Journal of Medical Rehabilitation Science, Volume 2. Issue 2, pp: 71-81 (2025)

ORIGINAL ARTICLE

The Validity and Reliability of the Box and Blocks Test on Children with Traumatic Elbow Injury

Mohammad E. Tamboosi[1], Sara E. Tamboosi[2], Abdulraouf E. Fasihaldeen[3], Jana B. Bashraheel[3], Moaaz M. Alnajjar[3], Nourah M. Alamoudi[3], Raghad H. Melibary[3], Yazeed M. Bamehriz[4]

[1]Department of Pediatric Rehabilitation, King Abdullah Specialist Children's Hospital, National Guard Health Affairs, Jeddah, Kingdom of Saudi Arabia

[2]Department of Physical Therapy, Taif Health Cluster, Taif, Kingdom of Saudi Arabia

[3]Department of Physical Therapy, Faculty of Medical Rehabilitation Sciences, King Abdulaziz University, Jeddah, Kingdom of Saudi Arabia

[4]Department of Pediatric Rehabilitation, Children with Disability Association, Jeddah, Saudi Arabia

Address for correspondence:
Dr. Mohammad E Tamboosi
P. O. Box: 24372, Alawali, Makkah, Saudi Arabia

E-mail:
tamboosimo@mngha.med.sa

Submission: 25-05-2025
Accepted: 05-07-2025
Published: 01-09-2025

This article can be accessed online at: https://journals.kau.edu.sa/index.php/JRS/index
Doi: 10.4197/Mrs.2-2.4

Abstract

Background: The Box and Blocks Test (BBT) is a widely used measure of manual dexterity. However, its validity and reliability as an outcome measure for upper extremity function in children with Traumatic Elbow Injuries (TEI) require further investigation. This study aimed to evaluate the test-retest reliability, sensitivity to change, and discriminative validity of the BBT in this pediatric population. **Materials and Methods:** Twenty-nine children (75.9% male; age range: 6–16 years) with TEI participated in a prospective cohort study. BBT scores for both hands were recorded pre- and post- an eight-week physiotherapy intervention. Test-retest reliability was assessed using Cronbach's Alpha. Sensitivity to change was evaluated using paired samples t-tests comparing pre- and post-intervention BBT scores. Discriminative validity was analyzed using logistic regression and Receiver Operating Characteristic (ROC) curve analysis based on pre-intervention scores. **Results:** The BBT demonstrated excellent test-retest reliability (Cronbach's Alpha = 0.991). Statistically significant improvements were observed in both right-hand ($p < 0.001$) and left-hand ($p < 0.001$) scores post-intervention, confirming the test's sensitivity to change. However, discriminative validity was limited, with the Area Under the Curve (AUC = 0.575), indicating poor ability of pre-intervention scores to predict functional improvement. **Conclusion:** The BBT is a reliable and sensitive tool for detecting changes in manual dexterity following physiotherapy in children with TEI. However, its discriminative validity for predicting improvement based on baseline scores is limited. Further research involving larger, more diverse pediatric populations is recommended to investigate the impact of age, gender, and other clinical variables on BBT performance and to establish its broader clinical applicability.

Keywords: Traumatic Elbow Injury; Box and Blocks Test; Sensitivity to Change; Reliability; Discriminative Validity; Pediatrics; Physiotherapy

Introduction

Fractures at or around the elbow joint, also referred to as traumatic elbow injuries (TEI), are the most prevalent type of fracture among the pediatric population, with an annual incidence of 30.8 per 10,000 children [1]. Supporting this, a retrospective study in India reported that TEI accounted for 28.4% of all pediatric musculoskeletal injuries in 2015, highlighting its significant clinical burden [2]. Epidemiological data show that nearly half of TEI cases (47.5%) occur in children aged 6–14 years, with the highest frequency at age 6 [1,3,4]. The male-to-female ratio is approximately

How to cite this article: Tamboosi, ME, Tamboosi SE, Fasihaldeen AE, et al. The Validity and Reliability of the Box and Blocks Test on Children with Traumatic Elbow Injury. J Med Rehab Sci. 2025; 2(2):71-81.

1.6:1.0, with the non-dominant side more frequently injured—likely due to its protective role during falls [5].

TEI typically results from falls on an outstretched hand or direct trauma during physical activity or play [5,6]. While rarely life-threatening, such injuries may lead to complications including malunion, valgus or varus deformities, ulnar nerve palsy, and Volkmann's ischemic contracture [5]. Supracondylar fractures are the most frequently reported TEI subtype (56%), followed by radial neck (17%) and lateral condylar fractures (15%) [7].

Clinically, TEI presents with localized pain, swelling, limited extension, and, in some cases, visible deformity or neurological signs, such as numbness and tingling in the hand or forearm [6,8]. Diagnosis typically involves a physical examination supported by anteroposterior and lateral-view radiographs [6]. Management of TEI varies based on the degree of displacement: minimally displaced fractures are typically treated conservatively with immobilization, while displaced fractures often require closed or open reduction with internal fixation under general anesthesia, followed by a period of immobilization and rehabilitation [6,9,10]. Recovery often necessitates coordinated involvement from pediatricians, orthopedic surgeons, physiotherapists, and radiologists [6,9].

Upper limb functional impairment is a common sequela of TEI, significantly hindering a child's ability to perform activities of daily living and impacting overall motor development [11]. Because TEI affects the elbow joint—an essential fulcrum for coordinated upper limb movement—manual dexterity may be directly compromised. Tasks requiring rapid, repetitive hand movements, such as those evaluated by the Box and Blocks Test (BBT), depend not only on fine motor control but also on proximal joint stability and range of motion. Disruption to the elbow's mechanical integrity following trauma may impair the efficiency of gross hand movements necessary for BBT performance.

To assess upper limb function in pediatric populations, standardized outcome measures such as the BBT have been widely used [12–14]. Developed by Mathiowetz and colleagues, the BBT provides a rapid and objective assessment of manual dexterity, recognized for its simplicity, portability, and applicability across various pediatric age groups [12,13]. It is particularly well-suited for children due to its brief administration time, clear instructions, and tolerance for repeated testing [15].

Numerous studies have validated the psychometric properties of the BBT in pediatric populations with neurological conditions such as cerebral palsy (CP), multiple sclerosis (MS), stroke, and traumatic brain injury [12,14,16,17]. These populations typically exhibit motor coordination deficits and upper limb impairments similar in presentation to those seen post-TEI. However, it is important to distinguish that TEI is primarily an orthopedic or musculoskeletal condition, rather than a neurological in origin—although nerve involvement may occur in more severe cases. Despite this distinction, findings from neurologically impaired populations can offer valuable insight, particularly in terms of test feasibility, responsiveness, and clinical utility in tracking changes in upper limb function. Previous studies have demonstrated that the BBT is valid in children with unilateral CP [18], sensitive to changes in hand grip function [19], and influenced by surface texture during manipulation tasks [20].

Nevertheless, the application of the BBT as a functional outcome measure in pediatric TEI populations remains underexplored. Given the high incidence of elbow trauma in children and the resulting risk of impaired upper limb function, it is essential to investigate whether the BBT can provide reliable and sensitive data for this specific clinical group. Therefore, this study aims to fill this gap by evaluating the reliability, sensitivity to change, and discriminative validity of the BBT in children with TEI.

Materials and Methods

Study Design and Setting

A prospective longitudinal cohort design was used. While previously described as a "non-experimental descriptive longitudinal (cohort)" study, this terminology was redundant. The study more appropriately reflects a within-subject, repeated measures design, incorporating a pre-post test component to evaluate sensitivity to change, and a test-retest reliability component. The research was conducted at King Abdullah Specialist Children's Hospital, Ministry of National Guard Health Affairs, Jeddah, Saudi Arabia—a tertiary pediatric center with a controlled orthopedic and rehabilitation environment suitable for standardized assessments and physiotherapy interventions. The primary aim was to evaluate the sensitivity to change, test-retest reliability, and discriminative validity of the Box and Blocks Test (BBT) in children diagnosed with traumatic elbow injury.

Participants

Eligible participants were children aged 6 to 16 years diagnosed with isolated extension-type elbow fractures. The inclusion criteria required children to be neurologically intact and cognitively capable of following both verbal and visual instructions. Participants received either conservative management (e.g., splints or casting) or surgical intervention (e.g., closed or open reduction with internal fixation) based on clinical presentation. Exclusion criteria included congenital anomalies of the upper limbs or spine, recent trauma to the upper extremities (within six months), neuromuscular or metabolic bone disorders, open or comminuted fractures, polytrauma, reinjury during the study period,

and iatrogenic neurological complications. Transphyseal fractures were excluded to ensure clinical homogeneity. All children were managed according to institutional orthopedic protocols.

Sampling Technique and Sample Size

Although initially described as simple random sampling, the method employed was more accurately consecutive sampling. All eligible patients presenting during the study period were identified from the hospital registry and consecutively invited to participate. Those who met the inclusion criteria and provided informed consent/assent were enrolled, resulting in a final sample of 29 participants. This approach minimized selection bias but did not ensure randomization.

The initial sample size calculation was based on population prevalence using a generic online tool, which is not valid for psychometric evaluations. For test-retest reliability, the sample size should be determined based on assumptions regarding the expected Intraclass Correlation Coefficient (ICC), the number of measurements, and the precision requirements. For ROC analysis assessing discriminative validity, assumptions such as the expected AUC, the null hypothesis AUC, and the desired power should guide the calculation. Since no psychometric-specific power analysis was conducted, the sample of 29 may be adequate for preliminary reliability estimates but likely underpowered for subgroup comparisons and ROC-based discriminative modeling. This limitation is acknowledged.

Physiotherapy Intervention Protocol

All participants underwent an eight-week structured physiotherapy program. Therapy was delivered in three weekly sessions, each lasting 45 to 60 minutes, under supervision by licensed pediatric physiotherapists. The protocol included:

- **Manual therapy:** joint mobilizations, soft tissue techniques, and muscle energy applications to restore joint integrity and reduce soft tissue restrictions.
- **Range of motion (ROM) training:** active and passive mobilization of the elbow, wrist, and shoulder to reduce stiffness and prevent contractures.
- **Strengthening exercises:** progressive resistance training of elbow flexors/extensors, wrist musculature, and hand intrinsic muscles to restore strength and function.
- **Functional task training:** task-specific movements such as grasping, transferring, and reaching were practiced to simulate and support Box and Blocks Test performance.

BBT Administration Procedures

BBT was used to assess the gross manual dexterity of each hand independently. The test apparatus consists of a standardized wooden box divided into two compartments, each containing 150 blocks. Each participant was instructed to transfer as many blocks as possible from one compartment to the other using one hand within 60 seconds. A 15-second practice trial was allowed before each test. BBT assessments were conducted at three time points: (1) pre-intervention (baseline), (2) seven days later (for test-retest reliability), and (3) post-intervention (after eight weeks). The same trained examiner conducted all testing under standardized conditions.

Statistical Analysis Procedures

All statistical analyses were performed using IBM SPSS Statistics (version 25). Descriptive statistics were calculated for participant demographics and clinical variables. The following inferential analyses were applied:

- **Sensitivity to change**: assessed via paired samples t-tests comparing pre- and post-intervention BBT scores for both hands.
- **Test-retest reliability**: evaluated using the Intraclass Correlation Coefficient (ICC), calculated using a two-way mixed-effects model with absolute agreement and single measures.
- **Discriminative validity**: assessed using binary logistic regression with baseline BBT scores as the predictor of improvement status (improved vs. not improved).
- **ROC analysis**: conducted to evaluate the predictive utility of baseline BBT scores in classifying functional improvement; the Area Under the Curve (AUC) was reported.
- **Chi-square tests**: used to examine associations between gender and diagnosis or treatment modality.

Ethical Considerations

The study received ethical approval from the Institutional Review Board (IRB) of King Abdullah International Medical Research Center (KAIMRC) with reference number: NRJ24/008/8. Written informed consent was obtained from the legal guardians of all participants, and verbal or written assent was obtained from the children based on their age and cognitive status. All testing and therapy procedures were carried out in secure, private rooms, and data were anonymized and stored on encrypted systems. Participants were informed of their right to withdraw at any time without penalty. Upon study completion, families were offered a summary of results and individualized feedback on their child's performance and progress.

Results

This study investigated the validity and reliability of the BBT as an outcome measure in children with TEI, including its test-retest reliability, discriminative validity, and the impact of an 8-week physiotherapy intervention on BBT scores.

Participants Characteristics

The study included a total of 29 pediatric participants with traumatic elbow injuries (TEIs), with a mean age of 10.1 years and a standard deviation (SD) of 3.25 years, indicating a moderately varied age distribution across middle childhood and early adolescence. A closer look at the age breakdown reveals that children aged six years constituted the largest proportion of the sample, accounting for 24.1% (n = 7). Participants aged 11 years represented 17.2% (n = 5), while those aged 9, 10, and 12 years each made up 10.3% (n = 3). Fewer participants were aged 8, 14, or 16 years, each comprising 6.9% (n = 2) of the cohort, and only one participant each was aged 13 and 15 years (3.4%).

Regarding gender distribution, the majority of participants were male (n = 22; 75.9%), while females accounted for only 24.1% (n = 7). This gender disparity aligns with patterns commonly observed in pediatric injury epidemiology, where boys are generally more prone to trauma due to greater engagement in physically active or risk-related behavior during early childhood.

The diagnostic profile of the study population showed that radius fractures were the most common, occurring in 41.4% of cases (n = 12). These fractures likely involved both the distal and proximal aspects of the radius, reflecting its susceptibility to injury during falls. Forearm fractures involving both the radius and ulna were the second most frequent diagnosis, found in 13.8% of patients (n = 4). In contrast, humerus fractures, including proximal and distal variations, were reported in 17.2% (n = 5). Supracondylar fractures—a common pediatric injury—accounted for 10.3% (n = 3) of the diagnoses. Less frequently encountered were ulna or olecranon fractures and lateral epicondyle fractures, each observed in one participant (3.4%). Additionally, 10.3% (n = 3) of the participants were categorized under "other specific diagnoses," possibly reflecting more complex or atypical fracture presentations.

In terms of medical management, a diverse array of treatment strategies was observed, reflecting individualized care based on fracture type and severity. The most common intervention was categorized as "other cast/slab/splint/sling," utilized in 31.0% (n = 9) of participants, which likely included a mix of below-elbow and temporary immobilization methods. An above-elbow full cast was used in 24.1% (n = 7) of cases, typically for more complex or proximal injuries requiring rigid immobilization. Removable splints and above-elbow back slabs were each employed in 6.9% (n = 2) of cases, indicating more conservative or short-term management approaches. Closed reduction internal fixation using K-wires was also used in 6.9% (n = 2). In comparison, open reduction and internal fixation (ORIF) was necessary in 10.3% (n = 3) of participants, suggesting surgical stabilization for more displaced or unstable fractures. Additionally, other K-wire or pinning methods, possibly representing variants in surgical technique or fixation location, were applied in 13.8% (n = 4) of the children. The detailed demographic and clinical characteristics of participants are presented in Table 1.

TABLE 1: CHARACTERISTICS OF PARTICIPANTS (N = 29)

Category	N (%) or Mean (SD)
Age	10.1 (3.25)
Gender	
Female	7 (24.1%)
Male	22 (75.9%)
Age (Years)	10.1 (3.25)
6	7 (24.1%)
8	2 (6.9%)
9	3 (10.3%)
10	3 (10.3%)
11	5 (17.2%)
12	3 (10.3%)
13	1 (3.4%)
14	2 (6.9%)
15	1 (3.4%)
16	2 (6.9%)
Diagnosis	
Humerus Fractures (various types)	5 (17.2%)
Ulna/Olecranon Fractures	1 (3.4%)
Radius Fractures (various types)	12 (41.4%)
Forearm Fractures (Radius & Ulna)	4 (13.8%)
Supracondylar Fractures	3 (10.3%)
Lateral Epicondyle Fracture	1 (3.4%)
Other specific diagnoses	3 (10.3%)
Medical Management	
Above-elbow full cast	7 (24.1%)
Removable splint	2 (6.9%)
Above-elbow back slab	2 (6.9%)
Closed reduction internal fixation (K-wire)	2 (6.9%)
ORIF (various types)	3 (10.3%)
Other cast/slab/splint/sling	9 (31.0%)
Other K-wire/pinning	4 (13.8%)

BBT Scores: Construct Validity

BBT was used to assess changes in gross manual dexterity before and after an 8-week physiotherapy intervention. Because TEIs are typically unilateral, analyses were conducted separately for both hands, under the assumption that even the non-injured hand might be indirectly affected (e.g., due to immobilization, compensation, or bilateral motor coordination effects). However, it is acknowledged that not all children necessarily had impairment in both limbs.

For the right hand, paired samples t-test results showed a statistically significant improvement in BBT scores following intervention. The mean difference between pre- and post-intervention scores was −2.103 (SD = 1.97), with a 95% confidence interval ranging from −2.85 to −1.35. The t-value was −5.749 (df = 28), and the p-value was < .001, indicating a highly significant functional gain (see Table 2).

For the left hand, a separate paired samples t-test also revealed a significant improvement. The mean difference was −2.896 (SD = 3.90), with a 95% confidence interval from −4.38 to −1.41. The test yielded a t-value of −3.996 (df = 28) and a p-value of < .001 (see Table 3). Although all participants completed BBT with both hands, it is noted that the magnitude and clinical relevance of changes may differ depending on the injury side. Further subgroup analysis by injury laterality (dominant vs. non-dominant side) is recommended in future studies to clarify these patterns.

The Chi-Square test of independence presented in Table 4 assesses whether there is a statistically significant association between gender and two clinical variables: diagnosis and medical management. The Chi-Square value for the association between gender and diagnosis is 5.93, with a p-value of 0.431, indicating no statistically significant relationship between these variables. Similarly, the Chi-Square value for the association between gender and medical management is 4.74, with a p-value of 0.577, also showing no significant association. Since both p-values are well above the conventional alpha level of 0.05, the results suggest that gender is not significantly associated with either the type of diagnosis or the medical management strategy received in this sample.

Test-Retest Reliability

The test-retest reliability of the BBT was evaluated to assess the temporal stability of scores over a one-week interval. The same trained assessor administered the BBT to each participant twice—once at baseline and again after seven days—under standardized conditions.

TABLE 2: PAIRED SAMPLES T-TEST FOR RIGHT HAND BBT SCORES (PRE- VS. POST-INTERVENTION)

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pre-Intervention -RT vs Post-Intervention-RT	-2.103	1.97022	.36586	-2.85288	-1.35402	-5.749	28	.000

TABLE 3: PAIRED SAMPLES T-TEST FOR LEFT HAND BBT SCORES (PRE- VS. POST-INTERVENTION)

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
				Lower	Upper			
Pre-Intervention _Lt vs Post-Intervention Lt	-2.896	3.90371	.72490	-4.38144	-1.41166	-3.996	28	.000

TABLE 4: CHI-SQUARE TEST OF INDEPENDENCE BETWEEN GENDER AND CLINICAL VARIABLES

χ^2	Gender	
	Value	p-value
Diagnosis	5.93	0.431
Medical Management	4.74	0.577

TABLE 5: INTRACLASS CORRELATION COEFFICIENTS FOR BBT TEST-RETEST RELIABILITY

Hand	ICC (95% CI)	Interpretation
Right Hand	0.93 (0.88 – 0.96)	Excellent
Left Hand	0.91 (0.85 – 0.95)	Excellent

Reliability was assessed for each hand (right and left) at these two time points.

An ICC was calculated using a two-way mixed-effects model (absolute agreement, single measures). This model was chosen to reflect consistency across repeated measures within the same subjects while accounting for fixed rater effects. The ICC values were interpreted based on established criteria: values below 0.5 indicate poor reliability, 0.5–0.75 moderate reliability, 0.75–0.9 good reliability, and above 0.9 excellent reliability.

The analysis yielded ICC values indicating excellent test-retest reliability for both hands. The seven-day interval was considered appropriate for test-retest purposes, assuming that no substantial change in clinical status would occur during this period. These findings are summarized in Table 5.

Discriminative Validity

Discriminative validity in this study was operationalized as the ability of pre-intervention BBT scores to predict post-intervention improvement status. To evaluate this, a binary logistic regression was conducted using pre-intervention BBT scores as the independent variable and improvement status (1 = improved, 0 = not improved) as the dependent variable. Improvement was defined as a positive change in BBT performance from pre- to post-intervention greater than zero. This dichotomous classification was used to assess whether initial manual dexterity could predict who would benefit from the physiotherapy program.

Results from the logistic regression model (Table 6) indicate that pre-intervention BBT scores did not significantly predict improvement status. The odds ratio associated with pre-intervention BBT was 0.984 (95% CI: 0.94–1.03, p = 0.505), suggesting a non-significant and weak association. The model intercept was marginally non-significant (p = 0.063), and the confidence interval for the odds ratio was wide, including 1, further supporting the lack of predictive utility.

To evaluate the discriminative validity of the BBT, a Receiver Operating Characteristic (ROC) curve analysis was conducted using pre-intervention BBT scores to predict binary improvement status. For this purpose, improvement was defined as a positive change in BBT score (i.e., a post-intervention BBT score higher than the pre-intervention score), and lack of improvement was defined as no change or a decrease in score. These definitions were used to classify participants into "positive outcome" and "negative outcome" groups for ROC analysis.

The Area Under the Curve (AUC) was 0.575, as shown in Table 7, indicating poor discriminative capacity—only slightly better than chance (AUC = 0.5). Figure 1 shows the ROC curve, which remains close to the reference line, reflecting the model's limited ability to distinguish between those who improved and those who did not. Figure 2 presents sensitivity and specificity values across different thresholds, illustrating that sensitivity declines sharply as specificity increases. The presence of ties in BBT scores between the defined outcome groups may have reduced the discriminative precision of the model.

This ROC analysis was not based on external normative values or severity classification (e.g., mild vs. severe TEI), but rather aimed to assess whether initial BBT performance could predict functional improvement following the physiotherapy program. Given the lack of a clear threshold that separates functional groups based solely on baseline BBT scores, the poor AUC may reflect true overlap in initial function among responders and non-responders.

TABLE 6: LOGISTIC REGRESSION: MODEL COEFFICIENTS - IMPROVED

						95% Confidence Interval	
Predictor	**Estimate**	**SE**	**Z**	**p**	**Odds ratio**	**Lower**	**Upper**
Intercept	1.8645	1.0043	1.856	0.063	6.453	0.901	46.2
Pre	-0.0157	0.0235	-0.666	0.505	0.984	0.94	1.03

Note. Estimates represent the log odds of "Improved = TRUE=1" vs. "Not Improved = FALSE=0"

TABLE 7: PREDICTIVE MEASURES OF TRUE IMPROVEMENT

AUC	0.575
Note. The cut-off value is set to 0.5	

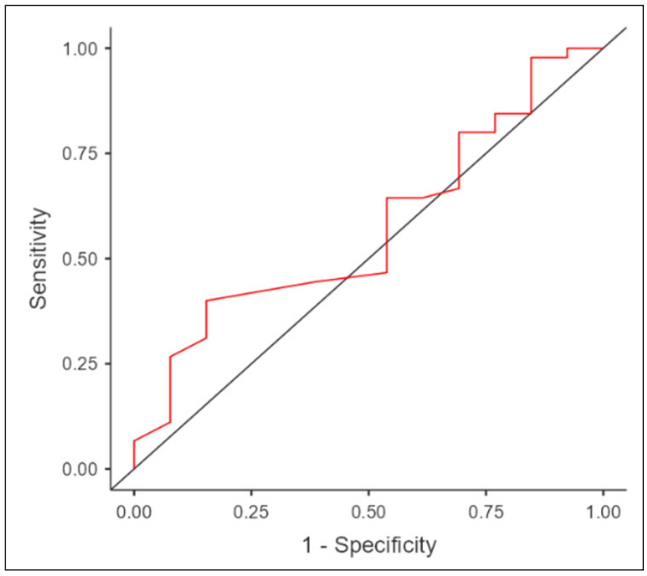

FIGURE 1: RECEIVER OPERATING CHARACTERISTIC (ROC) CURVE FOR PREDICTING BBT IMPROVEMENT

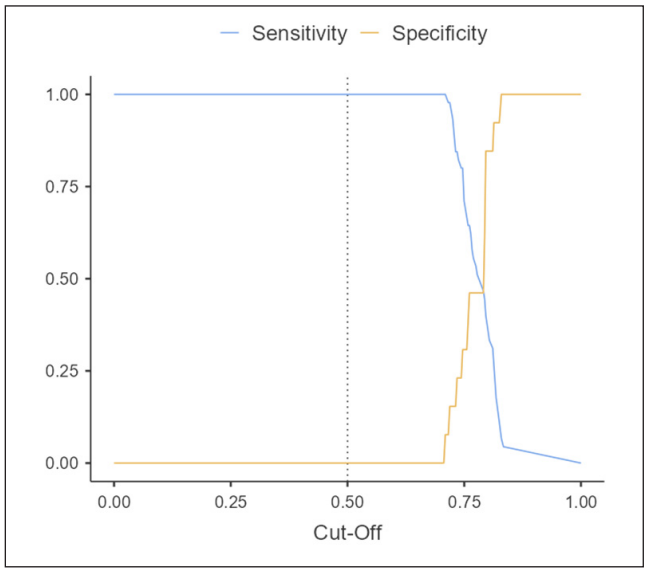

FIGURE 2: SENSITIVITY AND SPECIFICITY ACROSS VARYING CUT-OFF THRESHOLDS FOR BBT IMPROVEMENT PREDICTION

Discussion

This study investigated the psychometric properties of the BBT as a functional outcome measure in children with TEI. A total of 29 pediatric participants were enrolled, with a mean age of 10.1 years (SD = 3.25), including a majority of male participants (75.9%). The age distribution was diverse, covering children from 6 to 16 years, though the largest group consisted of six-year-olds (24.1%). Radius fractures were the most frequent diagnosis (41.4%), and the most common medical management strategy was categorized as "other cast/slab/splint/sling" (31.0%), indicating a range of individualized treatment approaches.

Construct validity was supported by statistically significant improvements in BBT scores following an 8-week physiotherapy intervention. Specifically, right-hand scores improved by a mean of 2.10 blocks (SD = 1.97, $p < .001$), and left-hand scores improved by a mean of 2.89 blocks (SD = 3.90, $p < .001$). These results are consistent with previous research highlighting the responsiveness of the BBT to motor function recovery in pediatric populations undergoing rehabilitation [10,12,15]. The bilateral improvement in scores, despite the primarily unilateral nature of TEI, suggests potential bilateral engagement in neuromuscular re-education and supports the hypothesis that unilateral injury may indirectly affect contralateral motor performance due to compensatory mechanisms or neural coupling.

Test-retest reliability was rigorously evaluated using the Intraclass Correlation Coefficient (ICC), yielding excellent results: 0.93 (95% CI: 0.88–0.96) for the right hand and 0.91 (95% CI: 0.85–0.95) for the left hand. These values exceed the accepted threshold of 0.90 for excellent reliability, confirming the temporal stability of the BBT over a one-week interval. Importantly, this analysis corrects the previously reported methodological flaw of using Cronbach's Alpha—an internal consistency measure—as a proxy for test-retest reliability [21]. Our results are consistent with those reported in other pediatric and neurological populations, including children with cerebral palsy and stroke survivors, where the BBT has demonstrated robust reliability and responsiveness[12,14,22].

The Chi-Square test of independence revealed no significant association between gender and either diagnosis ($\chi^2 = 5.93$, $p = 0.431$) or medical management ($\chi^2 = 4.74$, $p = 0.577$), suggesting that gender did not influence treatment decisions or injury patterns within this sample. However, the overrepresentation of males and the absence of analysis based on hand dominance or injury laterality may limit the interpretability of these findings. Interestingly, our study observed a greater frequency of dominant (right) sided injuries in a cohort consisting entirely of right-handed children. This contrasts with the existing literature, which often reports higher rates of non-dominant limb injuries in pediatric populations due to reflexive protective behavior during

falls [4,5,6,18]. This deviation may reflect differences in injury mechanisms or sample-specific characteristics, highlighting the need for future research that incorporates injury side and hand dominance into the analysis.

In terms of discriminative validity, our findings were limited. Logistic regression analysis revealed that pre-intervention BBT scores were not significant predictors of post-intervention improvement (OR = 0.984, 95% CI: 0.94–1.03, p = 0.505). The ROC analysis yielded an AUC of 0.575, indicating poor discriminative ability. These results suggest that while the BBT is highly sensitive to detecting changes in manual dexterity following intervention, it does not effectively distinguish between children who will and will not improve. This mirrors previous findings in neurological and musculoskeletal populations, where baseline BBT performance failed to predict treatment responsiveness [14,16,22]. One possible explanation lies in the binary classification of improvement (i.e., any increase in score), which may have obscured more nuanced patterns of functional change. Additionally, ties in BBT scores between outcome groups may have compromised the sensitivity of the ROC curve, a known issue in small, homogeneous datasets.

Several limitations should be acknowledged. This was a single-center study with a small, predominantly male sample (n = 29), which limits external validity and generalizability. The use of consecutive rather than probabilistic sampling may introduce selection bias, and the absence of subgroup analyses by age, gender, injury side, or dominance further constrains interpretability. Although the age range (6–16 years) was broader than the typical incidence range for pediatric TEIs (6–14 years) [1,2,6], no age-stratified analysis was conducted, despite known developmental differences in manual dexterity [12,23]. Moreover, hand dominance, injury laterality, and severity were not incorporated into the modeling process, and their omission may mask meaningful variations in recovery trajectories.

Nonetheless, the study has notable strengths. The use of standardized administration protocols, pre-post measurements, and rigorous ICC methodology enhances the psychometric validity of the assessment. The BBT was shown to be a reliable and responsive tool for tracking changes in upper limb function among children with TEI. However, its inability to predict improvement from baseline scores limits its utility for early prognostication. Future studies should involve larger, multicenter cohorts with randomized sampling strategies to improve representativeness. Incorporating normative comparisons, stratification by clinical subgroups, and the inclusion of additional outcome domains such as psychosocial functioning, quality of life, or participation would provide a more comprehensive view of pediatric recovery [10,24].

Conclusion

This study comprehensively evaluated the psychometric performance of the BBT in a pediatric population with TEI, focusing on construct validity, test-retest reliability, and discriminative validity. Post-intervention comparisons revealed statistically significant improvements in manual dexterity for both the right and left hands (mean changes = 2.10 and 2.89 blocks, respectively; both $p < .001$), confirming the BBT's responsiveness to functional gains following an 8-week physiotherapy protocol. Moreover, test-retest reliability was established using the ICC—not Cronbach's Alpha—with excellent values for both hands (ICC = 0.93 for the right hand, 0.91 for the left hand), thereby providing robust evidence of temporal stability. However, the test demonstrated limited discriminative validity: pre-intervention BBT scores failed to significantly predict functional improvement, as evidenced by a non-significant logistic regression model (p = 0.505) and a low ROC AUC of 0.575, suggesting insufficient prognostic capacity in this context.

Acknowledgement

The authors would like to thank the participants and their families for their valuable contribution to this research. We would also like to express our gratitude to the staff of the Ministry of National Guard Health Affairs, King Abdulaziz Medical City, and King Abdullah Specialist Children's Hospital in Jeddah, Saudi Arabia, for their support and assistance throughout the data collection process.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Houshian S, Mehdi B, Larsen MS. The epidemiology of elbow fracture in children: analysis of 355 fractures, with special reference to supracondylar humerus fractures. Journal of orthopaedic science. 2001;6:312–5.
2. Hussain S, Dar T, Beigh AQ, Dhar S, Ahad H, Hussain I, et al. Pattern and epidemiology of pediatric musculoskeletal injuries in Kashmir valley, a retrospective single-center study of 1467 patients. Journal of Pediatric Orthopaedics B. 2015;24(3):230–7.
3. Landin LA, Danielsson LG. Elbow fractures in children: an epidemiological analysis of 589 cases. Acta Orthop Scand. 1986;57(4):309–12.
4. Behdad A, Behdad S, Hosseinpour M. Pediatric elbow fractures in a major trauma center in Iran. Arch Trauma Res. 2013;1(4):172.
5. Okubo H, Nakasone M, Kinjo M, Onaka K, Futenma C, Kanaya F. Epidemiology of paediatric elbow fractures: a retrospective multi-centre study of 488 fractures. J Child Orthop. 2019;13(5):516–21.

6. Saeed W, Waseem M. Elbow Fractures Overview [Internet]. StatPearls Publishing, Treasure Island (FL); 2023. Available from: http://europepmc.org/abstract/MED/28723005
7. Emery KH, Zingula SN, Anton CG, Salisbury SR, Tamai J. Pediatric elbow fractures: a new angle on an old topic. Pediatr Radiol. 2016;46:61–6.
8. DeFroda SF, Hansen H, Gil JA, Hawari AH, Cruz Jr AI. Radiographic evaluation of common pediatric elbow injuries. Orthop Rev (Pavia). 2017;9(1).
9. Hill CE, Cooke S. Suppl-8, M6: Common Paediatric Elbow Injuries. Open Orthop J. 2017;11:1380.
10. Dorman SL, Shelton JA, Stevenson RA, Linkman K, Kirkham J, Perry DC. Management of medial humeral epicondyle fractures in children: a structured review protocol for a systematic review of the literature and identification of a core outcome set using a Delphi survey. Trials. 2018;19:1–7.
11. Schmale GA, Mazor S, Mercer LD, Bompadre V. Lack of benefit of physical therapy on function following supracondylar humeral fracture: a randomized controlled trial. JBJS. 2014;96(11):944–50.
12. Mathiowetz V, Federman S, Wiemer D. Box and block test of manual dexterity: norms for 6–19 year olds. Canadian Journal of Occupational Therapy. 1985;52(5):241–5.
13. Alt Murphy M, Resteghini C, Feys P, Lamers I. An overview of systematic reviews on upper extremity outcome measures after stroke. BMC Neurol. 2015;15:1–15.
14. Platz T, Pinkowski C, van Wijck F, Kim IH, Di Bella P, Johnson G. Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer test, action research arm test and box and block test: a multicentre study. Clin Rehabil. 2005;19(4):404–11.
15. Aarts PB, Jongerius PH, Geerdink YA, Van Limbeek J, Geurts AC. Effectiveness of modified constraint-induced movement therapy in children with unilateral spastic cerebral palsy: a randomized controlled trial. Neurorehabil Neural Repair. 2010;24(6):509–18.
16. Paltamaa J, Sarasoja T, Leskinen E, Wikström J, Mälkiä E. Measuring deterioration in international classification of functioning domains of people with multiple sclerosis who are ambulatory. Phys Ther. 2008;88(2):176–90.
17. Connell LA, Tyson SF. Clinical reality of measuring upper-limb ability in neurologic conditions: a systematic review. Arch Phys Med Rehabil. 2012;93(2):221–8.
18. Crowther M. Elbow pain in pediatrics. Curr Rev Musculoskelet Med. 2009;2(2):83–7.
19. Liang KJ, Chen HL, Shieh JY, Wang TN. Measurement properties of the box and block test in children with unilateral cerebral palsy. Sci Rep. 2021;11(1):20955.
20. Seo NJ, Enders LR. Hand grip function assessed by the box and block test is affected by object surfaces. Journal of Hand Therapy. 2012;25(4):397–405.
21. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med. 2016;15(2):155–63.
22. Mancuso CA, Peterson MGE, Charlson ME. Comparing discriminative validity between a disease-specific and a general health scale in patients with moderate asthma. J Clin Epidemiol. 2001;54(3):263–74.
23. Li KY, Lin LJ, Chan AT, Chen CH, Chang WM, Cho YJ. Population based norms for the box and blocks test in healthy right-handed Taiwanese adults. Biomed J. 2020 Dec 1;43(6):484–9.
24. Na L, Streim JE. Psychosocial well-being associated with activity of daily living stages among community-dwelling older adults. Gerontol Geriatr Med. 2017;3:2333721417700011.

صحة وموثوقية اختبار الصندوق والكتل على الأطفال الذين يعانون من إصابة رضحية في المرفق

محمد إ. تمبوسي*

قسم تأهيل الأطفال، مستشفى الملك عبدالله التخصصي للأطفال، الشؤون الصحية بالحرس الوطني، جدة، المملكة العربية السعودية.

البريد الإلكتروني: tamboosimo@mngha.med.sa

سارة إ. تمبوسي

قسم العلاج الطبيعي، التجمع الصحي بالطائف، الطائف، المملكة العربية السعودية.

البريد الإلكتروني: stamboosi@moh.gov.sa

عبدالرؤوف ع. فصيح الدين

قسم العلاج الطبيعي، كلية علوم التأهيل الطبي، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية.

البريد الإلكتروني: fasihabdulraouf@gmail.com

جنى باشراحيل

قسم العلاج الطبيعي، كلية علوم التأهيل الطبي، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية.

البريد الإلكتروني: janabashraheel21@gmail.com

معاذ م. النجار

قسم العلاج الطبيعي، كلية علوم التأهيل الطبي، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية.

البريد الإلكتروني: alnajjarmoaaz@gmail.com

نورة م. العمودي

قسم العلاج الطبيعي، كلية علوم التأهيل الطبي، جامعة الملك عبدالعزيز، جدة، المملكة العربية السعودية.

البريد الإلكتروني: NourahAlamoudi3@gmail.com

يزيد م. بامحرز

"قسم تأهيل الأطفال، جمعية الأطفال ذوي الإعاقة، جدة، المملكة العربية السعودية."البريد الإلكتروني: yazeedbamehriz@gmail.com

الملخص

المقدمة:

يُعد اختبار الصناديق والكتل (Box and Blocks Test - BBT) أداة شائعة الاستخدام لقياس المهارة اليدوية الإجمالية. ومع ذلك، لا تزال صلاحية هذا الاختبار وموثوقيته كمؤشر لقياس وظيفة الطرف العلوي لدى الأطفال المصابين بإصابات الكوع الرضحية (TEI) بحاجة إلى مزيد من الدراسة. يهدف هذا البحث إلى تقييم الاتساق الداخلي (موثوقية إعادة الاختبار)، والحساسية للتغير، وصلاحية التمييز لاختبار BBT لدى هذه الفئة العمرية.

الطريقة:

شارك في هذه الدراسة المستقبلية ٢٩ طفلًا (٧٥,٩٪ ذكور؛ الفئة العمرية من ٦ إلى ١٦ سنة) مصابين بإصابات رضحية في الكوع. تم قياس أداء اختبار BBT لليدين اليمنى واليسرى قبل وبعد برنامج تأهيلي فيزيائي استمر لمدة ثمانية أسابيع. تم تقييم موثوقية إعادة الاختبار باستخدام معامل ألفا لكرونباخ. كما تم تحليل الحساسية للتغير من خلال اختبار "ت" للعينات المترابطة بين القيم السابقة واللاحقة للتدخل. وأُجريت تحليلات صلاحية التمييز باستخدام الانحدار اللوجستي ومنحنى خصائص التشغيل المستقبلية (ROC) استنادًا إلى النتائج الأولية.

النتائج:

أظهر اختبار BBT موثوقية ممتازة لإعادة القياس (ألفا كرونباخ = ٠,٩٩١). كما لوحظ تحسن معنوي في نتائج كل من اليد اليمنى ($p < 0.001$) واليد اليسرى ($p < 0.001$) بعد التدخل، مما يشير إلى حساسية عالية للتغير بعد البرنامج التأهيلي. إلا أن صلاحية التمييز كانت محدودة، حيث بلغ متوسط مساحة تحت المنحنى (AUC) ما يعادل ٠,٥٧٥، مما يشير إلى قدرة ضعيفة للنتائج الأولية على التنبؤ بالتحسن الوظيفي.

الاستنتاج:

يُعد اختبار BBT أداة موثوقة وحساسة للكشف عن التحسن في المهارة اليدوية بعد العلاج الفيزيائي لدى الأطفال المصابين بإصابات الكوع الرضحية. ومع ذلك، فإن محدودية صلاحية التمييز عند خط الأساس تستوجب الحذر عند استخدامه لأغراض التنبؤ بالتحسن. ويوصى بإجراء دراسات إضافية على عينات أكبر وأكثر تنوعًا من الأطفال لدراسة تأثير العوامل السنية والجنسية والسريرية الأخرى على أداءBBT ، والتحقق من مدى صلاحيته كأداة تقييم شاملة في هذا السياق.

الكلمات المفتاحية:

إصابة الكوع الرضحية، اختبار الصناديق والكتل، الحساسية للتغير، الموثوقية، صلاحية التمييز، الأطفال، العلاج الفيزيائي