The Gender-Neutral Timed Obstacle Course: a valid test of police fitness?

C. A. Jackson¹ and D. Wilson²

¹Occupational Health Psychology, Division of Psychology, School of Social Sciences, Birmingham City University, Birmingham B42 2SU, UK, ²Criminology, Centre for Applied Criminology, Birmingham City University, Birmingham B42 2SU, UK.

Correspondence to: C. A. Jackson, Occupational Health Psychology, Division of Psychology, School of Social Sciences, Birmingham City University, Perry Barr, Birmingham B42 2SU, UK. Tel: +44 (0)121 3315000 E-mail: Craig.Jackson@bcu.ac.uk

Background

In the light of the Winsor review, UK police forces have been urged to use fitness tests as large-scale cost-effective measures of officers’ fitness to work. One test is the Gender-Neutral Timed Obstacle Course (GeNTOC), which must be completed within 3 min 45 s, regardless of sex.

Aims

To investigate if obstacle courses, mimicking a range of police-type activities, can provide a suitable and valid basis for identifying fit and unfit officers and if any other factors may influence test performances.

Methods

Five years of GeNTOC records were randomly sampled, providing data for 1701 officers. Pass/fail rates were analysed alongside demographics and obstacle performance.

Results

Of 1701 candidates, 24% (397) failed GeNTOC (7 and 42% of males and females, respectively). Females failed in two specific obstacles significantly more often than males: the ‘body drag’ and ‘gate weave’. Errors made on these obstacles alone accounted for 49% of obstacle errors made by females. GeNTOC success was significantly associated with candidates who were male, younger, taller, heavier and of lower body mass index (BMI). Of all candidates, 42% were overweight, and 8% were obese.

Conclusions

The GeNTOC was not a useful screening tool and worked independently of BMI groups. Too few candidates were appropriately screened out; too many of those failing were female; and too many who passed were overweight or obese. GeNTOC was unfair to female candidates and favoured overweight or obese males. Recommendations are made to adopt alternative fitness tests or to modify the GeNTOC obstacles, testing procedures and data collection.

Key words

Fitness standards; fitness tests; police officers; selection.

Introduction

Given the range of tasks undertaken by police officers and an increasing variety of hazards encountered in police work [1], good physical health and fitness are highly important. Most industrialized nations have populations who are living and working longer and who are becoming more sedentary and fatter [2,3]. Defining ‘obese’ and ‘overweight’ by body mass index (BMI; kg/m²) is not a perfect guide to obesity status [4], but despite some vagaries of BMI and its relationship to ‘fitness’, it retains a high level of usefulness and good predictive ability for ‘fitness to do the job’ with many active, emergency and uniformed workforces [5–9]. In March 2012, the final part of a UK-independent review of police officer and staff remuneration and conditions [10] made radical suggestions for the 43 police forces in England and Wales. The review acknowledged that policing is more complex and specialized than ever before and that staff development should be modified accordingly. The Metropolitan (London) Police administered a health check to 11 548 officers and staff in 2011 (Health Diagnostics Ltd, unpublished results), and the Winsor report [10] made much of obesity among 6900 male and 4600 female London Metropolitan Police staff (61% uniformed officers; 32% police staff; 7% police and community support officers). Among males, 52% were overweight, 22% were obese and 1% were morbidly obese. For females, figures were 32, 16 and 2%, respectively. Thirty-five per cent of staff (50% of females
and 25% of males) had a normal BMI. Associations exist between high BMI, sickness absence and poor performance, and there are concerns for the workability of emergency services staff in the context of rising obesity [11]. In addition, the Winsor review [10] recommended increasing the retirement age for police officers to 60, suggesting a need for continual health promotion.

Many concerns around obesity and fitness to work are directed towards future generations, who already appear to have levels of obesity likely to take them past current obesity levels by the time they reach adulthood [12]. Male Dutch police officers have obesity levels higher than the national average [13] and these increased with age. Similarly, Royal Canadian Mounted Police officers were rated as less fit than criminals in a study 10 years earlier [14]. Other emergency services are affected by obesity and with US federal laws preventing pre-employment screening of the fitness of firefighters, findings such as an aerobic capacity below that deemed appropriate for work in a sample of firefighters [7] are not entirely surprising. Another study looking at almost 7000 Dutch police [15] found young males completed fitness tests quicker than females and older candidates, concluding that older officers and females were physically ‘less tailored’ to their police jobs than younger male colleagues. Such fitness levels provide good reason for appropriate pre-employment physical testing in safety critical occupations, providing the tests are appropriate, worthwhile and accurate. However, in the UK where policing requires much face-to-face community interaction and ‘consensual policing’, fitness requirements are different from those in other countries with different policing styles [16].

One attempt to develop a bespoke fitness test for UK constabularies was made by the then Royal Ulster Constabulary (RUC) as a ‘job-related and completely gender-free’ test. The need for gender-free assessments is illustrated by the 1997 case of Alcock versus Chief Constable of Hampshire Constabulary. Alcock, a male candidate for a dog-handler job, completed a fitness test within the female cut-off point (17 min), but outside the male cut-off point (16 min), and consequently failed the test. This was considered to be direct sexual discrimination in contravention of the Sexual Discrimination Act (1975 (amended 1986)), despite the defence that males have greater aerobic and anaerobic ability potential than females [17], and that such differences required compensation. The RUC developed a Gender-Neutral Timed Obstacle Course (GeNTOC) after identifying a range of work activities performed by sampling 500 officers. It agreed on a final gender-neutral cut-off time of 3 min 45 s for successful completion of three laps of the test. An industrial tribunal (Dougan versus Chief Constable of the RUC (1997, 1998)) considered that this cut-off time was entirely arbitrary and, therefore, represented indirect sexual discrimination. This judgement was made in the case of a female job applicant who twice failed the GeNTOC by completing the test outside this cut-off time but within that formerly applied to females of 3 min 54 s.

The complexities of discrimination legislation means there will be few certainties for UK constabularies using fitness tests to select people for employment. If tests are used at a pre-employment screening stage or to stream for specialist positions, tests with identical cut-off times for both sexes can have an adverse impact on females, and this could be viewed as indirect sex discrimination. The Winsor report [10] contained numerous recommendations concerning wide-ranging aspects of police officers’ working lives, including the requirement for them to undergo annual fitness testing, suggesting that this should include rigorous testing of officers’ physical abilities to crawl, climb stairs and to carry traffic cones, such as those included in the GeNTOC. Further, it was recommended that any officers who failed such tests should face loss of earnings and subsequent dismissal after three consecutive fails. This study attempted to investigate whether obstacle courses that mimic a range of police activities, such as the GeNTOC, can provide a suitable and valid basis for distinguishing fit officers from unfit ones and whether any other biases may play a part in test performances.

**Methods**

We obtained data concerning performances of police officers and new recruits from a single English constabulary using the GeNTOC. Data were obtained in hard copy by randomly selecting a number of box-files from storage using random number generation. The analysis was based on a random selection of data from individuals who undertook the GeNTOC over 113 different testing occasions from 2004 to 2009. Data were anonymous and contained no personal details, medical or psychological information (other than biometric measurements to allow BMI calculation), and, consequently, the constabulary involved determined that the research did not need to be evaluated by a local research ethics committee. BMI classifications were calculated differently for Asian participants in order to reflect physiological differences [18]. All data were double entered into the Statistics Package for Social Sciences and checked for integrity. Checks were made by comparing the two databases and any detected anomalies were corrected.

The circuit run component of the GeNTOC consisted of three identical laps of an obstacle course involving eight different obstacles, encountered in the following order: ‘Crawl’, involving crawling on hands and knees for several yards underneath a low grid; ‘Jump’, involving making a forwards jump from a stationary standing position and clearing a length of ~1 m; ‘Stairs’, involving climbing a set of steps, running along a flat platform, then down another set of steps; ‘Gate’, involving climbing down a small flight of stairs while ducking under horizontal bars set at head
height; ‘Balance beam’, involving walking along a three metre balance beam ~6 inches above the ground, and then walking along a second identical beam set at 90° to the first beam; ‘Weave’, involving passing through a series of slalom gates set ~3 feet apart without touching the gate sides; ‘Cone carry’, involving picking up a weighted traffic cone in each hand and carrying them in a loop over a distance of ~16 feet and returning them to their original positions and ‘Body drag’, involving dragging in a circular pattern, a life-size dummy on the floor and placing it back in its original position. If errors were made on any obstacle, the candidate had to repeat it until it was completed without error. No additional time penalties were added to completion times as a result of any errors.

**Results**

A total of 1701 candidates were included in the analysis, of which 54% (913) were male, 88% were white British, 6% were Indian and 3% were Pakistani. The remaining 3% were from 11 other ethnic groups with no single group accounting for >1% of the entire sample. The age distribution of candidates was biased towards younger ages and was skewed to the right (kurtosis = 0.74).

Table 1 shows descriptive details of the sample with 51% being overweight or obese. Height, weight and BMI were all normally distributed.

Of 1701 candidates, 98% completed three laps of the GeNTOC circuit, while 2% (25) failed to do so. As GeNTOC had to be completed within 3 min 45 s, completing candidates were divided into groups: ‘in-time’ completion (passes) and ‘outside-time’ completion (fails). Of 1676 candidates completing three GeNTOC laps, 22% (372) failed to complete within time (fails) and 78% passed, giving an overall failure rate (non-completers and fails) of 24% (n = 397). The mean total lap completion time for the 1676 candidates was 3 min 33 s (±21 s). For those who passed, the mean completion time was 3 min 24 s (±13 s), and for those who completed outside the time limit, it was 4 min 03 s (±17 s). Final lap times correlated weakly with age (r = 0.12); weight (r = −0.15); height (r = −0.40) and BMI (r = 0.12). Significant differences in BMI distribution between passes and fails are shown in Table 2. Of 1304 successfully passing the GeNTOC, 65% (845) were male. Of 374 who completed GeNTOC outside the time limit, 83% (310) were female.

The sample represented 5053 completed laps of the course, in which 40 504 individual obstacle attempts were made. Of those, 1811 obstacle errors were recorded, giving an error rate of 4.5%. Some obstacles were more difficult for candidates to complete as shown in Table 1. Females made significantly more errors than males on the obstacles of ‘body drag’ (P < 0.01) and ‘weave’ (P < 0.001).

The results of stepwise multiple regression to investigate the pattern of variables for predicting completion rates (pass versus fail) for those attempting the GeNTOC, including ‘sex’; ‘errors made’; ‘BMI’ and ‘age’, are shown in Table 3. To explore whether a modified gender-neutral time cut-off point should be established, we examined the effect of extending the GeNTOC cut-off time to 1.96 SD above the mean completion time (03:24 ±13 s) for the participants who passed. Using normal distribution theory, 1.96 SD (1.96 ×13 s) would add 25 s to the mean completion time, resulting in a guideline cut-off time of 03:49, thereby extending the current cut-off by 5 s. Using this modified cut-off would alter the total number of passes from 78% (1304) to 81% (1365) and fails from 22% (372) to 19% (311). This modified cut-off would...
increase the percentage of males passing by 3% (from 93 to 96%) and the percentage of females passing by 5% (from 60 to 65%). Increasing the cut-off time alone would not, therefore, make the GeNTOC gender neutral.

**Discussion**

The main findings of this study were firstly that in the GeNTOC ‘gender-neutral’ obstacle course, females performed significantly worse than males. Secondly, success in performance was most closely related to the number of obstacle errors made with those participants who passed making significantly fewer errors than those who failed. Thirdly, females made significantly more errors than males on the ‘body drag’ and ‘weave’ obstacles and such obstacles combined accounted for almost half of all errors made. Height, age, weight and BMI were not significantly independently associated with satisfactory GeNTOC performance in the way that might be expected. A regression model of errors made, age, sex and BMI combined could only account for ~30% of variability in completion rates. As a screening tool assessing the fitness of police officers, GeNTOC appears to possess content validity in that it seems to be a measure of fitness conducted in realistic settings emulating a range of police activities. However, its simplicity may mask inherent unfairness as a selection tool.

GeNTOC could be considered unfair to females who are significantly disadvantaged (relative to males) on obstacles which account for success/failure. The ‘body drag’ obstacle may prove difficult because of physiological and anatomical differences between males and females in relation to pulling and dragging dead weights, and also because of the placing of this obstacle at the end of the lap, when fatigue may also be a factor. Should police forces wish to retain the current GeNTOC, they should, in our opinion, modify the test procedures and obstacles used, rather than increase the gender-neutral cut-off time. We recommend that (i) the ‘body drag’ obstacle becomes the first obstacle in each lap to counter the effects of end-of-lap fatigue; (ii) the ‘weave’ obstacle uses wider gates and (iii) more data be collected from candidates concerning health, lifestyle, physiology, demographics and suitability/consistency of clothing.

Over 76% of candidates passed the GeNTOC successfully, and of those, the majority were males. However, over 42% were overweight, and a further 7% were obese. The low number of candidates being screened out by this tool and the high percentage of those passing being overweight or obese poses concerns as the usefulness of a screening tool must be questioned if it screens out ~25% of candidates, with a disproportionately large number of those failing being female, while a significant proportion of those who pass are overweight. Although BMI can be unreliable in the categorization of obesity, especially in individuals and occupations with extreme or high levels of muscular density, other studies have found increased BMI to correlate well with poorer physical test performance times in many occupations, including firefighters, the military and police officers [5–8,14], as well as being a useful measure in screening candidates for safety critical jobs. Such studies have also found BMI to reveal

<table>
<thead>
<tr>
<th>Variable, mean (SD); n (%)</th>
<th>Fail n = 372</th>
<th>Pass n = 1304</th>
<th>$X^2/F$</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>68 (18)</td>
<td>845 (65)</td>
<td>279</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>28 ± 7</td>
<td>27 ± 6</td>
<td>9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Height (cm)</td>
<td>167.7 ± 8.1</td>
<td>174.0 ± 9.0</td>
<td>144</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Weight (kg)</td>
<td>72.5 ± 13.2</td>
<td>75.8 ± 13.2</td>
<td>18</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI</td>
<td>25.6 ± 3.5</td>
<td>24.9 ± 3.2</td>
<td>15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lap 1 time</td>
<td>01:19 ± 00:07</td>
<td>01:11 ± 00:08</td>
<td>&lt;1</td>
<td>NS</td>
</tr>
<tr>
<td>Lap 2 time</td>
<td>02:46 ± 00:28</td>
<td>02:18 ± 00:21</td>
<td>46</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lap 3 time</td>
<td>04:03 ± 00:17</td>
<td>03:24 ± 00:13</td>
<td>1916</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Underweight</td>
<td>3 (1)</td>
<td>12 (1)</td>
<td>&lt;1</td>
<td>NS</td>
</tr>
<tr>
<td>Normal weight</td>
<td>162 (44)</td>
<td>652 (50)</td>
<td>6</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>Overweight</td>
<td>159 (42)</td>
<td>555 (43)</td>
<td>&lt;1</td>
<td>NS</td>
</tr>
<tr>
<td>Obese</td>
<td>48 (13)</td>
<td>85 (6)</td>
<td>20</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

NS, not significant; SD, standard deviation.

*Yates’ continuity for correction t-test.*

<table>
<thead>
<tr>
<th>Predictor variables r</th>
<th>$r^2$</th>
<th>b</th>
<th>$\beta$</th>
<th>T</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.40</td>
<td>−0.33</td>
<td>−0.39</td>
<td>19.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Errors made</td>
<td>0.53</td>
<td>−0.11</td>
<td>−0.35</td>
<td>17.63</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI</td>
<td>0.56</td>
<td>−0.01</td>
<td>−0.13</td>
<td>6.43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>0.57</td>
<td>−0.00</td>
<td>−0.10</td>
<td>5.16</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

$r^2 = 0.32$; adjusted $r^2 = 0.31$; $r = 0.56$. 

---

**Table 2.** Comparisons between GeNTOC fails and passes (completers only), $n = 1676$

<table>
<thead>
<tr>
<th>Predictor variables</th>
<th>r</th>
<th>$r^2$</th>
<th>b</th>
<th>$\beta$</th>
<th>T</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>0.40</td>
<td>0.16</td>
<td>−0.33</td>
<td>−0.39</td>
<td>19.49</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Errors made</td>
<td>0.53</td>
<td>0.29</td>
<td>−0.11</td>
<td>−0.35</td>
<td>17.63</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>BMI</td>
<td>0.56</td>
<td>0.31</td>
<td>−0.01</td>
<td>−0.13</td>
<td>6.43</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Age</td>
<td>0.57</td>
<td>0.32</td>
<td>−0.00</td>
<td>−0.10</td>
<td>5.16</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

$r^2 = 0.32$; adjusted $r^2 = 0.31$; $r = 0.56$. 

---

**Table 3.** Multiple regression of predictors of pass versus fail classification on GeNTOC
high levels of overweight or obese staff in safety critical emergency services [6,9]. In the median year of the data sample (2008) in this study, the geographical region in which it was based had higher than UK national average rates of overweight males and females (42 and 33%, respectively) [19], while the number of obese candidates in the sample was just below the regional average for the same period (24 and 25%, respectively).

By using a random sample of real-world data, this study should possess sound content validity. However, it would have been preferable to conduct a longitudinal follow-up study of the health trajectory of candidates in their police careers. The quality of the data was sufficient to allow such linkage at a future date. While the motives of constabularies operating physical fitness tests are doubtless commendable [20], current standards in operation are not necessarily helpful. If a screening test allows substantial numbers of overweight and unfit subjects to remain within a police force, there exists a risk of officers being made ill by their work or of current health problems being exacerbated further. For example, officers repeatedly carrying greater physical mass (in and out of patrol cars) or working in unnatural postures for prolonged periods could incur greater musculoskeletal risks, as well as risks posed by overweight workers not adapting to shift working [21,22]. The ability of police officers to do their job adequately is a public safety concern, with additional concerns for the safety of officers in respect of personal protective equipment, shields and public order equipment, as well as concerns that larger officers may place themselves at risk by being bigger and slower moving ‘targets’. There remains a challenge for emergency services to provide healthy workforces [23].

In conclusion, we support the recommendation of the Winsor report [10] to overhaul police fitness, but only if appropriately validated measurements of fitness are utilized for this purpose.

Key points

- One in four candidates (24%) failed a police force recruitment obstacle test (the Gender-Neutral Timed Obstacle Course) and of those who passed it, 49% were overweight or obese.
- Test success was closely attributed to obstacle performance, and females were significantly more error-prone than males on two distinct obstacles, probably hindered by female physiology.
- The Gender-Neutral Timed Obstacle Course discriminated against females when competing to an identical time cut-off point for both sexes and as such we consider Gender-Neutral Timed Obstacle Course is not a useful screening tool for identifying unfit police officers or for highlighting ‘fitness’ among police officers.

Acknowledgements

This work was undertaken by the authors on a pro bono publico basis with a view to assisting all UK police forces with their employment practices. No fee was paid by any individual police force, or accepted by the authors or by their employer, Birmingham City University.

Conflicts of interest

None declared.

References


---

**Wanted: Questionnaires and Questionnaire Reviewers**

*Occupational Medicine*, the journal of the Society of Occupational Medicine, is running a series of articles covering questionnaires used in OH clinical practice. If you use a particular questionnaire in your practice and would be willing to review it and submit it for consideration for publication please contact Angela Burnett at om@som.org.uk to check we haven’t already got a review of that questionnaire underway and for guidance on the review content we are looking for.