Measuring physical fitness in persons with severe or profound intellectual and multiple disabilities
Waninge, Aly

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2011

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 5
Psychometric quality of a graded treadmill exercise test for people with severe or profound intellectual and visual disabilities.

C.H. Sickler
A. Waninge
T. Takken
C.P. van der Schans
R. van Wijck

Submitted
Abstract

Introduction Exercise tests using treadmills are valuable tools for assessing exercise capacity. However, a treadmill protocol for persons with severe or profound intellectual and visual disabilities (severe or profound intellectual and multiple disabilities, SIMD) is not yet available.

Objective The present study investigated primarily the feasibility, validity and test-retest reliability of a graded treadmill exercise test (GXT) for people with SIMD and GMFCS level I.

Method Thirty participants with SIMD and GMFCS level I performed a graded exercise test and retest. A supra maximal block test (SMBT) was administered to assess validity. Two physical therapists alternated as test leaders, assisted by one of seven specially trained physical education instructors.

Results The participants’ mean (sd) age was 41 years (11 yrs). Feasibility was sufficient for the test and retest of the GXT (86.6%). For the SMBT feasibility was less than sufficient (76.9%). Correlation of the peak heart rate ($HR_{peak}$) between the 1st GXT (GXT1) and 2nd GXT (GXT2) was good and significant (ICC=0.95; 95%CI 0.88-0.98) with good agreement (t-statistic p=0.5). Limits of agreement (LOA) were -16 to 14, which amounts to 19.5% of the mean $HR_{peak}$. The number of attained levels of the GXT1 and GXT2 showed a highly significant correlation (ICC=0.95; 95%CI 0.90-0.98). Correlation between $HR_{peak}$ GXT and $HR_{peak}$ SMBT was good (ICC=0.94; 95%CI 0.86-0.98) with good agreement (t-statistic p=0.7). LOA were from -17 to 15, being 20.5% of the mean $HR_{peak}$. Validity of GXT was good. As a secondary result, correlations and agreements between directly measured $HR_{peak}$ and estimated $HR_{peak}$ (using the Fernhall equation) were poor.

Conclusion A GXT performed on a treadmill is a feasible, reliable and valid means of determining $HR_{peak}$ as well as number of attained levels for people with SIMD and GMFCS level I. At the individual level, results may have fairly large variability. The Fernhall equation for estimating $HR_{peak}$ for people with SIMD systematically overestimated $HR_{peak}$. 
Introduction

Intellectual disability (ID) is characterized by significant limitations in both intellectual functioning and adaptive behavior as expressed in conceptual, social, and practical skills [1]. The disability originates before the age of 18 [1]. Intellectual disabilities are categorized in four groups: mild, moderate, severe and profound. Persons with severe or profound ID have a prevalence of visual impairments of 92% [2].

People with intellectual disabilities tend to have low activity and fitness levels which decline over the years when compared with those without a disability [3, 4]. Like individuals with intellectual disabilities, persons with visual impairments also display poor performance in locomotor skills [5] and have low levels of habitual activity [6]. Individuals who suffer from a combination of severe or profound intellectual and visual disabilities (severe or profound multiple disabilities, SIMD) are particularly at risk in terms of the potential development of deficits in both locomotor skills and daily functioning [7]. Furthermore, the combination of these deficits suggests that persons having SIMD are likely to have lower levels of habitual activity, than persons with ID without visual impairment.

Bouchard et al. [8] describes the relationship between physical activity, health-related fitness and health. Good physical fitness improves wellbeing and quality of life [9, 10], and decreases health risks, such as overweight and obesity [11]. Bouchard et al. described cardio-respiratory fitness as an important component of health-related fitness [8]. However, a considerable number of persons with SIMD achieved poor results compared to other persons with specific health conditions on health-related fitness as measured by the six minutes walking test [12]. These findings hence underscore the importance of gaining comprehensive insight into the health-related fitness of persons with SIMD, including the level of cardiorespiratory fitness.

Waninge et al. [12] performed a feasibility and reliability study for an adapted Shuttle Run Test (aSRT) in adults with SIMD. Feasibility and test-retest reliability of this aSRT over ground were good for participants classified on the Gross Motor Functional Classification Scale (GMFCS) as level I, yet not sufficient for those with GMFCS level II [12]. The peak heart rate (HR\text{peak}) for each participant was estimated according to the Fernhall equation [13]. However, a significant difference (p<0.001) was found between the mean (±sd) HR\text{peak} at the end of the aSRT (126 beats/min ± 20) over ground and the mean (±sd) estimated HR\text{peak} (172 beats/min ± 6) [12]. This finding suggests that the aSRT overground has a limited validity as so far as the objective of the test is to measure the HR\text{peak}. However, an alternative explanation of the significant difference between the achieved and the estimated HR\text{peak} may point to the formula designed by Fernhall et al [13], which could simply overestimate the exercise capacity of subjects with SIMD.

Treadmill tests are known to obtain valid data for cardiorespiratory fitness [14, 15], and thus may help to solve this question. Testing a group of participants with SIMD using a graded treadmill test will give us an answer as to whether their peak heart rates are lower than expected, thereby rendering the Fernhall equation [13] invalid, or whether the aSRT test [12] is not able to yield a peak heart rate for this target group.

Another advantage of using a treadmill is that this apparatus might counteract the well-known tendency of persons with SIMD to display a lower motivation for physical performance than those without ID [3, 16]. Environmental cues as well as the gradually increasing speed of the treadmill reduce the degree of freedom as described in theories about perception-action coupling and the ‘constraint-led approach’ [17], thereby facilitating physical performance by our SIMD participants.
In order to confirm the validity of the aSRT [12], results of both the treadmill and aSR test should be compared to a golden standard [18]. Midgley et al. [19] reviewed available literature regarding the verification phase following a graded exercise test and suggested performing a Supra Maximal Block Test (SMBT) to set a gold standard for HR\textsubscript{\text{peak}} on an individual basis [19]. If the heart rate of the SMBT differs by no more than two beats from the HR\textsubscript{\text{peak}} reached during the corresponding aSRT [12], the exertion during the aSRT [12] is scored as maximal [19].

The purpose of this study thus is twofold; it seeks to determine the feasibility, validity and test-retest reliability of a graded treadmill test (GXT) for participants with SIMD and GMFCS level I, using a Supra Maximal Block Test (SMBT) as golden standard. By doing so, this study will also assess whether the formula of Fernhall overestimates HR\textsubscript{\text{peak}} in persons with SIMD.

**Methods**

**Participants**

The participants were recruited from a residential care facility in the Netherlands, which is home to 200 persons with severe or profound intellectual and visual disabilities. Only persons functioning at GMFCS-I level were included, because Waninge et al. have indicated the aSRT performed over ground is only reliable for persons with GMFCS level I [12]. Probably due to visual impairments, the participants were not able to run and jump spontaneously [5, 7, 12]. Subsequently, classification was adjusted so as to include people with GMFCS-I who were able to increase walking speed [12].

A total of 30 participants were recruited (17 males, 13 females), all classified as GMFCS-I. Twenty-three participants had a severe ID and seven a profound ID. Some could not walk alone because of a visual impairment. Table 1 presents the participants’ characteristics.

**Table 1. Population**

<table>
<thead>
<tr>
<th></th>
<th>Male (n=17)</th>
<th>Female (n=13)</th>
<th>Total (n=30)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>40 (11)</td>
<td>42 (11)</td>
<td>41 (11)</td>
</tr>
<tr>
<td><strong>ID</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Severe ID (20\leq IQ&lt;35)</td>
<td>12</td>
<td>12</td>
<td>23</td>
</tr>
<tr>
<td>Profound ID (IQ&lt;20)</td>
<td>5</td>
<td>2</td>
<td>7</td>
</tr>
<tr>
<td><strong>Down Syndrome</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td><strong>Vision</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blind</td>
<td>11</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td>Impaired vision</td>
<td>6</td>
<td>5</td>
<td>11</td>
</tr>
<tr>
<td><strong>GMFCS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>17</td>
<td>13</td>
<td>30</td>
</tr>
</tbody>
</table>

n: sample size; SD: Standard Deviation; ID: Intellectual Disability; GMFCS: Gross Motor Function Classification System; IQ: Intelligence Quotient
Exclusion criteria were mental or physical health problems that prevented the client from participating. A physician, specialised in the care for people with ID, approved participation. Exclusion criteria at the moment of testing were illness or fever, recent epileptic insult, stress, wound, concussion or pain during movement.

**Design**
Participants were tested twice, the first graded treadmill test (GXT1) was followed by a second graded treadmill test (GXT2) at least one week later. Both tests lasted 15-25 minutes. The same instructor and the same test leader performed the testing procedure at the same time of day. The test leader scored \( \text{HR}_{\text{peak}} \), attained level and test time. A two-minute supra-maximal block test (SMBT) followed at random either GXT1 or GXT2 (convenience sampling). A Trimline T370HR treadmill (Tunturi, Almere, the Netherlands) was used for testing. A Polar RS 800 heart rate monitor (Polar Nederland, Almere, the Netherlands) measured heart rate during the test.

**Ethical statement**
The study was performed in agreement with the guidelines of the Helsinki Declaration as revised in 1975. Permission to carry out the study was obtained from an institutional ethics committee. All participants were unable to give consent. Therefore extra care and attention was given to:
1) Obtaining informed consent. Informed consent was obtained from legal representatives and caregivers of all participants;
2) The construction of the study group by formulating exclusion criteria and contraindications: We screened the participants based on the examination findings of a physician specialised in intellectual disabilities and also of a behaviour scholar;
3) The measurement procedure: The measurements were performed in accordance with the behavioral code section entitled 'Resistance among people with an intellectual disability in the framework of the Act Governing Medical-Scientific Research Involving Humans' [20]. Consistent distress or unhappiness was interpreted as a sign of lack of assent and further participation in the study was reconsidered.

**Protocols**
**Graded treadmill test (GXT)**
The adjusted SRT protocol of Waninge et al. was used and performed on a treadmill [12]. All participants practiced walking the treadmill at least twice before being tested. The specially trained physical education instructor (instructor) accompanied the participant to the treadmill and attached the safety cord. The instructor explained the procedure and attached the belt of the heart rate monitor to the participant. To ensure safety the instructor positioned himself behind the participant, with a foot on each side of the treadmill. The test leader checked the heart rate monitor and stayed one meter on the side of the participant during the whole test procedure. Rintala et al. reviewed the familiarization process in cardio-respiratory fitness testing in persons with mild to moderate intellectual disabilities, recommending a familiarization protocol [21]. To reduce stress due to unfamiliar situations, the participants walked the treadmill at least twice at regular walking speed before testing. Participants did not reach their \( \text{HR}_{\text{peak}} \) levels during these practice sessions.
The treadmill stood in a gymnasium where other clients were doing workouts. The test leader calibrated the treadmill once a week to guarantee reliability. The starting speed was 3 km/hour for all participants. Each minute the speed increased with alternating 0.3 and 0.2 km/hour increments to reach the next level. Levels were comparable with the levels in the Waninge study [12], in which 0.5 km/h increase of speed appeared not feasible for persons with SIMD.

The test continued until volitional exhaustion, similarly to the study of Rintala et al [16]. The test was finished in three possible ways; when the participant insisted he wanted to stop or refused to continue, when the predicted HR_{peak} [13] was reached, or when the safety pin was pulled. The level maintained during the last full minute stage counted as the highest attained level. After either GXT1 or GXT2, the participant rested for five minutes and went on to perform the SMBT.

Seven instructors participated in data collection. Two physical therapists acted as test leaders. They checked exclusion criteria at the moment of testing on a form filled out by the participant’s personal care professional.

Supra Maximal Block Test (SMBT)
The individual gold standard for HR_{peak} was set by SMBT [19]. Midgley et al [19] reported the utility of the verification phase, which was performed at a speed equivalent to one stage higher than that attained during the last completed stage of the incremental phase. They did not exactly define the incremental phases and corresponding speed, as they found that ‘despite the incremental phases being distinctly different, the mean maximal VO2max values attained in the appended verification phases were almost identical’ [19]. Therefore, in the present study, speed during the verification phase was 0.2 or 0.3 km/h higher than speed during the last completed stage, as in persons with SIMD only an increase of speed of 0.2 or 0.3 km/h is feasible [12]. This is considered an acceptable increase of speed, because Midgley et al [19] stated that ‘the verification phase should incorporate a workload higher than that attained in the incremental phase to conform to the original concept of VO2max’. If the heart rate of the SMBT differed by not more than two beats from the HR_{peak} reached during the corresponding GXT, the performance during the GXT was scored as maximal [19]. During the two minutes of the verification phase, the attained HR_{peak} was registered in beats per minute.

Two alternative protocols were developed. Criterion A was the SMBT [19]. The participant walked for another two minutes at a level one step (0.2 or 0.3 km/hr) higher than the level attained during the GXT. If the instructor thought, for instance for behavioural reasons, an increase in level was not feasible, the participant performed criterion B, which was a two minutes maximal block test at the highest previously attained level. The HR_{peak} was registered in beats per minute. Validity calculation did not include results from criterion B.

Motivation
Since people with ID tend to have lower motivation for physical activity, motivation was considered as a factor influencing the validity of the test [3, 16]. Paired modelling and positive reinforcement have a positive effect on compliance to treadmill walking for people with moderate to severe ID [22]. Participants were encouraged to continue walking using these techniques. Both the instructor and the test leader observed independently the amount of encouragement given as well as the compliance with the task. The observed motivation was defined as to how well the
participant had fulfilled the task. This was drawn out using a 100 mm Visual Analogue Scale (VAS) [23, 24]. Zero corresponded to no motivation and 100 to the best possible motivation.

Estimated peak heart rate
Several researchers have found that the \( HR_{\text{peak}} \) for people with ID systematically differs from the \( HR_{\text{peak}} \) of people without ID [25,26]. Therefore, Fernhall et al. developed an equation to more accurately predict \( HR_{\text{peak}} \) for people with ID [13]. The estimated \( HR_{\text{peak}} \) using the Fernhall formula [13] was calculated for each participant before testing.

Data Analysis
All statistical analyses were performed using the Statistical Package for Social Studies (SPSS) version 16.0 for Windows. All data was checked for normal distribution and homoscedasticity by statistical analysis. Homoscedasticity was defined as no relation between the error and the size of the measured value [27]. A p-value of < 0.05 indicated statistical significance for all tests.

Feasibility
The GXT and the SMBT were tested for feasibility. The percentage of the participants that finished the test successfully determined feasibility. Interpretation of the feasibility scores was taken from the Groningen fitness test for the elderly [28]. A 95% score meant “good” feasibility and an 80% score meant “sufficient” feasibility [28].

Test-retest reliability
The Intraclass Correlation Coefficient (ICC) (two-way random, absolute agreement) was calculated for \( HR_{\text{peak}} \) and for the attained GXT1 and GXT2 levels. An ICC<0.75 indicated poor or moderate reliability, 0.75≤ICC<0.90 good reliability and ICC0.90 very good reliability [29].

Agreement was analysed with a paired samples t-test for \( HR_{\text{peak}} \) and attained levels. In a Bland-Altman plot the Limits of Agreement (LOA) was determined [30]. The LOA expressed as a percentage of the mean described the variability of the results. The LOA represents individual variability. No criteria are available for ‘good’, ‘sufficient’ or ‘poor’ LOA. In literature researchers themselves are encouraged to judge whether the LOA is narrow enough for the test to be of practical use [27].

To assess test-retest reliability the Standard Error of Measurement (SEM), the Smallest Detectable Difference (SDD) and Effect Size (ES) were calculated for \( HR_{\text{peak}} \) and the attained GXT and GXT2 levels.

The Standard Error of Measurement (SEM) represents the standard deviation of measurement error. The SEM reflects the reliability of the response [29] and indicates the maximum difference that could be based on measurement error. SEM was calculated using the formula: SEM=SD x \( \sqrt{(1-\text{ICC})} \) [27].

The Smallest Detectable Difference (SDD) reflects the diversity of the participants. It is a measure of agreement and is based on measurement error. The SDD represents the smallest true difference. The SDD was calculated as SDD = 1.96 x \( \sqrt{2 x \text{SEM}} \) [27, 31].

Clinical relevance was estimated calculating the Effect Size (ES). The ES gives an objective and standardised measure for observed differences [32]. The formula for calculating ES: \( r = \sqrt{t^2/ (t^2+df)} \) [32], t being the t-statistic and df the degrees of freedom. The cut-off point for a small ES is 0.1, for a medium ES 0.3 and for a large ES 0.5 [32].
Validity
Validity calculation only includes results of the SMBT, performed according to criterion A (see Protocols). To evaluate validity we compared the HR_{peak} of the SMBT with the HR_{peak} of the corresponding GXT, which was either GXT1 or GXT2. The ICC was calculated for HR_{peak}. Agreement was analysed with a paired samples t-test.

In a Bland-Altman plot the LOA was determined [30]. The LOA expressed as a percentage of the mean described the variability of the results.

Furthermore, to assess validity the SEM, SDD and ES were calculated for the HR_{peak} of the SMBT and the HR_{peak} of the corresponding GXT.

The Spearman correlation was calculated for the motivation scored by the instructor and by the test leader.

The ICC (two-way random, absolute agreement) for HR_{peak} and level was used to compare the results of the aSRT over ground of Waninge et al. and the GXT on the treadmill in the same participants. Agreement was analysed with a Wilcoxon signed ranks test for both HR_{peak} and level.

Estimated peak heart rate
We analysed validity of the equation developed by Fernhall et al. in persons with SIMD, comparing measured HR_{peak} with the estimated HR_{peak} [13]. The HR_{peak} SMBT was used in a regression analysis to calculate β for each participant. To minimize the influence of outliers the five highest and five lowest heart rate scores of the SMBT were eliminated. The equation developed by Fernhall et al. was adjusted by a new β-constant useful for our sample.

Results
Feasibility
The test-retest feasibility of both GXTs was 86.6%, which can be considered as sufficient [28]. Twenty-six out of thirty participants completed both GXT1 and GXT2 successfully. One participant dropped out because of a heart rate irregularity during the first GXT. Two participants dropped out because the GXT1 caused stress and behavioural problems. One participant consistently stepped on the sides of the treadmill and was not able to perform the GXT.

The feasibility of the SMBT was 76.9%, which fails to be sufficient [28]. A total of twenty out of twenty-six participants who completed the GXT1 and GXT2 performed the SMBT successfully (criterion A). Two participants did not perform the SMBT because of a high stress level. Four persons performed the maximal block test according to criterion B.
Table 2. Descriptive results peak heart rate and levels

<table>
<thead>
<tr>
<th></th>
<th>HR_{peak} GXT1 (n=26)</th>
<th>HR_{peak} GXT2 (n=26)</th>
<th>HR_{peak} GXT (n=20)</th>
<th>HR_{peak} SMBT (n=20)</th>
<th>HR_{peak} calculated (n=26)</th>
<th>Level GXT1 (n=26)</th>
<th>Level GXT2 (n=26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (beats/min) (SD)</td>
<td>154 (16)</td>
<td>155 (17)</td>
<td>156 (19)</td>
<td>157 (16)</td>
<td>171 (7)</td>
<td>13.54 (3.9)</td>
<td>13.77 (3.7)</td>
</tr>
<tr>
<td>Range (beats/min)</td>
<td>118-181</td>
<td>117-186</td>
<td>118-181</td>
<td>118-181</td>
<td>159-184</td>
<td>6-22</td>
<td>4-22</td>
</tr>
</tbody>
</table>

HR: heart rate (beats/min); GXT1: 1st GXT; GXT2: 2nd GXT; GXT: test corresponding to the SMBT; SMBT: supra-maximal block test; SD: standard deviation (beats/min)

Test-retest reliability

Table 2 represents the descriptive statistics of the HR_{peak} and the attained levels of GXT1, GXT2 and SMBT. The HR_{peak} data in GXT1 and GXT2 were normally distributed.

Pearson’s correlation between the HR_{peak} GXT1 and the difference in HR_{peak} between GXT1 and GXT2 was low (r =0.11), which is indicative of the homoscedasticity of the results.

The ICC (ICC=0.95; 95%CI 0.88-0.98) for the HR_{peak} of the GXT1 and GXT2 was very good (Table 3). The t-test showed no significant difference between the measurements (p=0.5) (Table 3).

![Figure 1. Bland-Altman plot of HR_{peak} GXT1 and HR_{peak} GXT2 in beats/min; n=26; LOA -16 to 14 (19.4% of mean); mean difference -1; Standard Deviation of the difference 8 beats/min](image)
The LOA are determined in a Bland-Altman plot (Figure 1). The LOA for HR\textsubscript{peak} of GXT1 and GXT2 range from -16 to 14, which is 19% of the mean HR\textsubscript{peak} (mean HR\textsubscript{peak} = 155). This means that the individual variability represented in the region between the dotted lines has a width of 19% of the mean (Figure 1).

The achieved levels of GXT1 and GXT2 were normally distributed. The ICC between the attained levels was very good (ICC=0.95; 95%CI 0.90-0.98). The t-test showed no significant difference between the attained levels of GXT1 and GXT2 (p=0.56) (Table 3).

![Bland and Altman plot of reached levels GXT1 and GXT2; n=26; LOA -3.4 to 2.9; (43.4% of mean); mean difference -0.19](image)

Figure 2. Bland and Altman plot of reached levels GXT1 and GXT2; n=26; LOA -3.4 to 2.9; (43.4% of mean); mean difference -0.19

The LOA for the attained levels of GXT1 and GXT2 range from -3.4 to 2.9 (Figure 2). The mean level reached in both tests was 13.6.

The SDD was 2 levels indicating that persons had to increase or decrease their treadmill performance with 2 levels to have a relevant change in endurance time.

Table 3 presents the SEM, SDD and ES. The mean difference in HR\textsubscript{peak} between GXT1 and GXT2 was 1 beat /minute. The mean difference in HR\textsubscript{peak} between the SMBT and the corresponding GXT was also 1 beat /minute. These are both smaller than the SEM (5 beats/minute in both tests) (Table 3). The SDD for HR\textsubscript{peak} was 6 beats/minute. Thus one can only speak of a significant difference when confronted with a difference larger than 6 beats/minute. The Effect Size (ES) for both HR\textsubscript{peak} and the attained level ranges from small to moderate (0.12) between GXT1 and GXT2 [32].
Table 3. Test-retest reliability of $HR_{\text{peak}}$ and levels achieved

<table>
<thead>
<tr>
<th></th>
<th>$HR_{\text{peak} \ GXT1}$ - $HR_{\text{peak} \ GXT2}$ (n=26)</th>
<th>$HR_{\text{peak} \ GXT1}$ - $HR_{\text{peak} \ GXT} \ \text{(n=20)}$</th>
<th>$HR_{\text{peak} \ GXT1}$ - $HR_{\text{peak} \ \text{calculated}} \ \text{(n=26)}$</th>
<th>$HR_{\text{peak} \ GXT2}$ - $HR_{\text{peak} \ \text{calculated}} \ \text{(n=26)}$</th>
<th>$HR_{\text{peak} \ SMBT}$ - $HR_{\text{peak} \ \text{calculated}} \ \text{(n=20)}$</th>
<th>$HR_{\text{peak} \ GXT1}$ - $HR_{\text{peak} \ GXT2}$ (n=26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LOA % of mean</td>
<td>-16 to 14</td>
<td>-17 to 15</td>
<td>-45 to 13</td>
<td>-43 to 13</td>
<td>-42 to 14</td>
<td>-3.4 to 2.9</td>
</tr>
<tr>
<td>ICC (95%CI)</td>
<td>0.95 (0.88-0.98)</td>
<td>0.94 (0.86 - 0.98)</td>
<td>0.28 (-0.23 - 0.62)</td>
<td>0.39 (-0.21 - 0.72)</td>
<td>0.36 (-0.25 - 0.72)</td>
<td>0.95 (0.90-0.98)</td>
</tr>
<tr>
<td>t-test</td>
<td>-0.6 (p=0.5)</td>
<td>-0.3 (p=0.7)</td>
<td>-5.4 (p=0.001)</td>
<td>-5.3 (p=0.001)</td>
<td>-4.6 (p=0.001)</td>
<td>-0.6 (p=0.56)</td>
</tr>
<tr>
<td>SEM</td>
<td>5</td>
<td>5</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>SDD</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>ES</td>
<td>0.12</td>
<td>0.07</td>
<td>0.73</td>
<td>0.73</td>
<td>0.73</td>
<td>0.12</td>
</tr>
</tbody>
</table>

HR: heart rate; GXT: graded exercise test; LOA: Limits of Agreement; ICC: Intraclass Correlation Coefficient; SEM: Standard Error of Measurement; SDD: Smallest Detectable Difference; ES: Effect Size

Validity

Twenty participants performed the SMBT, according to criterion A. The ICC (ICC=0.94; 95%CI 0.86-0.98) between the $HR_{\text{peak} \ SMBT}$ and the corresponding $HR_{\text{peak} \ GXT}$ was very good (Table 3). The t-test (p=0.7) revealed no significant differences between the $HR_{\text{peak} \ SMBT}$ and the $HR_{\text{peak} \ GXT}$ (Table 3). The LOA results range from -17 to 15. The LOA represents 20.5% of the mean (157 beats/min) (Figure 3).

Eleven persons met the criterion of a maximal performance on the GXT, because the heart rate of the SMBT differed by not more than two beats from the HRpeak reached during the corresponding GXT. In four persons HRpeak SMBT was more than two beats more than HRpeak GXT, which means they did not meet the criterion of a maximal performance on the GXT. In five persons HRpeak SMBT was less than HRpeak GXT.
However, the ICCs between the estimated $HR_{peak}$ and $HR_{peak\ GXT1}$, $HR_{peak\ GXT2}$ and $HR_{peak\ SMBT}$ were all poor (Table 3). The t-statistics comparing estimated $HR_{peak}$ with $HR_{peak\ GXT1}$, $HR_{peak\ GXT2}$ and $HR_{peak\ SMBT}$ all showed a significant difference (Table 3). This means that the agreement between estimated $HR_{peak}$ against $HR_{peak\ GXT1}$, $HR_{peak\ GXT2}$, and $HR_{peak\ SMBT}$ was poor (Table 3).

The ES for $HR_{peak}$ between the SMBT and the corresponding GXT was small (0.07) [32]. The ES for $HR_{peak}$ between GXT1, GXT2 and SMBT with the estimated $HR_{peak}$ were all large (>0.5) [32]. This indicates a large difference between the measures for $HR_{peak}$ found and estimated $HR_{peak}$, which corresponds to the significant difference between these measures recorded by the t-test.

Spearman’s correlation of the VAS motivation scores between the instructor and test leader in GXT1 was low (rho 0.3; p=0.09) but significant for GXT2 (rho 0.7; p<0.001) and SMBT (rho 0.6; p=0.004). So, analysing a correlation between scored motivation and $HR_{peak}$ would only be appropriate for GXT2 and SMBT. Spearman’s correlation between scored motivation and $HR_{peak}$ was significant only for GXT2 for both the instructor (p=0.02) and the test leader (p=0.03) (Table 4).
All but one participant (96%) needed encouragement during GXT1 or GXT2. Fourteen (54%) needed encouragement during both tests and eleven participants (42%) needed encouragement during one test. Out of the total of fifty-two tests of GXT1 and GXT2, a lot of encouragement was given during twelve tests and some encouragement was given during twenty-seven tests. Both GXT1 and GXT2 lasted 14 minutes (SD 3.8) on average, with a range of 4 to 22 minutes.

Correlation between HRpeak of GXT1 on the treadmill and HRpeak of the over ground test was poor and not significant (ICC=0.21; 95% CI -0.24–0.59). For the GXT2 the correlation was also poor and not significant (ICC=0.24; 95% CI -0.22–0.63). The HRpeak for GXT1 and GXT2 treadmill were significantly higher than the tests performed over ground, as was expected.

Correlation between the levels of GXT1 treadmill and aSRT1 over ground was moderate but significant (ICC=0.67; 95% CI = 0.18–0.87). Correlation between the levels achieved for the GXT2 was moderate and also significant (ICC=0.73; 95% CI = 0.22–0.90). The wide confidence intervals indicate diversity among the participants. The achieved levels show a significant difference between the treadmill and the over ground test for the GXT1 (Wilcoxon signed ranks, p=0.03) and the GXT2 (Wilcoxon signed ranks, p=0.01).

Estimated peak heart rate
In our sample the ICC for HRpeak SMBT and the estimated HRpeak was poor (ICC=0.36; 95% CI -0.25–0.72). The t-statistic showed a significant difference (t=-4.6; p=0.001). Figure 4 shows a scatter plot of these findings. The equation seems to over-estimate HRpeak for people with SIMD. To adjust the equation to be relevant to our sample a new constant β=0.88 was calculated.
Chapter 5

Figure 4 Scatter plot calculated HR\textsubscript{peak} and HR\textsuperscript{supra-max} of supra-maximal block test (Pearson’s r=0.5; p≤0.05; t-statistic p=0.0001). The dotted line representing perfect correlation, line of identity and absolute agreement.

Discussion

This study investigated the feasibility, test-retest reliability and validity of the adapted Shuttle Run Test protocol performed on a treadmill (graded exercise test, GXT) for people with SIMD and GMFCS-I. The results indicate that the GXT protocol performed on the treadmill has sufficient feasibility, based on the 80% criterion [28]. The test-retest reliability and validity of the GXT treadmill were sufficient for the sample. Furthermore, the results show that the Fernhall equation predicting HR\textsubscript{peak} consistently overestimated the realized HR\textsubscript{peak}.

High correlation (ICC=0.95; 95%CI 0.88-0.98) and good agreement (p=0.5) has been showed in this study between the GXT1 and GXT2 for HR\textsubscript{peak}, indicating good test-retest reliability. This is in line with the results Fernhall et al. found in a reliability study performed with mentally retarded adults [33]. A SEM of 5 and an SDD of 6 with a small ES indicate that the margin for error is 5 and hence only a difference of more than 6 beats/minute indicates a significant difference. The results of the GXT1 and GXT2 are not significantly different and the ES are small. This confirms good test retest reliability.

A measure for variability is the LOA. With a 19.4% of the mean HR\textsubscript{peak}, the LOA can be said to be wide. A wide LOA points at a considerable variation at the individual level and may be due to difficulties in motivation, behaviour or an inability to cope with stress, all of which are typically prevalent in people with SIMD [1, 3, 33].

In order to check the validity of the test protocol, the HR\textsubscript{peak} SMBT was compared with the
preceding GXT results on $HR_{\text{peak}}$. Good correlation (ICC = 0.94; 95%CI 0.86-0.98) and agreement ($p=0.7$) for $HR_{\text{peak}}$ exists between SMBT and the preceding GXT. These results might serve as indicative for a gold standard according to Midgley et al. [18, 19]. Midgley et al. reviewed available literature regarding the verification phase following a graded exercise test. In the present study, speed during the verification phase was 0.2 or 0.3 km/h higher than speed during the last completed stage, as in persons with SIMD only an increase of speed of 0.2 or 0.3 km/h is feasible [12]. This is considered an acceptable increase of speed, because Midgley et al [19] stated that ‘the verification phase should incorporate a workload higher than that attained in the incremental phase to conform to the original concept of VO2max’. Furthermore, they indicated that the mean difference in $HR_{\text{peak}}$ between the test and the SMBT should be maximal 1.9 beats/minute (SD 1.7) [19]. The results of our study show a mean difference of 1 beat/minute and a standard deviation of 1, which is within the limits drawn by Midgley [19]. The attained performance levels thus were reached using maximal effort. When looked at the individual level, eleven persons met the criterion of a maximal performance on the GXT, whereas four persons did not meet this criterion.

The actual and estimated $HR_{\text{peak}}$ show low correlation. This could indicate that the equation developed by Fernhall et al. [13] is not applicable to our sample of participants. In the Fernhall study, 144 of the 276 participants had mild ID [13]. Similar to the findings of Kittredge et al. [34] we found the $HR_{\text{peak}}$ to be significantly lower than the estimated $HR_{\text{peak}}$ [34], suggesting that the formula's constant ($\beta=0.56$ in the Fernhall equation) should be corrected and take a higher value for people with SIMD. Hence, a recommendation for future research is to adjust the equation for estimating $HR_{\text{peak}}$ for individuals with SIMD, enabling a valid calculation of the estimated $HR_{\text{peak}}'$. This is crucial in assessing whether target heart rate has been reached for this specific group of participants.

Since both $HR_{\text{peak}}$ and attained levels are significantly higher in the treadmill test than in the over ground test, the treadmill test is valid for measuring maximal exercise capacity in people with SIMD and GMFCS-I. Environmental factors may explain the test results of the GXT protocol as performed on the treadmill [35]. The constraints-led approach limits the degrees of freedom during testing [35], resulting in a more restrictive test situation. The physical constraints of the treadmill, the sound of the running walking belt, the side bars and the instructor standing behind the participant all may stimulate the participant to continue walking.

Furthermore, issues of motivation, stress and the ability to understand test directions should be considered when interpreting test results [36]. For people with SIMD motivation for physical activity is low [36] and extrinsic encouragement and rewards often dictate activity performance [37]. This pattern of behavior is also evident in our study. All but one participant needed encouragement, fourteen during both test sessions and eleven during one of the test sessions.

Moreover, unfamiliar situations caused stress in several participants. Rintala et al. reviewed the familiarization process in cardiorespiratory fitness testing in persons with mild to moderate intellectual disabilities, recommending a familiarization protocol [21]. In our study the protocol consisted of walking the treadmill at least twice at regular walking speed before testing. Participants did not reach their $HR_{\text{peak}}$ levels during these practice sessions. Frey et al. described how people with ID are hardly challenged by their support systems to exert physically [3]. During data collection most participants were challenged up to volitional exhaustion [16] for the first time in their lives. This may put forward an explanation for the insufficient feasibility of the SMBT. By letting future participants practice at a sub-maximal exercise level, the feasibility of the SMBT may improve. Furthermore, in future studies a familiarization protocol should be established with
well defined criteria for advancement from one familiarization level to the next. Since people with intellectual disabilities tend to have lower motivation for physical activity [3, 16], we included a motivation score into the test protocol. An aspect that may have influenced the motivation scores was that both the test leader and the instructor were aware of the estimated \( \text{HR}_{\text{peak}} \). This may have influenced the encouragement given, and as a consequence, the scored motivation. A significant correlation between observed motivation and \( \text{HR}_{\text{peak}} \) existed during GXT2 for both the instructor and the test leader (Table 4). Nonetheless, in future studies the inter-observer reliability of scored motivation should be assessed as well.

Handrail support during steady-state treadmill exercise reduces the momentary aerobic demands [38, 39]. All but one participant held on to the handrail during the test procedure. When walking speed increased some participants leaned more heavily on the handrail which may have had an influence on the levels or \( \text{HR}_{\text{peak}} \) [39, 40]. The achieved levels in the GXT treadmill may have been relatively high as a consequence of leaning on the handrail.

A limitation of this study was how to decide when maximal exercise level or volitional exhaustion was reached. As for now, realizing the estimated \( \text{HR}_{\text{peak}} \) seems to be the only objective measure of maximal performance, which in the present study none of the participants achieved. Rintala et al. described volitional exhaustion [16] by signals such as heavy breathing, maximal heart rate, uncoordinated walking, sweating or verbal protest, which is too wide a range for a clear and workable measure. In future studies the volitional exhaustion has to be defined in a more accurate way.

The results of this project could be used to develop an experimental study investigating the trainability of exercise capacity in people with SIMD. Treadmill training could possibly improve health related physical fitness and thereby health for people with multiple disabilities. Further experimental research on training a population with SIMD is recommended.

**Conclusion**
The main conclusion of our study is that a GXT protocol performed on a treadmill is a feasible, reliable and valid test for determining \( \text{HR}_{\text{peak}} \) and exercise levels for people with SIMD and GMFCS-I. For this population, the GXT protocol has better validity for determining \( \text{HR}_{\text{peak}} \) and maximal level than the SRT over ground.

For future research, we recommend a revision of Fernhall’s equation so as to enable a better prediction of the \( \text{HR}_{\text{peak}} \) for people with SIMD.

Furthermore, future studies should comprise of a familiarization protocol with well-defined criteria so as to reduce the influence of stress, stemming from unfamiliarity with the test situation, on the test results. Moreover, volitional exhaustion should be defined more clearly using unambiguous variables. Finally, an evaluation of inter-tester reliability of scored motivation should be established.

**Acknowledgements**
The authors kindly acknowledge and thank the participants for their participation in this study, their representatives for giving permission for this and the physical education instructors for their effort. The authors are grateful for the supportive and valuable feedback. The authors have no conflict of interest to declare.
References


17 Davids K, Button C, Bennett S. Dynamics of Skill Acquisition, a constraints-led approach. 2008, Human Kinetics, Champaign, IL.


