Original Article

JHSMR Journal of Health Science and Medical Research

Test-Retest Reliability of Shoulder Function Capacity using the EvalTech System

Worawan Ekabutr, B.Sc.[1], Sutida Sakunkaruna, M.Sc.[1], Siriwan Yotsungnoen, B.Sc.[1], Prasert Sakulsriprasert, Ph.D.[1], Jim Richards, Ph.D.[2], Thanayos Sakunkaruna, Ph.D.[1]

[1]Faculty of Physical Therapy, Mahidol University, Phutthamonthon, Nakhon Pathom 73170, Thailand.

[2]Allied Health Research Unit, University of Central Lancashire, Preston, Lancashire, PR1 2HE, United Kingdom.

Received 3 April 2025 ● Revised 14 January 2026 ● Accepted 31 January 2026 ● Published online 18 September 2026

Abstract:

Objective: The EvalTech system is often used to evaluate an individual's functional shoulder performance in a work-related manner. However, there is a lack of information on the reliability of functional capacity evaluations (FCE) performed using the EvalTech system.

Material and Methods: This study considered the test-retest reliability of FCE for shoulder function in the push and pull and time measurement tests in 31 healthy young adults. Reliability was estimated using intraclass Correlation Coefficients ($ICC_{2,1}$).

Results: The shoulder level push test had good to excellent reliability (ICC=0.88-0.93), the shoulder level pull test showed good reliability (ICC=0.81-0.83), the cart level push showed excellent reliability (ICC=0.95-0.97), and the cart level pull test showed excellent reliability (ICC=0.90-092). However, the combination of time measurement and overhead reaching tests showed only moderate reliability (ICC=0.66-0.72).

Conclusion: This indicates that the EvalTech system has good to excellent reliability in the push and pull tests, and moderate reliability in time measurement tests. This data should be considered by assessors when interpreting changes due to interventions and return-to-work assessments.

Impact Statement: This may have relevance to physical therapists performing FCE and supports the wider use of the EvalTech system, as this offers greater reliability and shorter assessment times than alternative systems when considering return-to-work assessments.

Keywords: EvalTech, functional capacity evaluation, pushing, pulling, time measurement test

Contact: Thanayos Sakunkaruna, Ph.D.
Faculty of Physical Therapy, Mahidol University, Phutthamonthon,
Nakhon Pathom 73170, Thailand.
E-mail: thanayos.sak@mahidol.edu

J Health Sci Med Res
doi: 10.31584/jhsmr.20261427
www.jhsmr.org

Introduction

Musculoskeletal conditions associated with work-related injuries are increasing globally. For example, 44% of the general working population in the United Kingdom reported shoulder injuries and neck disorders[1], 37.5% of workers at an electronic waste recycling site in Africa also reported upper extremity injury, and lower back pain is increasing in the poultry industry in Asia[2,3]. Several work-related tasks, including tool handling, heavy lifting, and overhead reaching, have been shown to lead to musculoskeletal disorders[4]. Musculoskeletal injuries have been associated with inappropriate ergonomics and incorrect posture while lifting or moving heavy objects, with pushing and pulling tasks being shown to be associated with a high prevalence of injuries in industrial settings[5,6]. For this reason, injured workers often visit clinicians for rehabilitation and to determine return-to-work criteria. Adjusting the posture of workers has been shown to reduce the rate of work-related injuries[7]. Functional capacity evaluation (FCE) systems were commonly used to evaluate the progression and prognosis of injured workers, before, during, and after physical therapy rehabilitation, and to assess individuals' performance and safety. The systems can be used to optimize work abilities and rehabilitation training to improve their functional and muscle strength. Ergonomically, if workers can improve their performance, which is related to their work capacity, this may reduce work-related injuries, leading to a decline in stress at work, leave rate, and medical costs[8]. Thus, the validity and reliability of the mean assessment values are required when performing FCE.

The EvalTech System is commonly used for FCE in the United States[9]. Furthermore, it is used to evaluate an individual's performance and to inform return-to-work activities. The EvalTech System is used to perform standardized assessments such as static muscle strength using load cells and range of motion using single and dual-axis goniometers[9]. The Evaltech comprises functional strength tests related to work activities: pushing, pulling, gripping, and lifting. It can also be used to assess the time to complete several tasks that simulate several working positions, such as standing, stooping, and squatting.[9] The Evaltech system comprises vertical peg boards, which are frequently used for FCE to determine upper extremity dexterity, as reported using time measurement tests. Moreover, there were other forms of pegboard tests, such as the Purdue pegboard test, the 9-hole pegboard test, and the Groove pegboard test[10,11]. Previous studies have shown that the Purdue pegboard test has good to excellent reliability, and the 9-hole pegboard test has moderate reliability. Clinically, physical therapists often assess functional performance in a variety of ways, including goniometry, manual muscle testing, and grip strength. However, work-related functional evaluation systems for industrial workers were also required. There were a few systems that have been used to assess work-related FCE, including the Isernhagen Work Systems FCE (The WorkWell Systems, WWE FCE), which is the most commonly used in Europe and has been reported to have a high test-retest reliability, assessed using intraclass correlation coefficients $\geq$0.75, percentages of agreement $\geq$80% and kappa coefficients $\geq$0.60[12]. However, despite its wide usage, the test-retest reliability of the EvalTech system has not yet been reported. Therefore, this study evaluated the test–retest reliability of the EvalTech system during push-and-pull tasks at the shoulder and cart level, and time measurement tests during overhead reaching, measured with the pegboard test in healthy participants.

Material and Methods

A convenience sample of 31 healthy young adults aged between 18 and 23 years without symptoms of musculoskeletal, neurological, or cardiovascular conditions, and no underlying disease affecting full physical exertion. Individuals who had a limited range of motion in the shoulders, elbows, wrists, or fingers were excluded. This study was approved by the Mahidol University Central Institutional Review Board (MU-CIRB COA. NO. 2019/146.1309). Informed consent was obtained from all the participants before the study commenced.

The EvalTech functional testing system (BTE technology, Baltimore, Maryland, USA) was used for the assessment of pushing and pulling forces and time to complete tasks. The device was calibrated before testing. The pushing and pulling forces were measured by the Universal Task Master (UTM) of EvalTech (BTE), and the time measurements were recorded from the pegboard positional test. Assessments were conducted on day 1 and then repeated approximately 48 hours later (Figure 1).

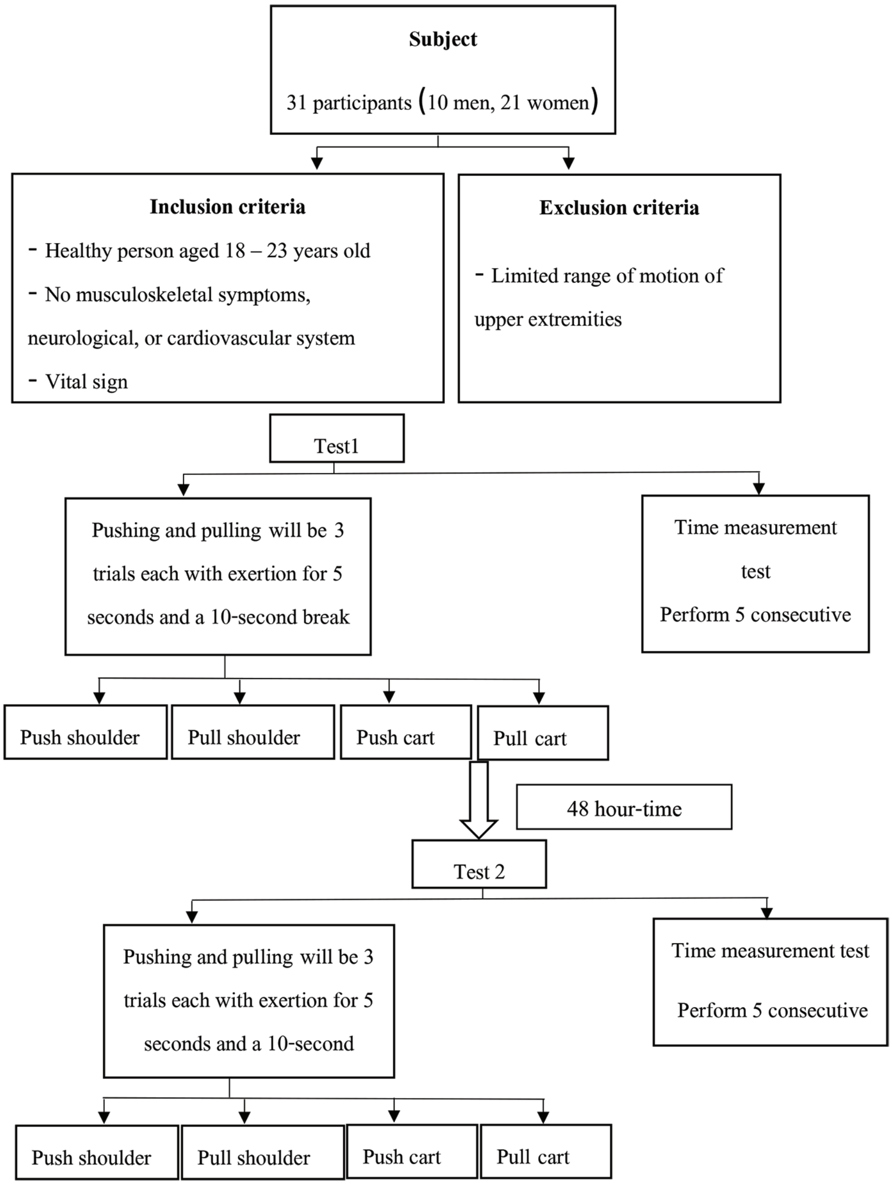

Figure 1 Consort diagram

Evaluation of function

The push and pull strength tests were performed at the shoulder and cart levels. For the shoulder level, the bar height was set to the same level as the acromion process and to just above the Iliac crest for the cart level for each participant, with the width of the handles set to the participant's shoulder width. The participants grasped the middle of the handles with their wrists in a neutral position and their arms parallel to the ground. The elbows were fully extended in the pushing test and the pulling test. Participants were in a step-standing position with their feet placed flat on the floor (Figure 2).

The participants were allowed to practice the pushing and pulling tasks before starting the tests. Each participant performed the tests in the same order: pushing test at shoulder level, pulling test at shoulder level, pushing test at cart level, and pulling test at the cart level. In each test, the participants were asked to perform three trials by exerting the maximum force for five seconds with a 10-second break between trials, and a 60-second rest interval between each test. For each test, the foot position was marked to ensure consistency on both test days, and the same researcher (TS) gave all participants the same instructions and commands.

Figure 2 Measuring the efficacy of a) pushing at shoulder level and b) pulling at shoulder level, and c) pushing at cart level and d) pulling at cart level using the universal task master (UTM), EvalTech (BTE)

The recommended test protocol by EvalTech includes an overhead reaching positioning task and a time measurement test. Each participant was asked to stand in front of the BTE FCE pegboard (Figure 3), and they were then asked to extend their right arm with fingers extended as high as they could while both feet touched the ground. The highest row that could be reached by the middle finger was recorded. The participants then moved a total of 5 rows of pegs from Panel 2 zone A (green box) to Panel 3 zone A (red box), Figure 3. To move the 15 pegs, they used their left hand to remove the pegs and pass them to the right hand, and then used the right hand to place the pegs in the new position, starting at the highest row first. When all 15 pegs were moved, all the pegs were then moved back to their starting positions. This was repeated a total of 5 times, and the time to complete the entire task was recorded (Figure 4).

Data analysis

The data were analyzed using the Statistical Package for the Social Sciences (SPSS version 23; IBM, New York, USA). Descriptive statistics included the mean and standard deviation. Test-retest reliability was determined using the intraclass correlation coefficient (ICC 2,1) with 95% confidence intervals. The criteria for determining reliability were 0.00–0.50, indicating poor reliability, 0.50–0.74, indicating moderate reliability, 0.75–0.90, indicating good reliability, and 0.90–1.00, indicating excellent reliability[13]. The standard error of measurement (SEM) was calculated using the $S.D._{\text{pooled variance}} \times$ square root of 1-ICC (23), the minimal detectable change (MDC) was calculated using 2.77×SEM (24), and the level of significance was set at p-value<0.05[13].

FCE=functional capacity evaluation

Figure 3 The BTE FCE pegboard

Figure 4 Measuring position with the overhead reaching test and time measurement test

Results

A convenience sample of 31 healthy young adults aged between 18 and 23 years (10 men and 21 women) was recruited with a body mass index (BMI) of 21.10±2.87 kg/m^2. The test-retest reliability results are shown in Table 1. The intraclass correlation coefficient (ICC 2,1) from the shoulder level push test, for individual and both hands, showed good to excellent reliability (ICC=0.88–0.93), SEM (3.56-10.54) and MDC (9.87-29.20 kg), with the shoulder level pull test showing good reliability (ICC=0.81-0.84), SEM (4.26–8.63) and MDC (11.81–23.90 kg). The cart level push test showed excellent reliability (ICC=0.95-0.97), SEM (2.75–4.38), and MDC (8.00–12.15 kg), with the cart level pull test showing moderate to excellent reliability (ICC=0.72-0.92), SEM (4.38–18.21), and MDC (12.14–50.46 kg). However, the time measurement test showed only moderate reliability (ICC=0.66), SEM (9.71), and MDC of 26.90 seconds.

Discussion

The test-retest reliability of the FCE performed using the EvalTech system showed excellent reliability for the functional push test at the shoulder level and cart level, while the functional pull test at the shoulder level and cart level showed good to excellent reliability, respectively. The time measurement test only showed moderate reliability. In addition, the EvalTech system showed better reliability than other evaluation systems, such as hip and knee strength assessed using a portable fixed dynamometry Evaluator (BTE technologies), which showed fair to high intrasession and intersession reliability, as previously reported in the literature[14]. Conversely, the test–retest reliability of the Isernhagen Work System (IWS), which assesses work-related activities, such as lifting, carrying, pushing, and pulling, was only moderate for static pushing (ICC=0.68) and was therefore considered unacceptable for functional

Table 1 Test-retest reliability ($ICC_{2,1}$) of the EvalTech functional capacity evaluation system (n=31)

Evaluation	Occasion 1 (Mean±S.D.)	Occasion 2 (Mean±S.D.)	SEM	MDC	$ICC_{2,1}$	95% CI	p-value
Shoulder level push test (kg)							
Left	41.07±15.65	39.6±14.44	3.56	9.87	0.93	0.87-0.97	<0.001
Right	50.2±14.46	48.15±16.24	4.01	11.11	0.93	0.87-0.97	<0.001
Combined	91.28±29.02	90.57±31.5	10.54	29.20	0.88	0.76-0.94	<0.001
Shoulder level pull test (kg)							
Left	42.13±10.6	38.99±9.84	4.26	11.81	0.81	0.61-0.90	<0.001
Right	53.19±14.16	50.1±14.23	5.58	15.46	0.84	0.68-0.92	<0.001
Combined	95.32±21.96	87.75±21.12	8.63	23.90	0.83	0.65-0.92	<0.001
Cart level push test (kg)							
Left	42.96±14.05	42.16±14.1	2.89	8.00	0.95	0.91-0.98	<0.001
Right	55.23±16.31	51.97±14.34	2.75	7.64	0.96	0.92-0.98	<0.001
Combined	98.19±29.21	95.65±26.71	4.38	12.15	0.97	0.94-0.98	<0.001
Cart level pull test (kg)							
Left	74.44±15.51	48.92±15.5	4.38	12.14	0.92	0.83-0.96	<0.001
Right	61.51±17.58	58.74±14.85	4.64	12.87	0.90	0.79-0.95	<0.001
Combined	113.67±33.64	110.79±34.43	18.21	50.46	0.72	0.41-0.86	<0.001
Time measurement test							
Time (seconds)	221.16±25.94	209.45±16.73	9.71	26.90	0.66	0.30-0.83	0.002

ICC=intraclass correlation coefficient, SEM=standard error of measurement=S.D.$\sqrt{(1\text{-}ICC)}$, MDC=minimal detectable change=2.77×SEM

capacity evaluation (FCE)[15]. In comparison, static pulling demonstrated good reliability (ICC=0.89) and was deemed acceptable for FCE purposes[15]. The discrepancy in reliability between pushing and pulling, relative to the findings of the present study, may be partially explained by the lack of standardization in foot placement and task height during pushing and pulling assessments in the IWS[15]. The present study demonstrated greater reliability than the Isernhagen Work System (IWS). This may be attributed, in part, to efforts to standardize foot positioning during assessment. Additionally, the EvalTech system includes only five tasks, comprising push and pull tests at the shoulder and cart levels and a timed measurement task, which may reduce variability. In contrast, the IWS consists of 28 tasks involving lifting, carrying, and bending, potentially introducing greater performance variability. Another important consideration is the duration of assessment, as the IWS may require up to two hours to complete, whereas the EvalTech assessment can be completed in less than 30 minutes; this difference may influence participants' ability to consistently repeat the tasks. Interestingly, assessment reliability was generally higher when the left and right sides were analyzed separately rather than combined.

In addition, the Worklab functional capacity evaluation was commonly used in Australian workplace injury management and has been shown to have excellent reliability (ICC=0.94-0.98)[16]. Similarly, we found good reliability of the pulling test at the shoulder level on the left (ICC=0.81) and right side (ICC=0.84). Besides, the reliability studies performed by Toonstra and Mattacola[17], which considered the reliability of portable fixed dynamometry (PFD), handheld dynamometry (HHD), and isokinetic dynamometry (IKD) machines to assess knee strength, showed that PFD and IKD were very reliable.

When considering the reliability of the time measurement test for the overhead reaching task showed moderate test-retest reliability (ICC=0.663), which is lower than the test–retest reliability of the Purdue Pegboard Test that was developed for occupational therapy assessment in the USA, and has been reported to have excellent reliability over five repetitions (ICC>0.80) but only poor to fair reliability when using one repetition[10]. Moreover, the Purdue Pegboard test was found to have high test-retest reliability in patients with carpal tunnel syndrome (ICC=0.91)[18]. The Purdue Pegboard Test showed similar reliability to our study in the time measurement tests, indicating measures of dexterity of hand function; however, the Purdue Pegboard Test assesses the time to perform tasks on a table, whereas the EvalTech assesses at the overhead level, which may be more pertinent to the assessment of back, shoulder, and neck injuries. Our study found that the time measurement test had moderate reliability, which may have been caused by the time interval between the two evaluations and the complexity of the test. Greater time intervals between tests could reduce recall bias[18-20]. For this reason, the authors suggest that in order to enhance accuracy and reliability, the tester should give standardized instructions and permit more time for practice before the trial.

The limitation of this study: we only tested reliability in young healthy participants aged between 18 and 23 years. Therefore, this finding may not be directly applicable to the clinical population and elder workers. Further studies should be conducted to determine the test-retest reliability of workers with musculoskeletal conditions. Additionally, the pegboard test was a complex task. Further work should explore the effect of standardized instructions and learning and practice between sessions before testing to determine whether test-retest reliability can be improved.

Practical implication

Health professionals, including physical therapists, rehabilitation physicians, and nurses, can use the EvalTech system as an assessment tool and for rehabilitation planning to evaluate functional abilities, such as push–pull performance at the shoulder and cart levels, as well as work-simulation capacity.

Conclusion

This study showed that the EvalTech system had good reliability in the pushing and pulling tests, and moderate reliability in the time measurement tests. The EvalTech system offers greater reliability and shorter assessment times than alternative systems when considering return-to-work assessments, and this should be considered by assessors when interpreting changes due to interventions and return-to-work assessments.

Acknowledgement

The authors would like to thank all the participants in this study. We would also like to thank the Faculty of Physical Therapy, Mahidol University, for providing the data collection venues. Special thanks to Dr. James Richards for providing manuscript revisions and suggestions.

Funding sources

This work was supported by research funding from the Faculty of Physical Therapy, Mahidol University, Thailand.

Conflict of interest

None.

References

1. Executive HaS. Work related musculoskeletal disorder statistics (WRMSDs) in Great Britain, 2020 [monograph on the Internet].

Great Britain: Health and Safety Executive; [cited 2021 May 14]. Available from: https://www.hse.gov.uk/statistics/causdis/msd.pdf

2. Acquah AA, D'Souza C, Martin BJ, Arko-Mensah J, Dwomoh D, Nti AAA, et al. Musculoskeletal Disorder Symptoms among Workers at an Informal Electronic-Waste Recycling Site in Agbogbloshie, Ghana. Int J Environ Res Public Health 2021;18:2055.
3. Laorujisawat M. Ergonomic risk factors and guidance on the management of work-related musculoskeletal disorders (WMSDs) in the poultry industry. Thai Journal of Ergonomics 2020;3:65-75.
4. Renberg J, Wiggen ØN, Tvetene PØS, Færevik H, Van Beekvelt M, Roeleveld K. Effect of working position and cold environment on muscle activation level and fatigue in the upper limb during manual work tasks. Int J Ind Ergon 2020;80:103035.
5. Pillastrini P, Mugnai R, Bertozzi L, Costi S, Curti S, Guccione A, et al. Effectiveness of an ergonomic intervention on work-related posture and low back pain in video display terminal operators: a 3 year cross-over trial. Appl Ergon 2010;41:436-43.
6. Acquah AA, D'Souza C, Martin B, Arko-Mensah J, Nti AA, Kwarteng L, et al., editors. Processes and challenges associated with informal electronic waste recycling at Agbogbloshie, a suburb of Accra, Ghana. Proceedings of the Human Factors and Ergonomics Society Annual Meeting; 2019: SAGE Publications Sage CA: Los Angeles, CA.
7. Ohnishi A, Takanokura M, Sugama A. Evaluation of interhandle distance during pushing and pulling of a four-caster cart for upper limb exertion. Saf Health Work 2016;7:237-43.
8. Buddenberg LA, Davis C. Test-retest reliability of the Purdue pegboard Test. Am J Occup Ther 2000;54:555-8.
9. Proud EL, Bilney B, Miller KJ, Morris ME, McGinley JL. Measuring hand dexterity in people with Parkinson's disease: reliability of pegboard tests. Am J Occup Ther 2019;73:7304205050p1-p8.
10. Chen JJ. Functional capacity evaluation & disability. Iowa Orthop J 2007;27:121.
11. Bieniek S, Bethge M. The reliability of WorkWell systems functional capacity evaluation: a systematic review. BMC Musculoskelet Disord 2014;15:1-13.
12. BTE Technologies. EvalTech: full-suite objective evaluation and testing system [homepage on the Internet]. Hanover (MD): BTE Technologies; [cited 2026 May 14]. Available from: https://www.btetechnologies.com/products/evaluationtesting/evaltech/
13. Portney LG, Watkins MP. Foundations of clinical research: applications to practice: Pearson/Prentice Hall Upper Saddle River, NJ; 2009.
14. Kollock Jr RO, Onate JA, Van Lunen B. The reliability of portable fixed dynamometry during hip and knee strength assessments. J Athl Train 2010;45:349-56.
15. Reneman M, Brouwer S, Meinema A, Dijkstra P, Geertzen J, Groothoff J. Test-retest reliability of the Isernhagen work systems functional capacity evaluation in healthy adults. J Occup Rehabil 2004;14:295-305.
16. James C, Mackenzie L, Capra M. Inter-and intra-rater reliability of the manual handling component of the WorkHab Functional Capacity Evaluation. Disabil Rehabil 2011;33:1797-804.
17. De Baets S, Calders P, Schalley N, Vermeulen K, Vertriest S, Van Peteghem L, et al. Updating the evidence on functional capacity evaluation methods: a systematic review. J Occup Rehabil 2018;28:418-28.
18. Amirjani N, Ashworth NL, Olson JL, Morhart M, Chan KM. Validity and reliability of the Purdue pegboard Test in carpal tunnel syndrome. Muscle Nerve 2011;43:171-7.
19. Zetterberg C, Heiden M, Lindberg P, Nylén P, Hemphälä H. Reliability of a new risk assessment method for visual ergonomics. Int J Ind Ergon 2019;72:71-9.
20. Gouttebarge V, Wind H, Kuijer PP, Sluiter JK, Frings-Dresen MH. Intra-and interrater reliability of the Ergo-Kit functional capacity evaluation method in adults without musculoskeletal complaints. Arch Phys Med Rehabil 2005;86:2354-60.