Study Protocol and Statistical Analysis Plan

Official Title: **Effects of Exercise and Cognitive Training on Executive Function in Parkinson's Disease (NCT 01156714)**

Document Date: **9/28/18 (date when document was most recently updated)**

Study Protocol:

Screening Methods:

Screening for participation in the study will be held in 1) Baltimore VAMC, and 2) UMPDC. Patients will go through initial evaluation (history, medical & neurological exams), mental exam, visual acuity exam, exercise fitness test and daily living questionnaire. Assessment will be done during the "ON period", when antiparkinsonian medications are working optimally. Below is a description of each assessment tool:

Informed Consent: Written consent for participation will be obtained from each participant at their first evaluation by trained study staff. Participants will undergo an Evaluation to Sign Consent. An assessment of patient understanding of the protocol is also required prior to study entry.

Hoehn & Yahr: The Hoehn & Yahr scale (57) will be used to assess general staging of PD. Classification ranges from 1 to 5, with higher scores indicating greater disability. The cut-point in the present study is stages 1- 3 (1 = Unilateral involvement only, usually with minimal or no functional impairment, 2 = Bilateral or midline involvement, without impairment of balance, 3= First sign of impaired righting reflexes; patients are physically capable of leading independent lives, and their disability is mild to moderate). This assessment will be performed by a movement disorders specialist.

UPDRS: The Motor subscale of the UPDRS (58) will be used to assess disease severity, with particular attention to items #28 (gait) and #30 (postural instability). To be included in this study, a subject must score between 1 and 3 on item #28, and 1 and 2 on item #30, indicating the some impairment of gait or balance but preserved ability to walk independently. On the motor UPDRS, scores each item ranges from 0-4 with a maximum of 108. Higher scores indicate greater disease severity. The motor UPDRS will be assessed by neurologists who specialize in movement disorders.

OARS: The OARS (55) is comprised of 14 questions including 7 ADLs and 7 IADLs. ADLs are self-care activities (walking, eating bathing, dressing, grooming, transferring in/out of bed and using the toilet), IADLs are activities necessary to function independently in the environment (housework, shopping, travelling, preparing meals, using the telephone, managing medication and handling money). For each ADL and IADL, the patient is asked if they are able to perform the task independently. In the original version of the OARS each activity is scored with a 3-point scale (1=no difficulty, 2=needs some help, 3=unable to perform). We will use a modified version of the OARS previously reported by our research group for patients with PD (63). The modified version adds two intermediate responses: slower/with greater difficulty and needs moderate help. To be included in the proposed study, a subject must score between 1 and 3 in the gait section of the OARS. The total maximum score is 70 indicating the highest level of disability.

Folstein Mini-Mental State: The Folstein Mini-Mental State (59) examination will be used to assess gross cognitive function and rule out dementia. The test will be administered by a trained research assistant.

Montreal Cognitive Assessment: The Montreal Cognitive Assessment will be used to assess mild cognitive impairment.

Corrected Visual Acuity: A corrected near visual acuity of > 20/70 will be required to ensure subjects can participate in computer-based cognitive training protocols and neuropsychological testing. Visual acuity will be measured with a vision chart developed by Ferris et al… This chart conforms to the standards of acuity testing proposed by the National Academy of Science-National Research Center Committee on Vision (NAS-NRC 1980). A force-testing procedure will be adopted, where the participant is required to guess even if the letters appear illegible until at least 4 of 5 letters on a row are named incorrectly. This test will be administered by a trained research assistant.

Entry TM Exercise Test: determines cardiopulmonary safety and neuromotor capacity to participate. To acclimate to the TM, subjects start at zero incline and 0.1 mph, increasing by 0.1 mph increments to a self-selected pace (with physician rating of gait stability).(1) Following a 15-minute rest, subjects perform a peak effort TM exercise test. The incline is gradually increased at the pre-selected constant velocity to assess cardiopulmonary response and safety of more strenuous exercise.(2) Patients minimize handrail support, gait safety belt is worn at all times, and tests are physician-supervised with continuous vital signs and ECG monitoring.

Cardiovascular Fitness Test: is conducted with open circuit spirometry, using the subject's pre-selected velocity at progressive inclines.(1-2) Since reliability of fitness testing is not established in PD, each subject performs 2 exercise tests, 1-2 weeks apart, to avoid confounding effects of fatigue. Highest of 2 values is accepted as V02 max or peak, if true max is not achieved. (3) Data are expressed as V02 ml/kg/min to account for any change in weight that may occur. Intra-class correlation coefficients will analyze reliability of repeated baseline tests. Peak ambulatory work load capacity (Mets) is estimated from exercise tests, as previously reported.(4) All tests are conducted in the VA Exercise & Robotics Center using velocity-calibrated TM's and Sensormedics Metabolic Systems.(2,5)

Interventions:

The intervention protocols will be implemented at VAMHCS sites. All 3 study interventions will be conducted by study personnel with the same training will continue for 3 months. When the intervention is completed, patients will receive two post-assessments within a week of the final training session.

TAEX Protocol: The exercise protocol of the TAEX group has been successfully implemented in patients with PD. Subjects will exercise 3 times per week for 3 months. Treadmill walking will start conservatively at 15 minutes with target heart rate 40–50% of maximal heart rate reserve (HRR), by the formula of Karvonen (64). Subjects incapable of continuous treadmill walking will perform intermittent exercise until 20 minutes is achieved. Exercise intensity and duration will be increased by 5 minutes every two weeks, and 0.1 MPH increments as tolerated until reaching a target of 30 minutes at 70-80% HRR. Speed and/or hill will be increased as tolerated every week to maintain the 70-80% HRR. Vital signs will be taken prior, during and after exercise. Heart rate monitoring (Polar) will be used to monitor exercise intensity. Training will

be discontinued for blood pressure drop > 20 mm Hg for medical evaluation. Logbooks will be used to document daily training parameters, complaints or concerns. All participants will be supported in a non-weight bearing fashion in a Biodex™ harness (Biodex Medical Systems, Inc Shirley, NY) to eliminate risk of falls. Health professionals credentialed in exercise training and cardiopulmonary resuscitation will monitor all training sessions at each intervention site (Baltimore VAMC and Parkinson & Movement Disorder Center).

TCOG Protocol: Subjects in the TCOG group will perform cognitive training 3 times per week for 3 months. A commercially available computer-based program will be used (Insight Program by Posit Science, San Francisco, CA). This program was chosen for our study because it has been validated in studies of older adults (14, 15) and is readily available for use. Training will consist of five visual-based exercises designed to improve divided attention, working memory, visual information processing speed, and useful field of view. The program is set up so that the level of difficulty is continuously adjusted to user performance, to maintain a correct rate of about 85%. The program uses a Bayesian algorithm to estimate and then adjust to the individual's performance on a trial-by-trial basis (66). Correct trials are rewarded with points and animations.

Each training session will last 40 minutes, and subjects will be monitored to spend 10 minutes on each of the 4 exercises per session. Training will be performed in a quiet room with monitoring by a trained research assistant to assure the subject's understanding of the tasks, the time spent on each exercise, and the level of motivation. Even though this could be executed by the subject alone, the presence of a trainer during cognitive exercises has been recommended in the literature (13). Adherence to the protocol will be monitored using electronic data upload to the Posit Science Research Database after each training session. Subject will complete cognitive remediation under an identification code. Data will be de-identified upon transmission. In addition, logbooks will be used daily to document subjects' complaints or concerns.

TAEX + TCOG Protocol: Subjects in the combined intervention (TAEX + TCOG) group will perform both treadmill exercise and cognitive training 3 times per week for 3 months. The order of exercise will be TAEX followed by TCOG. This order was chosen to take advantage of the immediate effects of TAEX on learning, which might augment the cognitive performance during TCOG training (16-18). The treadmill portion of the combined intervention will follow the same protocol as the TAEX group, and the cognitive portion of the combined intervention will follow the same protocol as the TCOG group. The overall time for the combined intervention will be 1 hour and 30 minutes. Two health professionals credentialed in exercise training and cardiopulmonary resuscitation will monitor the TAEX portion of the intervention and a research assistant will monitor the TCOG portion of the intervention.

Outcomes Testing:

Pre-tests and post-tests will be conducted at the Baltimore VAMC. All measurements will be performed by trained professionals (exercise physiologists, neuropsychologists and research assistants) blinded to the subject group. Below is a description of the outcome measures categorized into cognitive, cardiorespiratory function, behavioral, dual task, and daily activity measures.

Outcome Measures specific to Aim 1: Compare the efficacy of TAEX versus TCOG versus TAEX + TCOG to improve EF in PD.

Cognitive Measures:

Neuropsychological assessment will consist of both paper-and-pencil and computer-based tests of EF. Subtests from the ANAM test system were selected since: 1) our previous work validated these tests in PD (see preliminary data section a) 2) tests were required that are free of floor or ceiling effects, since a wide range of cognitive ability is expected in our sample, 3) tests must isolate the effects of the disease on cognition from the biasing effects of age-related cognitive decline, since older subjects with PD will participate in the study. Our selected test battery meets these qualifications. Following is a description of each selected test:

a. ANAM tests:
The ANAM was developed by the Department of Defense (DoD) as a computerized performance battery sensitive to cognitive change (67). It is administered via computer, uses a mouse interface and requires approximately 30 minutes of administration time. Seven subtests shown to be sensitive to cognitive change in PD a) will be administered, including Simple Reaction Time, Two-Choice Reaction Time, Procedural Reaction Time, Delayed Matching to Sample, Sternberg Memory Search, Running Memory Continuous Performance and the Tower Test. The battery utilizes a pseudo-randomization procedure to generate stimulus items such that each test session (pre-test and post-test) includes a different combination of items in order to minimize practice effects. Ratings of speed, accuracy, and general cognitive efficiency are generated for each subtest. Below is a brief description of each subtest.

Simple Reaction Time:
The test consists of pressing a specified response key each time a stimulus is presented on the computer screen. The test provides a measure of pure reaction time as well as a means to parse out the effects of motor response from cognitive processing time.

Two-Choice Reaction Time:
This test measures patients' ability to shift mental set. One of two stimuli are presented on the screen (e.g. + or *). Subjects press a specified response button on the keyboard corresponding to the presented stimulus. In order to obtain a measure of cognitive processing time, the Simple Reaction Time scores are subtracted from the Two-Choice Reaction Time scores.

Procedural Reaction Time:
This test measures processing speed and decision making speed. Subjects differentiate between two sets of characters. The test presents a stimulus on the screen (e.g. the number 2, 3, 4, or 5). Subjects are required to press the left mouse key if a 2 or 3 is presented, or the right mouse key if a 4 or 5 is presented. As for the previous test, in order to obtain a measure of cognitive processing time, the Simple Reaction Time scores are subtracted from the Procedural Reaction Time scores.
Delayed Matching to Sample:
This test assesses visual working memory. Subjects view a 4X4 red and white block design.

After a delay they are shown two designs from which they must select the original design.

Sternberg Memory Search:
This test assesses working memory and sustained attention. Subjects learn a 6-letter set then determine if individually presented letters were part of the set.

Running Memory Continuous Performance:
This test assesses working memory and sustained attention. Subjects determine if a letter appearing on the screen is the same or different from the letter immediately preceding it in a sequence.

The Tower Test:
This test assesses planning, problem solving and adherence to rules. Subjects are presented with a display of blocks situated on a "tower." Their goal is to move blocks from a random position to create a replica of the display tower in the fewest number of moves possible while adhering to stated rules.

b. Paper-and-Pencil Cognitive Tests:

Stroop Test:
The Stroop Color and Word Test (68) is a measure of selective attention and cognitive flexibility in which the subject must inhibit a preponderant response. Subjects are asked to complete three parts under timed conditions: (1) reading words describing colors that are written in black-and-white, (2) naming those colors when printed as Xs, and (3) naming the color ink when words describing the colors are mismatched with the colors (suppressing the verbal content). The scores for this test are the number of items completed correctly in 45 seconds.

Delis-Kaplan EF System (D-KEFS):
The D-KEFS sorting subtest (69) is a measure of problem-solving, concept-formation, and cognitive flexibility. Subjects are asked to sort six cards into two groups according to as many rules as possible. The outcome measure for this test is the number of correct sorts within 4 minutes. The ability to recognize the correct sorting rules is also assessed.

The North American Adult Reading Test (NAART):
The NAART provides an estimate of intelligence by assessing the ability to read 61 non-phonetic words (70).

Action Fluency:
Action Fluency is a measure of executive function in which the examinee is asked to spontaneously produce as many action words (i.e., verbs) as possible in one minute (71). Previous research has found this test to be particularly sensitive to cognitive decline in patients with PD (72). The total score is the number of different verbs produced in one minute.

Cardiorespiratory Function Measures:

a. Cardiorespiratory fitness test:
A cardiorespiratory fitness test with open circuit spirometry will be conducted in all study subjects (54, 65, 73). For detailed description see section 1.4.7. Data will be expressed as V02 in ml/kg/min to account for variation in body weight. Peak ambulatory work load capacity (METs) will also be estimated from exercise tests, as previously reported (54, 73). This test will be conducted in the VAMC Exercise Physiology Lab using non-rebreather masks (Hans Rudolph), velocity-calibrated treadmills and Sensormedics Metabolic Systems (54, 64, 73).

b. Six-minute walking test:
The six-minute walking test will be used as an additional measure of cardiopulmonary performance and gait performance. This test was originally developed for patients with pulmonary disease but has also been used in neurologic populations including stroke (74) and PD (75). The test consists of walking on a track for 6 minutes at a comfortable walking speed. The distance covered in 6 minutes is the outcome measure.

c. Submaximal Effort Treadmill Economy of Gait: Economy of gait is measured using open circuit spirometry during two standard constant load submaximal effort treadmill walking tasks at pre-established gait velocities (1mph and 2mph). The mean rate of VO2 is calculated based on the final 2.5 minutes of both 5-minute walks under steady state oxygen consumption conditions, as previously described. Patients not achieving a plateau in VO2 are re-tested at a lower velocity on a different date to eliminate potential confounding effects of fatigue on testing. These tests will be performed on the same day, in order to prevent fatigue becoming an issue, between each 5 minute test, the participant will have a 10 minute rest period.

d. Over-ground Gait Economy: Over-ground gait economy will be measured using a portable metabolic monitoring system, K4b2 (Cosmed; Rome, Italy) during a 6 minute walk, with subjects walking at their comfortable self-selected walking speed while open circuit spirometry collects break-by-break data. The K4b2 consists of a small battery pack and portable gas analyzer (weighing less than 1 kg) that participants wear on their chest. Attached to the portable system is a flexible rubber facemask with flowmeter used for breath-by-breath analysis. The mean rate of VO2 will be calculated based on the final 3 minutes of a 6-minute walk under steady state oxygen consumption conditions. A 6 minute walk is a distance most representative of community-based ambulatory capacity and is a sensitive outcome measure in exercise studies in chronic stroke subjects. Prior to performing the 6 minute walk, subjects will undergo a period of acclimatization so subjects can feel comfortable wearing the rubber facemask described above. Following this familiarization period, subjects will be seated for a 10 minute rest period prior to performing the 6 minute walk. Subjects will use the same assistive device and/or orthoses typically used and will be instructed to cover as much distance as they can over a flat 100-foot walking surface demarcated by traffic cones.

Behavioral Measures:

Psychiatric and behavioral features may interfere with cognitive performance (76, 77). We will control for the potential confounding effects of depression, apathy, fatigue, and self-efficacy on cognitive performance of our patients by assessing the following measures:

a. Beck Depression Inventory:
Depression will be assessed with the Beck Depression Inventory (BDI) (78). This measure has been shown to be valid for screening and diagnosis of depression in patients with PD (79).The BDI is a self-rating questionnaire that consists of 21 statements describing how you feel over the past week. The highest score on each item is 3 and for the full test (63), indicating extreme depression.

b. Apathy Evaluation Scale:
Apathy will be assessed with the Apathy Evaluation Scale (AES). The AES has been used to investigate apathy in many diseases including PD (77). The scale has 3 versions (clinician, informant and self-rated). We will use the self-rated version (AES-S) for this study. The scale is composed of 18 items, with a 4-point response format. Subjects are asked to base their answers on the past 4 weeks. Items are rated as follows: Not at all true=1, slightly true=2, somewhat true=3, Very true=4. Some items have positive and some have negative syntax. Higher scores reflect greater apathy.

c. Fatigue Severity Scale:
Fatigue will be assessed with the Fatigue Severity Scale (FSS). The FSS has been used to investigate fatigue in PD and has been shown to be independent of severity (80). The FSS is a self-rating questionnaire composed of 9 items. Each item is a statement on fatigue (e.g. “my motivation is lower when I am fatigued) and is rated from 1, “completely disagree” to 7, “completely agree”. Scoring is the sum of responses divided by 9, where the maximum score is 7 indicating severe fatigue.

d. Self-Efficacy Scale:
Self-efficacy, the sense of personal competence to deal efficiently with stressful situations, will be assessed with the Loring Self Efficacy Scale (SES) (81). The SES is a self-rating questionnaire that evaluates 8 domains of self-efficacy (total of 26 items), including confidence in 1) obtaining medical information (1 item), 2) obtaining help (4 items), 3) communicating with physicians (3 items), 4) managing disease (5 items), 5) doing chores (3 items), 6) doing social activities (2 items), 7) managing symptoms (5 items), and 8) exercising (3 items). The original Loring scale has a 10 point response set, which has been found to be burdensome for PD patients in our clinic, so responses have been modified to a 3-point scale (ranging from 1=less confident to 3=very confident). The total score is a sum of the score on all items.

Outcome Measures specific to Aim 2: Compare the efficacy of TAEX versus TCOG versus TAEX + TCOG to improve DT performance in PD.

Dual Task Measures:
Dual Tasking (DT) assessment will use a “walking while talking” paradigm. Dual tasking will be measured under 3 conditions:

1) Single task Talking- generating words that start with the letter “s” for 60 sec while comfortably seated, subjects will be tape recorded during word generation to ensure that timing of words is captured for analysis.

2) Single task gait- subjects will walk 50 feet at a comfortable speed, on the 24 foot Gaitrite mat, turning twice to complete 50 feet total

3) Dual task- generating words that start with letter "f" while walking 50 feet on Gaitrite, subjects will be tape recorded during task performance so that rate of word production relative to walking speed can be analyzed. Start of tape recorder will be synched to start of walking.

Walking spatial-temporal parameters will be measured using the Gaitrite 24 foot gait mat with existing hardware and software for analysis.

Outcome Measures specific to Aim 3: Investigate whether improvement in DT performance translates into improved function in IADLs.

Daily activity Measures:

a. OARS:
A detailed description of the OARS was provided in section 1.4.4. Validity and reliability for the OARS has been reported in the geriatric population (83). As described previously, the OARS is composed of ADL and IADL. Higher scores represent greater disability. For the purposes of this study (Aim III), we will focus our analysis on the IADL component of the scale. One limitation of this scale is that there may be a floor effect, due to the categorical nature of the scale (see preliminary data section f). This limitation is addressed by adding a second IADL measure: the Timed IADL test.

b. Timed IADLs:
Reliability and validity of Timed IADLs has been reported in older adults(56). We have selected this testing for the following reasons: 1) it generates continuous data (time in seconds), and therefore is free of ceiling or floor effects, and 2) it has been shown to be a sensitive measure to assess improvement in speed of processing following a cognitive training intervention in older adults84. The Timed IADL involves the timing of performance of 5 tasks that resemble everyday instrumental activities of daily living: 1) finding a telephone number for a given person in the telephone directory, 2) finding and counting out correct change from a group of coins, 3) finding and reading the ingredients on a food can label, 4) finding two food items in an array of food items on a shelf and 5) searching for and then reading the directions on a medicine container. For each task there is a 2 minute time limit, with the exception of the telephone number task which has a limit of 3 minutes. If the participant does not complete the task within the time limit, the task is terminated. The following error codes are assigned for each task: (1) completed without error and within the time limit; (2) completed with minor errors, or (3) not completed within the time limit or completed with major errors. For the tasks completed with minor errors, a time penalty is added to the completion time. This added time penalty is equal to 1 SD, based upon the data from the participants who completed the same item without error. The times for each of the tasks are transformed into Z scores which are then summed to form a composite (84). Higher scores mean worse performance.

c. Activity Monitor:
A microprocessor-linked step activity monitor (SAM) (Cyma Corporation; Mountlake Terrace,

Washington) will be used to assess home ambulatory function. Previous research by our group has validated this device for patients with PD and has shown that step count as measured by this device is highly correlated with disease severity in PD (85). Subjects will wear the SAM at home over 48 hours while performing their usual daily routines (removing it during bathing and sleep). Recording will be done on the days that immediately precede or follow the on-site clinical testing of the other outcome measures. Placement of the SAM will be taught to the subject on site during their first testing session. Individual calibration will be done based on each subject's height and gait (e.g. subjects will perform a 30-foot walk while an examiner records time and number of steps without the SAM). Two variables will be measured: mean step count per day and maximal number of steps per hour.

Statistical Analysis Plan (SAP):

The data from the study will be entered into a dataset and validated by the investigators. Descriptive statistics will be calculated for all variables using SAS version 9.1 (SAS institute, Cary, N.C.). Exploratory data analyzes will be performed to check the data for extreme values and to test for normality (Kolmogorov-Smirnov test). If our data is not normally distributed, we will try conventional normalizing and/or variance stabilizing transformations to obtain something close enough to a Gaussian distribution to enable the use of parametric statistics. A two-tailed $p<0.05$ will be considered indicative of a significant result. Below is a description of the statistical analyses to test each hypothesis of this study:

Hypothesis I: The combined intervention TAEX + TCOG group will show greater improvement in EF than either group alone.
EF performance will be stratified into domains and standardized as composite scores using z-score transformation. In addition, an overall composite variable will be derived from the ANAM battery, representing an Index of Cognitive Efficiency (ICE). The ICE will be derived by weighting throughput scores from the individual ANAM tests and then combining them into a single score reflecting overall performance on the test battery. The weighting will be done so that each score contributes equally to the overall index with higher scores indicative of better test performance. This method has been used in previous studies to combine data in an efficient manner to look at cognitive performance over time (86). Our dependent variables are: 1) a total z-score for each EF domain and 2) the ICE.

Each dependent variable will be evaluated with a parametric Mixed Models Repeated Measures Analysis of Variance (ANOVA), with Group as the between factor (TAEX, TCOG, TAEX+TCOG) and Time as the within factor (pre-test, post-test). Therefore, each intervention group will be compared against each other, and against the baseline phase (i.e. those undergoing delayed entry). We will adjust for potential confounding variables such as age, educational level, and behavioral variables (depression, fatigue, apathy and self-efficacy) by including these in the analysis as covariates. Post-hoc pair-wise comparisons will be done using Tukey's Honestly Significant Difference with an alpha=0.05 as the experiment-wise error rate. An intention-to-treat analysis paradigm will be used to account for missing data (e.g. subjects from whom we have obtained data at baseline, but who later drop out). The data collected from these subjects along with those who completed the study will be combined using multiple imputation. These analyses will be performed using the SAS procedures PROC MI and PROC MIANALYZE and

will provide insight into the effectiveness of our interventions.

Hypothesis II: The combined intervention TAEX + TCOG group will show greater improvement in DT than either group alone.
Our dependent variables on DT performance are: 1) the number of words generated per second, 2) spatiotemporal gait parameters including walking speed, stride length, cycle time, and variability of cycle time, 3) Cognitive Cost and 4) Gait Cost. Each dependent variable will be evaluated with a parametric Mixed Models Repeated Measures Analysis of Variance (ANOVA), with Group as the between factor (TAEX, TCOG, TAEX+TCOG) and Time as the within factor (pre-test, post-test). Therefore, each intervention group will be compared against each other, and against the delayed entry phase. Similar to the analysis described under Hypothesis I, we will adjust for potential confounding variables, run a Tukey's Honestly Significant Difference post-hoc test, and perform an intention-to-treat analysis for Hypothesis II.

Hypothesis III: Improvement in DT performance will be associated with improvement in IADLs function.
Our IADL measures are: 1) OARS IADLs, 2) Timed IADLs and 3) daily activity as measured by the SAM. To test Hypothesis III, these measures will be converted into change scores. Therefore our dependent measures are: 1) OARS change scores (post-test – pretest), 2) Timed IADLs (post-test – pretest), 3) SAM change scores (post-test – pretest). These measures will be correlated with DT performance measures, also represented as change scores.
Correlations between OARS IADLs change scores and DT change scores will be calculated using non-parametric statistics, Spearman Correlation Analysis, because OARS produces rank-ordered data. Correlations between Timed IADLs change and DT change, as well as correlations between SAM change and DT change will be calculated using parametric statistics: Pearson Product Moment Correlation Analysis.

Secondary analyses:
In order to investigate whether EF improvements are associated with cardiopulmonary fitness improvements, we will run Pearson Product Moment Correlations. Correlated variables will be: 1) change in EF performance (z-scores and ICE) where change is calculated as pre-test – post-test, and 2) change in cardiopulmonary fitness measures (Peak VO2 max, METs, 6 min walk test time) where change is calculated as pre-test – post-test.