EFFECTS OF SPEECH PASSAGE LENGTH ON ACCURACY OF PREDICTING METABOLIC THRESHOLDS USING THE TALK TEST

Megan M. Schroeder, Carl Foster, John P. Porcari, and Richard P. Mikat

Department of Exercise and Sport Science, University of Wisconsin-La Crosse, La Crosse, WI, USA

Abstract:

The Talk Test (TT) is a simple technique for prescribing exercise training intensity, based on the ability to ‘speak comfortably’ after reciting a standard speech provoking stimulus. This study compares the length of the speech provoking stimulus on power output (PO) at standard TT speech comfort markers: last positive (LP), equivocal (EQ), negative (NEG), in relation to objective markers of exercise intensity, the ventilatory (VT) and respiratory compensation (RCT) thresholds. Eighteen healthy subjects performed incremental (25W/2 min) exercise with concomitant gas exchange to measure VT and RCT. They also performed (random order) incremental exercise without gas exchange while repeating standard speech provoking stimuli of 31, 62 and 93 words to allow identification of the LP, EQ and NEG stages of the TT. The mean (±SD) PO at LP (139±38, 117±39 and 103±38 W), EQ (164±38, 142±38 and 128±33 W) and NEG (196±42, 189±43, and 174±39 W) stages of the TT were analyzed for 31, 62 and 93 word passages, in relation to the PO at VT (128±43 W) and RCT (175±39 W). PO at EQ and PO at NEG stages of the TT with the 93 word speech provoking stimulus were not significantly different than the PO at VT and PO at RCT, respectively. The mean error for predicting PO at VT approximated zero with the longer (93 word) speech passage duration for EQ (0.1±37 W) and for predicting the PO at RCT for NEG (0.1±29 W). The results suggest that a longer duration speech provoking passage optimizes accuracy of the TT estimation of VT and RCT.

Key words: exercise prescription, exercise test, cardiorespiratory test

Introduction

The American College of Sports Medicine (ACSM, 2017) and American Heart Association (Haskell, et al., 2007) have published physical activity guidelines for healthy adults, designed to improve overall health and decrease the risk of chronic diseases. These guidelines suggest that individuals accumulate at least 30 min of moderate-intensity activity on at least five days each week, or achieve at least 20 min of high-intensity aerobic activity on at least three days per week (Haskell, et al., 2007). The gold standard (ACSM, 2017) recommendation for exercise intensity is 40-60% heart rate reserve (%HRR) or VO₂R for moderate intensity exercise and 60-90% HRR or VO₂R for vigorous intensity exercise. These ranges are derived from anchor points obtained during maximal exercise testing, in which maximum heart rate (HRmax) and/or maximum oxygen uptake (VO₂max) are measured. More recent recommendations have suggested that exercise prescription is optimized when intensity is prescribed in terms of metabolic thresholds (Mezanni, et al., 2012), which also requires maximal exercise testing. However, in both preventive and clinical exercise settings, maximal exercise tests are performed infrequently (due to expense, requirement for trained professionals, and acute risk).

Given the reduction in the frequency of exercise testing for exercise prescription, subjective measurements for determining exercise intensity, including the Rating of Perceived Exertion (RPE) (Borg, 1998; Eston, 2012) and the Talk Test (TT) have become of interest. Both of these methods are simple and easy to perform. While the RPE scale is widely used (Eston, 2012) and has evidence of the effectiveness of RPE regulation to guide exercise training programs (Parfitt, Evans, & Eston, 2012), less evidence is available about the Talk Test.

The TT presents a uniquely attractive option as it is directly linked to the intensities at the ventilatory (VT) and respiratory compensation (RCT) thresholds (Dehart-Beverley, et al. 2000; Foster, et al., 2008; Recalde, et al., 2002; Rodriguez-Marroyo,
et al., 2013). These “threshold” intensities have been suggested as idealized indices for controlling the intensity of training (Mezzani, et al., 2012), which eliminate inconsistent responses typical of the relative percent concept (Katch, et al., 1978; Shar Hag-Rosenberger, et al., 2010). When subjects are able to “speak comfortably”, they are typically below the intensity at the VT (Dehart-Beverley, et al., 2000; Foster, et al., 2008). Likewise, the first point where subjects are distinctly less than fully comfortable during speech is an intensity that approximates VT (Dehart-Beverley, et al., 2000; Foster, et al., 2008). The relationship between the Talk Test and the VT has been demonstrated in several populations including well-trained athletes (Jeans, et al., 2011; Recale, et al., 2002; Rodriguez-Marroyo, et al., 2013; Woltmann, et al., 2015), patients with cardiovascular disease (Brawner, et al., 2006; Nielsen, et al., 2014, Petersen, et al., 2014; Voelker, et al., 2001; Zanettini, et al., 2013; Lyon, et al., 2014), patients with paraplegia (Cowan, et al., 2012) and sedentary individuals (Foster, et al., 2009; Loose, et al., 2012; Norman, et al., 2008). Additionally, the TT has been shown to be responsive to experimental manipulation of VT by blood loss and training (Foster, et al., 2008).

Although the TT is clearly a simple and cost-effective method for prescribing exercise intensity, more research needs to be done to standardize the TT protocol. Current research has utilized many methods for provoking speech. These include “hearing your breath” (Goode, et al., 1998), reciting a standard speech passage (Pledge of Allegiance or the Rainbow Passage) (Dehart-Beverley, et al., 2000), Foster, et al., 2008), the counting method (Loose, et al., 2012; Norman, et al., 2008) and responding to pre-recorded questions (Brawner, et al., 2006). The original research in our laboratory used a 101-word paragraph from the speech pathology literature (Dehart-Beverley, et al., 2000). However, it required a cue card to be used effectively. To avoid the need for a cue card, Shafer et al. (2000) suggested the use of the “Pledge of Allegiance” (POA) as the standard speech provoking stimulus for the TT. This was potentially an ideal speech passage because it is well known to most people raised in the USA and does not require subjects to read from a cue card during testing. However, it is not suitable for people not raised in the USA, and there has been some concern that it may be too short to really capture the difficulty of speaking during exercise (e.g. it may overestimate VT and RCT). Studies using the POA as a speech provoking stimulus have led to conflicting results relative to whether the last positive (LP) (Persinger, et al., 2004) or the equivocal (EQ) (Dehart-Beverley, et al., 2000; Foster, et al., 2008, 2009) stage of the TT was equivalent to the VT, suggesting that the very short length of the POA might be introducing systematic error into the TT.

Since speaking requires suppression of breathing frequency and indeed total ventilation at an intensity where the breathing frequency and total ventilation (Dou st & Patrick., 1981; Meckel, et al., 2002; Rotstein, et al., 2004) would naturally be increasing (e.g. VT), it would be reasonable to suspect that longer speech provoking stimuli might be better at detecting difficulties with comfortable speech. Therefore, the purpose of this study was to evaluate the effect of manipulating speech passage length on the TT response. It was hypothesized that as the speech passage length increased, there would be a decrease in power output (PO) at each stage of the TT, that the EQ stage would most closely align with the VT, and that the NE G stage would most closely align with the RCT.

Methods

The subjects for this study were eighteen healthy, moderately-active, adult volunteers (men=6, women=12). Descriptive statistics are provided in Table 1. The university human subjects committee reviewed and approved the protocol. The participants provided written informed consent and completed the Physical Activity Readiness Questionnaire (PAR-Q) (Bredin et al., 2013) to determine health status prior to their participation.

The subjects completed four incremental exercise tests on an electronically-braked cycle ergometer (Lode Excalibur, Groningen, The Netherlands), starting at a power output (PO) of 25 Watts, with an increase of 25 Watts at the end of each 2-minute stage. Heart rate (HR) was monitored throughout all the tests using radiotelemetry (Polar Electro-Oy, Finland). One of these tests was a maximal exercise test during which respiratory gas exchange was measured in order to objectively determine VT, RCT and VO2max. Respiratory gas exchange was measured using open circuit spirometry with a mixing chamber based metabolic system (AEI, Pittsburgh, PA). Standard calibration procedures were used and VT and RCT were

Table 1. Descriptive characteristics of subjects (mean±SD)

<table>
<thead>
<tr>
<th></th>
<th>Men (n=6)</th>
<th>Women (n=12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>29.3±17.97</td>
<td>22.1±1.56</td>
</tr>
<tr>
<td>Body height (cm)</td>
<td>180.6±8.36</td>
<td>167.2±5.56</td>
</tr>
<tr>
<td>Body weight (kg)</td>
<td>90.3±18.83</td>
<td>65.6±6.32</td>
</tr>
<tr>
<td>BMI</td>
<td>27.7±5.80</td>
<td>23.4±1.48</td>
</tr>
<tr>
<td>PPO (W)</td>
<td>222.0±51.32</td>
<td>212.1±44.40</td>
</tr>
<tr>
<td>VO2max (ml/kg/min)</td>
<td>43.7±10.84</td>
<td>43.6±6.69</td>
</tr>
<tr>
<td>Max HR (bpm)</td>
<td>177±30.8</td>
<td>184±8.6</td>
</tr>
<tr>
<td>PO @ VT (W)</td>
<td>145.8±3.06</td>
<td>118.6±41.96</td>
</tr>
<tr>
<td>PO @ RCT (W)</td>
<td>185.5±34.77</td>
<td>170.4±41.54</td>
</tr>
</tbody>
</table>

Note. BMI=body mass index, PPO=peak power output (last 60 s average PO), PO at VT=power output at the ventilatory threshold, PO at RCT=power output at the respiratory compensation threshold.
detected using both v-slope and ventilatory equivalent methods (Foster & Cotter, 1995). $\text{VO}_2\text{max}$ was accepted as the highest 30 s $\text{VO}_2$ during the test, provided the test was subjectively maximal and that the RER >1.15. A verification protocol for $\text{VO}_2\text{max}$ was not used (Foster, et al., 2007).

The other three tests followed the same exercise protocol, but without measuring respiratory gas exchange. Instead, subjects performed the Talk Test in the last 30 seconds of each stage. The speech passage was the “Pledge of Allegiance”, which was recited one (31 words), two (62 words) or three (93 words) times. The order of speech passage length was assigned in a counterbalanced method to eliminate an order effect. The experimenter asked the subject at the conclusion of each recitation if they could “speak comfortably”. The test was terminated when the subject could no longer speak comfortably (e.g. provided a negative response – NEG). Heart rate (HR), PO, duration of speech passage, and RPE (Borg, 1998) were measured at the conclusion of each exercise stage.

**Statistical analysis**

The data were analyzed using repeated measures ANOVA to test the effect of speech passage length on TT response. Tukey’s post-hoc tests (p<.05) were used when justified by ANOVA.

**Results**

The effects of the TT using 31-, 62- and 93-word speech provoking stimuli on power output at the three Talk Test stages (LP, EQ and NEG) are presented in Table 2. The LP stage of the TT occurred at a significantly greater PO than the PO at VT in the 31-word condition. The PO at the EQ stage of the TT was significantly greater than the PO at VT in the 31-word condition, but was not different in either the 62- or 93-word TT conditions. The numerical average of PO at VT and at the EQ TT stage in the 93-word condition was within 1 W (Figure 1).

There was no significant difference between the PO at RCT and the PO at the NEG stage in the 93-word TT condition. The numerical average of PO at RCT and at the NEG TT stage in the 93-word condition was within 1 W (Figure 1).

![Figure 1. Mean (+SE) errors of the power output at the VT and RCT in relation to speech passage duration and the stages of the Talk Test. The mean error at the EQ and NEG stage of the TT with the 93 word speech provoking passage was less than 1W for the VT and RCT, respectively.](image)

<table>
<thead>
<tr>
<th>Exercise Marker</th>
<th>Power output (W)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ventilatory threshold</td>
<td>127.7±43.12</td>
</tr>
<tr>
<td>Respiratory compensation threshold</td>
<td>175.4±39.06</td>
</tr>
<tr>
<td>31-word TT</td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>138.9±37.60 a</td>
</tr>
<tr>
<td>EQ</td>
<td>163.9±37.60 a</td>
</tr>
<tr>
<td>NEG</td>
<td>195.8±42.23 c</td>
</tr>
<tr>
<td>62-word TT</td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>116.7±38.35 ns VT</td>
</tr>
<tr>
<td>EQ</td>
<td>141.7±38.35 ns VT</td>
</tr>
<tr>
<td>NEG</td>
<td>188.9±43.07 c</td>
</tr>
<tr>
<td>93-word TT</td>
<td></td>
</tr>
<tr>
<td>LP</td>
<td>102.8±33.09 b</td>
</tr>
<tr>
<td>EQ</td>
<td>127.8±33.09 ns VT</td>
</tr>
<tr>
<td>NEG</td>
<td>173.6±38.80 ns RCT</td>
</tr>
</tbody>
</table>

a TT PO > VT (p<.05)
b TT PO < VT (p<.05)
c TT PO > RCT (p<.05)
The average duration of the 31-, 62- and 93-word speech stimuli at LP, EQ and NEG was ~9 s, ~20 s and ~28 s, respectively. Thus, the longer speech provoking passages did not change the basic pattern of speech production or the average rate of speaking. The PO at any given speech provoking stimulus length decreased systematically as the length of the speech passage increased (Table 2).

The errors in estimating PO at VT and RCT from different TT speech passage length stages (LP, EQ, NEG) and varying speech stimulus lengths (31, 62 and 93 words) are presented in Figure 1. When using the LP TT stage to predict VT, the PO at VT was overestimated when using a 31-word speech passage and underestimated when using a 62- or 93-word passage (11.2±29.67, -11.0±34.17, -24.9±34.65 W). When using the EQ TT stage to predict, the PO at VT was overestimated when using a 31- or 62-word passage (36.2±29.67, 14.0±34.17 W). However, there was less than 1 W average error when using a 93-word speech passage at the EQ TT stage to predict PO at VT (0.1±36.65 W). Similarly, when using the NEG TT stage to predict the PO at RCT, the PO was overestimated when using a 31- or 62-word TT (17.9±24.83, 11.0±23.23 W). However, when using the 93-word speech passage at the NEG TT stage to predict PO at RCT there was an average error of less than 1 W (-0.1±28.95 W).

Discussion and conclusions

The main result of this study is that the EQ and NEG stages of the TT can be used to predict the PO at VT and RCT most accurately when using longer speech passage durations (~90 words). Shorter speech provoking passages tend to overestimate the PO at both the VT and RCT. Previous research has shown that when subjects exercise at intensity at which they can speak comfortably (e.g., ≤ LP stage of the TT), they are reliably below the intensity at their VT (Dehart-Beverley, et al. 2000; Foster, et al., 2008, 2009). Likewise, if a subject is definitely unable to speak comfortably, he/she is reliably at an intensity > RCT (Jeans, et al., 2011; Recalde, et al., 2002; Rodriguez-Marroyo, et al., 2013; Wol�man, et al., 2015). The EQ stage of the TT corresponds to an intensity very close to VT (Dehart-Beverley, et al., 2000; Foster, et al., 2008, 2009; Rodriguez-Marroyo, et al., 2013; Goode, et al., 1998; Mertens, et al., 2001; Loose, et al., 2012; Petersen, et al., 2014; Quinn & Coons, 2011). The results of the present study show that the estimation of the PO at VT and RCT from the EQ and NEG stages of the TT is very accurate when longer speech provoking passages are used.

The present study’s results disagree with the results of Persinger et al. (2004) from our laboratory, in which subjects recited a 31-word speech passage. A 31-word speech passage, the POA, has been the passage of choice for many TT studies because of its simplicity (Shafer, et al., 2000). Persinger et al. (2004) concluded that the LP stage of the TT corresponded to VT, and that the EQ stage of the TT was consistently above VT. We are at a loss to explain the differences from the findings of Persinger et al. (2004), as this study was from our laboratory, using the same equipment and the same basic subject population. However, Persinger’s results did suggest to us, for the first time, that the POA might be too short to properly challenge speech comfort.

A technical issue that may have affected the interpretation of the present results is the influence of stage duration. Preliminary data from Xiong et al. (2015) using a 31-word speech provoking stimulus, examined the relationship between exercise stage duration and gas exchange (GE) thresholds. They showed that longer stage durations (1, 2 and 3 min) decreased PO at thresholds (VT, RCT) and at maximal exercise, but had no effect on the magnitude of physiological markers (HR, VO2, RPE). Further research needs to be done to determine the stage duration that most accurately predicts VT and RCT using a longer speech passage. In the present results we were able to achieve near zero errors in predicting the PO at VT and RCT with 2 min stages. The data of Xiong et al. (2015) suggest that stages shorter than 2 min may lead to unrealistically high values for the PO at VT and RCT.

Although the current results demonstrate that the TT can be used to predict the PO at VT and RCT with some accuracy, one of the practical realities is that to be used for exercise prescription, such data (including directly measured VT and RCT) need to be “translated” to allow appropriate exercise intensity during exercise training. This is based on the well-known tendency for heart rate (Foster, et al., 1986) and RPE (Foster, et al., 2009) to drift during sustained exercise. This led to the suggestion from Mezzani et al. (2012) that the PO at the VT during incremental exercise needed to be reduced by ~10% to achieve tolerable workloads during exercise training. Previous studies from our laboratory have suggested that sustained exercise at about the PO of the LP stage of the TT is well tolerated in well-trained individuals (Jeans, et al., 2011; Wol�mann, et al., 2015) and that the PO at the preceding stage (LP-1) in not so well-trained individuals or patients in rehabilitation programs (Foster, et al., 2009; Lyon, et al., 2014) may be appropriate adjustments to allow comfortable speech, appropriate HR and acceptable RPE during sustained exercise training. This is conceptually similar to the ~10% decrease in PO at VT recommended by Mezzani et al. (2012). In this regard, the LP stage of the TT, with the 93-word speech provoking stimulus, marks an exercise intensity that, while below VT, may be more or less an ideal starting training intensity for many individuals, and is ~10-15% lower than the

Discussion and conclusions

The main result of this study is that the EQ and NEG stages of the TT can be used to predict the PO at VT and RCT most accurately when using longer speech passage durations (~90 words). Shorter speech provoking passages tend to overestimate the PO at both the VT and RCT. Previous research has shown that when subjects exercise at intensity at which they can speak comfortably (e.g., ≤ LP stage of the TT), they are reliably below the intensity at their VT (Dehart-Beverley, et al. 2000; Foster, et al., 2008, 2009). Likewise, if a subject is definitely unable to speak comfortably, he/she is reliably at an intensity > RCT (Jeans, et al., 2011; Recalde, et al., 2002; Rodriguez-Marroyo, et al., 2013; Wol�man, et al., 2015). The EQ stage of the TT corresponds to an intensity very close to VT (Dehart-Beverley, et al., 2000; Foster, et al., 2008, 2009; Rodriguez-Marroyo, et al., 2013; Goode, et al., 1998; Mertens, et al., 2001; Loose, et al., 2012; Petersen, et al., 2014; Quinn & Coons, 2011). The results of the present study show that the estimation of the PO at VT and RCT from the EQ and NEG stages of the TT is very accurate when longer speech provoking passages are used.

The present study’s results disagree with the results of Persinger et al. (2004) from our laboratory, in which subjects recited a 31-word speech passage. A 31-word speech passage, the POA, has been the passage of choice for many TT studies because of its simplicity (Shafer, et al., 2000). Persinger et al. (2004) concluded that the LP stage of the TT corresponded to VT, and that the EQ stage of the TT was consistently above VT. We are at a loss to explain the differences from the findings of Persinger et al. (2004), as this study was from our laboratory, using the same equipment and the same basic subject population. However, Persinger’s results did suggest to us, for the first time, that the POA might be too short to properly challenge speech comfort.

A technical issue that may have affected the interpretation of the present results is the influence of stage duration. Preliminary data from Xiong et al. (2015) using a 31-word speech provoking stimulus, examined the relationship between exercise stage duration and gas exchange (GE) thresholds. They showed that longer stage durations (1, 2 and 3 min) decreased PO at thresholds (VT, RCT) and at maximal exercise, but had no effect on the magnitude of physiological markers (HR, VO2, RPE). Further research needs to be done to determine the stage duration that most accurately predicts VT and RCT using a longer speech passage. In the present results we were able to achieve near zero errors in predicting the PO at VT and RCT with 2 min stages. The data of Xiong et al. (2015) suggest that stages shorter than 2 min may lead to unrealistically high values for the PO at VT and RCT.

Although the current results demonstrate that the TT can be used to predict the PO at VT and RCT with some accuracy, one of the practical realities is that to be used for exercise prescription, such data (including directly measured VT and RCT) need to be “translated” to allow appropriate exercise intensity during exercise training. This is based on the well-known tendency for heart rate (Foster, et al., 1986) and RPE (Foster, et al., 2009) to drift during sustained exercise. This led to the suggestion from Mezzani et al. (2012) that the PO at the VT during incremental exercise needed to be reduced by ~10% to achieve tolerable workloads during exercise training. Previous studies from our laboratory have suggested that sustained exercise at about the PO of the LP stage of the TT is well tolerated in well-trained individuals (Jeans, et al., 2011; Wol�mann, et al., 2015) and that the PO at the preceding stage (LP-1) in not so well-trained individuals or patients in rehabilitation programs (Foster, et al., 2009; Lyon, et al., 2014) may be appropriate adjustments to allow comfortable speech, appropriate HR and acceptable RPE during sustained exercise training. This is conceptually similar to the ~10% decrease in PO at VT recommended by Mezzani et al. (2012). In this regard, the LP stage of the TT, with the 93-word speech provoking stimulus, marks an exercise intensity that, while below VT, may be more or less an ideal starting training intensity for many individuals, and is ~10-15% lower than the
PO estimated for VT from the Talk Test. Likewise, in more athletic individuals wishing to undertake systematic interval training, even though the RCT is better marked by the NEG stage of the TT, the EQ stage of the TT (with the 93-word speech provoking passage) may represent a PO that is sustainable during intervals at about the intensity of the RCT (Woltmann, et al., 2015).

In conclusion, the present study has compared TT responses to speech passages of varying lengths as surrogates of direct measurement of VT and RCT. Speech passages that are relatively longer in duration (≥ 93 words) most accurately predicted VT and RCT using the EQ and NEG stages of the TT. When subjects are first unsure of their ability to speak comfortably (EQ stage), they are very close to the PO at VT. When subjects first definitely cannot speak comfortably (NEG), they are very close to the PO at RCT. These data suggest the practical value of a very simple strategy for evaluating exercise capacity at threshold markers, which may then be easily translated into the appropriate intensity for either sustained submaximal training or for the hard segments of interval training.

References


