©Journal of Sports Science and Medicine (2024) **23**, 351-357
http://www.jssm.org DOI: https://doi.org/10.52082/jssm.2024.351

Research article

Assessment of Maximum Oxygen Uptake in Elite Youth Soccer Players: A Comparative Analysis of Smartwatch Technology, Yoyo Intermittent Recovery Test 2, and Respiratory Gas Analysis

Peter Düking [1]✉, Ludwig Ruf [2], Stefan Altmann [2,3], Maximiliane Thron [3], Philipp Kunz [4] and Billy Sperlich [4]
[1] Department of Sports Science and Movement Pedagogy, Technische Universität Braunschweig, Braunschweig, Germany; [2] TSG ResearchLab gGmbH, Zuzenhausen, Germany; [3] Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany; [4] Integrative and Experimental Exercise Science, Department of Sport Science, University of Würzburg, Würzburg, Germany

Abstract
The maximum oxygen uptake (VO_{2max}) is a critical factor for endurance performance in soccer. Novel wearable technology may allow frequent assessment of $\dot{V}O_{2max}$ during non-fatiguing warm-up runs of soccer players with minimal interference to soccer practice. The aim of this study was to assess the validity of VO2max provided by a consumer grade smartwatch (Garmin Forerunner 245, Garmin, Olathe, USA, Software:13.00) and the YoYo Intermittent Recovery Run 2 (YYIR2) by comparing it with respiratory gas analysis. 24 trained male youth soccer players performed different tests to assess VO_{2max}: i) a treadmill test employing respiratory gas analysis, ii) YYIR2 and iii) during a non-fatiguing warm-up run of 10 min wearing a smartwatch as recommended by the device-manufacturer on 3 different days within 2 weeks. As the device-manufacturer indicates that validity of smartwatch-derived VO_{2max} may differ with an increase in runs, 16 players performed a second run with the smartwatch to test this claim. The main evidence revealed that the smartwatch showed an ICC of 0.37 [95% CI: -0.25; 0.71] a mean absolute percentage error (MAPE) of 5.58% after one run, as well as an ICC of 0.54 [95% CI: -0.3; 8.4] and a MAPE of 1.06% after the second run with the smartwatch. The YYIR2 showed an ICC of 0.17 [95% CI: -5.7; 0.6]; and MAPE of 4.2%. When using the smartwatch for VO_{2max} assessment in a non-fatiguing run as a warm-up, as suggested by the device manufacturer before soccer practice, the MAPE diminishes after two runs. Therefore, for more accurate VO_{2max} assessment with the smartwatch, we recommend to perform at least two runs to reduce the MAPE and enhance the validity of the findings.

Key words: Data-informed Training, Digital Health, eHealth, Technology, Wearable, mHealth.

Introduction

Soccer imposes significant metabolic load on players (Stolen et al., 2005). Due to continuous change in playing direction and running speeds from low intensity to all-out efforts (Wing et al., 2020), both the aerobic and anaerobic system play an important factor in soccer specific fitness (Stolen et al., 2005). With maximum oxygen uptake (VO_{2max}) as the most prominent parameter reflecting the capacity of the aerobic system (Poole and Richardson, 1997), assessing this value in soccer players is crucial to quantify cardiovascular fitness. There is a close association of a sufficient VO_{2max} with running distances at high-intensity speed in official matches (Bradley et al., 2013) and is therefore often a primary focus in the training and assessment of soccer players (Slimani et al., 2019; Clemente et al., 2023; Wisloeff et al., 1998). Additionally, the assessment of VO_{2max} of soccer players is important to retrospectively elucidate effectiveness, and to prospectively prescribe training programs to enhance performance optimally (Wisloeff et al., 1998; Clemente et al., 2023; Helgerud et al., 2001).

The gold-standard to assess VO_{2max} is connected to ramp testing on a treadmill employing respiratory gas analysis. However, this procedure is time-consuming, costly and requires specialized equipment and personnel. Furthermore, ramp-testing necessitates the player's maximal effort, which can disrupt training and recovery routines. It has been noted that maximal or near-maximal fitness tests, like ramp-tests, are often considered unsuitable for regular use in practical settings by many practitioners, due to their interference with ongoing training schedules or competitions (Schimpchen et al., 2023).

To address these drawbacks, the scientific community developed and explored simpler methods. For instance, a submaximal cycle test designed to predict $\dot{V}O_{2max}$ was introduced by Åstrand and Ryhming in 1954 (Åstrand and Ryhming, 1954). In soccer, running based field tests such as the YoYo Intermittent Recovery Test 2 (YYIR2) have become popular for indirectly assessing VO_{2max}. The test is widely employed due to its simplicity, -cost-effectiveness and allows to test multiple players at the same time (Bangsbo et al., 2008). Additionally, the YYIR2 was designed to reflect the demands of team sports such as soccer with special regards to their intermittent nature (Thomas et al., 2006). The YYIR2 provides VO_{2max} with good-to-excellent test-retest reliability in team sport athletes as recently outlined in a review (Grgic et al., 2019). However, the YYIR2, also requires an all-out effort which impairs the (frequent) assessment of e.g. VO_{2max} within preparation phase and soccer season. Additionally, factors affecting the results of the YYIR2 include nutritional status (Grgic et al., 2019), time of day at which testing is conducted (Chtourou et al., 2012), provision of verbal encouragement (Currell and Jeukendrup, 2008), and it is currently unclear whether

Received: 10 January 2024 / Accepted: 19 April 2024 / Published (online): 01 June 2024

ground surface characteristics influence test results (Grgic et al., 2019). Consequently, the YYIR2 has some disadvantages in practice and overcoming (some of) these might allow more frequent and easier assessment of VO_{2max}.

With technological progress, particularly in chip miniaturization (Waldrop, 2016), contemporary wearable technologies like smartwatches are now available. These devices provide users with $\dot{V}O_{2max}$ if the smartwatch is worn during outdoor runs at submaximal intensities. (Garmin Ltd, 2019). Although assessing VO_{2max} with a smartwatch during submaximal outdoor runs appears promising, studies have revealed significant variability in the criterion-related concurrent validity of parameters provided by consumer-grade wearables (Düking et al., 2016; Peake et al., 2018). Many, although not all wearables do not provide valid parameters in many scenarios and populations (Düking et al., 2016; Peake et al., 2018) yet are marketed with bold marketing claims due to a largely unregulated market (Sperlich and Holmberg, 2017).

Therefore, evaluating the criterion-related concurrent validity of parameters from wearables is crucial for their application in sports practice. For VO_{2max} assessment, companies producing wearables such as smartwatches advancements smartwatch technologies claim to estimate VO_{2max} through non-exhaustive, easy-to-administer tests in their marketing. Recent analyses indicate that runners can accurately assess their VO_{2max} with a smartwatch during submaximal outdoor runs, with an error margin of 5.7% compared to respiratory gas analysis (Düking et al., 2022).

Currently there is no available evidence on the validity of the smartwatch "Garmin Forerunner 245" to assess VO2max using a non-fatiguing test performed as a warm-up prior to the usual soccer training.

Addressing this research gap could aid practitioners in selecting the most effective methodology for VO_{2max} assessment in highly trained soccer players, whether it be through respiratory gas analysis, the YYIR2 test, or smartwatch-assessed $VO_{2max.}$ Therefore, the aim was to assess the criterion related concurrent validity of VO_{2max} estimation provided by an end consumer grade smartwatch andthe YYIR2 in comparison to respiratory gas analysis. Based on prior studies of VO_{2max} assessment using smartwatches (which showed an error of 5.7% in runners) (Düking et al., 2022) and the YYIR2 (Grgic et al., 2019) in varied populations, our hypotheses are: i) the error in $\dot{V}O_{2max}$ assessment via smartwatch will be approximately 5 - 6%, and ii) the error margin in VO_{2max} estimation through smartwatch will be comparable to that of the YYIR2 test.

Methods

Experimental Approach to the Problem

All players visited the laboratory to determine anthropometric data, maximum values of heart rate (HR_{max}) and VO_{2max}, which was assessed during a ramp test treadmill protocol and by employing the criterion breath-by-breath gas analysis. On another day, subjects performed one outdoor YYIR2 as this test is often used by practitioners when no laboratory with respiratory gas analysis is available or if such assessments of $\dot{V}O_{2max}$ are deemed unfeasible e.g. due to time constraints. On two other days, the players performed outdoor runs during the warm-up period of the regular soccer training while wearing the smartwatch. Figure 1 illustrates the experimental approach.

Subjects

24 healthy and injury free national level male youth soccer players (mean age 17.3 ± 1.3 years, body height 178.1 ± 6.2 cm, body mass 71.5 ± 8.6 kg) and/or if necessary, their legal guardians were informed about all experimental procedures and consented to participate in the study. The players were recruited from a youth soccer academy. Players were eligible for inclusion if they had been part of the academy for over three years. However, players with wrist tattoos were excluded due to potential interference with the smartwatch's optical sensor (Bent et al., 2020). According to a recent classification framework, this group of players were categorized as Tier 3 athletes (highly trained/national level athletes) (McKay et al., 2022).

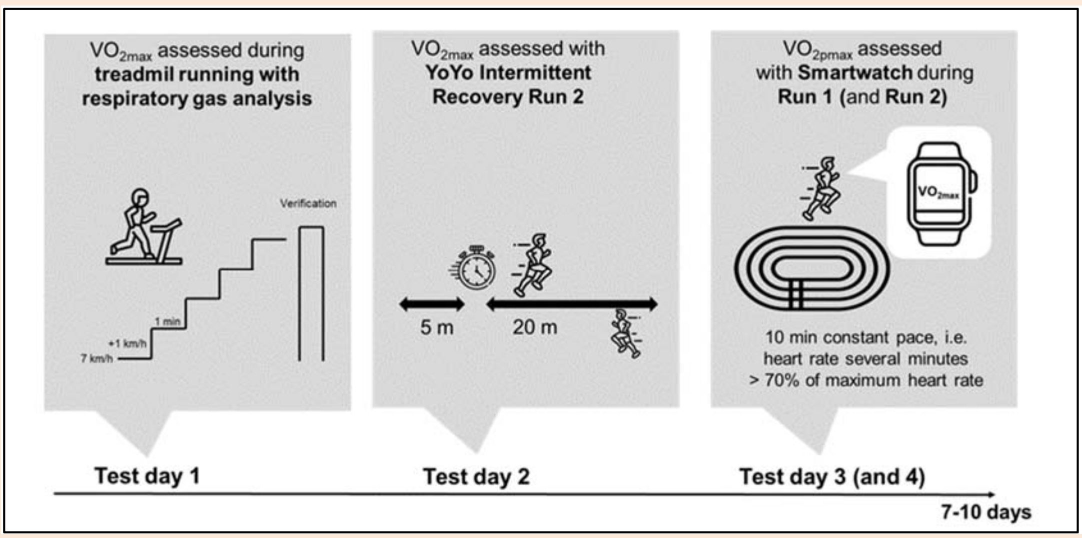

Figure 1. Illustration of the experimental procedures. Respiratory gas analysis on a treadmill, the YoYo Intermittent Recovery Run 2 and the Smartwatch Run 1 (and 2) took place on separate days within a period of 7-10 days.

Table 1. Intraclass Correlation Coefficient and the mean absolute percentage error comparing different methodologies assessing $\dot{V}O_{2max}$

	Intraclass Correlation Coefficient [95% CI]; "Interpretation"	Mean Absolute Percentage Error [%]
Smartwatch Run 1 vs. respiratory gas analysis	0.37 [-0.25; 0.71]; "poor"	5.58
Smartwach Run 2 vs. respiratory gas analysis	0.54 [-0.3; 8.4]; "moderate"	1.06
YoYo Intermittent Recovery Test 2 vs. respiratory gas analysis	0.17 [-5.7; 0.6]; "poor"	4.2

As there are no ICCs available prior to our study, we performed a sample size calculation following data collection in our study using Arifin´s web-based sample size calculator (Arifin, 2018) with parameters set as follows: ICC, ρ0 = 0.70 (Nunnally, 1978); ρ1 = 0.54 (own calculation, see Table 1), α = 0.05, 1 - β = 0.80, k = 2, dropout = 0% (as no participant dropped out of our study). A final sample size of 115 participants was calculated and our result should be seen as a convenient sample as highly trained, national level youth soccer players are few in numbers and difficult to motivate to participate in research studies. All experimental procedures took place within the players' usual training and testing routines. The study was approved by the Faculty's Exercise Science and Training Ethical Committee of the University of Würzburg and performed following the declaration of Helsinki.

Assessment of maximum oxygen uptake with respiratory gas analysis

All players performed a ramp protocol followed by a verification phase on a motorized treadmill (Mercury, h/p/cosmos sports and Medical GmbH, Nussdorf-Traunstein, Germany) to assess VO_{2max}. The initial treadmill speed was set to 7 km·h^{-1} increasing every minute by 1 km·h^{-1} until volitional exhaustion. Overall, exhaustion was verified when three of the four following criteria were met: 1) plateau in VO_2, that is, an increase < 1.0 mL·min^{-1}·kg^{-1} despite an increase in velocity; 2) respiratory exchange ratio >1.1; 3) ratings of perceived exertion > 18; and 4) peak blood lactate (peak lactate) > 6 mmol·L^{-1} after exercise. Addressing concerns that these criteria alone might not adequately determine VO_{2max}, a second trial to verify $\dot{V}O_{2max}$ was initiated three minutes after the completion of the ramp test, as suggested in existing literature (Poole and Jones, 2017). This trial was performed at a velocity of 1 km·h^{-1} higher than that achieved during the final velocity of the ramp test and all participants were encouraged verbally to run for as long as possible.

Portable breath-by-breath analysis (Metamax 3B, CORTEX Biophysik GmbH, Leipzig, Germany) employing standard algorithms assessed oxygen consumption serving as the criterion measure. The breath-by-breath analyzer was calibrated prior to each individual test using a calibration gas (15.8% O_2, 5% CO_2 in N; Praxair, Düsseldorf, Germany) targeting the range of anticipated fractional gas concentrations and a precision 3L syringe. The oxygen sensor of this portable breath-by-breath gas analyzer provides reliable data with technical measurement error below 2% (Macfarlane and Wong, 2012). Additionally, a recent article by van Hooren (van Hooren et al., 2024) examined error rates of various gas analyzers. Across different intensities, the Cortex Metamax 3B used in this study exhibited an error rate of 1.64 ± 1.87%, which is among the lowest error rates observed for the investigated devices.

Protocol for assessment of peak oxygen uptake with the YoYo Intermittent Recovery Run 2

The YYIR2 involved progressively increasing running speed over 2 × 20 m, interspersed with a 10-s period of jogging around a marker placed 5 m behind the finish line after each 40 m. The time frame was controlled by audio signals. The test ended when the participant stopped voluntarily or was unable to complete the shuttle run in time due to exhaustion on two consecutive occasions. The primary outcome of this test was the total distance covered (Bangsbo et al., 2008). A recent review summarized the literature on the reliability of the YYIR2 and found that ICCs ranged from 0.86 to 0.96 and coefficients of variation ranged from 4.2 to 12.7% (Grgic et al., 2019).

Protocol for assessment of maximum oxygen uptake with the smartwatch

We employed a multi-sensor wrist worn smartwatch (Forerunner 245, Garmin, Olathe, USA, Software:13.00). This smartwatch was chosen because it claims to provide the user with a $\dot{V}O_{2max}$ value and the manufacturer is among the top selling brands in the worldwide market for smartwatches (Statista Market Insights, 2023). The smartwatch features an optical heart rate (HR) sensor and a GPS receiver unit. To test the smartwatch in a scenario which resembles practice, we used and programmed the smartwatch as indicated by the manufacturer. Additionally, as the information which is provided by the manufacturer is also available to the soccer players, this is likely how the smartwatch would be used in real life scenarios. Each player's individual ramp-test-derived HR_{max} was inserted into the smartwatch's software. The specific algorithms utilized to assess VO_{2max} are not publicly disclosed by the manufacturer. However, according to the manufacturer, reliable HR and GPS-derived velocity data segments from individual runs are utilized to estimate VO_{2max} (Garmin Ltd, 2019). The manufacturer's instructions to assess $\dot{V}O_{2max}$ with the smartwatch indicates that a person should run outdoors for at least 10 min with a HR "several minutes" above 70% of the HR_{max} (Garmin Ltd, 2019). Therefore, each participant wore the smartwatch while performing a warm-up run prior to their usual soccer training on the soccer pitch. To meet the manufacturers recommendations, the warm-up run was performed at an individually constant pace (which was checked by the coach by assessing GPS-derived velocity on the smartwatch) and the experienced coach ensured that the individuals HR was above 70% HR_{max} for several minutes by checking the HR of players throughout the run. The manufacturer indicates that the VO_{2max} assessment might improve following "a couple" of runs (Garmin Ltd, 2019). Therefore, out of the tested athletes and prior to any data analysis, 16 randomly selected athletes wore the smartwatch on a second occasion approx. 3-5 days after their first run.

Statistical analysis

To investigate the validity of the VO_{2max} provided by the smartwatch and the YYIR2, we compared values of the smartwatch and YYIR2 against the respiratory gas analysis using different statistical metrics: i) mean absolute percentage error, ii) Bland-Altmann analysis and ii) Intraclass Correlation Coefficients (ICC).

Data was checked for normal distribution using the Kolmogorov-Smirnoff-Test which revealed normal distribution for all variables ($p \geq 0.05$).

To provide an indicator of the overall measurement error, mean absolute percent errors (MAPE) were calculated as previously performed (Lee et al., 2014). We followed previously performed statistical analysis to investigate validity to increase comparability of results between studies.

MAPE was calculated as average of absolute difference between the smartwatch and the respiratory gas analysis divided by the respiratory gas analysis values, multiplied by 100. The MAPE was calculated as it is a more conservative estimate of error that takes into account both over- and underestimation (Lee et al., 2014).

As previously performed (Mayorga-Vega et al., 2024), we calculated an ICC using a two-way random effects model with absolute agreement and single measurements [also known as ICC (2.1)] (Koo and Li, 2016) and was interpreted as follows:<0.5 poor, 0.5 - 0.75 moderate, 0.75 - 0.09 good, and 0.90 - 1.00 excellent (Koo and Li, 2016).

Bland–Altman plots were used to calculate the average difference and corresponding 95% limits of agreement.

Results

No missing data or dropouts were reported. VO_{2max} as measured by respiratory gas analysis was 56.6 ± 4.9 $ml \cdot min^{-1} \cdot kg^{-1}$ (Figure 2). The players achieved a VO_{2max} of 54.2 ± 1.7 $ml \cdot min^{-1} \cdot kg^{-1}$ as assessed by the YYIR2. Table 2 summarizes the duration and covered distance of the runs performed while the players wore the smartwatch.

The Bland-Altman analysis comparing the different $\dot{V}O_{2max}$ assessments are displayed in in Figure 3. The average difference and 95% limits of agreement revealed by the Bland-Altman analysis when comparing VO2max values from smartwatch run 1 with the respiratory gas analysis, smartwatch run 2 with the respiratory gas analysis and the YYIR2 with the respiratory gas analysis are -3.16 $ml \cdot min^{-1} \cdot kg^{-1}$(-15.7 $ml \cdot min^{-1} \cdot kg^{-1}$; 9.3 $ml \cdot min^{-1} \cdot kg^{-1}$); 0.12 $ml \cdot min^{-1} \cdot kg^{-1}$ (-8.94 $ml \cdot min^{-1} \cdot kg^{-1}$; 9.17 $ml \cdot min^{-1} \cdot kg^{-1}$) and -2.40 $ml \cdot min^{-1} \cdot kg^{-1}$ (12.06 $ml \cdot min^{-1} \cdot kg^{-1}$; 7.25 $ml \cdot min^{-1} \cdot kg^{-1}$), respectively.

Table 1 presents Intraclass Correlation Coefficient and the Mean Absolute Percentage Error.

Discussion

The aim of the present study was to assess the validity of VO_{2max} estimation provided by a consumer-grade smartwatch and the YYIR2 in comparison to breath-by-breath-analysis. The main findings of the present study showed that when performing two runs with the smartwatch to assess VO2max, the mean absolute percentage error decrease to 1.06% and agreement with respiratory gas analysis increases to 0.12 $ml \cdot min^{-1} \cdot kg^{-1}$ across the tested VO2max range, but comparably large confidence intervals indicate that individual values might show larger errors. Additionally, the mean absolute percentage error was 4.2% and showed to have poor agreement with respiratory has analysis.

The different statistical approaches employed in our study (i.e. mean absolute percentage error, intra-class correlation coefficient and Bland-Altmann analysis) are congruent, showing that it is favorable to perform two runs with the herein used Smartwatch to increase validity of provided VO_2max values.

Table 2. Duration, covered distance and mean heart rate of the players performing the runs with the smartwatch to assess $\dot{V}O_{2max}$ (mean +- SD)

	Duration [s]	Distance [m]	% maximum heart rate [bpm]	Smartwatch $\dot{V}O_{2max}$ [$ml \cdot min^{-1} \cdot kg^{-1}$]
Run 1 (n = 24)	665.2 ± 19.1	2121 ± 155	81.0 ± 4.3	53.4 ± 5.6
Run 2 (n = 16)	679.0 ± 20.5	2114 ± 120	78.2 ± 4.4	54.7 ± 4.3

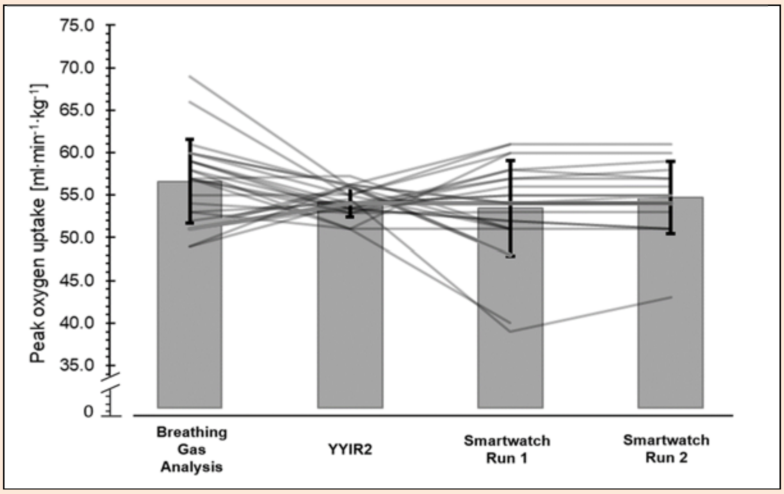

Figure 2. VO_{2max} [$ml \cdot min^{-1} \cdot kg^{-1}$] of each player derived from the different assessment methods.

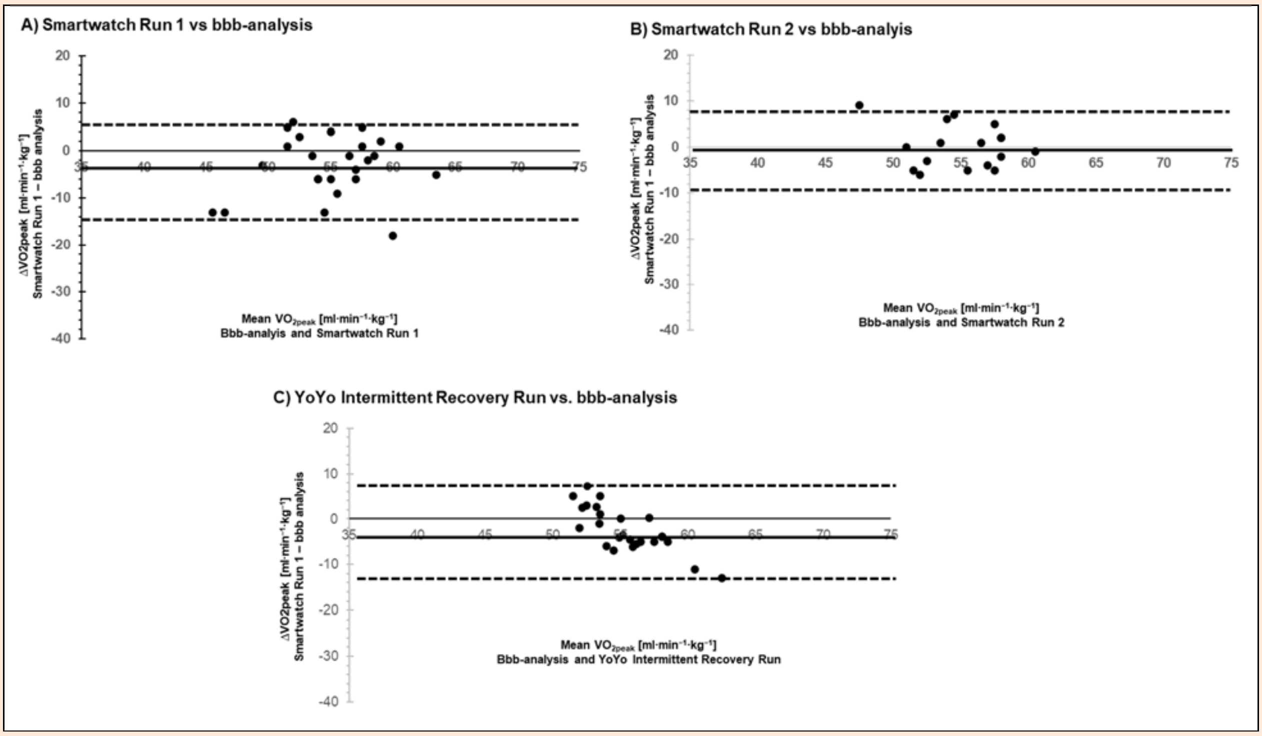

Figure 3. Bland-Altman Plots comparing the different methodologies to assess maximum oxygen uptake. A) Smartwatch derived VO_{2max} (first run) vs respiratory gas analysis ; B) Smartwatch derived VO_{2max} (second run) vs respiratory gas analysis; C) YoYo Intermittent Recovery Run 2 vs respiratory gas analysis.

Our findings expand upon existing research in different populations, which demonstrated a MAPE of 5.6% for $\dot{V}O_{2max}$ assessment using a smartwatch in runners (Düking et al., 2022). Interestingly, a previous study (Düking et al., 2022) revealed no decrease in MAPE with an increase in the number of runs as seen in our results. As the algorithms employed by the smartwatch are not disclosed by the manufacturer in detail, we can only speculate why in the present study the MAPE decreases with one more run. Potentially the algorithm needs a sufficient amount of individual data to calculate VO_{2max} properly which was not available in the present data, but in the data of previous publications. Another differentiating reason is that in our study, we included participants HR_{max}, which was not done in the previously published study with runners (Düking et al., 2022) which could affect VO_{2max} assessment. Consequently, it seems reasonable to manually insert HR_{max} when aiming to use the smartwatch to assess VO_{2max}, and based on our data, it appears advisable to conduct multiple runs with the smartwatch to reduce error rates. Remarkably, even with a relatively low MAPE, the ICCs for $\dot{V}O_{2max}$ assessment following one or two runs using the smartwatch are classified as "poor" and "moderate." This may be attributed to limited variability in our data, as suggested by the standard deviation in Table 1, and a small sample size, which are known to influence the ICC (Koo and Li, 2016). Consequently, future studies should investigate the ICC in larger groups of individuals and with more heterogenous data.

For the YYIR2, our study showed an ICC of 0.17 [95%CI: -5.7; 0.6] and a MAPE of 4.2%; ~2.3 ±·min^{-1}·kg^{-1}. A recent review summarized the literature on the reliability of the YYIR2 and found that ICCs ranged from 0.86 to 0.96 and coefficients of variation ranged from 4.2 to 12.7% (Grgic et al., 2019).

Our differing results for the YYIR2 compared to existing literature are speculative, but similar to our rationale regarding the ICCs for smartwatch-assessed $\dot{V}O_{2max}$, we assume that the lower ICC for YYIR2-assessed $\dot{V}O_{2max}$ in our study may be due to our small sample size and homogeneous data (Koo and Li, 2016).

When comparing $\dot{V}O_{2max}$ assessment errors between YYIR2 and the smartwatch, YYIR2 shows a lower MAPE with only one smartwatch run. However, with two smartwatch runs, the MAPE for smartwatch-assessed $\dot{V}O_{2max}$ is smaller than that for YYIR2-assessed $\dot{V}O_{2max}$.

Limitations of the study and future research perspectives

The results of the smartwatch derived VO_{2max} are limited to runs performed during a warm-up of highly trained soccer players and generalization to other populations or settings should be performed with caution. Given the improvements in validity-related parameters after two runs performed with the smartwatch shown in our study, it could be that validity further increases with additional runs. In our experiments, we limited the number of runs to two because a previous analysis did not demonstrate any improvement in the validity of VO_{2max} with additional runs (Düking et al., 2022). Therefore, we did not assume that increasing the number of runs would improve the validity of VO_{2max}. Future studies should verify if validity of smartwatch assessed VO_{2max} increases with additional runs. As our aim was to assess concurrent criterion related validity of smartwatch derived VO_{2max} values, we did not assess other relevant criteria, such as reliability (Currell and Jeukendrup,

2008) or sensitivity (Düking et al., 2018) which future studies need to assess. Our results are limited to the herein investigated smartwatch, software version and the investigated VO_{2max} parameter range and additionally should be seen with caution due to a small sample size. VO_{2max} assessments of other models or software versions and other VO_{2max} parameter ranges may differ from these results. As alterations in algorithms do not have to be disclosed in detail since the device is a non-medical grade product, future smartwatch versions might yield different results for VO_{2max} assessment. Future research needs to continuously evaluate newer models of smartwatches and would be eased if manufacturers would have to disclose algorithm details and potential alterations.

Conclusion

We showed that when performing two runs with the smartwatch to assess VO2max, the mean absolute percentage error decrease to 1.06%, and agreement with respiratory gas analysis increases to 0.12 $ml \cdot min^{-1} \cdot kg^{-1}$ across the tested $\dot{V}$O2max range, but comparably large confidence intervals indicate that individual values might show larger errors. Additionally, the mean absolute percentage error of the YYIR2 was 4.2% and showed to have poor agreement with respiratory has analysis. If VO_{2max} is assessed with the smartwatch, we advise to perform (at least) two runs to decrease the MAPE of the results. If practitioners are aware of the MAPE after two non-fatiguing runs during the warm-up prior to soccer practice and take this error into account, practitioners might use the herein investigated smartwatch (and software version) to assess VO_{2max} in soccer players within a VO_{2max} range of 49-69 $ml \cdot min^{-1} \cdot kg^{-1}$. To eliminate potential inaccuracies in individual measurements, professionals should meticulously review the $\dot{V}O_{2max}$ results displayed by the smartwatch. If the data appears implausible, verification through gold-standard procedures, such as an all-out ramp test using a respiratory gas analysis, is recommended.

Acknowledgements

The authors thank all subjects for their participation. The results of the present study do not constitute endorsement of the product by the authors. The authors declare no conflict of interest. The datasets generated and analyzed during the current study are not publicly available but are available from the corresponding author upon reasonable request.

References

Arifin, W.N. (2018) A Web-based Sample Size Calculator for Reliability Studies. *Education in Medicine Journal* **10(3)**. https://doi.org/10.21315/eimj2018.10.3.8

Åstrand, P.O. and Ryhming, I. (1954) A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during submaximal work. *Journal of Applied Physiology* **7(2)**, 218-221. https://doi.org/10.1152/jappl.1954.7.2.218

Bangsbo, J., Iaia, F.M. and Krustrup, P. (2008) The Yo-Yo intermittent recovery test: a useful tool for evaluation of physical performance in intermittent sports. *Sports Medicine* **38(1)**, 37-51. https://doi.org/10.2165/00007256-200838010-00004

Bent, B., Goldstein, B.A., Kibbe, W.A. and Dunn, J.P. (2020) Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digital Medicine* **3(1)**, 18. https://doi.org/10.1038/s41746-020-0226-6

Bradley, P.S., Carling, C., Diaz, A.G., Hood, P., Barnes, C., Ade, J., Boddy, M., Krustrup, P. and Mohr, M. (2013) Match performance and physical capacity of players in the top three competitive standards of English professional soccer. *Human Movement Science* **32(4)**, 808-821. https://doi.org/10.1016/j.humov.2013.06.002

Chtourou, H., Hammouda, O., Souissi, H., Chamari, K., Chaouachi, A. and Souissi, N. (2012) Diurnal variations in physical performances related to football in young soccer players. *Asian Journal of Sports Medicine* **3(3)**, 139. https://doi.org/10.5812/asjsm.34604

Clemente, F.M., Moran, J., Ramirez-Campillo, R., Beato, M. and Afonso, J. (2023) Endurance performance adaptations between SSG and HIIT in soccer players: A meta-analysis. *International Journal of Sports Medicine* **45(03)**, 183-210. https://doi.org/10.1055/a-2171-3255

Currell, K. and Jeukendrup, A.E. (2008) Validity, reliability and sensitivity of measures of sporting performance. *Sports Medicine* **38(4)**, 297-316. https://doi.org/10.2165/00007256-200838040-00003

Düking, P., Fuss, F.K., Holmberg, H.C. and Sperlich, B. (2018) Recommendations for Assessment of the Reliability, Sensitivity, and Validity of Data Provided by Wearable Sensors Designed for Monitoring Physical Activity. *Journal of Medical Internet Reearch Mhealth and Uhealth* **6(4)**, e102. https://doi.org/10.2196/mhealth.9341

Düking, P., Hotho, A., Holmberg, H.C., Fuss, F.K. and Sperlich, B. (2016) Comparison of Non-Invasive Individual Monitoring of the Training and Health of Athletes with Commercially Available Wearable Technologies. *Frontiers in Physiology* **7**, 71. https://doi.org/10.3389/fphys.2016.00071

Düking, P., van Hooren, B. and Sperlich, B. (2022) Assessment of Peak Oxygen Uptake with a Smartwatch and its Usefulness for Training of Runners. *International Journal of Sports Medicine* **43(07)**, 642-647. https://doi.org/10.1055/a-1686-9068

Grgic, J., Oppici, L., Mikulic, P., Bangsbo, J., Krustrup, P. and Pedisic, Z. (2019) Test-retest reliability of the Yo-Yo test: A systematic review. *Sports Medicine* **49**, 1547-1557. https://doi.org/10.1007/s40279-019-01143-4

Helgerud, J., Engen, L. C., Wisløff, U., & Hoff, J. A. N. (2001). Aerobic endurance training improves soccer performance. Medicine & Science in Sports & Exercise, 33(11), 1925-1931.

Koo, T.K. and Li, M.Y. (2016) A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine* **15(2)**, 155-163. https://doi.org/10.1016/j.jcm.2016.02.012

Lee, J.M., Kim, Y. and Welk, G.J. (2014) Validity of consumer-based physical activity monitors. *Medicine and Science in Sports and Exercise* **46(9)**, 1840-1848. https://doi.org/10.1249/MSS.0000000000000287

Garmin Ltd, G. (2019) Forerunner (r) 245/245 Music Benutzerhandbuch [Accessed 21.12.20202020].

Macfarlane, D.J. and Wong, P. (2012) Validity, reliability and stability of the portable Cortex Metamax 3B gas analysis system. *European Journal of Applied Physiology* **112(7)**, 2539-2547. https://doi.org/10.1007/s00421-011-2230-7

Mayorga-Vega, D., Casado-Robles, C., Guijarro-Romero, S. and Viciana, J. (2024) Criterion-Related Validity of Consumer-Wearable Activity Trackers for Estimating Steps in Primary Schoolchildren under Controlled Conditions: Fit-Person Study. *Journal of Sports Science and Medicine* **23(1)**, 79-96. https://doi.org/10.52082/jssm.2024.79

McKay, A.K.A., Stellingwerff, T., Smith, E.S., Martin, D.T., Mujika, I., Goosey-Tolfrey, V.L., Sheppard, J. and Burke, L.M. (2022) Defining Training and Performance Caliber: A Participant Classification Framework. *International Journal of Physiology and Performance* **17(2)**, 317-331. https://doi.org/10.1123/ijspp.2021-0451

Nunnally, J.C. (1978) Psychometric Theory: 2d Ed: McGraw-Hill.

Peake, J.M., Kerr, G. and Sullivan, J.P. (2018) A Critical Review of Consumer Wearables, Mobile Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in Physically Active Populations. *Frontiers in Physiology* **9**, 743. https://doi.org/10.3389/fphys.2018.00743

Poole, D.C. and Jones, A.M. (2017) Measurement of the maximum oxygen uptake Vo2max: Vo2peak is no longer acceptable. *Journal of Applied Physiology* **122(4)**, 997-1002. https://doi.org/10.1152/japplphysiol.01063.2016

Poole, D.C. and Richardson, R.S. (1997) Determinants of oxygen uptake: implications for exercise testing. *Sports Medicine* **24**, 308-320. https://doi.org/10.2165/00007256-199724050-00003
Schimpchen, J., Correia, P.F. and Meyer, T. (2023) Minimally invasive ways to monitor changes in cardiocirculatory fitness in running-based sports: A systematic review. *International Journal of Sports Medicine* **44(02)**, 95-107. https://doi.org/10.1055/a-1925-7468
Slimani, M., Znazen, H., Miarka, B. and Bragazzi, N.L. (2019) Maximum oxygen uptake of male soccer players according to their competitive level, playing position and age group: implication from a network meta-analysis. *Journal of Human Kinetics* **66**, 233. https://doi.org/10.2478/hukin-2018-0060
Sperlich, B. and Holmberg, H.C. (2017) Wearable, yes, but able…?: it is time for evidence-based marketing claims! *British Journal of Sports Medicine* **51(16)**, 1240. https://doi.org/10.1136/bjsports-2016-097295
Statista Market Insights (2023) Smartwatches. Available from URL: https://www.statista.com/outlook/hmo/digital-health/digital-fitness-well-being/fitness-trackers/smartwatches/worldwide#key-players [Accessed 7 February 2024].
Stolen, T., Chamari, K., Castagna, C. and Wisloff, U. (2005) Physiology of soccer: an update. *Sports Medicine* **35(6)**, 501-536. https://doi.org/10.2165/00007256-200535060-00004
Thomas, A., Dawson, B. and Goodman, C. (2006) The yo-yo test: reliability and association with a 20-m shuttle run and VO2max. *International Journal of Sports Physiology and Performance* **1(2)**, 137-149. https://doi.org/10.1123/ijspp.1.2.137
van Hooren, B., Souren, T. and Bongers, B.C. (2024) Accuracy of respiratory gas variables, substrate, and energy use from 15 CPET systems during simulated and human exercise. *Scandinavian Journal of Medicine and Science in Sports* **34(1)**, e14490. https://doi.org/10.1111/sms.14490
Waldrop, M.M. (2016) The chips are down for Moore's law. *Nature* **530(7589)**, 144-147. https://doi.org/10.1038/530144a
Wing, C.E., Turner, A.N. and Bishop, C.J. (2020) Importance of strength and power on key performance indicators in elite youth soccer. *The Journal of Strength & Conditioning Research* **34(7)**, 2006-2014. https://doi.org/10.1519/JSC.0000000000002446
Wisloeff, U., Helgerud, J.A., Hoff, J..M. (1998) Strength and endurance of elite soccer players. *Medicine & Science in Sports & Exercise* **30(3)**, 462-467. https://doi.org/10.1097/00005768-199803000-00019

Key points

- Assessing maximal oxygen consumption (VO_{2max}) is relevant for soccer players, but require expensive and time consuming procedures e.g. by employing respiratory gas analysis
- Selected Smartwatches provide players with VO_{2max} values if worn e.g. during non-fatiguing outdoor runs if sufficient GPS and heart rate data is available
- The mean absolute percentage error between the Smartwatch estimated VO_{2max} (after 2 runs) and a respiratory gas analyzer was 1.06% in our study.

✉ **Peter Düking**
Department of Sports Science and Movement Pedagogy Technische Universität Braunschweig Braunschweig, Germany

AUTHOR BIOGRAPHY

Peter DÜKING
Employment
Jun.-Professor at the Department of Sports Science and Movement Pedagogy, Technische Universität Braunschweig, Braunschweig, Germany
Degree
Jun.-Prof.; Dr.
Research interests
Exercise & Training; Technology
E-mail:
peterdueking@gmx.de

Ludwig RUF
Employment
TSG ResearchLab, Germany
Degree
Dr.
Research interests
Exercise physiology, performance testing, team sports, soccer
E-mail: ludwig.ruf@tsg-researchlab.de

Stefan ALTMANN
Employment
CEO at TSG ResearchLab and Head of Performance Diagnostics at Karlsruhe Institute of Technology, Germany
Degree
Dr.
Research interests
Exercise physiology, performance testing, team sports, soccer
E-mail: Stefan.altmann@kit.edu

Maximiliane THRON
Employment
Institute of Sports and Sports Science, Karlsruhe Institute of Technology, Karlsruhe, Germany
Degree
M.Sc.
Research interests
Exercise physiology, performance testing, team sports, soccer
E-mail: maximiliane.thron@kit.edu

Philipp KUNZ
Employment
Integrative and Experimental Exercise Science, Department of Sport Science, University of Würzburg, Würzburg, Germany
Degree
M.Sc.
Research interests
Exercise physiology, performance testing, team sports, soccer
E-mail: philipp.kunz90@outlook.com

Billy SPERLICH
Employment
Integrative and Experimental Exercise Science, Department of Sport Science, University of Würzburg, Würzburg, Germany
Degree
Prof., Dr.
Research interests
Exercise & Training
E-mail:
Billy.sperlich@uni-wuerzburg.de