Research Report

Reliability and Responsiveness of Two Physical Performance Measures Examined in the Context of a Functional Training Intervention

Background and Purpose. The reliability and responsiveness of 2 physical performance measures were assessed in this nonrandomized, controlled pilot exercise intervention. Subjects. Forty-five older individuals with mobility impairment (mean age=77.9 years, SD=5.9, range=70–92) were sequentially assigned to participate in an exercise program (intervention group) or to a control group. Methods. The intervention group performed exercise 3 times a week for 12 weeks that targeted muscle force, endurance, balance, and flexibility. Outcome measures were the 8-item Physical Performance Test (PPT-8) and the 6-minute walk test. Test-retest reliability and responsiveness indexes were determined for both tests; interrater reliability was measured for the PPT-8. Results. The intraclass correlation coefficient for interrater reliability for the PPT-8 was .96. Intraclass correlation coefficients for test-retest reliability were .88 for the PPT-8 and .93 for the 6-minute walk test. The intervention group improved 2.4 points and the control group improved 0.7 point on the PPT-8, as compared with baseline measurements. There was no change in 6-minute walk test distance in the intervention group when compared with the control group. The responsiveness index was .8 for the PPT-8 and .6 for the 6-minute walk test. Conclusion and Discussion. Measurements for both the PPT-8 and the 6-minute walk test appeared to be highly reliable. The PPT-8 was more responsive than the 6-minute walk test to change in performance expected with this functional training intervention. [King MB, Judge JO, Whipple R, Wolfson L. Reliability and responsiveness of two physical performance measures examined in the context of a functional training intervention. *Phys Ther.* 2000;80:8–16.]

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

Key Words: *Older people, Physical performance, Reliability, Responsiveness.*

Mary B King

James O Judge

Robert Whipple

Leslie Wolfson

Preventing loss of physical function is a primary goal of physical therapists who treat older people. The development of interventions that forestall functional deterioration is, therefore, an important area of study. Definitive trials to prevent functional decline are expensive because they involve large samples followed for several years. Before a definitive trial can be undertaken, there must be some evidence that an intervention proposed to prevent disability is both feasible and effective. Physical performance tests may be used to complement self-reports of functional status, and they can serve as surrogate outcome measures of function for use in preliminary trials to determine the potential efficacy of interventions. Performance scales usually use time (eg, time to complete a task, time maintaining a balance position) as a gauge of performance. Measures of physical performance, in addition to having construct and concurrent validity, must be stable on repeated measurement (ie, have test-retest reliability) and be sensitive to changes in performance resulting from an intervention (ie, responsive). The responsiveness of a measure, expressed as the responsiveness index (RI), is its ability to detect minimal clinically important differences after an intervention.[1]

In this pilot study, we used 2 performance tests—the 8-item Physical Performance Test (PPT-8)[2] and the 6-minute walk test[3]—as outcome measures in a 12-week exercise intervention for community-dwelling older people with mild mobility impairments. These tests were chosen from a long list of available performance measures because they measure the ability to do several common daily tasks (PPT-8) and endurance (6-minute walk test). Physical Performance Test (PPT) scores have been shown to predict adverse outcomes of nursing home placement or death.[4] The 7- and 9-item PPTs have been validated in older outpatients, demonstrating inter-

MB King, MD, is a geriatrician in the Hartford Hospital Geriatric Program, Hartford, Conn, and Assistant Professor of Medicine, University of Connecticut School of Medicine, Farmington, Conn. Address all correspondence to Dr King at Hartford Hospital Geriatric Program, Burlingame Bldg, 2nd Floor, 400 Washington St, Hartford, CT 06106 (USA) (mking@nso.uchc.edu).

JO Judge, MD, is Vice President of Medical Services, Masonicare, Wallingford, Conn, and Associate Professor of Medicine, University of Connecticut Health Center, Farmington, Conn.

R Whipple, PT, is Assistant Professor of Neurology, University of Connecticut Health Center.

L Wolfson, MD, is Chairman and Professor, Department of Neurology, University of Connecticut Health Center and Hartford Hospital.

Mark Thomas, MD, Odin Kuiper, MD, Asif Ali, MD, Doug Mangan, and Mawuena Quist assisted with the physical performance testing.

Concept and research design, writing, and consultation (including review of manuscript before submission) were provided by King, Judge, Whipple, and Wolfson. Data collection/analysis and subjects were provided by King and Judge; project management, by King, Judge, and Whipple; and facilities and equipment, by Judge, Whipple, and Wolfson. Wolfson provided fund procurement, and Whipple provided institutional liaisons. Cynthia Gruman, PhD, contributed to data analysis and consultation.

This study was approved by the University of Connecticut Health Center Institutional Review Board.

This work was supported by National Institute on Aging grants AG-09675 and AG-13631, and by funds provided by the University of Connecticut Center on Aging.

This article was submitted December 12, 1998, and was accepted September 9, 1999.

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

nal consistency for the individual items tested as well as construct and concurrent validity.[2] In the study describing the development of the PPT as a physical performance measure,[2] concurrent validity was established by finding that the PPT-7 and PPT-9 scores were highly correlated with the modified Rosow-Breslau scale scores, instrumental and basic activities of daily living scale scores, and the Tinetti gait score. Use of the PPT-8 as an outcome measure in an intervention study has not been reported. The 6-minute walk test was originally developed as a measure of exercise capacity in patients with chronic heart failure or pulmonary diseases.[3,5] Test-retest reliability has been determined in patients with peripheral arterial occlusive disease.[6] The 6-minute walk test has also been used as an outcome measure in studies to improve fitness in people with single chronic diseases (osteoarthritis of the knee,[7] chronic lung disease[8,9]), but it has not been used in studies of older people with varied or multiple causes of functional impairment.

In this article, we describe the reliability and responsiveness of these 2 performance measures. The aim of this pilot study was to test the sensitivity of these measures to detect change in performance after an intervention designed to improve the capacity to carry out common daily activities in older community residents. We did not examine the validity of these measures. Data from this study can be used to improve the accuracy of study size estimates for definitive intervention trials in older subjects with early mobility impairment.

Method

Study Design and Subjects

The study was a nonrandomized, controlled intervention trial. Letters introducing the study, with questionnaires about health and mobility, were mailed to 2,560 members of 2 senior centers in a suburb of Hartford, Conn. In response, 239 members, aged 70 years or older, expressed an interest in participation. Potential subjects were screened for early mobility impairment, first by their answers to questions on the mailed questionnaire and then by testing at introductory group meetings held at the senior center where testing and the intervention took place. The entry criteria were either: (1) moderate difficulty or inability to perform at least one of the following mobility-related activities, reported on the returned questionnaires[10,11]: walk 0.4 km (¼ mile); climb 1 flight of stairs; stoop, crouch, or kneel; push large objects; and carry 4.5 kg (10 lb); or (2) have a usual gait speed of less than 1.0 $m \cdot s^{-1}$ on walking 8 m. The 8-m walk was tested at the end of the introductory meeting. Each volunteer was timed walking a distance of 8 m, marked on a long hallway at the senior center.

Of 150 interested volunteers who attended the group meetings, 98 volunteers met the entry criteria. The first 55 eligible volunteers then gave consent to participate and were scheduled for a focused history and physical examination, which included cardiac, neuromuscular, and joint examinations. The Mini-Mental State Exam (MMSE) and a questionnaire about performance of basic activities of daily living were also given. The 43 remaining eligible volunteers were placed on a contact list for future studies. Exclusion criteria included: assistance with activities of daily living; cognitive impairment (MMSE score of <24); stroke, Parkinson disease, or other major neurologic deficit; coronary artery disease or congestive heart failure with symptoms during moderate activity; poorly controlled hypertension; inability to walk 8 m independently; use of neuroleptics or benzodiazepines; or current physical therapy. Ten volunteers did not complete entrance testing or were excluded by study criteria after the physical examination.

The first 26 volunteers who met the study criteria were assigned to the exercise program, and the 19 subsequent enrollees were assigned to a control group. The subjects' average age was 77.9 years (SD=5.9, range=70–92). Seventy-one percent of the subjects were women, and 79% had 12 or more years of education. Twenty-eight percent of the subjects reported having fallen in the past year, and 55% reported moderate difficulty with one or more of the mobility-related activities. Transportation was provided by senior center van to volunteers who were otherwise unable to get to the senior center where the exercise program and testing took place. Thirty-seven subjects (18 subjects in the intervention group, 19 subjects in the control group) completed the study.

Physical Performance Testing

All subjects were tested with the PPT-8 and the 6-minute walk test. The PPT is designed to measure the ability to use the upper and lower extremities in everyday activities, and it was developed as a 7- or 9-item test. The performance of the following tasks is timed and scored in the 7-item PPT: writing a sentence, simulated eating, lifting a book and putting it on a shelf, putting on and removing a jacket, picking up a small object from the floor, turning 360 degrees, and walking 15.2 m (50 ft). The 9-item PPT includes 2 additional items: climbing one flight of stairs and counting the number of flights of stairs the subject is able to ascend.[2] The ninth item (the number of flights of stairs) was dropped in this pilot study because we believed there was the potential for subjects to become fatigued with climbing multiple flights of stairs. Climbing a single flight of stairs was included along with the other 7 items because we believe it measures performance of an important (albeit more difficult) daily physical activity. An 8-item PPT has not been described in the literature, but because the 7- and 9-item PPTs have both been validated,[2] we decided, after discussion with the developer of the test (David B

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

Reuben, MD; personal communication), to use the 8-item test and to examine the results for internal consistency. Each PPT-8 item was scored from 0 points (lowest) to 4 points (highest) according to published test protocol.[2] The maximum total PPT-8 score was 32 points.

The 6-minute walk test was administered on an indoor 32-m course with low-pile carpeting. Subjects were instructed to walk as far as they could in 6 minutes. At 2 and 4 minutes, they were informed of the time elapsed and were given the standard encouragement, "You're doing a good job."[5] The distance (in meters) walked in 6 minutes was recorded. Electronic pulse monitors*[1] were worn around the chest and on the wrist and were used to record heart rate throughout the test. Heart rate at 6 minutes was compared from first baseline measurement to follow-up tests.

There were 4 testers for the PPT-8: 3 medical students who were masked to group assignment and an unmasked investigator who did not participate in the exercise training. A geriatrics fellow who was masked to group assignment was the single tester for the 6-minute walk test. Testing occasions are shown in Figure 1. Interrater reliability of the PPT-8 was checked by having 2 raters simultaneously time each of 18 subjects during the first baseline test. In order to establish test-retest reliability for the PPT-8 and the 6-minute walk test, 22 subjects in the intervention group were tested on 2 additional occasions, each 1 to 2 weeks apart, after the first baseline test and before beginning the intervention. Because we believed interrater reliability was excellent for the PPT-8, repeat testing was done either by the same rater or by a different rater. The 37 subjects who completed the 12-week study were tested with the same measures at the end of the intervention. Results were compared with the first baseline test results, as about half of the subjects had only one baseline test.

Intervention

The 12-week training program was held at the senior center 3 times a week, with classes lasting 75 minutes each. Classes were led by the study physical therapist and an exercise leader, with 6 to 10 subjects per class. Individual attention was given to the subjects, who progressed at their own pace. The program was a pilot trial of center-based exercise designed to improve factors that affect physical performance (ie, muscle force, endurance, flexibility, balance).[12] Muscle force and endurance were the focus of the first 6 weeks of training (phase 1). In the second 6 weeks (phase 2), subjects concentrated on the elements of balance and flexibility,

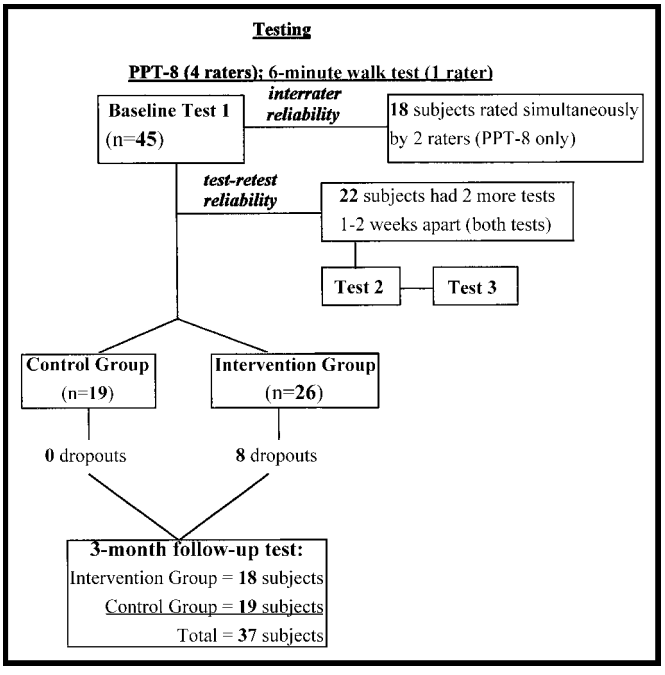

Figure 1.
Diagram showing sequence of testing at baseline and follow-up for the 8-item Physical Performance Test (PPT-8) and 6-minute walk test.

while continuing to maintain muscle force and endurance gains from phase 1.

In phase 1, each session began with brisk walking. Subjects were instructed to walk at a moderate intensity, using the Borg Rating of Perceived Exertion Scale[13] (intensity of 12–14 on the 6–20 Borg Rating of Perceived Exertion Scale). Walking duration was increased from 6 minutes in the first week to a maximum of 15 minutes in the third week, which was then maintained through week 6. Subjects who became tired were permitted to rest and then resume walking. Resistance exercises were performed using sandbags or dumbbells, and props such as chairs, doorways, mats, and walls. Exercises with the following movements were used: ankle plantar flexion and dorsiflexion, performed in a standing position; knee and hip extension by step-ups, using increasing step height (15.2–20.3 cm [6–8 in]) and adding weighted vests; elbow flexion/biceps curls with progressively increasing dumbbell weights; wall push-ups; seated "dips"; and hip flexion and abduction using increasing sandbag resistance at the ankles. Hip flexion was performed supine, with the inactive leg in hooklying. Hip abduction was performed in side-lying, with the knee and hip flexed on the down side.

The first week of training was a breaking-in period in which submaximal loading was used (ie, ≥15 repetitions at moderate effort). In free-weight exercises (elbow flexion, dorsiflexion, hip abduction, hip flexion), loads were estimated on the basis of 1 or 2 trial-and-error

* Polar Electro Inc, 370 Crossways Park Dr, Woodbury, NY 11797.

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

Table 1.
Summary of Phase 2 Exercises

Exercise Description	Objective	Protocol
15-min endurance circuit		
Brisk walk	Endurance	152.4 m (500 ft)
Stair climbing	Force enhancement/endurance	2–4 flights at a time
Lift heavy objects from floor to shelves	Force enhancement	1.1- to 3.3-kg (2.5- to 7.5-lb) sandbags, lift to waist, shoulder, overhead heights, 2 cycles
Push-ups	Force enhancement/endurance	Body inclined 45°–60° to wall, wide and narrow hand separations, 1 set of 15 repetitions
Sit-to-stands	Force enhancement/endurance	15 rapid chair rises and returns in a row
Static standing heel cord stretch	Flexibility	Single 1-min holds, each leg
Standing heel-ups	Force enhancement	15 repetitions bilateral; 15 repetitions twice, unilateral
Standing bilateral dorsiflexion	Force enhancement	Level achieved at end of phase 1; 15 repetitions
Anteroposterior limits of stability maximal leaning without bending at waist	Balance	2 min practice; hold for 10 s at anterior and posterior limits
1-leg standing	Balance	2 stands per leg; 1 min each
Lunge-to-kneel	Balance/force enhancement/flexibility	2 sets 15 repetitions/leg; knee-touch to variable-height raised mats
Static hamstring muscle stretch	Flexibility	Long-sitting, trunk propped in extension, isometric knee extension and dorsiflexion
All-4s ⇔ side-sitting	Flexibility/force enhancement	2 sets 15 repetitions/side
Supine ⇔ prone righting	Flexibility/force enhancement	15 repetitions, alternating
Quadrupedal arm/leg raises	Balance/force enhancement	2 sets 15 repetitions, reciprocating
Bridging	Balance/force enhancement	15 repetitions bilateral plus 15 repetitions each side
Kneel ⇔ stand	Balance/force enhancement/flexibility	5 repetitions maximum
Dips (body elevation by pushing down with arms)	Force enhancement	2 sets 15 repetitions; from long-leg sitting (or on chair) with variable-height blocks
360° standing turns	Balance	10 repetitions each direction

attempts, using sex, medical and exercise history, and the therapist's impressions of the subject's robustness and walking vigor as factors. During week 2, in conjunction with performance impressions from week 1, an 8- to 10-repetition maximum load level was estimated. Two sets were carried out with each muscle group, with a goal of 15 repetitions each before increasing the load by 0.45 to 1.13 kg (1.0–2.5 lb). Similar repetition criteria were used for exercises using principally body weight (heel-ups, push-ups, dips, step-ups); however, the load was increased by altering body position or by use of props. For example, plantar-flexion (heel-rises) progression was as follows: bilateral → bilateral on 3.8-cm (1.5-in) block → unilateral → unilateral on block → unilateral on block with weighted vest. Dips were done straight-sitting on a mat (or in a chair if the subject was unable to get to the floor) by pushing down against blocks of progressively increasing height. For step-ups, all subjects started with 15.2-cm blocks. Beginning with the third week, block height was increased to 20.3 cm, unless the subject developed knee pain. All subjects were candidates for vest loading. Vests could not accommodate more than 4.5 kg (10 lb). The starting load was always 0.9 kg (2 lb). The week in which initial loading began (not earlier than week 2), and the subsequent increase in loading, was determined by the exercise leader, taking into consideration factors such as subject height and weight, sex, exercise history, disease burden, history of knee pain and pathology, and perceived degree of exertion or discomfort during step-ups with no weight during week 1.

Phase 2 sessions (Tab. 1) began with 15 minutes of a self-paced 6-component endurance circuit that emphasized functional tasks requiring force and balance. The next 60 minutes was devoted primarily to balance and flexibility exercises. Some exercises were carried over or modified from phase 1 for purposes of maintaining force and endurance gains attained in phase 1.

Control subjects were given no intervention and were instructed not to begin an exercise program during the

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

Table 2.
Means and Intraclass Correlation Coefficients (ICCs) for Repeated Measures of the 8-Item Physical Performance Test (PPT-8) and the 6-Minute Walk Test in 22 Subjects at Baseline Testing[a]

	Baseline Test 1		Baseline Test 2		Baseline Test 3		Repeated-Measures ANOVA		
	$\bar{X}$	SD	$\bar{X}$	SD	$\bar{X}$	SD	$F_{2,19}$	*P*	ICC
PPT-8 score (0–32)	22.1	4.5	22.6	4.8	23.2	4.8	4.1	.03	.88
6-minute walk test (m)	321	105	341	107	322	108	NS		.93

[a] Repeated-measures analysis of variance (ANOVA) was done to determine whether there were differences among the 3 baseline tests. NS=not significant.

study. They were encouraged to maintain their usual level of physical activity; they did not report to the senior center during the intervention. At the conclusion of the 12-week intervention period, they were offered an exercise program similar to that described.

Dropouts and Adherence

Eight of the 26 subjects assigned to the intervention group dropped out of the study. Seven of those subjects dropped out within the first 4 weeks. No control group subjects dropped out of the study. Three subjects dropped out of the study because of surgery or intercurrent illness, 3 subjects did not tolerate the exercise program (increased joint or muscle pain), and 2 subjects quit because of other commitments. For those subjects who completed the exercise intervention, average attendance was 86.3% (SD=8.7%, range=55%–100%); subjects performed 97% of assigned exercises during the sessions. Two subjects with arthritis of the knee noted increased symptoms during training but continued to exercise with minor adjustments in their programs.

Data Analysis

SYSTAT 4.2[†2] and SPSS Release 6.0[‡3] software were used for the statistical analyses. Cronbach's alpha was used to determine internal consistency of the PPT-8 as a scale and was compared with that of the PPT-7 in this study. Interrater reliability of baseline measurements was determined using the intraclass correlation coefficient (ICC). The ICC described by Fleiss[14] was used to determine test-retest reliability of the baseline measurements, as the study conformed to a one-way random-effects model.

The data for all subjects who had follow-up testing were included in the analysis for responsiveness to change. A repeated-measures analysis of variance (ANOVA) was used to test for a time (repeated-measures) effect and an exercise (group × time) effect. The RI was determined by dividing the mean change in score (PPT-8) or distance walked (6-minute walk test) for the intervention group by the square root of twice the mean square error for the change in score or distance walked in the control group.[1] The relationship between PPT-8 score on the first baseline test and change in score at follow-up was measured using linear regression. A similar analysis was done on the 6-minute walk test results.

Results

Reliability

The Cronbach alpha for the PPT-8 scale scores was .785. For the PPT-7 (all items except for climbing a flight of stairs), the Cronbach alpha was .740. The ICC for interrater reliability on testing 18 subjects with the PPT-8 was .96. Mean test scores and ICCs for the 22 subjects who had 3 baseline tests for the PPT-8 and 6-minute walk test are shown in Table 2. Scores on the PPT-8 improved by 1.1 points from the first baseline test ($\bar{X}$=22.1 points, SD=4.5) to the third baseline test ($\bar{X}$=23.2 points, SD=4.8). The differences in 6-minute walk test distance did not reach statistical significance when tested by repeated-measures ANOVA. The ICC for the PPT-8 was .88, and the ICC for the 6-minute walk test was .93.

Sensitivity to Change

At baseline, the control group's scores on the 2 performance measures were slightly better than those of the intervention group (Tab. 3), but the differences were not statistically significant, as determined by the *t* test. The intervention group improved their performance on the PPT-8 by an average of 2.4 points, and the control group improved their performance on the PPT-8 by an average of 0.7 point (Tab. 3). The difference between groups in improvement in PPT-8 scores was statistically significant (ANOVA F=4.6; *df*=1,35; *P*=.04 for group × time effect). To determine whether this improvement in the intervention group was due to the PPT items that were practiced, a 5-item subset of the PPT-8 was examined (lifting a book and putting it on a shelf, picking up a cube from the floor, turning 360°, walking 15.2 m [50 ft], climbing one flight of stairs). The improvement was similar to that seen on the PPT-8 and was statistically significant (Tab. 3). Scores on the 3-item subset of tasks that were not practiced did not improve at follow-up testing in either group (Tab. 3).

[†] SYSTAT Inc, 1800 Sherman Ave, Evanston, IL 60201.
[‡] SPSS Inc, 444 N Michigan Ave, Chicago, IL 60611.

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

Table 3.
Summary of Performance Outcomes (PPT-8, PPT-5, PPT-3, and 6-Minute Walk Test) for Intervention and Control Groups[a]

Measure	Intervention Group (n=18)						Control Group (n=19)						ANOVA			
	Baseline			Follow-up			Baseline			Follow-up			Effect of Repeated Measures (Time Effect)		Effect of Exercise (Group × Time Effect)	
	$\bar{X}$	SD	Range	$\bar{X}$	SD	Range	$\bar{X}$	SD	Range	$\bar{X}$	SD	Range	F	*P*	F	*P*
PPT-8 score (0–32)	22.7	4.0	17–19	25.1	3.2	19–31†	23.6	4.0	14–29	24.3	4.0	14–30	15.9	<.001	4.6	.04
PPT-5 score (items used: 0–20)	14.9	2.8	10–19	16.6	2.3	10–20	15.7	2.9	8–19	16.1	2.9	7–20	11.1	.002	4.9	.03
PPT-3 score (items not used: 0–12)	7.8	2.1	3–12	8.5	1.4	5–11	8.1	1.8	5–12	8.5	1.4	6–11	5.1	.03	NS	
6-minute walk test distance (m)	337	94	179–498	367	97	161–526*	350	85	128–474	363	95	176–532	11.2	.002	1.9	.18

[a] Paired *t* test, 2-tailed, asterisk (*) indicates $P=.05$, dagger (†) indicates $P<.001$. PPT-8=8-item Physical Performance Test, PPT-5=5-item Physical Performance Test, PPT-3=3-item Physical Performance Test, ANOVA=analysis of variance, NS=not significant.

The 6-minute walk test distance did not increase (Tab. 3). There was no change in pulse at 6 minutes from the first baseline test to the follow-up test in either group.

A regression analysis was used to determine whether baseline performance predicted response to training. Subjects in the intervention group with poor performance on the PPT-8 at the baseline test had the greatest improvement in performance ($r=.38$, $P<.05$ for PPT-8; Fig. 2). There was no relationship between baseline distance (in meters) walked and improvement on the 6-minute walk test. The RI was calculated as .8 (2.4/3.1) for the PPT-8 and .6 (30/52) for the 6-minute walk test.

Discussion and Conclusions

In this pilot study, the PPT-8 was shown to have good internal consistency as an 8-item scale, with a Cronbach alpha of .785; there was excellent interrater reliability as well (ICC=.96). The addition of the item for time to climb one flight of stairs to the PPT-7 adds another test of lower-extremity function. There was a small increase in PPT-8 scores from the first baseline test to the third baseline test (ie, from 22.1 points [SD=4.5] to 23.2 points [SD=4.8]), which may have represented a testing effect. Despite this finding, the PPT-8 demonstrated excellent test-retest reliability, with a high ICC of .88. Subjects, for the most part, did not remember previous testing when tested at follow-up; therefore, it seems unlikely that the observed testing effect carried over for 12 weeks. There was a great deal of variability among subjects in the 6-minute walk test results, as evidenced by a large standard deviation, which resulted in the failure of the observed 20-m increase from the first baseline test to the second baseline test to reach statistical significance. The high ICC indicated excellent test-retest reliability for the 6-minute walk test.

The 12-week exercise intervention was designed to improve muscle force, endurance, flexibility, and balance, the key factors affecting physical function.[12] In our previous FICSIT (Frailty and Injuries: Cooperative Studies on Intervention Techniques) study in which we used an intensive, high-tech program of balance and muscle force exercises, we found that older individuals made gains in both balance and force production only if they had a program that incorporated both balance and resistance exercise.[15]

Our goal in this study was to pilot test an exercise program for older people beginning to have mobility impairment that could be carried out at a senior center without high-tech equipment. The objective was to improve flexibility and endurance as well as force and balance in functional activities. Although we believe that clinically meaningful gains were made by the interven-

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

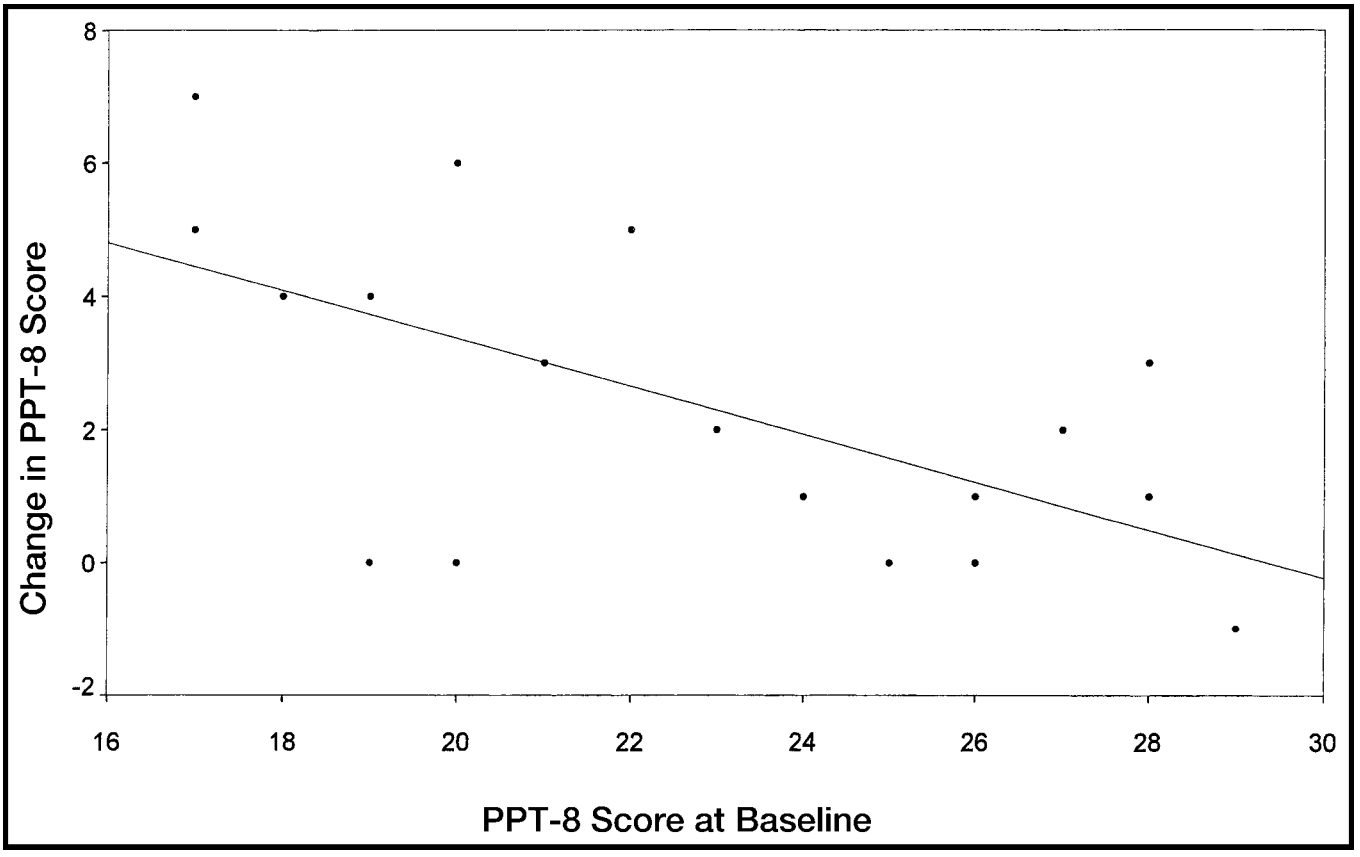

Figure 2.
Linear regression of baseline 8-item Physical Performance Test (PPT-8) scores with change in PPT-8 scores postintervention for the exercise group. Regression line shown, R^2=.38.

tion group after 12 weeks, we would expect even better performance after a longer intervention. The 2 tests used as outcome measures were chosen because they measured changes in performance that were expected with this exercise intervention. The PPT-8 measures performance of usual activities requiring force, balance, and flexibility, and the 6-minute walk test is a measure of endurance. When compared with the control group's scores, the improvement in the PPT-8 scores in the intervention group was clinically meaningful. Subjects in the intervention group with poor performance at baseline testing had the greatest improvement in PPT-8 scores. This finding may have been the result of a ceiling effect, with better performers at baseline testing already being close to the maximum achievable score. There was a greater improvement in the 6-minute walk test distance in the intervention group when compared with that of the control group at follow-up, but this difference did not achieve statistical significance.

The RI is a good indicator of the ability of a test to detect meaningful change due to an intervention. The responsiveness of a measure is determined both by the improvement in scores due to the intervention and by the variability of scores from baseline to follow-up in the control group. Thus, the RI increases as the difference in intervention group scores from baseline to follow-up increases and the variability in change in control group scores decreases.[1,16]

In our study, the RI of the 2 measures was calculated by using the change in scores (PPT-8) or meters walked (6-minute walk test) observed in the intervention group as the clinically meaningful difference. One danger of using this method to determine the RI is that the intervention may have produced an effect that was greater or less than the minimally clinically important difference or one that was statistically significant but clinically unimportant. We feel that the improvements obtained by the intervention group on both measures were clinically meaningful, based on data from our study and the results of previous studies. Although we used a new, untested intervention, the intervention was designed to improve the performance of the tasks tested by the performance measure. Five items tested by the PPT-8 (stair climbing, walking, stooping, lifting, and turning) were directly or indirectly practiced. When our analysis was limited to these 5 items, the results were similar to those obtained with the 8-item scale, suggesting that the PPT-8 was measuring meaningful improvements resulting from the intervention. The moderate intensity of the walking component of the intervention

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022

did not lead to improvement in 6-minute walk test distance in the intervention group when compared with the control group. Judging by the RI of .6, we would have needed almost 3 times as many subjects per group to show a statistically significant difference between groups. As this was a pilot intervention, we used this information in planning sample size for a larger and longer intervention project.

The results of previous studies indicate the changes seen after our exercise intervention were clinically meaningful. In one prospective study,[4] there was a baseline difference of 3.7 points on the PPT-7 between older people who, at 18 to 24 months follow-up after an initial health and function screening, were living at home and those who were living in a nursing home or had died. The 6-minute walk test has been used in intervention studies, although not previously in this population. An 8-week program of fitness walking in 102 patients aged 40 years or more with osteoarthritis resulted in a 70-m increase in 6-minute walk test distance; control subjects had a 17-m decrease.[7] Twenty-one subjects, aged 49 to 71 years, with chronic heart failure improved their 6-minute walk test distance by 18% after 16 weeks of aerobic training.[8] The 6-minute walk test distance improved by 57 m in a study of older patients, aged 70 to 89 years, with chronic obstructive airways disease who had a 12-week program of home exercise and respiratory rehabilitation.[9] Although the differences in PPT scores and improvements in 6-minute walk test distance were greater in the studies cited than in our study, we believe the 2.4-point improvement in PPT-8 scores that we found was a minimally clinically meaningful change. A 2-point improvement in PPT-8 scores would indicate that the person did 2 of the 8 tasks faster, or had decreased time for one task enough to improve by 2 points.

The relatively high dropout rate and number of musculoskeletal complaints in the intervention group highlight the challenges of intervening in older people with self-reported mobility impairments or impaired physical performance. Five of the 8 dropouts were for reasons unrelated to the intervention; 3 of these dropouts were for intercurrent illness or surgery—reasons that may be expected in frail older people. Five subjects in the intervention group complained of joint or muscle pain due to the intervention, and 3 of the 5 subjects dropped out because they could not tolerate the exercise program. Efforts were made to adjust the program for these individuals, and we were able to retain 2 of the 5 subjects. It is possible that the pace or intensity of the present intervention was too high for some subjects or that, in attempting to provide a uniform training stimulus, we were not sufficiently sensitive to preexisting musculoskeletal problems. A trade-off between providing a uniform stimulus (or intervention "dose") and possibly increasing side effects of training must be made in the design of intervention trials for frail older people. Efforts must also be made to assess the exercise goals of older people and their commitment to an exercise intervention prior to enrollment in a study. Given the small size of the study and the short duration of the training, the results reported here are very encouraging for intervention trials designed to improve performance.

References

1 Guyatt GH, Walter S, Norman G. Measuring change over time: assessing the usefulness of evaluative instruments. *J Chronic Dis.* 1987; 40:171–178.

2 Reuben DB, Siu AL. An objective measure of physical function of elderly outpatients: the Physical Performance Test. *J Am Geriatr Soc.* 1990;38:1105–1112.

3 Guyatt GH, Sullivan MJ, Thompson PJ, et al. The 6-minute walk: a new measure of exercise capacity in patients with chronic heart failure. *Can Med Assoc J.* 1985;132:919–923.

4 Reuben DB, Siu AL, Kimpau S. The predictive validity of self-report and performance-based measures of function and health. *J Gerontol.* 1992;47:M106-M110.

5 Guyatt GH, Pugsley SO, Sullivan MJ, et al. Effect of encouragement on walking test performance. *Thorax.* 1994;39:818–822.

6 Montgomery PS, Gardner AW. The clinical utility of a six-minute walk test in peripheral arterial occlusive disease patients. *J Am Geriatr Soc.* 1998;46:706–711.

7 Kovar PA, Allegrante JP, MacKenzie CR, et al. Supervised fitness walking in patients with osteoarthritis of the knee: a randomized, controlled trial. *Ann Intern Med.* 1992;116:529–534.

8 Kavanagh T, Myers MG, Baigrie RS, et al. Quality of life and cardiorespiratory function in chronic heart failure: effects of 12 months' aerobic training. *Heart.* 1996;76:42–49.

9 Roomi J, Johnson MM, Waters K, et al. Respiratory rehabilitation, exercise capacity, and quality of life in chronic airways disease in old age. *Age Ageing.* 1996;25:12–16.

10 Nagi SZ. An epidemiology of disability among adults in the United States. *Milbank Mem Fund Q Health Soc.* 1976;54:439–467.

11 Rosow I, Breslau N. A Guttman health scale for the aged. *J Gerontol.* 1966;21:556–559.

12 Jette AM, Branch LG, Berlin J. Musculoskeletal impairments and physical disablement among the aged. *J Gerontol.* 1990;45:M203-M208.

13 Borg GA. Psychophysical bases of perceived exertion. *Med Sci Sports Exerc.* 1982;14:377–381.

14 Fleiss JL. Reliability of measurement. In: Fleiss JL, ed. *The Design and Analysis of Clinical Experiments.* New York, NY: John Wiley & Sons Inc; 1986:1–15.

15 Wolfson L, Whipple R, Derby C, et al. Balance and strength training in older adults: intervention gains and Tai Chi maintenance. *J Am Geriatr Soc.* 1996;44:498–506.

16 Weiner DK, Bongiorni DR, Studenski SA, et al. Does functional reach improve with rehabilitation? *Arch Phys Med Rehabil.* 1993;74:796–800.

Downloaded from https://academic.oup.com/ptj/article/80/1/8/2842463 by guest on 16 August 2022