International Journal of Aquatic Research and Education

Volume 2 | Number 2 Article 6

5-1-2008

Reliability of Peak Cardiorespiratory Responses During Aquatic Treadmill Exercise

William M. Silvers
Eastern Washington University, msilvers@ewu.edu

Dennis G. Dolny
University of Idaho

Recommended Citation
Silvers, William M. and Dolny, Dennis G. (2008) "Reliability of Peak Cardiorespiratory Responses During Aquatic Treadmill Exercise," *International Journal of Aquatic Research and Education*: Vol. 2: No. 2, Article 6.
DOI: https://doi.org/10.25035/ijare.02.02.06
Available at: https://scholarworks.bgsu.edu/ijare/vol2/iss2/6

International Journal of Aquatic Research and Education, 2008, 2, 140-150

Reliability of Peak Cardiorespiratory Responses During Aquatic Treadmill Exercise

William M. Silvers and Dennis G. Dolny

Twenty-four college-age participants took part in 2 protocols on an aquatic treadmill (ATM) submerged to the xiphoid process. ATM speed was increased to 212.2 ± 19.2 m/min, and water-jet resistance was increased 10% every minute thereafter. Rest between sessions was at least 6 days. Oxygen consumption (VO_2), heart rate (HR), minute ventilation (V_E), tidal volume (V_T), breathing frequency (*f*), and respiratory-exchange ratio (RER) were measured continuously. Rating of perceived exertion (RPE) was recorded immediately after each test, and blood lactate (LA) was measured 3 min after. There were no significant differences for Trial 1 vs. Trial 2 for any variable. ICCs were very strong (r = .90–.99), and coefficients of variance (CVs) were low (1.3–4.7%) for VO_{2peak}, HR, V_E, and V_T; ICCs were moderate (r = .73–.76) and CVs were greater (2.5–9.3%) for *f*, RER, and LA. The ATM VO_{2peak} protocol used in this study produces consistent, reproducible VO_{2peak} values.

Keywords: water, hydrostatic, VO_{2peak}, cross-training, aerobic

The measurement of peak oxygen uptake (VO_{2peak}) has been a staple of performance testing in the field of exercise physiology for over 80 years. Not without its controversy (Howley, Bassett, & Welch, 1995; Noakes, 1997), this test is used to evaluate present aerobic capacity and adaptations to the cardiorespiratory system as a result of physical training. Many protocols for the maximal-oxygen-consumption test have been developed for the treadmill. With any physical test or performance measure the reproducibility, or reliability, of the measure should be examined (Atkinson & Neville, 1998; Hopkins, 2000). Performance-test reliability can be evaluated by testing for any significant changes in mean score, calculating the standard estimate of the mean (typical error), coefficient of variation (CV, typical error expressed as a percentage of the mean), intraclass correlation coefficients (ICC), and total error (TE) caused by technological or measurement error (Atkinson & Neville; Hopkins).

There have been numerous studies that established the reproducibility of VO_{2peak} testing on a land-based treadmill (TM; Froelicher et al., 1974; Harling,

Silvers is with the Dept. of Physical Education, Health and Recreation, Eastern Washington University, Cheney, WA 99004. Dolny is with the Dept. of Health, Physical Education, Recreation and Dance, University of Idaho, Moscow, ID 83844-2401.

Tong, & Mickleborough, 2003; Katch, Sady, & Freedson, 1982; McArdle, Katch, Pechar, Jacobson, & Ruck, 1972; Mitchell, Sproule, & Chapman, 1958; Taylor, 1944; Wilmore et al., 1980, 1985). Test–retest reliability coefficients have ranged from .90 to .96, with CVs ranging from 2.3% to 6.5%. Reliability is considered to primarily be influenced by biological variability, usually accounting for >90% of variability according to Katch et al. (1982), and to a lesser extent by technological error of measurement (TE). Baumgartner and Jackson (1991) recommend that for psychomotor tests ICCs should reach or exceed .80 to demonstrate strong reliability. Furthermore, low CVs and TEs provide insights specific to what would be considered the smallest worthwhile change (SWC) whereby researchers who measured this or a greater magnitude of change from one test to the next would determine this difference to be meaningful beyond that of the expected variability (Hopkins, 2000), thereby indicating the test performance to be unreliable.

Deep-water running and shallow-water running (SWR) have become popular as alternatives to land-based running, in part because of their potential to reduce repetitive strain and stress to the lower extremities from ground-reaction forces. Participants typically perform running actions in an indoor pool immersed to the level of the neck for deep-water running and around waist-deep for SWR. Underwater or aquatic treadmills (ATMs) are becoming more available. ATMs present an SWR option that mitigates frontal water resistance by eliminating forward locomotion through a body of water. Consequently, a more natural walking or running gait pattern is possible, which might enhance the specificity of SWR training. Recent advances in technology have improved the functionality of ATMs, offering broader flexibility in treadmill speeds, water-submersion levels, and external fluid resistance via water jets. We have previously determined that peak cardiorespiratory responses on ATMs were similar to those on land-based treadmills (TMs) in recreationally active male and female participants (Silvers, Rutledge, & Dolny, 2007). Less information is available regarding the reliability of aquatic VO_{2peak} protocols, especially those incorporating an ATM.

Therefore, the purpose of this study was to determine the reliability of peak cardiorespiratory responses elicited during maximal-effort protocols using an ATM. We hypothesized that the variability in peak measures would be comparable to previously reported reliability scores for TM exercise (Froelicher et al., 1974; Harling et al., 2003; Katch et al., 1982; McArdle et al., 1972; Mitchell et al., 1958; Taylor, 1944; Wilmore et al., 1980, 1985). We believe this might be the first study to verify the reliability of the cardiorespiratory response to any form of aquatic exercise (SWR, deep-water running, or ATM) at any exercise intensity. For that reason, this study represents a particularly significant and important contribution to the physiology and aquatic-exercise literature.

Method

Participants

Twenty-four recreationally competitive male (n = 13) and female (n = 11) runners participated in this investigation (age 25 ± 3 years, height 174.9 ± 11.1 cm, weight 68.8 ± 11.1 kg). Criteria for participation included at least 6 months of consistent aerobic training (≥3 sessions/week, ≥30 min/session). All participants completed informed-consent waivers consistent with the policy statement regarding the use of human participants and written informed consent as reviewed and approved by the University of Idaho Human Assurance Committee.

Experimental Procedures

A test–retest research design was employed to investigate the effects of repeated maximal-effort ATM protocols on cardiorespiratory endurance, ratings of perceived exertion (RPE), and blood lactate (LA) measures. Each participant completed two maximal-exertion running protocols on an ATM and was asked to refrain from eating for at least 4 hr before testing. Rest between testing sessions was 7 ± 1 day.

Equipment

ATM protocols were performed on a HydroWorx 2000 (HydroWorx, Middletown, PA) that consisted of a small pool kept at 28 °C with a treadmill built into an adjustable-height floor. Water jets inset at the front of the pool provided an adjustable water-flow resistance.

Expired air was analyzed using an automated metabolic system (TrueOne 2400, ParvoMedics, Sandy, UT) that was calibrated immediately before each testing session. The metabolic cart's reliability and validity have been reported elsewhere (Crouter, Antczak, Hudak, DellaValle, & Haas, 2006). Participants wore water-resistant chest-strap transmitters (Polar T31, Polar, Lake Success, NY) to monitor heart rate. We assessed perceived exertion immediately after each test using Borg's 15-point RPE scale (Borg, 1982). To measure postexercise LA, we used a handheld lactate analyzer (Lactate Pro, ARKRAY, Inc., Minami-Mu, Kyoto, Japan).

Testing Protocols

Table 1 summarizes the ATM testing protocols. Male participants wore spandex shorts, and female participants wore spandex shorts and a sports bra. We established the initial and final treadmill speeds based on information solicited from individual

Table 1 Descriptions of Both Aquatic-Treadmill Protocols

Trial	Initial workload, *M* (*SD*)	Progression	Final workload, *M* (*SD*)
1	149.3 (16.8) m/min, 40% jets	Increased speed 13.4 m/min every minute for 4–5 min, then increased jets 10% every minute to volitional fatigue	212.2 (19.2) m/min, 75.7% (13.8%) jets
2	149.3 (16.8) m/min, 40% jets	Increased speed 13.4 m/min every minute for 4–5 min, then increased jets 10% every minute to volitional fatigue	212.2 (19.2) m/min, 80.4% (11.9%) jets

participants relative to their typical daily workout running paces and, if available, best performance times in 5- to 10-km road races in the 3 months before testing. Water jets were directed at the participant's torso to provide an additional adjustable resistance during testing. Participants were submerged to the xiphoid process and positioned approximately 1 m away from the water jets to standardize the amount of fluid resistance. Underwater sagittal- and frontal-plane camcorders connected to video screens in front of the pool provided the investigators and participants real-time feedback about position in relation to the water jets and running gait, to ensure that the ATM protocol did not degrade participants' running form when they neared physiological exhaustion. Based on pilot-testing sessions, we chose to use 40% water-jet resistance as the beginning resistance for the first ATM speed to promote normal running gait and minimize "float time" over the treadmill belt. After a 4- to 6-min warm-up, participants began the test running at their predetermined initial speed with 40% water-jet resistance for 1 min. Thereafter, we increased speed 13.4 m/min every minute for 5–6 min to a predetermined maximum speed, while water-jet resistance stayed constant at 40%. Once participants reached maximum speed, we increased water-jet resistance incrementally 10% every minute until participants reached the point of volitional exhaustion. Air temperature and relative humidity in the room were maintained at 24 ± 1.0 °C and 43% ± 2.0%, respectively.

We applied the following criteria to verify that participants had achieved a valid maximal-effort test: heart rate [HR] ±5 beats/min of age-predicted maximum (220 – age), LA ≥9 mM, respiratory-exchange ratio (RER) ≥1.10, and VO_2 increases <0.15 L/min with increases in workload at end of test (Taylor, Buskirk, & Henschel, 1955). We selected these criteria because they represent typical criteria commonly used in previous TM protocols.

Variables

We continuously sampled data for VO_2, HR, tidal volume (V_T), ventilation ($V_{E[BTPS]}$), breathing frequency (*f*), and RER during testing. Four 15-s samples around the highest 15-s VO_2 sample were averaged to express peak 1-min values for each variable. Three minutes after completion of the ATM protocol, we drew 5 μl of whole blood from participants' fingertips and placed them on an analyzer testing strip. LA values were reported in mmol/L.

Statistical Analysis

Peak 1-min values for VO_2, VCO_2, HR, V_E, V_T, *f*, RER; postexercise LA and RPE; and test time were recorded for each trial. We checked the data for heteroscedasticity using plots of the raw and log-transformed data, with the change scores plotted against the mean scores and the uniformity of the scatter checked (Bland & Altman, 1986). If we detected heteroscedasticity, analyses were conducted with log-transformed data. We employed one-way repeated-measures ANOVAs to determine significant differences between trials for each variable, with the level of significance set at $p < .05$. Pearson's *r,* ICC, TE, and CV were calculated for each variable. TE was used to represent technological error of measurement. SWC was calculated as 0.2 × between-participants *SD* (Cohen, 1988; Hopkins, 2000).

Results

Means, standard deviations, and reliability measures for each variable are presented in Table 2. There were no significant differences between trials for any variable, with VO_{2peak} mean differences of 0.02 L/min (95% CI of –0.03 to 0.08 L/min), or 0.4% (95% CI of –0.8 to 1.9%); ICC of .99 (95% CI of 0.968–0.995); CV of 2.2% (95% CI of 1.4–3.0%); and TE of 2.2%. Linear-regression analysis of Trial 1 versus Trial 2 (intercept = 0.11, slope = 0.97) appeared similar to the line of identity (intercept = 0, slope = 1.00; Figure 1), with 95% limits of agreement within ±0.22 L/min.

Table 2 Test–Retest Reliability Parameters for All Variables

Variable	Trial 1, *M* (*SD*)	Trial 2, *M* (*SD*)	TE[a]	CV	*r*	ICC
VO_2 (L/min)	3.65 (0.80)	3.67 (0.80)	0.09	2.2%	.99	.99
HR (beats/min)	187 (13)	187 (14)	2.27	1.3%	.96	.97
VCO_2 (L/min)	4.09 (0.90)	4.11 (1.00)	0.12	2.9%	.98	.99
V_E (L/min)	128.1 (25.9)	129.4 (27.1)	5.30	3.7%	.90	.97
V_T (L/min)	2.39 (0.51)	2.42 (0.60)	0.12	4.7%	.97	.95
f (breaths/min)	54 (6)	54 (6)	2.70	4.9%	.76	.82
RER	1.12 (0.05)	1.12 (0.10)	0.03	2.5%	.74	.73
LA (mM)	12.1 (1.7)	12.4 (1.9)	0.80	6.8%	.75	.84
RPE (Borg 6–20)	19 (1)	19 (1)	0.49	2.2%	.60	.53
Test time (min)	10.0 (1.3)	10.2 (1.3)	0.69	7.0%	.69	.71

Note. TE = total error; CV = coefficient of variation; ICC = intraclass correlation coefficient; HR = heart rate; V_E = minute ventilation; V_T = tidal volume; RER = respiratory-exchange ratio; LA = blood lactate; RPE = rating of perceived exertion.

[a]Units for TE are the same as variable units.

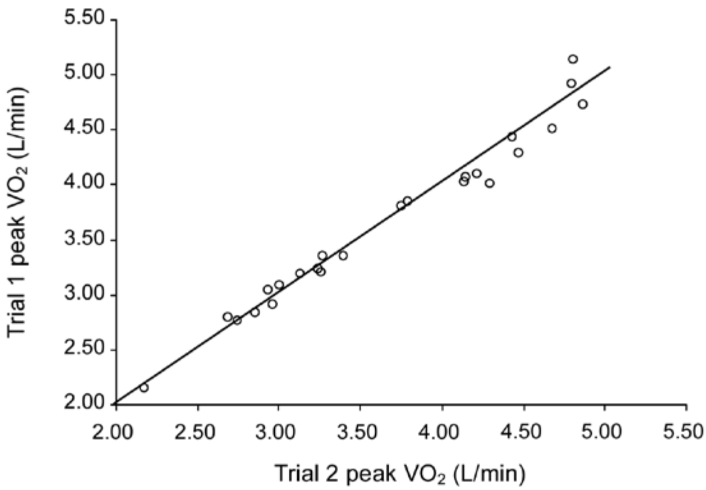

Figure 1 — Plot of VO_{2peak} for Trial 1 versus Trial 2.

Table 3 Percentage of Peak Criteria Met for Each Trial

	% of Trials in Which Criteria Were Met	
Criteria for peak test[a]	Trial 1	Trial 2
$VO_2 \leq 0.15$-L/min increase	71%	79%
Heart rate ±5 beats/min age-predicted maximum	47%	47%
Respiratory-exchange ratio ≥ 1.10	83%	87%
Blood lactate ≥ 9 mM	94%	96%

[a]Descriptions of criteria are listed in the Methods section.

Table 4 Summary of This and Previous Studies on Treadmill Reliability

Authors	Trial 1 VO_2, *M* (*SD*)	Trial 2 VO_2, *M* (*SD*)	CV	*r*
Current study	3.65 (0.8)	3.67 (0.8)	2.2%	.99
Froelicher et al. (1974; Bruce protocol)	3.41 (0.33)	3.50 (0.50)	5.4%	.82
Froelicher et al. (1974; Taylor protocol)	3.74 (0.44)	3.70 (0.50)	4.1%	.92
Froelicher et al. (1974; Balke protocol)	3.40 (0.39)	3.40 (0.40)	7.9%	.63
Harling et al. (2003)	4.31 (0.47)	4.39 (0.61)	2.9%	.96
McArdle et al. (1972)	2.15 (0.33)	2.17 (0.33)	2.5%	.92
Mitchell et al. (1958)	3.06 (0.46)	3.07 (0.44)	4.5%	.92
Wilmore et al. (1980)	3.27 (0.56)	3.29 (0.63)	—	.90
Wilmore et al. (1985)	3.54 (0.35)	3.54 (0.35)	—	.94

Note. A dash indicates the CV was not reported.

Using the selected criteria, on average participants reached a plateau in VO_2 in 75% (36/48) of the tests, and they achieved HR_{max} and peak RER and LA in 47%, 85%, and 95% of the tests, respectively (Table 3). ICCs were very strong (r = .90–.98), and TE and CV (1.3–6.0%) were low for peak values of HR, V_E, and V_T. ICCs were moderate (r = .73–.76), and TE and CV (2.5–9.3%) tended to be greater, for f, RER, and LA.

Discussion

We designed the current study to evaluate the reliability of peak cardiorespiratory responses elicited during maximal-effort ATM testing. Our results are in line with previously reported reliability data for VO_{2peak} TM protocols. The ICC of .99 (95% CI = 0.968–0.995) and CV of 2.2% are similar to the best previously reported TM reliability of ICCs = .90–.96 and CVs = 2.3–6.5% (Froelicher et al., 1974; Harling et al., 2003; Katch et al., 1982; McArdle et al., 1972; Mitchell et al., 1958; Taylor, 1944; Wilmore et al., 1980, 1985; Table 4). The average test duration for both trials was ~10 min, which is within the optimal range for TM VO_{2peak} testing (Astorino et al., 2004).

The ATM allowed us to administer workloads during the ATM protocol by manipulating treadmill speed and adjusting fluid resistance (i.e., water jets), as well as customizing water height to ensure that each participant was submerged to the xiphoid level. Therefore, we feel that the ATM afforded each participant an individualized opportunity to exercise up to his or her maximal potential.

We patterned the current protocol after previously established land-based TM protocols that initially increase running speed to a predetermined level, then increased TM incline to fatigue (Astrand & Rodahl, 1970). Using the same protocol in an earlier study, we demonstrated that ATM yielded VO_{2peak} values comparable to those of TM (Silvers et al., 2007). We believe that the incremental increase in treadmill speed followed by increases in water-jet resistance corresponds to a substantial degree to the increase in treadmill speed and incline experienced with land protocols. We would have to quantify the added external work of running against different jet-resistance settings before we could absolutely support this comparison with TM. Different protocols such as combining increases in treadmill speed and jet-resistance levels throughout all stages of testing might yield different results. All measures representing performance reliability indicated that the ATM protocol used in the current study provided a reliable way to assess peak cardiorespiratory indices for the measured variables.

The SWC identifies the magnitude of change required to elicit a meaningful or significant improvement in VO_{2peak}. The SWC, calculated as a proportion of the effect size, represents the magnitude of improvement in a variable as a function of the between-participants standard deviation of the particular cohort (Hopkins, 2000). Knowledge of the magnitude of the SWC helps a coach or athlete interpret the usefulness of a change in performance or physiological measurement. We chose an indirect method of estimating the SWC, using a small Cohen's effect size as suggested by Hopkins. For the current study, based on the SWC for VO_{2peak} of 0.16 L/min (4.4% of mean), mean bias in VO_{2peak} of 0.4% (–0.9 to 1.8% lower and upper 95% CIs), and TE of 2.2%, we would rate this ATM VO_{2peak} test as having sufficient statistical power to be able to detect real and meaningful within-participant changes. We are confident that in subsequent training interventions using a similar participant population and ATM testing procedures, a change in VO_{2peak} that meets or exceeds 4.4% for the change can be considered both meaningful and real.

Regarding the female participants, there is a strong possibility that there are both intra- (between test and retest) and interindividual variability (in each test) because of menstrual-cycle phases. Indeed, taking into account that both exercises were interspersed by a rest period of at least 6 days, the test and retest exercises might have been performed by most women with different hormonal environments. Most previous research has reported no difference in VO_{2peak} test results resulting from menstrual-cycle phase (Janse de Jonge, 2003).

We believe that the remaining variability between ATM trials for VO_{2peak} is a result of technological or instrumentation error. The current study used a Parvo-Medics TrueOne 2400 automated metabolic cart for all testing. This system has been shown to be very reliable and valid when compared with the Douglas-bag criterion procedure (Crouter et al., 2006). Crouter et al. had participants cycle from 50 to 250 W. VO_{2peak} reached 3.65 ± 0.18 L/min and yielded mean error values of 0.04, 0.03, and 1.34 L/min for VO_2, VCO_2, and $V_{E(BTPS)}$, respectively. Pearson's *r*s of .99 for VO_2, .99 for VCO_2, and .97 for $V_{E(BTPS)}$ with corresponding CVs of 4.7%, 5.7%, and 7.3% were more reliable than the author's Douglas-bag techniques.

Data-sampling rate appears to influence the existence of a VO_2 plateau at peak exercise. Myers, Walsh, Sullivan, and Froelicher (1990) reported that as much as 20% differences in measured VO_2 at the end of a test occur because of the method of sampling gas-exchange data (breath by breath, 15, 30, or 60 s) and that a VO_2 plateau was more commonly observed with a shorter sampling period. The current study used a 15-s sampling interval, which might have contributed somewhat to the observed VO_2 plateau rate of 75%.

For years there has been a debate in the exercise-physiology literature about whether a leveling off or plateau of VO_2 with an increase in workload at the end of the test is required for the measurement to be considered valid (Howley et al., 1995; Noakes, 1997). The occurrence of a VO_2 plateau near the end of a maximal exercise test ranges from a high of ~94% (Taylor et al., 1955) to a low of 7% (Froelicher et al., 1974) using the Bruce TM protocol. In a review of 29 studies on VO_{2max} testing, Howley et al. determined that only ~50% of all participants tested demonstrated a true VO_2 plateau. The authors did note that the lack of a plateau does not necessarily mean that VO_{2peak} was not reached.

Although a plateau of VO_2 near the end of a maximal oxygen-consumption test is the most traditional evidence that a true maximal effort has been achieved, researchers typically look for other indicators to verify a valid test, especially when no VO_2 plateau is observed. HR_{max}, RER, and LA are common variables cited. Although they are accepted as secondary criteria, there appears to be no consensus on what their cutoffs should be. Using our criteria cutoffs (Table 3) for VO_2 plateau, HR_{max}, RER, and LA, we observed that these criteria, when averaged over both trials, were achieved during 75%, 47%, 85%, and 95% of the tests, respectively. Of the total of 48 tests conducted, all four criteria were met in 10 tests, three criteria in 24 tests, two criteria in 12 tests, and one criterion in 2 tests. This compares favorably to the results from Duncan, Howley, and Johnson (1997), who used criteria similar to those used in the current study and reported that VO_2 plateau, HR_{max}, RER, and LA criteria were met for 50%, 40%, 90%, and 90% of the tests, respectively. Differences between studies for achieving criteria might be the result of differences in the criteria selected, testing protocols, and participant fitness levels.

HR_{max} has demonstrated reasonable reproducibility with repeated maximal TM testing, ranging from .76 (McArdle et al., 1972) to .81 (Taylor, 1944; Wilmore et al., 1985) and .82 (Wilmore et al., 1980), with HR CV of 1.5–4.0% (Froelicher et al., 1974; Katch et al., 1982). In the current study HR_{max} was very reproducible, as demonstrated by the high ICC (r = .96) and low CV (1.3%). RER has been demonstrated as less reliable, with ICCs ranging from .44 (Wilmore et al., 1985) to .48 (Wilmore et al., 1980) and .52 (McArdle et al.). The current study's RER data appear to be more reliable than those in most of these previous studies, suggesting that the cardiorespiratory responses of the participants in this study were very reproducible from Test 1 to Test 2.

Less is known regarding the reliability of LA and RPE during maximal testing. Krustrup et al. (2003) reported LA concentrations at exhaustion of two trials of an intermittent shuttle-running test (Yo-Yo test) to be similar (mean difference of 0.1 ± 0.6 mmol/L). At the same time, they found large intraindividual variations

DOI: https://doi.org/10.25035/ijare.02.02.06

(CV = 17%). Doherty, Smith, Hughes, and Collins (2001) determined that RPE solicited at 30-s intervals of a supramaximal running test lasting 3.3 ± 0.6 min administered on three separate occasions yielded ICCs of .78–.87 and CVs of 4.4–6.0%, which was in line with ICCs of .70–.90 previously reported (Eston & Williams, 1988). In the current study we solicited the RPE immediately after the test, once the treadmill stopped and we had removed the respiratory valve. Whether RPE solicited immediately after versus during the last minute of the test yields greater reliability awaits further study.

Although not considered a secondary criterion in maximal-exercise testing, $V_{e(BTPS)}$ has demonstrated moderate reliability, with ICCs of .53 (Taylor, 1944), .73 (Wilmore et al., 1980), and .78 (McArdle et al., 1972; Wilmore et al., 1985). The ICC of .90 and CV of 3.7% for $V_{e(BTPS)}$ in the current study suggest that $V_{E(BTPS)}$ was at least as reliable in ATM as in previous TM protocols and raises the possibility of examining the role of $V_{e(BTPS)}$ in evaluating the extent of effort during these tests.

We might have observed different results in the current study if we had used a different water depth. Gleim and Nicholas (1989) demonstrated that during running at 134.1, 147.5, and 160.9 m/min, VO_2 and HR were higher as water levels rose from ankle to patella to midthigh than with land running. They also reported that running in waist-deep water produced VO_2 values comparable to those seen during TM running at speeds of 134.1 m/min and faster. This result suggests that water-submersion level might considerably influence cardiorespiratory responses during ATM peak-exercise testing protocols. Immersion to the xiphoid process has been shown to decrease limb loading by 71% versus 57% and 85% for submersion to the waist and seventh cervical vertebra, respectively (Harrison, Hillman, & Bulstrode, 1992). When buoyancy is inadequate to provide substantial limb unloading, as is typically seen in water levels below the waist, drag forces imposed by fluid resistance substantially elevate the metabolic cost, as evidenced by increased VO_2, VO_2 cost per stride, and HR (Gleim & Nicholas; Pohl & McNaughton, 2003). Conversely, when water-submersion levels meet or exceed waist height, increases in buoyancy counteract concomitant increases in workload imposed by fluid resistance, resulting in similar or reduced VO_2 and HR (Gleim & Nicholas; Pohl & McNaughton). Therefore, readers should limit generalizing the results of the current study to participants who exercise in xiphoid-deep water.

Our results might not be comparable to results when an ATM protocol is conducted with different water temperatures. Our water temperature was 28 °C, which might be viewed as just below a thermoneutral temperature condition. Craig and Dvorak (1969) point out that exercise intensity can lower the acceptable level of thermoneutrality during moderate- to high-intensity exercise. With the high intensity of exercise performed in this study, we feel that water temperature was not a limiting factor for peak cardiorespiratory responses during ATM.

Within the limitations of the single water depth and the protocol we selected in this study and in light of our findings, it appears that ATM VO_{2peak} testing is comparable to land-based TM protocols in terms of reproducibility of physiological performance measures. These results confirm and extend our earlier findings that VO_{2peak} responses in ATM compare favorably with those in land-based TM exercise. ATM might be a viable training alternative to maintain or improve fitness level for injured and healthy individuals alike.

References

Astorino, T.A., Reitschel, J.C., Tam, P.A., Taylor, K., Johnson, S.M., Freedman, T.P., et al. (2004). Reinvestigation of optimal duration of VO_{2max} testing [Electronic version]. *Journal of Exercise Physiology, 7*(6), 1–8.

Astrand, P-O., & Rodahl, K.R. (1970). *Textbook of work physiology* (pp. 16–18). New York: McGraw Hill.

Atkinson, G., & Neville, A.M. (1998). Statistical methods for assessing measurement error (reliability) in variables relevant to sports medicine. *Sports Medicine (Auckland, N.Z.), 26,* 217–238.

Baumgartner, T.A., & Jackson, A.S. (1991). *Measurement for evaluation in physical education and exercise science* (4th ed.). Dubuque, IA; Wm. C. Brown.

Bland, J.M., & Altman, D.G. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet, 1*(8476), 307–310.

Borg, G.A. (1982). Psychophysical bases of perceived exertion. *Medicine and Science in Sports and Exercise, 14,* 377–381.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.

Craig, A.B., & Dvorak, M. (1969). Thermal regulation during water immersion. *Journal of Applied Physiology, 21*(5), 1577–1585.

Crouter, S.E., Antczak, A., Hudak, J.R., DellaValle, D.M., & Haas, J.D. (2006). Accuracy and reliability of the ParvoMedics TrueOne 2400 and MedGraphics VO2000 metabolic systems. *European Journal of Applied Physiology, 98,* 139–151.

Doherty, M., Smith, P.M., Hughes, M.G., & Collins, D. (2001). Rating of perceived exertion during high-intensity treadmill running. *Medicine and Science in Sports and Exercise, 33,* 1953–1958.

Duncan, G.E., Howley, E.T., & Johnson, B.N. (1997). Applicability of VO_{2max} criteria: Discontinuous versus continuous protocols. *Medicine and Science in Sports and Exercise, 29,* 273–278.

Eston, R.G., & Williams, J.G. (1988). Reliability of ratings of perceived effort regulation of exercise intensity. *British Journal of Sports Medicine, 22,* 153–155.

Froelicher, V.F., Brammell, H., Davis, G., Noruega, I., Stewart, A., & Lancaster, M.C. (1974). A comparison of three maximal treadmill exercise protocols. *Journal of Applied Physiology, 36,* 720–725.

Gleim, G.W., & Nicholas, J.A. (1989). Metabolic costs and heart rate responses to treadmill walking in water at different depths and temperatures. *American Journal of Sports Medicine, 17*(2), 248–252.

Harling, S.A., Tong, R.J., & Mickleborough, T.D. (2003). The oxygen uptake response running to exhaustion at peak treadmill speed. *Medicine and Science in Sports and Exercise, 35*(4), 663–668.

Harrison, R.H., Hillman, M., & Bulstrode, S.J. (1992). Loading of the lower limb when walking partially immersed: Implications for clinical practice. *Physiotherapy, 78*(3), 164–166.

Hopkins, W.G. (2000). Measures of reliability in sports medicine and science. *Sports Medicine (Auckland, N.Z.), 30,* 1–15.

Howley, E.T., Bassett, D.R., Jr., & Welch, H.G. (1995). Criteria for maximal oxygen uptake: Review and commentary. *Medicine and Science in Sports and Exercise, 27,* 1292–1301.

Janse de Jonge, X.A. (2003). Effects of the menstrual cycle on exercise performance. *Sports Medicine (Auckland, N.Z.), 33*(11), 833–851.

Katch, V.L., Sady, S.S., & Freedson, P. (1982). Biological variability in maximum aerobic power. *Medicine and Science in Sports and Exercise, 14,* 21–25.

Krustrup, P., Mohr, M., Amstrup, T., Rysgaard, T., Johansen, J., Steensberg, A., et al. (2003). The Yo-Yo intermittent recovery test: Physiological response, reliability, and validity. *Medicine and Science in Sports and Exercise, 35,* 697–705.

McArdle, W.D., Katch, F.I., Pechar, G., Jacobson, L., & Ruck, S. (1972). Reliability and interrelationships between maximal oxygen uptake, physical work capacity, and step test scores in college women. *Medicine and Science in Sports, 4,* 182–186.

Mitchell, J.H., Sproule, B.J., & Chapman, C.B. (1958). The physiological meaning of the maximal oxygen intake test. *The Journal of Clinical Investigation, 37,* 538–547.

Myers, J., Walsh, D., Sullivan, M., & Froelicher, V. (1990). Effect of sampling on variability and plateau in oxygen uptake. *Journal of Applied Physiology, 68*(1), 404–410.

Noakes, T.D. (1997). Challenging beliefs: Ex Africa simper aliquid novi. *Medicine and Science in Sports and Exercise, 29,* 571–590.

Pohl, M.B., & McNaughton, L.R. (2003). The physiological responses to running and walking in water at different depths. *Research in Sports Medicine, 11,* 63–78.

Silvers, W.M., Rutledge, E.R., & Dolny, D.G. (2007). Peak cardiorespiratory responses during aquatic and land treadmill exercise. *Medicine and Science in Sports and Exercise, 39*(6), 969–975.

Taylor, C.I. (1944). Some properties of maximal and submaximal exercise with reference to physiological variation and the measurement of exercise tolerance. *The American Journal of Physiology, 142,* 200–212.

Taylor, H.L., Buskirk, E., & Henschel, A. (1955). Maximal oxygen intake as an objective measure of cardiorespiratory performance. *Journal of Applied Physiology, 8,* 73–80.

Wilmore, J.H., Davis, J.A., Obrien, R.S., Vodak, P.A., Walder, G.R., & Amsterdam, E.A. (1980). Physiological alterations consequent to 20-week conditioning programs of bicycling, tennis, and jogging. *Medicine and Science in Sports and Exercise, 12,* 1–8.

Wilmore, J.H., Freund, B.J., Joyner, M.J., Hetrick, G.A., Hartzell, A.A., et al. (1985). Acute response to submaximal and maximal exercise consequent to beta-adrenergic blockade: implications for the prescription of exercise. *The American Journal of Cardiology, 55,* D135–D141.