Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions

Yoav Gimmon a, Grinshpon Jacob b, Constanze Lenoble-Hoskovec c, Christophe Büla c, Itshak Melzer a, *

a Department of Physical Therapy, Faculty of Health Sciences, Ben-Gurion University of the Negev, Beer-Sheva, Israel
b Geriatric Department, Soroka University Medical Center, Beer-Sheva, Israel
c Service of Geriatric Medicine and Geriatric Rehabilitation, University of Lausanne Medical Center, CHUV, Switzerland

ARTICLE INFO

Article history:
Received 29 January 2013
Received in revised form 2 February 2013
Accepted 7 February 2013
Available online xxx

Keywords:
Aging
Gait instability
Single task
Dual task
Reliability

ABSTRACT

Decline in gait stability has been associated with increased fall risk in older adults. Reliable and clinically feasible methods of gait instability assessment are needed. This study evaluated the relative and absolute reliability and concurrent validity of the testing procedure of the clinical version of the Narrow Path Walking Test (NPWT) under single task (ST) and dual task (DT) conditions. Thirty independent community-dwelling older adults (65–87 years) were tested twice. Participants were instructed to walk within the 6-m narrow path without stepping out. Trial time, number of steps, trial velocity, number of step errors, and number of cognitive task errors were determined. Intraclass correlation coefficients (ICCs) were calculated as indices of agreement, and a graphic approach called “mountain plot” was applied to help interpret the direction and magnitude of disagreements between testing procedures. Smallest detectable change and smallest real difference (SRD) were computed to determine clinically relevant improvement at group and individual levels, respectively. Concurrent validity was assessed using Performance Oriented Mobility Assessment Tool (POMA) and the Short Physical Performance Battery (SPPB). Test–retest agreement (ICC1,2) varied from 0.77 to 0.92 in ST and from 0.78 to 0.92 in DT conditions, with no apparent systematic differences between testing procedures demonstrated by the mountain plot graphs. Smallest detectable change and smallest real change were small for motor task performance and larger for cognitive errors. Significant correlations were observed for trial velocity and trial time with POMA and SPPB. The present results indicate that the NPWT testing procedure is highly reliable and reproducible.

© 2013 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Falls cause considerable mortality and morbidity among older adults (Rubenstein & Josephson, 2002). Deterioration of the postural control system with aging can lead to reduced ability to control balance while walking. Declines in both gait speed and gait stability have been associated with increased fall risk in older adults (Gunter, White, Hayes, & Snow, 2000; Hausdorff, Rios, & Edelberg, 2001; Rubenstein & Josephson, 2002; Shumway-Cook, Brauer, & Woollacott, 2000), with the majority of falls occurring during walking (Berg et al., 1997; Overstall, Exton-Simith, Imms, & Johnson, 1977; Prudham & Evans, 1981) and with unsteady gait (Rubenstein, 2006). Ganz, Bao, Shekelle, and Rubenstein (2007) found that two of the six most important clinically identifiable risk factors for falls are gait impairments and balance disorders. The latter finding highlights the importance of evaluating dynamic balance abilities during gait. However, the majority of clinical balance testing procedures have focused on either balance performance or on gait performance separately.

There is accumulating evidence that aging effects on balance may be accentuated in the mediolateral (ML) direction. For example, decreased frontal plane stability during up-right standing has been associated with increased fall risk in older adults (Lord, Rogers, Howland, & Fitzpatrick, 1999; Maki, Holliday, & Topper, 1994; Melzer, Benjuya, & Kaplanski, 2004; Melzer, Kurz, & Oddsson, 2010; Stel, Smit, Pluijm, & Lips, 2003). Maki et al. (1994) showed increased ML-sway during standing, and ML-sway was the single best predictor of future falling. Lord et al. (1999) found that subjects with a history of falls had increased ML-sway in a near-tandem stability test, and measures of ML-sway have been associated with recurrent falls (Stel et al., 2003). Melzer et al.
also found increased ML instability during narrow stance in older adults with a history of falls. Stabilogram diffusion analysis revealed significantly higher short-term diffusion coefficients and critical displacements in fallers compared to nonfallers in ML direction, indicating postural sway to drift away from an equilibrium point, unchecked by the postural control system (Melzer et al., 2010). Maki, Edmondstone, and McIlroy (2000) as well as McIlroy and Maki (1996) found a tendency in older adults to fall laterally toward the swinging leg during compensatory stepping, following the forward or backward step. Another study found that rapid voluntary stepping, which challenges ML stability, provided a better prediction of falls (Brauer, Burns, & Galley, 2000). Thus, incorporating assessment of ML stability during gait amongst older adults may be a useful tool in identifying individuals with an increased risk for falls during walking. The Narrow Path Walking Test was developed to challenge one’s balance when walking in a narrowed pathway (Kelly, Scharger, Price, Ferrucci, & Shumway-Cook, 2008).

Mobility during daily life often requires the performance of concurrent cognitive and motor tasks, such as walking and talking. The role of cognition in postural control and gait is being increasingly recognized. Kelly et al. (2008) added a concurrent cognitive task (i.e., dual task) to the NPWT to further challenge gait stability. They found that increasing age and the performance of a concurrent cognitive task were independently associated with decreased speed (Kelly et al., 2008). However, their study utilized expensive laboratory equipment (i.e., motion analysis system) that would not be available to most clinics; also, the reliability of the testing procedure was not reported.

For clinicians, the ability to identify gait instability is critical and needs to demonstrate sensitivity to clinically relevant change and remain invariant when there is no change in function. In this study, we present a clinical version of the NPWT. Before using this new test to examine gait instability and to evaluate the effects of balance training programs in older adults, it is important to establish the relative and absolute test–retest reliability of its testing procedure. Relative reliability is commonly evaluated with the ICC, which evaluates the level of agreement or reproducibility between two or more measurements (Shrout & Fleiss, 1979). The ICC value is of limited use to the clinician since it is not related to the actual scale of measurement and depends on the range of the subjects’ performance (Kelly et al., 2008). Absolute reliability, however, examines the variability or measurement error caused by repeated measurements, such as standard error of measurement (SEM) and smallest real difference (SRD) (Bland & Altman, 1990). SEM denotes the smallest change that indicates a real difference for a group of subjects, while SRD represents the smallest change that indicates a real improvement for a single subject; both have clinical importance when evaluating the effects of treatment or changes over time (Bland & Altman, 1990). In addition to relative and absolute reliability, a graphic approach was used for improved descriptive analyses of variation in scores from different trials in the same participants to guide the clinical interpretation of patterns of discordance; we constructed a folded cumulative distribution curve for the NPWT parameter called mountain plot (Monti, 1995). Mountain plot graphs provide better ability to gain an appreciation of the direction and magnitude of the summary score differences across two raters’ assessments. Another aim of the current study was to investigate the concurrent validity of NPWT with performance-based, widely used, clinical measures of balance and gait.

2. Methods

Thirty independent healthy older adults (20 women) aged 65–87 years (mean age 81.8 ± 6.1 years) were recruited. This supplementary study was part of the baseline assessment in a prospective study, the “Self Mobility Improvement of Elderly by counteractINg falls” (SMILING). The SMILING study is a small-medium, prospective multi-center randomized controlled study of the effect of a perturbation-training device on balance and gait parameters. This study was conducted with subjects of “Mishan Avot HaNegev”, a protected retirement home for older adults. Subjects were included based on the following criteria: age ≥65 years; being able to walk independently; having no serious visual impairments (i.e., disorders of retina, glaucoma, cataract, and blindness caused by injury); subjects were excluded if scoring less than 24 on Folstein’s mini-mental state examination (Folstein, Robins, & Helzer, 1983), suffering from severe cardiovascular disease, lower limb amputation, terminal diseases, Parkinson’s disease, multiple sclerosis, status post-stroke, cervical or lumbar stenosis, or Ménière’s disease. All subjects provided written informed consent, in accordance with approved procedures by the Helsinki ethics committee at Soroka University Medical Center, Israel.

To assess reproducibility of the testing procedure between examiners, participants performed the complete NPWT twice on two different occasions within 1 h, conducted by two trained examiners in random order. Two participants were tired after the first testing procedure, thus were re-tested within 10 days. Each participant’s performance on the NPWT was videotaped. Each examiner later rated the set of video clips he had conducted (“1st testing procedure” or “2nd testing procedure”). The examiners were blinded to each other’s results. Participants also completed Tinetti’s POMA (Tinetti, Williams, & Mayewski, 1986), and Short Physical Performance Battery (SPPB) (Guralnik et al., 1994). During these tests participants were scored on balance and gait performance tasks (maximum 28 in the POMA and 12 the SPPB; higher scores indicating better balance and gait functions).

2.1. Experimental protocol

During the NPWT participants were asked to walk at a comfortable pace within a narrow path (6 m long) both without and with concurrent cognitive task (single- and dual-task) (Kelly et al., 2008). To produce a similar challenge for individuals with different body morphologies, the width of the narrow path was normalized to 50% of the distance between the participant’s anterior superior iliac spines plus the width of the subject’s shoe. The narrow path was outlined by two narrow 6-m long black carpets. Participants were instructed to walk at their normal speed within the path between the carpets without stepping on them (Fig. 1). Participants were allowed to practice 2–3 times to become familiar with the test situation before assessment. For each testing procedure, participants performed 3 trials under single task conditions (ST) and 3 trials under dual task (DT) conditions. Three different cognitive tasks were used to prevent learning effect of the cognitive tasks (e.g., reciting the days of the week backwards, the months of the year backwards, and to count down in increments of 5 from 100 to 50 at first, second, and third DT trials, respectively). Participants were asked to perform both tasks as best as they can, without any instruction to focus on one of the tasks. All 6 trials were videotaped using a video-camera placed 1.5 m in front of the walking path at about 1 meter height, in prevision of subsequent analysis.

Parameters extracted during video analysis were: number of steps, trial time, step errors, and trial velocity. A step error was defined as every step where the participant’s shoe touched the carpet outline on the sides of the narrow path walkway. Cognitive task errors were also assessed, i.e., the number of cognitive task mistakes made during each ST trial (sitting trial) and each DT (walking trial).

2.2. Sample size

The sample size was calculated post hoc specifying that an ICC of 0.8 would be the minimum acceptable, and the null hypothesis
was an ICC of 0.4. With 80% power and \( p = 0.05 \), at least 16 subjects were required in each group (Donner & Eliasziw, 1987). Considering this, the sample was judged to be valid for performing the current reliability-analysis.

2.3. Data and statistical analysis

Based on the video clip analysis, average values of the parameters assessed were calculated separately for ST and DT conditions. Comparative measures of the path width for the 1st and the 2nd testing procedures were also done. Relative reliability of reproducibility of testing procedure between examiners was estimated using a two-way random model for ICC (Fleiss, 1986). ICCs (2,1) were used assuming that each subject was assessed by two different raters, in random order, and that reliability was calculated from an average measurement. Reliability was considered “high” for ICCs > 0.80, “substantial” for ICCs between 0.61 and 0.80, “moderate” for ICCs between 0.41 and 0.60, and “poor to fair” for ICCs < 0.4 (Bland & Altman, 1986).

Absolute reliability, i.e., measurement error, at group level was quantified by calculating SEM and SEM%; and at individual level through the SRD and SRD% (Bland & Altman, 1990) as described in the following equations:

\[
SEM = \sqrt{WMS}
\]

\(1\) where WMS is the mean square error term from ANOVA.

\(2\) \(SEM\% = (SEM/mean \times 100)\)

\(3\) \(SRD = 1.96 \times SEM \times \sqrt{2}\)

\(4\) \(SRD\% = (SRD/mean) \times 100\)

where the last equation is the mean of all the data from the two test occasions.

In addition, mountain plot graphs for the NPWT parameters were generated separately for ST and for DT trials: (A) trial time (s); (B) number of steps; (C) trial velocity (m/s); (D) number of step errors; and (E) number of cognitive task errors. The mountain plot graphs display the cumulative distribution of the difference between the scores in the two testing procedures. The graph is created by computing a percentile rank for ranked difference between respondents’ scores and “folded” at the 50th percentile rank (percentile rank is defined as the proportion of cases having

---

Please cite this article in press as: Gimmon, Y., et al., Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions. Arch. Gerontol. Geriatr. (2013), http://dx.doi.org/10.1016/j.archger.2013.02.001
lower or equal value to the score under consideration). The “mountain graph” plots cumulative percentages (y-axis values) against the ranked difference scores (x-axis values). In “folding” the graph, the percentile ranks for difference scores above the 50th percentile are obtained by subtracting the actual percentile rank from 100 (Monti, 1995). The mountain plot has several advantages, including locating the median immediately, easily determining symmetry, and observing outliers to determine central or tail percentiles. These advantages are especially useful in examining scoring differences for each NPWT parameter scored by the 1st and the 2nd testing procedure.

Concurrent validity of the NPWT was evaluated through the associations between NPWT parameters, POMA and SPB8 scores using partial Spearman correlations controlled for age (r) since the data were not normally distributed. Correlation magnitude was estimated as absent to little (r = 0.00–0.25), low (0.26–0.49), moderate (0.50–0.69), high (0.70–0.89); or very high (0.90–1.00) (Domholdt, 2005). All statistics were analyzed using SPSS (version 16, Chicago, IL, USA).

3. Results

3.1. Relative reliability

There were no significant differences between raters for all NPWT parameters (Table 1).

The relative reliability of the NPWT was “substantial” to “high” with ICC values ranging from 0.77 to 0.92 (p < 0.0001) for ST and DT conditions (Table 2). Similarly, “substantial” reliability was found for the measure of path width between raters (ICC = 0.79, p < 0.0001).

### Table 1

<table>
<thead>
<tr>
<th>Single task</th>
<th>Testing procedure #1 average ± SD (95% CI)</th>
<th>Testing procedure #2 average ± SD (95% CI)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of steps</td>
<td>15.0 ± 3.3 (13.7–16.2)</td>
<td>15.5 ± 3.3 (14.3–16.8)</td>
<td>0.51</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>11.1 ± 4.1 (9.6–12.7)</td>
<td>10.1 ± 4.5 (8.4–11.7)</td>
<td>0.12</td>
</tr>
<tr>
<td>Trial velocity (m/s)</td>
<td>0.6 ± 0.2 (0.5–0.7)</td>
<td>0.68 ± 0.2 (0.6–0.7)</td>
<td>0.11</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>4.9 ± 4.7 (3.2–6.6)</td>
<td>5.8 ± 4.4 (4.2–7.5)</td>
<td>0.26</td>
</tr>
<tr>
<td>Number of cognitive task errors</td>
<td>0.5 ± 1.0 (0.15–0.9)</td>
<td>0.7 ± 1.4 (0.2–1.2)</td>
<td>0.53</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dual task</th>
<th>Testing procedure #1 average ± SD (95% CI)</th>
<th>Testing procedure #2 average ± SD (95% CI)</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of steps</td>
<td>17.3 ± 5.1 (15.4–19.2)</td>
<td>16.8 ± 3.6 (15.5–18.2)</td>
<td>0.91</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>18.2 ± 9.5 (14.7–21.8)</td>
<td>14.6 ± 6.1 (12.3–16.9)</td>
<td>0.10</td>
</tr>
<tr>
<td>Trial velocity DT (m/s)</td>
<td>0.4 ± 0.16 (0.3–0.5)</td>
<td>0.47 ± 0.15 (0.4–0.5)</td>
<td>0.10</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>8.0 ± 7.3 (5.3–10.7)</td>
<td>8.6 ± 5.8 (6.5–10.8)</td>
<td>0.38</td>
</tr>
<tr>
<td>Number of cognitive task errors</td>
<td>1.5 ± 2.4 (0.6–2.4)</td>
<td>1.45 ± 1.8 (0.7–2.2)</td>
<td>0.31</td>
</tr>
</tbody>
</table>

p-Value measured using Mann–Whitney U test. s, seconds; m/s, meters per second.

### Table 2

Intraclass correlation (ICC2,1 and 95% CI), SEM, coefficient of variation (SEM%), and SRD, mean difference between the two testing procedures (d), 95% confidence interval of d (95% CI) and limits of agreement (LOA) for all parameters of the NPWT in ST and DT conditions.

<table>
<thead>
<tr>
<th>ST condition</th>
<th>ICC2,1 (95% CI)</th>
<th>SEM</th>
<th>SEM%</th>
<th>SRD</th>
<th>SRD%</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of steps</td>
<td>0.85 (0.68–0.93)</td>
<td>0.57</td>
<td>3.75</td>
<td>1.58</td>
<td>10.4</td>
<td>0.5</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>0.77 (0.51–0.89)</td>
<td>0.74</td>
<td>6.90</td>
<td>2.05</td>
<td>19.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Trial velocity (m/s)</td>
<td>0.88 (0.76–0.94)</td>
<td>0.36</td>
<td>0.56</td>
<td>0.1</td>
<td>15.6</td>
<td>0.08</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>0.92 (0.82–0.96)</td>
<td>0.78</td>
<td>14.4</td>
<td>2.1</td>
<td>39</td>
<td>0.09</td>
</tr>
<tr>
<td>Number of cognitive task errors</td>
<td>0.81 (0.59–0.91)</td>
<td>0.24</td>
<td>30</td>
<td>0.06</td>
<td>82.5</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DT condition</th>
<th>ICC2,1 (95% CI)</th>
<th>SEM</th>
<th>SEM%</th>
<th>SRD</th>
<th>SRD%</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of steps</td>
<td>0.89 (0.77–0.95)</td>
<td>0.75</td>
<td>4.4</td>
<td>2.1</td>
<td>12.3</td>
<td>0.5</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>0.84 (0.66–0.92)</td>
<td>1.04</td>
<td>6.3</td>
<td>3.77</td>
<td>22.9</td>
<td>3.6</td>
</tr>
<tr>
<td>Trial velocity (m/s)</td>
<td>0.89 (0.76–0.95)</td>
<td>0.89</td>
<td>6.9</td>
<td>0.78</td>
<td>17.9</td>
<td>0.07</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>0.92 (0.83–0.96)</td>
<td>1.12</td>
<td>13.3</td>
<td>3.4</td>
<td>40.5</td>
<td>0.6</td>
</tr>
<tr>
<td>Number of cognitive task errors</td>
<td>0.78 (0.54–0.90)</td>
<td>0.40</td>
<td>25</td>
<td>1.1</td>
<td>70.5</td>
<td>0.5</td>
</tr>
</tbody>
</table>

p-Value was <0.0001 for all test parameters. s, seconds; m/s, meters per second.

Please cite this article in press as: Gimmon, Y., et al., Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions. Arch. Gerontol. Geriatr. (2013), http://dx.doi.org/10.1016/j.archger.2013.02.001
The differences near indicate performance reported zero steps between time; errors; velocity therefore, Fig. 2. NPWT in ST condition for older adults. Mountain plot for agreement (testing procedure #1 vs. testing procedure #2): (A) trial time; (B) number of steps; (C) trial velocity (m/s); (D) number of steps errors; and (E) number of cognitive task errors (n = 30).

For the single task trials, the median difference score was near +1.5 for trial time, near −0.5 for the number of steps and the number of step errors, near −0.1 for trial velocity, and near zero for the number of cognitive task errors. These results indicate that no large systematic differences were detected between the first and second NPWT testing procedures (Fig. 2). Differences range from approximately +5 to −5 units for trial time; +3 to −3 for number of steps; +5 to −5 units for number of steps errors; +1 to −1 units for the number of cognitive task errors; and +0.2 to −0.2 for the trial velocity. The magnitude of such differences is small (about 1 standard deviation away from the median in both tails). For the “left tail” of the mountain plots, only two differences were greater than one standard deviation to the left of the median in trial time (5 units); there were four in the number of steps (3 units), three in trial velocity (0.2 units), two in the number of step errors (5 units), and three in number of cognitive task errors (1 unit) (Fig. 2). For the “right tail” of the mountain plots, the number of differences falling beyond one standard deviation to the right of the median were three in the trial time, two in the number of steps and number of cognitive task errors, and zero in the number of step errors and trial velocity. Therefore, most of the differences are within plus and minus one standard deviation.

Y. Gimmon et al. / Archives of Gerontology and Geriatrics xxx (2013) xxx–xxx e5

Please cite this article in press as: Gimmon, Y., et al., Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions. Arch. Gerontol. Geriatr. (2013), http://dx.doi.org/10.1016/j.archger.2013.02.001
Table 3  Correlations (r) and p-value of NPWT and POMA and SPPB balance and gait measures controlled for age.

<table>
<thead>
<tr>
<th></th>
<th>POMA balance</th>
<th>POMA gait</th>
<th>POMA overall</th>
<th>SPPB balance</th>
<th>SPPB gait</th>
<th>SPPB overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. ST condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of steps</td>
<td>-0.43 (0.012)</td>
<td>-0.52 (0.002)</td>
<td>-0.47 (0.005)</td>
<td>-0.34 (0.021)</td>
<td>-0.49 (0.003)</td>
<td>-0.58 (0.001)</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>-0.48 (0.004)</td>
<td>-0.62 (0.001)</td>
<td>-0.55 (0.001)</td>
<td>-0.35 (0.041)</td>
<td>-0.55 (0.001)</td>
<td>-0.67 (0.0001)</td>
</tr>
<tr>
<td>Trial velocity (m/s)</td>
<td>0.48 (0.004)</td>
<td>0.62 (0.0001)</td>
<td>0.55 (0.001)</td>
<td>0.35 (0.041)</td>
<td>0.55 (0.001)</td>
<td>0.67 (0.0001)</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>-0.23 (0.197)</td>
<td>-0.24 (0.166)</td>
<td>-0.22 (0.214)</td>
<td>-0.21 (0.227)</td>
<td>-0.27 (0.118)</td>
<td>-0.28 (0.11)</td>
</tr>
<tr>
<td>B. DT condition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of steps</td>
<td>-0.32 (0.062)</td>
<td>-0.42 (0.013)</td>
<td>-0.36 (0.037)</td>
<td>-0.43 (0.012)</td>
<td>-0.41 (0.016)</td>
<td>-0.48 (0.004)</td>
</tr>
<tr>
<td>Trial time (s)</td>
<td>-0.31 (0.074)</td>
<td>-0.51 (0.002)</td>
<td>-0.38 (0.028)</td>
<td>-0.33 (0.054)</td>
<td>-0.45 (0.008)</td>
<td>-0.56 (0.001)</td>
</tr>
<tr>
<td>Trial velocity (m/s)</td>
<td>0.31 (0.074)</td>
<td>0.51 (0.002)</td>
<td>0.38 (0.028)</td>
<td>0.33 (0.054)</td>
<td>0.45 (0.008)</td>
<td>0.56 (0.001)</td>
</tr>
<tr>
<td>Number of step errors</td>
<td>-0.28 (0.113)</td>
<td>-0.27 (0.118)</td>
<td>-0.3 (0.086)</td>
<td>-0.45 (0.007)</td>
<td>-0.21 (0.245)</td>
<td>-0.46 (0.006)</td>
</tr>
</tbody>
</table>

s, seconds; m/s, meters per second.

For the DT trials, the median difference scores was near +1 for trial time; near zero for the number of steps, trial velocity, and number of cognitive task errors; and near -1 for the number of step errors, indicating that no systematic differences were detected between the first and second NPWT testing procedures (Fig. 3). Similar to the single task trials, most of the differences were within plus and minus one standard deviation. For the “left tail”, there was one difference beyond one standard deviation in trial time (5 units), zero for the number of steps (3 units); three in the number of step errors (5 units), and four in number of cognitive task errors (1 unit) and trial velocity (0.15 units), falling beyond one standard deviation to the left of the median. For the “right tail”, there were seven differences in the trial time, four in the number of steps and number of cognitive task errors, and one in the number of step errors and trial velocity, falling beyond one standard deviation to the right of the median.

3.4. Concurrent validity

Significant negative correlations were found between POMA, SPPB, and the NPWT parameters under ST conditions (Table 3A). Highest coefficients were observed for trial velocity and time, while correlations for the number of step errors were the lowest and did not reach statistical significance. In general, NPWT parameters had higher correlation coefficients with gait than balance domains in POMA and SPPB.

Results found for DT trials were essentially similar except that the low correlations between number of step errors reached statistical significance with SPPB balance (r = -0.45, p = 0.007), and SPPB overall score (r = -0.46, p = 0.006) (Table 3B).

4. Discussion

This study’s results show that the NPWT has “substantial” to “high” absolute and relative reliability, as well as stability over time. In addition, low to moderate correlations were observed with other well validated and widely used clinical measures of gait and balance performance. This study also contributes to extending knowledge about the NPWT in providing values for meaningful changes in test performance. Overall, these results suggest that the NPWT is a reliable, effective, and clinically feasible method to evaluate gait instability and assess balance control during walking in older adults.

NPWT parameters showed substantial to high relative reliability in both ST and DT testing conditions (ICC ranging from 0.77 to 0.92 and 0.78 to 0.92, respectively), with ICC values similar to, and in some cases higher than, those reported for other functional tests of gait and balance (Berg, Wood-Dauphinee, Williams, & Gayton, 1989; Curb et al., 2006; Halvarsson et al., 2012; Harada, Chiu, & Stewart, 1999; Melzer, Shitlman, Rosenblatt, & Oddsson, 2007). Among NPWT parameters, trial velocity and number of step errors had the highest reliability in ST and DT conditions (ICC = 0.92 and ICC = 0.88–0.89, respectively). Furthermore, trial velocity also had the strongest association with POMA and SPPB (Table 3), two instruments widely used for gait and balance assessment. Overall, these results strongly suggest that trial velocity might be the single best NPWT parameter to estimate fall risk in older adults. This is further supported by results from previous studies that showed strong association between gait velocity and falls (Abellan van Kan et al., 2009; Hausdorff, 2005; Maki, 1997).

This study also points to other directions for future studies. For instance, the cross-sectional and prospective relationship between the number of step errors at NPWT and falls has not yet been investigated. Step errors might be occasioned by either inability to tightly control balance in ML direction during NPWT narrow base gait between the carpets, or performing a lateral protective step to avoid a fall when losing balance. Both hypothesized mechanisms suggest that step errors at NPWT might indeed reflect frontal plane instability, a feature that is not well captured by usual existing clinical performance-based measures of balance and gait. Thus, an important and specific contribution of the NPWT to the field of clinical gait and balance assessment might be to provide specific information on frontal plane instability. This seems especially relevant as frontal plane instability has also been related to a greater risk for falls in older adults (Brauer et al., 2000; Kelly et al., 2008; Lord et al., 1999; Maki et al., 1994, 2000; Mclroy & Maki, 1996; Melzer et al., 2004, 2010; Stel et al., 2003). Interestingly, the associations between the number of step errors and POMA as well as SPPB were significant only under DT conditions. This suggests that higher cognitive demand further impacts ML stability control during the NPWT because it requires activity planning and executing a motor task (i.e., placing the feet one in front of the other during the narrow path walking). The NPWT is a much more complex procedure that requires higher executive function abilities, something that neither the POMA nor the SPPB test appropriately.

The present study further contributes to the field of gait and balance assessment in providing cut-off values to define clinically relevant changes at the NPWT. Low SEM and SRD scores in the NPWT revealed that a reasonably small improvement might be enough to detect relevant changes in a group of older adults or a single elderly person. Individual improvement (or decline) for trial velocity in ST and DT conditions were under 15.9% and 17.9%, respectively, and group improvement (or decline) of less than 0.56% and 6.9%, respectively, can be considered as of no clinical relevance for this parameter. Generally, the smaller the SRD or SEM the better. In the present study, the SEM values were below 14.4% in ST and 13.3% for DT (apart from the cognitive task errors), which can be considered small (Smidt et al., 2002). Interestingly, Perera, Mody, Woodman, and Studenski (2006) found that substantial change near 0.10 m/s for gait speed is considered to be clinically meaningful, a value very similar to results of the current study.
The mountain plot graphs (Figs. 2 and 3) did not reveal any systematic bias between the two NPWT testing procedures. Median difference scores for all NPWT parameters in ST and DT were near zero, indicating high stability and no large systematic differences between the testing procedures. Skewness was found to be asymmetrical only in DT condition, with slightly longer “right tails” (“left tail” for trial velocity). Most likely, this results from a learning effect, with some few participants scoring better during the 2nd than the 1st testing procedure.

The small sample size of the study clearly limits the stability of the findings. Further studies are needed to examine the construct validity, responsiveness, and other psychometric properties of the NPWT. Moreover, assessment is needed of its relative and absolute reliability in other populations, such as fallers and patients suffering from neurologic and orthopedic impairments. Finally, the ability of NPWT to predict falls needs to be investigated in a prospective study.

5. Conclusion
The clinical version of the NPWT seems to have high relative and absolute reliability, as well as stability over time. It is a relatively simple and feasible tool for clinical use to measure gait stability in community-dwelling older persons, and measures under ST and DT conditions closely reflect “real life” situations. The present study provides values for meaningful changes in gait instability assessment in older adults and sets the stage for future use of NPWT in investigating gait evolution over time as well as the effects of balance training in older adults.

Authors’ contributions
YG was involved in conducting experiments, subjects recruitment, data analysis and interpretation as well as drafting the manuscript. GJ was involved in subject recruitment and testing as well as drafting the manuscript. CLH and CB were involved in the experimental design, statistical planning and statistical analysis as well as drafting the manuscript. IM was the PI of the research he was involved in planning experimental design, subjects’ recruitment and responsibility for the day-to-day operations of the research project, data analysis and interpretation as well as drafting the manuscript.

Conflict of interest statement
The authors declare that they have no competing interests, any financial and personal relationships with other people or organizations that could inappropriately influence (bias) their work.

Acknowledgments
The authors wish to acknowledge the contribution of the management of “Mishan Avot HaNegev”, protected retirement home for older adults, to the research study. This work was partially supported by a grant from the European Union “Self Mobility Improvement of elders by counteracting falls” (Grant number EU FP7, GA #215493).

References

Please cite this article in press as: Gimmon, Y., et al., Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions. Arch. Gerontol. Geriatr. (2013), http://dx.doi.org/10.1016/j.archger.2013.02.001


Y. Gimmon et al. / Archives of Gerontology and Geriatrics xxx (2013) xxx–xxx

Please cite this article in press as: Gimmon, Y., et al., Relative and absolute reliability of the clinical version of the Narrow Path Walking Test (NPWT) under single and dual task conditions. *Arch. Gerontol. Geriatr.* (2013), [http://dx.doi.org/10.1016/j.archger.2013.02.001](http://dx.doi.org/10.1016/j.archger.2013.02.001)