players; Sample 3 = female soccer players at second assessment; Sample 4 = male volleyball players; C1 ...filled dots, C2...empty dots, C3...switching between filled and empty dots

errors we controlled for the total number of player's attempts in each variable. This approach was chosen to overcome the higher likelihood of players showing higher performance scores the more often they tried/spent time on the court (e.g. higher likelihood to generate two service errors when serving ten times than four times). Thus, we controlled for the number of line-ups when analysing the correlation between DFT scores and the sum of scored performance points; and we controlled for the total number of serves/attacks, respectively when analysing the correlation between DFT scores and serve errors/attack errors. Performance data were transformed to square root values to address the skewed distributions of these variables. Alpha level was Bonferroni adjusted to the number of analysed DFT scores ($k=14$), and thus was $p<0.004$.

Since scoring points differed in their probability of occurrence, we controlled for playing positions (setters/liberoes versus hitters/blockers) and team membership, i.e. national team versus youth national team (cf. Vestberg et al., 2012). In addition, for the separate analyses of the total points, service errors, aces and attack errors we controlled for the total number of player's attempts in each variable. This approach was chosen to overcome the higher likelihood of play-

ers showing higher performance scores the more often they tried/spent time on the court (e.g. higher likelihood to generate two service errors when serving ten times than four times). Thus, we controlled for the number of line-ups when analysing the correlation between DFT scores and the sum of scored performance points; and we controlled for the total number of serves/attacks, respectively when analysing the correlation between DFT scores and serve errors/attack errors.

Results

■ **Table 2** shows the descriptive statistics of the test and retest DFT scores across samples.

Short-term and long-term test–retest reliability

■ **Table 3** displays the Spearman's rank correlation coefficients between t1 and t2 for each sample. Correlation coefficients varied between $r=-0.08$ and $r=0.72$. The only acceptable test–retest reliability is shown in the composite score of C1 and C2 in the number of correct design (C1+C2_sum correct). The highest correlations in single scores were obtained in the sum of correct de-

signs in condition 2 across groups, with coefficients between $r=0.61$ and $r=0.68$. In all samples, correlation coefficients for the sum of correct designs were found to be lowest in the third (between $r=0.00$ and $r=0.37$), compared to the first (between $r=0.29$ and $r=0.57$) and the second condition. This was most prominent in sample 4 (test–retest interval of a year), showing a zero correlation in condition 3. Set-loss designs and repeated designs showed poor to questionable test–retest coefficients across all samples and all conditions, varying between $r=-0.08$ and $r=0.67$ for set-loss and between $r=-0.05$ and $r=0.63$ for repeated designs. Remarkably weak correlations ($r<0.35$) were observed in the contrast measure. The other composite scores demonstrated poor to questionable test–retest correlations coefficients (between $r=0.17$ and $r=0.66$), with exception of the total sum of correct designs ($r=0.70$) and the sum of correct designs of condition 1 and 2 ($r=0.72$) of sample 1 (female students).

Changes in test–retest DFT scores

The results on differences between test and retest are reported for each sample in ■ **Tables 2 and 3**. The greatest number of significant alterations was obtained in

Table 3 Spearman's rank correlations and Z-scores on differences (Wilcoxon signed-rank test) between test and retest for each sample

	Spearman rank correlation				Wilcoxon signed-rank test			
	Sample 1	Sample 2	Sample 3	Sample 4	Sample 1	Sample 2	Sample 3	Sample 4
Parameter	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>Z</i>	<i>Z</i>	<i>Z</i>	<i>Z</i>
C1_sum correct	0.50*	0.29	0.42	0.57	-6.26*	-0.87	-3.77*	-3.22*
C1_sum set-loss	0.42*	0.25	0.12	0.67*	-0.24	-0.11	-0.89	-2.07
C1_sum repeated	0.28	0.42*	0.38	0.29	-4.42*	-3.64*	-3.17	-2.39
C2_sum correct	0.65*	0.61*	0.68*	0.61	-6.10*	-0.46	-4.85*	-3.12
C2_sum set-loss	-0.08	0.35*	0.18	0.32	-0.14	-0.05	-1.20	-0.58
C2_sum repeated	0.40*	0.50*	0.63*	-0.05	-2.89*	-0.49	-1.92	-0.99
C3_sum correct	0.37	0.25	0.34	0.00	-2.98*	-0.49	-1.26	-1.84
C3_sum set-loss	0.58*	0.39*	0.24	0.25	-0.72	-1.10	-1.80	-0.80
C3_sum repeated	0.48*	0.24	0.40	0.12	-1.96	-1.87	-1.97	-1.20
Total_sum correct	0.70*	0.47*	0.60*	0.61	-6.38*	-0.68	-4.73*	-3.36*
Total_sum set-loss	0.64*	0.44*	0.29	0.17	-0.42	-0.67	-1.11	-1.77
Total_sum repeated	0.54*	0.59*	0.60*	0.26	-4.39*	-2.91*	-3.24	-1.23
C1+C2_sum correct	0.72*	0.50*	0.66*	0.66*	-6.52*	-0.81	-4.98*	-3.33*
Contrast measure	0.22	0.24	0.30	0.05	-4.52*	-0.91	-3.20	-0.91

* $p < 0.004$

Sample 1 = 56 female students; Sample 2 = 66 female soccer players; Sample 3 = 38 female soccer players at second assessment; Sample 4 = 16 male volleyball players; C1...filled dots, C2...empty dots, C3...switching between filled and empty dots

the students' sample (sample 1) that had the shortest test-retest interval and a low cognitive load between measures. These significant changes in the students were observed in the sum of correct designs in all three conditions (C1: increase of 42.52%, C2: increase of 25.17%, and C3: increase of 11.66%), as well as in the sum of repeated designs in the condition 1 and 2. Furthermore, with exception of the total sum of set-losses, all composite scores yielded significant changes in sample 1. Soccer players of sample 2, having had a high cognitive load between assessments, demonstrated a significant increase from test to retest in the sum of repeated designs in condition 1 (increase of 52.87%), and in the total sum of repetitions. Female soccer players that were familiar with the DFT and completed a physically high demanding test battery between assessments (sample 3) revealed significant improvements in the number of correct designs in condition 1 (increase of 17.59%) and 2 (increase of 17.94%), as well as in both composite scores on correct designs. Regarding long-term alterations (sample 4), volleyball players demonstrated a significant improvement in the sum of correct designs in condition 1 (increase of 26.04%), and a significant improvement in the sum of correct

designs of condition 1 and 2 and the sum of correct designs of all three conditions.

Differential value

A MANOVA run on all single scores with age as a covariate yielded a significant difference between students and soccer players (sample 5 vs. sample 6), $F(9, 225) = 9.81$, $p < 0.001$, $\eta_p^2 = 0.28$. **Table 4** includes descriptive statistics and results of follow-up univariate analyses of variances. Soccer players scored significantly higher in the number of correct designs in condition 1 and 2, compared to students. Students, however, repeated significantly less designs in all three conditions.

Prospective value

The analysis of prospective power was based on primary measures and the contrast score as described in D-KEFS manual (Delis et al., 2001b). **Table 5** shows the partial correlations between performance data and the sum of correct designs in all conditions and the contrast measure in elite volleyball players.

Partial correlations showed that the number correct designs of condition 1 and 2, the sum of correct designs of

condition 1 and 2 as well as the total sum of correct designs overall conditions significantly predicted players' total points. However, the partial correlations between players' total points and the sum of correct designs in condition 3 turned out low ($r = 0.07$) and did not reach statistical significance. The remaining DFT parameters indicated no significant correlations with the number of total points, the number of aces, serve errors and attack errors.

Discussion

The aim of the present study was to assemble different analyses evaluating the psychometric properties of the DFT in the field of sports. Thus, test-retest correlations and changes between measurements were determined among different contexts to gain insights on random and non-random effects on the reproducibility of DFT scores in applied settings. Furthermore, differential and prospective aspects were evaluated to expand on previous empirical findings.

Table 4 Descriptive statistics of DFT parameters across the students' (sample 5; $n = 119$) and soccer players' sample (sample 6; $n = 117$), and results of differences between groups using univariate analyses of variances with age as a covariate

Parameter	Students		Soccer players		<i>F</i>	<i>Df</i>	η_p^2	<i>p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
C1_sum correct	9.90	2.92	12.56	4.06	34.54	1, 233	0.13	< 0.001*
C1_sum set-loss	0.73	0.92	1.11	1.36	6.33	1, 233	0.03	0.01
C1_sum repeated	0.65	1.43	2.50	3.26	31.88	1, 233	0.12	< 0.001*
C2_sum correct	11.39	2.92	13.86	3.59	35.34	1, 233	0.13	< 0.001*
C2_sum set-loss	0.82	0.99	0.96	1.42	0.71	1, 233	0.003	0.40
C2_sum repeated	1.57	1.94	4.49	3.88	53.21	1, 233	0.19	< 0.001*
C3_sum correct	9.24	2.92	10.04	2.86	5.80	1, 233	0.02	0.02
C3_sum set-loss	2.83	2.38	3.79	2.68	8.42	1, 233	0.04	< 0.01
C3_sum repeated	0.80	1.18	1.54	1.58	16.78	1, 233	0.07	< 0.001*

* $p < 0.006$

C1 ...filled dots, C2 ...empty dots, C3 ...switching between filled and empty dots

Table 5 Partial correlations between primary measures and the contrast score and performance data of the following season (sample 7; $n = 36$; elite volleyball players)

	<i>R</i> _{total points} ^a	<i>R</i> _{aces} ^b	<i>R</i> _{serve errors} ^b	<i>R</i> _{attack errors} ^c
C1_sum correct	0.51*	0.28	0.33	-0.09
C1_sum set-loss	0.22	-0.17	-0.06	-0.19
C1_sum repeated	0.34	0.08	0.10	-0.05
C2_sum correct	0.53*	0.29	0.16	-0.13
C2_sum set-loss	-0.17	0.00	-0.04	-0.13
C2_sum repeated	0.18	-0.05	-0.02	0.03
C3_sum correct	0.07	0.02	0.08	0.05
C3_sum set-loss	0.25	0.04	-0.06	-0.16
C3_sum repeated	0.17	0.11	0.25	-0.09
Total_sum correct	0.50*	0.38	0.21	-0.10
Total_sum set-loss	0.23	0.02	-0.07	-0.20
Total_sum repeated	0.34	0.06	0.14	-0.05
C1+C2_sum correct	0.56*	0.31	0.30	-0.13
Contrast measure	-0.33	-0.23	-0.17	0.13

* $p < 0.004$ ^acontrolled for playing position, team membership, line-ups^bcontrolled for playing position, team membership, serve attempts^ccontrolled for playing position, team membership, attacking attempt

Short-term and long-term test-retest reliability across different contextual situations

Three female adolescent samples were used for analysing how well the rank order of participants in the first test is replicated in the retest considering a short-term interval. Test-retest correlations showed poor to acceptable reliability coefficients (George & Mallery, 2003) in all samples, suggesting no specific impact of sample characteristics, duration of test-retest period, and activity between measures. The students who had the shortest time interval and no physical and low cog-

nitive load between measurements obtained similar coefficients as compared to the soccer players. Based on the consistent findings regarding the size of the test-retest correlations, it is assumed that the poor to acceptable test-retest reliability represents a general than a specific effect. This assumption converges with data on test-retest reliability of the D-KEFS manual (Delis et al., 2001a), reporting test-retest coefficients between $r = 0.58$ (C1), $r = 0.57$ (C2), and $r = 0.32$ (C3) for correct designs using an average test-retest interval of 25 days ($SD = 12.8$). Furthermore, this is corroborated with our findings on long-term test-retest re-

liability, even when the results have to be interpreted carefully, due to the small sample size. As a consequence, we have to emphasize that DFT application in sports seems critical, as we failed to provide clear evidence on its test-retest reliability.

Interestingly, the most stable relationship in a single score across samples was the sum of correct designs in condition 2 (between $r = 0.61$ and $r = 0.68$). The exposure to condition 1 may have led to practice effects, resulting in higher test-retest reliability scores of condition 2 in terms of correct designs. This finding appears to be of significance for the improvement of the DFT. An extension of the number of practice trials and/or the number of tasks within a condition might contribute to an increase of test-retest reliability.

Regarding all four samples, it is remarkable that the coefficients of condition 3 showed the lowest correlations in the sum of correct designs. The switching from the very similar conditions 1 and 2 to the more complex task of connecting filled and empty dots alternately resulted in a poor stability of the relationship of test-retest DFT scores. Hence, differences in the level of coefficients between conditions raises doubts regarding the usage of aggregate DFT scores across all conditions. Further studies are needed to address the divergent findings observed in condition 1 and 2, compared to condition 3. The poor correlations of the contrast measure throughout all four samples might result from these deviant correlations between conditions. This corroborates the notion of Crawford et al. (2008)

that the contrast score of the DFT is “uninterpretable” (p. 1072).

The results on test–retest correlations across all four samples raise the question whether the DFT measures temporally stable traits related to higher-order cognitive functions. A possible reason for the obtained test–retest correlations might result from performance variability and measurement error caused by a task requiring complex and effortful processes (Delis, Kramer, Kaplan, & Holdnack, 2004). Furthermore, it has to be clarified how different states such as achievement motivation, test anxiety, self-confidence and other possible moderators affects behavioural accuracy and variability in design fluency.

The design fluency task can be considered an open-ended task, with more than one correct outcome. Specifically, participants can perform the test by drawing the lines intuitively, using one specific strategy or switching between different strategies. Thus, intra-individual variance increases, which, in turn, decreases the correlation test–retest DFT scores. It is assumed that the condition 1 contributes to a decrease in performance variability by arousing the strategy of t1 that might result in higher test–retest reliability of condition 2.

Taking these poor to acceptable test–retest correlations into consideration, the application of the DFT in the field of team sports has to be regarded with caution. It is recommended to improve the DFT substantially before it is applied in team diagnostics as well as in individual diagnostics. The administration of only three practice trials per condition might not be enough to obtain stable data on design fluency performance. An extensive pretest practice phase consisting of the entire DFT would be useful, which allows participants to become more familiar with the test minimizing practice effects. Furthermore, a longer processing time per each condition might enhance test–retest reliability. Future studies are required to determine whether these recommendations will show the expected improvements in reliability.

Changes in test–retest DFT scores across different contextual situations

All four samples yielded systematic, however inconsistent, changes between test and retest. The found effects on the sum of correct designs of condition 1 and 2, are assumed to be due to practice effects, and converges with findings presented in the DFT manual ($M_{\text{interval}} = 25 \pm 12.8$ days), indicating significant changes for all measures with the exception of repetition errors (Delis et al., 2001a). It is worth mentioning that even after a year improvements are observable. This finding is in line with Rabbitt et al. (2004), showing practice effects in neuropsychological tests even after a period of 7 years. The higher number of significant increases of DFT scores in the students may be a consequence of the shorter time interval, and the lower activity between measures. The absent practice effects of soccer players who performed the DFT for the first time (sample 2) might be a result of the demanding cognitive sport psychological assessments between test and retest. Mental fatigue might have reduced practice effects, which is supported by the significant increase of repeated designs in condition 1. The inconsistent findings on changes between test and retest across samples emphasize the importance of controlling for mental and physical activity when assessing DFT performance.

Differential value

In condition 1 and 2, soccer players generated more correct designs than the students. This finding is in line with previous results (Vestberg et al., 2012, 2017), showing higher scores of total correct designs in soccer players, compared to a normative group. However, soccer players repeated significantly more correct designs in all three conditions. Hence, our study delivers first findings that scoring should not exclusively be limited to a composite measure of correct designs how it was applied in previous studies (Vestberg et al., 2012, 2017). It is recommended that future studies should focus on the interplay of correct designs, set-loss designs, and repeated designs of the DFT in order

to obtain a holistic pattern of cognitive performance.

An alternative explanation of the found differences is conceivable. It is possible that soccer players may have an advantage in design-fluency performance when tested for the first time due to the similarity of the DFT with game situations that are featured using soccer tactic boards. This assumption is supported by more significant changes between test and retest (section: Changes in test–retest DFT scores) and higher increases in the sum of correct designs from test to retest in the students sample (sample 1; ■ Table 3) as compared to both soccer samples (samples 2 and 3). Thus, the familiarity with similar scenarios as administered in the DFT needs to be addressed in future studies. A further reason for the superiority of soccer players in generating a higher number of correct designs could be the better fitness status that might have influenced DFT performance. There is strong evidence that there is a relationship between physical fitness and cognitive functions (Chaddock, Neider, Voss, Gaspar, & Kramer, 2011), and therefore has to be considered in research on EF.

Prospective value

The partial correlations between the parameters on correct designs—with the exception of condition 3—and sports performance turned out significant, with an explanation of variance between 25% and 31.36%. Volleyball players drawing more correct designs yielded more points. This relationship is in line with findings reported in soccer players (Huijgen et al., 2015; Vestberg et al., 2012). However, caution is needed, given the fact that test–retest correlation of the overall score of correct designs ($r = 0.61$) was of similar size as the correlation between the overall score and total play points ($r = 0.50$). Furthermore, it has to be mentioned critically that the relationships between DFT scores on set-loss repeated designs, and correct designs in condition 3 and sport performance resulted in some variables not moving in the expected direction. For example, positive, non-significant correlations were revealed between total points

scored and the repeated designs in all conditions. Another unexpected finding represents the positive, non-significant correlation between the sum of correct designs of condition 1 and 2 and serve errors.

Our findings deliver weak support for the prospective power of the DFT that need to be discussed against the background of the poor to acceptable test–retest reliability of DFT scores. If performance variability is high, as indicated by the poor to acceptable test–retest correlations, then the question arises whether it is appropriate to predict sports performance by just using one composite measure of design fluency. Future work may thus examine whether administering the DFT at various time intervals and calculating average composite scores across these intervals counteracts large individual performance variability, and results in comparable findings on prospective validity.

Limitations

The selection and assessment of participants in varying contextual situations for determining aspects of reliability might be assessed at first glance as a limitation of the study. However, the overall aim of this study was to show how reliable and valid the DFT is in applied sport psychological contexts in order to provide data of different settings, which occur in practical work. The assessment was done in the interest of the sport associations, and therefore is typical for the work in the field of applied sport psychology. It must be critically noted that the samples of this study consisted of females with exception of one male sample. A further shortcoming is the short time interval between test and retest, which could have resulted in memory effects for some designs. The findings of long-term reliability are based on a small sample of 16 elite volleyball players, and thus a replication in a larger sample is worthwhile. Regarding differences among soccer players and non-athlete high school students, it cannot be completely excluded that different intellectual abilities, motivation, and coping strategies had an influence on the findings. The sample size and the small

number of volleyball games may be regarded as a potential limitation of results on the prospective value of the DFT with respect to the generalisability of the results. Finally, the findings of this study are limited to differential and prospective validity. It is recommended to examine the internal structure of the DFT using exploratory and confirmatory factorial analyses to test the assumptions of the DFT measures (Delis et al., 2001b). Additionally, divergent and convergent validity using well-defined EF tasks that are distal and proximal to the skills proposed in the test manual (Delis et al., 2001b) as well as tactical decision making tests, creativity tests or even coaching ratings should be examined in order to explore which executive subdomains are called upon by the DFT.

Conclusions

The size of the test–retest correlations, and significant changes between test and retest lead to the conclusion that the application of the DFT in individual as well as in group diagnostics cannot be recommended. Based on our findings and the information provided by the D-KEFS manual (Delis et al., 2001a) on test–retest reliability, we draw the conclusion that previous findings on the prospective power of the DFT have to be regarded with caution (Huijgen et al., 2015; Vestberg et al., 2012, 2017). Further studies are needed to improve the psychometric properties of the DFT before thinking about its application in sports. It has to be mentioned that there is growing acceptance that soccer talent is of multidimensional nature, and therefore needs to be predicted by multidisciplinary test batteries (Murr, Feichtinger, Larkin, O'Connor, & Höner, 2018). Finally, it is suggested (1) to select scientific instruments not only considering face validity, and (2) to use more sophisticated research approaches instead of simple research logic as in previous studies on the DFT and soccer expertise.

Corresponding address

Thomas Finkenzeller
Department of Sport and
Exercise Science, University of
Salzburg
Schlossallee 49,
5400 Hallein/Rif, Aus-
tria
thomas.finkenzeller@
sbg.ac.at

Acknowledgements. This work was facilitated through the cooperation of the Austrian Football Association (ÖFB) and the Department of Sport and Exercise Science of the University of Salzburg. The authors would like to thank all the students, soccer players and volleyball players for their participation.

Funding. Financial support was received by the ÖFB and the Austrian Volleyball Federation (ÖVV) for conducting the sport psychological tests.

Funding. Open access funding provided by Paris Lodron University of Salzburg.

Compliance with ethical guidelines

Conflict of interest. T. Finkenzeller, B. Krenn, S. Würth and G. Amesberger declare that they have no competing interests.

For this article no studies with human participants or animals were performed by any of the authors. All studies performed were in accordance with the ethical standards indicated in each case.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alves, H., Voss, M., Boot, W. R., Deslandes, A., Cossich, V., Inacio Salles, J., & Kramer, A. F. (2013). Perceptual-cognitive expertise in elite volleyball players. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2013.00036>.
- Asterios, P., Kostantinos, C., Athanasios, M., & Dimitrios, K. (2009). Comparison of technical skills effectiveness of men's national volleyball teams. *International Journal of Performance Analysis in Sport*, 9(1), 1–7.

- Chaddock, L., Neider, M.B., Voss, M.W., Gaspar, J.G., & Kramer, A.F. (2011). Do athletes excel at everyday tasks? *Medicine and science in sports and exercise*, 43(10), 1920–1926.
- Chang, Y.-K., Labban, J.D., Gapin, J.I., & Etnier, J.L. (2012). The effects of acute exercise on cognitive performance: a meta-analysis. *Brain Research*, 1453, 87–101. <https://doi.org/10.1016/j.brainres.2012.02.068>.
- Crawford, J. R., Sutherland, D., & Garthwaite, P. H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society*, 14(6), 1069–1073.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001a). *The Delis-Kaplan executive function system: technical manual*. San Antonio: The Psychological Corporation.
- Delis, D.C., Kaplan, E., & Kramer, J.H. (2001b). *Delis Kaplan D-KEFS executive function system*. London: Pearson.
- Delis, D.C., Kramer, J.H., Kaplan, E., & Holdnack, J. (2004). Reliability and validity of the Delis-Kaplan executive function system: an update. *Journal of the International Neuropsychological Society*, 10(2), 301–303.
- Etnier, J.L., & Chang, Y.-K. (2009). The effect of physical activity on executive function: a brief commentary on definitions, measurement issues, and the current state of the literature. *Journal of Sport and Exercise Psychology*, 31, 469–483.
- Furley, P., & Memmert, D. (2010a). Differences in spatial working memory as a function of team sports expertise: the Corsi Block-tapping task in sport psychological assessment. *Perceptual and motor skills*, 110(3), 801–808.
- Furley, P., & Memmert, D. (2010b). The role of working memory in sport. *International Review of Sport and Exercise Psychology*, 3(2), 171–194. <https://doi.org/10.1080/1750984X.2010.526238>.
- Furley, P., & Memmert, D. (2015). Creativity and working memory capacity in sports: working memory capacity is not a limiting factor in creative decision making amongst skilled performers. *Frontiers in Psychology*, 6, 115.
- Furley, P., Schul, K., & Memmert, D. (2017). Das Experten-Novizen-Paradigma und die Vertrauenskrise in der Psychologie [The expert-novice paradigm and the crisis of confidence in psychology]. *Zeitschrift für Sportpsychologie*, 23, 131–140.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: a simple guide and reference 11.0 update*. Boston: Allyn and Bacon.
- Homack, S., Lee, D., & Riccio, C.A. (2005). Test review: Delis-Kaplan executive function system. *Journal of Clinical and Experimental Neuropsychology*, 27(5), 599–609.
- Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine*, 30(1), 1–15.
- Huijgen, B. C. H., Leemhuis, S., Kok, N. M., Verburgh, L., Oosterlaan, J., Elferink-Gemser, M. T., & Visscher, C. (2015). Cognitive functions in elite and sub-elite youth soccer players aged 13 to 17 years. *PLoS ONE*, 10(12), e144580.
- Jacobson, J., & Matthaeus, L. (2014). Athletics and executive functioning: How athletic participation and sport type correlate with cognitive performance. *Psychology of Sport and Exercise*, 15(5), 521–527.
- Krenn, B., Finkenzeller, T., Würth, S., & Amesberger, G. (2018). Sport type determines differences in executive functions in elite athletes. *Psychology of Sport and Exercise*, 38, 72–79.
- Lambourne, K., & Tomporowski, P. (2010). The effect of exercise-induced arousal on cognitive task performance: a meta-regression analysis. *Brain Research*, 1341, 12–24. <https://doi.org/10.1016/j.brainres.2010.03.091>.
- Lundgren, T., Högman, L., Näslund, M., & Parling, T. (2016). Preliminary investigation of executive functions in elite ice hockey players. *Journal of clinical sport psychology*, 10(4), 324–335.
- McMorris, T., & Hale, B.J. (2012). Differential effects of differing intensities of acute exercise on speed and accuracy of cognition: a meta-analytical investigation. *Brain and Cognition*, 80(3), 338–351. <https://doi.org/10.1016/j.bandc.2012.09.001>.
- Miyake, A., Friedman, N.P., Emerson, M.J., Witzki, A.H., Howerter, A., & Wager, T.D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cognitive psychology*, 41(1), 49–100.
- Montuori, S., D’Aurizio, G., Foti, F., Liparoti, M., Lardone, A., Pesoli, M., et al. (2019). Executive functioning profiles in elite volleyball athletes: Preliminary results by a sport-specific task switching protocol. *Human movement science*, 63, 73–81.
- Murr, D., Feichtinger, P., Larkin, P., O’Connor, D., & Höner, O. (2018). Psychological talent predictors in youth soccer: a systematic review of the prognostic relevance of psychomotor, perceptual-cognitive and personality-related factors. *PLoS ONE*, 13(10), e205337. <https://doi.org/10.1371/journal.pone.0205337>.
- Patsiaouras, A., Moustakidis, A., Charitonidis, K., & Kokaridas, D. (2011). Technical skills leading in winning or losing volleyball matches during Beijing olympic games. *Journal of Physical Education and Sport*, 11, 39–42.
- Peña, J., Guerra, J., Buscà, B., & Serra, N. (2013). Which skills and factors better predict winning and losing in high-level men’s volleyball? *Journal of strength and conditioning research/ National Strength & Conditioning Association*, 27(9), 2487–2493.
- Rabbitt, P., McInnes, L., Diggle, P., Holland, F., Bent, N., Abson, V., et al. (2004). The University of Manchester longitudinal study of cognition in normal healthy old age, 1983 through 2003. *Aging Neuropsychology and Cognition*, 11(2–3), 245–279.
- Schweizer, G., Furley, P., Rost, N., & Barth, K. (2020). Reliable measurement in sport psychology: the case of performance outcome measures. *Psychology of Sport and Exercise*, 48, 101663.
- Shunk, A.W., Davis, A.S., & Dean, R.S. (2006). TEST REVIEW: Dean C. Delis, Edith Kaplan & Joel H. Kramer, Delis Kaplan Executive Function System (D-KEFS), The Psychological Corporation, San Antonio, TX, 2001. \$415.00 (complete kit). *Applied neuropsychology*, 13(4), 275–227.
- Suchy, Y., Kraybill, M.L., & Larson, J.C.G. (2010). Understanding design fluency: motor and executive contributions. *Journal of the International Neuropsychological Society*, 16(1), 26–37.
- Swanson, J. (2005). The Delis-Kaplan executive function system: a review. *Canadian Journal of School Psychology*, 20(1/2), 117.
- Tillman, C.M., & Wiens, S. (2011). Behavioral and ERP indices of response conflict in Stroop and flanker tasks. *Psychophysiology*, 48(10), 1405–1411.
- Verburgh, L., Königs, M., Scherder, E.J., & Oosterlaan, J. (2013). Physical exercise and executive functions in preadolescent children, adolescents and young adults: a meta-analysis. *British journal of sports medicine*, 48, 973–979.
- Verburgh, L., Scherder, E.J., van Lange, P.A., & Oosterlaan, J. (2014). Executive functioning in highly talented soccer players. *PLoS ONE*, 9(3), e91254.
- Vestberg, T., Gustafson, R., Maurex, L., Ingvar, M., & Petrovic, P. (2012). Executive functions predict the success of top-soccer players. *PLoS ONE*, 7(4), e34731.
- Vestberg, T., Jafari, R., Almeida, R., Maurex, L., Ingvar, M., & Petrovic, P. (2020). Level of play and coach-rated game intelligence are related to performance on design fluency in elite soccer players. *Scientific reports*, 10(1), 1–10.
- Vestberg, T., Reinebo, G., Maurex, L., Ingvar, M., & Petrovic, P. (2017). Core executive functions are associated with success in young elite soccer players. *PLoS ONE*, 12(2), e170845.