Validity and reliability of the 20-m run, horizontal leap, and four-bound tests measuring high-level mobility in neurologically impaired patients

Marek Gorski, Master of Education (Physical Education)*, Grant Scroggie, Master of Health Science, Terrence Haines, PhD

Monash Health, Kingston Centre, Cheltenham, Victoria, Australia

Abstract Background: High-level mobility (HLM) training including running forms an integral part of physical rehabilitation for neurologically impaired patients. Objective: This study examines the validity and reliability of three quickly administrable measures of HLM, namely, the 20-m run, horizontal leap, and four-bound tests in patients with neurological disorders. Methods: This is a retrospective data audit of 62 patients (23 women, 37.1%; 39 men, 62.9%) participating in the HLM (running retraining) task. All participants were recovering from neurological conditions such as stroke, brain injury, brain/spinal tumour, Guillain–Barré syndrome, and cerebral palsy complications. Results: High levels of test–retest reliability of the investigated tests (interclass correlation coefficient > 0.95) were obtained. The 95% minimum detectable changes were as follows: 20-m run, 1.9 seconds; horizontal leap, 0.20 m; four-bound test, 0.57 m. The area under the receiver-operated characteristic curve was 0.96 for the 20-m run, 0.90 for the horizontal leap, and 0.91 for the four-bound test, which suggests high validity of the tests to discriminate between participants who were classified as “running” and those as “not running”. Participants performing at < 7.2 seconds for the 20-m run test or ≥ 0.75 m for the horizontal leap test or 4.0 m for the four-bound test were most likely classified as running. Conclusion: The 20-m run, horizontal leap, and four-bound tests are valid and reliable objective measures of HLM when administered in people with neurological conditions.
Introduction

Physical rehabilitation for people with neurological dysfunction focuses primarily on re-educating lower-level mobility required for walking and performance of activities of daily living. Nonetheless, the importance of high-level mobility (HLM) training (including running retraining) should not be overlooked as it enhances safety, independence, and participation in recreational and social activities [1–3]. HLM, a construct that currently lacks a concrete definition in the literature, includes mobility activities beyond normal gait such as climbing, running, jogging, and jumping [4]. This aspect of rehabilitation, however, is often given limited consideration, especially in patients with neurological conditions suffering from dysfunctions such as mild ataxia, mild hemiparesis, spasticity, or decreased motor control. Many patients throughout their rehabilitation process reach the stage in which they are able to walk independently and subsequently would benefit from participation in HLM training (including running retraining) [3,4].

Fundamental to modern-day rehabilitation is the ability to reliably and accurately measure a patient’s progress [5]. Although many clinical objective measurement tools exist to quantify walking and low-level mobility, few have been developed for the quantification of HLM incorporating running ability. For a test to have a high degree of clinical utility, it should be quick, easy to administer and interpret, be portable and inexpensive, and have minimal need for specialized equipment or training [6]. Several clinical objective measures of walking ability are available, such as the 10-m walk test [7], the timed up and go test (TUG) [8], or the 6-minute walk test (6MWT) [9]. These tests are used widely to quantify walking ability [6], but the specific physical demands associated with running, such as shock absorption, control of vertical collapse in weight acceptance, and energy generation associated with forward and upward propulsion [10,11], necessitate the use of different tests.

Although different assessment tools have been developed to assess HLM, their practical application is often limited in the clinical environment. The Community Balance and Mobility Scale [12] and the Rivermead Mobility Index [13] are two tools with adequate psychometric properties [6,14] that quantify HLM and contain items pertaining to running ability. However, these contain 13 and 15 items, respectively, and thus, administering, analysing, and interpreting their measures are time consuming, which consequently reduces their clinical utility. It takes 10–30 minutes to administer, analyse, and interpret the Rivermead Mobility Index and >30 minutes for the Community Balance and Mobility Scale [6]. The High-Level Mobility Assessment Tool (HiMAT) [15] is another tool that exhibits robust psychometric properties [6,15,16]; however, it still contains eight items for testing and takes >30 minutes to administer, analyse, and interpret the results [6].

The aims of this study were to examine the validity and reliability of three quickly administrable clinical objective measures of HLM, namely, the 20-m run, horizontal leap, and four-bound tests. These measures were derived from tests used extensively in athletics [17–19] due to their ability to specifically address the unique and specific requirements of running mentioned earlier. The tests can be administered, analysed, and interpreted in <10 minutes.

The objectives of this study were to determine the test–retest reliability of the 20-m run, horizontal leap, and four-bound tests; to evaluate the discriminant validity of each investigated new test by examining video footage to see how closely the investigated new tests correspond to whether the patient was classified as “running” at the assessment; to examine the concurrent validity of the investigated new tests by comparing the score for each test with the HiMAT, 6MWT, step test, and TUG test (validated tests).

Methods

Design

The study examined the reliability and validity of the 20-m run, horizontal leap, and four-bound tests through a retrospective audit of routinely collected data (investigated new tests and validated existing tests) of patients participating in the HLM (running retraining) task conducted at a rehabilitation hospital and rehabilitation outpatient’s clinic between 2003 and 2012.

Ethics

Ethical approval was sought from the Research Directorate of Monash Health, Melbourne, Australia, and the project was approved on February 23, 2012. The Research Directorate deemed this project a quality-assurance exercise involving collection, use, and disclosure of data in a deidentified format. It did not raise ethical concerns and did not fall within the category of a research project within the National Statement on Ethical Conduct in Human Research (NHMRC 2007) and World Medical Association-Declaration of Helsinki: ethical principles for medical research involving human participants. This project thus did not require submission to the Human Research Ethics Committee. The HLM retraining is routinely offered to patients admitted to the rehabilitation facilities at Monash Health. An informed written consent was obtained on admission from all patients admitted to the rehabilitation facilities at Monash Health.

Participants and setting

Participants were recovering from neurological conditions such as stroke, hypoxic brain injury, brain/spinal tumour, Guillain–Barré syndrome, and cerebral palsy complications. They were offered the opportunity to participate in the HLM task (HLM group) when they achieved independent walking and consented to the program. The primary purpose of the task was to teach participants how to run. The exercise programs were individually designed for each participant to address his/her limitations pertaining to the specific demands of running such as shock absorption, control of vertical collapse in weight acceptance, and energy generation associated with forward and upward propulsion. The transition from walking to running occurs when
periods of double limb support during the stance phase of gait (both feet on the ground) give way to two periods of double float [10].

A total of 62 (23 female, 39 male) participants were included in the HLM group between 2003 and 2012 (55 patients at the rehabilitation hospital and 7 patients at the outpatients rehabilitation clinic of Monash Health). The assessors were senior exercise physiologists and physiotherapists who routinely conducted the HLM sessions at Monash Health.

Procedure

The participants were tested (investigated new tests) on their first attendance to the HLM session. If weather conditions were unsuitable and the 20-m run test could not be safely conducted, participants were tested on their second or third attendance. The participants were retested on a regular basis (every 6 weeks) and video footage of running was taken each time. All tests were conducted after 5-minute “warm-up” exercises, which engaged major muscle groups of the lower and upper limbs but did not incorporate components involved in the testing. Two of the investigators (M.G. and G.S.) administered these tests to all 55 participants at the rehabilitation hospital over the entire 9-year period. A third therapist administered these tests to seven participants from the outpatients’ rehabilitation clinic.

Each patient’s testing sequence commenced with the horizontal leap test, followed by the four-bound test and the 20-m run test. Each test was repeated three times and scores were documented. Each participant completed all three repetitions of a given test before attempting the next test. The four-bound and horizontal leap tests had a recovery period of approximately 30 seconds between test repetitions. The 20-m run test had a recovery period of approximately 1 minute between test repetitions. The recovery time was the approximate time it took the participant to return to the start line after performing the test. The tests were conducted on a flattened surface outdoors or in the indoor physiotherapy treatment area if the weather conditions outdoors were not suitable. However, the 20-m run test was only conducted outdoors as no indoor location was suitable.

Data regarding patient’s demographics (age, sex, diagnosis, other comorbidities, previous participation in sports and recreation activities, and dates of admission and discharge), HIMAT, 6MWT, step test, and TUG scores (validated tests) were extracted in 2012 from medical records for discharge, HiMAT, 6MWT, step test, and TUG scores (validated tests) were extracted in 2012 from medical records for 55 participants at the rehabilitation hospital over the entire 9-year period. A third therapist administered these tests to seven participants from the outpatients’ rehabilitation clinic.

20-m run test, horizontal leap test, and four-bound test

The 20-m run test involved a patient being timed running (or moving if unable to run) between two markers placed 20 m apart with 3-m markers on either end (outside the 20 m) to allow for speeding up and slowing down. The participant was instructed to run (or move if unable to run) at their highest speed on the command “go”. If there were safety concerns with a participant, a therapist would run alongside them. The participants were asked “are you ready?” while standing at the start marker, followed by the command “go”. Participants were asked to continue running (or moving if unable to run) until they reached the final marker 26 m away. An assessor at the start line would drop his arm when the person crossed the start line to indicate for the other assessor at the finish line with a stopwatch to commence the stopwatch. If only one assessor was involved, the assessor would run alongside the participant and start/stop a stopwatch when the participant crosses the start/finish lines. Participants were encouraged while performing the 20-m run test (e.g., “go faster” or “almost there”). They were also given positive feedback upon completion of each test repetition (e.g., “you did well!”). When conducting this test, the investigators found that at least 30 m of running space in total is required to allow participants to slow to a stop at the conclusion of this test.

In the horizontal leap test, the participants were asked to jump forward from a standing position with both feet as far as they could. The distance was measured from the start line to the heel position of the closest leg to the start line at the landing point. If the participant moved forward after landing, the assessor used his/her visual estimate of the landing position for the measurement. If the participant moved backward after landing, the test was declared invalid and only the remaining valid repetitions were scored.

In the four-bound test, the participants stood with their toes immediately behind a start line. They were asked to bound (leap from one leg and land on the other) four times in succession. The total distance of the four bounds was measured from the start line to the heel position at the completion of the bounds. If the participant moved forward after landing, the assessor used his/her visual estimate of the landing position for the measurement. If the participant moved backward after landing, the test was declared invalid and only the remaining valid repetitions were scored.

Validated tests

The HiMAT examines walking, walking backward, walking on toes, walking over an obstacle, running, skipping, hopping, bounding, and walking up and down stairs. The test has demonstrated reliability and validity in traumatic brain injury patients [15].

The step test evaluates speed of performing a dynamic single-limb stance task. It involves stepping one foot on and off a block as quickly as possible in a set time (15 seconds). The test exhibits good reliability and validity for healthy older participants and stroke patients [20].

The 6MWT measures endurance of gait. It exhibits excellent retest reliability in community-dwelling stroke patients [interclass correlation coefficient (ICC) = 0.99] [21] and validity, showing that mobility training poststroke has a positive impact on 6-minute walk distance [22,23].
Significant correlation exists between distance on 6MWT and VO₂ peak as a percentage of VO₂max in subacute stroke patients \( r = 0.84 \) [24]. The TUG test measures speed during several functionally important tasks, which potentially threaten balance [8]. The patient was observed and timed while she/he rose from an armchair, walked 3 m, turned, walked back, and sat down again. Very high inter-rater reliability (ICC = 0.99) has been shown in a sample of day hospital patients [8] along with responsiveness to change in acute stroke rehabilitation [25].

These existing tests are routinely conducted as standard procedures by the treating physiotherapists at Monash Health.

Classification of running

The video footage was used for the classification of running versus not running. Video was captured in both sagittal and coronal plane views. The video footage was assessed by an independent assessor (physiotherapist) who was blinded to all test scores of participants. This assessor was asked to classify whether the patient was running or not. The assessor was allowed to use slow motion and forward or rewind the video to make the determination. Running was defined to occur when periods of double limb support during the stance phase of gait (both feet on the ground) give way to two periods of double float [10]. The participant had to have only two periods of double float to be classified as running. The results provided by the independent assessor were reviewed by one of the study authors (M.G.). An agreement was found in 60 of the 62 cases. A discussion was held with a third independent assessor to reach consensus on the other two cases.

Analysis

The intraclass correlation coefficient \( (\text{ICC}_{3,1}) \) was used to examine the test—retest reliability of each investigated new test. The standard error of measurement \( \text{SEM} \) is calculated as the square root of 1 minus the ICC multiplied by the standard deviation of the scores. The minimal detectable change at 95% confidence; MDC equals 1.96 \( \times \) SEM \( \times \) square root of 2) [26] were calculated.

The discriminant validity of each investigated new test was evaluated by seeing how closely the test results corresponded to whether the patient was classified as “running” at the time of this assessment. For this purpose, the video footage taken during the assessment of each patient was reviewed and scored (running vs. not running) by an assessor blinded to the scores of the tests. The investigators treated each assessment as an individual data point ignoring dependence between admission and discharge scores within individual. The sensitivity and specificity of each test in predicting the running versus not-running classification by the physiotherapist were calculated along with the area under the receiver operating characteristic curve (ROC). The sensitivity of a diagnostic test is the proportion of participants for whom the outcome is positive (running) that is correctly identified by the test. The specificity is the proportion of participants for whom the outcome is negative (not running) that is correctly identified by the test. The test score cut-off points at which the sum of sensitivity and specificity was maximized were also identified for each investigated new test by examining the Youden Index (which equals sensitivity + specificity minus 1) [27]. Examination of the Youden Index allowed the investigators to identify the scores at which the participants were definitely running as well as the scores at which the participants were definitely not running. The maximized Youden Index allowed the investigators to identify the scores at which the participants were most likely to be classified as running. All available test occasions for each individual were used.

The concurrent validity of the three new tests was examined by correlating the scores with the validated tests using pairwise Pearson \( r \) calculations. All analyses were undertaken using available patients’ data. Dependency between admission and discharge assessments within an individual was ignored for the purpose of these analyses. The reliability estimates were calculated from within-session test—retest testing. The scores for the different trials of each participant were averaged for calculation of concurrent and discriminant validity. Each test analysis was conducted using the overall sample and then again using only scores of participants with diagnosis of stroke.

Results

Demographic details of participants are presented in Table 1. The majority of participants were men, and the most common diagnosis was stroke. The outcomes from analysis of test—retest reliability are presented in Table 2. In these results, ICC > 0.95 indicated high levels of test—retest reliability for each investigated new test (ICC > 0.90, high; 0.90 > ICC > 0.75, good; ICC < 0.75, poor to moderate) [28]. The ability of each test to determine which participants would be classified as running by independent video-based assessment was excellent with all tests displaying sensitivity and specificity between 80% and 90% for cut-off scores at the maximal Youden Index. Patients performing at < 7.2 seconds for the 20-m run or ≥ 0.75 m for the horizontal leap or 4.0 m for the four-bounds test were most likely to be classified as running. In the stroke subgroup analysis, similarly high predictive accuracy was demonstrated for all tests; however, the horizontal leap test showed a slightly reduced specificity at the maximized Youden Index (71% for stroke subgroup versus 84% for all participants).

The three investigated new tests were used for the video classification of whether the participant was running or not to identify a test cut-off point that maximized overall classification accuracy according to the Youden Index. These cut-off points are presented in Table 3.

The maximized Youden Index for the 20-m run was 0.76 compared with 0.64 for the horizontal leap and 0.67 for the four-bound tests. The area under the ROC curve was 0.96 for the 20-m run, 0.90 for the horizontal leap, and 0.91 for the four-bound test, suggesting high validity of the tests to
discriminate between those participants who were classified as running or not running (1.0 is a perfect prediction and 0.5 is a random chance) [28].

Correlations between the three new tests and the validated tests are presented in Table 4. High correlations were identified between the three investigated new tests ($r > 0.76$) and between the HiMAT and the four-bound test. Medium correlations ($0.51 < r < 0.75$) were identified between the HiMAT, the horizontal leap, and 20-m run tests. There is also a medium correlation between the TUG and 20-m run tests. A low degree of relationship ($0.25 < r < 0.50$) was observed between the 6MWT and the three new tests. The step test appeared to be poorly correlated with both the 20-m run test and the HiMAT. A post hoc power analysis indicated that 62 participants provided 80% power to identify significant correlations (alpha $\alpha = 0.05$) as low as $r = 0.35$. For the stroke subgroup analysis, a sample size of 27 provided 80% power to detect significant correlations as low as $r = 0.52$.

**Discussion**

The three new HLM tests appear to be valid and reliable for use in clinical practice. The ability of these tests to determine which participants would be classified as running by independent video-based assessment was high for all participants in the sample and the stroke subgroup. All three tests were highly correlated within themselves and had high-to-medium correlation with the HiMAT.

The test—retest reliability was sufficiently high to allow development of relatively narrow MDC95 thresholds for each test. Scores at or above the MDC95 threshold can be interpreted as improvement on the test rather than a measurement error. No previous research has investigated MDC95 thresholds for these investigated new tests, and therefore comparisons could not be made.

Several predictive tests of running ability were examined by Williams and Goldie [29] in a traumatic brain injury population. Walking on toes, stepping backward up a step with affected leg, timed single-leg standing, and bounding forward onto a single leg were all found to be moderate-to-strong predictors; however, the presence of nonsupport phase during the single bounding task was found to be the strongest predictor of running ability. Although the presence of a nonsupport phase during the bounding task was not specifically examined in this study, the ability of the four-bound test to accurately determine the ability to run was strong. The high correlation between the four-bound test and the HiMAT ($r = 0.85$;...
suggested that the four-bound test may be an excellent clinical tool of HLM. This strong correlation remained in the stroke subgroup analysis ($r = 0.88; p < 0.001$).

Interestingly, the step test exhibited only medium level correlation with the horizontal leap and four-bound tests and was very poorly correlated with the 20-m run test (Table 4). This indicated that poor performance on the step test, a dynamic single-limb stance balance task, did not prevent a participant from achieving a good score on a HLM test and vice versa. Although running required the sequential action of moving from one single-limb stance to another, the ability to maintain single-limb standing balance did not appear to correlate with the ability to run in this population.

With the exception of the 20-m run test, which exhibited medium correlations with TUG, relatively weak correlations existed between each of the three new HLM tests, the 6MWT, and TUG, suggesting that the new tests were in fact measuring different constructs. The specific physical demands of running (discussed previously) assessed by the three new tests were likely to be less relevant for the dynamic balance requirements imposed by the TUG and the endurance requirements of the 6MWT.

### Limitations

The study examined the reliability and validity of the 20-m run, horizontal leap, and four-bound tests through a retrospective audit of routinely collected data. The data for the three new tests were collected by the investigators (M.G. and G.S.) using standardized procedures. Data for the validated tests, however, were gathered retrospectively from patient records. These data were collected routinely by other therapists in the usual course of therapy following standard procedures published with these tests but not with a specific standardized data-collection protocol for this research. As such, it is possible that some variability may have existed in testing procedures, instructions, and/or the equipment used. The horizontal leap and four-bound tests (2 of the investigated new tests) were conducted both indoors (floorboards) and outdoors (hardened ground with grassy patches). It is possible that results of these tests could vary due to the different testing conditions. The patients were offered to participate in the HLM task when they were able to walk independently and able to provide consent. No standardized tests were applied to determine eligibility of the patient to participate in the HLM task. Not separating clients according to age, diagnostic category, or severity allowed a larger sample size for pooled analysis, which may improve generalizability. However, patients with different neurological conditions may yield different results. Our stroke sub-group analysis indeed found that the specificity was lower for the horizontal leap test at the maximal Youden Index when compared with the overall analysis. However, sensitivity, specificity, and correlations remained high across all other areas during this subgroup analysis. Further investigation into age and/or severity may be needed. Because this is the first study to investigate these clinical tests of HLM, the ability to draw comparisons with other research is limited.

### Conclusion

Therapists are offered many choices for the objective evaluation of walking and lower level mobility activities. However, in cases where HLM is concerned, these choices become more restricted and the currently available measurement instruments are often scales with multiple items to assess, and thus are lengthy to administer. Results of this research suggested that the three new tests (20-m run test, horizontal leap, and four-bound tests) are reliable and valid alternatives to the current measurement instruments. Further investigation into the specific characteristics of running and the relationship with balance and endurance is recommended to better understand the clinical utility of these tests.
Table 4  Pearson correlation matrix across the investigated new tests for high-level mobility.

<table>
<thead>
<tr>
<th></th>
<th>20-m run test (average)</th>
<th>Horizontal leap test (average)</th>
<th>Four-bound test (average)</th>
<th>6MWT</th>
<th>HiMAT</th>
<th>Step test (R)</th>
<th>Step test (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal leap test (average)</td>
<td>-0.755 *</td>
<td>0.913 *</td>
<td>0.485 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-bound test (average)</td>
<td>-0.806 *</td>
<td>0.440 **</td>
<td>0.854 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6MWT</td>
<td>-0.458 **</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HiMAT</td>
<td>-0.624 **</td>
<td>0.715 *</td>
<td>0.585 **</td>
<td>0.289</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step test (R)</td>
<td>-0.398</td>
<td>0.675 *</td>
<td>0.625 *</td>
<td>0.106</td>
<td>0.933 *</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step test (L)</td>
<td>-0.349</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>0.710 *</td>
<td>-0.562 **</td>
<td>-0.481 ***</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stroke subgroup</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Horizontal leap test (average)</td>
<td>-0.666 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Four-bound test (average)</td>
<td>-0.730 *</td>
<td>0.938 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6MWT</td>
<td>-0.590 **</td>
<td>0.777 *</td>
<td>0.807 *</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HiMAT</td>
<td>-0.501</td>
<td>0.644 **</td>
<td>0.878 *</td>
<td>0.850</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step test (R)</td>
<td>-0.560</td>
<td>0.870 *</td>
<td>0.782 **</td>
<td>0.839</td>
<td>0.513</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Step test (L)</td>
<td>-0.384</td>
<td>0.788 **</td>
<td>0.611 ***</td>
<td>0.862</td>
<td>0.147</td>
<td>0.893 *</td>
<td></td>
</tr>
<tr>
<td>TUG</td>
<td>0.886 **</td>
<td>-0.703 ***</td>
<td>-0.555</td>
<td>-0.905</td>
<td></td>
<td>-0.422</td>
<td>-0.884 **</td>
</tr>
</tbody>
</table>

* $p < 0.001$.  
** $p < 0.01$.  
*** $p < 0.05$.  

6MWT = 6-minute walk test; HiMAT = high-level mobility assessment tool; TUG = timed up and go test.
horizontal leap test, and four-bound test) may be used as valid, reliable, quickly administered objective measures of HLM (including running ability) in neurologically impaired patients.

Conflicts of interest

None.

Funding/support

The researchers would like to acknowledge the financial support they received from the Lions Australia John Cockayne Fellowship Trust Fund (grant number 12053 Q), Melbourne, Australia.

References