ORIGINAL ARTICLE

Test-retest Reliability of Unilateral and Bilateral Carrying Protocol Based on Joule’s Functional Capacity Evaluation

Hanif Farhan Mohd Rasdi1, Nurul Syahirah Md Senin2

1 Centre Rehabilitation Sciences and Special Needs (iCaRehab), Occupational Therapy Programme, Faculty of Health Sciences, Universiti Kebangsaan Malaysia, 50300 Jalan Raja Muda Abdul Aziz, Kuala Lumpur, Malaysia
2 Flying Kids Occupational Therapy, No 6, Jalan horizon Perdana 1, Horizon Hills, 79150 Iskandar Puteri, Johor, Malaysia.

ABSTRACT

Introduction: Load carriage is one of the most frequent manual material handlings in industrial settings. However, due to the biomechanical risk exposures, carrying could lead to work-related musculoskeletal disorders. Joule’s Functional Capacity Evaluation is an FCE system used widely in Malaysia in return-to-work programs to ensure work readiness and prevent re-injury. However, the literature regarding the test-retest reliability of the carrying protocols is minimal, leading to a questionable level of consistency. Therefore, this study aimed to investigate the test-retest reliability of the carrying protocols in Joule’s FCE.

Methods: A cross-sectional study was conducted to evaluate test-retest reliability where the carrying protocol of Joule’s Functional Capacity Evaluation was used among healthy university students (N=30). The participants were asked to perform a retest of the carrying protocol after one week.

Results: The ICC values for the carrying protocols were good for dominant unilateral carrying was (ICC: 0.82; 95% CI: -6.00 to 5.81), non-dominant carrying was (ICC: 0.74; 95% CI: -6.78 to 6.02) and bilateral carrying was (ICC: 0.85; 95% CI: -6.26 to 5.51). Bland and Altman’s plot indicated no visible distribution patterns of the differences without exceeding the 95% limits of agreement. The standard error of measurements (SEMs) was relatively small for all carrying protocols.

Conclusion: The test-retest reliability of carrying protocols in Joule’s FCE was good. Future studies are needed to replicate this study in real patients to further verify the reliability for clinical settings.

Keywords: Return-to-work, Load carriage, Work readiness, Ergonomics, Occupational therapy

INTRODUCTION

Any accidents that cause injury, illness, or death in the workplace are considered occupational injuries (1). According to the literature, manual handling activities such as carrying and lifting are the main risk factors for waist soreness and various back disorders (2, 3). According to the Fourth European Working Conditions Survey (2005) carried out in the EU-27, 35% of all workers were exposed to carrying or moving heavy loads for at least a quarter of their working time (4). The factors that make manual handling hazardous are the load (e.g., too heavy, oversized, challenging to reach or grasp, unbalanced) and the task (e.g., too strenuous, involves awkward postures or movements) (5, 6).

The Social Security Organization (SOCSO) is a statutory body representing the employment accident insurance scheme and the invalidity pension scheme. This organization introduced a Return-to-Work (RTW) program in 2007, a comprehensive rehabilitation program for its insured persons with a disability due to work injury (7). When a worker has recovered from injury and is permitted to come back to work, this organization needs to assess the workers’ level of work readiness. Therefore, a Functional Capacity Evaluation report is usually required to determine workers’ physical capacity and disability (8).

Functional Capacity Evaluation (FCE) is a measurement tool that assists in determining safe, tolerable levels of function and predicting when an individual is ready to return to work duties (9). In addition, it also can provide a comprehensive assessment in determining the level of functionality of an individual in their work (10). Rehabilitation professionals frequently use performance-
The FCE is considered the gold-standard assessment in a return-to-work program (11). The FCE has been practiced for years by occupational therapists, physiotherapists, vocational assessors and psychologists (9, 12). However, to the researcher’s knowledge, the measurement properties, such as the validity and reliability of some of the FCE systems, are still limited. Generally, reliability can be defined as how an assessment tool can produce a stable and consistent decision (13). If a measurement can be replicated consistently within two or more time intervals, the tool is said to have good test-retest reliability (14). On the other hand, inaccurate FCE can increase the risk of re-injury. Therefore, professionals involved in work rehabilitation need to assess their patients’ physical capabilities before returning to their previous work (15). According to the literature, the FCE had been criticized for lacking legitimacy, credibility and questionable outcomes unrelated to continuous work (16-18).

Thus, it is essential to determine and report the validity and reliability of the FCE to allow accurate clinical decision-making. For instance, in the RTW program established by the SOCSO, the FCE report is critical to determine the ability of a client to return to work and for legal documentation purposes. A study has shown that intra-rater reliability between the assessors for Joule’s FCE is high in healthy adults (10), but no study investigating the reliability in Malaysia. Therefore, the purpose of this study was to examine the test-retest reliability of the carrying protocol in Joule’s FCE.

MATERIALS AND METHODS

Study design
This cross-sectional study was conducted on healthy university students recruited using convenience sampling (n=33). The initial sample consisted of 33 participants: 16 males and 17 females. However, only 30 participants performed the retest (13 males and 17 females). All declared to be healthy, i.e., having no medical condition that would restrict them from performing maximally. The data were collected within three months.

Sample size
The sample size determination for this study was conducted based on the power analysis method. The power set for this study was 80%, while the alpha value was set at 5%. The effect size used for this sample size calculation was based on the value of the Intra-class Correlation Coefficient (ICC) of 0.87 (19, 20). After considering 10% of dropout rates, 30 participants were required in this study.

Instruments
This study used two instruments, namely the Joule’s Functional Capacity Evaluation (FCE) (Valpar International Corporation) and the International Physical Activity Questionnaire Short Form (IPAQ-SF). Joule’s FCE was selected due to the limited literature about this FCE. This instrument can be used to assess the ability of a person to function in manual handling activities and demonstrated high intra-rater reliability (ICC > 0.90) for the Joule’s FCE in healthy adults (10). The IPAQ-SF was used to determine the level of physical activity in the last seven days (i.e., inactive, moderately active or active). This questionnaire was selected due to the ease of assessing participants’ physical activity over the last 7-days. The IPAQ-SF consisted of seven items, including open-ended questions about the physical activity of the previous seven days. It is a self-assessment report, paper-pencil version or verbally. The correlation coefficients of the IPAQ-SF show ranged from 0.16 to 0.35 indicating moderate criterion validity, while (ICC = 0.71–0.89) indicated acceptable reliability (21, 22).

Procedures
The main parameters of this study are the carrying load limit for unilateral and bilateral carrying protocols in Joule’s FCE. The carrying protocols were conducted based on the standardized methods as stated in the Joule FCE module. The participants were introduced to the general FCE procedures for carrying protocol and then signed informed consent. Then, they were asked to fill in the IPAQ-SF and performed the carrying protocol with progressive loads (i.e., 1.5 kg increment). The loads were placed into standard containers with dividers. The load arrangement in the dividers was based on the Joule’s load progression flipchart. The participants were briefly instructed on the required performance. The evaluator demonstrated each test of carrying protocol (i.e., unilateral and bilateral carrying). After a one week, carrying protocols were retested. For the unilateral carrying activity, the participants lifted a container using the dominant hand, carried it from point A to point B (7 meters), and returned to point A. After that, the participants changed to their non-dominant hands, took the load from point A to point B, and returned to point A. For the bilateral carrying activity, the participants lifted the container using both hands, carrying the load from point A to point B and returning to point A. Finally, each test was terminated for three reasons: 1) the participants wished to do so, 2) heart rate exceeded 85% age-related target heart rate, and 3) unsafe action, defined as a situation in which the participants were not in complete control of him or herself or the load. The evaluator recorded the time and safe maximum carrying load limit directly after each test.
Data Analysis
The data were analyzed using IBM SPSS Statistics version 22. The sociodemographic characteristics and level of physical activity were analyzed using descriptive analysis. The Intra-class Correlation Coefficient (ICC) was used to determine the level of agreement between the measurements. An ICC greater than 0.7 was considered good, and an ICC over 0.9 was excellent (23). The Bland-Altman (B&A) plot was used to determine the 95% limit of agreement between the measurements. Bland and Altman introduced a method to describe the limits of agreement between two quantitative measurements (24). They had established a way to quantify the agreement between two quantitative measures by constructing limits of agreement. These statistical limits are calculated using the mean and the standard deviation (s) of the differences between the two measurements. Bland and Altman recommended that 95% of the data points lie within ± 2s of the mean difference (26). Finally, the following formula was used to determine the standard error of measurement (SEM), where the SD (diff) is the standard deviation of the difference between the measurements, and the ICC is the value of the Intraclass Correlation Coefficient:

\[
\text{SEM} = \sqrt{\frac{1}{N} - \text{ICC}^2} \times \text{SD (diff)}
\]

Ethical Clearance
The ethical clearance was obtained from the UKM Secretariat of Ethics (ethics numbers: NN-2017-187).

RESULT
Thirty-three participants met the inclusion criteria, and all returned the consent forms. However, only 30 participants return on retest (13 males & 17 females). The descriptions of the 30 participants are presented in Table I. The means, standard deviations, ICCs and standard error of measurement (SEM) are shown in Table 2. The ICC values of the dominant hand, non-dominant hand and bilateral carrying task were 0.82, 0.74 and 0.85, respectively. The standard error of measurement (SEM) was relatively minor for all types of carrying (SEM dominant was 1.27kg; SEM non-dominant = 1.678; SEM bilateral = 1.15) (Table II).

Bland and Altman’s plot indicated that the dominant hand’s carrying had an outlier beyond the upper limit and the lower limit for the limits of agreement. For the non-dominant hand, Bland and Altman’s plot indicated two outliers above the upper limit, while bilateral carrying had one outlier beyond the lower limit. The differences show a heteroscedasticity pattern for dominant unilateral. However, non-dominant unilateral and bilateral carrying shows a linear pattern of differences. In addition, the mean differences for dominant and non-dominant unilateral and bilateral carrying were -0.0933, -0.3833 and -0.3767, respectively, which are relatively close to zero, indicating a good level of agreement.

### Table I: Participants’ characteristics (N=30)

<table>
<thead>
<tr>
<th>Gender</th>
<th>Freq. (%)</th>
<th>Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>17 (57%)</td>
<td>22.60 (1.248)</td>
</tr>
<tr>
<td>Male</td>
<td>13 (43%)</td>
<td></td>
</tr>
<tr>
<td>Race</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malay</td>
<td>16 (53%)</td>
<td></td>
</tr>
<tr>
<td>Chinese</td>
<td>12 (40%)</td>
<td></td>
</tr>
<tr>
<td>Indian</td>
<td>2 (7%)</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>2 (7%)</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>6 (20%)</td>
<td>22.60 (1.248)</td>
</tr>
<tr>
<td>23</td>
<td>17 (57%)</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>4 (13%)</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>1 (3%)</td>
<td></td>
</tr>
<tr>
<td>First IPAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>8 (27%)</td>
<td>1.93 (0.6911)</td>
</tr>
<tr>
<td>Minimal active</td>
<td>16 (53%)</td>
<td></td>
</tr>
<tr>
<td>HEPA active</td>
<td>6 (20%)</td>
<td></td>
</tr>
<tr>
<td>Second IPAQ</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inactive</td>
<td>3 (10%)</td>
<td>2.10 (0.548)</td>
</tr>
<tr>
<td>Minimal active</td>
<td>21 (70%)</td>
<td></td>
</tr>
<tr>
<td>HEPA active</td>
<td>6 (20%)</td>
<td></td>
</tr>
</tbody>
</table>

### Table II: Test-retest reliability of carrying task Joule’s Functional Capacity Evaluation

<table>
<thead>
<tr>
<th>Type of carrying</th>
<th>ICC</th>
<th>95% CI Lower</th>
<th>95% CI Upper</th>
<th>SEM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dominant-hand unilateral carrying</td>
<td>0.82</td>
<td>-6.00</td>
<td>5.81</td>
<td>1.27</td>
</tr>
<tr>
<td>Non-dominant hand unilateral carrying</td>
<td>0.74</td>
<td>-6.78</td>
<td>6.02</td>
<td>1.68</td>
</tr>
<tr>
<td>Bilateral carrying</td>
<td>0.85</td>
<td>-6.26</td>
<td>5.51</td>
<td>1.15</td>
</tr>
</tbody>
</table>

*ICC = Intra-Class Correlation Coefficient.

### DISCUSSION
The findings from this study indicated that the test-retest reliability of the carrying protocol in Joule’s FCE was acceptable, indicating that the carrying protocols are consistent from one testing occasion to another. In addition, the findings were consistent with a previous study that reported good test-retest reliability of carrying task (25). For unilateral carrying, the dominant hand indicated better reliability than non-dominant unilateral carrying, indicating a slight difference in reliability between both unilateral carrying protocols. Reneman, Brouwer (27) also reported differences in the reliability score between unilateral carrying.
Furthermore, when comparing between types of carrying, bilateral load carriage showed a higher level of agreement than both unilateral carrying protocols. This finding may suggest that the bilateral carrying was relatively stable compared with unilateral carrying. Carrying loaded bilaterally will displace one's centre of gravity posteriorly. Hence, increasing an individual's pelvic tilt or forward lean will keep them in an upright, vertical position (28-30). Even if the total load weight doubles during bilateral carrying, it does not increase the spine load (31). However, when observing the biomechanics of unilateral carrying, the trunk bends towards the contralateral side (unload side) to compensate for the perturbation from the external load (32). This action causes higher forces toward the spinal while walking, leading to increased muscular activity and more significant spinal shear and compressive force (30, 33). Furthermore, continuously carrying a load on one side of the body can lead to muscle fatigue due to increased muscle use on opposite sides since carrying one hand results in high spine compression when bilateral carrying (34). This fatigue phenomenon can affect the consistency of the result and then influence the reliability of this unilateral carrying task (35). Nevertheless, bilateral carrying becomes the most demanding task compared with other carrying on the worksite (36). Hence, it is vital to ensure this carrying is stable enough as an ergonomic intervention to prevent injury on the worksite.

According to the limits of agreement, most of the differences were within the upper and lower limit of agreement. However, there were some outliers in both types of carrying protocols. Two participants (outliers) in the dominant and non-dominant unilateral carrying, and one participant scored outside of the 95% limits of agreement. Based on the level of physical activity before the test and retest, all the participants had no changes in their physical activity level. However, slight differences in Metabolic Equivalents (METs) were indicated, either increase or decrease. Dodds, Kuh (37) stated that increasing physical activity benefits grip strength, while Stenholm, Tiainen (38) stated becoming sedentary was associated with a decline in muscle strength. Hence, this evidence correlated with participants carrying a heavier load on retest when increasing METs before the retest. However, this situation is not associated with one participant (ID=29) who had a heavier load while decreasing METs before the retest.

Pre-existing fatigue during the retest session can also explain the differences between carrying activities. It isn't easy to maintain physical activity as students since they have to participate in numerous co-curricular activities. Therefore, prolonged physical activity between the tests may lead to fatigue. A previous study stated intensive training of muscles causes a decline in performance (35). For instance, Ahmed (39) suggested avoiding upper extremity fatigue among basketball players since it may reduce grip strength and affect performance. Fatigue has two types which are central and peripheral fatigue. During the prolonged exercise of moderate intensity, a decrease in the blood glucose level due to depletion of the liver glycogen stores is one factor known to affect the central nervous system and cause fatigue (40). Another factor that can cause central fatigue during dynamic exercise is the increased neurotransmitter release in the brain (41-43). However, such adaptations may contribute to the delay of central fatigue during sustained exercise. This fatigue indicates a reduction in the contractile strength of the muscle fibres and changes in the mechanisms underlying the transmission of muscle action potentials (44).

This study’s standard error of measurements (SEMs) was relatively small for all carrying protocols, suggesting a clinically tolerable measurement error for the evaluation. Bilateral carrying indicated the smallest SEM rather than unilateral carrying, indicating better quality, accuracy, and reliability (45). Even though the SEM of unilateral carrying was slightly higher than bilateral carrying, it can still be acceptable since all measurements contain some error. No similar SEM data for FCE carrying tests have been previously reported. Thus, the SEM’s cut-off value for clinical tolerance cannot be retrieved from literature and should be based on the practitioner’s knowledge of the carrying tests. However, compared with the SEM for other functional assessments, the study also shows relatively small SEM within 0.19 to 0.66 with high reliability of measurements (46). In addition, the SEM lifting tests varied between 1.9 and 8.6 kg, and the SEM for carrying lifting strength test was 3.4 (47). Thus, this study indicated a sufficient level of agreement with the lifting tests.

To our knowledge, this is the first study that has investigated the test-retest reliability of the carrying protocol of Joule’s FCE in Malaysia. This study has found that the carrying protocol of Joule’s FCE has good test-retest reliability. Therefore, it is essential to ensure this instrument is safe, reliable, valid, practical, and utility (12, 43). However, some difficulties were faced in carrying out this study, such as setting up participants’ appointments for their retest sessions. In addition, most of them were university students with busy schedules due to classes and co-curricular activities. In addition, the time to access the instrument was limited as it was located in a teaching clinic. Therefore, it can only be accessed when no patients or classes are being held in the room. However, despite these challenges, an adequate sample size was successfully achieved.

Although the number of participants in this study achieved a sample size considered adequate for a fair quality study, a larger sample size would have improved the quality of the study results (48). Besides, this study was conducted among healthy university students, which can pose different effects on retest due to several
factors such as differences in lifestyles. However, the influence of physical activity on the carrying load limit was not investigated in detail throughout the study due to the statistical limitations. Therefore, future studies are needed to improve the statistical power and replicate this study in clinical situations to verify the reliability further.

CONCLUSION

The primary purpose of this research was to investigate the test-retest reliability for the carrying protocol of Joule’s FCE among healthy adults. The findings show the carrying protocol in Joule’s FCE was good test-retest reliability. Since the carrying protocol in Joule’s FCE is reliable, it can help clinicians make decisions in clinical settings to prevent re-injury among the workers. As this study was conducted on healthy participants, future studies are needed to replicate this study in clinical situations such as work-related musculoskeletal disorders (WRMD) to verify the reliability.

ACKNOWLEDGEMENTS

Much appreciation is given to Mr. Lim Wen Xing for helping in the analyzed data collection and member checking process. Many thanks to all participants, students at the Universiti Kebangsaan Malaysia (Kuala Lumpur Campus), for participating in this study.

REFERENCES

34. Corrigan L. The effect of unilateral load carriage on the muscle activities of the trunk and lower limbs of young healthy males during gait: University of Ottawa (Canada); 2012. http://dx.doi.org/10.20381/ruor-6220
38. Stenholt, S., Tiainen, K., Rantanen, T., Sainio, P., Heliovaara, M., Impivaara, O., & Koskinen, S. Long - term determinants of muscle strength decline: Prospective evidence from the 22-year mini -


